JOURNAL OF TRANSPORTATION
of the Institute of Transportation Engineers

VOLUME I, ISSUE I

Advancing transportation knowledge and practices for the benefit of society

EDITORIAL BOARD

ASSOCIATE EDITORS

Preface

Welcome to the inaugural edition of the *Journal of Transportation of the Institute of Transportation Engineers*. This Journal was conceived to provide the transportation community with the results of research in the areas of planning, operations, design, system management and related policy. It also provides a venue for academicians and practicing professionals to share research results. All papers submitted to the Journal are subject to an extensive peer review process prior to acceptance for publication.

We are heartened by the response to our call for papers, as we received close to 40 submissions. We express our sincere appreciation to the Senior Editor and the Associate Senior Editor for guiding the review process and selecting the four papers that appear in this edition. A group of highly qualified Associate Editors read and rated each of the submitted papers. The highly-rated papers that were not included in this issue will be considered for future editions. I thank all the authors who responded to the call for papers, and encourage them and all others to consider submitting for future editions. Please watch for future calls-for-papers.

We will be sharing this and future editions of the *Journal of Transportation of the Institute of Transportation Engineers* with key transportation organizations around the world and we are hopeful that we will be able to share papers that reflect international experiences in future editions. ITE's core purpose is to advance transportation knowledge and practices for the benefit of society. This is one more way that ITE is facilitating the exchange of ideas to accomplish that purpose.

Robert C. Wunderlich, P.E. (F)
International President of ITE

JOURNAL OF TRANSPORTATION
of the Institute of Transportation Engineers

VOLUME I | NUMBER I | MARCH 2011

CONTENTS

INSTITUTE OF TRANSPORTATION ENGINEERS 1627 Eye Street, NW, Suite 600, Washington, DC 20006 USA
P: 202.785.0060 **F:** 202.785.0609 **www.ite.org**

Bias and Uncertainty in Traffic Sign Retroreflectivity

By Stephen M. Remias, Sarah M. L. Hubbard, Eric A. Hulme, E.I., Alexander M. Hainen, Grant D. Farnsworth, and Darcy M. Bullock

Effective January 2015, the *Manual on Uniform Traffic Control Devices* (MUTCD) requires that a sign meet minimum retroreflectivity standards based on color and sheeting type.[1] Retroreflectivity standards are defined as absolute thresholds; however, a sign may exhibit a range of retroreflectivity readings due to measurement uncertainty in sign-face sheeting properties, as well as uncertainty associated with the retroreflectometer and operator. ASTM E 1709-08 defines procedures for measuring retroreflectivity of traffic signs and notes these factors are sources of error, but the MUTCD standard and the literature are silent on how much uncertainty one can expect.[3]

Abstract

In an effort to provide guidance to agencies on expected uncertainty in readings, this paper examines the range of bias and coefficient of variation (COV) of retroreflectivity readings for Type I (engineering-grade beaded sheeting) and Type III (high-intensity beaded sheeting) signs in lab and field settings. In the lab, retroreflectivity readings were taken on 22 stop signs using three different retroreflectometers and four different operators. In the field, retroreflectivity readings were taken for 87 red, white, and yellow signs.

The median COV for the retroreflectivity readings of all of the signs was 5.6 percent. When different calibrated retroreflectometers were compared, median bias of up to 40 cd/lx/m^2 was observed. Understanding the bias and uncertainty in retroreflectivity readings is important as agencies are developing their protocol for compliance with the MUTCD standards on minimum retroreflectivity.[4, 5, 10] Finally, in response to potential litigation alleging signs failing to meet current standards, it is important to recognize that there is some uncertainty associated with retroreflectivity measurements.

Table 1: Minimum retroreflectivity standards (adapted from MUTCD Table 2A-3).

Sign Color	Retroreflectivity Standard by Sheeting Type (cd/lx/m^2)		Additional Criteria
	I	III, IV, VI, VII, VIII, IX, X	
White on Green	W* G ≥ 7	W ≥ 120 G ≥ 15	Ground-Mounted
Black on White	B = 0 W ≥ 50		-
Black on Yellow	B = 0 Y*	B = 0 Y ≥ 75	Text and fine symbols measuring less than 48 inches
White on Red	W ≥ 35 R ≥ 7		Minimum contrast ratio (White : Red) ≥ 3:1

Motivation

The *Manual on Uniform Traffic Control Devices* (MUTCD) has defined standards for sign retroreflectivity (Table 1) that become effective January 2015.[1] It is important to understand the expected bias and uncertainty of retroreflectivity readings. Currently, the section in ASTM Standard E 1709-08 on precision and bias of retroreflectivity states, "These data are under development."[3] This paper reports on a study to evaluate the measurement uncertainty and bias associated with using a retroreflectometer to measure the retroreflectivity of a traffic sign.

Understanding typical uncertainty and bias is particularly important with respect to potential litigation regarding sign retroreflectivity. Consequently, this paper seeks to provide the objective documentation characterizing the expected bias and measurement uncertainty an agency might expect to see.

Background

The ASTM Standard E 1709-08 provides numerous details on the measurement of retroreflectivity using a portable retroreflectometer. The standard includes a procedure for measuring and reporting retroreflectivity results and sources of error when measuring the signs.[3]

The literature is sparse in regards to uncertainty and bias of retroreflectivity measurements. Several authors have reported on the high variance in signs' retroreflectivity observed after signs have been deployed in the field for several years, but it is unclear whether this variance is due to different environmental effects, operator effects, or equipment effects.[6, 7, 8] Perhaps the most relevant paper is a pavement marking repeatability study completed by Holzschuher et al. in 2009. That study reported on the precision repeatability of a mobile retroreflectometer unit (MRU) for pavement marking retroreflectivity and reported a coefficient of variance of 2.4 percent.[2] Similar strategies were adopted to measure the repeatability of a sign retroreflectometer in this study.

The current MUTCD now has language requiring agencies to meet the minimum retroreflectivity standards by January 2015. Although most agencies will likely use a scheduled replacement program based on expected sign life and not regularly measure retroreflectivity, it is still important to understand the concept of retroreflectivity and recognize the expected uncertainty and bias that can be introduced during measurement of retroreflectivity. Understanding typical uncertainty and bias is particularly important with respect to potential litigation regarding sign retroreflectivity. Consequently, this paper seeks to provide the objective documentation characterizing the expected bias and measurement uncertainty an agency might expect to see.

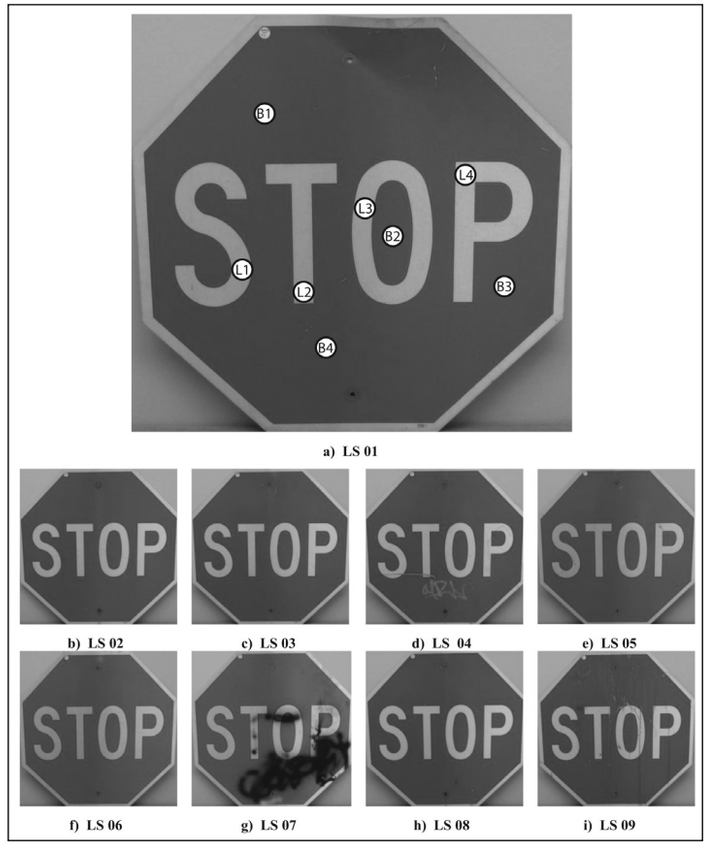
Table 2: Color and sheeting types of signs used in both the lab and field study.

Setting	Red		White		Yellow		Total
	Type I	Type III	Type I	Type III	Type I	Type III	
Lab	11	11	-	-	-	-	22
Field	15	15	15	10	18	14	87

Data Collection Procedure

Data collection was performed using Type I engineering-grade beaded sheeting and Type III high-intensity beaded sheeting in a controlled laboratory environment as well as sheeting installed in the field (Table 2). Measurements were taken at four predetermined locations on each sign. These locations were chosen based on an MUTCD recommendation to measure in a diagonal pattern, focusing on key visibility points on a sign.[1] Figure 1a shows an example of the retroreflectivity reading points on a stop sign; each type of sign had its own reading points distributed in a similar manner.

Figure 1: Type I stop signs (lab).



a) LS 01 d) LS 04 g) LS 07
b) LS 02 e) LS 05 h) LS 08
c) LS 03 f) LS 06 i) LS 09

Lab Study

In the lab study, retroreflectivity measurements were taken for 22 stop signs that had been removed from service in Richmond, Indiana, USA. These stop signs consisted of 11 signs of Type I sheeting and eleven signs of Type III sheeting. The Type I lab signs can be seen in Figure 1. Stop signs were selected for study because they are one of the most crucial types of sign managed by an agency. The 11 Type I stop signs were measured by four different operators (designated O1, O2, O3, O4) using three different retroreflectometers (designated R0, R1, R2). R0 was a retroreflectometer manufactured in 1999; R1 and R2 were identical models manufactured in 2009. Different operators and different retroreflectometers were used to determine whether uncertainty or bias in retroreflectivity readings occurred due to a change in the operator or measuring equipment.

Field Study

In the field study, retroreflectivity measurements were taken for 87 signs, which varied in sheeting type and color as shown in Table 2. The signs for the field study were all located in West Lafayette, Indiana, on or near the campus of Purdue University. These signs were measured by four different operators (designated

O1, O2, O3, O4) using two different retroreflectometers (designated R0, R1). The field data were used to provide a representative sampling of different colors and sheeting types commonly found in the field.

Data Analysis

Using the retroreflectivity measurements from both the lab and field studies, calculations were made to quantify both bias and uncertainty in the measurements. The lab study was used to evaluate bias in the sign measurements because the same signs were measured under tightly controlled conditions with different operators and retroreflectometers.

The field study was used to provide a larger dataset that was more representative of the data that an agency might collect. The coefficient of variance (COV) was used to compare the uncertainty in all of the sign measurements:

$$COV = \frac{variance}{mean}$$

This COV normalizes the variance by dividing it by the mean, which makes it an appropriate way to compare different sign colors and sheeting types.

Observations on Retroreflectivity Variation

A number of retroreflectivity measurements were taken to explore the uncertainty of measurements in the controlled conditions of the lab. Initial comparisons are described below and illustrated in Figure 2, Figure 3, and Figure 4:

- Three trials by one operator using one reflectometer (shown in Figure 2);
- Trials by one operator using three reflectometers (shown in Figure 3); and
- Trials by four operators using one reflectometer (shown in Figure 4).

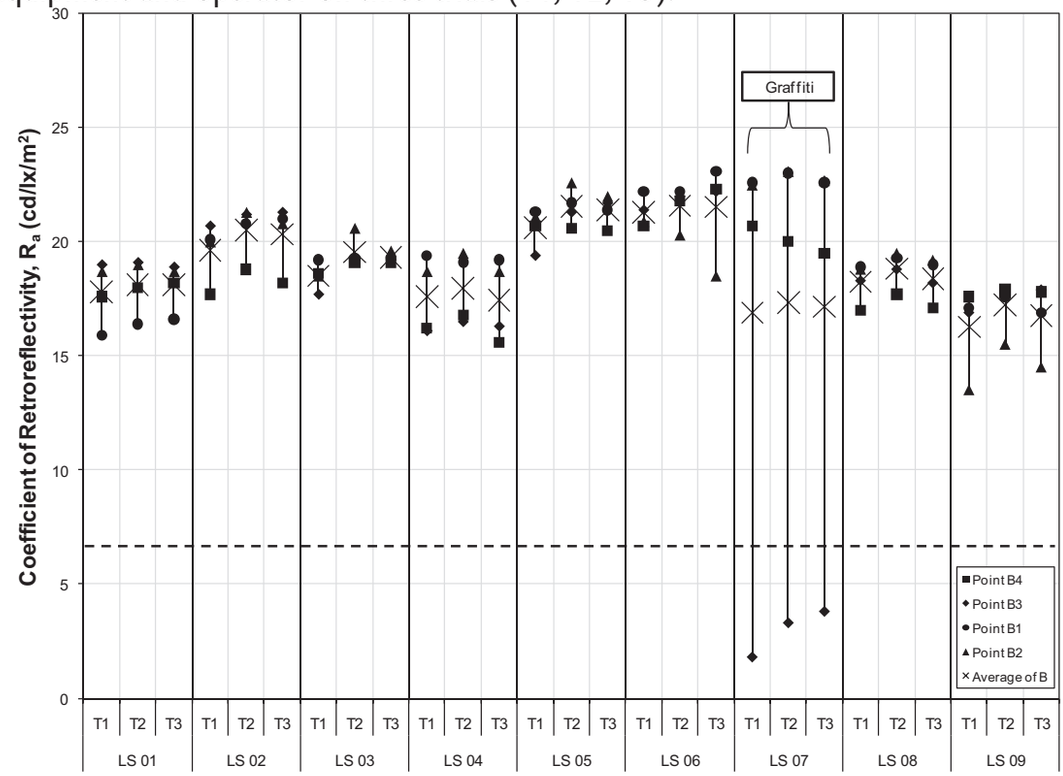
In Figure 2 through Figure 4, only nine of the 11 Type I signs included in the study are shown due to space constraints. The MUTCD minimum retroreflectivity standard is represented by the dashed line in these figures.

The first comparison (shown in Figure 2) was a true repeatability test, comparing three trials with measurements taken under the exact same circumstances: one operator using the same retroreflectometer on the same set of signs. Ideally, the uncertainty would be zero and each of these trials would produce identical results. Figure 2 illustrates the range of the actual measurements; clearly, there is measurement uncertainty even when a single operator uses the same retroreflectometer on three separate trials under controlled conditions in the laboratory. The vertical lines for each lab sign (designated LS _) and each trial (designated T_) in Figure 2a show the range of retroreflectivity readings for the background sheeting; the vertical lines for each sign in Figure 2b show the range of retroreflectivity readings for the sign legend.

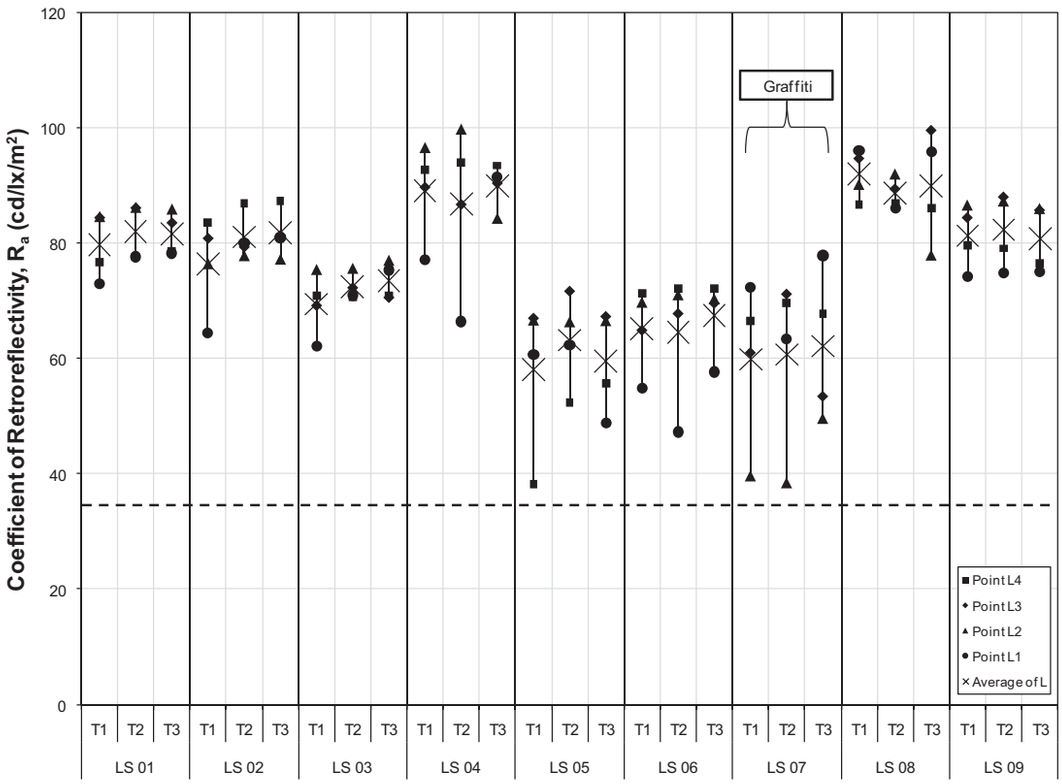
Upon examination, it is possible to discern each of the four readings for each trial. For example, inspecting sign LS 07 in Figure 2a, the lowest retroreflectivity reading for the background sheeting was observed on trial 1 (T1) at point B3; this reading was between 2 and 4 cd/lx/m^2. The highest reading for sign LS 07 on trial 1 (T1) was observed at point B1; this reading was approximately 23 cd/lx/m^2. The average reading for trial 1 (T1) was approximately 17 cd/lx/m^2. For sign LS 07, the wide range of values is attributable to graffiti as illustrated by the photograph of this sign in Figure 1g.

The second comparison (shown in Figure 3) illustrates the measurement uncertainty due to use of different equipment, and the third comparison (shown in Figure 4) illustrates the measurement uncertainty due to different operators. The uncertainty associated with different operators and

Figure 2: Lab retroreflectivity readings for nine signs (LS 01...LS 09) with same equipment and operator on three trials (T1, T2, T3).

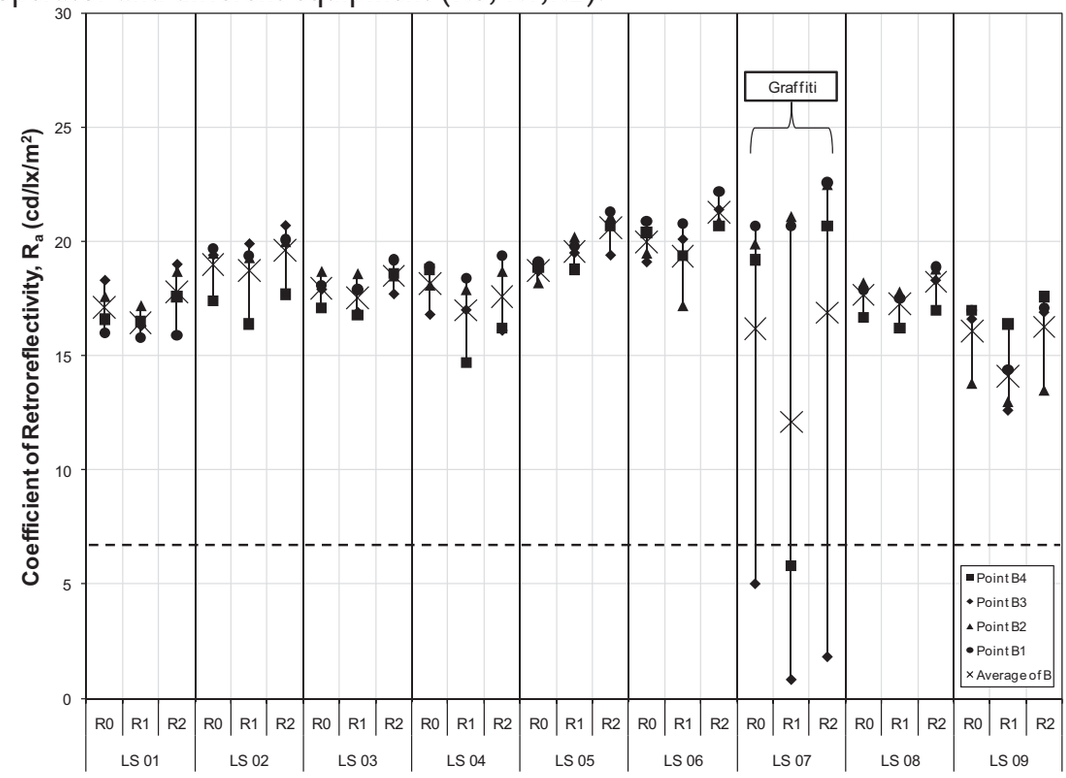


a) Range of observed background retroreflectivity.

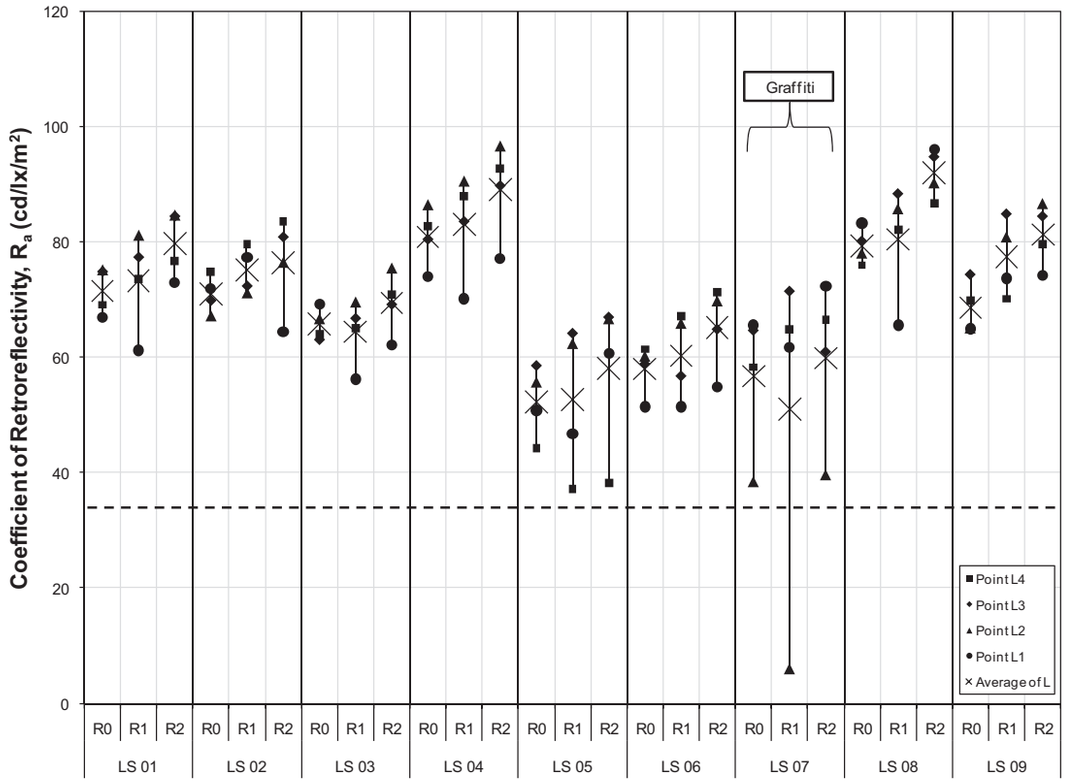


b) Range of observed legend retroreflectivity.

Figure 3: Lab retroreflectivity readings for nine signs (LS01...LS09) with same operator and different equipment (R0, R1,R2).

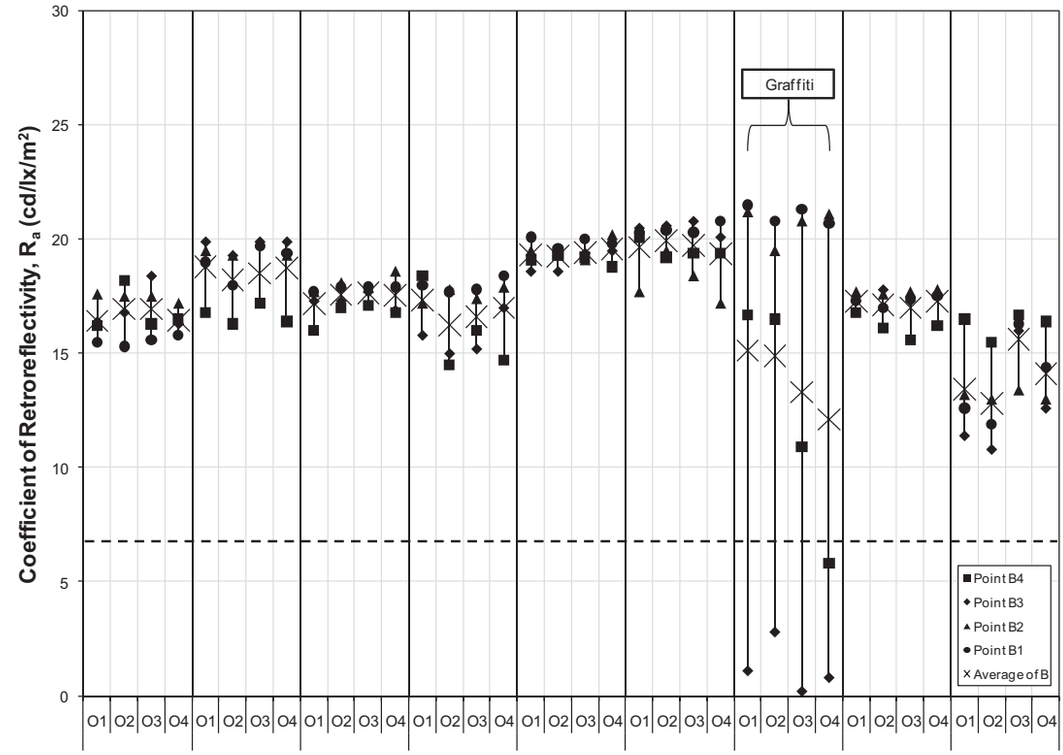


a) Range of observed background retroreflectivity.

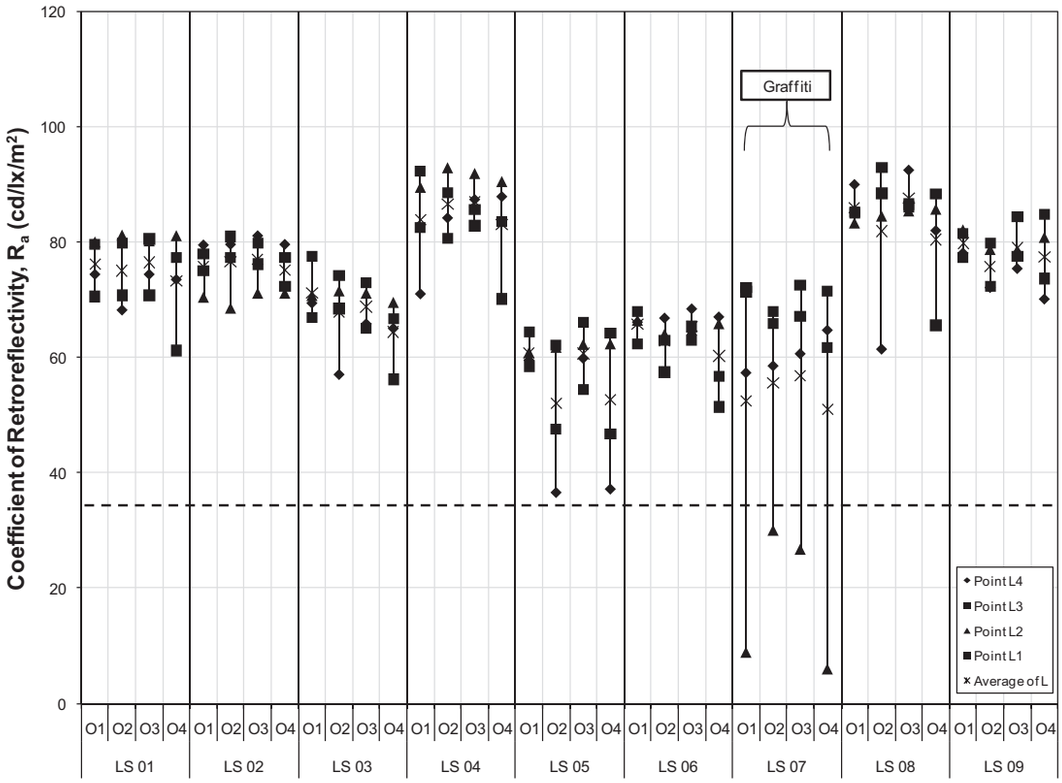


b) Range of observed legend retroreflectivity.

Figure 4: Lab retroreflectivity readings for nine signs (LS 01...LS 09) for four operators (O1...O4) with same equipment.



a) Range of observed background retroreflectivity.



b) Range of observed legend retroreflectivity.

