

Effects of Reheating and Mixture Composition on Balanced Mix Design Test Results

<https://vtrc.virginia.gov/media/vtrc/vtrc-pdf/vtrc-pdf/26-R42.pdf>

KAZUO KUCHIISHI, Ph.D.
Research Scientist

ILKER BOZ, Ph.D., P.E.
Senior Research Scientist

Final Report VTRC 26-R42

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 26-R42		2. Government Accession No.:		3. Recipient's Catalog No.:	
4. Title and Subtitle: Effects of Reheating and Mixture Composition on Balanced Mix Design Test Results				5. Report Date: May 2026	
				6. Performing Organization Code:	
7. Author(s): Kazuo Kuchiishi, Ph.D., Ilker Boz, Ph.D., P.E.				8. Performing Organization Report No.: VTRC 26-R42	
9. Performing Organization and Address: Virginia Transportation Research Council 530 Edgemont Road Charlottesville, VA 22903				10. Work Unit No. (TRAIS):	
				11. Contract or Grant No.: 126581	
12. Sponsoring Agencies' Name and Address: Virginia Department of Transportation Federal Highway Administration 1221 E. Broad Street 400 North 8th Street, Room 750 Richmond, VA 23219 Richmond, VA 23219-4825				13. Type of Report and Period Covered: Final	
				14. Sponsoring Agency Code:	
15. Supplementary Notes: This is an SPR-B report.					
16. Abstract: <p>Since 2019, the Virginia Department of Transportation (VDOT) has conducted several field trials with balanced mix design (BMD) mixtures. In 2024, these efforts culminated with the statewide application of VDOT's BMD specification to all 9.5- and 12.5-mm surface mixtures with A and D designations in maintenance resurfacing contracts. Several Virginia Transportation Research Council (VTRC) research projects helped address challenges during the specification development and implementation. These projects revealed significant and practical differences in BMD results between lots or samples of the same mixture and across producers and VDOT districts, highlighting the influence of mixture handling practices and component materials on BMD laboratory performance. A better understanding of how factors such as reheating and mixture composition affect BMD test results is essential to ensure consistent asphalt mixture production and robust quality measurement practices.</p> <p>This study's objective was to quantify the effect of reheating and mixture composition on VDOT's BMD test results and to identify the volumetrics and gradation parameters that can significantly influence Cantabro, indirect tensile cracking test (IDT-CT), Asphalt Pavement Analyzer (APA), and indirect tensile at high temperature (IDT-HT) results. The scope of work included a literature review, statistical analysis of BMD datasets from past VTRC studies and VDOT's 2024 BMD statewide production, a statewide stakeholder survey to gather insights regarding typical production-level practices for BMD testing, and BMD testing of four plant-produced asphalt mixtures to verify the findings.</p> <p>The ratio in cracking tolerance index (CT_{Index}) performance criteria between 100 (reheated) and 130 (non-reheated) was validated. Significant differences in Cantabro mass loss and IDT-HT strength between reheated and non-reheated conditions were observed, and preliminary non-reheated criteria were proposed. Statistical analysis revealed that asphalt content, gradation parameters (percent passing sieves No. 8, No. 30, and No. 200), aggregate bulk specific gravity, voids in total mixture, and absorbed asphalt content are significant factors influencing BMD results, with subsets affecting individual tests. A CT_{Index} predictive equation framework was developed, and the case study showed its potential as a tool to support decisions in requesting or waiving additional testing during production, especially under reduced testing frequency. Laboratory testing of four plant-produced mixtures showed no significant differences in BMD tests between lag times of 1 and 4 days, which the survey revealed to be typical lag times in Virginia. The laboratory testing also confirmed the adequacy of the proposed non-reheated performance criteria for Cantabro and IDT-HT and validated the significant volumetrics and gradation parameters, given the agreement between measured and predicted values.</p> <p>This study recommends that VDOT should (1) continue specifying CT_{Index} criteria of 100 and 130 for reheated and non-reheated mixtures during production, respectively; (2) benchmark Cantabro, IDT-CT, and IDT-HT production results during future construction seasons to evaluate the feasibility of the proposed non-reheated performance criteria for Cantabro and IDT-HT tests; (3) consider incorporating the proposed CT_{Index} predictive equation framework into a guidance document to support the current BMD specification when reduced testing frequency is applied during production, with refinement as additional data become available; and (4) continue monitoring and evaluating the effect of mixture composition on BMD test results.</p>					
17. Key Words: Asphalt mixture, performance criteria, Cantabro, IDT-CT, APA, IDT-HT, predictive equation, production			18. Distribution Statement: No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.		
19. Security Classif. (of this report): Unclassified		20. Security Classif. (of this page): Unclassified		21. No. of Pages: 68	22. Price:

FINAL REPORT
**EFFECTS OF REHEATING AND MIXTURE COMPOSITION ON BALANCED MIX
DESIGN TEST RESULTS**

Kazuo Kuchiishi, Ph.D.
Research Scientist
Virginia Transportation Research Council

Ilker Boz, Ph.D., P.E.
Senior Research Scientist
Virginia Transportation Research Council

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

Charlottesville, Virginia

May 2026
VTRC 26-R42

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

Copyright 2026 by the Commonwealth of Virginia.
All rights reserved.

ABSTRACT

Since 2019, the Virginia Department of Transportation (VDOT) has conducted several field trials with balanced mix design (BMD) mixtures. In 2024, these efforts culminated with the statewide application of VDOT's BMD specification to all 9.5- and 12.5-mm surface mixtures with A and D designations in maintenance resurfacing contracts. Several Virginia Transportation Research Council (VTRC) research projects helped address challenges during the specification development and implementation. These projects revealed significant and practical differences in BMD results between lots or samples of the same mixture and across producers and VDOT districts, highlighting the influence of mixture handling practices and component materials on BMD laboratory performance. A better understanding of how factors such as reheating and mixture composition affect BMD test results is essential to ensure consistent asphalt mixture production and robust quality measurement practices.

This study's objective was to quantify the effect of reheating and mixture composition on VDOT's BMD test results and to identify the volumetrics and gradation parameters that can significantly influence Cantabro, indirect tensile cracking test (IDT-CT), Asphalt Pavement Analyzer (APA), and indirect tensile at high temperature (IDT-HT) results. The scope of work included a literature review, statistical analysis of BMD datasets from past VTRC studies and VDOT's 2024 BMD statewide production, a statewide stakeholder survey to gather insights regarding typical production-level practices for BMD testing, and BMD testing of four plant-produced asphalt mixtures to verify the findings.

The ratio in cracking tolerance index (CT_{Index}) performance criteria between 100 (reheated) and 130 (non-reheated) was validated. Significant differences in Cantabro mass loss and IDT-HT strength between reheated and non-reheated conditions were observed, and preliminary non-reheated criteria were proposed. Statistical analysis revealed that asphalt content, gradation parameters (percent passing sieves No. 8, No. 30, and No. 200), aggregate bulk specific gravity, voids in total mixture, and absorbed asphalt content are significant factors influencing BMD results, with subsets affecting individual tests. A CT_{Index} predictive equation framework was developed, and the case study showed its potential as a tool to support decisions in requesting or waiving additional testing during production, especially under reduced testing frequency. Laboratory testing of four plant-produced mixtures showed no significant differences in BMD tests between lag times of 1 and 4 days, which the survey revealed to be typical lag times in Virginia. The laboratory testing also confirmed the adequacy of the proposed non-reheated performance criteria for Cantabro and IDT-HT and validated the significant volumetrics and gradation parameters, given the agreement between measured and predicted values.

This study recommends that VDOT should (1) continue specifying CT_{Index} criteria of 100 and 130 for reheated and non-reheated mixtures during production, respectively; (2) benchmark Cantabro, IDT-CT, and IDT-HT production results during future construction seasons to evaluate the feasibility of the proposed non-reheated performance criteria for Cantabro and IDT-HT tests; (3) consider incorporating the proposed CT_{Index} predictive equation framework into a guidance document to support the current BMD specification when reduced testing frequency is applied during production, with refinement as additional data become available; and (4) continue monitoring and evaluating the effect of mixture composition on BMD test results.

TABLE OF CONTENTS

INTRODUCTION	1
PURPOSE AND SCOPE.....	2
METHODOLOGY	3
Literature Review.....	3
Reheating and Mixture Composition Analyses	3
Laboratory Evaluation of Storage and Mixture Conditioning Protocols.....	7
RESULTS AND DISCUSSION.....	12
Literature Review.....	12
Effect of Reheating on Balanced Mix Design Test Results.....	15
Effect of Asphalt Mixture Composition on Balanced Mix Design Test Results.....	25
Laboratory Evaluation of Storage and Mixture Conditioning Protocols.....	47
Summary of Findings.....	54
CONCLUSIONS.....	56
RECOMMENDATIONS.....	57
IMPLEMENTATION AND BENEFITS	57
Implementation	57
Benefits	58
ACKNOWLEDGMENTS	58
REFERENCES	59

FINAL REPORT

EFFECTS OF REHEATING AND MIXTURE COMPOSITION ON BALANCED MIX DESIGN TEST RESULTS

Kazuo Kuchiishi, Ph.D.
Research Scientist
Virginia Transportation Research Council

Ilker Boz, Ph.D., P.E.
Senior Research Scientist
Virginia Transportation Research Council

INTRODUCTION

In 2017, the Virginia Department of Transportation (VDOT) initiated the first steps toward implementing the balanced mix design (BMD) concept for its dense-graded asphalt surface mixtures with A and D designations. Several field trials using BMD mixtures were constructed during the 2019 and 2020 paving seasons. Throughout 2021 and 2022, approximately 300,000 tons of BMD mixtures were placed across 23 maintenance schedule contracts in most districts (Diefenderfer et al., 2023a). In 2024, VDOT's BMD specification was applied statewide to all 9.5- and 12.5-mm surface mixtures with A and D designations in maintenance resurfacing contracts, totaling 2.36 million tons of BMD mixtures (Diefenderfer et al., 2026).

The Virginia Transportation Research Council (VTRC) has conducted multiple research projects to address challenges associated with developing the BMD specification and supporting its implementation in an efficient and responsible manner in Virginia. The data analysis from these projects revealed significant differences in results from a given BMD test method between lots or samples of the same mixture. In some cases, these differences were statistically significant, and in others, they were also practically significant, where one lot or sample produced a passing result, and another failed. Additional discrepancies were observed between VDOT district and producer test results, with mixtures from the same production subplot passing for one organization but failing for the other. These observations highlight the need to quantify the effect of factors such as reheating and mixture composition on BMD test results.

In 2022, a preliminary analysis focusing on the available cracking tolerance index (CT_{Index}) data from the indirect tensile cracking test (IDT-CT) was conducted, specifically considering the reheating factor to help address these concerns (Boz et al., 2022). This analysis involved BMD mixtures and their controls placed during the 2019 through 2021 paving seasons. The findings revealed a clear effect of asphalt mixture reheating on test results, showing a statistically significant reduction in the CT_{Index} due to reheating compared with the non-reheat case. As a result, a recommendation was made to set a CT_{Index} threshold of 95 for non-reheated, or hot-compacted, asphalt mixtures during production. Similar analyses are still needed for other

tests used in VDOT's BMD specification, including the Cantabro, Asphalt Pavement Analyzer (APA), and indirect tensile at high temperature (IDT-HT) tests.

