Evaluation of Bias in the Hamburg Wheel Tracking Device

Final Report
For
RB00-010

October 2013

Sponsored by:
Iowa Department of Transportation
And
The Federal Highway Administration
SP&R Project # RB00-010

By
Scott Schram

Iowa Department of Transportation
Office of Construction and Materials
800 Lincoln Way
Ames, IA 50010
515-239-1604

Iowa Department of Transportation

Disclaimers

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the sponsors.

The sponsors assume no liability for the contents or use of the information contained in this document. This report does not constitute a standard, specification, or regulation. The sponsors do not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objectives of the document.

Statement of Non-Discrimination

Federal and state laws prohibit employment and/or public accommodation discrimination on the basis of age, color, creed, disability, gender identity, national origin, pregnancy, race, religion, sex, sexual orientation or veteran's status. If you believe you have been discriminated against, please contact the Iowa Civil Rights Commission at 800-457-4416 or Iowa Department of Transportation's affirmative action officer. If you need accommodations because of a disability to access the Iowa Department of Transportation's services, contact the agency's affirmative action officer at 800-262-0003.

1. **Report No.** **RB00-010**	2. **Government Accession No.** Optional	3. **Recipient Catalog No.** Optional

4 **Title and Subtitle** Evaluation of Bias in the Hamburg Wheel Tracking Device	5 **Report Date** September 25, 2013 6 **Performing Organization Code** Optional
7. **Author(s)** Scott Schram, Iowa DOT R. Christopher Williams, Iowa State University	8 **Performing Organization Report No.** Optional
9 **Performing Organization Name and Address** Iowa Department of Transportation Office of Construction and Materials 800 Lincoln Way Ames, Iowa 50010	10 **Work Unit No. (TRAIS)** Not Required 11 **Contract or Grant No.** Optional
12 **Sponsoring Organization Name and Address** Iowa Department of Transportation 800 Lincoln Way Ames, Iowa 50010 Federal Highway Administration	13 **Type of Report and Period Covered** Final Report 14 **Sponsoring Agency Code** **RB00-010**

15 **Abstract**

As the list of states adopting the HWTD continues to grow, there is a need to evaluate how results are utilized. AASHTO T 324 does not standardize the analysis and reporting of test results. Furthermore, processing and reporting of the results among manufacturers is not uniform. This is partly due to the variation among agency reporting requirements. Some include only the midpoint rut depth, while others include the average across the entire length of the wheel track.

To eliminate bias in reporting, statistical analysis was performed on over 150 test runs on gyratory specimens. Measurement location was found to be a source of significant variation in the HWTD. This is likely due to the non-uniform wheel speed across the specimen, geometry of the specimen, and air void profile. Eliminating this source of bias when reporting results is feasible though is dependent upon the average rut depth at the final pass. When reporting rut depth at the final pass, it is suggested for poor performing samples to average measurement locations near the interface of the adjoining gyratory specimens. This is necessary due to the wheel lipping on the mold. For all other samples it is reasonable to only eliminate the 3 locations furthest from the gear house. For multi-wheel units, wheel side was also found to be significant for poor and good performing samples. After eliminating the suggested measurements from the analysis, the wheel was no longer a significant source of variation

16 **Key Words** Hamburg Wheel Tracking Device, Bias, LVDT Location	17 **Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

18 **Security Classification (of this report)** Unclassified	19 **Security Classification (of this page)** Unclassified	20 **No. of pages** 18	21 **Price** N/A

DOT Form F1700

(This page left intentionally blank.)

INTRODUCTION

The Hamburg Wheel Tracking Device (HWTD) has been used for nearly three decades worldwide, and has recently gained popularity among researchers and state highway agencies (SHA) in the U.S. The device measures the combined effects of rutting potential and moisture-induced damage by rolling a steel wheel across the surface of an asphalt specimen immersed in a temperature controlled water bath. The only measurement collected by the HWTD is cumulative impression depth per wheel pass.

In 2007, 2 states reported the use of HWTD in their specifications. The most recent survey shows this number has now grown to 7. In addition to these 7, several agencies and consultants have purchased a device for research and evaluation. It is now commonly used in experimental designs for local and national research projects, qualified products evaluations, and as a mix design tool. More states are expected to implement HWTD in a quality assurance capacity.

