## Quantifying the Variability of Production of Asphalt Mixtures through newly implemented Performance Tests for the Wisconsin Department of Transportation

by

Rachel Taylor Cousins

A thesis submitted to the Graduate Faculty of Auburn University in partial fulfillment of the requirements for the Degree of Master of Science

> Auburn, Alabama May 5, 2023

Keywords: Balanced Mix Design, Indirect Tensile Asphalt Cracking Index, Hamburg Wheel Tracking Test

Copyright 2023 by Rachel Taylor Cousins

Approved by

Randy C. West, Chair, PhD, Civil and Environmental Engineering Raquel Moraes Puchalski, PhD, Civil and Environmental Engineering Carolina Rodezno, PhD, Civil and Environmental Engineering Fan Yin, PhD, Civil and Environmental Engineering

#### Abstract

 This study aims to determine an appropriate standard deviation of Balanced Mix Design performance tests for Wisconsin specifications based on field-produced mixes. Identifying variability is an essential aspect of the pavement materials and construction industry. Federal Highway Administration (FHWA) and State Department of Transportation (DOTs) quantify the variability of material properties to manage quality. Typical asphalt mixture properties measured to assess variability have been binder content, aggregate gradation, and mix volumetrics. New performance tests are being used to assess the quality of an asphalt mix. These new performance tests included in the balanced mix design are rutting and cracking indicators on an asphalt mix.

This study used mixtures from ten shadow projects from various locations across Wisconsin to obtain representative production variability data to determine the within-lot pooled standard deviation. For this study, two performance tests, Hamburg Wheel Tracking Test (HWTT) and Indirect Tensile Asphalt Cracking Index (IDEAL-CT), were performed for the mixtures from the ten shadow projects and their representative lots/sublots.

The analysis methods used to quantify the variability include the standard deviation, coefficient of variation, normality, outlier test, and cumulative distribution function of production standard deviations. The final conclusions of this study indicated that asphalt content was the least variable quality characteristic measured with a COV of 2.8%. The most variable quality characteristic measured was HWTT passes to 12.5 mm with a COV of 16.6%. IDEAL-CT had a mean COV of 13.1%. The mean COV for air voids was 10.4%. The mean COV for HWTT CRD20k was 10.9%.

#### Acknowledgments

 I want to thank my advisor, Dr. Randy West, for his support throughout this project and the academic opportunities he provided me during my time at Auburn. Special thank you to Dr. David Timm for the semester I spent under his guidance and the opportunities to be a part of the test track instrumentation team. A thank you to my outstanding committee members, Dr. Raquel Moraes, Dr. Fan Yin, and Dr. Carolina Rodezno, for the time they spent reviewing this Thesis and the guidance they provided.

 I want to thank my fellow graduate students at the NCAT laboratory for all the support. My NCAT friends Megan Foshee, Elizabeth Turochy, and Tiana Lynn, whom I had the pleasure of getting to know and work with throughout this research journey. My roommate from France, Amelie Martin, who spent a semester with us at NCAT.

A special thank you to the research engineers, Adam Taylor, Nathan Moore, and Jason Moore, for their guidance in the laboratory. I would also like to thank Mrs. Vickie Adams for coming in early so I would have a lab buddy and for the countless times we went to get fried chicken and mac and cheese at the best spot in Auburn.

 The lifelong friends I made through Auburn Ducks Unlimited who made Auburn a good time and the support system they provided. They are indeed the best people and mean the world to me.

 I try to tell my Montana ranch family how much they have meant to me over these past two years, but I don't think it's possible to thank them enough. Mike and Maria, Mary and Dustin, Stephanie and the kiddos (who provided lovely drawings for my office), Rose and Derek, Joey, and the countless others in the Shields Valley community for their encouragement and for allowing me to spend my summers/school breaks in the best part of the country.

3

 I want to thank my parents, Lee and Karen, and my sister, Natalie, for the love, support, encouragement, and guidance they have given me through this journey.

 Lastly, a special thank you to the civil engineering department here at Auburn University for the financial support to continue seeking higher education. Thank you to the Wisconsin Department of Transportation and the National Center for Asphalt Technology for funding the research.

## Table of Contents







## List of Tables



# List of Figures





## List of Abbreviations



#### Chapter 1 – Introduction

## 1.1 Background

This research aims to quantify the overall variability of asphalt mixture performance test results for Balanced Mix Design (BMD) tests being considered for use in Quality Assurance (QA) by the Wisconsin Department of Transportation (WisDOT). Overall variabilities of traditional quality characteristics such as binder content, aggregate gradation, and mixture volumetrics properties have been well documented in previous studies. Overall production variability is used to measure product quality. However, very little research has been reported on the overall variability of new performance tests used in BMD tests. The WisDOT has selected the indirect tensile asphalt cracking test (IDEAL-CT) and Hamburg wheel tracking test as their two performance tests for BMD.

## 1.2 Research Objective

The main objectives of this thesis were to:

- Statistically analyze the overall variability of performance test results from ten shadow projects
- Recommend appropriate standard deviations for the BMD performance tests for Wisconsin specifications based on field-produced mixes

#### 1.3 Scope of Work

Several steps were accomplished to meet the research objectives. First, ten shadow projects were selected across Wisconsin from which asphalt mixtures were sampled during production at the same time as traditional QA samples for each sublot for two to three lots. A shadow project is a project on which additional (BMD) tests are conducted at a frequency similar to existing acceptance quality characteristics. The additional test results are only used for research purposes

and not used to influence the production process or material acceptance decisions. In total, 134 mixture samples were obtained for this study. These mixture samples were shipped to NCAT for performance testing. These mixture samples were volumetrically verified for specific gravity and air voids, and samples were compacted to specification requirements for performance testing. HWTT and IDEAL-CT tests were conducted on each sublot sample. The data from the performance tests were statistically analyzed using various methods to provide overall variability statistics for developing QA specifications for performance tests in Wisconsin.

#### 1.4 Organization of Thesis

This thesis has been organized into five chapters. Chapter One is the introduction: including the background, research objectives, scope of work, and the organization of this thesis. Chapter Two is a literature review on the background of BMD, variabilities in the production of asphalt paving materials, and test methods on IDEAL-CT and HWTT. Chapter Three focuses on the research plan, explaining the selection of shadow projects, testing plan, and method of analysis. Chapter Four discusses the results and the impact of the outlier analysis on the variability results. Chapter Five provides the conclusions and recommendations for this project.

#### Chapter 2 – Literature Review

#### 2.1 Background on Balanced Mix Design

Most state highway agencies currently use the Superpave mix design method (AASHTO M 323) and associated criteria for asphalt mix design specifications. Since many of these agencies have been dissatisfied with Superpave mixtures' cracking and durability performance, recent research efforts proposed moving toward a new mix design approach that directly assesses a mixture's resistance to prevalent distresses. Balanced Mix Design (BMD) is defined as "asphalt mix design using performance tests on appropriately conditioned specimens that address multiple modes of distress taking into consideration mix aging, traffic, climate, and location within the pavement structure" (AASHTO PP105-20). BMD usually includes two or more performance tests, such as rutting and cracking tests, to determine how well the mixture resists common forms of distress in asphalt pavements (West, R et al., 2021).

In 2021 the Wisconsin Department of Transportation (WisDOT) developed a draft special provision, the *HMA Pavement Balanced Mix Design*, to implement BMD (Wisconsin, 2021). The performance tests used in the special provision are the Hamburg Wheel-Tracking Test (HWTT) to evaluate the mixture for rutting resistance and moisture resistance and the Indirect Tensile Asphalt Cracking Test (IDEAL-CT) for cracking resistance (Wisconsin, 2021).

## 2.2 Variability in the Production of Asphalt Materials

Variability in the production of asphalt paving mixtures is an important measure to assess quality. One definition of quality states that "quality is inversely proportional to variability" (Montgomery, D). AASHTO R 9 Acceptance Sampling Plans for Highway Construction, recommends quantifying the "overall" variability of quality characteristics for QA programs (AASHTO R 9-05).