Variations in BMD results may also arise from differences in asphalt loose mixture storage time. This storage time, commonly referred to as *lag time*, represents the time interval between asphalt mixture sampling during production and fabrication of BMD specimens. Producers typically experience shorter lag times because of the proximity of their laboratories to asphalt plants and the 48-hour requirement to submit BMD test results under VDOT's special provision. In contrast, VDOT district laboratories may experience longer lag times because of the number of plants within a given district and the distances between plants and district laboratories. Taylor (2025) observed no statistical difference in CT_{Index} values for 90% of the evaluated plant-mixed, laboratory-compacted (PMLC) mixtures for lag times ranging from 2 days to 2 months. Boz et al. (2025a) evaluated the effect of lag time on IDT-HT test results for laboratory-mixed, laboratory-compacted mixtures and observed that the differences between different lag times were insignificant when the single-operator precision estimate of the test was considered. Further investigations are warranted to assess the effect of lag time on other BMD tests using PMLC mixtures.

Recent studies have also been conducted to assess the effect of production variability on BMD test results in Virginia (Bowers et al., 2023; Diefenderfer et al., 2024). From these studies, it was observed that changes in asphalt content (AC) and gradation have a complex influence on BMD results, indicating the potential effect of other volumetric parameters. In addition, given the material variability between design and production, some mixtures that were balanced at the design stage became unbalanced during production by failing to meet some of the performance thresholds. For instance, the study highlighted that mixtures designed with average performance values that fall within the single-operator precision limits of performance thresholds may still face a risk of failing during production because of variability in materials and testing (Diefenderfer et al., 2024).

Therefore, proper consideration of the effects of reheating and mixture composition on BMD test results is essential to maintain consistency and quality in asphalt mixture production. Such efforts will help ensure robust quality measurement practices and responsible BMD implementation in Virginia.

PURPOSE AND SCOPE

The purpose of this project was to quantify the effects of reheating and mixture composition on BMD test results and to identify the volumetrics and gradation parameters that may significantly influence asphalt mixture laboratory performance. The scope of work included a literature review, statistical analysis of BMD datasets from past VTRC studies and VDOT's 2024 BMD statewide production, a statewide stakeholder survey to gather insights regarding typical production-level practices for BMD testing, and testing four plant-produced asphalt mixtures to assess the effect of lag time on BMD results and to verify the findings related to reheating and mixture composition.

METHODOLOGY

Literature Review

A brief literature review was conducted to gather information related to effects of reheating and mixture composition on asphalt mixture laboratory performance. Multiple databases and search engines were used, including TRID, Scopus, Web of Science, Science Direct, Google Scholar, ROSA-P, and Google.

Reheating and Mixture Composition Analyses

Datasets

Three major datasets (Datasets 1 to 3), including asphalt mixture volumetric parameters, gradation, and performance test results, were used. Dataset 1 was obtained from previous VTRC projects and consisted of PMLC specimens tested at VTRC. Dataset 1 included data from 30 job mix formulas (JMFs) totaling 104 subplot production results. Dataset 1 comprised BMD mixtures designed in accordance with VDOT’s BMD special provision, as well as their corresponding control mixtures (if available) designed using VDOT’s Superpave specification. More details of each mixture can be found elsewhere (Diefenderfer et al., 2021, 2023a, 2023b).

Dataset 2 included test results from 81 mixtures evaluated for volumetric properties, binder characteristics, and asphalt mixture performance using laboratory-mixed, laboratory-compacted specimens. Each mixture incorporated varying types and quantities of component materials. These combinations were determined based on the tolerance limits established by VDOT specifications (VDOT, 2020) and were developed systematically to isolate the influence of individual factors within the selected combinations. More details of each mixture can be found elsewhere (Bowers et al., 2023; Diefenderfer et al., 2024).

Dataset 3 included laboratory test results from PMLC specimens obtained from VDOT’s 2024 statewide BMD production. Dataset 3 included volumetrics, gradation, and performance test results from producers and VDOT districts for surface mixtures produced under maintenance schedule projects. Dataset 3 included 81 JMFs from eight VDOT districts, corresponding to production results from 758 sublots. All asphalt mixtures in Dataset 3 were designed following VDOT’s BMD specification. Table 1 summarizes the range of values within Datasets 1 to 3, along with VDOT specification limits.

Table 1. Range of Values for Volumetrics, Gradation, and Performance Results for Datasets 1 to 3

Parameter	Minimum–Maximum			VDOT Specification	
	Dataset 1	Dataset 2 ^a	Dataset 3	9.5 mm	12.5 mm
General Information					
Number of Job Mix Formulas	30	81	81	--	--
NMAS (mm)	9.5–12.5	9.5–12.5	9.5–12.5	--	--
RAP Content (%)	15–40	30–40	16–30	--	--
Binder PG	58-28, 64S-22, 64E-22	58-28, 64S-22	58-28, 64S-22	--	--

Parameter	Minimum–Maximum			VDOT Specification	
	Dataset 1	Dataset 2 ^a	Dataset 3	9.5 mm	12.5 mm
Mixtures Containing Recycling Agents?	Yes	Yes	No	--	--
Volumetric Parameters					
AC (%)	4.1–6.6	5.0–6.5	5.1–6.9	--	--
Rice SG, G_{mm}	2.444–2.720	2.422–2.711	2.387–2.766	--	--
Aggregate Bulk SG, G_{sb}	2.651–2.906	2.684–2.935	2.583–3.031	--	--
Absorbed AC, P_{ba} , (%)	0.01–0.57	0.03–0.71	0.01–0.73	--	--
VTM (%)	0.9–5.5	1.0–7.1	0.8–7.0	2–5	2–5
VMA (%)	13.8–19.0	14.0–19.7	14.6–19.7	≥ 16	≥ 15
VFA (%)	66–94	63.9–94.0	63–95	70–85	68–84
FA Ratio	0.7–1.4	0.68–1.35	0.7–1.3	0.7–1.3	0.7–1.3
Gradation Parameters (Percent Passing)					
No. 4 (4.75 mm)	43.6–75.2	50.1–67.5	44–81	58–80	58–80
No. 8 (2.36 mm)	31.6–49.0	31.1–46.3	30–49	38–67	38–50
No. 30 (0.6 mm)	16.4–28.5	14.3–25.5	15–32	≤ 23	≤ 23
No. 200 (0.075 mm)	3.7–8.3	4.2–7.0	3.3–8.0	2–10	2–10
Performance Parameters					
Cantabro Mass Loss (%)	1.7–10.5	3.6–13.9	0.7–8.4	≤ 7.5	≤ 7.5
CT_{Index}^b	41.7–263.9	19.4–185.7	53.6–281.0	≥ 70 or 95	≥ 70 or 95
APA Rut Depth (mm)	2.0–7.7	1.7–9.4	--	≤ 8.0	≤ 8.0
HT Strength (kPa)	109.9–302.7	--	58.9–201.0	≥ 100	≥ 100

-- = data not available; AC = asphalt content; APA = Asphalt Pavement Analyzer; CT = cracking tolerance; D = dataset; FA = fines-to-asphalt; HT = high temperature; NMAS = nominal maximum aggregate size; PG = performance grade; RAP = reclaimed asphalt pavement; SG = specific gravity; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture. ^a Parameters exceeding VDOT specifications were intentional, aimed at studying the effect of production variability on performance test results. ^b CT_{Index} criterion is equal to 70 for design or production (reheat) or 95 for production (non-reheat).

Effect of Reheating on Balanced Mix Design Test Results

The analysis was conducted in two parts. The first part involved validating the CT_{Index} thresholds for the reheated and non-reheated conditions using Dataset 3. Specifically, the previously used thresholds of 70 (reheated) and 95 (non-reheated), as well as the thresholds of 100 (reheated) and 130 (non-reheated) planned for implementation in the 2026 construction season, were validated. The second part of the analysis involved evaluating the need for separate reheated and non-reheated thresholds for Cantabro, APA, and IDT-HT tests, followed by validation and risk assessment analyses. VDOT’s BMD special provision currently does not differentiate reheated and non-reheated performance criteria for these tests as it does for IDT-CT.

Dataset 1 was used to develop preliminary thresholds for reheated and non-reheated Cantabro mass loss (CML) and APA test results. For IDT-HT, data from the mixtures collected in 2025 for this study (M1 to M4) were combined with data from VTRC Project No. 25-R14 for further analysis (Boz et al., 2025a). Dataset 1 was not used for the IDT-HT analysis because it included only dry strength results instead of wet strength values, which is the current conditioning method specified in Virginia Test Method 145 (VDOT, 2025). Finally, Dataset 3 was used for validation and risk assessment analyses for the CML and IDT-HT analyses. Corresponding analysis for APA was not conducted because no data were available in Dataset 3, given that APA is only conducted during production on request. Dataset 2 was not used to

evaluate the effect of reheating practices because no data were available from PMLC mixtures under reheated and non-reheated conditions.

The IDT-CT, CML, and IDT-HT validation analyses with Dataset 3 were conducted by comparing reheated (VDOT districts) and non-reheated (producers) test results. Producers that reported reheated IDT-CT, CML, and IDT-HT results were not included in the analysis. Although most of Dataset 3 specified whether reheated or non-reheated conditions were adopted, part of the dataset did not include such information. To help overcome this limitation, the survey responses were leveraged to supplement Dataset 3 with information regarding reheating or non-reheating practices during production.

Prior to conducting any analysis with Dataset 3, a preprocessing step was conducted to help remove production samples that failed (1) the volumetric process tolerance limits defined by VDOT specifications (VDOT, 2020), (2) the multilaboratory variability limits for volumetrics data (i.e., $d2s$ limits between producer and district), and (3) single-operator precision limits for IDT-CT (i.e., coefficient of variation of 18.3%) and the IDT-HT test (i.e., range of 35.6% of the test results average). This screening step was important to help eliminate problematic data potentially caused by asphalt mixture production issues at the plant. The analysis was conducted only for IDT-CT and IDT-HT results, given that no precision statements are yet available for the Cantabro test. In addition, a two-level analysis (local and global) using the interquartile range (IQR) method was conducted to detect and remove any potential outliers from IDT-CT and IDT-HT results. Outliers were removed within a given JMF (local) and across the entire dataset (global). This step was critical to help remove abnormal BMD test results, likely caused by testing issues or reporting errors. Outlier removal procedures were also performed for Datasets 1 and 2.

For the statistical analysis and graph generation, JMP® Software version 12.2.0 (JMP Statistical Discovery LLC, 2015), R version 4.5.2 (R Core Team, 2025), and Microsoft Excel programs were used.

Effect of Mixture Composition on Balanced Mix Design Test Results

Dataset 2 was used to evaluate and identify the significant volumetrics and gradation parameters that affect BMD test results. Dataset 2 was selected because the test results were obtained from a single operator using the same set of equipment, reducing the potential variability in test results and minimizing the existence of confounding factors in the statistical analysis. Given that Dataset 2 included only results for Cantabro, IDT-CT, and APA, data from VTRC Project No. 23-R11 were used to conduct the IDT-HT analysis (Boz et al., 2023). It is worth noting that the IDT-HT test results in this dataset were obtained from specimens subjected to dry conditioning because this study was conducted prior to the evaluation of the wet conditioning as an alternative expedited protocol. Although the current specification requires wet conditioning, it is not expected that different conditioning methods (i.e., dry versus wet) would result in different correlations between high temperature strength and volumetrics and gradation parameters. In all cases, the volumetric parameters were obtained from the volumetric pills (i.e., 150-mm diameter and 115-mm tall) used in mixture acceptance and quality control protocols

during production. In this sense, volumetrics data from individual BMD specimens were not included in the analysis.