Among the state agencies using or evaluating the HWTD are:

- California (specification)
- Colorado (specification)
- Florida
- Illinois (special provision)
- Iowa (specification)
- Louisiana
- Montana (specification)
- Oklahoma
- Nevada
- Texas (specification)
- Utah (specification)
- Wyoming

PROBLEM STATEMENT

As the list of states adopting the HWTD continues to grow, there is a need to evaluate how results are utilized. A testing protocol was adopted by AASHTO (AASHTO T 324 Standard Method of Test for Hamburg Wheel-Track Testing of compacted Hot-Mix Asphalt (HMA)). In practice, a maximum rutting threshold is commonly set to define a test failure. However, T-324 does not standardize the analysis and reporting of test results. For example, the Montana Department of Transportation averages the middle seven points (1) as the tests' final rut depth, while the Colorado DOT uses the midpoint rut depth (2). The Utah DOT considers the maximum deformation in any location but disregards two inches on either end of the wheel path (3). Both the TxDOT and Caltrans test methods do not specify how the rut depth at failure is to be obtained (4, 5), while others simply rely on the results as reported by the manufacturers' software.

Rut depth is not the only HWTD metric used to evaluate asphalt mixtures. The stripping inflection point (SIP) has been increasingly used to evaluate moisture susceptibility (NCHPR 9-49), yet its calculation is clearly and consistently defined. Other metrics used include stripping slope, creep/rut slope, and slope ratio (6).

Processing and reporting of the results among manufacturers is not uniform. This is partly due to the variation among agency reporting requirements. Analysis is needed to assess these differences and reduce or eliminate any bias in the reported results. Information is limited regarding how the location of each measurement contributes to bias. Furthermore, bias may exist in multi-wheel units. Some address this by setting thresholds for the maximum difference between both wheels. This paper provides recommendations for reducing or eliminating bias from reported HWTD results.

DESIGN OF EXPERIMENT

Samples collected over 2012 and 2013 were tested in the IDOT Central Materials laboratory with a two-wheel HWTD manufactured by Precision Metal Works. Linear variable displacement transducers (LVDTs) measure rut depths at eleven locations across the wheel track per pass. Measurements were recorded to the nearest 0.01 mm every 20th pass for the first 1,000 passes. The frequency was reduced to every 50th pass thereafter. The target wheel speed was 50 passes per minute. All samples were tested submerged in a 50°C ± 0.5°C water bath.

Of the 153 tests, 142 were run on adjoining gyratory compacted specimens, while the remaining samples were adjoining field cores. Replicates were simultaneously tested in both wheels. All gyratory specimens were compacted to 7.0% ± 1.0% air voids with the exception of 2, which were compacted to a void level matching that of a companion field core. To reduce the size of the dataset, only rut depths at 6 pass counts were considered (1, 2, 5, 10, 15, and 20 thousand passes). The objective was to perform a complete statistical analysis to assess the following:

1. Statistical differences between the two wheels
2. Statistical differences among the eleven measurement locations.
3. Statistical differences among the eleven measurement locations within each wheel.

While material properties were not included as a factor, the distribution among mix design levels and binder grades is shown in Table 1. Mixes were selected to ensure an a variation in performance could be observed in frequencies large enough to support statistical analyses across pass counts.

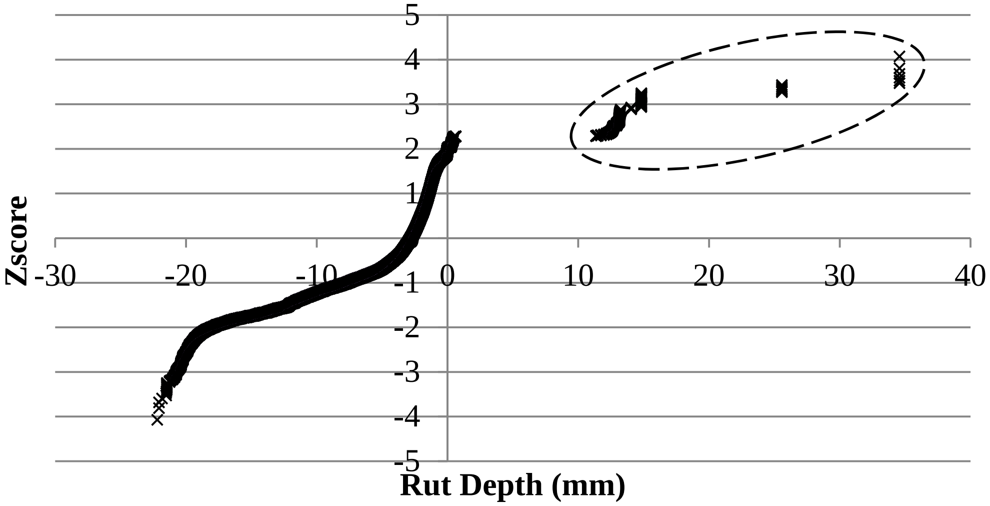
TABLE 1 Sample Distribution

Design ESALS (millions)	Design Gyrations	Number of Tests Binder Performance Grade (PG)			
		58-28	64-22	64-28	70-22
1M	76	7	7	0	0
3M	86	42	31	7	0
10M	96	0	40	8	0
30M	109	0	0	10	10

Normality Check

A preliminary check for normality on the rutting data across all measurement locations and 6 pass levels is shown in Figure 1. The disparity among mixture performance is apparent and expected. Because not all mixes survive an equal number of passes, the frequency at larger rut depths will be less than that at smaller rut depths. In some test runs, positive impressions are reported (typically at the beginning and ending of the test). These results are considered erroneous. Rut depth is considered synonymous with impression depth throughout the paper and will be considered a negative value.