Hughes (1996) described overall production variability to consist of four components: testing variability, sampling variability, materials variability, and construction variability. Mathematically, overall variance  $(\sigma_o^2)$ , is the sum of the testing variance  $(\sigma_t^2)$ , sampling variance  $(\sigma_s^2)$ , materials variance  $(\sigma_m^2)$ , and construction variance  $(\sigma_c^2)$ , shown as Equation 1.

$$
\sigma_o^2 = \sigma_t^2 + \sigma_s^2 + \sigma_m^2 + \sigma_c^2
$$
 Equation 1

Where:



Between 1956 and 1962, construction materials tests were conducted at the AASHO Road Test, which documented the overall variabilities of material qualities encountered during the construction of the pavements (Hughes, 1996). After the results of the AASHO Road Tests were published, many highway agencies established statistically based specifications using the variabilities of typical materials and construction processes (Hughes, 2005).

## 2.2.2 Typical Variability in Asphalt Mixtures

In NCHRP Synthesis 232 completed in 1996, Hughes summarized variabilities of common acceptance quality characteristics such as laboratory compacted, air voids, gradation, and asphalt content based on data obtained through random sampling procedures (Hughes, 1996). By the 1970s, statistically, based-specifications had been incorporated into QA programs with a strong dependence on statistical analysis (Halstead, 1979). Other asphalt material

properties often studied include: gradation, asphalt material viscosity, and asphalt binder penetration (Solaimanian et al. 1995).

#### 2.2.3 Uses of Variability to Establish Specification Limits

AASSHTO R 9, Standard Practice for Acceptance Sampling Plans for Highway Construction, explains types of acceptance plans and states that a "statistical acceptance plan is one based on analysis of either variables or attributes" (AASHTO R 9-05). The standard gives an example acceptance plan using the Percent Within Limits (PWL) procedure based on population, estimates of central tendency, and variability (AASHTO R 9-05). The example given in AASHTO R 9 Appendix X1 includes an analysis of "target miss" variability; this is an important case where quality characteristic has a specified target value and upper and lower specification limits are set above and below the target value (West et al., 2023).

## 2.2.4 Current Quality Control/Quality Acceptance in Wisconsin

 WisDOT developed its hot mix asphalt (HMA) quality management program (QMP) in the early 1990s. QMP is considered a best construction practice to ensure that an agency receives quality construction materials produced by a contractor (Faheem et al., 2018). Developing a QMP specification involved identifying key asphalt mixture parameters related to long-term pavement performance and the development of the agency's quality assurance (QA) program, including procedures for quality assurance (QA) and quality verification (QV) (Faheem et al., 2018). The asphalt pavement acceptance quality characteristics in Wisconsin's QMP are aggregate gradation, asphalt content, air voids, voids in the mineral aggregate, and in-place density (Faheem et al., 2018).

## 2.3 Test Methods

#### 2.3.1 IDEAL-CT

The IDEAL-CT is an asphalt mixture performance test to assess cracking resistance using laboratory-prepared cylindrical specimens developed for mix design and quality assurance testing (Zhou 2019). According to the test method, ASTM D8225-19, a cylindrical specimen is centered in the indirect tensile test fixture, and a load is applied at a rate of 50.0+2.0 mm/min. The load and the vertical displacement measured during the test are used to calculate the  $CT_{Index}$ (Figure 2.3.1). The  $CT_{Index}$  is calculated from failure energy, the post-peak slope of the loaddisplacement curve, and deformation at 75% of the peak load (shown in Equation 2.3.1).

$$
CT_{Index} = \frac{t}{62} * \frac{l_{75}}{D} * \frac{G_f}{|m_{75}|} * 10^6
$$
 Eq. 2.3.1

Where:

 $CT_{Index}$  = cracking tolerance index

 $G_f$  = fracture energy (J/m<sup>2</sup>)

 $|m_{75}|$  = absolute value of the post-peak slope m<sub>75</sub> (N/m)

 $l_{75}$  = displacement at 75% of the peak load after the peak (mm)

 $D$  = specimen diameter (mm)

 $t =$  specimen thickness (mm)



Figure 2.3.1: Load vs. LLD Data (Zhou, 2019)

#### 2.3.2 IDEAL-CT Variability

A Texas A&M Transportation Institute study reported the testing variability (repeatability) of the IDEAL-CT test based on its sensitivity to asphalt mix characteristics and conditions. The CT<sub>Index</sub> was sensitive to RAP and RAS content, asphalt binder type, binder content, and aging conditions. The highest within-lab COV was 23.5%, and most COVs were less than 20% (Zhou, 2019).

The Utah Department of Transportation conducted a study comparing the IDEAL-CT and I-FIT cracking tests to determine a feasible candidate for the cracking test in their BMD implementation. The study compared within and between lab COVs. They found that the IDEAL-CT COV ranged between 15 and 25% within and between labs and concluded that was an acceptable range of variability for a cracking test (VanFrank et al., 2020).

The National Center for Asphalt Technology (NCAT) compared results from six different IDEAL-CT machines (Moore et al., 2021). They stated that consistent specimen preparation is key to achieving low variability (Moore et al., 2021). The results of tests with different machines were compared using an equivalence limit of 20% of the average  $CT_{Index}$  (Moore et al., 2021).

In 2018, NCAT conducted a round-robin study on performance tests being considered for the BMD implementation. This study was broken into two phases, and fifteen different labs completed IDEAL-CT testing. The within-lab COV for phase one was 19.5%, and the betweenlab COV was 35.3%. For phase two of this project, the IDEAL-CT within-lab COV was 18.8%, and the between-lab COV was 20.2%. The difference between phase one and phase two was that all of the specimens were made in a single laboratory for phase two, while each laboratory made its own specimens in phase one. The difference in between-lab COV drops between the studies

18

highlights the importance of consistent sample preparation for  $CT$  index results (Taylor et al., 2022).

COVs of CTindex range from 15% to 35% for within and between-lab. Results were found to be sensitive to RAP content, asphalt content, asphalt binder type, and aging conditions, according to the study completed by the Texas A&M Transportation Institute study. Studies recommended testing an adequate number of replicate samples; the NCAT study did approximately 50 replicate samples per machine when performing statistical analysis on mixes, and that sample preparation is an essential step in reducing variability.

## 2.3.3 Hamburg Wheel Tracking Test

The Hamburg Wheel Tracking Test (HWTT) is a performance test used to evaluate asphalt mixtures rutting resistance and moisture susceptibility. According to AASHTO T324, a pair of laboratory-compacted specimens, 62mm in thickness and 150mm in diameter, is loaded using a reciprocating steel wheel. The test specimens are submerged in a temperature-controlled water bath, and the total deformation is measured and plotted as a function of the number of wheel passes.

For this study, the HWTT raw data were analyzed for rutting using two methods. The first method, referred to as the corrected rut depth method, isolates deformation due to rutting from deformation due to moisture damage, as illustrated in Figure 2.3.2 (Yin et al., 2014). This separation of the HWTT specimen damage according to the two distress mechanisms is necessary since the remedies for the two distresses are different. Thus, the corrected rut depth at 20,000 passes ( $CRD<sub>20k</sub>$ ) is a better indicator of mixture rutting resistance than the traditional HWTT rutting parameters of total rut depth or passes to 12.5mm rut depth (West et al., 2021).

19

However, to relate the HWTT results of this study to previous research, the total passes to 12.5 mm rut depth was also recorded.



Figure 2.3.2: HWTT Data Analysis, CRD<sub>20k</sub> (West et al., 2021)

## 2.3.4 Hamburg Wheel Tracking Test Variability

The Texas Transportation Institute studied the variability of seven HWTT devices, all manufactured by Precision Metal Works, in three laboratories in Texas. The two-way analysis of variance (ANOVA) showed that the variability within and between machines increased with the increase in load cycles (Chowdhury et al., 2004).

A round-robin study conducted by the University of California Pavement Research Center (UCPRC) involved twenty laboratories around California. Each lab conducted four HWTT tests. Two tests were conducted on specimens made by UCPRC and the other two were conducted on specimens compacted by each participating laboratory. The laboratories reported test results at rut depths after 5,000, 10,000, 15,000, and 20,000 passes, the number of passes to 12.5 mm rut depth, creep slope, stripping slope, and stripping inflection point. An outlier analysis was conducted if a lab's average differed considerably from the other labs. An ANOVA analysis was also conducted to determine variance components that influenced test results. The

study concluded that the type of HWTT device used was significant only for the rut depth after 5,000 and 10,000 passes. Single-operator variability was measured to be relatively low. Between-lab variability was relatively high for all results measured (Mateos, 2017).