Table 3: Coefficient of variation (%) for lab stop signs.

a) Type I stop signs

		Reference Figure	LS 01	LS 02	LS 03	LS 04	LS 05	LS 06	LS 07	LS 08	LS 09	LS 10[1]	LS 11[1]
i	1 Operator 1 Machine Background[2]	3a	6.3	6.2	3.6	8.7	3.9	5.7	50.4	4.5	8.8	7.4	4.8
ii	1 Operator 3 Machines Background [2]	4a	6.1	6.5	4.0	7.2	4.2	6.2	63.5	4.4	13.6	5.6	6.3
iii	4 Operators 1 Machine Background [3]	5a	5.7	7.2	3.6	8.2	2.5	5.5	62.5	3.9	14.2	4.5	5.0
iv	1 Operator 1 Machine Legend [2]	3b	5.6	7.4	5.5	10.3	16.2	12.3	21.2	6.6	6.7	7.4	4.8
v	1 Operator 3 Machines Legend [2]	4b	7.8	7.0	7.0	8.4	15.8	8.9	39.0	9.0	8.5	9.2	9.2
vi	4 Operators 1 Machine Legend [3]	5b	7.7	5.5	8.3	8.0	16.7	7.3	42.1	10.2	5.6	6.4	8.8

[1]Signs LS 10 and LS 11 were not shown in Figures 3 - 5 due to space reasons
[2]Based on 12 Readings per sign (4 readings per sign * [(1 Student * 3 Repetitions)])
[3]Based on 16 Readings per sign (4 readings per sign * [(4 Students * 1 Repetition)])

b) Type III stop signs

		LS 12	LS 13	LS 14	LS 15	LS 16	LS 17	LS 18	LS 19	LS 20	LS 21	LS 22
i	1 Operator 1 Machine Background[1]	10.5	4.0	1.4	5.7	2.8	2.8	4.0	9.3	6.5	37.5	4.2
ii	1 Operator 3 Machines Background [1]	17.5	4.5	2.6	5.2	3.3	3.5	3.1	9.5	4.5	33.0	4.2
iii	4 Operators 1 Machine Background[2]	17.4	3.8	1.0	4.6	2.7	3.0	2.3	7.5	2.2	36.5	4.3
iv	1 Operator 1 Machine Legend[1]	24.1	3.2	1.4	4.5	4.4	6.9	3.7	4.2	1.6	30.3	1.5
v	1 Operator 3 Machines Legend[1]	31.4	7.1	7.3	4.6	7.0	8.6	8.4	7.8	7.1	29.9	7.0
vi	4 Operators 1 Machine Legend [2]	33.1	6.5	5.8	4.3	4.3	5.6	6.6	11.9	1.7	32.2	2.1

[1]Based on 12 Readings per sign (4 readings per sign * [(1 Student * 3 Repetitions)])
[2]Based on 16 Readings per sign (4 readings per sign * [(4 Students * 1 Repetition)])

equipment is similar to the uncertainty associated with one operator but perhaps with some bias associated with the reflectometer.

The COV for each lab sign is based on all of the measurements taken. The results are shown in Table 3a for the Type I signs and in Table 3b for the Type III signs. The first comparison of measurements in the lab study, in row i and iv of Table 3a and 3b, shows the COV for one operator using one retroreflectometer. This COV for this comparison is based on 12 measurements (one operator on three trials with four readings per sign). The uncertainty ranges from a maximum COV of 50.4 for LS 07 to a minimum COV of 3.6 for LS 03 for the background of Type I stop signs.

The second comparison of measurements in the lab study, in row ii and v of Table 3a and 3b, shows the COV for one operator using three different retroreflectometers. Three retroreflectometers were calibrated and used to measure the exact same set of signs. The COV for this comparison is based on 12 measurements (one operator with three retroreflectometers with four readings per sign). The uncertainty ranges from a maximum COV of 63.5 for LS 07 to a minimum COV of 4.0 for LS 03 for the background of Type I stop signs. It is not uncommon to see a 6 percent uncertainty in the measurements of a sign for this comparison.

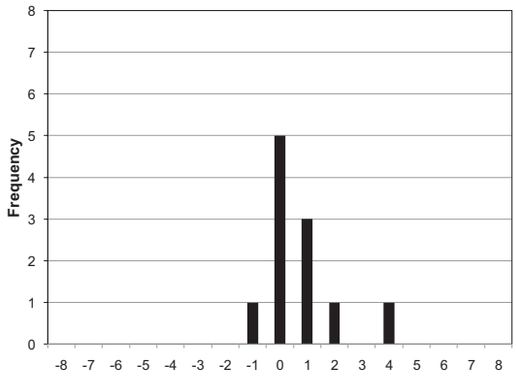
The third and final comparison of measurements in the lab study, in rows iii and vi of Table 3a and 3b, shows the COV for four different operators using the same retroreflectometer. The COV for this comparison is based on 16 measurements (four operators with one reflectometer with four readings per sign). The measurement uncertainty ranges from a maximum COV of 62.5 for LS 07 to a minimum COV of 3.6 for LS 03 for the background of Type I stop signs. Overall, the COV values in Table 3 clearly indicate that nontrivial uncertainty should be expected when taking retroreflectivity measurements.

Observations on Reflectometer Bias

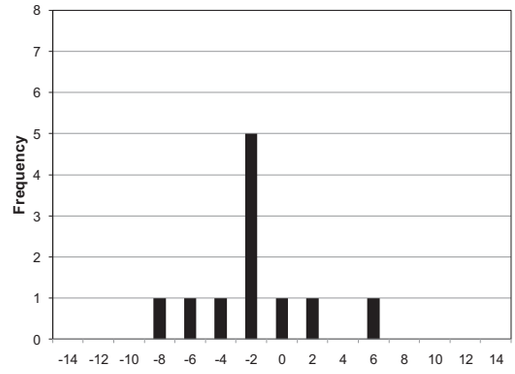
It was also desirable to explore the potential bias associated with the reflectometers (R0, R1, R2). To assess potential bias of one retroreflectometer to another, the difference between the mean retroreflectometer readings was calculated for each sign. To illustrate the bias, histograms were developed and are shown in Figure 5 and Figure 6 for Type I and Type III signs, respectively. If there were no bias, the histograms in Figure 5 and Figure 6 would be randomly distributed about 0; however, this does not appear to be the case, and the figures suggest that at least some bias is present. For example, Figure 5 and Figure 6 suggest that retroreflectometer R2 is likely to result in readings that are consistently higher than retroreflectometer R0, as evidenced for Type I red background with a mode of -1 (Figure 5e), for Type I white legend with a mode of -8 and -12 (Figure 5f), for Type III red background with a mode of -3 and -4 (Figure 6e), and Type III white legend with a mode of -45 (Figure 6f).

A comparison of the bias associated with the reflectometers is illustrated in a cumulative frequency diagram (Figure 7). The median bias values have a substantial range, from -40 to 2 cd/lx/m^2, and vary by sheeting type and color. The Type III sheeting material has a higher bias than Type I sheeting; however, this is not surprising because the retroreflectivity values are higher for Type III sheeting. Similarly, retroreflectometer R0 has a higher bias than the other retroreflectometers; however, this also is not surprising because it is an older unit.

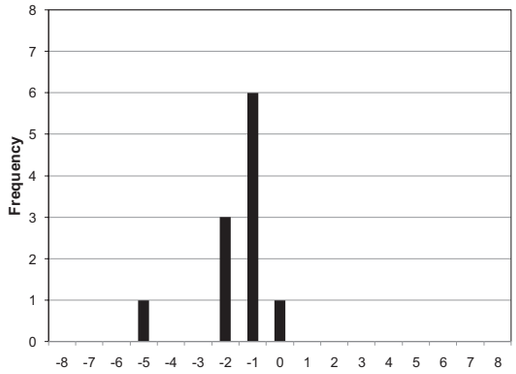
Figure 5: Lab retroreflectivity bias for type I stop sign measurements with same operator using different equipment (R0, R1, R2).



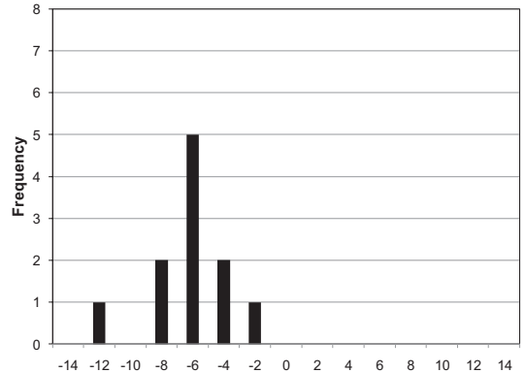
a) Type I Red Background R0-R1 (cd/lx/m^2).



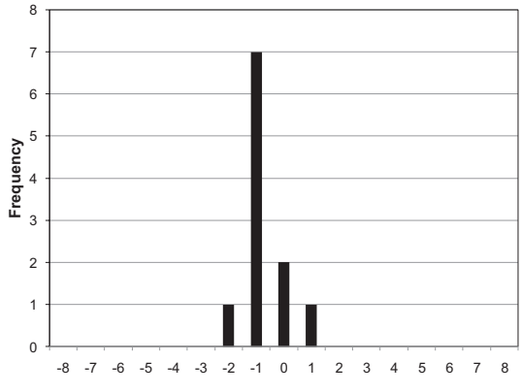
b) Type I White Legend R0-R1 (cd/lx/m^2).



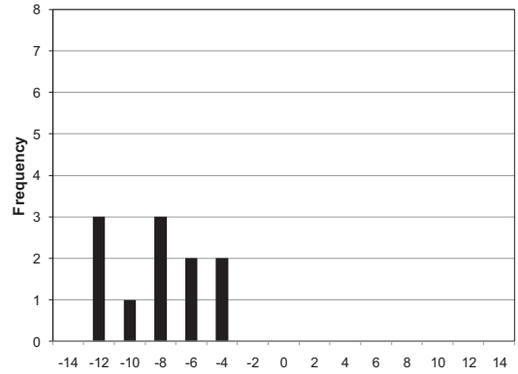
c) Type I Red Background R1-R2 (cd/lx/m^2).



d) Type I White Legend R1-R2 (cd/lx/m^2).

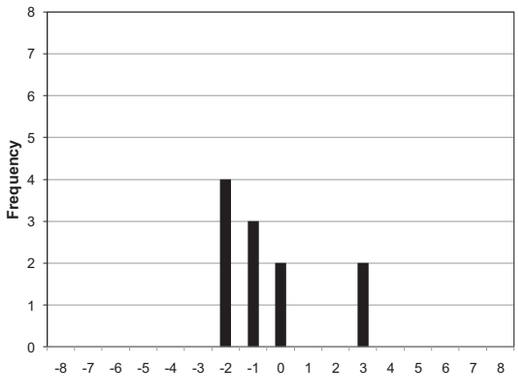


e) Type I Red Background R0-R2 (cd/lx/m^2).

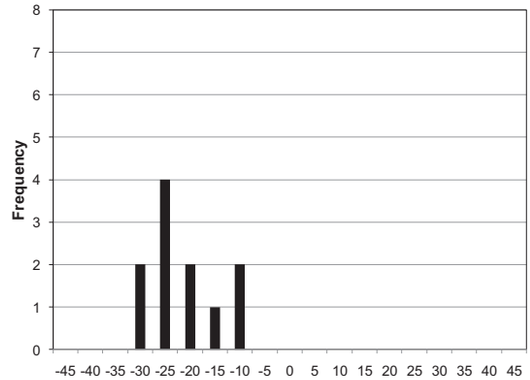


f) Type I White Legend R0-R2 (cd/lx/m^2).

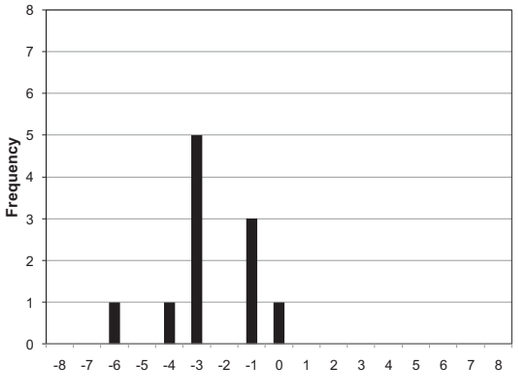
Figure 6: Lab retroreflectivity bias for type III stop sign measurements with same operator using different equipment (R0, R1, R2).



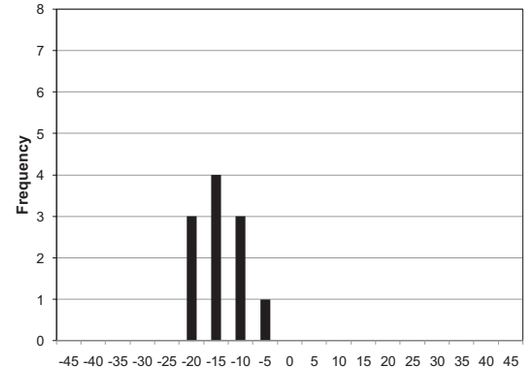
a) Type III Red Background R0-R1 (cd/lx/m^2).



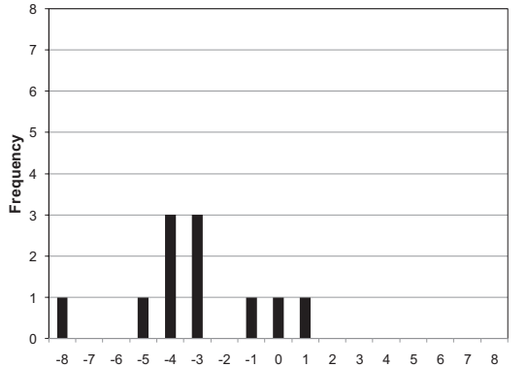
b) Type III White Legend R0-R1 (cd/lx/m^2).



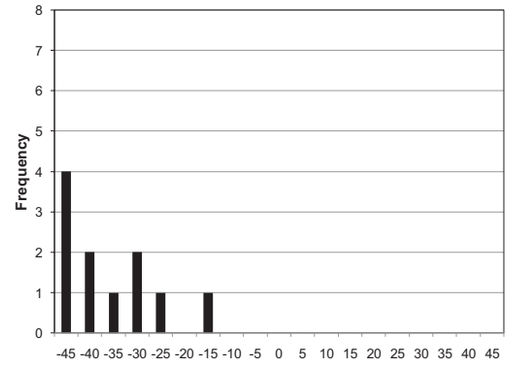
c) Type III Red Background R1-R2 (cd/lx/m^2).



d) Type III White Legend R1-R2 (cd/lx/m^2).

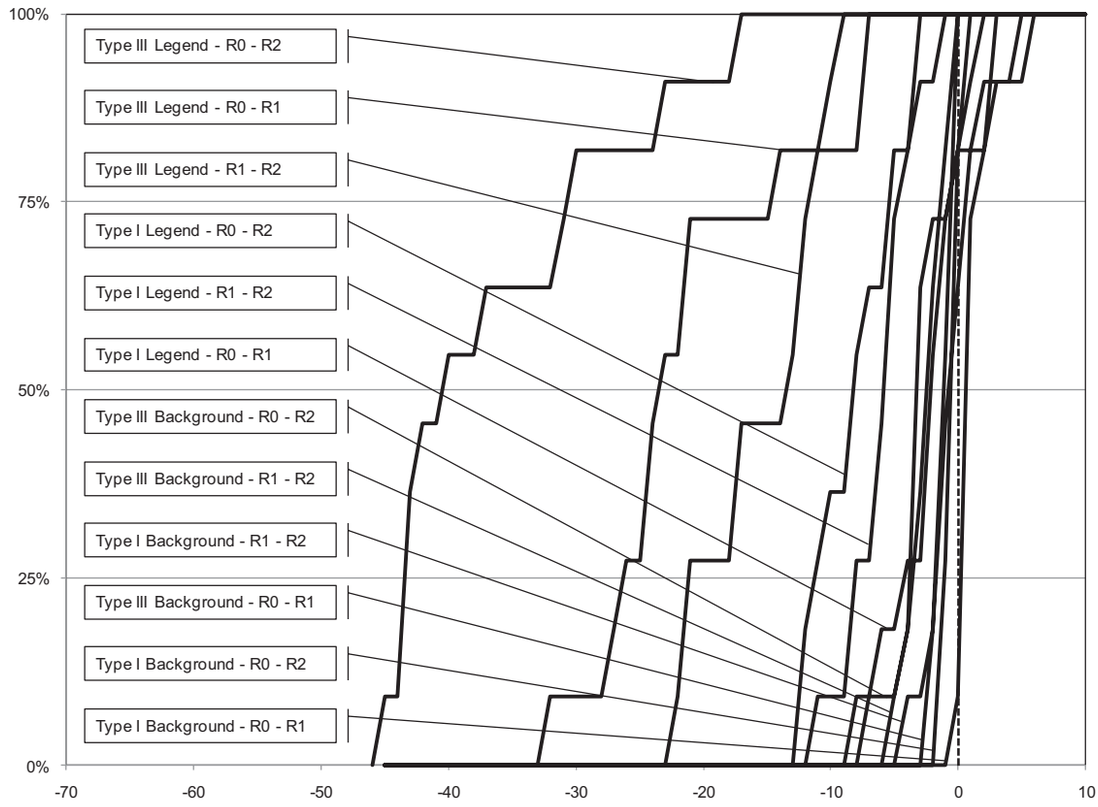


e) Type III Red Background R0-R2 (cd/lx/m^2).



f) Type III White Legend R0-R2 (cd/lx/m^2).

Figure 7: Absolute cumulative frequency diagram for retroreflectometer bias (cd/lx/m^2).



Measurement of Total Coefficient of Variation

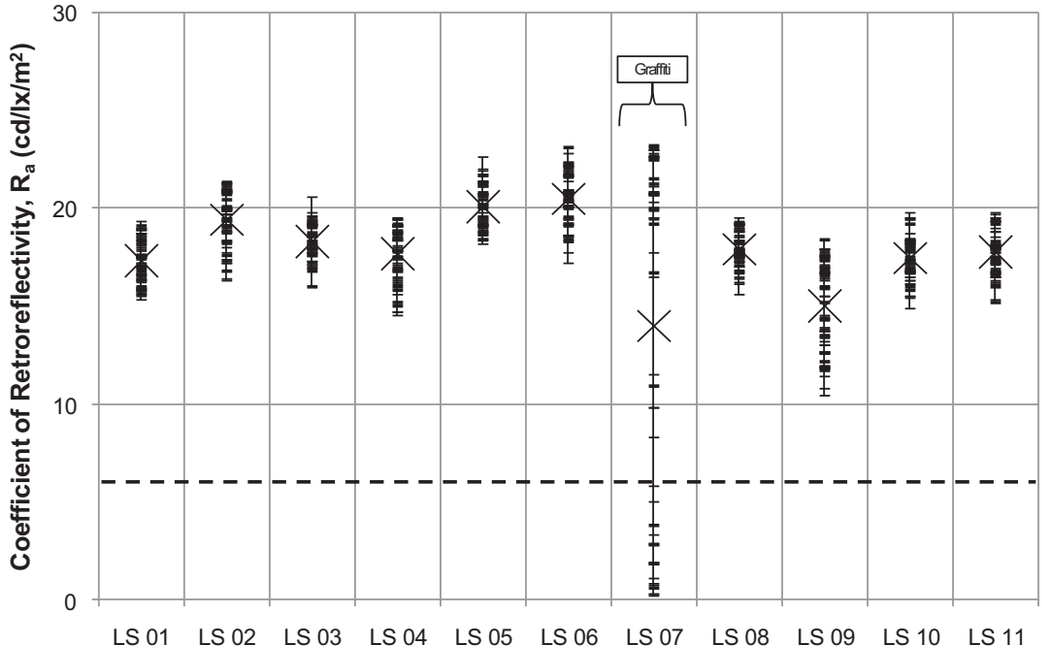
In Figure 8 and Figure 9, the sample pool of 56 measurements for each sign are shown for the background and legend for Type I and Type III sheeting. These plots display the range of readings taken for the same sign. This range is important because if an agency takes a single reading that is at the low end of the range, it may incorrectly assume that a sign needs to be replaced for compliance with the new MUTCD retroreflectivity standards. A very wide range of measurements is often indicative of a sign that has graffiti on it or has otherwise been damaged (Figure 8a, 8b LS07). Signs that have had graffiti cleaned off may also exhibit unusually wide ranges in observed retroreflectivity as a result of the chemicals that may be used to clean the graffiti off the sign. A wide range typically correlates with a large COV, which is shown below each sign in Figure 9.

Field Study

The field data were assessed in a similar manner to the lab data with calculations of the range of retroreflectivity measurements for each sign, the COV for each sign, and the bias for each retroreflectometer relative to the other reflectometer. The results of the analysis of the field data are consistent with the findings in the lab study and suggest that uncertainty and bias should be expected when measuring the retroreflectivity of signs. The results suggest that the range and COV for retroreflectivity measurements in the field are generally consistent with those in the lab. Both the lab data and the field data were pooled to create a histogram of the COV of 161 observations (four readings per observation) in Figure 10.

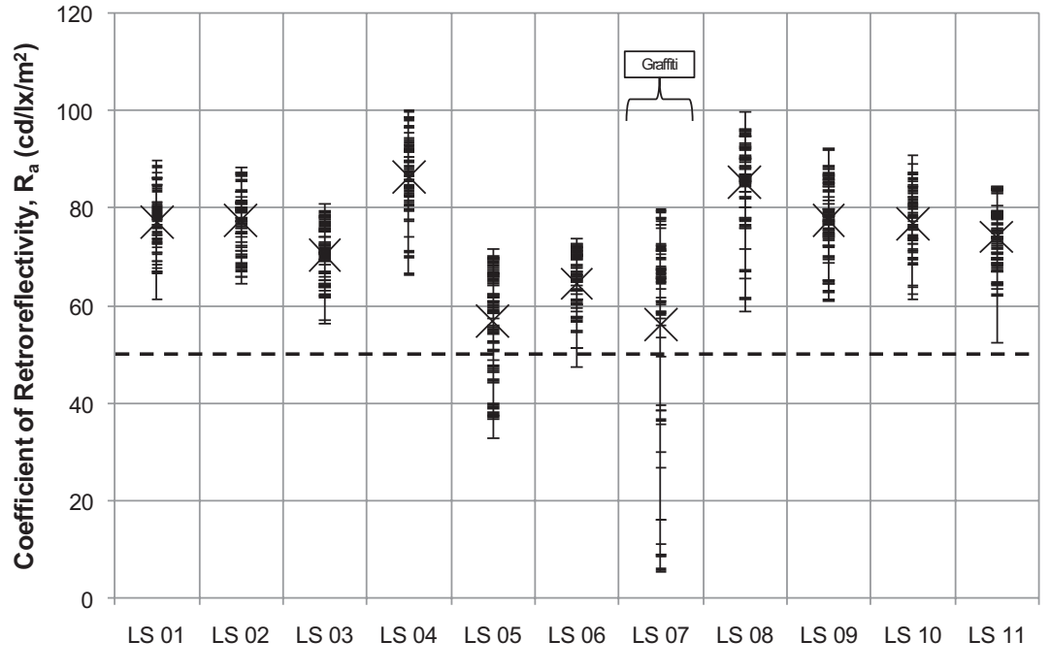
Figure 8: Graphical summary of all Type I lab stop signs' retroreflectivity.

Sign	LS 01	LS 02	LS 03	LS 04	LS 05	LS 06	LS 07	LS 08	LS 09	LS 10	LS 11
COV	6%	7%	5%	8%	6%	7%	64%	5%	15%	6%	6%



a) Sign background retroreflectivity.

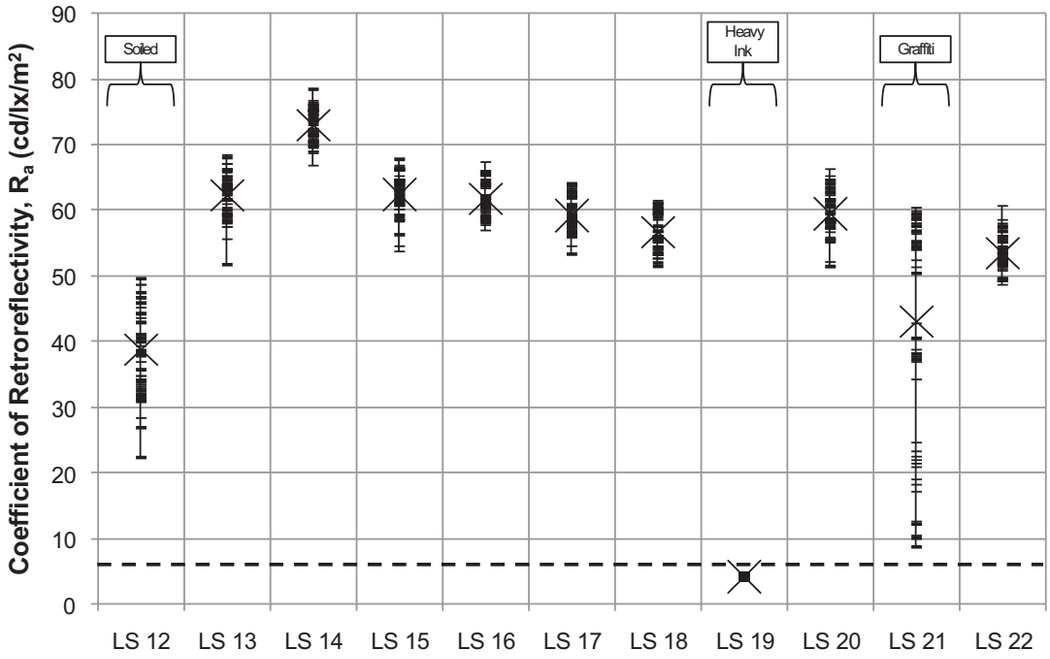
Sign	LS 01	LS 02	LS 03	LS 04	LS 05	LS 06	LS 07	LS 08	LS 09	LS 10	LS 11
COV	9%	8%	8%	9%	18%	10%	38%	10%	9%	9%	9%



b) Sign legend retroreflectivity.

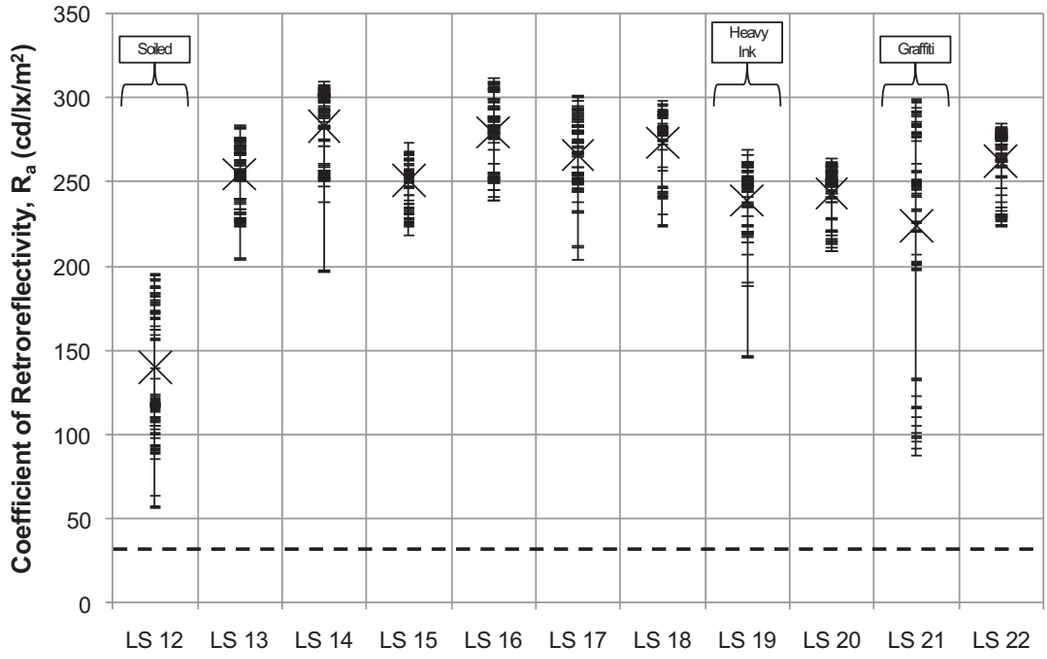
Figure 9: Graphical summary of all Type III lab stop signs' retroreflectivity

Sign	LS 12	LS 13	LS 14	LS 15	LS 16	LS 17	LS 18	LS 19	LS 20	LS 21	LS 22
COV	16%	5%	4%	5%	4%	5%	5%	9%	6%	39%	5%



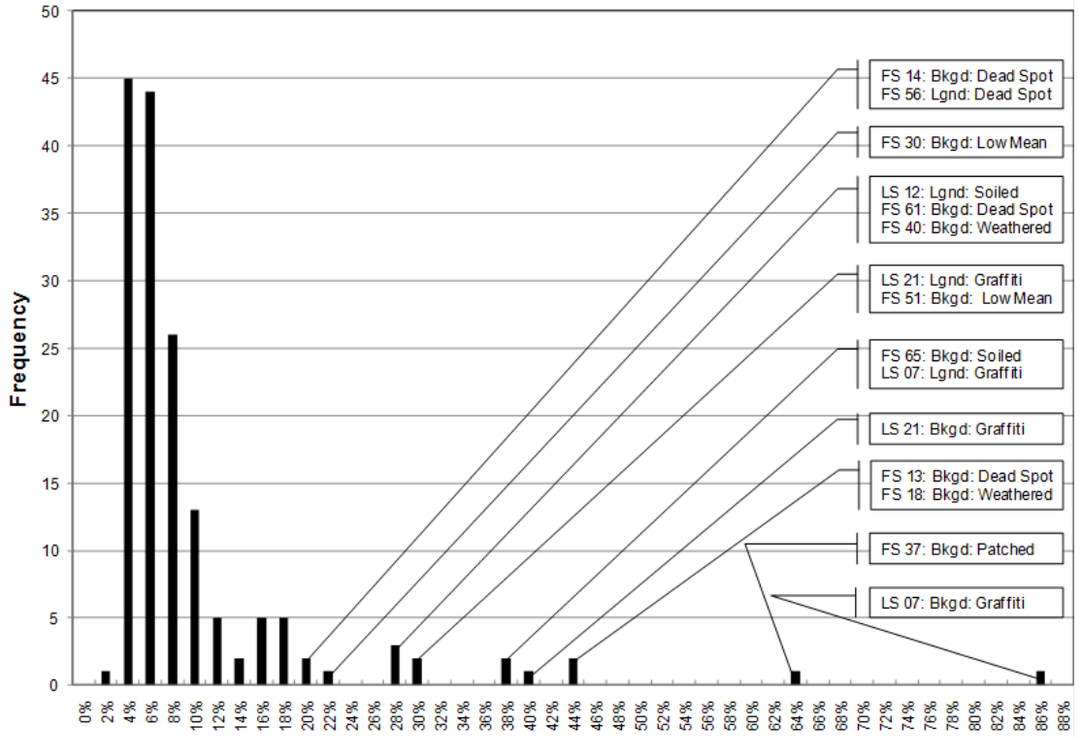
a) Sign background retroreflectivity.