(a) Normality Plot



(b) Histogram

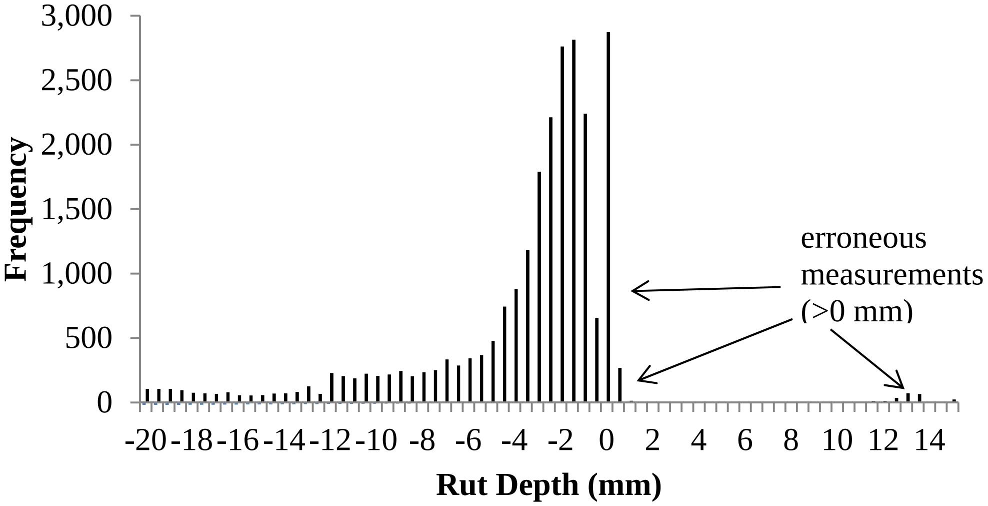


FIGURE 1 Normality Plot and Histogram of Rutting

When the data populations are blocked by their surviving pass count, the distribution becomes increasingly skewed per pass count as shown in Figure 2. Mixes surviving the full 20,000 passes had observed rut depths from 20 mm to 0.5 mm. It was decided a more rational approach to judging normality was to only consider the data captured at the final pass, which is how results are typically applied in practice.

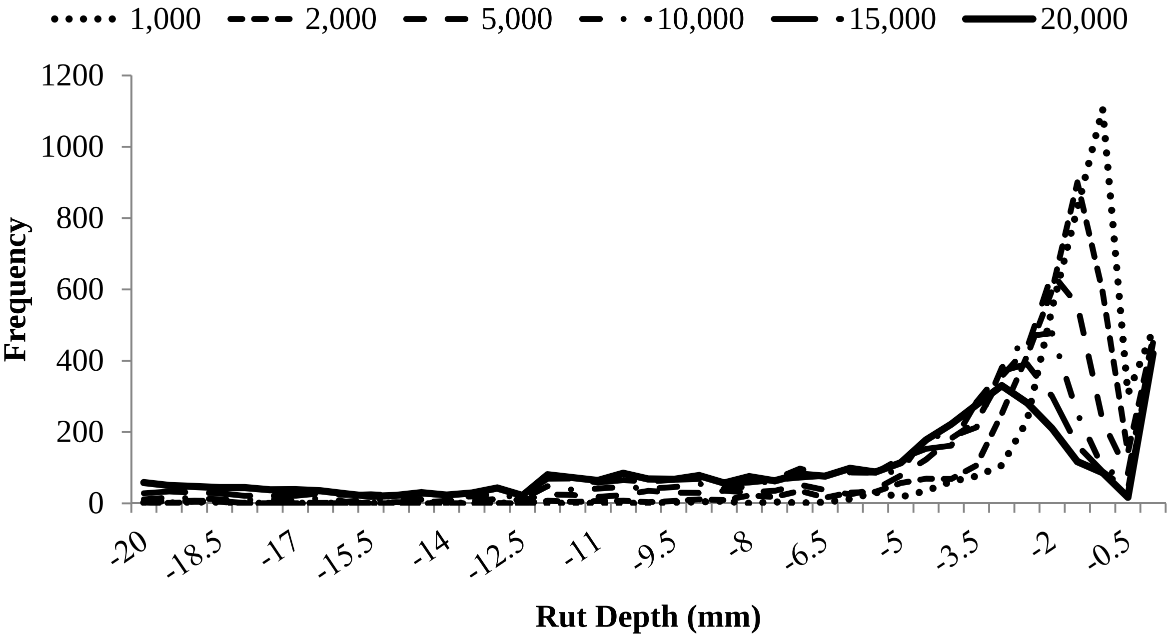


FIGURE 2 Frequency Distribution by Pass Count

When checking the normality of the dataset at the final pass a non-normal distribution was revealed in Figures 3 and 4. The "humps" in the normality plot confirm a tri-modal distribution, better illustrated in Figure 4.