In the 2018 NCAT round-robin study, the HWTT variability was analyzed since this is the most popular rutting performance test being considered for BMD implementation. Thirty-two labs participated in the first phase of the round-robin study; four different HWTT machines were used between the labs. At 10,000 passes, two of the thirty-two labs were shown as outliers; at 20,000 passes, four of the thirty-two labs were shown as outliers. The within and between-lab COV were reported for 10,000 and 20,000 passes. The within-lab COV for 10,000 passes was 9.0%, and for 20,000 passes, it was 9.4%. The between-lab COV for 10,000 passes was 21.1%, and for 20,000 passes, the COV was 25.9%. It is stated in the study that the COV results for within-lab repeatability are good, and the between-lab COV is reasonable (Taylor et al., 2022).

NCHRP project 20-07/Task 361, Hamburg Wheel-Track Test Equipment Requirements and Improvements to AASHTO T 324, evaluated the capabilities of available HWTT devices and identified issues with the AASHTO T 324 standard. The study concluded that there are differences in machines in the waveform, temperature range, and reporting parameters (Mohammad et al., 2015). Recommendations for fixing HWTT devices proposed addressing equipment capabilities, data collection, data analysis, and reporting to address the differences between machines (Mohammad et al., 2015).

A study completed by the AASHTO Materials Reference Laboratory studied the precision estimates for AASHTO T 324. The results proposed several changes to AASHTO T 324 to improve the repeatability and reproducibility of the HWTT machines. These changes included: starting location of the wheel, alignment of the wheel with respect to the specimen,

21

measurement locations used in the analysis, variability in the cutting of the gyratory specimens, potentially increasing the specimen length, designing a new mold in terms of material and reducing the joint space between the two specimens (Azari, 2014). Precision estimates were reported for the number of passes to a threshold rut depth for single-operator COV of 16.6% and for multi-laboratory COV of 24.2% (Azari, 2014).

In summary, HWTT variability increases with increasing cycles based on the Texas Transportation Institute study. AASHTO T 324 has several parameters, waveform, temperature range, and reporting parameters, that can be improved upon to improve the repeatability and reproducibility of HWTT results. The NCAT round-robin study reported within-lab COVs of rut depths at 10,000 passes to be 9.0%, and at 20,000 passes to be 9.4%.

#### Chapter 3 – Research Plan

## 3.1 Selection of Shadow Projects

For this study, ten shadow projects were chosen from various locations across Wisconsin to represent the state's diversity in aggregate type, binder grades, and mix types. Wisconsin contractors obtained the surface mixture samples for the research while they also sampled mix for QC testing. For WisDOT, random samples are taken every 750 tons, representing a sublot. A typical lot in Wisconsin consists of five sublots; this gave 10 to 15 mix samples per shadow project. Table 3.1 summarizes the shadow project county locations, the region in Wisconsin, route, mix design number, mix type, and contractor. Figure 3.1 shows a map of the ten shadow project locations. The mix designs for each project can be found in Appendix 1. All of the mixtures were designed using the Superpave method.

				WisDOT Mix		Contractor
Project	County	Region	Route	Design ID	Mix Type	
1	Ozaukee	Southeast	IH 43	250-0032-2021	4 MT 58-28 S	Payne $\&$ Dolan
$\overline{2}$	Florence	North Central	<b>STH 139</b>	250-0263-2021	4 LT 58-28 S	Payne $\&$ Dolan
3	Grant	Southwest	<b>STH 011</b>	601-21-4MTR301	4 MT 58-28 S	Mathy
$\overline{4}$	Kewaunee	Northeast	<b>STH 029</b>	250-0035-2022	4 MT 58-28 S	Northeast Asphalt
5	Waukesha	Southeast	<b>STH 067</b>	250-0051-2022	4 MT 58-28 S	Rock Road
6	Lacrosse	Southwest	<b>STH 016</b>	147-21-4MTR301	4 MT 58-28 S	Mathy
$\overline{7}$	Bayfield	Northwest	<b>USH 063</b>	158-22-5MTRW301	5 MT 58-34 V	Mathy
8	Iowa	South Central	<b>USH 018</b>	0-250-0025-2021	4 HT 58-28 S	Payne & Dolan
9	Barron	Northwest	<b>USH 008</b>	360-22-4MTRW301	4 MT 58-34 V	Mathy
10	Waushara	Central	IH 039	250-0107-2022	4 HT 58-28 S	American Asphalt

Table 3.1 Project Summary



Figure 3.1 Project Locations in Wisconsin

## 3.2 Testing Plan

Two performance tests were conducted for this study: IDEAL-CT and Hamburg Wheel Tracking Test (HWTT).

## 3.2.1 Mixture Processing

The asphalt mixtures used in this study were sampled during plant production while the contractor was sampling for regular QC/QA testing for WisDOT projects. For each project, sublots for two to three lots were sampled, resulting in approximately 15 samples per project. Two five-gallon buckets of asphalt mix for each sublot were obtained to ensure sufficient material for testing. The contractors also provided the results of their QC tests corresponding to each sample. The mixes were shipped from their respective Wisconsin contractor to NCAT for performance testing. Each bucket of loose asphalt mix was heated to compaction temperature and reduced to testing size per AASHTO R47-19 Standard Practice for Reducing Sample of Asphalt Mixtures to Testing Size. A Quartermaster quartering device, shown in Figure 3.2.1, was used to reduce the sample size while ensuring representative samples for consistent laboratory results. As shown in Figure 3.2.2, a quartering template was used to further reduce the sampled mix to size. This sample-reducing method produced four maximum specific gravity  $(G_{mm})$ samples, two bulk specific gravity  $(G<sub>mb</sub>)$  samples, and approximately fifteen test specimens per sublot.



Figures 3.2.1 and 3.2.2 : Quartering Devices (AASHTO R47-19)

Once the loose plant mix was reduced to the testing size, the samples were stored in sealed, labeled plastic bags to be compacted later. Each specimen was compacted to 62 mm in height and 150 mm in diameter using a gyratory compactor, following ASTM D6925-15. Each sample was made by the same engineer, scale, oven, and gyratory compactor to reduce specimen variability. The theoretical maximum specific gravity  $(G_{mm})$ , known as the Rice test, was determined for each mix. A trial specimen was made using the previously reduced samples to

determine the mass needed to achieve7.0±.05 air voids, 150 mm in diameter and 62 mm in thickness.

## 3.2.2 Summary of Testing Plan

Figure 3.2.3 shows a flow diagram of the testing procedure performed. Across the ten projects in this research study, a total of 134 sets of four samples were subjected to IDEAL-CT and HWTT testing. The maximum specific gravity  $(G_{mm})$  and bulk specific gravity  $(G_{mb})$  were verified to be consistent with the contractors' data using the multi-lab d2s limits in AASTO T 209 and T 166, respectively.



Figure 3.2.3 Testing Plan Flow Diagram

## 3.2.3 HWTT Testing Procedure

Mix was reheated to the compaction temperature to compact the HWTT specimens. Each specimen's air voids were checked using AASHTO T166, Standard Method of Test for Bulk Specific Gravity (G<sub>mb</sub>) of Compacted Asphalt Mixtures Using Saturated Surface-Dry Specimens. Each HWTT specimen was cut to fit into the HWTT mold. All HWTTs were conducted following AASTHO T 324 using the Troxler machine shown in Figure 3.2.4.



Figure 3.2.4 Troxler HWTT Machine at NCAT Laboratory

## 3.2.4 IDEAL-CT Procedure

For the IDEAL-CT test specimens, the loose plant mix samples were long-term aged for 6 hours at 275°F. This aging procedure was recommended to simulate in-service aging in a previous WHRP project using Wisconsin mixtures in 2018 (Bahia, 2018). This is similar to the "critical aging" procedure recommended by NCAT (Chen et al, 2018). A maximum specific gravity  $(G_{mm})$  test and a bulk specific gravity  $(G_{mb})$  test were performed on asphalt samples produced from the aged mixture. Once the quantity of loose mix needed to produce 150 mm diameter compacted samples to a height of 62 mm with 7.0% +/- 0.5% air voids, four specimens were compacted for IDEAL-CT testing. The IDEAL-CT test was conducted according to ASTM D8225 using the Troxler IDEAL Plus, shown in Figure 3.2.5.