Sign	LS 12	LS 13	LS 14	LS 15	LS 16	LS 17	LS 18	LS 19	LS 20	LS 21	LS 22
COV	28%	7%	8%	6%	7%	9%	8%	9%	7%	31%	7%

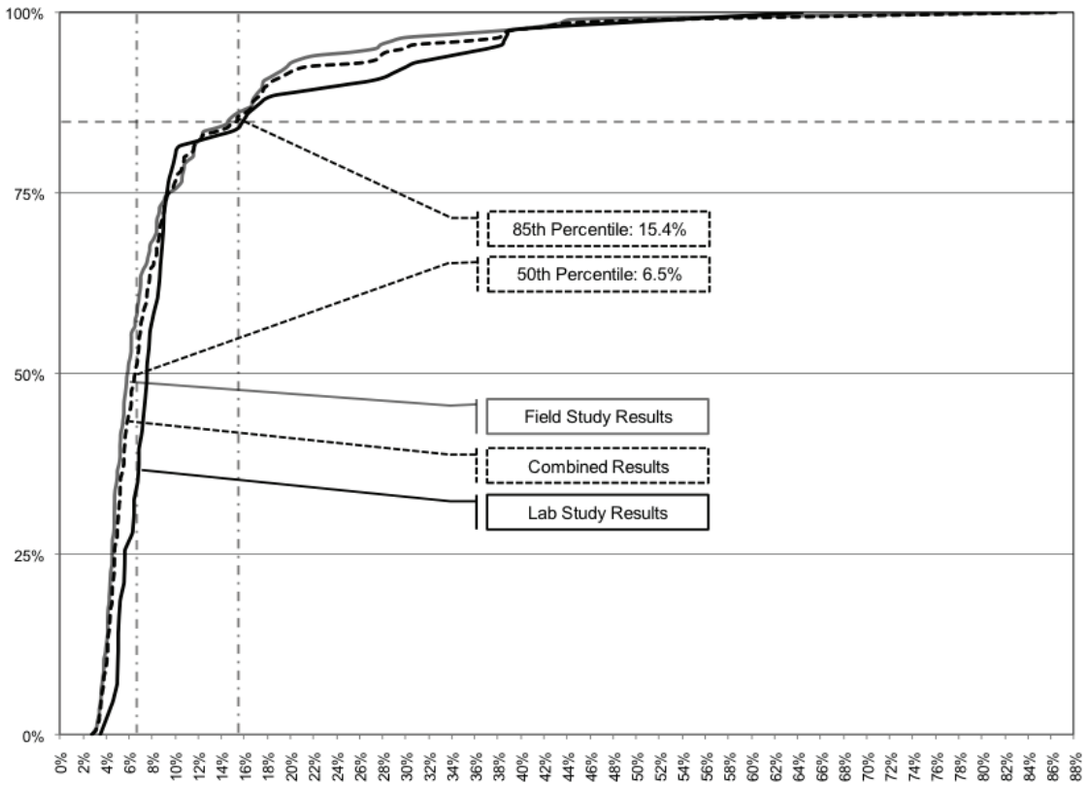


b) Sign legend retroreflectivity.

Figure 10: Coefficient of variation summary for all 161 readings (lab and field).



a) Histogram.



b) Cumulative frequency diagram.

Using a set of 109 traffic signs (161 total readings, including measurements of background sheeting and sign legends), four different operators, and three different retroreflectometers, the COV was determined for all readings for each individual sign. The study found that 85 percent of the measured signs have a COV of less than 15 percent. The callouts on Figure 10a illustrate that signs with COV values of greater than 15 percent are typically the result of damages or vandalism. It is reasonable to assume that the COV for an individual sign will be between 4 and 15 percent. In this study, an expected median COV was determined to be 6.5 percent for all signs as shown in Figure 10b. The COV results for both the lab and field measurements were plotted as a CFD, along with a combined summary line. The lab and field results closely mirrored each other, varying by less than 2 percent COV at the median, showing the similarities of the lab and field signs and measurement techniques.

The COV results for both the lab and field measurements were plotted as a CFD, along with a combined summary line. The lab and field results closely mirrored each other, varying by less than 2 percent COV at the median, showing the similarities of the lab and field signs and measurement techniques.

Conclusions

The text in ASTM Standard E 1709-08 – section 12 on the precision and bias of the retroreflectometer states, "These data are under development."[3] Using the data collected in this lab and field study, it was determined that there is nontrivial bias and uncertainty observed when measuring the retroreflectivity of traffic signs. This is the first study characterizing the bias and uncertainty of retroreflectometer measurements for traffic signs.

With Regard to Character Bias

Using a set of 22 stop signs and three different retroreflectometers in a controlled laboratory test, the range of median bias was determined for Type I and Type III sheeting for both the legend and background (white and red):

- Type I background ranged from 1 to 3 cd/lx/m^2;
- Type III background ranged from 2 to 4 cd/lx/m^2;
- Type I legend ranged from 3 to 12 cd/lx/m^2; and
- Type III legend ranged from 15 to 40 cd/lx/m^2.

With Regard to Character Variability

Analyzing both the field and laboratory data, it is reasonable to assume that the COV for an individual sign will be between 4 and 14 percent. In this study, the median COV was determined to be 5.6 percent.

It is important to recognize that there is uncertainty associated with retroreflectivity measurements in response to potential litigation alleging signs failing to meet current standards. Further data from a broader geographic range should be acquired to confirm this paper's findings with a larger sample and to determine an exact value that can be expected; however, the results of this study present a compelling case that bias and uncertainty need to be accurately characterized so that agencies can develop guidelines for their target tolerance thresholds.

Acknowledgments

This work was supported by the Indiana Local Technical Assistance Program. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.

References

1. *Manual on Uniform Traffic Control Devices for Streets and Highways.* 2009 Edition. FHWA, U.S. Department of Transportation, Washington, DC, December 2009, pp 30-44.

2. Holzschuler, C., Choubane, B., Fletcher, J., Severance, J., and Lee, H.S. Repeatability of Mobile Retroreflectometer Unit for Measurement of Pavement Markings. In TRB 2010 Annual Meeting CD-ROM, No. 10-1731, pp. 2-16.

3. ASTM. E 1790-08: Standard Test Method of Measurement of Retroreflective Signs Using a Portable Retroreflectometer at a 0.2 Degree Observation Angle. 2008.

4. Rogoff, M. J., A. S. Rodriguez, and M. B. McCarthy. Using Retroreflectivity Measurements to Assist in the Development of a Local Traffic Sign Management Program. In *ITE Journal*, Vol. 75, No. 10, ITE, October 2005, pp. 28-32.

5. Harris, Elizabeth A., W. Rasdorf, J. E. Hummer, and C. Yeom. Analysis of Traffic Sign Asset Management Scenarios. In *Transportation Research Record 1993*, TRB, National Research Council, Washington, DC, 2007, pp. 9-15.

6. Nuber, Luke, and D. M. Bullock. Comparison of Observed Retroreflectivity Values with Proposed FHWA Minimums. In *Transportation Research Record 1794*, TRB, National Research Council, Washington, DC, 2002, pp. 29-37.

7. Bischoff, Austin, and D. M. Bullock. *Sign Retroreflectivity Study.* Final Report, Report FHWA/IN/JTRP-2002/22. Indiana Department of Transportation Research Division, West Lafayette, IN 2002.

8. Immanei, V. P. K., J. E. Hummer, W. J. Rasdorf, E. A. Harris, and C. Yeom. Synthesis of Sign Deterioration Rates across the United States. In *Journal of Transportation Engineering*, Vol. 135, No. 3, March 2009, pp. 94-103.

9. McGee, H. W. and J. A. Paniati. *An Implementation Guide for Minimum Retroreflectivity Requirements for Traffic Signs.* Report FHWA-RD-97-052. Office of Safety and Traffic Operations R&D, FHWA, U.S. Department of Transportation, McLean, VA, April 1998, pp. 39-45.

10. Carlson, Paul J., Hawkins Jr., H. Gene, Schertz, Greg F., Mace, Douglas J., and Opiela, Kenneth S. Developing Updated Minimum In-Service Retroreflectivity Levels for Traffic Signs. In *Transportation Research Record 1824*, TRB, National Research Council, Washington, DC, 2003, pp. 133-143.

STEPHEN M. REMIAS received his BSCE from Michigan State University and is currently a doctoral student in the School of Civil Engineering at Purdue University. Steve has completed several projects in the area of traffic signal systems, origin-destination measurement techniques, and airport security wait time measurements.

SARAH M.L. HUBBARD is a professional engineer with experience in transportation planning, traffic engineering, and safety.

ERIC A. HULME, E.I., recently joined Area Wide Protective as a project manager. He received his BSCE from Purdue University and conducted postgraduate research at Purdue University in the areas of traffic operations, work zone analysis, and traffic sign retroreflectivity. He is a member of ITE.

ALEXANDER M. HAINEN received his BSCE from Michigan Technological University and is currently a doctoral student in the School of Civil Engineering at Purdue University. Alex 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.

GRANT D. FARNSWORTH received his BSCE from Brigham Young University and MSCE from Purdue. Grant is currently employed by HNTB as a design engineer. Grant has completed several projects in the area of retroreflective management programs, special-event traffic planning, and maintenance asset management.

DARCY M. BULLOCK 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.

The Influence of Underreported Crashes on Hotspot Identification

By Aaron Truong, Giovanni Bryden Jr., Wen Cheng, Ph.D. P.E., Xudong Jia, Ph.D., P.E., and Simon Washington, Ph.D.

Hotspot identification (HSID) plays a significant role in improving the safety of transportation networks. Numerous HSID methods have been proposed, developed, and evaluated in the literature. The vast majority of HSID methods reported and evaluated in the literature assume that crash data are complete, reliable, and accurate. Crash underreporting, however, has long been recognized as a threat to the accuracy and completeness of historical traffic crash records. As a natural continuation of prior studies, this paper evaluates the influence that underreported crashes exert on HSID methods.

Abstract

To conduct the evaluation, five groups of data gathered from Arizona Department of Transportation (ADOT; USA) over the course of three years are adjusted to account for 15 different assumed levels of underreporting. Three identification methods are evaluated: simple ranking (SR), empirical Bayes (EB), and full Bayes (FB). Various threshold levels for establishing hotspots are explored. Finally, two evaluation criteria are compared across hotspot identification (HSID) methods.

The results illustrate that the identification bias—the ability to correctly identify at risk sites—underreporting is influenced by the degree of underreporting. Comparatively speaking, crash underreporting has the largest influence on the FB method and the least influence on the SR method. Additionally, the impact is positively related to the percentage of the underreported property damage only (PDO) crashes and inversely related to the percentage of the underreported injury crashes. This finding is significant because it reveals that, despite PDO crashes being least severe and costly, they have the most significant influence on the accuracy of HSID.

Background

The identification of hotspots, sites with promise, blackspots, or accident-prone locations is an important task in road and traffic safety that seeks to screen potentially hazardous locations in a roadway network for further improvement. The importance of this task has long been recognized and, as a result, numerous HSID methods have been developed and proposed to better assist traffic safety engineers in correctly locating potentially problematic locations. Although methods vary in terms of analytical mechanics, underlying assumptions, and hazard indicator selection, most of them rely on the accuracy and completeness of historical traffic crash records. The evaluation of such methods to date, moreover, has assumed that crash records are complete. Importantly, as noted by Ogden, crash data are usually subject to various limitations including coding errors, location errors, missing data, discontinuities of measuring devices or techniques, delay of crash record-updating, and many others.[1] An obvious unanswered question is whether and to what extent data underreporting affects the accuracy of HSID methods.

Of particular interest here is incomplete reporting of road crashes, which has been recognized in many countries as a significant problem for road safety evaluation and policy making.[2–4] Due to its significant impact on safety studies, there is a substantial amount of literature dedicated to the study of this issue. Hauer and Hakkert conducted a review of 18 different studies.[5] By comparing police-reported crashes and crash cases reported to local hospitals, insurance companies, and garages, the authors reported considerable variability in the degree of nonreporting. According to their study, approximately 95 percent of fatalities were reported, while 20 percent of injuries that required hospitalization and half of all injuries sustained in motor vehicle crashes were not to be found in official statistics. They concluded that the accuracy with which road safety can be measured depends on the proportion of accidents reported and on the accuracy with which this proportion is known. Taking into account differences in the definitions of reportable accidents, reporting levels, and data sources used, Elvik and Mysen carried out a meta-analysis of road accident reporting studies conducted in 13 different countries.[6] They found that the mean reporting levels for fatal injuries, serious injuries (admitted to hospital), slight injuries (treated as outpatients), and very slight injuries (treated outside hospital) to be 95 percent, 70 percent, 25 percent, and 10 percent, respectively. Reporting was highest for car occupants and lowest for cyclists. A recent study documenting the seriousness of underreporting was conducted by Blincoe et al.[7] Based on several data sources, including the Fatality Analysis Reporting System (FARS), the Crashworthiness Data System (CDS), the General Estimates System (GES), the National Automotive Sampling System (NASS), and the National Health Interview Survey (NHIS), the authors estimated that in the year 2000, 21 percent of injury crashes and half of PDO crashes were not reported, whereas virtually 100 percent of fatal crashes were reported.

These past studies demonstrate that underreported crashes constitute a sizeable portion of motor vehicle crash populations that occur. While idealistically, all crashes will be identified and reported to police, the safety community must recognize that underreporting is a reality we must confront and account for in safety management practices. As found by Hauer and Hakkert, the probability that a crash is reported depends on a series of factors, including severity of the outcome, the age of the victim, the number of vehicles involved, and so on.[5] For example, crash nonreporting sometimes occurs when the at-fault injured party does not wish to involve police due to concerns about legal repercussions. Sometimes drivers are unaware that they are required to file a police report.[8] In order to reduce the administrative costs associated with crashes, jurisdictions on occasion raise crash reporting thresholds, which increases the proportion of nonreported crashes. A study of reporting thresholds in multiple states by Zegeer et al. found that the use of a tow-away reporting threshold would eliminate nearly half of the available crash data (48.3 percent).[9] In summary, given that not all crashes are reportable and not all reportable crashes are reported, it is imperative to estimate the impact of the underreporting of crashes in HSID methods.

There is relatively little research conducted to account for the impact of crash underreporting on HSID methods and modeling. A recent study conducted by Yamamoto et al. examined the effects of severity underreporting on model parameter estimates.[10] Through the use of both sequential binary probit and ordered-response probit models, mean parameter bias due to underreporting was found to be significant. Oh et al. revealed that underreporting of PDO crashes has significant influence on the accuracy of crash frequency models compared to weighted frequency models, accounting for crash severity.[11]

Considering that HSID methods rely on reported crashes, there is an underreporting need to examine the potential bias in HSID methods that may result. Hence, as a natural continuation of the study by Yamamoto et al., the main objective of this paper is to assess the crash underreporting influences in HSID methods under a variety of assumptions. This paper features a number of important differences and unique contributions. First, three identification methods are evaluated: simple ranking (SR), empirical Bayes (EB), and full Bayes (FB). Second, various assumed levels of crash underreporting are examined. Third, crash data from five different types of roadway sections are utilized. Finally, two evaluation criteria are used to compare methods.

The remainder of this paper first describes the HSID methods compared in the analysis. The details of the evaluation experiment consisting of the data used, assumed levels of underreporting, and evaluation criteria employed are then provided. Presented next are the results of the evaluation, followed by conclusions and recommendations.

The SR method is one of the most common HSID methods in practice. The reason for its popularity is due to its simplicity. To apply the SR method, a set of roadway locations or sections is ranked in descending order of crash frequency.

HSID Method Used in Comparison

Numerous HSID methods have been proposed and developed. Some methods are based on accident counts or frequencies,[12] some rely on accident rates,[13–16] and others adopt the joint use of accident frequency and rate to flag hazardous locations.[17] Instead of using overall accident counts at sites, some researchers have suggested using accident reduction potential (ARP) to identify sites with promise.[18–21] The ARP method is based on the assumption that only "excess" accidents over those expected from similar sites can be prevented by applying appropriate treatments, and thus it is argued that the potential for reduction is a superior method for identifying sites with promise. Again, for purposes here, three HSID methods based on crash counts are used: simple ranking (SR), empirical Bayes (EB), and full Bayes (FB). The details of each method are now described.

The SR method is one of the most common HSID methods in practice. The reason for its popularity is due to its simplicity. To apply the SR method, a set of roadway locations or sections is ranked in descending order of crash frequency. The top-ranked sites are then identified as potentially unsafe or hazardous sites for further engineering investigations. The number of top-ranked sites in the list that will be treated depends on the availability of funding resources to audit and possibly remediate the sites. Although the SR method has been widely applied by various agencies, it has one major disadvantage in that it is subject to regression to the mean (RTM) bias, a problem caused by random variations in crash count during the observation period resulting in diminished identification accuracy. In addition, it ignores traffic volumes or exposure and so favors the identification of sites with relatively low traffic volumes. In practice this second limitation is mitigated by using traffic volume categories to group sites for within-group comparisons.

To correct for RTM bias associated with typical SR methods, researchers have proposed using EB techniques.[22, 23–29] This method rests on several assumptions: Crash occurrence at a given location

obeys the Poisson probability law, while the probability distribution of the expected safety of the population of sites is approximately gamma distributed. In addition, it combines two clues as to the expected safety performance of a site—its crash history and the safety performance of similar sites. On the basis of these assumptions, the probability that a site has a random number of crashes is approximated by the negative binomial (NB) probability distribution. With the EB method gradually becoming the standard of professional practice, Hauer et al. provided a detailed tutorial on EB, which features a series of examples.[30]

In the EB method, the expected safety of a site λ_i is expressed as follows:

$$\lambda_i = wE[\lambda] + (1-w)\,x_i \tag{1}$$

where w is a weighting factor, $E[\lambda]$ is the expected safety of a reference population, and x_i is the observed count history for site i. The w (weighting factor) can be calculated through the following equation:

$$w = E[\lambda]/\{E[\lambda] + VAR[\lambda]\} \tag{2}$$

where $VAR[\lambda]$ is the corresponding variance of the expected safety of a reference population. If a safety performance function (SPF) for the reference population that relates crashes to covariates can be developed, w can be rewritten as follows:[29]

$$w = [1 + (\mu * Y) / \varphi]^{-1} \tag{3}$$

where μ is expected number of accidents/km-year on similar segments or accidents/year expected on similar intersections, Y is the number of years of accident count data used, and φ is the overdispersion parameter, which is a constant for the SPF and is derived during the regression calibration process.

A full Bayesian approach is an alternative method that rests upon Bayes' theorem. The fundamental difference between the EB and FB methods is that the former uses a point estimate for the expected safety performance of similar sites (μ in Equation 3), whereas the latter makes use of the entire distribution of safety performances of similar sites. Like the EB method, the FB method has also enjoyed wide applications in safety analysis,[34] especially with the availability of the software package WinBUGS.[35] Even though numerous studies have illustrated favorable results yielded by the EB method,[36–39] some researchers have also noticed the limitations associated with the EB approach.[40, 41] In EB analysis, an SPF needs to be calibrated based on sites with similar traits. Researchers have argued that in some cases limited reference samples can significantly affect the validity of the analysis results using the EB method. Another criticism is leveled on the EB's inadequate capability to explicitly account for the "uncertainty" of model parameter coefficients. Once the SPF is developed, all the model parameters and coefficients are treated as constant values and then are incorporated into the point estimates of the long-term safety of candidate sites. Through empirical analyses and/or comparisons, a set of studies reveals the potential advantages of the FB approach relative to the EB one: its capability to seamlessly integrate prior information and all available data into a posterior distribution (rather than point estimates), its capability to provide more valid safety estimates in smaller data samples, and its capability to allow more complicated model specifications.[40–44] In addition to the normal Poisson-Gamma distribution, the FB models are also capable of accommodating the Poisson-Lognormal distribution and various hierarchical Poisson distributions that can address the serial and spatial correlations among the sites.

For purposes here, to enable straightforward comparison of the evaluation results of alternate methods and to clearly identify the impacts of the underreporting of crashes, the FB estimation process employs the standard Poisson-Gamma (NB) distribution, which is consistent with the

assumptions of the EB method. Alternative FB specifications, not examined here, are discussed in various other sources.[40, 41, 45] The FB Poisson-Gamma model is usually expressed as follows:

$$y_{it} | \lambda_{it} \sim Poisson\ (\varepsilon_i \lambda_{it}) \quad (4)$$

$$\ln(\lambda_{it}) = X'_{it}\beta \quad (5)$$

$$\varepsilon_i \sim \text{Gamma}\ (\alpha, \tfrac{1}{\alpha}) \quad (6)$$

where i is the site index, t is the time period index, y is the recorded crash number, λ is the expected crash number, X' is the matrix of risk factors, β is the vector of model parameters, ε is the random effects, and α is the hyper-parameter of the model.

Description of Study

The levels of crash underreporting, data description, modeling results, and the selection of evaluation criteria are presented in this section.

Determination of the Levels of Crash Underreporting

As discussed previously, levels of crash underreporting are dependent on a multitude of influential factors. As shown in previous studies, crashes of low severity generally are reported less often than more severe crashes. Estimates of reporting for fatal crashes tend to exhibit the least variance, with virtually all incidents reported. In contrast, reporting estimates for PDO crashes vary considerably. Based on the study by Elvik and Mysen, there is substantial variability in the degree of underreporting among different countries.[6] In recognition of the difficulty of evaluating the full range of crash underreporting found in various countries, this paper focuses on the range of crash underreporting documented in the United States. Specifically, the findings of the U.S. DOT study performed by Blincoe et al.[7] are used as the basic building blocks of the study: 0 percent of fatal crashes, 21 percent of injury crashes, and 50 percent of PDO crashes not reported. It is important to note that these are national estimates and as such state and local agencies might experience different underreporting conditions that vary as a function of factors described previously. To capture variability that might be observed in practice, multiple (15) levels of crash underreporting are utilized in this evaluation. The 15 experimental conditions are shown in Figure 1.

The underreporting levels shown in Figure 1 are meant to reflect a realistic range of crash underreporting across the United States, with percentages ranging from 10 percent to 60 percent for underreported PDO crashes and from 10 percent to 50 percent for underreported PDO crashes.

Figure 1: Levels of underreporting of injury crashes and PDO crashes.

Levels of underreporting: (percent of underreported injury crashes, percent of underreported PDO crashes)
60 percent PDO underreporting: (10 percent, 60 percent); (20 percent, 60 percent); (30 percent, 60 percent); (40 percent, 60 percent); (50 percent, 60 percent)
50 percent PDO underreporting: (10 percent, 50 percent); (20 percent, 50 percent); (30 percent, 50 percent); (40 percent, 50 percent)
40 percent PDO underreporting: (10 percent, 40 percent); (20 percent, 40 percent); (30 percent, 40 percent)
30 percent PDO underreporting: (10 percent, 30 percent); (20 percent, 30 percent)
20 percent PDO underreporting: (10 percent, 20 percent)

The selected crash underreporting levels cover generally assumed conditions and are vulnerable to several limitations. First, the limits of the percentage of underreported PDO and injury crashes are somewhat arbitrary, and other underreported percentages are possible. Second, for both PDO and injury crashes, the percentage values increase by an interval of 10 percent. A smaller percentage increase (e.g., 5 percent) might lead to different results, and linear interpolation may not be appropriate. Third, injury crashes typically include various severity levels such as minor injury, moderate injury, and so forth. However, this study uses only one percentage to represent the overall underreporting condition of various injury severity levels. It is expected that the employment of different percentage values for different injury severity levels would yield more accurate results.

Data Description and Model Results

Three-years of crash data (2000–2002) of five functional classifications of rural roadway sections were obtained from the Arizona DOT for HSID analysis and underreporting impact assessment. The data consist of both highway and accident data. The highway data include information on pavement type, speed limit, section length, and so on. The crash data include detailed crash information, such as crash locations, crash collision type, crash severity, and so forth. Table 1 shows the statistics for roads of various functional classifications. The one potential disadvantage is that the sample size for minor collector road classification is relatively small.

The data shown in Table 1 represent the original crash data in the study. For the purpose of evaluating crash underreporting impacts on HSID methods, 15 new datasets were created for each functional classification of roadway based on the 15 different levels of crash underreporting discussed previously. The process of creating new crash data is straightforward. Assume one road section has three fatal crashes, 10 injury crashes, and 20 PDO crashes. If the underreported percentage for injury and PDO are 30 percent and 40 percent, the new crash number of different severities for the site would be 3, 13 (=10x130 percent), and 28 (=20x140 percent), respectively. The same adjustment process is applied for each site of the same classification. Overall, based on the original five datasets, there are 75 new datasets created with underreported crashes taken into consideration.

As discussed previously, safety performance functions are required when the EB method is used to identify the hotspots. Considering that the level of service and design standards vary significantly among various functional classifications of roadway, the SPFs were developed for each of the 80 datasets (five original ones plus 75 new sets). Due to overdispersion of crashes observed on various classifications of road segments, a negative binomial regression model was utilized. The model functional form used for the five SPFs is given as:

$$\lambda_i = \alpha * EXP(SL)\beta * (AADT)^{\lambda} \quad (7)$$

where λi is the expected safety of a site, *SL*, and is the independent variable representing the road section length and α, β, and λ are estimated parameters. It should be noted that *SL* is never zero and

Table 1: Statistics for rural roads of various functional classifications.

Functional Classification	Description	Number of Sections	Length (km)	Crashes (2000–2002)	Average AADT
1	Principal arterial-interstate	403	996.113	8122	23810
2	Principal arterial-others	441	1115.128	7012	7603
3	Minor arterial	436	1132.092	5261	5483
4	Major collector	628	1856.064	5285	2637
5	Minor collector	100	365.4	416	684

so it is not technically problematic to have a non-zero prediction for the case when *SL*=0. Hence, the model form including an intercept is used to develop the SPFs.

Two measures were used to assess the goodness of fit of such models: ρ^2 statistic and Akaike's information criterion (AIC). The ρ^2 statistic is a statistical test specifically designed to assess the goodness of fit of the logistic regression model. It is used in a similar manner as R^2 in linear regression analysis and is calculated as

$$\rho^2 = 1 - \frac{L(\beta) - K}{L(0)} \tag{8}$$

where $L(\beta)$ is the value of the log likelihood function when all parameter are zero, $L(0)$ is the value of the log likelihood function at its maximum, and K is the degree of freedom.

AIC is another effective measure of the goodness of fit of an estimated statistical model. It suggests that the better fitting model will yield smaller AIC values. In a general case, AIC is expressed as:

$$AIC = 2k - 2\ln(L) \tag{9}$$

where k is the number of parameters in statistical model and L is the maximized value of the likelihood function for the estimated model.

For illustration purposes, Table 2 shows a summary of SPFs calibrated for the five original functional classifications of road sections including estimated parameter values, the t-statistics for all independent variables, dispersion parameters, and measures of goodness of fit. The table reveals that all SPFs reveal good statistical fit, and also signs of all the estimated coefficients agree with expectations. All the independent variables are significant with 95 percent confidence.

As was done for the EB method, 80 FB Poisson-Gamma models were also calibrated across the datasets. The FB analyses were performed via the software package WinBUGS, which uses MCMC algorithm.[45] To reflect the lack of strong information of the values of various parameters and coefficients, prior distribution for all the regression coefficients (β) were assumed N (0, 103), and dispersion parameter α was assumed gamma (0.001, 0.001). As previously, the results of the five models for original datasets are shown in Table 3 for the FB models.

Table 2: Summary of safety performance function model results for the EB method.

Functional Classification	Estimated Parameters				Goodness of Fit	
	Intercept	SL	Ln(AADT)	Dispersion	ρ^2	AIC
1	-8.38373(-10.585)	0.27171 (-16.477)	0.93809 (12.208)	3.8883	0.4	1821
2	-3.12899(-4.899)	0.18884 (13.653)	0.47595 (6.775)	2.9287	0.36	1751
3	-4.30404(-8.215)	0.20146 (13.010)	0.61753 (10.514)	3.7999	0.36	1476
4	-1.9286(-4.796)	0.14255 (12.177)	0.36392 (7.299)	2.9555	0.32	1675
5	0.045(1.221)	0.059 (3.468)	0.073 (4.968)	2.9978	0.33	1465

Notes: SL is the variable of road section length (km); Ln (AADT) is the logarithm transformation on variable AADT. Values shown in parentheses represent t-statistic values associated with various independent variable coefficients.

Table 3: Summary of Poisson-Gamma model parameter estimation for the FB method.