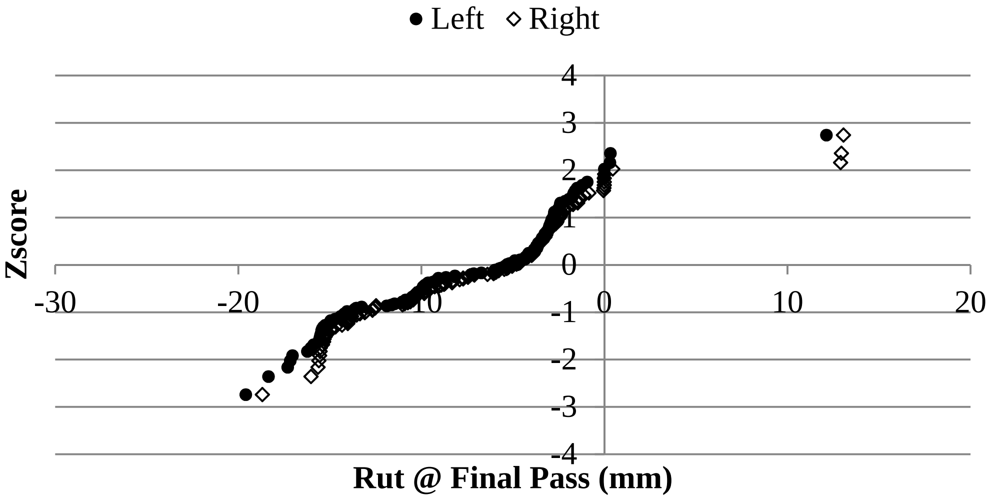


Figure 3 Normality Plots of Rutting at Final Pass

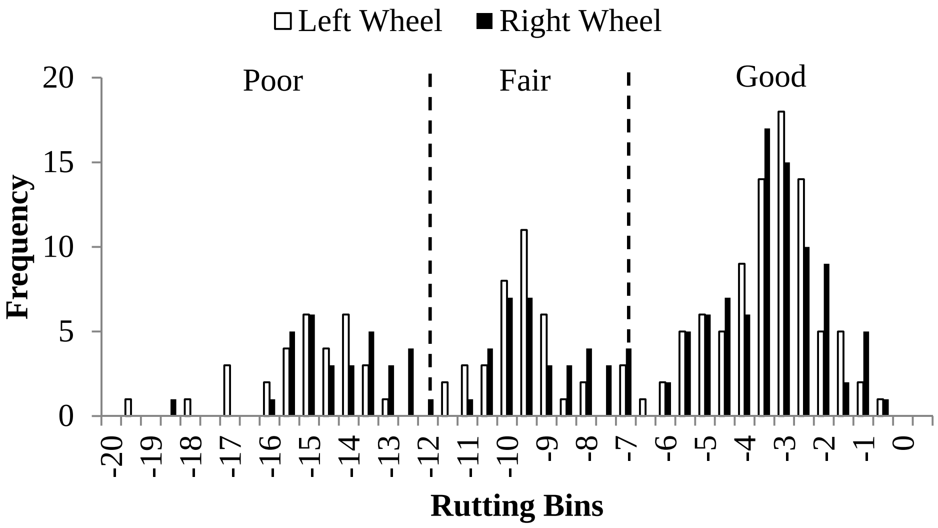


Figure 4 Histogram of Rutting at Final Pass

To utilize analysis of variance (ANOVA), the normality assumption can be satisfied by treating the dataset as three subsets. The subsets were determined by the following rutting ranges at the final pass:

- Subset A (Poor) > 12 mm rutting
- Subset B (Fair) 8 – 12 mm rutting
- Subset C (Good) ≤ 7 mm rutting

Because the response variable (rut depth) changes with each pass, the traditional approach would be to use analysis of covariance while treating pass count as a covariate. However, this approach assumes a linear relationship, which is not appropriate in case of typical nonlinear HWTD rut curves. Instead, a 2-Factor repeated measures ANOVA was used, with the different samples serving as the subjects in Table 2. Wheel and measurement location served as the factors. The measurement location is numbered in reverse order as the table reaches full extension from position #11 to position #1. The experiment was repeated for each of the three subsets.