Figure 3.2.5 Troxler IDEAL Plus machine at NCAT Laboratory

## 3.3 Method of Analysis

For this experiment, statistical analysis was performed on the data collected from the HWTT and IDEAL-CT tests. Statistical analysis was also performed on the percent binder  $(P_b)$  and air voids (Va) from the contractor's QC data. For the IDEAL-CT tests on each sublot, the average and standard deviation was calculated from four replicates. The HWTT data was used to determine the corrected rut depth (CRD) and the number of passes to reach a rut depth of 12.5 mm for the left and right wheels. The results for the left and right wheels were averaged to yield an average CRD, and an average passes to 12.5 mm rut depth for each sublot.

## 3.3.1 Calculation of Sample Mean, Standard Deviation, and Coefficient of Variation

The sample mean (Eq 3.1), standard deviation (Eq. 3.2), and coefficient of variation (Eq. 3.3) were calculated from the five sublots based on each lot. The calculations are as follows:

$$
\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}
$$
 Equation 3.1: Sample Mean  

$$
s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}
$$
 Equation 3.2: Standard Deviation  

$$
COV = \frac{\overline{x}}{s}
$$
 Equation 3.3: Coefficient of Variation

Where:

- $\overline{x}$  = Sample Mean
- $s =$ standard deviation

 $COV = coefficient of variation$ 

- $x_i$  = each value from the sample population
- $n =$  number of items in the sample

#### 3.3.2 Outlier Evaluation of a Lot

An outlier can be defined as "one that appears to deviate markedly from other members of the sample in which it occurs" (ASTM E178-21). This outlier procedure used in this study was developed by the Maine Department of Transportation and is an adaptation from ASTM E178 Dealing with Outlying Observations. This calculation procedure is based on a "two-tail t-test" with a level of significance ( $\alpha$ ) of 5%. The calculation steps are as follows:

- 1. Calculate the sample average  $(\bar{x})$  and standard deviation (s) of the results in the lot.
- 2. Find the critical t value " $t_{crit}$ " from Table 3.2 using the total number of samples (n) in the sample set.
- 3. Determine the total allowable deviation (D) on either side of the sample average by multiplying ' $t_{\text{crit}}$ ' by s.
- 4. Establish values for Max and Min by adding and subtracting D to and from  $\overline{x}$ .
- 5. Any results greater than the Max or less than the Min are determined to be an outlier.

n	$t_{crit}$
	1.155
	1.481
	1.715
	1.887
	2.020

**Table 3.2**  $t_{crit}$  values for a 5% Significance Level

This outlier evaluation procedure was performed on results from the IDEAL-CT, HWTT, the contractor's reported air voids, and asphalt content for each lot.

## 3.3.3 Normality Procedure

An Anderson-Darling (AD) test was performed using Minitab software to assess the normality of the results of quality characteristics from each project. This test compares the empirical cumulative distribution function of the data with the distribution expected if the data was normal ("Test for Normality'). The AD statistic measures how well the data follows the normal distribution; the better the distribution fits the data, the smaller the AD statistic. The null hypothesis for the Anderson-Darling test was that the data followed a normal distribution. If the p-value was less than 0.05, the null hypothesis was rejected and it was concluded that the data were found to be not normally distributed. The AD test was performed on the ten projects individual sublot data for IDEAL-CT, HWTT, air voids, and asphalt content.

## 3.3.4 CDF of Production Standard Deviation and Coefficients of Variation

The cumulative distribution frequency (CDF) for each project's lot standard deviations and coefficients of variation were plotted using Minitab software. Cumulative distribution frequencies are used to evaluate the distribution of a dataset. They can help analyze the percentage of the data that lie above or below a particular value, and the steepness or slope of the CDF can indicate how close the observations are to the mean (Cumulative, 2020).

## 3.3.5 Percent Within Limits Calculations

Percent within limits (PWL) calculations were conducted on each project's lot for IDEAL-CT and HWTT data based on current WisDOT specification criteria, as shown in Table 3.3. PWL is the percentage of a lot falling within a set specification limit based on simple statistics. Since the HWTT and IDEAL-CT criteria are minimum values (one-sided criteria), only the lower quality index was calculated following Equation 3.4.

$$
Ql_L = \frac{(\overline{x} - LL)}{s}
$$
 Eq. 3.4

Where:

 $QI_L$  = Lower Quality Index

 $\bar{x}$  = mean of test results

 $LL =$  Lower specification Limit (tolerance)

S = Standard Deviation

The "Q" table from AASHTO R 42 was used to find the PWL corresponding to each Ql.





2. For SMA, increase the minimum CT<sub>Index</sub> criterion to 80 for all binder designation levels.

## 3.4 Summary

Once the asphalt mixture samples were received at NCAT from the Wisconsin

contractors, they were reduced, compacted, and evaluated using various AASHTO procedures.

The two performance tests completed on each sublot sample were the IDEAL-CT and HWTT.

All specimen preparation and tests were conducted at NCAT by the same engineer using the same equipment to minimize variability. For each lot on each of the ten mixes, the average, standard deviation, and coefficient of variation were calculated for CT<sub>index</sub> and HWTT CRD and passes to 12.5 mm rut depth, and the results were analyzed for outliers, normality, cumulative distribution frequency, and percent within limits.

#### Chapter 4 – Results and Discussion

 This chapter summarizes the results from the performance testing conducted at NCAT on mixtures from the ten shadow projects sampled by WisDOT contractors. These mixes were tested using IDEAL-CT and HWTT performance tests, along with the asphalt content and air voids from the contractor's QC data to evaluate the production variability of the properties. 4.1 Summary of Averages, Standard Deviations, and Coefficients of Variation

 For mixes of each shadow project, the contractors provided samples for two or three lots, and for each lot, there were five sublots. Therefore, the tests' average, standard deviation, and COV s were calculated from the results of five sublots.

Table 4.1.1 summarizes the asphalt content of each project. The asphalt content has the lowest overall COVs among the evaluated quality characteristics, with an average COV of 2.8%; the maximum COV was observed to be 7.2%.

Table 4.1.2 summarizes the air voids for each project. The air voids had an average COV of 10.4%.

Table 4.1.3 summarizes the  $CT_{Index}$  for each project. The average COV for  $CT_{Index}$  was 13.1%, with the minimum COV being 1.3% and the maximum COV being 39.7%.

Table 4.1.4 summarizes the  $CRD_{20k}$  calculated for each project. The average COV for  $CRD<sub>20k</sub>$  was 10.9%, with a maximum COV of 26.4%, and a minimum COV of 4.1%. Table 4.1.5 summarizes the HWTT passes to reach a 12.5 mm rut depth. The average COV for passes to 12.5 mm was 16.6%, with a maximum of 35.8% and a minimum of 2.8%.

33

<b>Asphalt Content</b>					
Project	Lot	Average	Std. Dev.	<b>COV</b>	
	Lot 1	6.1	0.2	2.6%	
$\mathbf{1}$	Lot 2	6.3	0.2	3.8%	
$\overline{2}$	Lot 2	5.6	0.1	1.3%	
	Lot 3	5.7	0.1	2.5%	
	Lot 2	5.8	0.2	2.6%	
$\mathfrak{Z}$	Lot 3	6.0	0.4	7.2%	
	Lot 4	5.9	0.3	5.9%	
	Lot 2	5.9	0.1	1.7%	
$\overline{4}$	Lot 3	6.0	0.1	1.9%	
	Lot 4	6.0	0.1	1.4%	
	Lot 4	5.7	0.2	3.4%	
5	Lot 5	5.8	0.2	3.1%	
	Lot 6	5.8	0.1	1.9%	
6	Lot 9&11	6.0	0.1	2.2%	
	Lot 10	5.9	0.2	2.8%	
	Lot 3&6	6.6	0.1	1.3%	
$\tau$	Lot 4	6.7	0.1	1.8%	
	Lot 5	6.8	0.1	1.3%	
	Lot 3	5.8	0.2	4.0%	
8	Lot 4	5.8	0.1	1.9%	
	Lot 5	5.7	0.2	2.7%	
	Lot 8	5.6	0.1	2.1%	
9	Lot 9	5.6	0.1	2.3%	
	Lot 10	5.4	0.3	5.8%	
	Lot 8	6.2	0.2	2.7%	
10	Lot 9	6.2	0.1	2.1%	
	Lot 10	6.2	0.2	3.0%	