Functional Classification	Estimated Parameters			
	Intercept	SL	Ln(AADT)	Dispersion
1	-8.305(-9.037,-7.786)	0.267(0.229,0.295)	0.929(0.881,0.998)	4.022(3.090,5.159)
2	-2.805(-4.043,-2.004)	0.178(0.150,0.207)	0.448(0.362,0.584)	2.986(2.371,3.782)
3	-4.063(-4.641,-3.630)	0.20146(0.163,0.225)	0.585(0.538,0.649)	3.905(2.942,5.257)
4	-1.636(-2.373,-0.857)	0.136(0.112,0.161)	0.321(0.225,0.408)	3.075(2.379,3.916)
5	0.076(-1.653,1.676)	0.054(-0.011,0.116)	0.069(0.168,0.304)	3.085(2.183,4.012)

Notes: SL is the variable of road section length (km); Ln (AADT) is the logarithm transformation on variable AADT. Values shown in parentheses represent Bayesian Credible Interval (95%BIC) values associated with various independent variable coefficients.

Evaluation Criteria

The selection of appropriate evaluation criteria plays a significant role in identifying the effects of crash underreporting. Compared with the large number of studies focused on the development of various HSID methods, considerably less research has been dedicated to devising the evaluation criteria for comparing the performance of various methods. Hauer and Persaud proposed the use of false identifications, consisting of false negatives and false positives, to measure the performances of various methods for HSID.[47] Based on these two statistics, Elvik presented two diagnostic criteria including sensitivity and specificity.[48] Subsequently Cheng and Washington developed four new evaluation criteria containing the Site Consistency Test, the Method Consistency Test, the Total Rank Differences Test, and the Poisson Mean Differences Test.[39]

In this study the three HSID methods were applied to both original crash datasets and new adjusted datasets. The influences of crash underreporting on HSID were determined through the examination and comparison of the identified locations. Assessment criteria from a prior study by Cheng and Washington—the Site Consistency Test and the Total Rank Differences Test—were chosen as the basis to examine the potential negative effects of underreporting. For improved illustration, the Site Consistency Test was slightly modified to the Site Inconsistency Test, which is able to reveal the number of different sites identified based on unadjusted and adjusted crash data, respectively. To facilitate the understanding of the mechanics of the tests and the effects of underreporting, a sample of 30 road sections from Functional Classification 1 is presented in Table 4. In addition to the original reported crashes, a new dataset is shown using the underreporting combination of 40 percent, 60 percent. The sites are ranked as hotspots in terms of crash number. In this example only the SR method is used to identify potential hotspots. For the other two methods, the ranking is based on the EB-estimated and FB-estimated safety performances respectively.

Total Rank Difference Test

This test relies on site ranking to evaluate the influence of crash underreporting in hotspot identification. The test is conducted by calculating the sum of total rank differences (absolute value) of the hazardous road sections identified between the original and each of the new datasets.

Table 4: Crash information of the sample of 30 road sections using SR method.

Reported crash data			New data set example (40%, 60%) underreporting condition		
Ranking	**Site #**	**Total Crash**	**Ranking**	**Site #**	**Total Crash**
1	146	22	1	316	32
2	330	22	2	863	32
3	868	22	3	28	33
4	1627	22	4	868	33
5	247	23	5	1627	33
6	863	23	6	47	33
7	1679	23	7	1679	33
8	47	24	8	643	34
9	218	24	9	247	34
10	643	24	10	56	35
11	649	24	11	649	35
12	56	25	12	277	35
13	277	25	13	218	36
14	57	26	14	57	39
15	202	27	15	1483	39
16	1206	27	16	202	40
17	1483	27	17	1775	40
18	1775	27	18	1206	41
19	144	28	19	764	42
20	764	28	20	144	43
21	69	32	21	653	48
22	493	32	22	69	48
23	954	32	23	493	48
24	653	33	24	954	49
25	221	36	25	1646	54
26	1646	36	26	221	54
27	215	40	27	428	60
28	***428***	***42***	***28***	***215***	***63***
29	***463***	***44***	***29***	***945***	***64***
30	***945***	***72***	***30***	***463***	***109***

Notes: 1. The new dataset is created using the underreporting condition of (40%, 60%) as an example;
2. Only the total crash numbers are shown in the table for brevity. They are calculated as the sum of crashes of different severity levels.

The smaller is the total rank difference, the less influence underreporting has on HSID—reflecting consistent ranking of sites across datasets.

In equation form, this test is given as:

$$T1_j = \sum_{k=n-n\alpha}^{n} (\Re(k_{j,i}) - \Re(k_{j,i+1})) \tag{10}$$

where

n = total number of sites being compare

$\Re$ = ranking order of site k in dataset j for method j

α = threshold of identified high-risk site (e.g., $\alpha = 0.05$ corresponding with top 5 percent of n sites identified as high risk)

j = HSID method being compared (e.g., $j = 1$ could be the SR, $j = 2$ the EB, etc.)

i = dataset (e.g., i = original dataset, i+1 = adjusted dataset 1, etc.)

To illustrate this test, consider the small sample of data presented in Table 4. The top 10 percent sections—Sites 428, 463, and 945, identified by the SR method—possess the 28th, 29th, and 30th rankings in the original dataset and 27th, 30th, and 29th rankings in new data. The total ranking differences of accident frequency as a result is 3 (|28-27| + |29-30|+|30-29|).

Site Inconsistency Test

While the previous test takes into account the rankings of safety performances of road sections, this test relies on the number of sites that are inconsistently identified as potential hotspots between the original and underreported datasets, with the intent to evaluate the effects of the crash underreporting on HSID. As underreporting increases so does the number of inconsistent sites; the test rests on the assumption that a site identified as high risk based on reported crash data should also reveal inferior safety performance based on the underreported crash data, if the underreported crash number exerts no influence on HSID.

In equation form, this test is given as:

$$T2_j = \{k_{n-n\alpha}, k_{n-n\alpha+1}, \ldots, k\}_{j,i} \neq \{k_{n-n\alpha}, k_{n-n\alpha+1}, \ldots, k_n\}_{j,i+1} \tag{11}$$

where

k = ranked site in dataset j for method j

j = HSID method being compared (e.g., $j = 1$ could be the SR, $j = 2$ the EB, etc.)

i = dataset (e.g., i = original dataset, i+1 = adjusted dataset 1, etc.)

Revisiting the sample in Table 4, if we assume that the top 10 percent of the ranked sites are true hotspots, then Sites 428, 463, and 945 would be screened out for further investigation based on reported crash data. However, based on the underreported crash data set, Sites 215, 945, and 463 are identified as hotspots. Comparing the two HSID results reveals a 33 percent site inconsistency when crash underreporting is considered.

Underreporting Impact Evaluation Results

Establishing reliable and fair comparisons among the identification results using different datasets is paramount. One consideration in this regard is the selection of the threshold level used to establish hazardous locations. Two different levels are employed in the evaluations; the top 10 percent and 5 percent of all sites are identified as potentially hazardous. In practice the threshold level typically corresponds with the availability of resources for remediation.

The parameters of this evaluation now include levels of underreporting (15), functional classifications of roadway sections (5), and cutoff levels (top 10 percent and 5 percent). Three HSID methods are assessed, SR, EB, and FB. The evaluation criteria include the number of inconsistently identified sites and total rank differences of identified sites.

The evaluation consisted of the following steps:

- Road sections were categorized so that the safety of similar sites could be estimated. In this study five functional classifications of road section in the state of Arizona were used;
- For each classification of roadway section, 15 new datasets were created to represent the 15 different crash underreporting conditions. In total, 75 new datasets with the underreported crashes were taken into consideration;
- Each of the HSID methods is applied to both the original reported datasets and revised "underreported data sets." Road sections in each dataset are sorted in terms of their hotspot status. Specifically, the SR method ranks sites based on the total crash number, whereas the EB and FB approaches utilize the EB-estimated and FB-estimated crash numbers, respectively. Both the top 10 percent and 5 percent of all sites are selected under each method for further investigation;
- For each group of roadway sections, the two evaluation tests are conducted to compare the identification results of the original dataset and each of the 15 underreported datasets. Each test is performed 75 times; and
- For each level of crash underreporting, the evaluation results are accumulated for all the five classifications of sites under each HSID method. The impacts of the crash underreporting are then assessed based on the accumulated results.

It is important to note that when creating the underreported datasets, the sites with no crashes in the original crash dataset need to be excluded. Crash numbers for these sites remain the same, or 0, after the underreported crashes are considered. Inclusion of these sites in the analysis would diminish the accuracy of results. With these sites excluded, the numbers of sites for each functional classification of roadway are 352, 333, 310, 355, and 42, respectively. The total number of sites is 1,392.

The detailed evaluation results are shown in the following subsections.

Total Rank Differences Test Results

Table 5 summarizes the total ranking differences test results. The results are presented with levels of crash underreporting being arranged in ascending order.

As shown in Table 5, the total rank differences across various methods and underreporting levels are somewhat large in general. In the case of the top 5 percent, the maximum rank differences are 288, while in the case of the top 10 percent the maximum value reaches 832, representing 5.9 rank differences per identified hazardous locations. The relative large values indicate the strong influence of crash underreporting on HSID. In addition, under each method, the total ranking differences

Table 5: Accumulated results of total rank differences test of various methods for all classifications of highways.

	TOP (5%)			TOP (10%)		
(INJ%, PDO%)	SR	EB	FB	SR	EB	FB
(10%,20%)	14 (0.2)	108 (1.5)	228 (3.3)	42 (0.3)	560 (4.0)	752 (5.4)
(10%,30%)	14 (0.2)	149 (2.1)	240 (3.4)	54 (0.4)	703 (5.0)	784 (5.6)
(10%,40%)	29 (0.4)	193 (2.8)	243 (3.5)	108 (0.8)	732 (5.2)	810 (5.8)
(10%,50%)	41 (0.6)	209 (3.0)	279 (4.0)	149 (1.1)	752 (5.4)	819 (5.9)
(10%,60%)	57 (0.8)	216 (3.1)	288 (4.1)	174 (1.2)	808 (5.7)	832 (5.9)
(20%,30%)	14 (0.2)	137 (2.0)	237 (3.4)	38 (0.3)	575 (4.1)	779 (5.6)
(20%,40%)	24 (0.3)	164 (2.3)	242 (3.5)	68 (0.5)	719 (5.1)	790 (5.6)
(20%,50%)	38 (0.5)	200 (2.9)	268 (3.8)	122 (0.9)	731 (5.2)	801 (5.7)
(20%,60%)	44 (0.6)	206 (2.9)	271 (3.9)	140 (1.0)	774 (5.5)	797 (5.7)
(30%,40%)	16 (0.2)	160 (2.3)	237 (3.4)	63 (0.5)	658 (4.7)	770 (5.5)
(30%,50%)	31 (0.4)	190 (2.7)	250 (3.6)	100 (0.7)	673 (4.8)	791 (5.7)
(30%,60%)	37 (0.5)	203 (2.9)	263 (3.8)	115 (0.8)	744 (5.3)	793 (5.7)
(40%,50%)	30 (0.4)	186 (2.7)	245 (3.5)	74 (0.5)	662 (4.7)	789 (5.6)
(40%,60%)	25 (0.4)	205 (2.9)	263 (3.8)	102 (0.7)	703 (5.0)	795 (5.7)
(50%,60%)	32 (0.5)	205 (2.9)	263 (3.8)	86 (0.6)	706 (5.0)	803 (5.7)

Notes: 1. (INJ%, PDO%) represents the percentage of unreported injury and PDO crashes respectively;
2. The figures in the parentheses represent the average ranking difference per identified site.

significantly differ across the underreporting levels. For instance, for the FB method, the summed ranked differences range from 228 to 832, whereas the range is from 108 to 808 for the EB method. The considerable variability among the rank differences suggests that the impacts on HSID are related with the level of underreporting.

Another prominent characteristic illustrated in Table 5 is that the values of total rank differences also vary substantially among the alternative HSID methods. In comparison with EB and FB, the SR method yields smaller total ranking differences. In some cases, the values of SR could be as low as 6 to 8 times the values generated by EB and FB, respectively. This is explained by noting that when using the naïve ranking method to sort candidate sites, the underreported crash numbers merely affect each individual site and its ability to get identified. In contrast, both EB and FB methods rely also on the crash statistics of similar locations; hence the adjustment of the crash numbers on other sites would to some degree impact the ranking of each individual location and therefore lead to the much larger rank differences under the Bayes methods. Interestingly, the table also shows the big difference between FB and EB, with the FB method yielding higher total ranking differences in all situations. The phenomenon might be explained through one of the differences between the EB and FB methods noted by other researchers: With the EB method, once the SPF is developed, all the model parameters and coefficients are treated as constant values and then are incorporated into the point estimates of the long-term safety of candidate sites. However, the FB method addresses all the "uncertainty" via integrating the prior information and all available data into posterior distributions. Part of the uncertainty emerging in the crash data adjustment process might lead to the total different results between EB and FB methods.

Table 6: Accumulated results of total rank differences test of various methods for all classifications of highways.

PDO\INJ	10%	20%	30%	40%	50%
Simple Ranking					
20%	14 (42)	—	—	—	—
30%	14 (54)	13 (38)	—	—	—
40%	29 (108)	24 (68)	16 (63)	—	—
50%	41 (149)	38 (122)	31 (100)	30 (74)	—
60%	57 (174)	44 (140)	37 (115)	25 (102)	32 (86)
Empirical Bayesian					
20%	108 (560)	—	—	—	—
30%	149 (703)	137 (575)	—	—	—
40%	193 (732)	164 (719)	160 (658)	—	—
50%	209 (752)	200 (731)	190 (673)	186 (662)	—
60%	216 (808)	206 (774)	203 (744)	205 (703)	205 (706)
Full Bayesian					
20%	228 (752)	—	—	—	—
30%	240 (784)	237 (779)	—	—	—
40%	243 (810)	242 (790)	237 (770)	—	—
50%	279 (819)	268 (801)	250 (791)	245 (789)	—
60%	288 (832)	271 (797)	263 (793)	263 (795)	265 (803)

Note: 1. The figures outside the parentheses represent the test results in the case of top 5%; 2. The figures in the parentheses represent the test results in the case of top 10%.

For improved illustration of impacts of different underreporting levels, the same results of Table 5 are rearranged and shown in Table 6. In Table 6, the underreporting levels for PDO crashes increase from the top down for each HSID method, whereas for injury crashes they increase from left to right. Table 6 clearly demonstrates that the first underreporting situation (10 percent, 20 percent) generates the lowest total ranking differences in all scenarios. However, counterintuitively, the 15th underreporting condition (50 percent, 60 percent) does not produce the largest value of rank differences as we expect, even though it represents the largest underreporting levels in the study. The highest rank difference value in fact occurs in the fifth situation (10 percent, 60 percent), which represents the highest underreported PDO crashes but lowest underreported injury crashes. Further investigation of the table reveals that in general (with very few exceptional cases), the summed ranked differences increase with increasing percentage of underreported injury crashes and decrease with increasing percentage of underreported PDO crashes Therefore, the influence of underreporting in HSID is not proportional to the simultaneous increase of both underreported injured and PDO crashes. The potential reason for the trends might be due to the larger proportion of PDO crashes for most sites. When compared with injury or fatal crashes, the number of PDO crashes is the major contributor of the relatively large number of crashes (especially for top-ranked sites). Therefore, when increasing the number of PDO crashes for each site, the ranking of various sites is significantly affected. However, with the increase of underreported injury crashes, the major impact of ranking due to the underreported PDO crashes are somewhat offset.

Site Inconsistency Test Results

Tables 7 and 8 provide a summary of the site inconsistency test results. Again, the results in both tables are similar. They are presented differently with the intention to better illustrate the trends of results in various scenarios. Inspection of Table 7 shows that the maximum site inconsistency is 21, representing that 15.3 percent of the sites identified using original and adjusted data are different from each other. Both tables illustrate the similar phenomena as found in Tables 5 and 6. Compared to EB and FB methods, the SR method generates the least site inconsistency in most cases. As for the EB and FB methods, the site inconsistency associated with the latter one is generally larger than the EB approach. This indicates the underreported crashes have the largest impact on the FB method, whereas the influence on the traditional simple ranking method is least. When comparing the influence of underreporting levels on HSID, the underreporting condition (10 percent, 20 percent) under the SR method reveals the lowest values. However, the largest site inconsistency value, 21, goes to the condition of (10 percent, 60 percent) under the FB method (top 10 percent case). In general, the site inconsistency number is positively related with the percentage of the underreported PDO crashes and inversely related to the percentage of the underreported injury crashes.

It is important to note that even though similar trends are revealed in Tables 7 and 8 and Tables 5 and 6, the characteristics in Tables 7 and 8 are not so remarkable as those shown in Tables 5 and 6. The phenomenon might be due to the difference between the total rank differences and the site inconsistency tests. The former one records the shifting of all the relative positions of different hotspots, while the latter one merely counts the number of sites crossing a fixed threshold line.

Table 7: Accumulated results of site inconsistency test of various methods for all classifications of highways.

	TOP (5%)			TOP (10%)		
(INJ%, PDO%)	SR	EB	FB	SR	EB	FB
(10%, 20%)	3 (4.3)	7 (10.0)	10 (14.3)	4 (2.9)	8 (5.7)	12 (8.6)
(10%, 30%)	4 (5.7)	8 (11.4)	10 (14.3)	5 (3.6)	10 (7.1)	14 (10.0)
(10%, 40%)	4 (5.7)	8 (11.4)	11 (15.7)	5 (3.6)	11 (7.9)	16 (11.4)
(10%, 50%)	4 (5.7)	9 (12.9)	12 (17.1)	7 (5.0)	11 (7.9)	20 (14.3)
(10%, 60%)	5 (7.1)	9 (12.9)	14 (20.0)	7 (5.0)	13 (9.3)	21 (15.0)
(20%, 30%)	3 (4.3)	7 (10.0)	11 (15.7)	5 (3.6)	10 (7.1)	12 (8.6)
(20%, 40%)	3 (4.3)	8 (11.4)	11 (15.7)	5 (3.6)	10 (7.1)	16 (11.4)
(20%, 50%)	4 (5.7)	8 (11.4)	11 (15.7)	6 (4.3)	11 (7.9)	16 (11.4)
(20%, 60%)	4 (5.7)	8 (11.4)	14 (20.0)	7 (5.0)	11 (7.9)	16 (11.4)
(30%, 40%)	3 (4.3)	6 (8.6)	11 (15.7)	4 (2.9)	8 (5.7)	13 (9.3)
(30%, 50%)	4 (5.7)	6 (8.6)	11 (15.7)	6 (4.3)	9 (6.4)	13 (9.3)
(30%, 60%)	4 (5.7)	7 (10.0)	11 (15.7)	7 (5.0)	9 (6.4)	13 (9.3)
(40%, 50%)	4 (5.7)	5 (7.1)	11 (15.7)	5 (3.6)	8 (5.7)	12 (8.6)
(40%, 60%)	4 (5.7)	7 (10.0)	11 (15.7)	6 (4.3)	8 (5.7)	13 (9.3)
(50%, 60%)	3 (4.3)	6 (8.6)	12 (17.1)	6 (4.3)	8 (5.7)	13 (9.3)

Notes: 1. (INJ%, PDO%) represents the percentage of unreported injury and PDO crashes respectively;
2. The figures in the parentheses represent the percentage of inconsistent sites of the total identified sites.

Table 8: Accumulated results of site inconsistency test of various methods for all classifications of highways.

PDO\INJ	10%	20%	30%	40%	50%
Simple Ranking					
20%	3 (4)	—	—	—	—
30%	4 (5)	3 (5)	—	—	—
40%	4 (5)	3 (5)	3 (4)	—	—
50%	4 (7)	4 (6)	4 (6)	4 (5)	—
60%	5 (7)	4 (7)	4 (7)	4 (6)	3 (6)
Empirical Bayesian					
20%	7 (8)	—	—	—	—
30%	8 (10)	7 (10)	—	—	—
40%	8 (11)	8 (10)	6 (8)	—	—
50%	9 (11)	8 (11)	7 (9)	5 (8)	—
60%	9 (13)	8 (11)	7 (9)	7 (8)	6 (8)
Full Bayesian					
20%	10 (12)	—	—	—	—
30%	10 (14)	11 (12)	—	—	—
40%	11 (16)	11 (16)	11 (13)	—	—
50%	12 (20)	11 (16)	11 (13)	11 (12)	—
60%	14 (21)	14 (16)	12 (13)	11 (13)	12 (13)

Note: 1. The figures outside the parentheses represent the test results in the case of top 5%; 2. The figures in the parentheses represent the test results in the case of top 10%.

Conclusions and Recommendations

This research is focused on evaluating the impact of crash underreporting on the performance of HSID. An experimental evaluation was designed to examine four functional classifications of roadway sections, 15 levels of crash underreporting, two hotspot cutoff levels, three HSID methods, and two evaluation criteria. The study has yielded some noteworthy insights, leading to important conclusions:

- Under both evaluation tests, the results indicate that underreported crashes have the greatest impact to the FB method, whereas the influence to the traditional SR method is least significant;
- In most cases, the underreporting condition of (10 percent, 20 percent) shows the least impact to HSID, whereas the underreporting condition of (10 percent, 60 percent) exhibits the strongest influence. In general, the influence is positively related with the percentage of the underreported PDO crashes, but inversely related to the percentage of the underreported injury crashes; and
- Overall, the identification bias due to crash underreporting may be significant. Specially, under the FB method, if 10 percent of injury crashes are underreported and 60 percent of PDO crashes are underreported, 20 percent of sites identified as most risky will be incorrect.

Due to the significant impact exhibited by crash underreporting, it is recommended that underreported crashes be taken into consideration when conducting HSID to the extent that data

are available. At the very least, and in cases when the extent of crash underreporting is unknown, a sensitivity analysis of safety management activities involving modeling and analysis should be undertaken to acknowledge the extent of the error introduced by underreporting.

The analysis methods touted as most robust and capable of dealing with regression to the mean effects are the most affected by underreported crash data. This presents opposing motivations for using or avoiding advanced HSID methods, a quandary that deserves additional attention and quantification in future work.

It is hoped that the research findings of this study can further increase awareness among safety professionals regarding the potential negative influences of crash underreporting. HSID is not the only methodology that is affected, and other safety management analytical techniques may be subject to large negative influences as well.

There are limitations of this study worth repeating. First, even though a relative large number of different levels of underreporting were applied in the study, other underreported percentages are also possible. Second, a range of possible SPFs were not considered, say, ones with only ADT as a predictor and others will full complement of predictors. It is unknown how the range of possible SPFs might affect the outcomes presented here. Finally, other metrics of performance could be used to assess the performance, as described in Cheng and Washington.[39]

REFERENCES

1. K. W. Ogden, *Safer Roads: A Guide to Road Safety Engineering*, Averbury Technical Press, Ashgate Publishers, 1996.

2. Salifu, M. and W. Ackaah, *Underreporting of Road Traffic Crash Data in Ghana*, proceeding of the 4th IRTAD Conference, Seoul, Korea, 2009.

3. Hvoslef, H. Underreporting of Crashes Recorded by the Police, at the International Level. Paper prepared for IRTAD, OECD, Paris, France, 1994.

4. Amosros, E., J.-L. Martin, B. Laumon. Underreporting of road crash causalities in France, *Accident Analysis and Prevention*, 38, pp. 627–635, 2006.

5. Hauer, E. and A. S. Hakkert. Extent and Some Implications of Incomplete Accident Reporting. In *Transportation Research Record 1185*, TRB, National Research Council, Washington, DC, pp. 1-10. 1988.

6. Elvik, R. and A. B. Mysen. Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. In *Transportation Research Record 1665*, pp. 133–140, 1999.

7. Blincoe L., A. Seay, E. Zaloshnja, T. Miller, E. Romano, S. Luchter, and R. Sicer. *The Economic Impact of Motor Vehicle Crashes*, 2000. U.S. Department of Transportation, National Highway Traffic Safety Administration. 2002.

8. Sposito, B. and S. Johnston, *Three-Cable Median Barrier Final Report* (Report No. OR-RD-99-03). Oregon Department of Transportation, Research Unit, Salem, OR. 1998.

9. Zegeer, C., H. Huang, J. Stewart, R. Pfefer, and J. Wang. Effects of a Towaway Reporting Threshold on Crash Analysis Results. In *Transportation Research Record 1635*, TRB, National Research Council, Washington, DC, pp. 15–21, 1998.

10. Yamamoto, E., J. Hashiji, and V. Shankar. Underreporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models, *Accident Analysis and Prevention*, 40, pp. 1320–1329, 2008.

11. Oh, Jutaek, S. Washington, and D. Lee, (2010). Property Damage Crash Equivalency Factors for Solving the Crash-Severity Dilemma: Case Study on South Korean Rural Roads. *Transportation Research Record 2148* (2010), pp. 83–92.

12. Deacon, J. A., C. V. Zegeer, and R. C. Deen. Identification of Hazardous Rural Highway Locations. In *Transportation Research Record 543*, TRB, National Research Council, Washington, DC, 1975, pp. 16–33.

13. Norden, M., J. Orlansky, and H. Jacobs. *Application of Statistical Quality-Control Techniques to Analysis of Highway-Accident Data*. Bulletin 117, HRB, National Research Council, Washington, DC, 1956, pp. 17–31.

14. Rudy, B. M. *Operational Route Analysis.* Bulletin 341, HRB, National Research Council, Washington, DC, 1962, pp. 1–17.

15. Morin, D. A. Application of Statistical Concepts to Accident Data. *Highway Research Record 188*, HRB, National Research Council, Washington, DC, 1967, pp. 72–79.

16. Stokes, R. W. and M. I. Mutabazi. Rate-Quality Control Method of Identifying Hazardous Road Locations. *Transportation Research Record 1542*, TRB, National Research Council, Washington, DC, 1996, pp. 44–48.

17. Laughland, J. C., L. E. Haefner, J. W. Hall, and D. R. Clough. *NCHRP Report 162: Methods for Evaluating Highway Safety Improvements*. TRB, National Research Council, Washington, DC, 1975.

18. Hakkert, A. S. and D. Mahalel. Estimating the Number of Accidents at Intersections from a Known Traffic Flow on the Approaches. *Accident Analysis and Prevention*, Vol. 10, No. 1, 1978, pp. 69–79.

19. McGuigan, D. R. D. The Use of Relationships between Road Accidents and Traffic Flow in "Black-Spot" Identification. *Traffic Engineering and Control*, Aug.–Sept. 1981, pp. 448–453.

20. McGuigan, D. R. D. Nonjunction Accident Rates and their Use in "Black-Spot" Identification. *Traffic Engineering and Control*, Feb. 1982, pp. 45–56.

21. Persaud, B. N., C. Lyon, and T. Nguyen. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1665, TRB, National Research Council, Washington, DC, 1999, pp. 7–12.

22. Hauer, E. Bias-by-Selection: Overestimation of the Effectiveness of Safety Countermeasures Caused by the Process of Selection for Treatment. *Accident Analysis and Prevention*, Vol. 12, No. 2, 1980, pp. 113–117.

23. Abbess, C., D. Jarret, and C. C. Wright. Accidents at Blackspots: Estimating the Effectiveness of Remedial Treatment, with Special Reference to the "Regression-to-the-Mean" Effect. *Traffic Engineering and Control*, Vol. 22, No. 10, 1981, pp. 535–542.

24. Hauer, E. On the Estimation of the Expected Number of Accidents. *Accident Analysis and Prevention*, Vol. 18, No. 1, 1986, pp. 1–12.

25. Hauer, E. and B. N. Persaud. How to Estimate the Safety of Rail-Highway Grade Crossings and the Safety Effects of Warning Devices. *Transportation Research Record 1114*, TRB, National Research Council, Washington, DC, 1987, pp. 131–140.

26. Hauer, E., J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. *Transportation Research Record* 1185, TRB, National Research Council, Washington, DC, 1988, pp. 48–61.

27. Hauer, E., B. N. Persaud, A. Smiley, and D. Duncan. Estimating the Accident Potential of an Ontario Driver. *Accident Analysis and Prevention*, Vol. 23, No. 2–3, 1991, pp. 133–152.

28. Higle, J. L. and J. M. Witkowski. Bayesian Identification of Hazardous Locations. *Transportation Research Record 1185*, TRB, National Research Council, Washington, DC, 1988, pp. 24–36.

29. Persaud, B. and C. Lyon. Empirical Bayes before–after safety studies: lessons learned from two decades of experience and future directions. *Accident Analysis & Prevention* 39 (3), 2007, pp. 546–555.

30. Hauer, E., D. W. Harwood, F. M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, Transportation Research Board of the National Academies, Washington, DC, 2002, pp. 126–131.

31. Davis, G. A., S. Yang, Bayesian Identification of High-risk Intersections for Older Drivers via Gibbs Sampling. *Transportation Research Record 1746*, 2001. pp. 84–89.

32. Aguero-Valverde, J., Jovanis. P.P. *Bayesian Multivariate Poisson Log-normal Models for Crash Severity Modeling and Site Ranking*. Transportation Research Board 88th Annual Meeting. 2009.

33. Lan, B., B. Persaud, C. Lyon, and R. Bhim, Validation of a Full Bayes Methodology for Observational Before–after Road Safety Studies and Application to Evaluation of Rural Signal Conversions. *Accident Analysis & Prevention* 41 (3), 2009. pp. 574–580.

34. Washington, S. and J. Oh. Bayesian Methodology Incorporating Expert Judgment for Ranking Countermeasures Effectiveness under Uncertainty: Example Applied to at Grade Railroad Crossings in Korea. *Accident Analysis & Prevention* 38, 2006, pp. 234–247.

35. Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2003. *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit, Cambridge, MA. http://www.mrc-cam.ac.uk/bugs.