TABLE 2 Experimental Design

Factor	Levels
A (Wheel)	2 (Left/Right)
B (Location)	11 (1/2/3/4/5/6/7/8/9/10/11)

RESULTS AND ANALYSIS (RUT DEPTH)

The ANOVA results are shown in Table 3 (α = 0.05). The common significant sources of variation among subsets are sample and measurement location. The sample's contribution to the variance is expected, since the rut depth is dependent upon material components and their response to loading. Measurement location and wheel are examined further.

TABLE 3 ANOVA Results For All Subsets (p-value)

Subset	A (Poor)	B (Fair)	C (Good)
Sample Size	39	36	78
BLOCK(Mix)	***0.00000***	***0.0000***	***0.0000***
A (Wheel)	***0.03824***	0.0747	***0.0083***
B (Location)	***0.00000***	***0.0000***	***0.0000***
AB	0.98098	0.5018	0.9661

Measurement Location

Measurement location was found to be a source of significant variation. This is likely due to the non-uniform wheel speed across the specimen. The geometry of the specimen and variation in air voids may also play a role. Figure 5 shows the rut depth across all measurement locations for the three subsets. For the poor performing tests, locations 7, 8, and 9 show the largest rut depth on average, while the first few locations appear to underestimate the rut depth observed throughout the majority of the specimen. This can be attributed to the wheel traveling up onto the mold as the wheel arm is fully extended. For the fair performing subset, the measurement location also appears to show bias near the end of the wheel track. The effect was less pronounced in the good performing subset.

Good

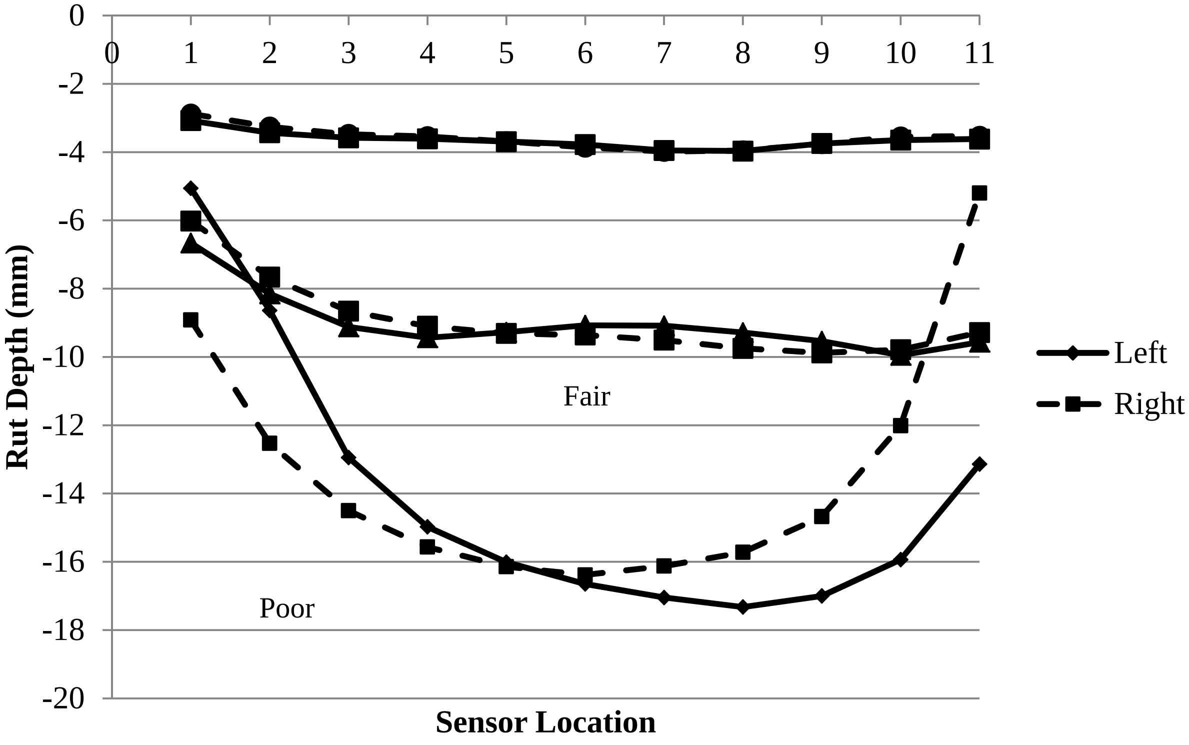


FIGURE 5 Rut Depth @ Final Pass (Subset A)

To determine which locations are statistically the same, Table 4 shows the results of a paired t-test between left and right wheels for each location. While the magnitude of the mean rut depth appears to be different between wheels at some measurement locations, a paired t-test reveals these differences are not significant for any location in the poor subset. For the fair and good subsets, the locations nearest the end of the wheel track are significantly different when comparing between wheels.