The analysis was divided into four parts. First, the correlation between the Cantabro, IDT-CT, APA, and IDT-HT results and the individual volumetrics and gradation parameters were evaluated in terms of direction (i.e., positive or negative) and magnitude (using the Pearson correlation coefficient, r). This analysis provided initial insights into the individualized effect on mixture performance from parameters typically measured as part of the quality assurance processes during asphalt mixture production.

Second, the effect of binder quantity and quality on BMD tests was evaluated. For the binder quantity, the volume of effective binder (VBE) was used in the correlation analysis with Cantabro, IDT-CT, and APA results. The goal of this analysis was to investigate whether VBE can help explain variations in performance for mixtures with similar AC. In terms of binder quality, information regarding reclaimed asphalt pavement (RAP) content, virgin binder performance grade (PG), and recycling agents (RAs) were used as categorical factors to indirectly estimate the binder quality. These parameters were selected because Dataset 2 did not include actual binder properties and other rheological parameters from the extracted and recovered binder that are typically used to correlate with asphalt mixture performance.

Third, statistical analysis was conducted to identify the significant volumetric and gradation factors that affect the test results for Cantabro, IDT-CT, APA, and IDT-HT. Preliminary assessment was conducted to verify the collinearity among different factors and remove them from the statistical analysis to avoid introducing confounding variables. For the collinearity analysis, pairwise correlation matrices were developed, using r to quantify the degree of correlation among variables. For this analysis, the degree of correlation was categorized as strong ($|r| \geq 0.7$), moderate ($0.5 \leq |r| < 0.7$), and weak ($|r| < 0.5$). To mitigate collinearity, when two factors demonstrated moderate to strong intercorrelation, the factor retained was the one exhibiting the stronger association with the corresponding BMD test result. After removing the collinear factors, a forward stepwise multiple linear regression was conducted for each BMD test, with a p-value of 0.05 as the stopping criterion for factors to both enter and leave the model. This approach was selected because only the significant volumetrics and gradation parameters would remain in the model, and the sign of the fitting coefficients would provide information regarding the correlation directionality between such parameters and the BMD test results. Additional analysis using the variance inflation factor (VIF) was conducted to ensure the significant volumetric and gradation parameters were not collinear with each other. VIF values below 5 typically suggest low collinearity, whereas VIF values greater than 10 suggest a high collinearity level.

Finally, a case study was conducted to evaluate the potential of using performance-predictive equations as practical decision-support tools to help determine when additional testing is needed during production. This effort is particularly relevant considering the changes in the 2026 construction season specifications, which reduce IDT-CT and Cantabro testing frequency during production from every 2,000 tons to every 4,000 tons, provided the first three consecutive lots meet all performance criteria. Although Cantabro, IDT-HT, and IDT-CT are performed during production, the analysis focused solely on the cracking performance (i.e., CT_{Index}) for

three reasons. First, cracking is considered the most critical distress. Second, rutting is generally not a surface mixture performance issue in Virginia. Third, although Cantabro test frequency is also reduced, volumetric pills are routinely fabricated for quality assurance purposes at a higher frequency (i.e., every 1,000 tons). This observation suggests that, if needed, Cantabro tests can be more easily performed than IDT-CT, given that no extra sampling and specimen fabrication would be required.

Therefore, CT_{Index} predictive equations were developed using Dataset 2 under a tiered approach (Tier 1 through Tier 3) to reflect varying testing frequencies for volumetrics and gradation data during production. A forward stepwise linear regression approach was adopted to obtain the predictive equations. Two stopping criteria were selected—p-value and minimum Akaike information criterion (AIC). The p-value criterion includes only significant factors at a significance level of 0.05, whereas the AIC criterion balances the model goodness of fit and complexity by penalizing the regression model when additional independent variables are considered in order to minimize overfitting risks. The selected equation between the p-value and AIC criterion was the one that yielded improved predictability. Verification of the CT_{Index} equations' predictability was conducted using Dataset 3. Lastly, a risk assessment analysis was conducted to quantify the potential risk of adopting such predictive equations when used as a support tool to determine whether additional testing is needed during production. If effective, the analysis framework presented herein may help practitioners identify potential mix design issues early and take timely corrective actions.

For the statistical analysis and graph generation, JMP® Software version 12.2.0 (JMP Statistical Discovery LLC, 2015), R version 4.5.2 (R Core Team, 2025), and Microsoft Excel programs were used.

Laboratory Evaluation of Storage and Mixture Conditioning Protocols

Survey

A survey was distributed on October 21, 2024, to gather information on the practices of producers and VDOT districts regarding the sampling of plant-produced asphalt surface mixtures, material conditioning (reheat or non-reheat), and specimen fabrication practices for BMD tests during production. For this project, the goal of the survey was threefold: (1) to evaluate typical time intervals between mixture sampling and specimen fabrication (i.e., lag time) and between specimen fabrication and BMD test completion (i.e., dwell time); (2) to refine the laboratory evaluation of storage and mixture conditioning protocols based on typical lag times reported in the survey; and (3) to provide additional context regarding specimen fabrication practices to help explain observed trends in VDOT's 2024 BMD production data. Forty-six survey responses were submitted between October 21, 2024, and November 5, 2024. Participants included 39 producers and 7 districts, with responses provided by laboratory technicians, laboratory managers, or quality assurance and quality control managers.

Materials

Four additional PMLC asphalt surface mixtures (M1 to M4) were collected in 2025 for asphalt mixture and binder performance testing to quantify the potential effect of different lag times and the mixture conditioning protocols during production (i.e., non-reheat versus reheat). Three sets of specimens per mixture were fabricated under three different conditions: (1) non-reheated (NR); (2) reheated after 1-day lag time (R1d); and (3) reheated after 4-day lag time (R4d). The 1- and 4-day time intervals were selected based on the survey responses and represent the range of typical lag times reported by producers and districts. For the NR case, the loose mixture was sampled into metal pans at the plant while hot and brought to the plant's laboratory to fabricate the volumetric pills (further used for Cantabro testing) and BMD specimens (for IDT-CT, APA, and IDT-HT testing). For the R1d and R4d, the loose mixture was sampled at the plant into cardboard boxes (10 inches by 10 inches by 10 inches) and brought to VTRC. After cooling to room temperature, the mix was held for 1 day (R1d) or 4 days (R4d), then reheated to compaction temperature and used to fabricate BMD specimens. In all cases, dwell time was kept to a minimum—typically 1 to 2 days, depending on the day of the week and on laboratory logistics—to reduce the effect of extended specimen storage on test results. All NR specimens were fabricated by the producer, and all R1d and R4d specimens were fabricated at VTRC.

Volumetrics and gradation parameters are presented in Table 2 for mixtures M1 and M2 and Table 3 for M3 and M4. Overall, volumetrics and gradation parameters during production were within process tolerance limits established by VDOT specifications for both R1d and R4d conditions (VDOT, 2020). The only exception was M4, which consistently showed voids filled with asphalt (VFA) values higher than the maximum limit of 85% and variable results between R1d and R4d for percent passing the No. 200 (0.075 mm) sieve and fines-to-asphalt (FA) ratio.

Test Methods

Volumetrics and Gradation Parameters

The asphalt binder content of the asphalt mixture was determined using the ignition oven in accordance with Virginia Test Method 102, *Determination of Asphalt Content from Asphalt Paving Mixtures by the Ignition Oven Method* (VDOT, 2025). The sieve analysis of the recovered aggregates after ignition of the asphalt mixture was conducted in accordance with American Association of State Highway and Transportation Officials (AASHTO) T 11, *Standard Method of Test for Materials Finer Than 75- μ m (No. 200) Sieve in Mineral Aggregates by Washing* (AASHTO, 2025), and AASHTO T 27, *Standard Method of Test for Sieve Analysis of Fine and Coarse Aggregates* (AASHTO, 2024). The theoretical maximum specific gravity (G_{mm}) of the mixture was obtained following AASHTO T 209, *Standard Method of Test for Theoretical Maximum Specific Gravity (G_{mm}) and Density of Asphalt Mixtures* (AASHTO, 2025). For mix verification, the loose mix was conditioned in the oven at the compaction temperature and then compacted at N_{design} of 50 gyrations using a Superpave gyratory compactor in accordance with AASHTO T 312, *Preparing and Determining the Density of Asphalt Mixtures Specimens by Means of the Superpave Gyratory Compactor (SGC)* (AASHTO, 2022). The specimen height and other volumetric parameters were then determined, including bulk specific gravity (G_{sb}), voids in total mixture (VTM), voids in the mineral aggregate (VMA), VFA, and others.

Table 2. Volumetrics and Gradation Information for Mixtures M1 and M2

Mix ID (VTRC Log Number)	M1 (25-1049)				M2 (25-1050)			
Mix Type	SM-12.5D (BMD P+VO)				SM-9.5D (BMD P+VO)			
District	Fredericksburg				Richmond			
NMAS (mm)	12.5				9.5			
RAP Content (%)	30				30			
Virgin Binder PG	64S-22				64S-22			
Additives	Zycotherm SP2 (0.03%)				Zycotherm SP2 (0.05%)			
Sample	JMF	Limits ^a	Production (1-day)	Production (4-day)	JMF	Limits ^a	Production (1-day)	Production (4-day)
Volumetric Parameters								
AC (%)	5.60	±0.43	5.68	5.55	5.80	±0.43	6.13	6.04
Rice SG, G_{mm}	2.598	-	2.608	2.610	2.489	-	2.473	2.473
VTM (%)	4.0	2.0–5.0	2.6	2.8	3.1	2.0–5.0	3.3	3.3
VMA (%)	17.3	Min. 15	16.4	16.2	16.6	Min. 16	17.1	16.9
VFA (%)	76.9	68-84	84.4	83.0	81.5	70–85	80.7	80.4
FA Ratio	0.9	0.7–1.3	1.1	1.2	0.9	0.7–1.3	0.9	0.9
Mixture Bulk SG, G_{mb}	2.492	-	2.541	2.538	2.300	-	2.391	2.391
Aggregate Effective SG, G_{se}	-	-	2.873	2.868	-	-	2.722	2.718
Aggregate Bulk SG, G_{sb}	-	-	2.866	2.861	-	-	2.708	2.704
Absorbed AC, P_{ba} (%)	-	-	0.09	0.09	-	-	0.20	0.20
Effective AC, P_{be} (%)	-	-	5.59	5.47	-	-	5.95	5.86
Gradation Parameters (Percent Passing)								
1 in. (25.0 mm)	-	-	100.0	100.0	-	-	100.0	100.0
¾ in. (19.0 mm)	100	±0.0	100.0	100.0	-	-	100.0	100.0
½ in. (12.5 mm)	96	±5.7	95.4	95.5	100	±0.0	99.6	99.5
3/8 in. (9.5 mm)	88	±5.7	86.0	86.6	93	±5.7	92.5	90.3
No. 4 (4.75 mm)	-	-	55.9	56.2	62	±5.7	63.5	60.2
No. 8 (2.36 mm)	40	±5.7	38.0	37.8	45	±5.7	46.9	44.6
No. 16 (1.18 mm)	-	-	28.6	28.7	-	-	36.0	34.4
No. 30 (600 µm)	-	-	21.7	21.9	-	-	24.0	23.1
No. 50 (300 µm)	-	-	15.5	15.9	-	-	13.4	12.9
No. 100 (150 µm)	-	-	9.9	10.2	-	-	8.2	7.9
No. 200 (75 µm)	5.4	±1.4	6.2	6.4	5.2	±1.4	5.5	5.3

- = data not available; AC = asphalt content; BMD = balanced mix design; FA = fines-to-asphalt; JMF = job mix formula; NMAS = nominal maximum aggregate size; RAP = reclaimed asphalt pavement; PG = performance grade; SG = specific gravity; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture. ^a Limits from production thresholds from Table II-14 and process tolerance for two tests from Table II-15 (VDOT, 2020).