36. Higle, J. L. and M. B. Hecht. A Comparison of Techniques for the Identification of Hazardous Locations. *Transportation Research Record 1238*, TRB, National Research Council, Washington, DC, 1989, pp. 10–19.

37. Maher, M. J. and L. J. Mountain. The Identification of Accident Blackspots: A Comparison of Current Methods. *Accident Analysis and Prevention*, Vol. 20, No. 2, 1988, pp. 143–151.

38. Cheng, W. and S. Washington. Experimental Evaluation of Hotspot Identification Methods. *Accident Analysis and Prevention*, Vol. 37, No. 5, 2005, pp. 870–881.

39. Cheng, W. and S. Washington. New Criteria for Evaluating Methods of Identifying Hot Spots. *Transportation Research Record 2083*, TRB, National Research Council, Washington, DC, 2008, pp. 76–85.

40. Huang, H., H.C. Chin and M. Haque. *Empirical Evaluations of Alternative Approaches in Identifying Crash Hotspots: Naive Ranking, Empirical Bayes and Full Bayes*. TRB 2009 Annual Meeting CD-ROM, National Research Council, Washington, DC, 2009.

41. Persaud, B., B. Lan, C. Lyon and R. Bhim. Comparison of Empirical Bayes and Full Bayes Approaches for Before-After Road Safety Evaluations. *Accident Analysis & Prevention* 42, 2010, pp. 38–43.

42. Miranda-Moreno, L.F. and L. Fu. *Traffic Safety Study: Empirical Bayes or Full Bayes?* The 86th Annual Meeting of the Transportation Research Board, TRB 2007. Paper 07-1680.

43. Miaou, S. and D. Lord. 2003. Modeling Traffic Crash-flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods. Transportation Research Record, pp.31–40.

44. Pawlovich, M.D., W. Li, A. Carriquiry and T. Welch. Iowa's Experience with "Road Diet" Measures: Impacts on Crash Frequencies and Crash Rates Assessed Following a Bayesian Approach. *Transportation Research Record: Journal of the Transportation Research Board 1953*, TRB, National Research Council, Washington, DC 2006, pp.163–171.

45. Congdon, P. *Bayesian Models for Categorical Data*. John Wiley & Sons, Ltd. 2005.

46. Gilks, W. R., S. Richardson and D.J. Spiegelhalter. *Markov Chain Monte Carlo Methods in Practice*. Chapman & Hall, New York. 1995.

47. Hauer, E. and B. N. Persaud. Problem of Identifying Hazardous Locations Using Accident Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 975, TRB, National Research Council, Washington, DC, 1984, pp. 36-43.

48. Elvik, R. *State-of-the-art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks*. Report 883. Institute of Transport Economics, Oslo, Norway. 2007.

AARON TRUONG is currently a graduate student in the Civil Engineering Department at California State Polytechnic University, Pomona. He is also a transportation engineer at Parsons Brinckerhoff in Orange, California, USA, where he is involved in a wide range of projects related to transportation planning, signal operations, traffic simulation and traffic impact, and corridor studies. He is a member of ITE.

GIOVANNI A. BRYDEN JR. is a student researcher in the Civil Engineering Department at California State Polytechnic University, Pomona. His research interests include road safety and the effect of pedestrians and bicyclists on traffic flow characteristics. As a cycling enthusiast, he hopes to develop ways to reduce the vulnerability of bicyclists on shared road networks.

WEN CHENG, Ph.D., P.E., PTOE is an assistant professor in the Department of Civil Engineering at California State Polytechnic University, Pomona. He received his B.S. from Tongji University, Shanghai, China, an M.S. from the University of Arizona, and a Ph.D. from Arizona State University. His research focuses on highway safety, statistical modeling, traffic signal design, and traffic simulation. He is a member of ITE.

XUDONG JIA, Ph.D., P.E., is a professor in the Department of Civil Engineering at California State Polytechnic University, Pomona. He received his B.S. and M.S. from Northern Jiaotong University, Beijing, China, an M.A.Sc. from the University of Toronto, and a Ph.D. from the Georgia Institute of Technology. His research focuses on ITS, GIS, highway design, and traffic safety.

SIMON WASHINGTON, Ph.D., is a member of the faculty of Built Environment and Engineering Centre for Accident Research and Road Safety (CARRS-Q) and of the faculty of Health Queensland University of Technology (QUT). He also serves as the Queensland Transport and Main Roads (TMR) Chair of Transport. Prior to joining QUT, Professor Washington was director of the Safe Transportation Research & Education Center at the University of California, Berkeley.

Evaluating Global Freight Corridor Performance for Canada

By William L. Eisele, Ph.D., P.E., Louis-Paul Tardif, Juan C. Villa, David L. Schrank, Ph.D., and Tim Lomax, Ph.D., P.E.

As part of Transport Canada's Gateways and Trade Corridors Initiative, the Directorate of Economic Analysis was interested in developing freight performance measurements for goods using Canada's international gateways and traveling along its freight transportation corridors. These performance indicators—termed "fluidity" measures—will assist Transport Canada in painting a clear picture of system efficiency for its freight-significant corridors. The indicators will ultimately aid Transport Canada in identifying to what extent the government of Canada's policies and investment in infrastructure are being leveraged and operated to support trade and economic prosperity.

Abstract

Transport Canada contracted with the Texas Transportation Institute (TTI) to develop and apply the indicators for measuring freight system performance. Researchers created two "fluidity indicators" using an index approach. One indicator captures average conditions (Fluidity Index), while the other indicator captures daily variation in travel time (Planning Time Index). Because freight moves according to both travel time and delivery requirement schedules, and because travel time varies according to mode, the performance measures use a normalizing concept to allow comparisons within a mode and across an entire supply chain.

This paper describes the development and application of the measures. The paper includes two applications. One application demonstrates how the fluidity measures are computed and presented for truck shipments. In the second application, researchers demonstrate the use of the fluidity

measures for monitoring freight system performance for an international and multimodal corridor from China to Canada.

The measures, application, and findings documented in this paper are valuable for practitioners and freight movement stakeholders interested in monitoring freight system efficiency.

Introduction

Global Trade Trends

Over the last few decades, the global economy has changed dramatically. To meet increasing consumer demand, China, India, and other Asian nations have become major manufacturing centers, producing many of the products consumed around the world. The establishment of the North American Free Trade Agreement (NAFTA) and the European Union (EU) has aided in creating a more efficient means of transporting goods across international boundaries in both North America and Europe. Accordingly, world trade has increased substantially over this same time period.

Canadian International Freight Flows

As the value of international trade on a global scale has grown, so too has the value of imports and exports in Canada. In 2008, the World Trade Organization reported that Canadian imports totaled more than $418 billion, while exports reached in excess of $456 billion.[1]

Increasing trans-Pacific trade volumes is one of the reasons why Canada has experienced substantial growth in both imports and exports over the last decade. In 2008, China ranked second ($42.6 billion) and fourth ($10.6 billion) in terms of Canada's total imports and exports, respectively.[1] From 2001 to 2006, Canada's exports and imports with China showed an average annual growth rate of 12 percent and 22 percent, respectively. Between 1996 and 2006, marine exports to China nearly tripled to reach $7 billion, while marine imports grew four times their 1996 level to approximately $15 billion.[2]

Cargo volumes have risen substantially over the past decade at all of the major ports on North America's Pacific Coast. The Port of Vancouver, located in British Columbia, is no exception to this trend. Over the last 10 years, throughput at the Port of Vancouver has more than doubled. According to the American Association of Port Authorities (http://aapa-ports.org/), in 2008 nearly 2.5 million 20-foot equivalent units (TEUs) passed through the Port of Vancouver.

Taking advantage of increasing trans-Pacific container volumes, the Port of Prince Rupert began container operations in 2007. This port, located approximately 480 miles north of the Port of Vancouver, is uniquely situated to provide the fastest marine transport times between Asia and North America. Container throughput at the Port of Prince Rupert has also risen dramatically since container operations began in 2007. In 2009, the Port of Prince Rupert handled more than 265,000 TEUs, a 46 percent increase over 2008 volumes.[3]

The Role of Supply Chains

As trans-Pacific trade volumes continue to grow, the role that freight transportation plays in Canada's economy will become more and more important. In 2008, commercial transportation services alone accounted for 4.1 percent of Canada's value-added gross domestic product (GDP).[1] Because freight transportation is such an integral part of Canada's economy, it has become necessary for the Canadian government to formulate policies and make infrastructure improvements to facilitate the seamless movement of goods to and from the country (particularly to and from Asia and Europe)

and within the country itself. Additionally, the need to facilitate trade is amplified by the fact that ports such as Los Angeles and Long Beach are approaching or exceeding existing capacity, placing more strain on North America's transportation system.

As international trade expands, supply chains become increasingly complex. Because Canada's population is highly concentrated in the provinces of Quebec and Ontario, a good majority of the freight entering Canada through its West Coast ports will ultimately need to be distributed to those two provinces.

There are two main options for transporting freight from the Port of Vancouver to the Toronto metro area: truck and rail. In 2006, 81 percent of loaded imports (approximately 662,000 TEUs) entering the Port of Vancouver were transported via railroad to Montreal (Quebec) and Toronto (Ontario) metro areas.[4] As trans-Pacific trade continues to increase, these numbers will only rise, placing even more importance on this freight corridor.

Domestic Canadian Freight Flows

In 2007, an estimated 1.2 billion metric tons of freight were transported within Canada's geographic borders.[5] Of that total, approximately 530 million metric tons of freight were moved via truck, 225 million metric tons were moved via rail, 67.3 million metric tons were moved by water, and 0.5 million metric tons were moved via air transport. Pipelines transported an additional 380 metric tons.

Geographically, a large majority of the freight entering or exiting Canada moves through the provinces of Ontario, Quebec, and British Columbia. These provinces accounted for approximately $443 billion, $117 billion, and $85 billion, respectively, in 2006.

Most of the freight exiting Canada through both Ontario and Quebec is destined to the United States. A substantial portion of the trade entering Canada through these two provinces also comes from the United States, largely due to the automotive industry in Detroit, Michigan. Because the province of British Columbia is the third-largest province for international trade, and the imports and exports are distributed much more evenly between U.S. and non-U.S. trading partners, this paper focuses on applying the indicators to freight moving through British Columbia.

Basic Concepts of the Fluidity Indicators

As part of Transport Canada's Gateways and Trade Corridors Initiative, the Directorate of Economic Analysis was interested in developing freight performance measurements for goods using Canada's international gateways and traveling along its freight transportation corridors. The primary objective of the research described here was to develop performance indicators to assist Transport Canada in painting a clear picture of freight system efficiency for its freight-significant corridors without compromising the sensitive nature of private industry data.

These indicators will ultimately aid Transport Canada in identifying to what extent the government of Canada's policies and investment in infrastructure are being leveraged and operated to support trade and economic prosperity. The measures should also be relevant to private sector operators and their investment and operating decisions. The idea behind this initiative is that if one cannot systematically measure the efficiency of Canada's supply chain system, it cannot be improved. This proactive approach is designed to stay ahead of the curve in a world where business practices change at an alarmingly fast rate.

Understanding the economic impact of freight transportation on Canada's economy, Transport Canada gathers a large amount of detailed freight transportation data on a regular basis. Researchers applied these data to create the Fluidity Index and Planning Time Index described in this paper.

Conceptually, the Fluidity Index and Planning Time Index will serve as a gauge that measures the overall efficiency of Canada's transportation system at a corridor level. Specific objectives of the measures include the following:

- Providing transparent reliability information pertaining to every one of Canada's freight-significant corridors;
- Identifying bottlenecks in the Canadian logistics system; and
- Ensuring Canada's overall competitiveness in the global marketplace.

These objectives are met using a set of principles that are applicable for all freight modes. Because the available data for this paper primarily consisted of truck freight data, the paper principally uses that mode to illustrate the concepts, but the measures were developed considering aspects of all travel modes. A more detailed explanation of the measures and calculation methods is included in the next section, but the basic elements are described below:

- An index-based approach is being used to calculate the fluidity indicators rather than simply using transit times. Travel times are important and very useful in many contexts. It is difficult, however, to compare conditions in corridors of varying lengths by only using travel times. An index that removes the effect of trip lengths allows for easier comparisons between corridors and within corridor sections and leads to better understanding about the location and timing of problems;
- The index concept meets the needs of both technical and nontechnical audiences and users with a small amount of explanation, but communication must accompany the performance measures. Something as simple as "an index value of 1.30 indicates a 20-hour trip by the fastest group of drivers on uncongested roads will require 26 hours by slower drivers or when considering congestion" have been used to explain indexes;
- The indexes use "rapid" trips as the comparison benchmark. This recognizes that ship, air, rail, and truck travel times are going to be different and should not be compared to each other. It also allows geographic and topographic constraints to be explicitly included, thereby highlighting the operational and infrastructure effects;
- The benchmark can also be developed with a goal of comparing similar populations to each other. For example, differentiating between the team-driver approach used with some urgent cargo and the regular single driver may allow a better, more accurate story to be told about the corridor travel conditions;
- Travel time is not enough. To understand the meaning of the fluidity indicators, analysts must know the "whys" associated with travel time. For example, is the slow travel time in an area caused by traffic congestion or by many trucks leaving the main highway to access a truck stop?
- Consistent definitions are important for accurate understanding; in some cases this will necessitate discussions between Transport Canada, operators, and data providers to resolve or understand the differences; and
- The application of the fluidity measures will likely evolve as more specific data are gathered, as more questions are asked, and as more issues are identified. The keys in the first phase are to show that the measures produce reasonable and understandable results. A few specific uses are depicted in this paper, but these will grow as the data and measures are more widely disseminated. Using the measures always has the effect of improving the data quality and targeting additional data collection on important elements, features, or policies.

Proposed Fluidity Indicators

It follows from the previous discussion that information about both the average travel conditions and the reliability of the container trips are important. The reliability is especially important for those trips of most interest—the high-priority trips. To characterize the transportation system in Canada, TTI proposed two fluidity indicators for Transport Canada—the Fluidity Index and the Planning Time Index.

A form of these measures has been used in TTI's *Urban Mobility Report* and the U.S. Department of Transportation Federal Highway Administration's *Urban Congestion Report* for communicating travel time mobility and reliability.[6, 7] A form of these measures is also described more fully in *The Keys to Estimating Mobility in Urban Areas* as it relates to passenger car mobility and reliability estimation.[8]

Fluidity Index

The Fluidity Index (FI) is a dimensionless quantity that compares the average travel time in the time period of interest to travel time during unconstrained or free-flow conditions (see Equation 1). Free-flow conditions might be represented with the low end of the quick trips, such as the 5th percentile travel time. For example, an FI of 1.50 indicates that a shipment that takes 2 days during free-flow conditions may take 3 days (2 days x 1.50) during the time period of interest. The FI provides an assessment of the average travel conditions of a freight shipment.

$$\text{Fluidity Index} = \frac{\text{Average Travel Time (days or minutes)}}{\text{Freeflow Travel Time (days or minutes)}} \quad (1)$$

Planning Time Index

The Planning Time Index (PTI) addresses the need for certain high-priority urgent shipments to be on time most of the time. It represents the total travel time that should be planned to ensure on-time delivery. It compares near-worst-case travel time to a travel time in free-flow conditions (see Equation 2). For example, a Planning Time Index of 2.00 means that for a two-day trip in light traffic, the total time that should be planned for the trip is 4 days (2 days x 2.00 = 4 days). The Planning Time Index is useful because it can be directly compared to the Fluidity Index on a similar numeric scale. The PTI provides an assessment of the travel time reliability of a freight shipment.

$$\text{Planning Time Index} = \frac{\text{95th Percentile Travel Time (days or minutes)}}{\text{Freeflow Travel Time (days or minutes)}} \quad (2)$$

Comparison Baseline

Using an index approach requires that a baseline travel time be determined. Sometimes this is described as free-flow or low-volume conditions. Basically, the concept is to identify a relatively fast travel time that can be used to judge the times from a group with similar characteristics. In this research, researchers defined the free-flow travel time as the 5th percentile travel time for a given trip type. The 5th percentile travel time is one of the shortest travel times, thus providing a good benchmark for freight shipments.

Form and Application of Fluidity Indicators

The two fluidity indicators are ratios of time (typically days or minutes depending upon the trip length). These two fluidity indicators are valuable because they are scalable for analysis at the following levels:

- **Trip priority:** High-priority container trips can be separated out and compared to all trips or trips during uncongested travel periods. FI and PTI values can be calculated for each trip priority.
- **Location:** Container trips at different locations along the supply chain can be assessed as well as trips along different adjacent corridors. Both indicators can be used for corridor travel as well as travel within a "node" (e.g., terminal, port).
- **Mode:** Container trips by a single mode or multiple modes can be compared.
- **Time:** Trips by start time of day or start day of week can be analyzed and compared.
- **Commodity:** If desirable and/or available, the fluidity indicators can be computed and analyzed at the commodity level.

Shippers and receivers (consignees) are not always in immediate need for the goods being transported. The range of travel times, in many cases, is caused by this difference in "urgency."

Incorporating Trip Urgency

Shippers and receivers (consignees) are not always in immediate need for the goods being transported. The range of travel times, in many cases, is caused by this difference in "urgency." The use of indexes allows the analyst to evaluate different freight shipment groups relative to free-flow for each shipment group. For example, it will be discussed further in later sections of this paper that the roadway travel time data suggest three groups of freight shipments.

- **Team drivers:** More than one driver in a truck can achieve relatively low travel times.
- **Single driver:** One driver must make stops to stay within the hours of service regulations.
- **Low-priority shipments:** Longer travel times can indicate delivery is not a high priority.

A later application in this paper demonstrates how individual Fluidity Indexes and Planning Time Indexes are computed for each of these freight shipment groups.

Application to the Supply Chain

Analysts can compute an FI and PTI for each "link" (e.g., roadway for truck or rail line for rail) and/or "node" (e.g., ports, terminals, or distribution centers) of a supply chain. An aggregate FI or PTI for the entire trip can be computed by weighting the TEU hours at each step of the supply chain — the amount of time spent traveling the segment by 20-foot equivalent containers. This is illustrated in an application at the end of this paper.

The interested reader is encouraged to review prior research that has demonstrated a conceptual framework for freight mobility, along with a similar methodology to estimate tonnage and dollar amount of commodities by weighting with traffic volume.[9]

Available Data

Desired Information for Decision Making

The desired levels of information for decision making for the FI and PTI can be summarized at several levels as identified by Transport Canada:

- **Level One**—shows travel time in days from a foreign port to its final destination in Canada.
- **Level Two**—shows the travel times split into the various modes required to complete the trip with each of the modes reported in days.
- **Level Three**—shows even greater detail for each mode.
- **Level Four**—the most disaggregated level, shows the times in hours and includes the average times as well as several percentiles that are used in calculating the Fluidity Index and the Planning Time Index.

The audience for each of these levels of information is different. Policymakers may use Level One or Level Two information, while transportation planners and engineers may concern themselves with the causes for delays that are highlighted in Level Three or Level Four data.

Researchers used the most disaggregated level possible to show the most detail. Only the truck data provided by Transport Canada had enough detail to analyze at Level Four. Data for the remaining modes generally could only be reported at Level Two because only average travel times by mode were available. For the application to a supply chain presented at the end of this paper, researchers created sample data to demonstrate the use of the FI and PTI for the supply chain.

Available Truck Data

The truck freight and operational data were provided by Transport Canada in two datasets—dispatch and global positioning system (GPS). Due to the anonymous nature of the data, it is not possible to link the truck-to-truck data included in these two datasets.

Dispatch Truck Data

The dispatch data contains information about specific truck trips. This database serves as a log of each truck trip and includes the following:

- Trip origin—address, date and time, GPS position;
- Trip destination—address, date and time, GPS position;
- Number of pieces of freight;
- Total cargo weight;
- Distance traveled;
- Type of trailer(s) used; and
- Whether it is a long-haul or short-haul trip.

The dispatch data include tremendous details about the truck trip endpoints. It includes details pertaining to the beginning and ending point of the trip as well as how much is being hauled and what type of trailer is used for the trip. It does not provide any specific information on the actual path taken for the individual trip.

The truck dispatch data file provided by Transport Canada to TTI had a sample of about 3.5 million dispatch records for calendar years 2008 and 2009. The data included trip origins and destinations from all of Canada. Because the corridor analyzed was Canada Highway 1, which runs from Vancouver, British Columbia to Toronto, Ontario, the dispatch records needed to be narrowed to focus only on this corridor. In many cases, the city listed in the database was not one of these central cities such as Toronto but was a suburb of the metropolitan area (e.g., Mississauga, which is a suburb of Toronto). The project team included these "suburb cities" within the metropolitan area as part of the central area.

GPS Truck Data

Trucking companies enter into agreements with GPS companies to help track the location of their fleet vehicles. The GPS companies determine the location of a truck at regular time intervals that are agreed upon with the trucking company. In addition to these regular intervals, the truck driver will send location information when the driver is making an operational change to his or her situation, such as stopping for fuel or making a delivery. The GPS company provides the following data every time a truck's position is ascertained:

- Truck identification code;
- Date and time stamp; and
- GPS position including longitude and latitude.

Researchers estimated travel time from the individual truck GPS readings. To accomplish this task there were several issues with the data:

- The initial truck GPS datasets consisted of about 450 million truck records, which presented several challenges with processing the information. A data record occurs every time a signal is sent to the satellite showing the truck's location;
- The truck GPS data consisted of truck geographic location details. However, the latitude and longitude were not linearly referenced to the route on which the truck was traveling. Lack of linear referencing made it difficult to find the exact distance traveled by a truck between two time periods along the route. Algorithms were developed to "match" these GPS points to points on the roadway network; and
- The format of the date field and truck ID field were incompatible with certain software, such as ArcGIS 9.3, which were used for linear referencing of the truck data. When the truck GPS data were opened in ArcGIS, the truck IDs were read as a scientific format and were truncated.

Working with 450 million truck records was overwhelming. To have a more manageable dataset, the data from October 2008 were chosen as the sample to be analyzed. The initial 35 million truck records from this dataset were scattered on several different routes all over Canada. The data were processed using SAS software to filter out the unwanted data that were not on the corridor. Next, researchers linearly referenced the truck GPS data records to get measurements between two truck GPS data points along a route. ArcGIS 9.3 software was used to achieve this task.

Combining Data Sources

Multiple datasets can provide better explanations of why travel times are distributed the way they are. However, evaluating different data sources requires an understanding of the original use and preparation of the data. For example, truck travel time distributions using the truck dispatch data often show a "mini-peak" at 24 hours due to a large number of drivers checking in as "available" to

the dispatchers the day after a delivery. Such "mini-peaks" due to driver availability do not appear in the real-time GPS data because the data are collected differently. This is demonstrated with an application in the next section of this paper.

Application: Truck Shipments

The application that follows describes the use of the truck data to estimate the Fluidity Index and the Planning Time Index. *The index values shown in the tables and example applications that follow are only meant to demonstrate the methodology and how the fluidity indicators can be calculated and presented. These results are not intended for system monitoring at this time.*

Investigating the truck dispatch data, researchers found that there are three types (groups) of truck trips occurring along the Vancouver to Toronto corridor (see Table 1). Researchers assumed that these three groups of truck trip types are team driving, solo driving, and low-priority shipments. Team driving, where multiple drivers are assigned to one truck, is utilized in the case of a critical time delivery. The trailer contents must reach a destination in a shorter time period than one truck driver can safely and legally drive due to hours-of-service constraints of 10-hour maximum driving time without rest. Solo driving occurs when the delivery time is not quite as constrained and a single driver can safely and legally get the trailer to its destination in the required amount of time. "Low-priority" travel has the longest travel times of the three types. This type of driving occurs when the delivery date is set out in the future longer than it would take for a single driver to safely make the trip.

Table 1 shows the shortest and longest travel times by trip type for each city pair. Travel times include influences of congestion, roadway length, roadway geometry, fuel station locations, and so forth.

A statistical analysis was performed on each of these trip types for each city pair. Several statistics were calculated including the average travel time as well as the 5th, 10th, and 95th percentile travel times. As described previously, the 5th percentile for each trip type was used as the free-flow travel time for that trip type. The average travel time is divided by the 5th percentile travel time to calculate the Fluidity Index for each trip type (per Equation 1).

Table 1: Shortest and longest travel times by city pair and trip type (adapted from reference 10).

		Team Driving (Hours)		Solo Driving (Hours)		Low-Priority (Hours)	
City Pair	Direction	Shortest	Longest	Shortest	Longest	Shortest	Longest
Vancouver-Toronto	Eastbound	48	90	91	144	145	192
Vancouver-Calgary	Eastbound	11	18	19	31	32	48
Calgary-Regina	Eastbound	7	16	17	24	25	40
Regina-Winnipeg	Eastbound	4	10	11	19	20	34
Winnipeg-Toronto	Eastbound	24	30	31	50	51	72
Toronto-Vancouver	Westbound	40	90	91	144	145	220
Calgary-Vancouver	Westbound	11	15	16	28	29	48
Regina-Calgary	Westbound	6	12	13	22	23	38
Winnipeg-Regina	Westbound	4	10	11	19	20	32
Toronto-Winnipeg	Westbound	24	38	39	52	53	72

The 95th percentile travel time shows one of the longest travel times and slowest overall speeds to complete the trip. There can be several causes for this longer trip length, including congestion on the roadway, poor weather conditions, or tourist travel. The 95th percentile travel time is divided by the 5th percentile travel time in the calculation of the Planning Time Index for each trip type (per Equation 2).

The truck dispatch does not contain enough detail about each trip to identify where bottlenecks may exist in the transportation network. The data can demonstrate the volatility associated with trips over longer distances. For a more detailed analysis at a more local level, the GPS truck data provides the necessary detail.

For the analysis findings that follow, it should be noted that the truck dispatch data were extensive enough to provide all of the necessary information to calculate the FI and PTI. With the other modes, only the average travel time values were provided. For these other modes, the FI and PTI could not be calculated, so the average travel time was reported alone without the FI and PTI.

In many instances there were not sufficient data to calculate the index values at the monthly level. Researchers found that the dispatch data had a great number of erroneous entries. Selected reasons for erroneous entries include those with very fast or very slow travel times, data entered into the incorrect field, or situations where the origin or destination could not be identified. Such records were removed, but it is possible, even after some quality control of the data, that some of the remaining data could still contain errors.

Index Values

Table 2 shows the Fluidity and Planning Time Indexes calculated using the dispatch truck data for city pairs along the study corridor (Vancouver to Toronto) in the eastbound direction.

The largest FI occurred in the Vancouver to Toronto city pair with a 2.28. The PTI was 3.27 for this city pair. These relatively higher values over the longer corridor are intuitive because the variation in trips over long distances will be greater than that of trips with shorter distances. The relatively smaller sample size from Vancouver to Toronto may also contribute to these relatively larger values. The lowest FI occurred in the Winnipeg to Toronto city pair in October with a 1.40. Note that researchers included all trip types (e.g., solo driving) in all the calculations.

City Pair Data Descriptions

The effects of factors such as distance, operating conditions, and traffic volume that might be encountered in corridor or operational analysis are easier to see in graphical form at the city-pair level of data. This section provides an example of the findings that might be drawn from the available data, as well as illustrates the measure values. The example provided has several of the key analytical values noted, as well as the summary performance measures for that section of the Vancouver to Toronto corridor. All trip type groups are combined in the graph; a subsequent analysis provides a discussion of the separation into trip urgency groups. Additional city-pair examples are provided in the full report.[10]

Regina to Winnipeg

Performance characteristics, including the fluidity indicators, are shown in Figure 1. Figure 1 has a shape similar to a lightly congested urban roadway with many trips near to the free-flow condition and a few "bad days" with weather or crashes that extend travel times upward. This "long tail" distribution diminishes the usefulness of statistics that assume a statistically "normal" bell curve shape (e.g., standard deviation).

The average travel time of 10.1 hours yields an FI of 2.02 while the PTI is a relatively high 5.50, owing to the 24-hour and 29-hour travel time "mini-peaks." The "mini-peak" at 24 hours is likely

Table 2: Summary of Fluidity Index and Planning Time Index by city pair and month—eastbound (adapted from reference 10).

	City—Pairs									
	Vancouver to Toronto		Vancouver to Calgary		Calgary to Regina		Regina to Winnipeg		Winnipeg to Toronto	
Month	FI	PTI	FI	PTI	FI	PTI	FI	PTI	FI	PTI
Annual Avg	2.28	3.27	1.83	3.11	1.93	3.00	2.02	5.50	1.56	2.34
January	—	—	1.94	3.56	1.93	2.82	2.04	5.60	1.57	2.21
February	—	—	1.91	3.07	1.93	2.82	2.00	5.50	1.60	2.34
March	—	—	1.83	2.81	1.80	2.82	2.02	5.60	1.61	2.49
April	—	—	1.91	3.19	1.91	2.82	1.92	4.40	1.58	2.42
May	—	—	1.65	2.41	1.81	2.65	1.78	3.60	1.68	2.60
June	—	—	1.63	2.78	2.06	2.82	1.74	3.70	1.58	1.98
July	—	—	1.86	3.30	2.44	3.00	1.76	3.60	1.55	2.08
August	—	—	2.06	3.15	2.35	2.88	1.76	3.50	1.51	2.08
September	—	—	1.76	2.81	1.89	2.82	2.04	5.40	1.46	2.04
October	—	—	1.72	2.81	1.84	2.82	2.24	5.56	1.40	2.13
November	—	—	1.90	3.19	1.89	2.82	2.20	5.70	1.48	2.09
December	—	—	2.02	3.19	1.87	2.82	2.08	5.60	1.65	2.38

Note: "—" indicates limited sample sizes of data points, so measure could not be computed.