Table 4.1.1 Asphalt Content Summary

Air Voids					
Project	Lot	Average	Std. Dev.	<b>COV</b>	
1 - P&D Jackson	Lot 1	3.3	0.4	12.8%	
	Lot 2	3.1	0.3	10.2%	
2 - NEA Popple	Lot 2	2.8	0.1	4.1%	
River	Lot 3	2.9	0.2	6.7%	
	Lot 2	2.9	0.2	7.9%	
3 - Mathy Plant 1	Lot 3	2.9	0.1	4.5%	
	Lot 4	2.8	0.4	15.8%	
	Lot 2	2.9	0.3	10.3%	
4 - NEA Denmark	Lot 3	3.0	0.1	2.9%	
	Lot 4	3.2	0.1	1.7%	
	Lot 4	3.3	0.3	8.3%	
5 - Rock Road	Lot 5	3.2	0.2	6.8%	
	Lot 6	3.3	0.3	$9.0\%$	
6 - La Crosse	Lot 9&11	2.9	0.3	10.1%	
	Lot 10	2.4	0.4	15.5%	
	Lot $3&6$	3.1	0.5	16.5%	
7 - Drummond	Lot 4	2.8	0.5	16.8%	
	Lot 5	2.6	0.3	9.9%	
	Lot 3	3.0	0.4	14.9%	
8 - Dodgeville	Lot 4	2.7	0.6	21.3%	
	Lot 5	2.9	0.6	19.5%	
	Lot 8	3.1	0.2	6.2%	
9 - Turtle Lake	Lot 9	2.8	0.4	14.6%	
	Lot 10	3.0	0.3	10.8%	
	Lot 8	3.0	0.1	3.6%	
10 - Coloma	Lot 9	2.8	0.3	10.6%	
	Lot $10$	2.8	0.3	10.0%	

Table 4.1.2 Air Voids Summary

<b>IDEAL CT</b>					
Project	Lot	Average	Std. Dev	<b>COV</b>	
1 - P&D Jackson	Lot 1	47.0	7.4	15.6%	
	Lot 2	48.0	4.0	8.4%	
2 - NEA Popple	Lot 2	58.2	9.1	15.7%	
River	Lot 3	62.8	19.6	31.1%	
	Lot 2	62.7	6.4	10.2%	
3 - Mathy Plant 1	Lot 3	69.7	27.7	39.7%	
	Lot 4	73.3	17.8	24.3%	
	Lot 2	86.2	7.6	8.8%	
4 - NEA Denmark	Lot 3	83.8	10.7	12.8%	
	Lot 4	89.0	6.0	6.7%	
	Lot 4	40.1	4.3	10.7%	
5 - Rock Road	Lot 5	44.3	8.8	19.9%	
	Lot 6	51.3	5.2	10.1%	
6 - La Crosse	Lot 9&11	46.2	3.6	7.8%	
	Lot 10	51.2	7.7	15.1%	
	Lot 3&6	106.7	16.8	15.7%	
7 - Drummond	Lot 4	113.5	7.8	6.9%	
	Lot 5	120.4	8.9	7.4%	
	Lot 3	45.1	2.0	4.4%	
8 - Dodgeville	Lot 4	51.0	4.6	9.1%	
	Lot 5	43.4	0.6	1.3%	
	Lot 8	51.5	8.9	17.2%	
9 - Turtle Lake	Lot 9	58.9	5.2	8.8%	
	Lot 10	57.5	5.5	9.5%	
	Lot 8	113.2	11.6	10.3%	
10 - Coloma	Lot 9	118.4	14.5	12.2%	
	Lot $10$	119.5	16.4	13.7%	

Table 4.1.3 CT<sub>Index</sub> Summary
HWTT - Corrected Rut Depth 20,000 passes						
Project	Lot	Average	<b>COV</b>			
	Lot 1	10.7	2.2	20.4%		
1 - P&D Jackson	Lot 2	11.0	1.4	13.1%		
2 - NEA Popple	Lot 2	16.4	2.8	16.8%		
River	Lot 3	16.2	0.7	4.4%		
	Lot 2	9.0	0.4	4.1%		
3 - Mathy Plant 1	Lot 3	11.0	0.4	4.1%		
	Lot 4	10.6	1.2	11.7%		
	Lot 2	15.9	1.6	10.3%		
4 - NEA Denmark	Lot 3	16.2	1.3	8.0%		
	Lot 4	17.3	3.0	17.6%		
5 - Rock Road	Lot 4	10.5	1.0	9.9%		
	Lot 5	11.2	0.7	5.8%		
	Lot 6	10.5	0.7	7.0%		
6 - La Crosse	Lot 9&11	11.3	1.0	8.7%		
	Lot 10	11.6	1.6	13.5%		
	Lot 3&6	11.7	0.7	5.6%		
7 - Drummond	Lot 4	13.1	3.4	26.4%		
	Lot 5	16.4	3.3	20.1%		
	Lot 3	10.2	1.2	11.9%		
8 - Dodgeville	Lot 4	10.2	1.0	10.0%		
	Lot 5	8.4	1.2	14.3%		
9 - Turtle Lake	Lot 8	9.7	0.9	9.3%		
	Lot 9	11.0	1.1	9.6%		
	Lot 10	12.0	1.3	10.6%		
	Lot 8	11.6	0.6	4.9%		
10 - Coloma	Lot 9	13.3	1.4	10.6%		
	Lot $10$	12.6	0.6	5.1%		

Table 4.1.4 CRD<sub>20k</sub> Summary

HWTT - Passes to 12.5 mm Rut Depth						
Project	Lot	Average	Std. Dev	<b>COV</b>		
1 - P&D Jackson	Lot 1	11416	4085.8	35.8%		
	Lot 2	9662	3298.4	34.1%		
2 - NEA Popple	Lot 2	5670	952.13	16.8%		
River	Lot 3	4785	905.8	18.9%		
	Lot 2	14580	2752.5	18.9%		
3 - Mathy Plant 1	Lot 3	11188	909.21	8.1%		
	Lot 4	11642	2575.7	22.1%		
	Lot 2	5682	523.12	9.2%		
4 - NEA Denmark	Lot 3	6200	734.81	11.9%		
	Lot 4	4800	486.16	10.1%		
	Lot 4	13972	2768.7	19.8%		
5 - Rock Road	Lot 5	10662	2670.8	25.0%		
	Lot 6	11266	1864.5	16.5%		
6 - La Crosse	Lot 9&11	8460	608.6	7.2%		
	Lot 10	9018	1618.3	17.9%		
	Lot 3&6	7192	215.1	3.0%		
7 - Drummond	Lot 4	6592	1294.2	19.6%		
	Lot 5	4726	1197.2	25.3%		
8 - Dodgeville	Lot 3	9188	2272.4	24.7%		
	Lot 4	9056	1568.8	17.3%		
	Lot 5	11278	1973.5	17.5%		
9 - Turtle Lake	Lot 8	10870	1189.3	10.9%		
	Lot 9	9278	1828.5	19.7%		
	Lot $10$	9370	1463.9	15.6%		
	Lot 8	9990	1406.6	14.1%		
10 - Coloma	Lot 9	8302	235.4	2.8%		
	Lot 10	8840	553.2	6.3%		

Table 4.1.5 Passes to 12.5 mm Rut Depth Summary

## 4.2 Normality Test

For the Normality Test, the Anderson-Darlington (AD) statistic and the probability plots of the sublot averages of IDEAL-CT, HWTT, asphalt content, and air voids were calculated for each of the ten shadow projects. Figures 4.2.1 through 4.2.5 display the probability plot and the AD test

statistic for Project 1 as examples. Tables 4.2.1 through 4.2.5 display the summary results of the AD statistic and the corresponding p-values for all 10 projects.



Figure 4.2.1 Probability Plot of CT<sub>index</sub> Project 1





From Table 4.2.1, it can be seen that the p-value for the Anderson-Darling test of normality was greater than 0.05 except for Project 3, indicating that the  $CT_{index}$  results from most projects were normally distributed. For Project 3, the AD normality test may have been influenced by the high CT<sub>index</sub> results from a four sublots compared to the average for all 15 sublots. For this project, the average CTindex for all sublots was 68.6, but CT<sub>index</sub> results for sublot 3-4, 3-5, 4-2, and 4-3 were 106.2, 92.2, 86.4, and 95.9, respectively. This example brings to light a limitation of assessing normality with small data sets.