Table 3. Volumetrics and Gradation Information for Mixtures M3 and M4

Mix ID (VTRC Log Number)	M3 (25-1061)				M4 (25-1072)			
Mix Type	SM-9.5A (BMD P+VO)				SM-9.5D (BMD P+VO)			
District	Northern Virginia				Salem			
NMAS (mm)	9.5				9.5			
RAP Content (%)	30				30			
Virgin Binder PG	64S-22				64S-22			
Additives	PC 2250 (0.05%)				Ad-here 62-40			
Sample	JMF	Limits ^a	Production (1-day)	Production (4-day)	JMF	Limits ^a	Production (1-day)	Production (4-day)
Volumetric Parameters								
AC (%)	5.70	±0.43	5.53	5.33	5.90	±0.43	6.23	6.25
Rice SG, G_{mm}	2.648	-	2.678	2.673	2.596	-	2.601	2.605
VTM (%)	3.4	2.0–5.0	5.2	4.8	3.3	2.0–5.0	2.4	2.2
VMA (%)	17	Min. 16	18.2	17.4	17.4	Min. 16	17.5	17.4
VFA (%)	80	70–85	71.6	72.5	81.1	70–85	86.1	87.3
FA Ratio	1.2	0.7–1.3	1.2	1.2	1.2	0.7–1.3	1.5	1.0
Mixture Bulk SG, G_{mb}	2.557	-	2.539	2.545	2.508	-	2.538	2.547
Aggregate Effective SG, G_{se}	-	-	2.954	2.936	-	-	2.894	2.900
Aggregate Bulk SG, G_{sb}	-	-	2.933	2.915	-	-	2.884	2.890
Absorbed AC, P_{ba} (%)	-	-	0.25	0.25	-	-	0.12	0.12
Effective AC, P_{be} (%)	-	-	5.29	5.09	-	-	6.11	6.13
Gradation Parameters (Percent Passing)								
1 in. (25.0 mm)	-	-	100.0	100.0	-	-	100.0	100.0
¾ in. (19.0 mm)	-	-	100.0	100.0	-	-	100.0	100.0
½ in. (12.5 mm)	100	±0.0	100.0	100.0	100	±0.0	100.0	99.9
3/8 in. (9.5 mm)	95	±5.7	95.3	94.6	96	±5.7	97.0	95.6
No. 4 (4.75 mm)	61	±5.7	54.5	52.2	57	±5.7	60.7	56.8
No. 8 (2.36 mm)	40	±5.7	35.7	33.9	40	±5.7	44.2	40.8
No. 16 (1.18 mm)	-	-	26.0	25.1	-	-	38.7	35.6
No. 30 (600 µm)	-	-	19.2	18.7	-	-	31.3	28.3
No. 50 (300 µm)	-	-	13.8	13.5	-	-	19.9	17.0
No. 100 (150 µm)	-	-	9.4	9.2	-	-	11.6	8.6
No. 200 (75 µm)	6.3	±1.4	6.3	6.1	6.7	±1.4	9.0	6.0

- = data not available; AC = asphalt content; BMD = balanced mix design; FA = fines-to-asphalt; JMF = job mix formula; NMAS = nominal maximum aggregate size; PG = performance grade; RAP = reclaimed asphalt pavement; SG = specific gravity; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture. ^a Limits from production thresholds from Table II-14 and process tolerance for two tests from Table II-15 (VDOT, 2020).

Cantabro Abrasion Loss

The Cantabro abrasion loss test was conducted following AASHTO T 401, *Standard Method of Test for Abrasion Loss of Asphalt Mixture Specimens* (AASHTO, 2022), and Virginia Test Method 144, *Cantabro Abrasion Loss of Asphalt Mixture Specimens* (VDOT, 2025). All specimens were fabricated with 150-mm diameter and 115 ± 5 -mm height and tested at room temperature, approximately at 25°C. The CML was calculated with lower CML, suggesting greater durability.

Indirect Tensile Cracking Test

IDT-CT was conducted in accordance with the American Society for Testing Materials (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, 2019), and Virginia Test Method 143, *Determination of Cracking Tolerance Index for Asphalt Mixture Using the Indirect Tensile Cracking Test (IDT-CT) at Intermediate Temperature* (VDOT, 2025). The test was performed on specimens with 150-mm diameter and 62-mm height, compacted at $7.0 \pm 0.5\%$ air voids, and conditioned in a water bath at 25°C for 2 hours prior to testing. The CT_{Index} was calculated from the load-displacement curve, with higher CT_{Index} indicating improved cracking resistance.

Asphalt Pavement Analyzer

The APA test was conducted in accordance with AASHTO T 340, *Standard Method of Test for Determining the Rutting Susceptibility of Hot Mix Asphalt (HMA) Using the Asphalt Pavement Analyzer (APA)* (AASHTO, 2023), and Virginia Test Method 142, *Method of Test for Determining Rutting Susceptibility of Hot Mix Asphalt Using the Asphalt Pavement Analyzer (APA)* (VDOT, 2025). The test was performed on specimens with 150-mm diameter and 75-mm height, compacted at $7.0 \pm 0.5\%$ air voids, and conditioned at 64°C for 6 hours prior to testing. The test was conducted at 64°C, and the average rut depth (RD) was calculated after the application of 8,000 loading cycles. Lower RD corresponds to improved rutting resistance.

Indirect Tensile at High Temperature Test

The IDT-HT test was conducted following Virginia Test Method 145, *Method of Test for Determining Rutting Susceptibility of Asphalt Mixtures Using the Indirect Tensile at High Temperature (IDT-HT) Test* (VDOT, 2025). The test was performed on specimens with 150-mm diameter and 62-mm height, compacted at $7.0 \pm 0.5\%$ air voids, and conditioned in a water bath at 54.4°C for 1 hour prior to testing. The IDT-HT strength was then calculated, with higher strength values suggesting improved rutting resistance.

RESULTS AND DISCUSSION

Literature Review

Effect of Reheating

Reheating plant-produced asphalt mixtures is a common laboratory practice in which loose mix is allowed to cool following production and later reheated to compaction temperature or a workable state prior to specimen fabrication and testing. This practice is primarily driven by logistical considerations. Plant-produced mixtures are typically sampled at the plant or job site and transported to a laboratory, where immediate compaction is often not feasible because of travel time, laboratory scheduling, or equipment availability. Although reheating facilitates controlled specimen preparation and repeatable testing, research consistently shows that it is not mechanically neutral. Additional thermal exposure during reheating promotes oxidative aging of the binder and can measurably influence mixture performance characteristics, such as stiffness, rutting resistance, cracking tolerance, and fatigue performance.

Multiple studies demonstrate that reheating introduces additional oxidative aging beyond that occurring during plant production. Xiao et al. (2015) evaluated plant-produced foamed warm-mix asphalt mixtures and reported that reheated specimens exhibited higher recovered binder failure temperatures and increased $|G^*|/\sin(\delta)$ values compared with non-reheated plant specimens. Lower phase angles were also observed, indicating a shift toward more elastic and less relaxational behavior. These findings are consistent with the fundamental expectation that reheating increases binder stiffness and reduces viscoelastic relaxation capacity.

Consistent with asphalt binder stiffening, reheating generally increases asphalt mixture stiffness and improves measured rutting resistance. Al-Qadi et al. (2012) reported higher dynamic modulus values and improved Hamburg rutting resistance for reheated stone-matrix asphalt mixtures relative to specimens compacted without reheating. Wang et al. (2013) observed aging (stiffening) ratios between approximately 1.2 and 2.0 for flow number and rut depth values, demonstrating that reheating can substantially reduce rutting susceptibility. Kidd et al. (2019) similarly reported increases in flexural modulus typically ranging from 20% to 50%, with some cases approaching 70%, depending on temperature and loading frequency. Xiao et al. (2015) also found lower APA rut depths and higher indirect tensile strength in reheated mixtures relative to non-reheated plant specimens.

In contrast to rutting resistance, reheating generally reduces cracking tolerance. Leng et al. (2012) reported increased brittleness in reheated plant-produced stone-matrix asphalt mixtures quantified by the semi-circular bend test. Reheating also altered comparative rankings among mixture types, demonstrating that additional aging can influence performance interpretation. Daniel et al. (2015; 2019) observed that reheated plant mixtures were stiffer and exhibited altered fatigue damage predictions compared with non-reheated plant-compacted specimens. Although relative mixture rankings were often preserved, predicted damage magnitudes differed, indicating the sensitivity of cracking-related indices to reheating. Similarly, Yin et al. (2017) showed statistically significant reductions in the indirect tensile N_{flex} factor values for reheated specimens compared with minimally reheated (hot production) specimens. Mixtures containing

rejuvenators and higher recycled asphalt shingle contents were particularly sensitive, suggesting that additional aging exposure may diminish rejuvenator effectiveness. Oshone et al. (2020) likewise demonstrated reduced fatigue resistance in reheated mixtures relative to laboratory-produced mixtures that were not reheated.

More recent evaluations using simplified cracking indices, such as CT_{Index} , provide additional perspective on reheating effects. Yin et al. (2020) evaluated five plant-produced mixtures reheated for approximately 2 to 3 hours at compaction temperature to restore workability, with specimen portions reheated until reaching compaction temperature prior to compaction. When CT_{Index} results were compared between non-reheated and reheated specimens, no uniform trend was observed across mixtures. Reported percent differences in CT_{Index} ranged from approximately -40% to +50%, depending on mixture composition. In most cases, reheating did not produce a statistically significant reduction in CT_{Index} ; however, measurable reductions were observed for selected mixtures.

Taylor (2025) further examined the combined effects of production lag time and reheating dwell time on CT_{Index} results for plant-produced mixtures. Mixtures were subjected to controlled lag periods prior to reheating, followed by reheating at compaction temperature for specified dwell durations prior to compaction. Results indicated that extended lag time combined with prolonged reheating dwell time could produce measurable reductions in CT_{Index} relative to non-reheated specimens, particularly for mixtures with higher RAP contents. In contrast, reheating durations representative of typical laboratory practice produced relatively small changes in CT_{Index} . The magnitude of CT_{Index} reduction increased with cumulative thermal exposure, indicating that CT_{Index} is sensitive not only to reheating temperature but also to total aging time prior to compaction.