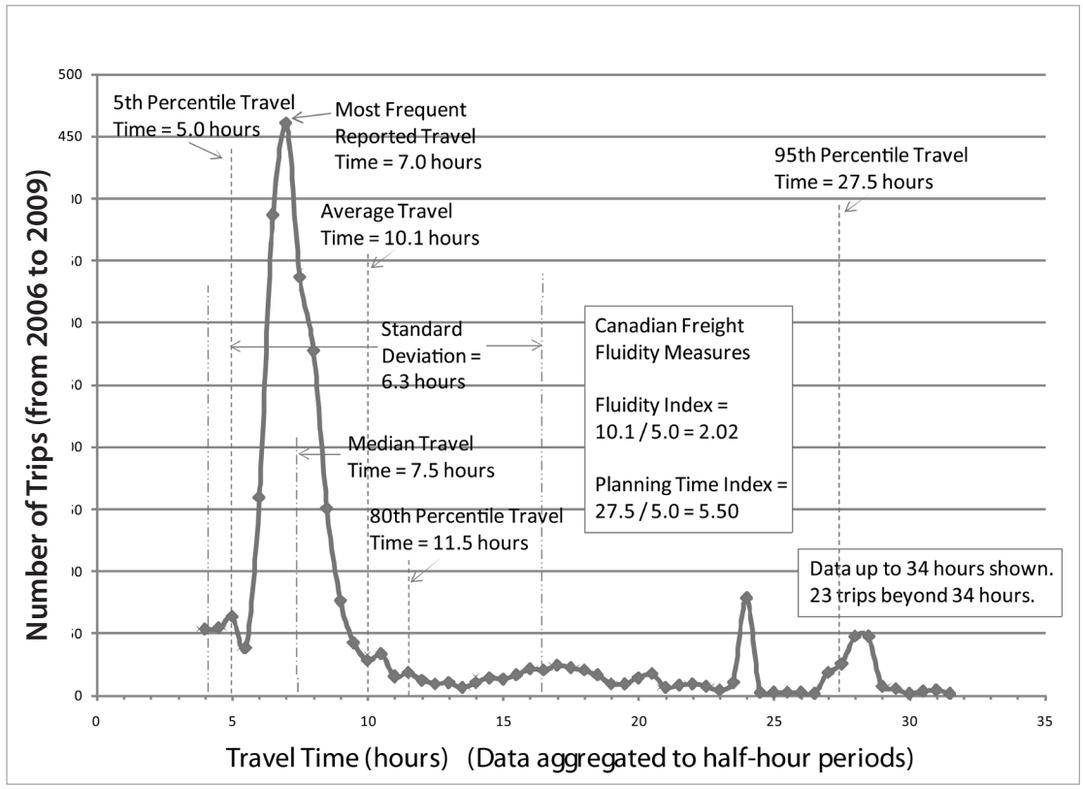
due to a large number of drivers checking in as "available" to the dispatchers the day after a delivery. This "mini-peak" at 24 hours is a data quality issue when using the dispatch data for performance monitoring because it can skew the distribution. The median travel time of 7.5 hours is unusually far from the average travel time (on a percentage basis), also due to the effect of the mini-peaks "pulling" the average travel time value higher.

Figure 2 presents the results of an analysis that divides the trip data into one of three urgency levels for the Vancouver to Calgary trip. The designation of urgency was accomplished by drawing dividing lines through the low point in the number of trips after the peak in the first two groups; this is somewhat arbitrary for the purposes of illustrating the concept.

The Fluidity Index and Planning Time Index values all decline markedly using this approach. This might be closer to the manner in which the index values would be used by shippers and manufacturers who would know the group in which their drivers operate. The values in Figure 2 suggest that, on average, they should allow 14 to 24 percent more time than the fastest drivers within their group (see Figure 2 FI values). For trips that must be on time, a budget of between 35 and 52 percent above the fastest trips in their group should be used (see Figure 2 PTI values).

Without confirmation from the trucking companies or a code in the data, the group definitions used in this paper are only speculation and are only valid for data experimentation; they should not be used for comprehensive performance evaluation at this time.

Figure 1: Performance statistics using truck dispatch data—Regina to Winnipeg, eastbound (575 kilometers and 2,739 trips). (Adapted from reference 10).



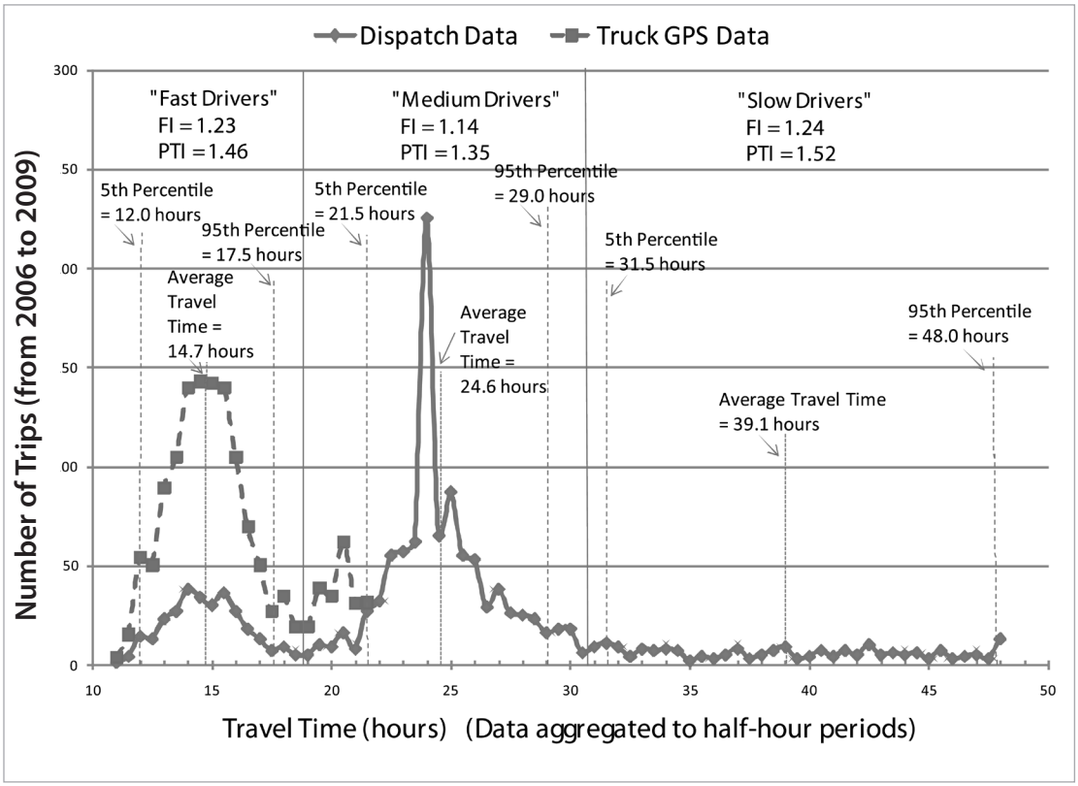
The dispatch data shows approximately 80 percent of the trips are less than 12 hours. The other 1 trip in 5, however, is spread over a long time increasing the standard deviation and the 95th percentile travel time. The average trip time is well above the median, suggesting a need to reexamine the travel time cutoff point for "long" trips (32 hours was used in this example) or to identify the cause of so many long travel times (e.g., weather, construction, dispatcher urgency).

Summary: A trip that takes 5 hours in the best conditions will take 10 hours (2.02 x 5 = 10.1 hours) on average. Shippers should allow 28 hours (5.50 x 5 = 27.5 hours) for trips with an important arrival time.

Combining Data Sources—Better Explanations and Data Quality Considerations

Multiple datasets can provide an indication of the "why" information related to travel times, but it also requires an understanding of the original use and preparation of the data. Figure 2 also compares the travel time data from the dispatchers (solid line; used in preceding sections) with a limited set of data from truck location monitoring devices (dashed line; tracked by GPS). The truck devices identify location and time, data that might be considered directly relevant. The average travel time in the GPS dataset is 14.9 hours, and the 95th percentile is 19.9 hours. Note that these values are not shown in Figure 2. This closely resembles the distribution of the "fast drivers" in Figure 2 from the truck dispatch data with an average of 14.7 hours and 95th percentile travel time of 17.5 hours. As mentioned previously, the dispatcher data, however, appear to be compiling a driver availability dataset because of the "mini-peaks" at 24 hours. The average time is more than 24 hours, and the 95th percentile is 48 hours. As one would expect, clearly there are differences in the methods used to compile the two datasets, differences that are important in the preparation of an accurate picture of freight travel conditions.

Figure 2: Distribution of trips by trip urgency group also showing comparison of truck GPS and truck dispatch data (Vancouver to Calgary, 915 kilometers). (Adapted from reference 10).



Calculating the performance measures for each trip urgency group provides a picture of a less congested, more reliable system.

Application—Multimodal Supply Chain

After investigating the use and calculation of the measures using the truck data, researchers demonstrate here the application of the fluidity indexes to an international and multimodal supply chain. Researchers investigated a supply chain from Shanghai, China to Toronto, Ontario, Canada. The multimodal trip includes the following elements:

- Ocean travel time (A);
- Port dwell time (B);
- Port drayage time (C);
- Rail dwell time awaiting departure (D);
- Rail travel time (E);
- Rail dwell time upon arrival (F);
- Rail drayage time to a distribution center (G);
- Truck travel time (H); and
- Truck dwell time upon arrival (I).

Note that these results are intended as a demonstration of the power of the fluidity indicators for decision making and are not intended for system monitoring at this point.

Table 3 has three different shades of cells depending on the source of the data as follows:

1. The non-shaded "Avg" (average) cells contain monthly data (when available).

2. The light-gray shaded cells contain annual average information that was used to populate all of the monthly cells (i.e., there is no monthly variation).

3. The dark-gray shaded cells contain data that were created for demonstration purposes.

This example assumes that 2 million TEUs move from Shanghai to the Port of Vancouver in a year. Of these TEUs, 1.5 million move by rail to Toronto with the remainder being transported by truck. Researchers calculated a Fluidity Index (FI) and Planning Time Index (PTI) for each segment of the trip. Table 3 also shows the average, 5th percentile, and 95th percentile travel times by month and mode.

The headings in Table 3 for each component of the supply chain are labeled with a letter from A to I. The rail trip includes the components labeled A, B, C, D, E, F, and G, and the truck trip includes the components labeled A, B, C, H, and I. These letter labels match the letters provided in parentheses in the previous bulleted list of each supply chain component.

The individual trip FI and PTI must be weighted based on their contribution to the entire supply chain to calculate a representative "weighted" overall FI and PTI for the entire supply chain. Researchers used TEU hours to weight the individual trip indexes. Since rail carries 1.5 million TEUs as opposed to 0.5 million TEUs on truck, the rail indexes will have a greater weighted effect on the overall indexes. The assumption was made that a greater percentage of TEUs are moved in March, July, September, and October than the other months for illustration purposes.

The ocean travel time takes the largest percentage of the entire trip with about 347 hours. The port drayage time takes the shortest percentage of the trip with just over an hour (see Table 3).

Table 4 illustrates the final weighted Fluidity Index and Planning Time Index for the supply chain application. In this example, the overall Fluidity Index for the entire supply chain is 1.28. The month with the longest travel time is January with a FI of 1.37. January also has the least reliable trip on average with a 1.66 Planning Time Index as opposed to the annual average PTI of 1.55. The quickest and most reliable trips occur during the month of May with an FI of 1.23 and a PTI of 1.49.

This example application demonstrates how the FI and PTI can provide supply chain performance information by mode and include a method to create annual measures by weighting monthly values by cargo amounts.

Table 3: Example of calculating the Fluidity Index and Planning Time Index for an international and multimodal supply chain. (Adapted from reference 10).

Port Metro Vancouver (2009)	A–Ocean Travel Time (hours or index)					B–Port Dwell Time (hours or index)					C–Port Drayage Time (hours or index)				
	Avg	5th	95th	FI	PTI	Avg	5th	95th	FI	PTI	Avg	5th	95th	FI	PTI
Annual Average	346.7	271.6	424.2	1.28	1.56	60.8	40.0	73.0	1.52	1.85	1.3	1.0	2.0	1.30	2.00
January	346.7	271.6	424.2	1.28	1.56	78.3	40.0	94.0	1.96	2.35	1.3	1.0	2.1	1.30	2.10
February	346.7	271.6	424.2	1.28	1.56	65.3	40.0	78.4	1.63	1.96	1.3	1.0	2.0	1.30	2.00
March	346.7	271.6	424.2	1.28	1.56	53.8	40.0	64.6	1.35	1.61	1.3	1.0	2.2	1.30	2.20
April	346.7	271.6	424.2	1.28	1.56	71.8	40.0	86.2	1.80	2.15	1.3	1.0	2.0	1.30	2.00
May	346.7	271.6	424.2	1.28	1.56	44.2	40.0	53.0	1.11	1.33	1.3	1.0	2.1	1.30	2.10
June	346.7	271.6	424.2	1.28	1.56	57.6	40.0	69.1	1.44	1.73	1.3	1.0	2.0	1.30	2.00
July	346.7	271.6	424.2	1.28	1.56	67.7	40.0	81.2	1.69	2.03	1.3	1.0	2.2	1.30	2.20
August	346.7	271.6	424.2	1.28	1.56	50.2	40.0	60.2	1.26	1.51	1.3	1.0	2.0	1.30	2.00
September	346.7	271.6	424.2	1.28	1.56	57.1	40.0	68.5	1.43	1.71	1.3	1.0	2.1	1.30	2.10
October	346.7	271.6	424.2	1.28	1.56	61.7	40.0	74.0	1.54	1.85	1.3	1.0	2.0	1.30	2.00
November	346.7	271.6	424.2	1.28	1.56	56.9	40.0	68.3	1.42	1.71	1.3	1.0	2.2	1.30	2.20
December	346.7	271.6	424.2	1.28	1.56	65.5	40.0	78.6	1.64	1.97	1.3	1.0	2.0	1.30	2.00

Port Metro Vancouver (2009)	D–Rail Dwell Time—Departure (hours or index)					E–Rail Travel Time (hours or index)					F–Rail Dwell Time—Arrival (hours or index)					G–Drayage Time to Distrib. Center (hours or index)				
	Avg	5th	95th	FI	PTI	Avg	5th	95th	FI	PTI	Avg	5th	95th	FI	PTI	Avg	5th	95th	FI	PTI
Annual Average	14	11	16	1.29	1.48	108	100	119	1.08	1.19	21	18	27	1.19	1.48	14	12	19	1.19	1.55
January	15	11	18	1.45	1.66	118	100	130	1.18	1.30	25	18	32	1.41	1.76	18	12	23	1.46	1.90
February	14	11	16	1.33	1.53	118	100	130	1.18	1.30	24	18	29	1.31	1.63	15	12	19	1.24	1.61
March	15	11	17	1.40	1.61	118	100	130	1.18	1.30	22	18	28	1.22	1.53	16	12	21	1.36	1.77
April	15	11	17	1.42	1.63	105	100	116	1.05	1.16	19	18	24	1.07	1.34	13	12	17	1.06	1.38
May	14	11	17	1.37	1.58	105	100	116	1.05	1.16	20	18	25	1.11	1.39	13	12	16	1.04	1.35
June	14	11	16	1.29	1.48	105	100	116	1.05	1.16	22	18	27	1.21	1.51	13	12	18	1.12	1.46
July	14	11	16	1.33	1.53	105	100	116	1.05	1.16	19	18	24	1.07	1.34	13	12	18	1.12	1.46
August	11	11	13	1.05	1.20	105	100	116	1.05	1.16	19	18	24	1.07	1.33	15	12	20	1.26	1.64
September	13	11	15	1.22	1.40	105	100	116	1.05	1.16	19	18	24	1.08	1.35	14	12	18	1.14	1.48
October	13	11	15	1.20	1.38	105	100	116	1.05	1.16	21	18	26	1.17	1.46	14	12	18	1.18	1.53
November	12	11	14	1.17	1.35	105	100	116	1.05	1.16	22	18	27	1.22	1.52	14	12	18	1.14	1.48
December	13	11	15	1.27	1.46	105	100	116	1.05	1.16	24	18	29	1.31	1.63	14	12	18	1.14	1.48

Table 3 (continued): Example of calculating the Fluidity Index and Planning Time Index for an international and multimodal supply chain. (Adapted from reference 10).

Port Metro Vancouver (2009)	H–Truck Travel Time (hours or index)					I–Truck Dwell Time-Arrival (hours or index)				
	Avg	5th	95th	FI	PTI	Avg	5th	95th	FI	PTI
Annual Average	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
January	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
February	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
March	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
April	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
May	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
June	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
July	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
August	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
September	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
October	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
November	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00
December	127.4	90.0	159.0	1.42	1.77	3.0	2.0	4.0	1.50	2.00

Note: Some values are rounded in table.

Table 4: Final weighted Fluidity Index and Planning Time Index for an international and multimodal supply chain application.

Port Metro Vancouver (2009)	Weighting Factors—TEU-Hours (in thousands)										
	A–Ocean	B–Port Dwell	C–Port Drayage	D–Rail Dwell Departure	E–Rail Travel	F–Rail Dwell Arrival	G–Rail Final Drayage	H–Truck Travel Time	I–Truck Dwell Time	FI	PTI
Annual Total	693,400	121,610	2,600	20,355	162,308	32,011	21,397	63,700	1,500	1.28	1.55
January	43,338	9,788	163	1,425	11,063	2,381	1,643	3,981	94	1.37	1.66
February	43,338	8,163	163	1,313	11,063	2,203	1,395	3,981	94	1.31	1.58
March	86,675	13,450	325	2,756	22,125	4,125	3,060	7,963	188	1.28	1.54
April	43,338	8,975	163	1,397	9,844	1,809	1,193	3,981	94	1.31	1.59
May	43,338	5,525	163	1,350	9,844	1,875	1,170	3,981	94	1.23	1.49
June	43,338	7,200	163	1,266	9,844	2,043	1,260	3,981	94	1.26	1.53
July	86,675	16,925	325	2,625	19,688	3,619	2,520	7,963	188	1.30	1.57
August	43,338	6,275	163	1,031	9.844	1,800	1,418	3,981	94	1.24	1.50
September	86,675	14,275	325	2,400	19,688	3,638	2,565	7,963	188	1.26	1.52
October	86,675	15,425	325	2,363	19,688	3,938	2,655	7,963	188	1.28	1.54
November	43,338	7,113	163	1,153	9,844	2,053	1,283	3,981	94	1.26	1.53
December	43,338	8,188	163	1,247	9,844	2,203	1,283	3,981	94	1.29	1.56

Note: Final FI and PTI values are weighted by TEU hours to obtain systemwide supply chain performance, which includes rail and truck movements. Some values are rounded.

Conclusions and Discussion

World trade will continue to grow, and Canada's competitiveness depends on a reliable and efficient freight transportation system to import and export commodities the most efficient way. It is important to systematically measure the efficiency of Canada's supply chain system to be able to make improvements to the systems and increase competitiveness.

The Measures

The Fluidity Index (FI) and Planning Time Index (PTI) will serve as measures of the overall efficiency of Canada's transportation system at a corridor level. Specific objectives of the measures are to:

- Provide transparent reliability information pertaining to every one of Canada's freight significant corridors;
- Identify bottlenecks in the Canadian logistics system; and
- Ensure Canada's overall competitiveness in the global marketplace.

The index-based approach is being proposed to calculate fluidity rather than simply using travel times. Information from transportation providers is imperative to the overall success of the FI and PTI. The successful implementation of the FI and PTI in Canada may encourage the other nations in North America to participate in similar initiatives.

The FI and PTI will have many factors in common with other mobility and congestion indicators. Experience from previous applications of the mobility principles provide analysts and planners with information to identify important factors and problems, communicate the benefits of investments and policy changes, and prioritize possible actions that affect the motoring public and freight operations. Some of the relevant principles include the following:

- Both average travel time conditions (FI) and reliability of travel time (PTI) are needed to measure freight system efficiency;
- No single measure will satisfy all fluidity monitoring needs, and no single measure can identify all aspects of fluidity. Mobility and freight fluidity are complex and in many cases requires more than one measure, more than a single data source, and more than one analysis procedure;
- The FI and PTI are described as ratios relative to travel time in "free-flow" conditions. They can both be used as multimodal transportation system measures, and they can be calculated for a range of corridors and regional systems; and
- At least six months, and desirably one year, of travel time data should be compiled to identify typical and unusual conditions. One year of data allows seasonal and market variations to be incorporated into the performance measures. If data sources are changed, analysis methods altered, or data quality standards adjusted, it will be useful to compile measures using both old and new standards for several months to understand the "translation" between the measures.

Data Quality

Information from shippers and carriers is a key element to develop a reliable FI and PTI. As with any analysis, the results of the FI and PTI calculation are only as good as the data used in the process. Transport Canada has many rich sources of data for this freight mobility analysis. However, as with any data, there were some issues identified with the various datasets. Many of the problems are related to consistency. In most cases, it is better to have less data of high quality rather than to have a lot of data with data quality issues (e.g., typos, inconsistent data entry, missing fields). Much of these

data were eliminated to perform an analysis, and if the fluidity indicators concept is to be widely deployed, reducing the amount of resources needed to clean and manage the data will be important.

Proof of Concept

The methodology and applications shown in this paper were used to calculate both FI and PTI values where sufficient data were available. The FI and PTI were calculated based on all trips and for several different types of truck trips because these different types of trips have different requirements of and uses for the transportation system.

The measure tables showed the results at different levels of aggregation from a travel time for the entire supply chain to the amount of time required at particular nodes along a trip (e.g., port dwell or truck drayage). The suggested freight unit to compare across the modes is the 20-foot equivalent unit (TEU) as estimated from total freight flows (not simply the volume of sampled travel times).

The detailed GPS data can be used to perform analyses at a very detailed level. This GPS information can supplement other information such as the less detailed dispatch data to gather a complete picture as to how the transportation network is performing and where potential problems exist at a very microscopic level.

Future Work

The corridor and modes selected for the proof of concept described in this paper include an emphasis on roadway, due to the detailed information available for this mode. However, given that Canada's largest trade partner is the United States, there is substantial movement of freight between these two countries. Efficient cross-border trade by truck between Canada and the United States affects Canada's competitiveness; therefore, future work should include the analysis of a corridor that has a land port of entry.

Acknowledgments

The authors would like to thank Transport Canada for its sponsorship of this project. The authors would also like to thank the staff of Transport Canada's Economic Analysis and Research Group for providing the data and insights related to the data throughout the project.

References

1. *Transportation in Canada: An Overview.* Report No. TP 14816. Transport Canada. 2008. Available: http://www.tc.gc.ca/media/documents/policy/overview_1.pdf. Last Accessed November 12, 2010.

2. *International Trade and Trade Flows.* Transport Canada. 2006. Available: http://www.tc.gc.ca/eng/policy/report-aca-anre2006-2b_trade-eng-1958.htm. Last Accessed November 12, 2010.

3. *Prince Rupert Records 12 year-high Cargo Volumes in 2009*. Prince Rupert Port Authority Press Release. Prince Rupert, British Columbia, Canada. January 25, 2010. Available: http://www.rupertport.com/pdf/newsreleases/ppr_records_12yearhigh2009cargovolumes.pdf. Last Accessed November 12, 2010.

4. *The Use of Containers in Canada.* Prepared for Transport Canada by MariNova Consulting Ltd. & Partners. December 2006. Available: http://www.transportation.alberta.ca/Content/docType56/Production/UseofContainersinCanadaFinalReport_0.pdf. Last Accessed November 12, 2010.

5. North American Transportation Statistics Database. *Domestic Freight Activity by Mode.* Available: http://nats.sct.gob.mx/nats/sys/tables.jsp?i=3&id=15. Last Accessed November 12, 2010.

6. *2010 Urban Mobility Report.* Texas Transportation Institute, College Station, TX. 2010. Available: http://mobility.tamu.edu/ums.

7. *Urban Congestion Report.* U.S. Department of Transportation. Federal Highway Administration. Washington, DC. December 2010. Available: http://www.ops.fhwa.dot.gov/perf_measurement/ucr/index.htm.

8. Schrank, D. L., W. L. Eisele, and T. J. Lomax. *The Keys to Estimating Mobility in Urban Areas: Applying Definitions and Measures that Everyone Understands.* White paper prepared for the Urban Transportation Performance Measure FHWA Study. Texas Transportation Institute, Second Edition, May 2005. Available: http://mobility.tamu.edu/resources/estimating_mobility.stm.

9. Eisele, W. L. and D. L. Schrank. Conceptual Framework and Trucking Application to Estimate the Impact of Congestion on Freight. *Transportation Research Record 2168*. Transportation Research Board, National Research Council, Washington, DC, 2010.

10. *Methodology for the Development of Fluidity Indicators for Freight Transportation.* Transport Canada Contract #T8080-09-0134. Texas Transportation Institute, College Station, TX. June 2010.

WILLIAM L. EISELE, Ph.D., P.E. is a research engineer at the Texas Transportation Institute. He has been actively involved in urban mobility, performance measurement, and access management research for more than 16 years. He has been part of a team on several projects to develop freight mobility methodologies that can inform mobility decisions. He is a member of ITE.

LOUIS-PAUL TARDIF is the director of Economic Analysis and Research at Transport Canada. He is responsible for the development of indicators providing a measurement of the performance of the transportation sector in Canada. Prior to joining Transport Canada in 2007, he served as a private consultant to private and public sector clients and has held positions with the Canadian Trucking Alliance, Quebec Ministry of Transport, and the Canadian Transport.

JUAN C. VILLA is a research scientist and program manager at the Texas Transportation Institute. He has 30 years of professional experience in transportation and logistics projects. He has experience in consulting, research, and engineering projects in Mexico, Canada, the United States, and other Latin American counties. He is a member of ITE.

DAVID L. SCHRANK, Ph.D., P.E. is an associate research scientist at the Texas Transportation Institute. He has been working in urban mobility research for almost 20 years as coauthor of the *Urban Mobility Report*. He has also been involved with several projects in freight mobility to develop solutions to mobility problems through improved performance measurement for decision making. He is a member of ITE.

TIM LOMAX, Ph.D., P.E. is a research engineer at the Texas Transportation Institute. He has been extensively involved in urban mobility research for more than 30 years. He has been part of a team of researchers that has developed and applied methodologies to assess traffic congestion levels and costs. He is a fellow of ITE.

Assessing Signal Timing Plans for Winter Conditions

By Thomas M. Brennan Jr., Christopher M. Day, Jason S. Wasson, James R. Sturdevant, and Darcy M. Bullock

Coordinated arterial signal timing plans are typically designed for normal weather conditions based on a number of assumed traffic flow characteristics. Qualitatively, during winter operations vehicle speeds decrease, arrival time of platoons shift, vehicle headway increases, platoon dispersion increases, and saturation flow rates decrease. As a result, good weather signal timing plans may be less than optimal during winter operations. Although there has been some research conducted regarding the effects of weather on some traffic flow parameters, these efforts have been based upon manual field observations. There has been little work on developing automated methods of measuring these parameters or assessing of the benefits that could be achieved by implementing alternative timing plans during winter events.

Abstract

This paper presents findings from the automated collection of high-resolution signal controller data and Bluetooth® probe vehicle travel times to characterize both the microscopic and macroscopic operation of a four-intersection signalized arterial during winter weather conditions (snow and ice on pavement) as well as clear pavement conditions. An 83-second increase in median travel time through the system was measured during winter conditions. Platoon shifts of 15, 25, and 30 seconds were measured at three intersection links that corresponded to an approximate reduction in design speed of 7 to 11 miles per hour (mph) on a corridor with a posted speed of 55 mph. Alternative offsets were calculated that showed an opportunity to decrease overall vehicle delay by 26.7 percent for southbound vehicles during the a.m. peak snow event.

Introduction and Concept

To determine whether coordination will benefit a signalized arterial, one of the factors that needs to be known is how dispersed a platoon of vehicles becomes as it travels through the system. As quoted in the *Traffic Control Systems Handbook*, page 3.35: [1]

> *"When a platoon of vehicles is released from a traffic signal, the degree to which this platoon has dispersed at the next signal (difference from profiles at releasing signal) in part determines whether significant benefits can be achieved from signal coordination."*

Various models and equations are currently used to determine the degree of dispersion based on factors such as travel speed, headway, and volume.[1, 2, 3] The traffic flow characteristics used for design are generally based upon normal operational conditions (i.e., clear weather). In addition, direct field observation of platoon arrivals during peak volume periods are often used to field tune traffic signal offsets. However, field observation and tuning of traffic signal offsets during winter weather conditions are not scalable and raise safety concerns.

This paper defines procedures to use high-resolution, event-based traffic controller data to directly measure traffic flow characteristics and coordinated platoon characteristics. The use of high-resolution data allows the visual display of the interaction of signal phase information with respect to advance detection cycle by cycle. This intersection-level data is coupled with Bluetooth® probe travel time data to assess system-level impact. The following three flow characterization concepts provide an indication of platoon presence and its associated dispersion:

- What is the change in corridor travel speed?
- What is the change in travel link approach headway?
- What is the change in platoon formation and dispersion?