Figure 4.2.2 Probability Plot of HWTT CRD<sub>20k</sub> Project 1

<b>HWTT CRD20k</b>						
Project	N (number of sublots)	<b>AD</b> Statistic	p-value			
1 - P&D Jackson	10	0.474	0.186			
2 - NEA Popple River	9	0.282	0.548			
3 - Mathy Plant 1	15	0.482	0.196			
4 - NEA Denmark	15	0.642	0.076			
5 - Rock Road	15	0.389	0.339			
6 - La Crosse	10	0.21	0.807			
7 - Drummond	15	0.81	0.027			
8 - Dodgeville	15	0.305	0.526			
9 - Turtle Lake	15	0.357	0.407			
10 - Coloma	15	0.725	0.046			

Table 4.2.2 Normality Test Results for HWTT CRD<sub>20k</sub>

Table 4.2.2 shows that the p-value for the Anderson-Darling test of normality for HWTT CRD20k results was greater than 0.05 for all projects except for Project 7, indicating that the HWTT CRD<sub>20k</sub> results from most projects were normally distributed. For Project 7, the high CRD20k results for sublots 4-5 and 5-1 were 18.8 mm and 21.4 mm, respectively, compared to the average of 13.7 mm for all 15 sublots. This example again demonstrates a limitation of assessing normality with small data sets.

Similarly, Table 4.2.3 summarizes the Anderson-Darling results for HWTT passes to 12.5 mm rut depth. For this quality characteristic, the AD test p-value was less than 0.05 for Project 7 and 10 and was just above 0.5 for Project 3. The other seven projects had p-values well above 0.5 indicating that the HWTT results of the majority of projects were normally distributed.



Figure 4.2.3 Probability Plot of HWTT Passes to 12.5 mm Project 1

HWTT Passes to 12.5 mm					
Project	<b>AD</b> Statistic	p-value			
1 - P&D Jackson	10	0.316	0.482		
2 - NEA Popple River	9	0.338	0.412		
3 - Mathy Plant 1	15	0.691	0.056		
4 - NEA Denmark	15	0.145	0.958		
5 - Rock Road	15	0.329	0.476		
6 - La Crosse	10	0.253	0.653		
7 - Drummond	15	1.248	0.005		
8 - Dodgeville	15	0.400	0.319		
9 - Turtle Lake	15	0.263	0.647		
10 - Coloma	15	0.964	0.011		

Table 4.2.3 Normality Test Results for HWTT Passes to 12.5 mm



Figure 4.2.4 Probability Plot of Asphalt Content Project 1

<b>Asphalt Content</b>					
Project	<b>AD</b> Statistic	P-Value			
1 - P&D Jackson	10	0.231	0.732		
2 - NEA Popple River	9	0.393	0.298		
3 - Mathy Plant 1	15	0.46	0.223		
4 - NEA Denmark	15	0.759	0.037		
5 - Rock Road	15	0.268	0.631		
6 - La Crosse	10	0.405	0.285		
7 - Drummond	15	0.974	0.010		
8 - Dodgeville	15	0.419	0.285		
9 - Turtle Lake	15	1.108	0.005		
10 - Coloma	15	0.508	0.167		

Table 4.2.4 Normality Test Results for Asphalt Content

Table 4.2.4 shows that the p-values for the Anderson-Darling test of normality of asphalt content results was greater than 0.05 for seven of the ten shadow projects, indicating that asphalt content results for most projects were normally distributed. Project 4, 7 and 9 had p-values less

than 0.05, indicating that their asphalt content results were not normally distributed. For Project 4, asphalt contents were very consistent with all 15 sublots having asphalt contents between 5.8% and 6.1%. Likewise, Project 7 had very consistent asphalt contents ranging from 6.5% to 6.8%, with six of the 15 sublot results at 6.8%. Project 9's asphalt contents ranged from 5.1% to 5.7% but five of the 15 results were 5.7% which does not seem to follow a normal distribution, but the analysis is limited by the small data set.



Figure 4.2.5 Probability Plot of Air Voids Average

Table 4.2.5 shows that the p-values for the Anderson-Darling test of normality of air void content results was greater than 0.05 for nine of the ten shadow projects, indicating that air voids were normally distributed for most projects. Only Project 4 had a p-values less than 0.05, indicating that their asphalt content results were not normally distributed. For this project, air

voids were very consistent, ranging from 2.7% to 3.3%, with a mean of 3.0%. However, five of the 15 sublots had air voids of 3.1% and four sublots had air voids of 3.2% which does not seem to follow a normal distribution, but the analysis is limited by the small data set.

Air Voids						
Project	N (number of sublots)	<b>AD</b> Statistic	p-value			
1 - P&D Jackson	10	0.645	0.065			
2 - NEA Popple River	9	0.367	0.348			
3 - Mathy Plant 1	15	0.458	0.227			
4 - NEA Denmark	15	0.873	0.019			
5 - Rock Road	15	0.269	0.628			
6 - La Crosse	10	0.139	0.960			
7 - Drummond	15	0.622	0.085			
8 - Dodgeville	15	0.285	0.576			
9 - Turtle Lake	15	0.327	0.455			
10 - Coloma	15	0.576	0.112			

Table 4.2.5 Normality Test Results for Air Voids

## 4.3 Evaluation of Outliers

 The research project evaluated 27 lots and 134 sublots across the ten shadow projects. Within each sublot, replicates of CT<sub>index</sub> were evaluated for outliers. Four replicates of IDEAL-CT specimens were individually tested; these four replicates were averaged to obtain the sublot value. The IDEAL-CT data was assessed for outliers within the lot based on the average CT<sub>index</sub> for each sublot. The HWTT data had a left and right wheel that were averaged to give one value for each sublot.

 For air voids, there was only one observed outlier within project one. Table 4.3.1 shows an example calculation for project one, lot two. The green shaded cells mean it is within the outlier range, while the red cell marks the outlier.

Table 4.3.1 Outlier in Air Voids



 Overall, one outlier was observed across the projects for all the IDEAL-CT lots tested in project 8, lot 5.

4.4 Examining Potential Relationships between Variability of Asphalt Content and Air Voids with IDEAL-CT and HWTT

 To determine if asphalt content and air voids variability influenced the variability of IDEAL-CT and HWTT results, their respective calculated COVs of each lot were plotted against each other in scatterplots. Best-fit linear regression equations were determined for these correlation plots using Excel. The scatterplots of  $CT_{index}$  COV versus asphalt content and air voids COVs can be seen in Figure 4.3.1 and Figure 4.3.2, respectively. The scatterplots of HWTT CRD<sub>20k</sub> COV versus asphalt content and air voids COVs can be seen in Figure 4.3.3 and Figure 4.3.4, respectively. The coefficient of determination  $(R^2)$  indicates how well the regression equation explains the relationship between the two variables.  $R^2$  can be interpreted as the percentage of the change in the dependent variable,  $CT_{index}$  COV, in this case, which can be attributed to the independent variable, asphalt content COV in this case. In general, the low  $R^2$ 

indicates that the variabilities of asphalt content and air voids had little to no influence on the variabilities of CTindex and HWTT CRD<sub>20k</sub>.



Figure 4.4.1 IDEAL-CT vs. AC COV



Figure 4.4.2 IDEAL-CT vs. Va COV



Figure 4.4.3 HWTT CRD20k vs. AC COV



Figure 4.4.4 HWTT CRD 20k vs. Air Voids COV

## 4.5 Percent Within Limits based on Current WisDOT Specification

Percent within limits (PWL) were calculated for each lot and project based on current WisDOT specifications. For the IDEAL-CT test, the minimum specification limit for CT<sub>index</sub> was set at 30. Table 4.5.1 summarizes the PWL calculations, where it can be seen that the only lots with PWLs below 100% were projects 3, lot 3 and project 5, lot 5.



Table 4.5.1 IDEAL-CT PWL

For HWTT, the number of passes to a rut depth of 12.5 mm for both the left and right wheels were recorded and averaged. The WisDOT specification required a minimum of 10,000 passes to reach 12.5 mm rutting for low and medium traffic projects and a minimum of 20,000 passes to reach 12.5 mm for high traffic projects. Table 4.5.2 summarizes the HWTT PWL calculations. In this table, the projects are colored according to traffic category (yellow for low traffic, blue for medium traffic, and green for high traffic). Several projects had failing HWTT

results resulting in negative Quality Index (QI) values, and PWL of less than 50% (highlighted in red).