The severity of reheating effects is strongly influenced by the reheating protocol (Lemke et al., 2018). Factors such as reheating duration, temperature, sample geometry (e.g., thin pans versus bulk containers), oven air circulation, and covered versus uncovered conditioning all affect the extent of aging (Lemke et al., 2018; Boz et al., 2025c). Overall, the literature indicates that reheating plant-produced asphalt mixtures prior to laboratory compaction introduces measurable additional aging. Reheating generally increases binder and mixture stiffness, improves measured rutting resistance, reduces cracking tolerance, and alters fatigue response. The magnitude of these effects depends on several factors, including binder grade, recycled content, rejuvenator use, and the severity and duration of reheating, and is often both test and mixture specific.

Effect of Mixture Composition

Volumetric properties and aggregate gradation are essential parameters in asphalt mixture design and serve as the basis for mixture acceptance and quality control during production. These factors not only ensure compliance with specifications but also have a significant influence on asphalt mixture performance in terms of durability, cracking, and rutting resistance. However, as the industry increasingly considers incorporating alternative materials in asphalt mixtures, such as higher contents of recycled material, polymer-modified binders, and various

additives, the direct relationship between volumetrics, gradation, and performance becomes less straightforward.

To address these uncertainties, performance-oriented mix design approaches have been evaluated and proposed in recent years. One such approach is BMD, which incorporates performance tests and criteria to optimize mix designs and balance resistance to different distress mechanisms. Although BMD offers a performance-only alternative that does not require volumetric compliance (i.e., Approach D), volumetric and gradation parameters remain essential in current practice for both mixture design and production control. In this sense, understanding their effect on performance is crucial.

Previous research has explored these relationships extensively. Zaumanis et al. (2018) conducted an extensive literature review on the relationship between mixture composition and performance, noting that higher AC, less aged binders, and lower nominal maximum aggregate size reduce rutting resistance but improve cracking resistance. Rocha et al. (2024) analyzed Brazilian mixtures using fundamental laboratory tests and found that less viscous binders and higher AC led to improved fatigue resistance, whereas binders with higher viscosity and coarser mixtures resulted in improved rutting resistance. Wang et al. (2022) developed a fatigue index parameter (S_{app}) using the viscoelastic continuum damage model, and besides the strong correlation with field performance, it was found that lower VTM, lower binder stiffness, and the use of polymer-modified binders improved the fatigue resistance. Other studies also investigated the relationship between mixture composition and mixture performance measured from fundamental tests (Wang et al., 2019; Etheridge et al., 2019; Veeraragavan et al., 2022).

Studies focused on BMD testing have also highlighted the role of mixture components. Mivehchi et al. (2022) evaluated 69 mix designs from the state of Washington with various compositions and found that higher CT_{Index} values were typically associated with higher AC and coarser gradations, whereas VTM was not statistically significant. From Hamburg wheel tracking tests, higher binder PG, lower AC, and greater fine aggregate angularity improved rutting resistance. Other studies reported similar trends for APA rutting tests (Tarefder et al., 2003; Bennert, 2026). Machine learning approaches have also been employed and confirmed that AC, VTM, and VMA strongly influence asphalt mixtures' cracking and rutting resistances (Rivera-Perez et al., 2023).

Recent studies conducted in Virginia have also investigated the effect of mixture composition on performance. Bowers et al. (2023) showed that higher VTM led to increased CML values using different material sources in Virginia. This observation confirmed findings from Cox et al. (2017) that demonstrated that other factors, such as RAP content and fines-to-asphalt ratio, can also have a detrimental effect on CML. Boz et al. (2023) correlated mixture composition factors with laboratory performance of 17 PMLC mixtures and observed that G_{sb} , non-recoverable creep compliance ($J_{nr3.2}$), and gradation parameters (percent passing the 3/8", No. 16, No. 30, and No. 200 sieves) significantly affect IDT-HT results, and percent passing the No. 30 sieve, G_{sb} , VMA, and $J_{nr3.2}$ significantly affect APA results. Tong et al. (2025) employed machine learning techniques to develop CT_{Index} predictive equations using data from six mix designs and observed that AC and percent passing the No. 200 sieve are the factors with the highest level of importance in CT_{Index} predictions. Notably, recent studies observed correlations

between BMD results and field performance, including CT_{Index} and percent cracking in the field (Diefenderfer and Boz, 2025), as well as IDT-HT strength and field rutting (Taylor et al., 2026).

Ongoing national-level studies are also being conducted as part of the National Cooperative Highway Research Program (NCHRP) Project 09-72, in which the sensitivity of BMD performance test results is being quantified with respect to variations in the mix design and test variables. Past NCHRP projects also investigated the correlation between mix design variables and performance, highlighting the importance of such factors (Kandhal and Cooley Jr., 2003; Prowell et al., 2005; Christensen and Bonaquist, 2006). Overall, the literature indicates that volumetrics, gradation, and binder properties have a strong influence in the asphalt mixtures' performance.

Effect of Reheating on Balanced Mix Design Test Results

Validation of the Ratio in CT_{Index} Performance Criterion Between Reheated and Non-reheated Conditions

Figure 1a shows the comparison between CT_{Index} results obtained under non-reheated (producer) and reheated (district) conditions from Dataset 3. Each marker corresponds to the average BMD test result. Note that non-reheated CT_{Index} values are consistently higher than reheated values. Because the non-reheated condition involves fabricating specimens by minimizing loose mix cool down, the oxidative aging induced onto the asphalt binder within the mix is reduced compared with the specimens prepared under the reheated condition. In this sense, reheating the asphalt mix leads to an embrittlement effect, reducing its cracking resistance (i.e., lower CT_{Index} values). This observation corroborates the findings from previous VTRC studies (Diefenderfer et al., 2021; Boz et al., 2022; Diefenderfer et al., 2023a; Diefenderfer et al., 2023b; Habbouche et al., 2022). Moreover, Figure 1a shows that the higher CT_{Index} values for the non-reheated condition resulted in 5 out of 17 samples (29%) falling outside the multilaboratory reproducibility limit ($d2s\%$). This observation highlights that different conditioning methods can lead to statistically different test results from a multilaboratory perspective.

To validate the ratio in CT_{Index} performance criteria between the two conditions, the non-reheated CT_{Index} values in Figure 1a were multiplied by a correction factor to estimate their corresponding reheated results. The correction factor corresponds to the ratio between reheated and non-reheated CT_{Index} thresholds. Two sets of correction factors were used, one for the 70 and 95 thresholds (i.e., $70 \div 95 = 0.737$) and the other for the 100 and 130 thresholds (i.e., $100 \div 130 = 0.767$). The data were considered validated if the slope between the estimated reheated CT_{Index} and the measured reheated CT_{Index} was close to one. Figure 1b presents the two sets of estimated reheated CT_{Index} values versus measured reheated CT_{Index} values along with the corresponding linear regression curve. The slope of the linear regression curve for the 70 and 95 thresholds and 100 and 130 thresholds was equal to 1.049 and 1.003, respectively, with the corresponding regression curves coinciding with the line of equality (LOE). In addition, after applying the correction factors, all data points fell within the multilaboratory reproducibility limits ($d2s\%$), suggesting statistical similarity. Therefore, the analysis presented herein validated the ratio in CT_{Index} performance criteria between reheated and non-reheated conditions.

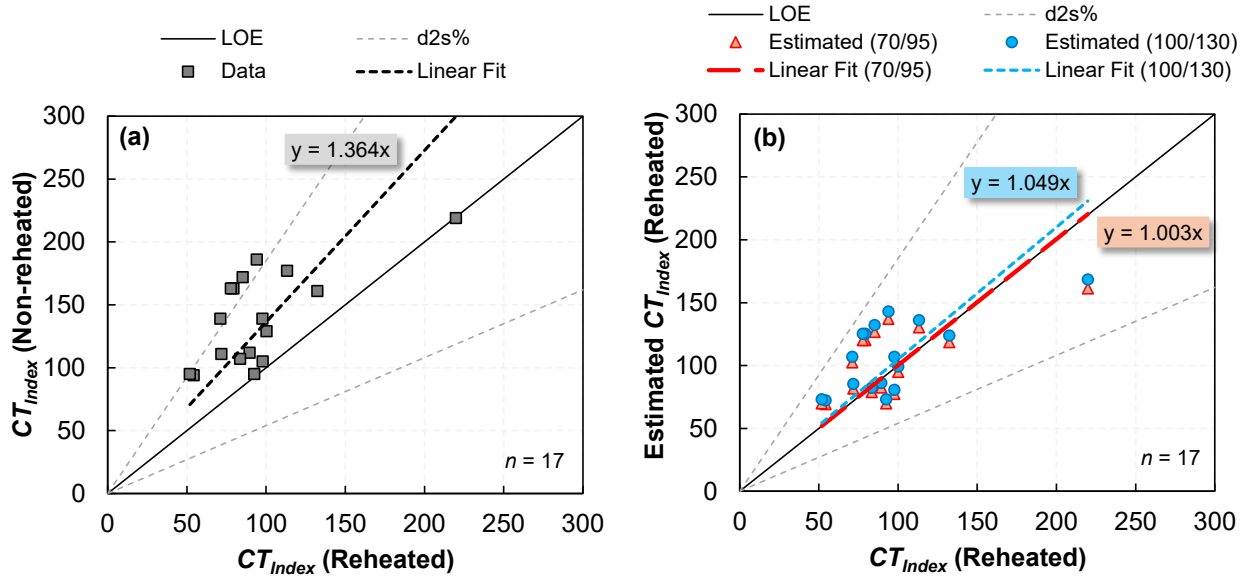


Figure 1. Comparing Reheated and Non-reheated CT_{Index} Results Using Dataset 3: (a) Measured Non-reheated CT_{Index} (Producer) and Measured Reheated CT_{Index} (District); (b) Estimated Reheated CT_{Index} (Producer) and Measured Reheated CT_{Index} (District). CT = cracking tolerance; LOE = line of equality; $d2s\%$ = multilaboratory reproducibility limit for indirect tensile cracking test; n = number of data points.

Effect of Reheating on Cantabro, APA, and IDT-HT Test Results

Statistical Analysis Comparing Non-reheated and Reheated BMD Results

Figure 2a through 2c show the comparison between non-reheated and reheated test results for the Cantabro, APA, and IDT-HT tests, along with the slope of the linear fit with a forcing intercept of zero, to help quantify the non-reheated to reheated relationship. Each marker corresponds to the average BMD test result.