TABLE 4 Pairwise Comparisons of Rut Depth Between Left vs. Right Wheels (p-value)

	Measurement Location										
	1	2	3	4	5	6	7	8	9	10	11
Poor	0.38	0.58	0.64	0.64	0.71	0.72	0.67	0.39	0.38	0.37	0.27
Fair	***0.01***	***0.04***	0.10	0.21	0.69	0.78	0.50	0.43	0.62	0.60	0.50
Good	***0.03***	0.08	0.24	0.39	0.72	0.86	0.82	0.57	0.78	0.36	0.33

Comparisons between locations were also made within wheels as shown in Figures 6 and 7. Circles of the same size and pattern appear above locations that are statistical the same ($p \geq 0.05$). For example, location #1 is significantly different than all other locations and therefore a circle of similar size is not found above any of the other 10 measurement locations. This is true of all subsets for the first location (nearest the end of the wheel track). For the poor performing subset, the mean rut depth at locations 6, 7, 8, and 9 are not significantly different for either wheel. In the fair subset, with the exception of the first 3 sensors, it is reasonable to consider the mean rut depth among the remaining sensors statistically the same for both wheels. For the good performing subset, with the exception of the first 2 sensors and the sensors near the interface of the gyratory specimens, it is reasonable to consider the mean rut depth among the remaining sensors statistically the same for both wheels. Similar results were found for the fair subset. At an a 0.10 level of significance, locations