Corridor travel speed is obtained from the Bluetooth® probe travel, while the headway and platoon characteristics are obtained from high-resolution, event-based traffic control data. These concepts are described in subsequent sections and are used to evaluate the need for signal coordination during an adverse weather event. As would be expected, there was an increase in corridor travel time and link approach headways during observed snow events. Observed traffic patterns during the snow event as compared to a clear-weather day indicate that there is an opportunity to improve coordination during a snow event. To quantify this observed opportunity, an optimization algorithm to minimize delay is applied to determine whether the arterial would benefit from alternative signal offsets during a snow event.[4, 5]

Literature Review

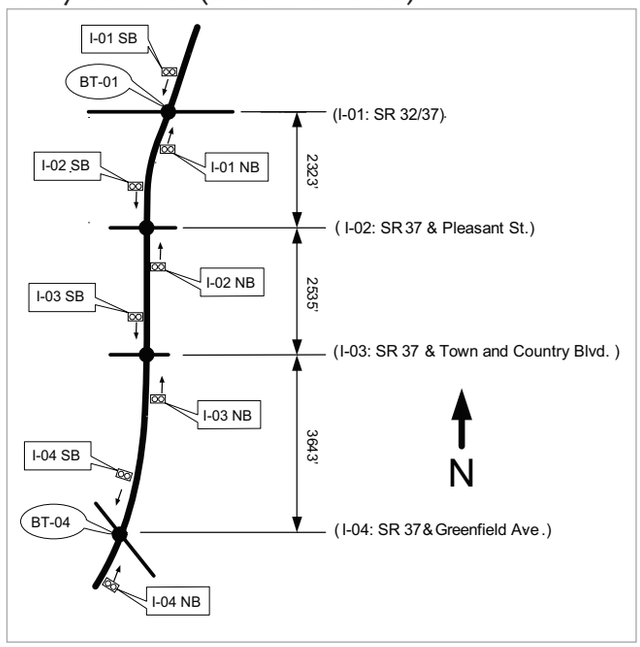
Two notable inclement weather plans have been developed for the Utah Department of Transportation (USA) and the city of Anchorage, AK, USA. [6, 7] In these studies, alternative timing plans are proposed to improve winter weather vehicle progression. A comprehensive literature review of several winter operation studies had been conducted to document weather impacts on arterial traffic flow.[6, 8, 9] All of the studies support the notion that adverse weather conditions, such as snow, affect the traffic flow characteristics of the corridor. Qualitatively, it is intuitive that during winter conditions vehicle speeds decrease, arrival time of platoons shift, platoon dispersion increases, vehicle headway increases, and saturation flow rates decrease.[7, 10,11, 12, 13, 14] As a result, good weather signal timing plans may be less than optimal during winter events. Therefore, improvements to the signal coordination are possible. [6, 7, 9, 11, 12, 13, 14] Although there has been some research conducted regarding both the effects of weather on some traffic flow characteristics and how changes to the signal offset and timing can improve traffic progression, there has not been any comprehensive measurement of both traffic flow characteristics and system arterial travel time during winter weather events to esti-

mate the benefits (if any) that could be obtained by implementing alternative winter weather timing plans.

Indiana Instrumented Corridor

Figure 1 shows a 1.6 mile (mi) (2.6 kilometer [km]) portion of SR 37 in Noblesville, IN, USA, located northeast of Indianapolis, IN. This is a coordinated system made up of four intersections. A portion of the coordinated phases are also actuated.[4, 5] Each intersection in the arterial features advance detectors 405 feet (ft) (123 meters [m]) upstream of the stop bar on all coordinated northbound and southbound movements. Each of the intersections—I-01, I-02, I-03, and I-04—has the capability of logging high-resolution controller data (phase and detector status changes) at a resolution of 0.1 seconds.[16] Two permanent Bluetooth® sensors for collecting unique media access control (MAC) addresses (BT-01 and BT-04) were also deployed along the arterial at intersections I-01 and I-04, respectively.[4, 17, 18] High-resolution controller and Bluetooth® probe data were remotely collected via a wireless private Internet connection.

Figure 1: Location of advance detectors on study corridor (State Route 37).



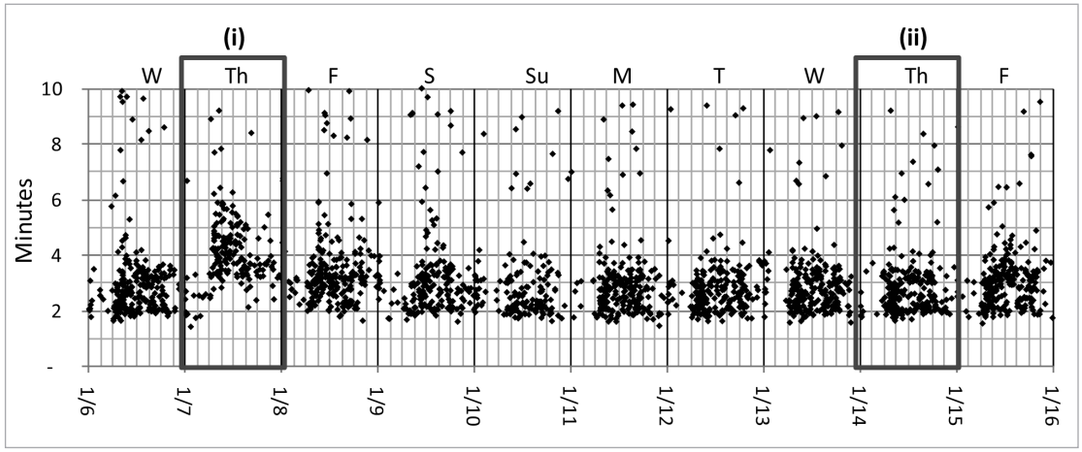
In Figure 2, where "N" is equated to the number of probe vehicles, plots of travel time measurements of vehicles traveling between BT-01 and BT-4 are shown. These Bluetooth® monitoring stations are located at the signalized intersection cabinets. Although midblock mounting provides more precise characterization of system travel time, the availability of power and communication at a signalized intersection provides a much more cost-effective, long-term data collection environment necessary for studies conducted over several weeks when the presence of a significant winter storm is highly variable. Despite suboptimal mounting, other studies have compared midblock with cabinet mounting in a side-by-side manner and have shown that intersection mounting of Bluetooth® probe monitoring sites provides reasonable travel time characterization over longer distances.[4]

Measurement and Assessment

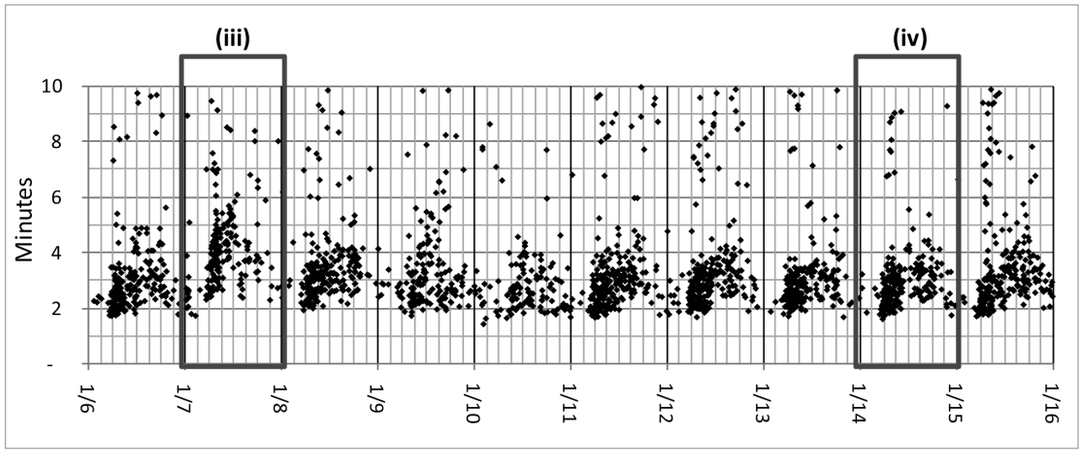
Travel Time

Ten days of travel time are displayed for the northbound (Figure 2a) and southbound (Figure 2b) directions. During the 10-day data collection, a snow event occurred on Thursday, January 7, 2010, which can be compared to a clear-weather day on Thursday, January 14, 2010. Comparing the travel time plots for these two days, the snow event travel times (Figure 2.a.i and Figure 2.a.iii) have a noticeable increase compared to a clear-weather (Figure 2.a.ii and Figure 2.b.iv) for both the northbound and southbound directions. The coordinated arterial is optimized to minimize the delay in the southbound direction in the a.m. peak under clear-weather conditions. The southbound direction serves the majority of the vehicle volume during the a.m. peak period (an approximate 70/30 directional split during normal a.m. operations).

Figure 2: 10-Day (Jan. 6, 2010 – Jan. 15, 2010) probe vehicle travel times (minutes) along SR37; (i & iii) snow conditions; (ii & iv) clear weather conditions.



a) Northbound BT-04 to BT-01 (vehicle volume [N]: Ni = 179, Nii = 241)

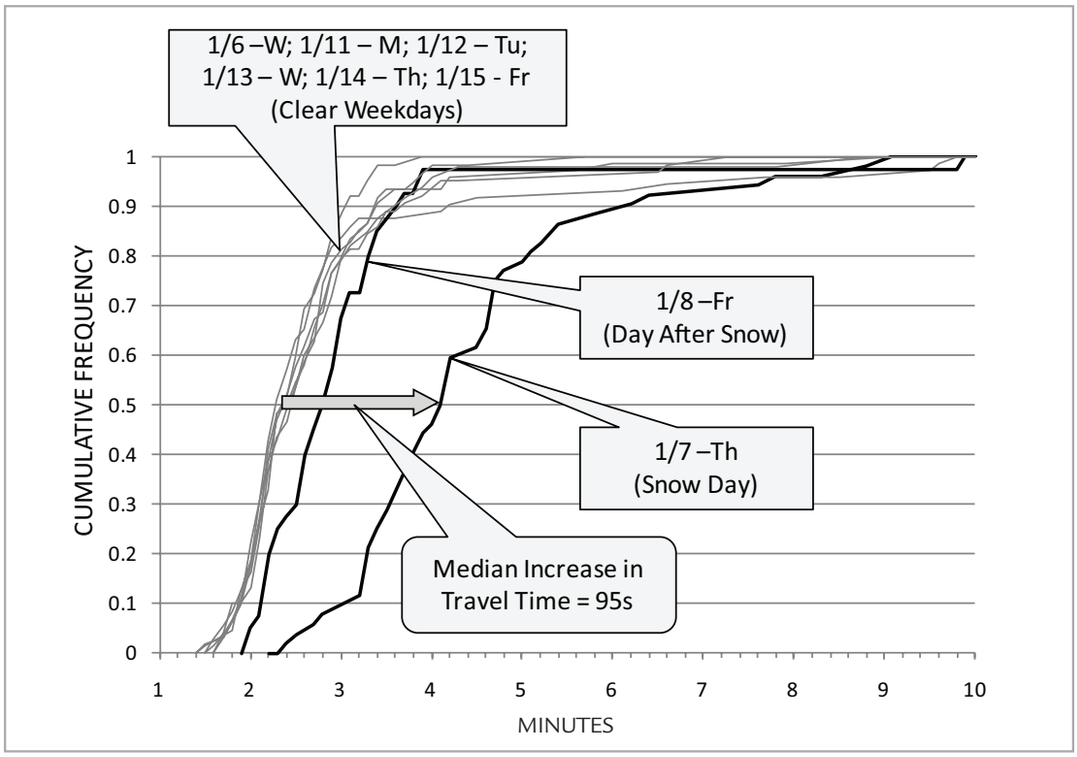


b) Southbound BT-01 to BT-04 (vehicle volume [N]: Niii = 199, Niv = 217)

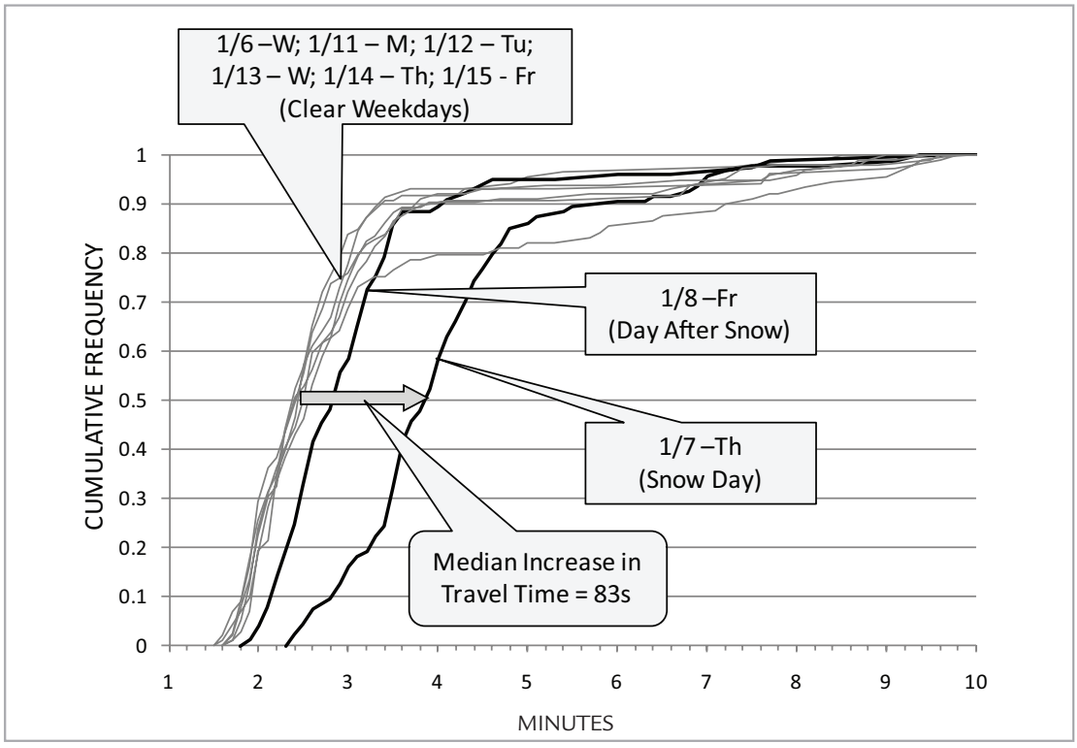
These travel times are perhaps more effectively compared during particular time periods using cumulative frequency distributions (CFDs), as shown in Figure 3.

- Figure 3.a. shows northbound weekday a.m. peak (06:00 – 09:00) travel times measured from probe vehicle data going from station BT-04 to BT-01 (Figure 1). A median travel time increase of 95 seconds is observed in the northbound travel direction during the snow event.
- Figure 3.b. shows southbound weekday a.m. peak (06:00 – 09:00) travel times measured from probe vehicle data going from BT-01 to BT-04 (Figure 1). A median travel time increase of 83 seconds is observed in the southbound direction during the snow event.

Figure 3: Cumulative frequency distribution of probe vehicle travel times along SR 37 during a.m. peak 06:00–09:00 (<= 10 minutes).

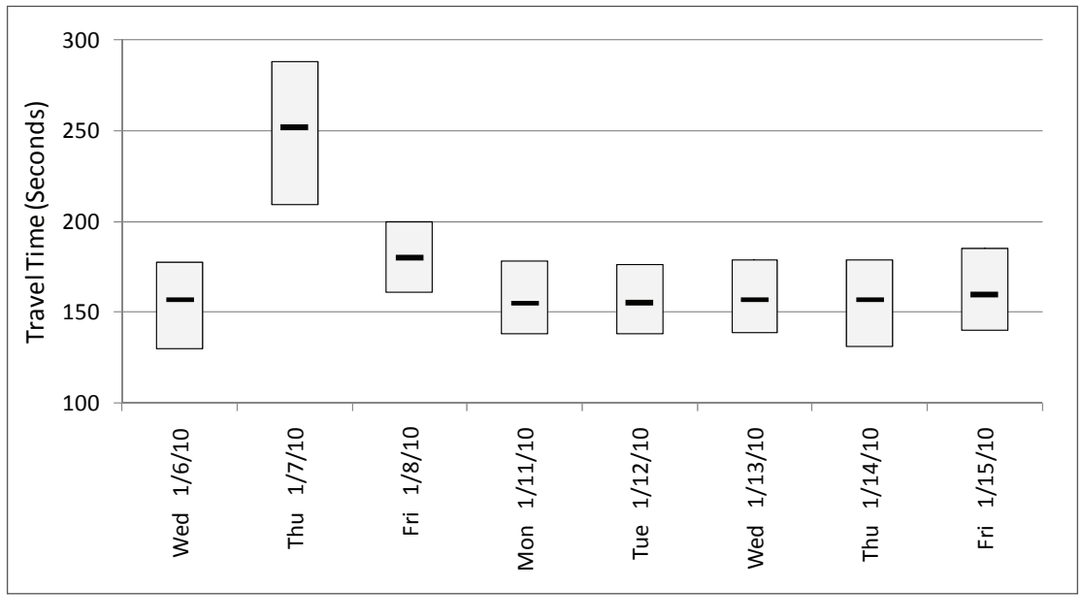


a) Northbound BT-04 to BT-01

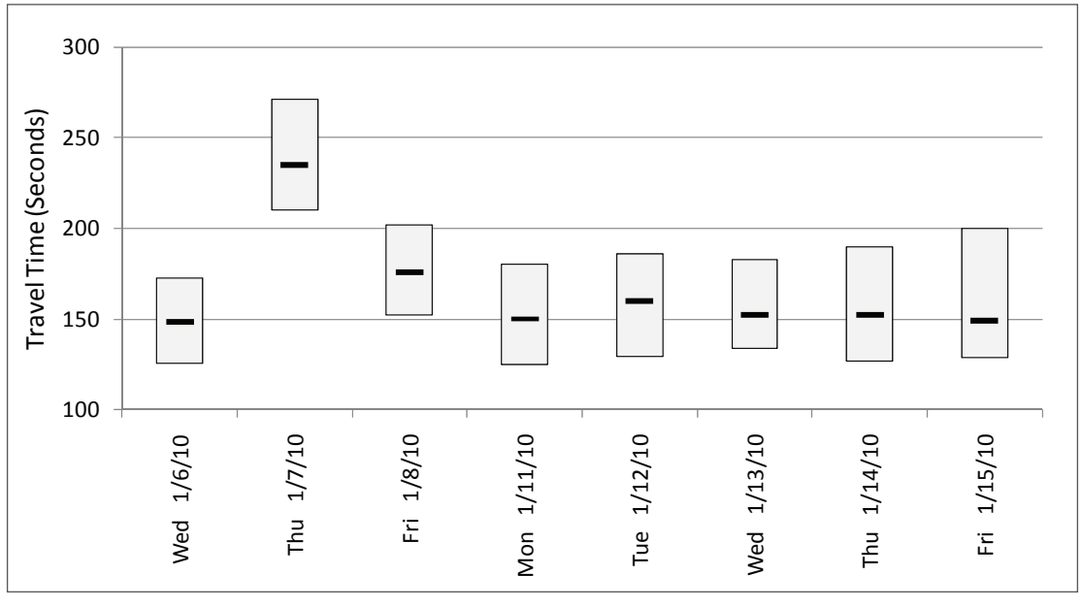


b) Southbound BT-01 to BT-04

Figure 4: 25th, 50th, and 75th quartile box-plot distribution of vehicle travel times along SR 37 corridor a.m. peak 06:00 – 09:00 (<= 10 minutes).



a) Northbound BT-04 to BT-01



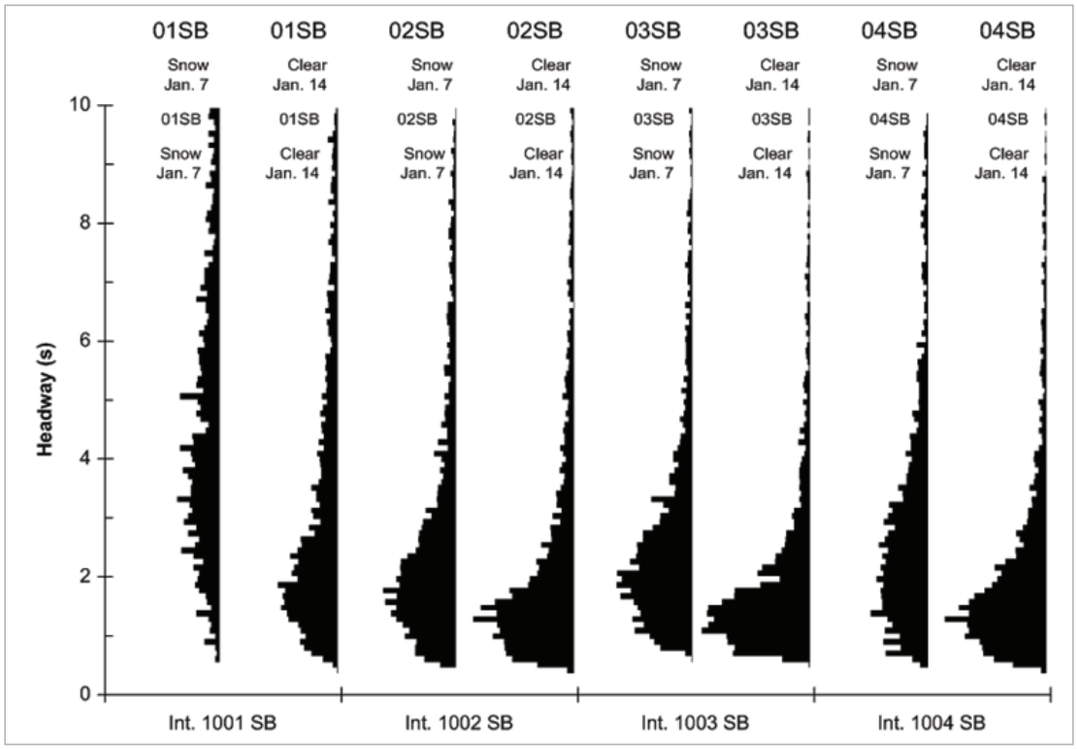
b) Southbound BT-01 to BT-04

Table 1: Probe vehicle travel times for SR 37 corridor.

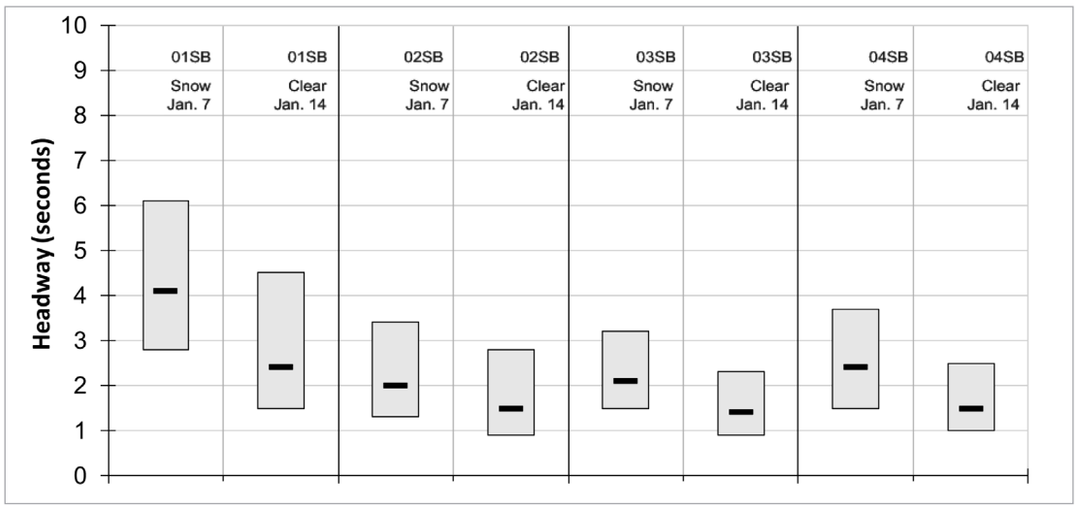
Dates	N (Veh)	Mean Travel Time (s)	S.D.	Percentiles (s) 25th	50th	75th	Interquartile Range (s)
Southbound BT-01 to BT-04 a.m. peak (06:00 – 09:00)							
W; 1/6/10	86	166	1.21	126	149	173	47
(SNOW) Th; 1/7/10	94	252	1.33	210	235	271	61
F; 1/8/10	77	191	1.29	152	176	202	50
M; 1/11/10	99	175	1.58	125	150	181	56
Tu; 1/12/10	102	183	1.52	129	160	186	57
W; 1/13/10	103	179	1.64	134	152	183	50
Th; 1/14/10	97	184	1.64	127	152	190	63
F; 1/15/10	89	208	2.17	129	149	200	71
Northbound BT-04 to BT-01 a.m. peak (06:00 – 09:00)							
W; 1/6/10	52	164	0.82	130	157	178	48
(SNOW) Th; 1/7/10	52	268	1.45	210	252	288	79
F; 1/8/10	32	181	0.46	161	180	200	39
M; 1/11/10	37	171	1.20	138	155	178	40
Tu; 1/12/10	50	156	0.48	138	156	177	38
W; 1/13/10	53	165	0.80	139	157	179	40
Th; 1/14/10	61	170	1.06	131	157	179	48
F; 1/15/10	49	168	0.70	140	160	185	45

In addition, a slight increase in travel time is observed on January 8, 2010, the day after the snow event occurred. This could be due to residual snow and ice remaining from the previous day's snow event. A statistical box-plot distribution, which displays the quartiles (25th, 50th, and 75th percentiles) of the probe vehicle observations, is shown in Figure 4. Table 1 provides the summary statistics of the probe data used to construct the CFDs in Figure 3 and box-plot distribution in Figure 4, where "N" is equated to the number of probe vehicles.

Figure 5: Southbound SR37 approach headways during Thursday snow (Jan. 7, 2010) and clear weather (Jan. 14, 2010) conditions.



a) Aggregate southbound a.m. peak 06:00 – 09:00 approach headways.



b) 25th, 50th, and 75th quartile box-plot distribution of a.m. peak 06:00 – 09:00 approach headways (<=10 seconds).

Table 2: Arrival headway for each intersection a.m. peak 06:00–09:00 southbound during snow (Jan 7) and clear weather (Jan 14) conditions. (<= 10 seconds).

Intersection	Direction	N (Veh)	N^h<10s (Veh)	Mean Headway (s)	S.D.	Percentiles (s) 25th	50th	75th	Interquartile Range (s)
January 7, 2010 (Snow) AM Peak									
I-01	SB	1434	769	3.05	1.15	2.2	3.1	4.0	1.8
I-02	SB	2983	2407	2.01	1.11	1.2	1.8	2.6	1.4
I-03	SB	2584	2145	2.15	1.02	1.4	2.0	2.7	1.3
I-04	SB	2331	1811	2.25	1.15	1.3	2.1	3.0	1.7
January 14, 2010 (Clear) AM Peak									
I-01	SB	2386	1703	2.22	1.20	1.3	2.0	3.0	1.7
I-02	SB	3585	3007	1.67	1.10	0.9	1.4	2.2	1.3
I-03	SB	3006	2585	1.60	0.97	0.9	1.3	2.0	1.1
I-04	SB	2896	2444	1.65	1.02	0.9	1.4	2.1	1.2

Headway

The approach headway is the amount of time it takes successive vehicles to pass through a fixed point. Approach headway was measured at the eight approach detectors (I-01 SB, I-02 SB, I-03 SB, I-04 SB, I-01 NB, I-02 NB, I-03 NB, and I-04 NB) that are 405′ back from the stop bar. The locations of these sensors are shown in Figure 1. These approach headways are represented graphically in Figure 5:

- Figure 5.a. shows the aggregated frequency of southbound headways observed during the a.m. peak (06:00 – 09:00) measured at the advanced detectors (Figure 1) for January 7, 2010 and January 14, 2010.
- Figure 5.b. shows the statistical box-plot distribution of the southbound headways observed during the a.m. peak (06:00 – 09:00) measured at the advanced detectors (Figure 1) for January 7, 2010 and January 14, 2010 for all headways less than or equal to 10 seconds.

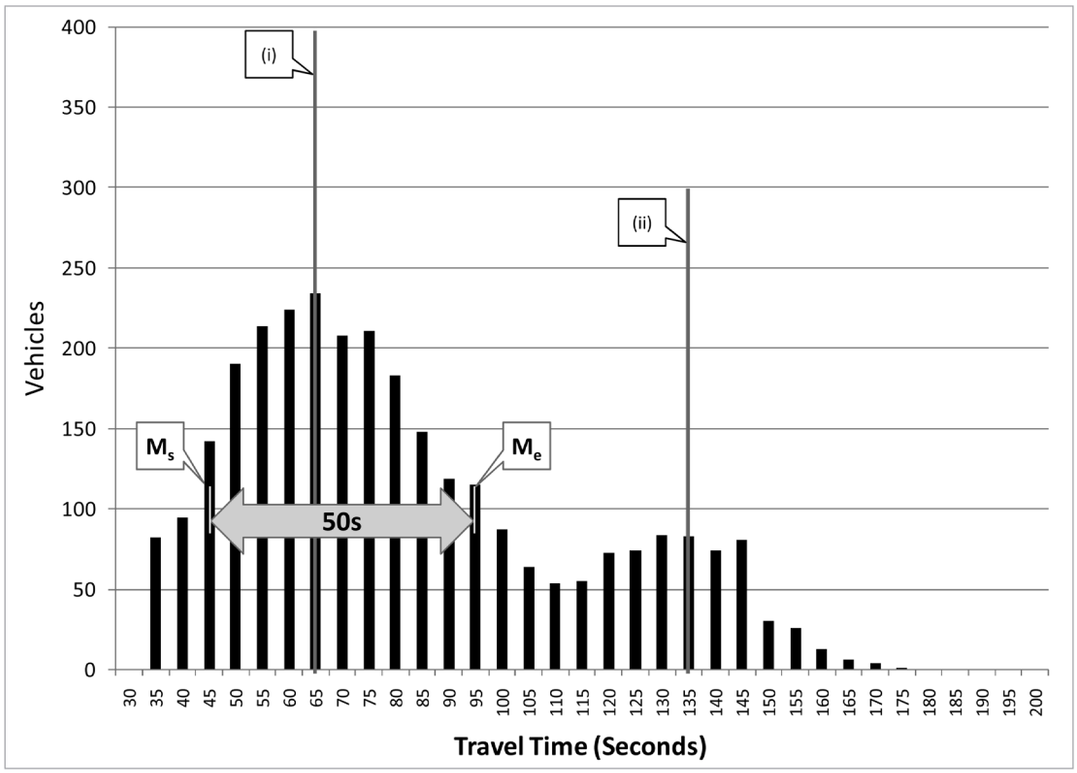
Table 2 shows the summary statistics for Figure 5.b., where "N" equals the total number of vehicles observed. An aggregated total of 7,132 headway observations less than 10 seconds (shown as "N^h<10" in Table 2) were made on January 7, and 9,739 for January 14. It is clear from the graphs that there is an overall increase in the amount of headway observed for each link during the snow event. This increase appears to be greatest for southbound vehicles first entering the coordinated system at I-01.