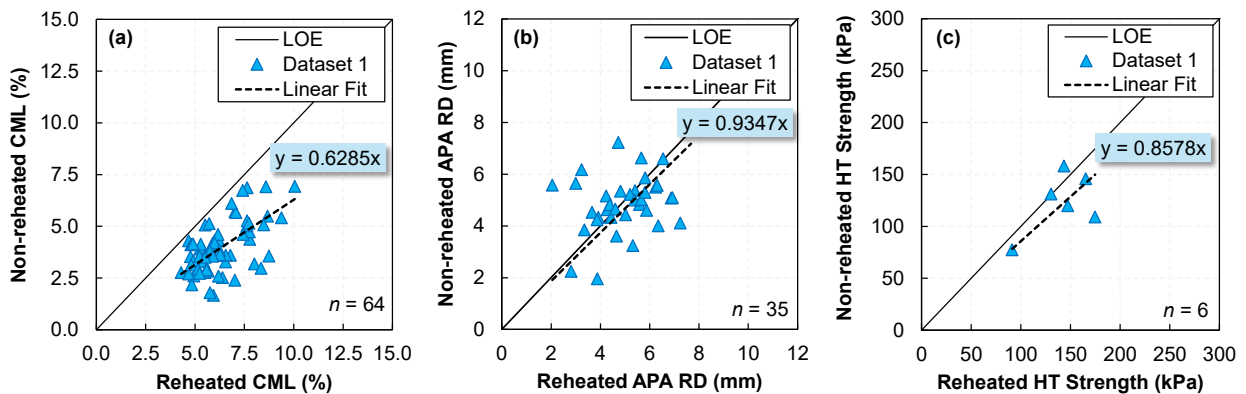


Figure 2. Relationship between Non-reheated and Reheated Results for (a) CML, (b) Asphalt Pavement Analyzer Rut Depth, and (c) HT Strength. APA = Asphalt Pavement Analyzer; CML = Cantabro mass loss; HT = high temperature; LOE = line of equality; n = number of data points; RD = rut depth.

For the Cantabro test, Figure 2a shows systematically higher CML values for the reheated condition compared with the non-reheated condition. The linear regression slope suggests that non-reheated CML values are approximately 63% of the reheated values. This trend is explained

by the additional aging induced onto the asphalt binder within the mix for the reheated condition at compaction temperature compared with the non-reheated condition. Higher aging levels increase the mixture embrittlement, making the specimens more prone to breakage inside the drum during the test, and leading to higher mass loss values.

For the APA test, Figure 2b shows no clear trend in rut depth measurements between reheated or non-reheated conditions, with data points evenly scattered around the line of equality. The slope of the linear regression suggests the non-reheated rut depth is 93% of the reheated rut depth.

For the IDT-HT test, although only six mixtures were available for comparison, the results in Figure 2c suggest that high temperature strength values are consistently higher at the reheated condition than the non-reheated condition. The additional aging at the reheated condition increases the asphalt specimen stiffness, leading to an increase in strength value. The slope of the linear regression suggests that the non-reheated high temperature strength values are 86% of the reheated condition.

Statistical analyses were conducted with the data in Figure 2 to assess whether the differences between non-reheated and reheated test results were significant or not. A mixed-effects model was used with “condition” as a fixed-effect factor (i.e., non-reheated or reheated) and “mixture” as a random-effect factor at a significance level of 0.05. The mixed-effects model was selected to account for mixture-to-mixture variability and correlation among specimen replicates of the same mix due to shared mixture characteristics (e.g., volumetrics, gradation, binder PG, source). The null hypothesis was that the coefficient multiplying the condition factor within the mixed-effects model is equal to zero. If the p-value is below the significance level of 0.05, the null hypothesis is rejected, suggesting that the “condition” factor is statistically significant. All assumptions regarding residuals’ normal distribution and homoscedasticity (constant variance) were met.

Table 4 shows the summary of the mixed-effects model for CML, APA rut depth, and IDT-HT strength results. The percentage of total variance for the mixture factor was greater than zero and equal to 61%, 36%, and 68% for the Cantabro, APA, and IDT-HT results, respectively, justifying the mixed-effects model selection with the mixture factor as a random effect. For the condition factor, the statistical analysis showed reheating has a significant effect on the results of the three tests, with a marginal significance for APA.

Table 4. Summary of Mixed-effects Model Analysis for Cantabro Mass Loss, APA Rut Depth, and IDT-HT Strength Results

Test	Number of Observations	Mixture Factor (Random Effect)		Condition Factor (Fixed Effect)			
		Percentage of Total Variance	Variance Ratio	Num. DF	Den. DF	F Ratio	p-value
Cantabro	342	60.91%	1.558	1	284	589.136	< 0.001
APA	100	35.99%	0.562	1	74	4.186	0.0443
IDT-HT	36	68.20%	2.145	1	29	10.900	0.0026

APA = Asphalt Pavement Analyzer; Den. = denominator; DF = degrees of freedom; IDT-HT = indirect tensile at high temperature; Num. = numerator.

Development of Preliminary Non-reheated Performance Criteria

Given the statistical difference in BMD test results between reheated and non-reheated conditions, preliminary non-reheated thresholds were developed for Cantabro, APA, and IDT-HT tests. Despite the marginal significance, the analysis was also conducted for APA to evaluate the feasibility of having a separate threshold for the non-reheated condition.

To estimate the non-reheated performance criterion for the aforementioned tests, a correction factor that quantifies the relationship between non-reheated and reheated test results was determined and further multiplied by the existing criterion for each test. This approach was selected considering that the performance criteria for Cantabro, APA, and IDT-HT tests were developed based on test results obtained from reheated PMLC mixtures. Three methods were evaluated to determine the correction factor given the scatter observed in the data. The result that yielded the least conservative value was selected as a preliminary non-reheated threshold to potentially facilitate implementation. A description of each correction factor calculation method is presented below.

- **Method 1:** *Slope of the linear regression curve.* The slope of the linear regression curve that fits the non-reheated versus reheated test results, as shown in Figure 2, was used to quantify the relationship between the two conditioning methods.
- **Method 2:** *Average ratio.* The ratio between the average non-reheated and the average reheated test result was calculated for each mixture evaluated in Figure 2. The average across all mixtures was then calculated.
- **Method 3:** *Median ratio.* The ratio between the average non-reheated and the average reheated test result was calculated for each mixture evaluated in Figure 2. The median across all mixtures was then calculated to address the potential skewness in the ratio values distribution.

Table 5 presents the correction factors calculated for each test using Methods 1 to 3, along with the estimated non-reheated threshold and the average value across all three methods. Overall, the estimated non-reheated values were similar among the three different methods. For CML, the average estimated non-reheated threshold was 4.80%, with values ranging from 4.71% to 4.83% (variation of 2.8% of the average). For APA, the average estimated non-reheated threshold was 7.97 mm, with values ranging from 7.5% to 8.4% (variation of 12.0% of the average). For IDT-HT, the average estimated non-reheated threshold was 86.7 kPa, with values ranging from 85.8 kPa to 87.9 kPa (variation of 2.4% of the average).

The preliminary non-reheated performance criteria proposed for the Cantabro, APA, and IDT-HT tests are maximum 5.0% mass loss, maximum 7.5 mm rut depth, and minimum 85 kPa strength, respectively. Note that, except for APA, these numbers correspond to the least conservative non-reheated threshold among the three methods evaluated and followed by rounding for potential ease of specification and implementation. Although the 8.4 mm would be the least conservative value for APA, proposing a non-reheated criterion with a higher rut depth limit than the reheated condition would be counterintuitive; therefore, the 7.5 mm value was selected instead.

Table 5. Preliminary Non-reheated Performance Criteria Proposed for Cantabro, APA, and IDT-HT Tests

Parameter	Method	Cantabro	APA	IDT-HT
Correction Factor	Method 1	0.6285	0.9347	0.8578
	Method 2	0.6465	1.0538	0.8790
	Method 3	0.6438	0.9989	0.8645
BMD Reheated Criterion	--	7.5% (Max.)	8.0 mm (Max.)	100 kPa (Min.)
Estimated Non-reheated BMD Test Result	Method 1	4.71%	7.5 mm	85.8 kPa
	Method 2	4.85%	8.4 mm	87.9 kPa
	Method 3	4.83%	8.0 mm	86.5 kPa
Preliminary Non-reheated Performance Criterion	--	5.0% (Max.)	7.5 mm (Max.)	85 kPa (Min.)

APA = Asphalt Pavement Analyzer; BMD = balanced mix design; IDT-HT = indirect tensile at high temperature; Max. = maximum; Min. = minimum.

The preliminary non-reheated thresholds for the Cantabro and IDT-HT tests are substantially different from the corresponding reheat ones. Nonetheless, these numbers are similar for the APA test (7.5 mm for non-reheat and 8.0 mm for reheat), reflecting the trends observed in Figure 2. Verification and risk assessment analyses were conducted to evaluate the feasibility of the Cantabro and IDT-HT non-reheated thresholds using VDOT’s statewide 2024 production data (Dataset 3). Given the lack of APA data in Dataset 3, these analyses were not conducted for APA test. More details of the verification and risk assessment analyses are presented next.

Verification Analysis of Non-reheated Performance Criteria

Figure 3 presents the non-reheated and reheated results for CML (Figure 3a) and IDT-HT tests (Figure 3b). Despite data scattering around the line of equality, the reheated CML and IDT-HT test results are greater than the corresponding non-reheated results, as indicated by the linear regression slopes (i.e., less than 1). This observation agrees with the results in Figure 2 for Dataset 1. Nonetheless, the difference in magnitude between non-reheated and reheated results in Figure 3 is not as exacerbated as in Figure 2. In Figure 3, the slopes for CML and high temperature test results are equal to 0.9383 and 0.9302, respectively. In Figure 2, the corresponding values are 0.6285 and 0.8578, suggesting a more pronounced distinction between non-reheated and reheated results. The smaller differences observed in Figure 3 can be attributed to the fact that the statewide data (Dataset 3) has a higher inherent variability because it was obtained from different districts, asphalt plants, operators, and test equipment. Dataset 1, on the other hand, includes data obtained in a more controlled environment, with all BMD tests conducted at VTRC.

Risk Assessment of Non-reheated Performance Criteria

Dataset 3 was also used to conduct a risk assessment analysis to evaluate the feasibility of the preliminary non-reheated performance criteria for CML and IDT-HT tests.

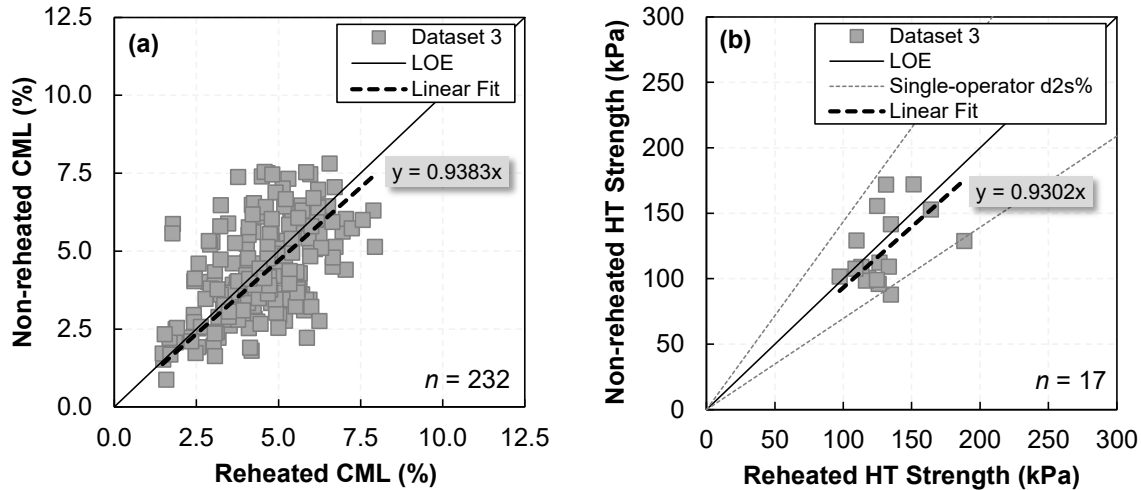


Figure 3. Verification of Non-reheated and Reheated Relationship for CML and HT Strength Results Using Dataset 3. CML = Cantabro mass loss; HT = high temperature; LOE = line of equality; n = number of data points.