It is important to note that the HWTT machine used for this project differs from the machine used in previous research for Wisconsin. There are a couple of reasons that the HWTT results for the mixtures evaluated in this project are worse than from previous studies. One possible explanation is that a calibration check of the machine used for this project near the end of testing for this study revealed that the wheel paths of this machine did not follow a sinusoidal form as required in the AASHTO standard. Another difference is that the HWTT analysis software programs used with the HWTT machines use different seating passes before the initial rut depth is established. Currently, the AASHTO standard does not address seating passes. It should also be noted that the current WisDOT specification was not the criteria that the previous research completed at NCAT recommended. Even though the mixes here did not meet the WisDOT specifications, this study was conducted on variability.

HWTT - Corrected Rut Depth 20,000 passes							
Project		Average	Std. Dev	<b>COV</b>	$\mathbf n$	Ql	<b>PWL</b>
1 - P&D Jackson	Lot 1	11416.0	4085.8	35.8%	5	0.35	62
	Lot 2	9662.0	3298.4	34.1%	5	$-0.10$	50
2 - NEA Popple River	Lot 2	5670.0	952.1	16.8%	5	$-4.55$	50
	Lot 3	4785.0	905.8	18.9%	$\overline{4}$	$-5.76$	50
	Lot 2	14580.0	2752.5	18.9%	5	1.66	99
3 - Mathy Plant 1	Lot 3	11188.0	909.2	8.1%	5	1.31	92
	Lot $4$	11642.0	2575.7	22.1%	5	0.64	72
	Lot 2	5682.0	523.1	9.2%	5	$-8.25$	50
4 - NEA Denmark	Lot 3	6200.0	734.8	11.9%	5	$-5.17$	50
	Lot 4	4800.0	486.2	10.1%	5	$-10.70$	50
	Lot 4	13972.0	2768.7	19.8%	5	1.43	95
5 - Rock Road	Lot 5	10662.0	2670.8	25.0%	5	0.25	59
	Lot 6	11266.0	1864.5	16.5%	5	0.68	74
6 - La Crosse	Lot 9&11	8460.0	608.6	7.2%	5	$-2.53$	50
	Lot 10	9018.0	1618.3	17.9%	5	$-0.61$	50
	Lot 3&6	7192.0	215.1	3.0%	5	$-13.05$	50
7 - Drummond	Lot 4	6592.0	1294.2	19.6%	5	$-2.63$	50
	Lot 5	4726.0	1197.2	25.3%	5	$-4.41$	50
	Lot 3	9188.0	2272.4	24.7%	5	$-4.76$	50
8 - Dodgeville	Lot 4	9056.0	1568.8	17.3%	5	$-6.98$	50
	Lot 5	11278.0	1973.5	17.5%	5	$-4.42$	50
9 - Turtle Lake	Lot 8	10870.0	1189.3	10.9%	5	0.73	75
	Lot 9	9278.0	1828.5	19.7%	5	$-0.39$	50
	Lot 10	9370.0	1463.9	15.6%	5	$-0.43$	50
	Lot 8	9990.0	1406.6	14.1%	5	$-7.12$	50
10 - Coloma	Lot 9	8302.0	235.4	2.8%	5	$-49.69$	50
	Lot $10$	8840.0	553.2	6.3%	5	$-20.17$	50

Table 4.5.2 HWTT PWL

# 4.6 Cumulative Distribution Frequency of Production Standard Deviation and COV

 Cumulative distribution frequencies were plotted for each lot standard deviation and COV for the  $CT_{Index}$ ,  $CRD_{20k}$ , air voids, and asphalt content.

Figure 4.6.1 and Figure 4.6.2 show the CT<sub>Index</sub> standard deviation and COV, respectively, for each lot. The 50<sup>th</sup> percentile for the CT<sub>index</sub> standard deviation is 7.5 The 50<sup>th</sup> percentile for CTindex COV was 13.2%, with approximately 80% of the lots tested having a COV under 20%.



Figure 4.6.1 CDF of Std. Dev. CT<sub>Index</sub>



Figure 4.6.2 CDF of COV for CT<sub>Index</sub>

Figures 4.6.3 and 4.6.4 show the CDF plots of standard deviation and COV for HWTT  $CRD<sub>20k</sub>$ , respectively. The standard deviation had a 50<sup>th</sup> percentile of 1.3 The COV had a 50<sup>th</sup> percentile of 10.9%. Figures 4.6.5 and 4.6.4 show the CDF plots of standard deviation and COV for HWTT passes to 12.5 mm, respectively. The standard deviation had a  $50<sup>th</sup>$  percentile of 1554 passes. The COV has a 50<sup>th</sup> percentile of 16.6%.



Figure 4.6.3 CDF of Std. Dev. HWTT CRD<sub>20k</sub>



Figure 4.6.4 CDF of COV for HWTT CRD<sub>20k</sub>



Figure 4.6.5 CDF of Std. Dev. for HWTT passes to 12.5 mm



Figure 4.6.6 CDF of COV for HWTT passes to 12.5 mm

Figures 4.6.7 and 4.6.8 display the CDFs of the standard deviation and COV of asphalt content, respectively. Figures 4.6.9 and 4.6.10 show the CDF of the standard deviation and COV of air voids, respectively.



Figure 4.6.7 CDF of Std. Dev. for Asphalt Content



Figure 4.6.8 CDF of COV for Asphalt Content



Figure 4.6.9 CDF of Std. Dev. for Air Voids



Figure 4.6.10 CDF of COV for Air Voids

A combined plot of CDFs for standard deviations of CT<sub>index</sub>, HWTT CRD<sub>20k</sub>, asphalt content, and air voids is shown in Figure 4.6.11. A combined plot of CDFs of COVs of CT<sub>index</sub>, HWTT CRD<sub>20k</sub>, HWTT passes to 12.5 mm, asphalt content, and air voids are shown in Figure 4.6.12. It can be seen in Figure 4.6.12 that the asphalt content COV was the lowest, air voids and HWTT CRD<sub>20k</sub> has very similar COV distributions, and the HWTT passes to 12.5 mm COV was the highest.



Figure 4.6.11 CDF of Std. Dev. for CT<sub>Index</sub>, HWTT CRD<sub>20k</sub>, asphalt content, air voids



Figure 4.6.12 CDF of COV for CT<sub>Index</sub>, HWTT CRD<sub>20k</sub>, HWTT passes to 12.5 mm, asphalt

content, air voids

#### Chapter 5: Conclusions and Recommendations

## 5.1 Summary

The main objective of this thesis was to statistically analyze the overall variabilities of the performance test results for ten shadow projects. A recommended appropriate standard deviation for the BMD performance test is also provided based on the Wisconsin fieldproduced mixes. In total, 134 mixture samples were obtained and sent to NCAT for performance testing. The statistical analysis included a summary of the averages, standard deviations, and coefficients of variation. The Anderson-Darlington statistic and probability plots of the ten shadow projects sublots were calculated to check for normality. Results were analyzed to determine outliers on the IDEAL-CT,  $HWTT$   $CRD<sub>20k</sub>$ , asphalt content, and air voids. The percent within limits was calculated based on current WisDOT specifications on the IDEAL-CT and HWTT passes to 12.5 mm. Finally, CDF plots were made on production standard deviations and COVs.

## 5.2 Conclusions

After statistical analysis, the following conclusions are made:

• AC content was the least variable quality characteristic, with a mean COV of 2.8%. HWTT passes to 12.5 mm rut depth was the most variable quality characteristic with a mean COV of 16.6%. IDEAL-CT had a mean COV of 13.1%. The mean COV for air voids was 10.4%. The mean COV for HWTT CRD20k was 10.9%. From the literature reviewed the COV results achieved at NCAT were lower. These parameters were measured across ten shadow project's production variability, one reason these might be lower is due to the extreme carefulness that was taken during sample preparation.

60

Another reason is that there was a small sample size going into each shadow project's lot (five sublots).

- Normality testing showed that HWTT passes to 12.5 mm followed the most normal distribution, and the least normal distribution was the CTindex.
- The PWL analysis of  $CT_{index}$  indicated that 25 of the 27 lots had 100% PWL over the ten projects. HWTT passes to 12.5 mm rut depth did not meet the PWL, partly due to the HWTT machine not having a sinusoidal wavelength.

## 5.3 Plans for Future Research

 Studies should follow up with these ten projects over the years to document the field cracking and rutting to determine correlations between lab and field performance.