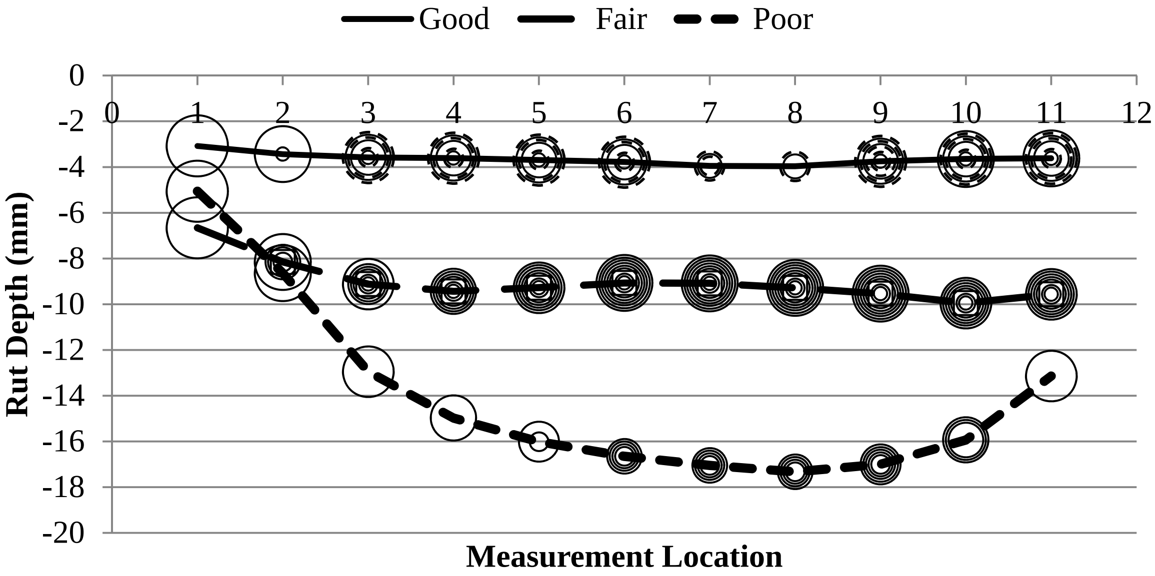


FIGURE 6 Rut Depth Comparison within Left Wheel

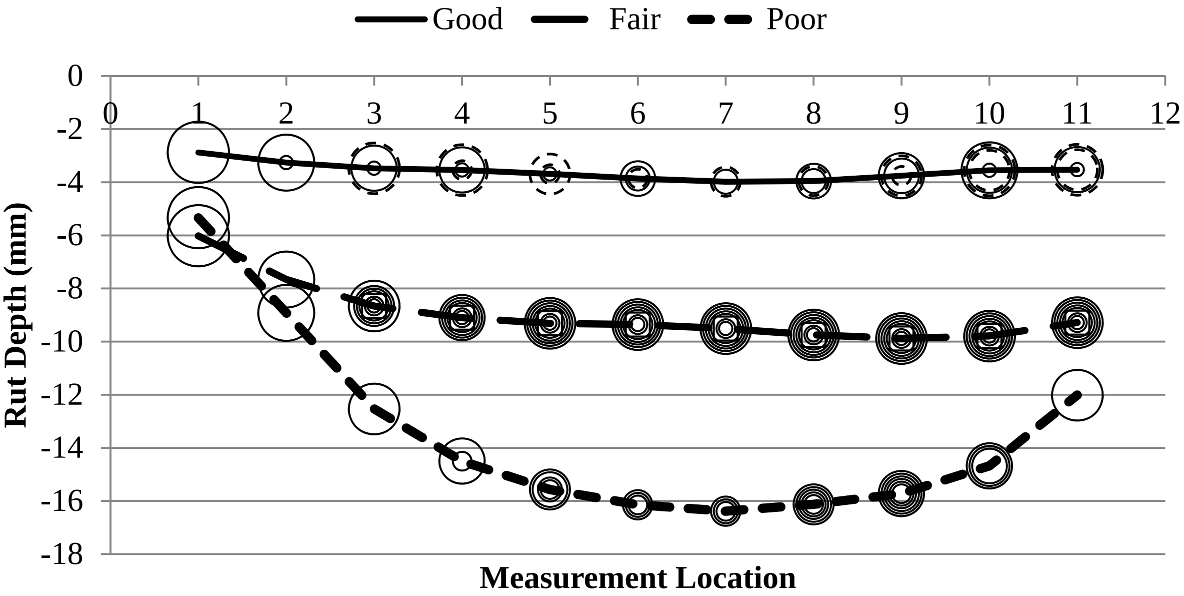


FIGURE 7 Rut Depth Comparison within Right Wheel

Alternatively, the bubble chart shown in Figure 11 depicts which locations are statistically different than. The larger the bubble size, the more locations that were found to be statistically different.

Alternatively, the bubble chart shown in Figure 7 depicts which locations are statistically different than. The larger the bubble size, the more locations that were found to be statistically different.

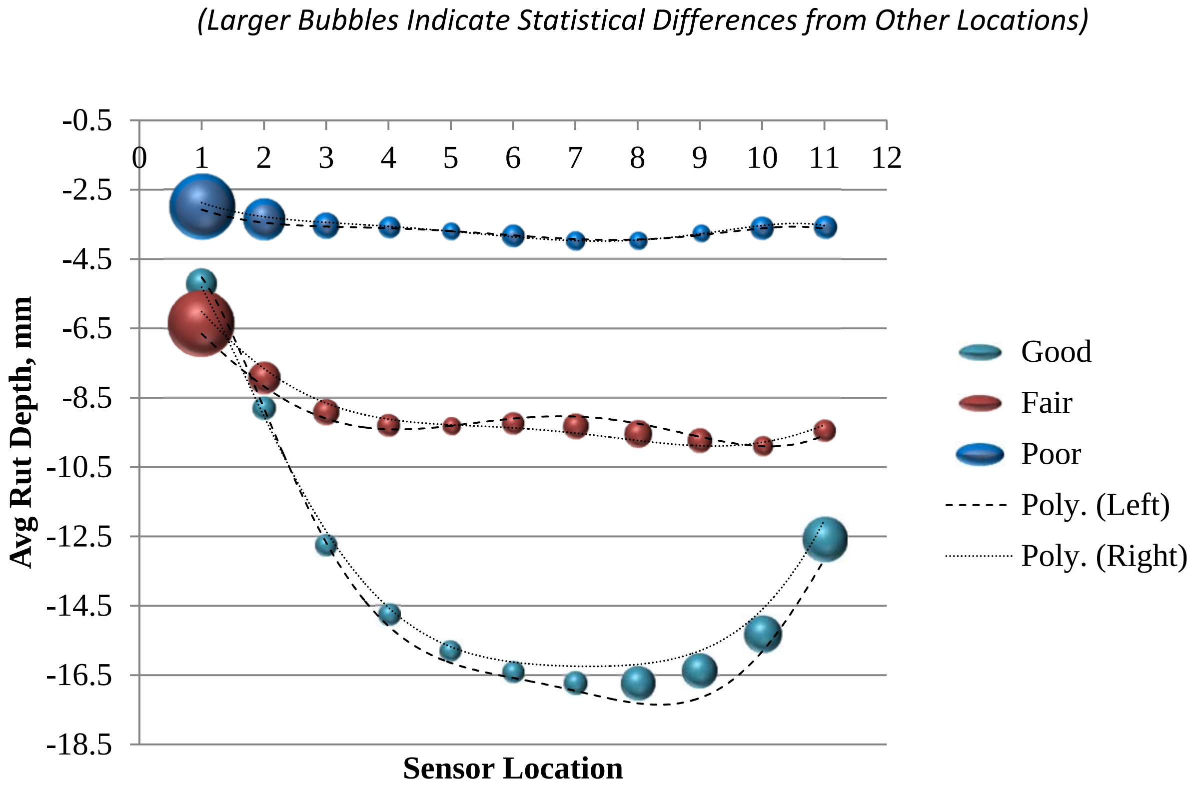


FIGURE 7 Bubble Chart for Rut Depth Comparisons Within Wheels (Subset A)

Rut depths at measurement locations 1, 2, 3, 4, 5, 10, and 11 are significantly different than the other locations. Locations 1 and 2 are the last two measurements furthest from the gear house, while locations 7 and 8 occur immediately after crossing the interface between the two gyratory specimens.