Platoon Shift and Dispersion

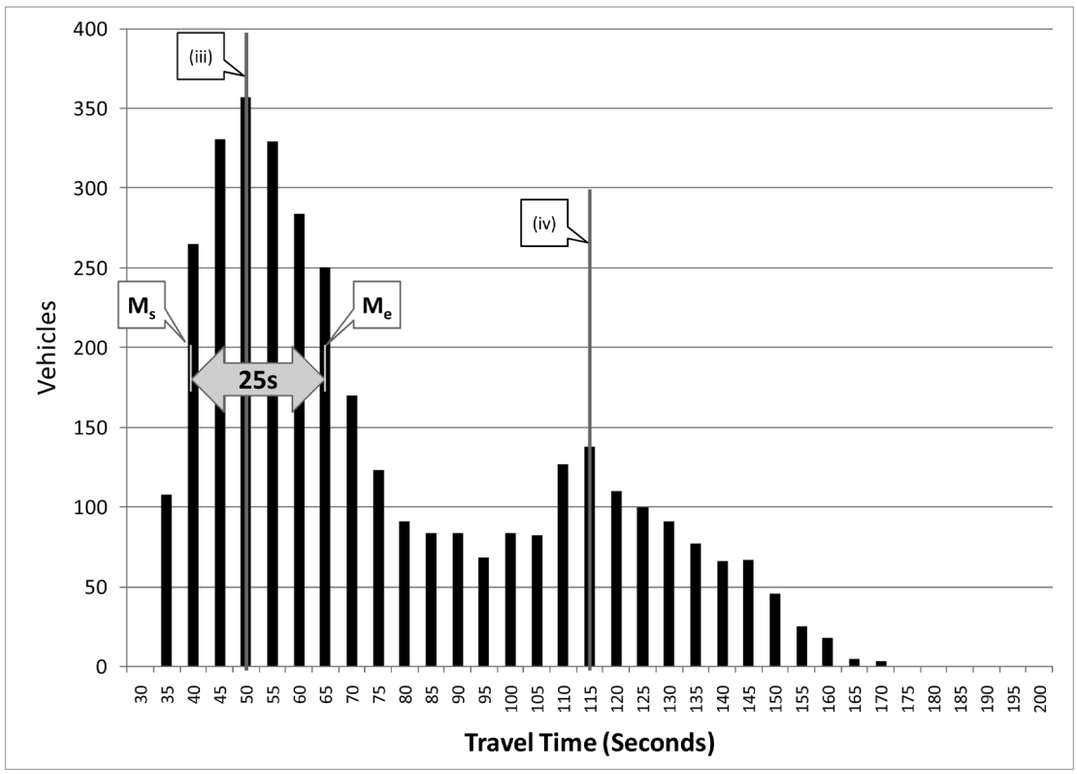
A direct measure of the aggregated platoon flow profile during the southbound a.m. peak is presented in Figure 6.2 The increase in headway, along with the increase in travel speed shown in Figure 3, Figure 4, and Figure 5, supported the snow event platoon dispersion observed in Figure 6.

- Figure 6.a. shows the aggregated platoon profile during the a.m. peak (06:00 – 09:00) snow event measured at the advanced detectors I-02 (Figure 1) for January 7, 2010 as referenced from the upstream BOG (I-01). Callout (i) represents the modal travel time for the coordinated platoon peak, while callout (ii) represents the modal travel time for the secondary platoon peak of vehicles entering the corridor from side streets.

Figure 6: Observed platoon shift and dispersion at I-02 under snow and clear conditions.



a) Southbound arrival time at I-02 SB Detector referenced to I-01 SB BOG, a.m. peak 06:00 – 09:00 (snow event on Thursday, Jan. 7, 2010).



b) Southbound arrival time at I-02 SB Detector referenced to I-01 SB BOG, a.m. peak 06:00 – 09:00 (clear weather on Thursday, Jan. 14, 2010).

Table 3: Flow profile intersection arrival times for a.m. peak southbound referenced to upstream start of green.

(Weather) Date	Link	Modal Travel Time (s)	Shift in Platoon Peak Travel Time (s)	Effective Link Speed (MPH)	Delta Speed (MPH)	Full width at Half-Max $M_e - M_s$ (s)
(SNOW) Th; 1/7/10	I-01 to I-02	65	15	24.3	7.3	50
(CLEAR) Th; 1/14/10		50		31.6		25
(SNOW) Th; 1/7/10	I-02 to I-03	75	25	23.0	11.6	65
(CLEAR) Th; 1/14/10		50		34.6		40
(SNOW) Th; 1/7/10	I-03 to I-04	115	30	21.6	7.6	75
(CLEAR) Th; 1/14/10		85		29.2		55

- Figure 6.b. shows the aggregated platoon profile during the a.m. peak (06:00 – 09:00) non-snow event measured at the advanced detectors I-02 (Figure 1) for January 7, 2010 as referenced from the upstream BOG (I-01). Callout (iii) represents the modal travel time for the coordinated platoon peak, while callout (iv) represents the modal travel time for the secondary platoon peak of vehicles entering the corridor from side streets.

- Based on the measured increase in link travel speed and approach headway, the expected modal arrival time of the platoon shifted resulting in the arrival 15 seconds later in Figure 6.a., callout (i) than in Figure 6.b., callout (iii). Also represented in the graph is a full-width half-max measurement (FWHM),

$$FWHM_i = M_e - M_s \quad (1)$$

where M_e is the end of the maximum width and M_s is the start of the maximum width measured at half the peak value. This measurement is generally used to measure the distribution of data with respect to its peak. Similarly, a platoon can be characterized as a peak in a distribution. The FWHM, in its simplest form as shown in Equation 1, was used to quantify the change in platoon dispersion due to snowfall. A summary table of these findings is found in Table 3. For each travel link, an increase in the platoon modal arrival time at the advance detector was measured. The FWHM increased on each of the travel links. These measures indicate that arrival times shifted (corresponding to longer travel times) and platoons were more dispersed during the January 7, 2010 a.m. peak snow event.

Coordination, Visualization, and Optimization

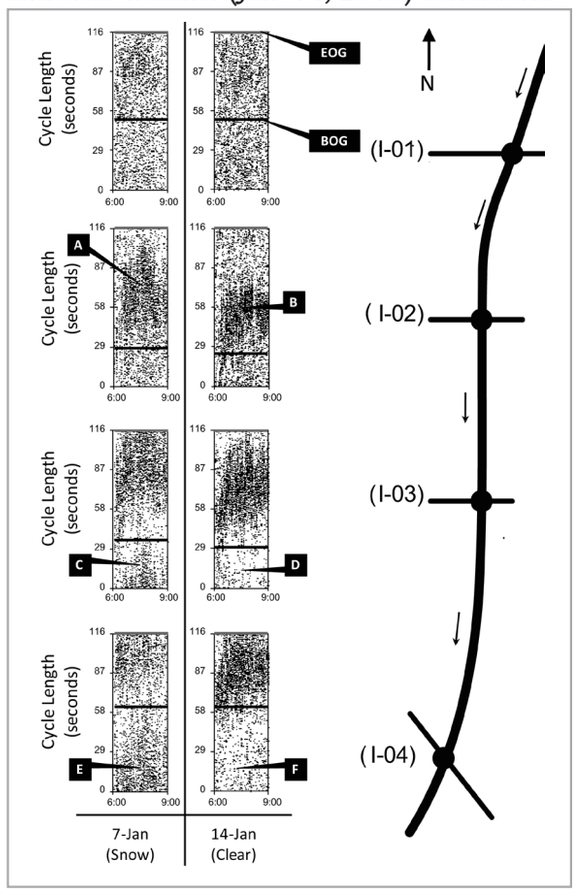
Visualization

Because the timing plan for the corridor was designed for clear-weather conditions with normal travel time speeds and approach headways, it is desirable to know whether this timing plan performed well under adverse weather conditions, particularly the offsets establishing the coordination pattern. This section examines the performance of the coordination pattern using the Purdue Coordination Diagram (PCD), a tool recently proposed for visualization of coordination events.[4, 5]

In Figure 7, eight PCDs for the southbound movements on SR 37 under snow (January 7) and clear (January 14) conditions are shown. In these graphs, each point represents a vehicle arrival. The vertical

position of the point shows the time in the cycle that the vehicle arrived, while the horizontal position is related to the time of day and hence the cycle in which the arrival occurred. The two horizontal lines indicate the average beginning of green (BOG) and end of green (EOG) times during the analysis period. The time between zero and BOG is when the coordinated phase is in red, while between BOG and EOG the phase is green. When vehicles are clustered together (callouts "A" and "B"), this is evidence of a platoon of vehicles arriving at an advance detector with respect to the BOG and EOG. In Figure 7, platoons are observed on each movement as inferred by the clusters of vehicle arrivals except for southbound at I-01, which has random arrivals. As shown in Figure 7, the majority of the platoons (callouts "A" and "B") arrive on green, as compared to callout "E" where the majority of the platoon arrives on red.

Figure 7: PCD comparison of a.m. peak 06:00 – 09:00 during Thursday snow (Jan. 7, 2010) and clear weather (Jan. 14, 2010) conditions.



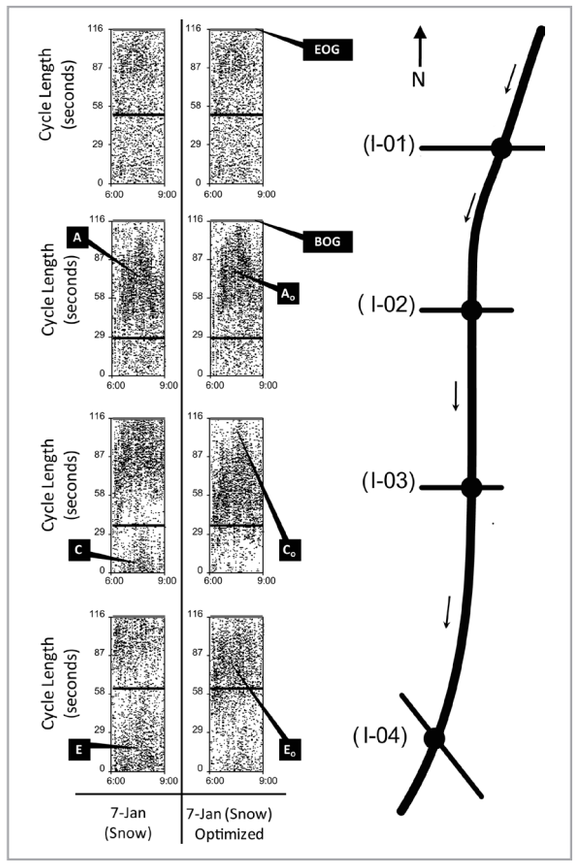
By comparing the PCDs from the snow event (in Figure 7, left column of PCDs) to clear weather (right column), it is possible to observe changes in the travel times and dispersion of platoons. At I-03, for example, during clear conditions the platoon arrives almost entirely within the green band, with no platoons observed to arrive on red (callout "D"). During snow conditions at I-03, the considerably more spread out platoon begins to arrive later in the cycle; in fact, the platoon is cut off by the end of green, resulting in a portion of the platoon arriving on red (callout "C"). The changes to the platoons are even more pronounced at I-04. During clear conditions most of the southbound platoons are captured within the green band, with a minimal part of the platoon arriving on red (callout "F"). This was also observed at I-03 (callout "D"). However, during the snow event, the platoon arrival characteristics were almost totally inverted, with the majority of the platoon arriving during red (callout "E"). From these graphs, it is apparent that there are opportunities to adjust the coordinated signal offsets times to promote platoon arrivals on green and thus improve coordination during a snow event.

Offset Optimization

To determine whether a benefit could be obtained by changing the signal offsets, the data collected during the snow conditions were used to optimize the offsets. The optimization objective used was delay minimization. Delay was calculated using an input-output procedure that effectively estimated the area between the arrival and departure curves.[19, 20] More specifically, the number of arrivals and number of departures (estimated by assuming saturation flow during green time) were tabulated for an average observed cycle, enabling an estimate of the queue length by numerically tracking the number of vehicles entering and exiting the system. Estimated delay is calculated by summing the estimated queue length over the cycle.

Having thus calculated delay for the coordinated approaches in the system, it was possible to use this information to optimize the offsets. To summarize that process, the combination method was applied to high-resolution data using a customized algorithm for a chain of intersections on a two-way arterial.[5] This algorithm finds the global optimum by executing a limited enumerative search on successive links in the system, while preserving previously optimized link flows by applying the additional adjustments to each additional intersection to all previously optimized intersections. More extensive details of the optimization algorithm and method for estimating delay based upon high-resolution data are described elsewhere.[4, 5] Snow event offsets were optimized to minimize delay for the southbound movements only at I-02, I-03, and I-04, with vehicles arriving at I-01 being random (Figure 7). Offsets were also calculated to minimize delay for the entire system in the northbound and southbound direction. Vehicles arriving at I-04 northbound and I-01 southbound were found to be random and therefore not considered in the optimization algorithm.

Figure 8. Comparison of southbound SR37 PCDs for normal offset times and proposed optimal offsets to minimize delay for a.m. peak 06:00 – 09:00 during Thursday snow (Jan. 7, 2010) conditions.



The results of the optimization procedure are shown in Figure 8, with corresponding data in Table 4, where "N" is the number of total vehicles, and "N_g" is the total number of vehicles arriving on green. In Figure 8, the left column of PCDs represents the observed conditions during the snow event, while the right column shows the predicted platoon arrivals if the optimal offsets (southbound only) were implemented. The optimal offsets would cause arrivals at I-03 to be shifted several seconds earlier, preventing the tail end of the platoon (callout "C") from being cut off with the prediction that the majority of the platoon would arrive in the green band (callout "C_o"). At I-04, the predicted offset shift needed to minimize delay is associated with a similar shift. The platoon previously cut off at the end of green (callout "E") is predicted to shift as shown by callout "E_o." A more pronounced potential improvement to the coordination is seen at I-04, where the cluster called out by "E" is almost entirely shifted into the green band "E_o."

Figure 9 shows the vehicle arrival patterns as cyclic flow profiles. In each plot, the black bars represent the probability of a vehicle arrival during a particular time in cycle. This is equivalent to a TRANSYT flow profile.[2] The shaded region represents the probability of green for that time in cycle. The green bands are considerably wider at I-02 and I-03 because of the low demand for side-street phases, meaning that the coordinated phase receives considerable additional green time from gap outs of preceding phases. Nevertheless, delay is minimized at I-03 by shifting the platoon arrival (Figure 9, callout "A") to an earlier part of the cycle ("A_o"). At I-04, the green band is relatively narrow due to high side-street demand. In this case, it is not possible to capture the entire platoon without increasing the coordinated phase split.

Table 4: Optimal offsets to minimize delay in the southbound direction in the a.m. peak (06:00 – 09:00) during a Thursday snow event (Jan. 7, 2010).

Optimized for SB Only (does not account for NB direction)						
			N_g (Veh)		Link Delay (s/veh)	
Links (Southbound)	N (Veh)	Proposed Offsets (s)	Before Optimization	After Optimization	Before Optimization	After Optimization
Arrivals at I-0I	2133	0	796	796	25.1	25.1
I-01 to I-02	2981	108	2405	2384	0.5	0.2
I-02 to I-03	2582	23	1938	2056	2.8	0.1
I-03 to I-04	2331	57	976	1274	21.6	11.8
SB System	10027		6115	6510	11.2	8.2
	SB Percent Change		6.9 percent		-26.7 percent	
			N_g (Veh)		Link Delay (s/veh)	
Links (Southbound)	N (Veh)	Proposed Offsets (s)	Before Optimization	After Optimization	Before Optimization	After Optimization
I-02 to I-0I	756	0	363	319	7.9	25.1
I-02 to I-02	1475	108	816	762	4.4	0.2
I-04 to I-03	1486	23	1174	967	0.1	0.1
Arrivals at I-04	1252	57	425	425	13.2	11.8
NB System	4969		2778	2473	5.9	5.8
	NB Percent Change		-11.3 percent		-2.2 percent	
SB and NB Total System Percent Change			1.0 percent		-22.1 percent	

Nevertheless, there is a considerable opportunity to reduce delay by simply changing the offset to cause those vehicles to arrive at the beginning of green (Figure 9, callout "C_o") rather than the end of green ("C"). These potential improvements, which amount to a possible 26.7 percent decrease (3.0 seconds per vehicle) in overall corridor delay for the southbound direction, are summarized in Table 4, where "N" is the number of total vehicles, and "N_g" is the total number of vehicles arriving on green.

The greatest opportunity to reduce delay in the southbound direction was found at I-04, where a reduction of 9.8 seconds per vehicle could be achieved. Optimizing only for the southbound direction did result in a northbound vehicle delay decrease of 2.2 percent. Alternatively, when applying the algorithm to minimize vehicle delay in both the northbound

Figure 9: Comparison of southbound SR37 flow profiles for normal offset times and proposed optimal offsets to minimize delay for a.m. peak 06:00 – 09:00 during Thursday snow (Jan. 7, 2010) conditions.

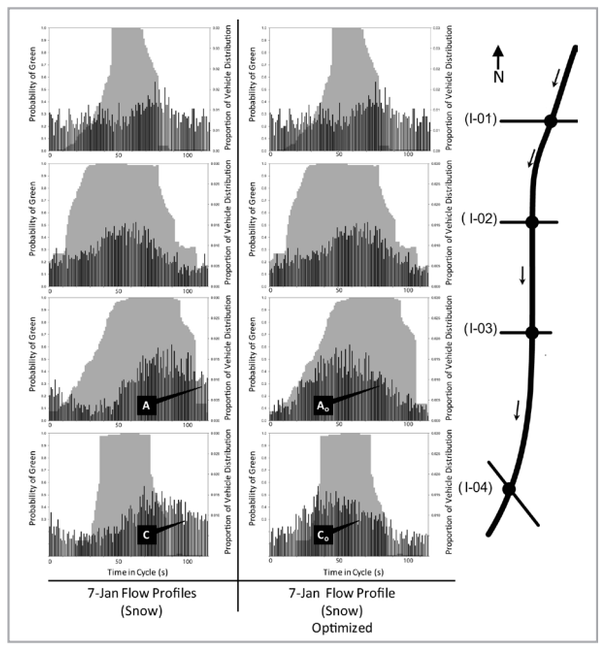


Table 5: Optimal offsets to minimize delay for north and southbound direction in the a.m. peak (06:00 - 09:00) during a Thursday snow (Jan. 7, 2010).

Optimized for SB and NB						
		Proposed	N_g (Veh)		Link Delay (s/veh)	
Links (Southbound)	N (Veh)	Offsets (s)	Before Optimization	After Optimization	Before Optimization	After Optimization
Arrivals at I-01	2133	0	796	796	25.1	25.1
I-01 to I-02	2981	112	2405	2400	0.5	0.3
I-02 to I-03	2582	35	1938	1913	2.8	0.4
I-03 to I-04	2331	67	976	1292	21.6	11.8
System	10027		6115	6401	11.2	8.3
SB System Percent Change			**4.6 percent**		**-25.9 percent**	
		Proposed	N_g (Veh)		Link Delay (s/veh)	
Links (Northbound)	N (Veh)	Offsets (s)	Before Optimization	After Optimization	Before Optimization	After Optimization
I-02 to I-01	756	0	363	340	7.9	8.1
I-02 to I-02	1475	112	816	833	4.4	1.3
I-04 to I-03	1486	35	1174	994	0.1	0.3
Arrivals at I-04	1252	67	425	386	13.2	13.9
System	4969		2778	2553	5.9	5.2
NB System Percent Change			**-8.1 percent**		**-11.7 percent**	
SB and NB Total System Percent Change			**0.7 percent**		**-23.1 percent**	

and southbound directions, the results shown in Table 5 indicate a possible delay reduction of 11.7 percent northbound and 25.9 percent southbound. When analyzing the overall corridor system delay, the results indicate a potential 22.1 percent reduction in total vehicle delay when optimizing in the southbound direction as opposed to considering northbound and southbound, which resulted in a potential reduction of 23.1 percent.

Conclusions and Recommendations

This paper demonstrated a set of measures to directly quantify traffic flow characteristics along an actuated-coordinated arterial. The system characteristics were measured using a combination of Bluetooth® probe data and high-resolution data (0.1 second) traffic signal controllers with logging capability. This data was used to characterize the following:

- Distributions of travel time along the corridor;
- Approach headways at the advance detectors; and
- Formation and dispersion of platoons referenced from the upstream traffic signal beginning of green.

These measures were made on a 1.6 mile (2.6 km), four intersection actuated-coordinated arterial to compare a snow event (Thursday, January 7, 2010) and clear weather (Thursday, January 14, 2010) for the southbound a.m. peak (06:00 – 09:00). The continuous Bluetooth® MAC address matching was used to identify increases in travel time related to adverse winter weather conditions. A southbound travel time increase of 83 seconds was measured during the snow event. This increase in corridor

travel time resulted in a measured increase (positive shift) of the modal platoon travel times by 15 seconds for I-02, 25 seconds for I-03, and 30 seconds for I-04. Platoon dispersion was quantified by calculating the full width at half max (FWHM) of the peak value in the platoon profiles. FWHM was found to increase by 25 seconds, 25 seconds, and 20 seconds respectively at I-02, I-03, and I-04, corresponding to more dispersed platoons. Increased platoon dispersion was further substantiated by measuring the median approach headway at the advance detectors, which increased from 1.4 to 2.0 seconds for the system. The PCD visually demonstrated both the shift and dispersion of the southbound a.m. peak platoons at each intersection during the snow event.

This paper outlined a set of metrics that can be measured and can provide a visual queue to determine platoon presence, if coordination is necessary during inclement weather conditions, and can determine whether there are opportunities to improve operations by adjusting offsets to account for winter weather-related increases in travel time. For this case study, running in coordination with adjusted offset times appears to be the best option. This recommendation is derived from the optimization of the offsets through the system resulting in predicted improved traffic flow. Based on the optimization algorithms developed,[5] an overall system delay reduction of 32,832 seconds (23.1 percent) can be obtained by minimizing delay for the northbound and southbound directions; a 22.1 percent reduction is possible from just minimizing southbound delay. Previous research had predicted a similar total delay reduction of 23 percent.[6] Focusing on the southbound direction, a greater reduction in the southbound delay of 30,784 seconds (26.7 percent) or approximately 3.0 seconds per vehicle could be achieved. Based on this finding, it is recommended that the corridor continue to run in coordination with alternative offsets programmed that minimize delay in the southbound direction. These offsets are based on the 70/30 directional split heavily in favor of southbound and can be enabled when a winter weather event is eminent during the weekday a.m. peak.

This work demonstrates opportunities for significant improvements in arterials during winter conditions through the changes to the offsets times. Further studies will need to be conducted to determine if changes to cycle lengths, split times, and gap acceptance are needed in addition to the offsets to yield additional improvements in the coordination through the system during a snow event. Although clearance intervals had been noted as an important parameter for improving winter timing6[6] this research has focused on coordination plans.

Acknowledgments

This work was supported by the Joint Transportation Research Program administered by the Indiana Department of Transportation and Purdue University. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the National Academy of Sciences, American Association of State Highway and Transportation Officials, Federal Highway Administration, or the Indiana Department of Transportation. These contents do not constitute a standard, specification, or regulation.

The controller hardware used in this research was the ASC/3 controller manufactured by Econolite Control Products, Inc.

References

1. Gordon, R. L. and T. Warren, (2005) Federal Highway Administration, *Traffic Control Systems Handbook*, Report Number. FHWA-HOP-06-006.

2. Robertson, D. I., (1969) "'TRANSYT' Method for Area Traffic Control." *Traffic Engineering and Control*, Vol. 11, pp. 276–281.

3. *Highway Capacity Manual* (HCM). (2000). "Highway Capacity Manual." SR, 209, Transportation Research Board, National Research Council, Washington, DC.

4. Day, C. M., R. Haseman, H. Premachandra, T.M. Brennan, J. Wasson, J. R. Sturdevant, and D. M. Bullock, (2010). "Visualization and Quantitative Assessment of Arterial Progression Quality and Travel Time Using High Resolution Signal Event Data and Bluetooth MAC Address Matching," Paper ID: 10-0039 in press.

5. Day, D. M. and D. M. Bullock, (2010). "Arterial Performance Measures, Volume 1: Performance Based Management of Arterial Traffic Signal Systems", Federal Highway Administration, Washington, DC.

6. Perrin, H. J., P. T. Martin, and B. G. Hansen, (2001). Modifying Signal Timing During Inclement Weather, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1748, TRB, National Research Council, Washington, DC, pp. 66–71.

7. *Anchorage Signal System Upgrade—Final Report* (1995). Bernardin Lochmueller and Associates, Inc., Anchorage, Alaska.

8. Goodwin, L. C., & P. A. Pisano, (2004). Weather-Responsive Traffic Signal Control. Institute of Transportation Engineers. *ITE Journal*, 74(6), pp. 28–33.

9. Mahmassani, H.S., J. Dong, J. Kim, R. B. Chen, and B. Park, (2009). "Incorporating Weather Impacts in Traffic Estimation and Prediction Systems", Federal Highway Administration, Washington, DC. Report # FHWA-JPO-09-065.

10. Botha, J. L. and T. R. Kruse, (1992). "Flow Rates at Signalized Intersections under Cold Weather Conditions," *Journal of Transportation Engineering*, pp. 439–450.

11. Berry, D. S., and P. K. Gandhi, (1973). "Headway Approach to Intersection Capacity" *Highway Research Record 453*, HRB, National Research Council, Washington, DC, pp. 56–60.

12. Agbolosu-Amison, S. J., A. W. Sadek, and E. Wael, (2004). "Inclement Weather and Traffic Flow at Signalized Intersections." *Transportation Research Record No. 1867*, TRB, National Research Council, Washington, DC, pp. 163–171.

13. Agbolosu-Amison, S. J., A. W. Sadek, and B. Henry, (2005). "Factors Affecting Benefits of Implementing Special Timing Plans for Inclement Weather Conditions." *Transportation Research Record* No. 1925, TRB, National Research Council, Washington, DC, pp. 146–155.

14. Al-Kaisy, A. and Z. Freedman, (2006). "Weather-Responsive Signal Timing: Practical Guidelines." *Transportation Research Record No. 1978*, TRB, National Research Council, Washington, DC, pp. 49–60.

15. Lieu, H. C. and S. Lin, (2004) Benefit Assessment of Implementing Weather-Specific Signal Timing Plans by Using CORSIM, *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1867*, TRB, National Research Council, Washington, DC, pp. 202–209.

16. Smaglik, E. J., A. Sharma, D.M. Bullock, J. R. Sturdevant, and G. Duncan, (2007). "Event-Based Data Collection for Generating Actuated Controller Performance Measures." In *Transportation Research Record No. 2035*, Transportation Research Board of the National Academies, Washington, DC, pp. 97–106.

17. Brennan, T. M., J. M. Ernst, C. M. Day, D. M Bullock, J. V. Krogmeier, and M. Martchouk "Influence Of Vertical Sensor Placement On Data Collection Efficiency From Bluetooth® MAC Address Collection Devices," Paper No. TEENG-418, in Press.

18. Wasson, J. S., J. R. Sturdevant and D. M. Bullock (2008) "Real-Time Travel Time Estimates Using Media Access Control Address Matching." *ITE Journal*, 78(6), 20–3.

19. Sharma, A., D. M. Bullock, and J. Bonneson. "Input-Output and Hybrid Techniques for Real-Time Prediction of Delay and Maximum Queue Length at a Signalized Intersection." *Transportation Research Record No. 2035*, Transportation Research Board of the National Academies, Washington, DC, pp. 25–33, 2006.

20. Sharma, A. and D. M. Bullock. "Field Evaluation of Alternative Real-Time Methods for Estimating Delay at Signalized Intersections." Proceedings of the 10th International Conference on Applications of Advanced Technologies in Transportation, Athens, Greece, May 27–31, 2008.

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.

CHRISTOPHER M. DAY, Ph.D. is a senior research scientist in the Joint Transportation Research Program at Purdue University. His research has been on traffic control systems performance measurement and optimization. He is a member of ITE.

JASON S. WASSON, P.E. is the director of the Indiana Department of Transportation Research & Development Division. He has extensive experience in the area of traffic operations, highway incident management, and special event management and is a past president of the Indiana Section of ITE.

JAMES R. STURDEVANT has worked in traffic operations for 15 years in positions with Indiana Department of Transportation Division and the Federal Highway Administration's National Resource Center Operations Team. Sturdevant is a licensed professional engineer in Indiana. 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.

Institute of Transportation Engineers
1627 Eye Street, NW, Suite 600
Washington, DC 20006
www.ite.org