Cantabro Test. Figure 4 presents the cumulative frequency of CML values for non-reheated (producers) and reheated samples (districts), along with the 5.0% preliminary non-reheated threshold and the current 7.5% threshold for Dataset 3. The 7.5% threshold results in high acceptance rates—99.3% for non-reheated and 98.3% for reheated samples—potentially raising concerns regarding their ability to screen out low-durability mixtures. A VTRC survey conducted as part of Project No. 126084, entitled *Virginia’s 2024 Implementation of Balanced Mix Design*, revealed that approximately 60% of producers and VDOT districts consider the test ineffective (Diefenderfer et al., 2026). However, these high acceptance rates may be attributed to two factors. First, BMD mixtures are designed to meet the 7.5% limit, so high acceptance during production is expected. Second, although the 7.5% threshold was developed based on reheated PMLC results, it is applied to both reheated and non-reheated data. Because non-reheated CML values are consistently less than the reheated ones, as shown with Dataset 1 (Figure 2) and verified with Dataset 3 (Figure 3), applying the same threshold may inflate acceptance rates and undermine the perceived test effectiveness, as seen with the 99.3% producer acceptance rate. Notably, when the preliminary 5.0% non-reheated criterion is applied instead, the acceptance rate reduces to 70%.

Additional analysis was conducted to identify potential correlations within the dataset and help explain the lower acceptance rate obtained (i.e., 70%) when applying the preliminary 5.0% criterion. The non-reheated CML data in Dataset 3 were combined with the corresponding non-reheated CT_{Index} data. A total of 19 pairwise data points from nine different JMFs and five VDOT districts were identified. Given the scatter in Dataset 3 and to help with trend visualization, the data points were clustered into three groups based on their CML values: CML < 2.5% (Group 1); 2.5% < CML ≤ 5.0% (Group 2); and CML > 5.0% (Group 3). Note that Group 3 aims to capture production sublots that would have failed the preliminary non-reheated CML threshold. The CML and CT_{Index} values were then averaged and presented in Figure 5 along with the corresponding non-reheated performance criteria. Despite the upcoming changes in the CT_{Index} criteria, the 95 threshold is shown because it corresponds to the specification in effect when these mixtures were produced.

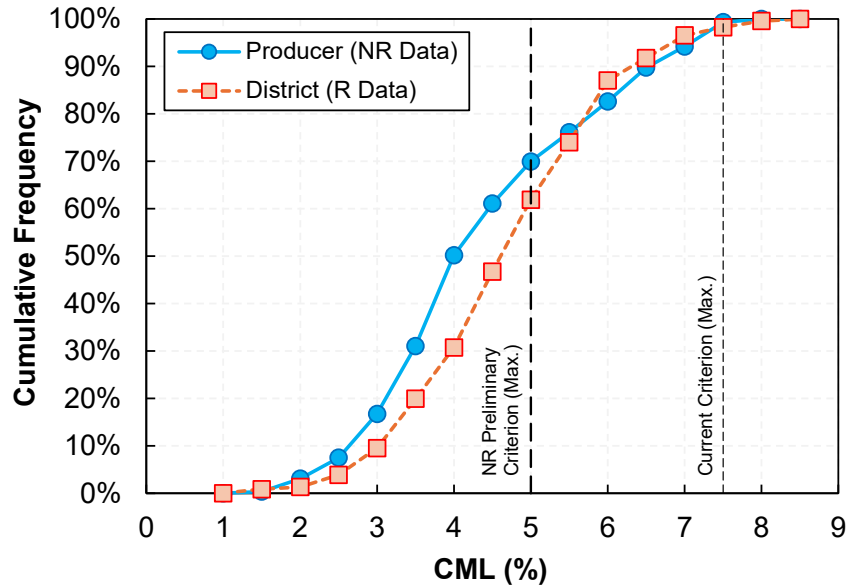


Figure 4. CML Cumulative Distribution for the Risk Assessment of CML Preliminary Non-reheated Performance Criterion Using Dataset 3. CML = Cantabro mass loss; Max. = maximum; NR = non-reheated; R = reheated.

Overall, Figure 5 shows that the performance ranking in terms of CML values is in agreement with CT_{Index} . In other words, cracking-resistant mixtures (higher CT_{Index}) are also the more durable ones (lower CML). Conversely, mixtures with lower cracking resistance (lower CT_{Index}) showed lower durability (higher CML). Notably, the data points that failed the preliminary 5.0% non-reheated CML criterion (Group 3) were also close to the non-reheated CT_{Index} threshold, confirming the equivalence between poor durability and marginal cracking performance. This trend becomes even more evident when the average CML and average CT_{Index} within each group is compared instead. Therefore, this analysis suggests that the CML results could be used as a cracking performance indicator during production to support the decision of requesting or waiving additional testing, especially when the BMD testing frequency is reduced. The analysis also suggests that the preliminary 5.0% non-reheated CML criterion may be a promising threshold in identifying poor-performing mixtures from non-reheated test results. It is worth noting that CML should not be used as a replacement for IDT-CT during production.

Given the negative correlation between CML and CT_{Index} in Figure 5, the upcoming changes in CT_{Index} performance criteria may have a detrimental effect on CML acceptance during production if no changes are made. For instance, increasing the CT_{Index} performance criteria (from 70/95 to 100/130 for reheated and non-reheated conditions, respectively) without changing the CML thresholds may lead to a less effective CML threshold, hindering its ability to screen out mixtures with low durability during production. In this sense, to confirm the negative correlation between CML and CT_{Index} , data from the optimum condition within Dataset 2 (i.e., design AC and gradation) were used for further analysis. These data were selected because they included data from laboratory-mixed, laboratory-compacted specimens tested under more controlled laboratory conditions (i.e., single operator and same test equipment) during the mix design stage, which can reduce data variability and minimize the effect of confounding factors.

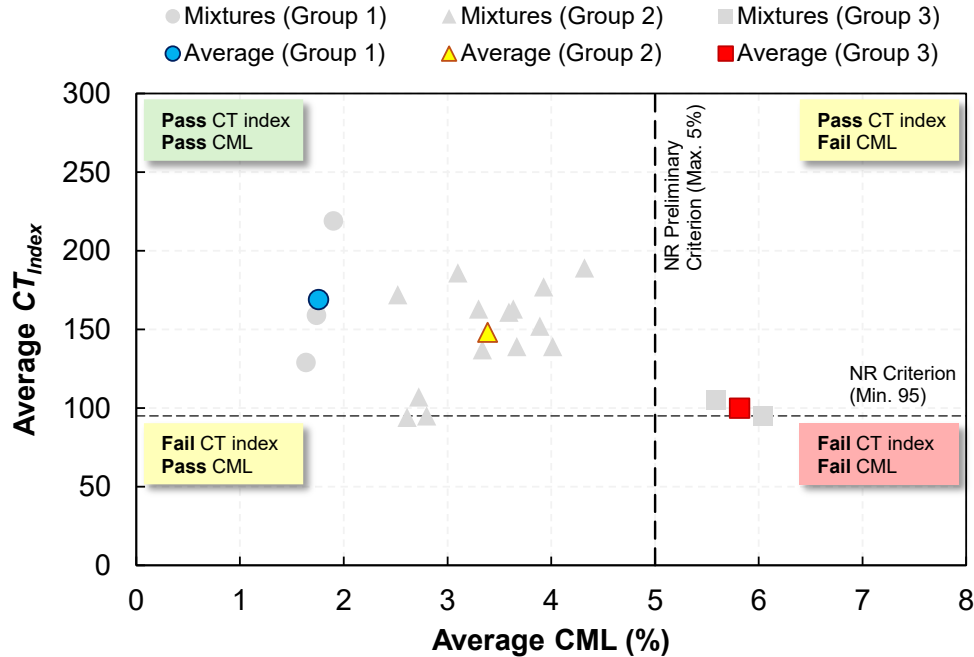


Figure 5. Relationship between Non-reheated CT_{Index} and Non-reheated CML for Production Samples in Dataset 3. CML = Cantabro mass loss; CT = cracking tolerance; Max. = maximum; Min. = minimum; NR = non-reheated. Groups 1, 2, and 3 include mixtures with CML < 2.5%, 2.5% < CML ≤ 5.0%, and CML > 5.0%, respectively.

Figure 6 shows a moderate to strong negative correlation ($R^2 = 0.61$) between CML and CT_{Index} following a non-linear power relationship. In fact, when the CT_{Index} reheated performance criterion of 70 is used as input into the power function, a corresponding CML of 7.6% is obtained, which is similar to the current CML threshold of 7.5%. The non-linear power function also suggests that mixtures failing the 7.5% mass loss criterion are also the ones failing the minimum CT_{Index} of 70. These observations collectively confirm the negative correlation observed in Figure 5 and highlight the agreement between IDT-CT and Cantabro performance criteria, suggesting that changes in CT_{Index} thresholds should be paired with corresponding modifications in CML performance criteria. However, the effect on mixture performance from the upcoming changes in the BMD specification starting in the 2026 construction season remains unknown. These changes include increasing CT_{Index} criteria and implementing the BMD specimen preparation VTM. Because of these changes, a phased approach, including continued monitoring of the asphalt mixtures' performance during production starting in 2026, should be conducted prior to any adjustments to CML thresholds. This approach will ensure a successful and responsible BMD implementation in Virginia.

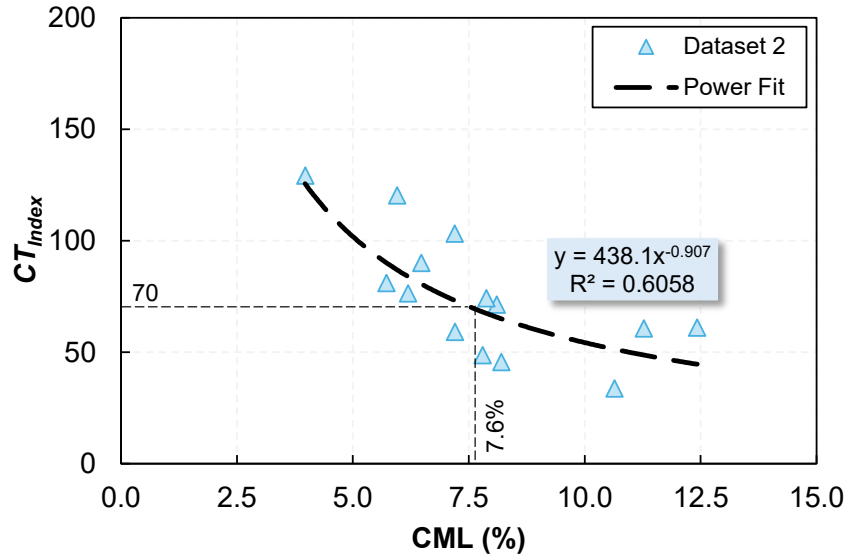


Figure 6. Correlation between CT_{Index} and CML Using Data from Dataset 2 Obtained at the Design Asphalt Content and Gradation. CML = Cantabro mass loss; CT = cracking tolerance.