Wheel

Wheel side was found to be a significant source of variability when rutting at the final pass is more than 12 mm. While a significant difference in variance was identified from the ANOVA, there was no difference in means as shown in Table 5. While measurement location plays a role in the wheel's significance at lower pass levels a broader perspective was examined. Intuitively, some of the variability originates from the specimen, particularly the aggregate structure and air void level. This variability is more pronounced at early stages of the test when the material is undergoing what is known as seeding. The rate of rutting will vary early in the test among measurement locations as the aggregate particles orient. Once the particles are seeded, the rutting rate stabilizes in the creep zone until stripping ensues. Figure 8 shows typical seeding in the HWTD. The rate of seeding will vary across locations and wheels, contributing to the variability. It is therefore postulated that the significant difference in variability due to the wheel up to 5,000 passes can be attributed to seeding and measurement noise, especially in locations #1 and #11.

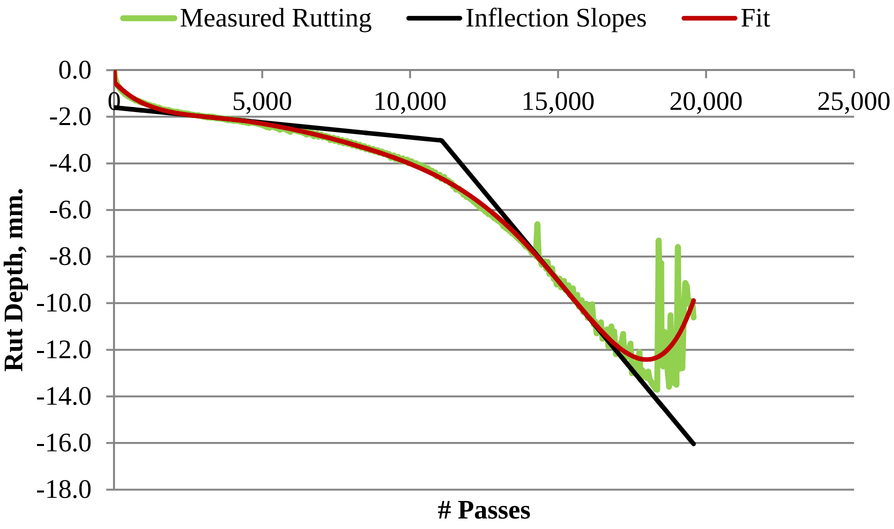


FIGURE 8 Typical Seeding of Specimen in HWTD

Subset B (8 – 12 mm Rutting)

Wheel

Unlike Subset A, the wheel side was not significant when the test results in an average rut depth between 8 and 12 mm.

Subset C (≤ 7 mm Rutting)

Wheel

The wheel was determined significant, solely due to measurement location 1. Removing the measurement from the analysis rectifies this difference.

CONCLUSIONS AND RECOMMENDATIONS

Measurement location was found to be a source of significant variation in the HWTD. This is likely due to the non-uniform wheel speed across the specimen, geometry of the specimen, and air void profile. Eliminating this source of bias when reporting results is feasible though is dependent upon the average rut depth at the final pass. When reporting rut depth at the final pass, it is suggested to average measurement locations 6 through 9 for poor performing samples. For all other samples it is reasonable to only eliminate the 3 locations furthest from the gear house. Wheel side was also found to be significant for poor and good performing samples. After eliminating the suggested measurements from the analysis, the wheel was no longer a significant source of variation. Furthermore, when analyzing SIP, measurement location #8 was significantly different between the left and right wheels.

REFERENCES

1. "Method of Test For Hamburg Wheel Track Testing of Compacted Bituminous Mixtures", MT 334-110-090211, Montana DOT, 2011.
2. "Hamburg Wheel-Track Testing of Compacted Bituminous Mixtures", CP-L 5112, Colorado DOT, 2009.
3. Method of Test for Hamburg Wheel-Track Testing of Compacted HMA, MOI 990, 2011.
4. "Test Procedure for Hamburg Wheel-Tracking Test", TEX-242-F, 2009.
5. Method of Test for Hamburg Wheel- Track Testing of Compacted HMA, Section 39-1.01D), Utah DOT, 2011.
6. Moisture Sensitivity Testing of HMA, I.M. 319 Iowa DOT, 2013.