#### References

- AASHTO PP 105-20. Standard Practice for Balanced Design of Asphalt Mixtures. American Association of State Highway and Transportation Officials, Washington, D.C., 2020.
- AASHTO R 9-05. Acceptance Sampling Plans for Highway Construction. American Association of State Highway and Transportation Officials, Washington, D.C., 2018.
- AASHTO R 30-22. Standard Practice for Mixture Conditioning of Asphalt Mixture. American Association of State Highway and Transportation Officials, Washington, D.C., 2022.
- AASHTO R 42-06. Standard Practice for Developing a Quality Assurance Plan for Hot Mix Asphalt (HMA). American Association of State Highway and Transportation Officials, Washington, D.C., 2016.
- AASHTO R 47-19. Standard Practice for Reducing Sample of Asphalt Mixture to Testing Size. American Association of State Highway and Transportation Officials, Washington, D.C., 2019.
- AASHTO T 166-22. Standard Method of Test for Bulk Specific Gravity (Gmb) of Compacted Asphalt Mixtures Using Saturated Surface-Dry Specimens. American Association of State Highway and Transportation Officials, Washington, D.C., 2022.
- AASHTO T 209-22. Theoretical Maximum Specific Gravity ( $G_{mm}$ ) and Density of Asphalt Mixtures. American Association of State Highway and Transportation Officials, Washington, D.C., 2022.
- AASHTO T 324-22. Standard Method of Test for Hamburg Wheel-Track Testing of Compacted Asphalt Mixtures. American Association of State Highway and Transportation Officials, Washington, D.C., 2022.
- American Society of Testing and Materials (2019). ASTM D8225-19: Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature. ASTM International.
- American Society of Testing and Materials. ASTM E 178-21: Standard Practice for dealing with Outlying Observations. ASTM International. 2021.
- Azari, H. 2014. Precision Estimates of AASHTO T 324, Hamburg Wheel-Track Testing of Compacted Hot Mix Asphalt (HMA). Transportation Research Board. Washington, D.C.

Bahia et al. Field Aging and Oil Modification Study. WHRP Project 0092-17-04, 2018.

- Chen, C., F. Yin, P. Turner, R. West, and N. Tran, 2018, Selecting a Laboratory Loose Mix Aging Protocol for the NCAT Top-Down Cracking Experiment, Transportation Research Record I-13, Transportation Research Board.
- Chowdhury, A., Button, Joe., Wikander, J. 2004. Variability of Hamburg Wheel Tracking Devices. Texas Transportation Institute.
- Christensen, D. W., Bonaquist, R.F. 2006. Volumetric Requirements for Superpave Mix Design. NCHRP Report 567. Washington, D.C.
- "Cumulative Frequency." Calcworkshop, 20 Sept. 2020, https://calcworkshop.com/exploringdata/cumulative-

frequency/#:~:text=A%20cumulative%20frequency%20graph%20shows,the%20running %2Dtotal%20of%20frequencies.

Faheem, Ahmed, and Arash Hosseini. Evaluation of WisDOT Quality Management Program (QMP) Activities and ... Oct. 2018, https://wisconsindot.gov/documents2/research/0092- 15-05-final-report.pdf.

Halstead, W.J. 1979. NCHRP Synthesis of Highway Practice 65: Quality Assurance.

Transportation Research Board, National Research Council. Washington, D.C., 42 pp.

- Hughes, Charles. 1996. Synthesis of Highway Practice 232: Variability in Highway Pavement Construction. Transportation Research Board. Washington, D.C.
- Hughes, Charles. 2005. NCHRP Synthesis 346: State Construction Quality Assurance Programs. Transportation Research Board. Washington, D.C.
- Mateos, A., Jones, D. 2017. Support for Superpave Implementation: Round Robin Hamburg Wheel-Track Testing. University of California Pavement Research Center. UC Davis.
- Mohammad, L., Elseifi, M., Raghavendra, A., Ye, M. 2015. NCHRP Web-Only Document 219: Hamburg Wheel-Track Test Equipment Requirements and Improvements to AASHTO T 324. Transportation Research Board. Washington, D.C.
- Montgomery, D.C. (2009), Introduction to Statistical Quality Control (Chapter 1), Wiley, NY.
- Moore, N., Steger, R., Bowers, B., Taylor, A. 2021. Investigation of IDEAL-CT Device Equivalence: Are All Devices Equal? Asphalt Paving Technology.
- Solainmanian, M., Kennedy, T. 1995. Production Variability Analysis of Hot-Mixed Asphalt Concrete Containing Reclaimed Asphalt Pavement. Texas Department of Transportation.
- Taylor, A., Moore, J., Moore, N. 2022. NCAT Performance Testing Round Robin. NCAT Report 22-01. Auburn, AL.
- "Test for Normality." Minitab, https://support.minitab.com/en-us/minitab/20/help-and-howto/statistics/basic-statistics/supporting-topics/normality/test-for-normality/.
- VanFrank, K., Romero, P. 2020. Balanced Asphalt Concrete Mix Performance in Utah, Phase IV: Cracking Indices for Asphalt Mixtures. Utah Department of Transportation Research & Innovation Division.
- West, R., Hand, A., Weiss, J., Musselman, J., Moore, N. 2023. NCHRP Project 10-116: Quantifying Variability in Quality Characteristics of Pavements. Transportation Research Board. Washington, D.C.
- West, R., Yin, F., Rodezno, C., Leiva, F. 2021. Balanced Mixture Design Pilot and Field Sections. Wisconsin Highway Research Program.
- West, R., Yin, F., Rodezno, C., Taylor, A. 2021. Balanced Mixture Design Implementation Support. Wisconsin Highway Research Program.
- Wisconsin DOT. 2021. WI State-of-the-Practice Approach A. National Asphalt Pavement Association.
- Yin, F., E. Arambula, R. Lytton, A. E. Martin, and L. G. Cucalon. 2014. Novel Method for Moisture Susceptibility and Rutting Evaluation Using Hamburg Wheel Tracking Test. Transportation Research Board.
- Zhou, F. 2019. Development of an IDEAL Cracking Test for Asphalt Mix Design, Quality Control, and Quality Assurance. NCHRP IDEA Project 195, Transportation Research Board.

# Appendix

## Mix Design Project 1









## MATHY CONSTRUCTION CO. **GENERAL CONTRACTORS**

POST OFFICE BOX 189 ONALASKA, WI 54650

PHONE 608-781-4683 FAX 608-781-4694

**Report of Bituminous Mix Design** 







#### **Aggregate Gradations**



Test Methods: D312, T176/D2419, T11/C117, T27/C136, D4791, D5821, T304/C1252, T96/C131, T209/D2041, T166/D2726



# MATHY CONSTRUCTION CO.

GENERAL CONTRACTORS

820 10<sup>TH</sup> AVE N POST OFFICE BOX 189 ONALASKA, WI 50056

PHONE #N6-781-4883 FAX 888-781-4694

Report of Bituminous Mix Design





**Mix Properties** 







 $\overline{\mathbf{z}}$ 

#### **March Photograph**





dinos this design is staterist upsoffs, the constantons and recommendations contained within any

content and they make the submitted to end outpleted to the resultant states to construct the contents and the<br>Adjustments may become nanouslay when field informing size, which is considered from plant produced with<br>the gu

som g $\angle$  5. g Cert. No. 100163 Date: 8/27/2021



70

# Mix Design Project 4

## Reviewed By: Jeffery. R. Anderson



# Mix Design Project 5








Mathy Construction Co. -- MTE



# Mix Design Project 7

### Reviewed By: Jeffery, R. Anderson

### WHOLE LAW DECICAL CEANINARY DATA INDUT FORM (DEDORT 340



# Mix Design Project 8

Reviewed By: Jeffery. R. Anderson

#### WISDOT MIX DESIGN STANDARD DATA INPUT FORM/REPORT 249

C-5 4HT 5023211





# MATHY CONSTRUCTION CO.

POST OFFICE BOX 189 ONALASKA, WI 54650

PHONE 608-781-4683 FAX 608-781-4694

**Report of Bituminous Mix Design** 



920 10<sup>TH</sup> AVE N



### **Aggregate Sources**



### **Aggregate Gradations**



Test Methods: 0312, T176/D2419, T11/C117, T27/C136, D4791, D5821, T304/C1252, T98/C131, T209/D2041, T166/C2726





Turtle Lake - Cameron USH 8

# Mix Design Project 10

Review ed By: Jeffery, R. Anderson

## WHOOT MIX DESIGN STANDARD DATA INDIT FORM (REPORT 249

114522-4HT