Indirect Tensile at High Temperature Test. Figure 7 presents the cumulative frequency of non-reheated IDT-HT test results for producers and VDOT districts, along with the 85 kPa minimum strength (preliminary non-reheated threshold) and 100 kPa minimum strength (current threshold) for Dataset 3. Although unusual, some districts reported conducting non-reheated IDT-HT tests for a select number of JMFs. Notably, the IDT-HT performance criterion was originally developed from reheated PMLC mixtures test results under dry conditioning. This criterion was further revised for specimens subjected to wet conditioning (Boz et al., 2025a). Applying the current 100 kPa threshold to the non-reheated data in Dataset 3 results in 12.5% and 29.7% of samples failing for producers and districts, respectively (Figure 7). Notably, the difference between producer and district rejection rates may be because the districts' cumulative frequency plot includes data from only two districts. This reduced number may have contributed to shifting the districts' cumulative frequency distribution compared with the producers' plot, which includes data from five districts. Nonetheless, when the preliminary non-reheated performance criterion of 85 kPa is used instead, the failing rate is reduced to 4.7% and 5.4% for producers and districts, respectively. Additional analysis was conducted to help identify the reason behind the failing samples and assess the efficacy of the high temperature, non-reheated performance criterion in screening out mixtures with poor rutting resistance.

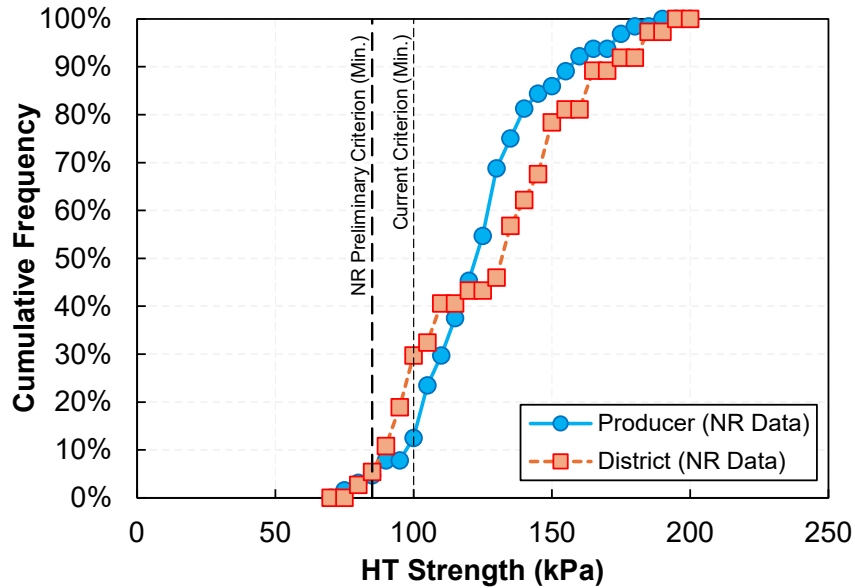


Figure 7. HT Strength Cumulative Distribution for the Risk Assessment of HT Strength Preliminary Non-Reheated Performance Criterion Using Dataset 3. HT = high temperature; Min. = minimum; NR = non-reheated.

Figure 8 presents the total AC versus RAP content for the individual production sublots subjected to the non-reheated IDT-HT test. Because asphalt mixtures with higher rutting susceptibility are often associated with finer gradations and lower binder stiffness (Veeraragavan et al., 2022, Mivehchi et al., 2023; Rocha et al., 2024), total AC and RAP content were used to indirectly infer about the mixtures' binder stiffness. The binder stiffness refers to the stiffness of the binder blend (virgin binder and RAP binder), which relates to the recycled binder ratio (RBR) of the mix—that is, the ratio between the RAP binder content and the total AC. In mixtures with higher RBR, the recycled AC is higher relative to the total AC, increasing the binder blend stiffness compared with a mix with lower RBR. Because Dataset 3 does not include information regarding RAP binder content, RBR was qualitatively evaluated as a function of total AC and RAP content. In other words, it was assumed that higher total AC and lower RAP content result in low RBR values or lower binder stiffness, whereas lower total AC and higher RAP content result in high RBR values or higher binder stiffness.

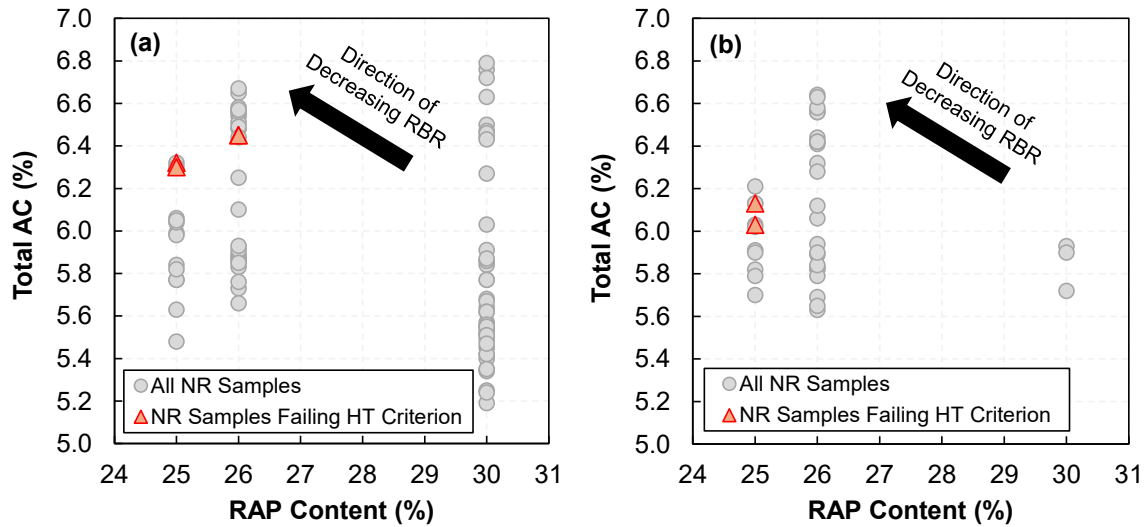


Figure 8. Evaluating Recycled Binder Ratio of Production Non-reheated Samples in Dataset 3 in Terms of Total AC and RAP Content for (a) Producers and (b) Districts. AC = asphalt content; HT = high temperature; NR = non-reheated; RAP = reclaimed asphalt pavement; RBR = recycled binder ratio.

In Figure 8, it was observed that mixtures that failed the 85 kPa criterion (red triangles) were also the ones with lower RBR values, depicted by higher total AC values and/or lower RAP content. If the 85 kPa criterion is proven to be effective in identifying mixtures with low binder stiffness tested under non-reheated conditions, this observation would confirm findings from NCHRP IDEA Project 245, which concluded that the IDT-HT test is sensitive to changes in binder stiffness and correlates with the asphalt mixture extracted binder PG (Boz et al., 2025b). Despite the apparent correspondence between samples that failed the 85 kPa criterion and the low binder stiffness depicted by lower RBR, it is worth noting that only a limited number of samples failed the 85 kPa criterion, suggesting that further analysis is warranted to confirm the findings. In terms of gradation, no clear trend was observed—mixtures that failed the high temperature, non-reheated criterion of 85 kPa were fine-graded on the producer side (Figure 8a) but coarse-graded in the district side (Figure 8b). Similar to the CML analysis, correlations with corresponding CT_{Index} non-reheated data were explored, but no strong relationship was observed.

Given the limited number of mixtures tested, additional PMLC mixtures should be evaluated to verify the relationship between non-reheated and reheated high temperature strength values.

Effect of Asphalt Mixture Composition on Balanced Mix Design Test Results

Effect of Volumetrics and Gradation

Table 6 summarizes the r values for each plot, and Figures 9 to 12 present the respective linear regression fit curves for Cantabro, IDT-CT, APA, and IDT-HT results. Although asphalt binder quality is a significant factor influencing asphalt mixture laboratory performance, it was not included as a variable because the analysis was focused on parameters typically measured during production. Including asphalt binder rheological parameters as part of this analysis may change the significance of the factors in Table 6.

Table 6. Pearson Correlation Coefficient Relating Cantabro, IDT-CT, APA, and IDT-HT Results with Volumetrics and Gradation Parameters

Parameter	Pearson Correlation Coefficient (r)			
	Cantabro	IDT-CT	APA	IDT-HT ^a
Asphalt Content (AC)	-0.54	0.85	0.70	-0.46
Percent Passing at Sieve No. 4	-0.38	0.02	0.24	0.34
Percent Passing at Sieve No. 8	-0.57	0.15	0.50	0.46
Percent Passing at Sieve No. 30	-0.57	0.25	0.63	0.39
Percent Passing at Sieve No. 200	-0.07	-0.18	0.03	-0.09
Maximum Specific Gravity (G_{mm})	0.31	-0.69	-0.59	0.73
Bulk Specific Gravity (G_{mb})	0.03	-0.54	-0.30	0.59
Aggregate Bulk Specific Gravity (G_{sb})	0.21	-0.58	-0.49	0.70
Aggregate Effective Specific Gravity (G_{se})	0.27	-0.64	-0.55	0.66
Voids in Total Mixture (VTM)	0.67	-0.29	-0.67	0.26
Voids in the Mineral Aggregate (VMA)	0.28	0.23	-0.17	0.14
Voids Filled with Asphalt (VFA)	-0.70	0.36	0.73	-0.26
Absorbed Asphalt Content (P_{ba})	0.40	-0.46	-0.60	-0.45
Effective Asphalt Content (P_{be})	-0.55	0.85	0.75	-0.38
Fines-to-Asphalt Ratio (FA)	0.30	-0.65	-0.44	0.17

APA = Asphalt Pavement Analyzer; IDT-CT = indirect tensile cracking test; IDT-HT = indirect tensile at high temperature. ^aIDT-HT analysis was conducted with data from VTRC Project No. 23-R11. Cantabro, IDT-CT, and APA analyses were conducted using Dataset 2. Cells highlighted in dark blue represent strong correlation with $|r| \geq 0.7$. Cells highlighted in light blue represent moderate correlation with $0.5 \leq |r| < 0.7$. Cells with no highlight (white) represent weak correlation with $|r| < 0.5$.

For Cantabro (Figure 9), VFA was the only factor strongly correlated with CML ($r = -0.70$), suggesting that higher AC filling the voids of the mix tends to reduce the mass loss during the test, given the improved asphalt mixture cohesion. VTM showed a moderate positive correlation ($r = 0.67$), suggesting that samples with higher air void content, and potentially lower cohesion, exhibit higher abrasion rates during Cantabro test. Moderate negative correlations were observed with gradation data, specifically with percent passing at the No. 8 ($r = -0.57$) and No. 30 ($r = -0.57$) sieves. Moderate negative correlations were also found between AC ($r = -0.54$) and P_{be} ($r = -0.54$). These correlations suggest that finer gradations and additional asphalt may improve the asphalt mixture cohesion, thus reducing its abrasion potential during the test. Cox et al. (2017) reported similar trends between volumetrics and gradation properties and CML.

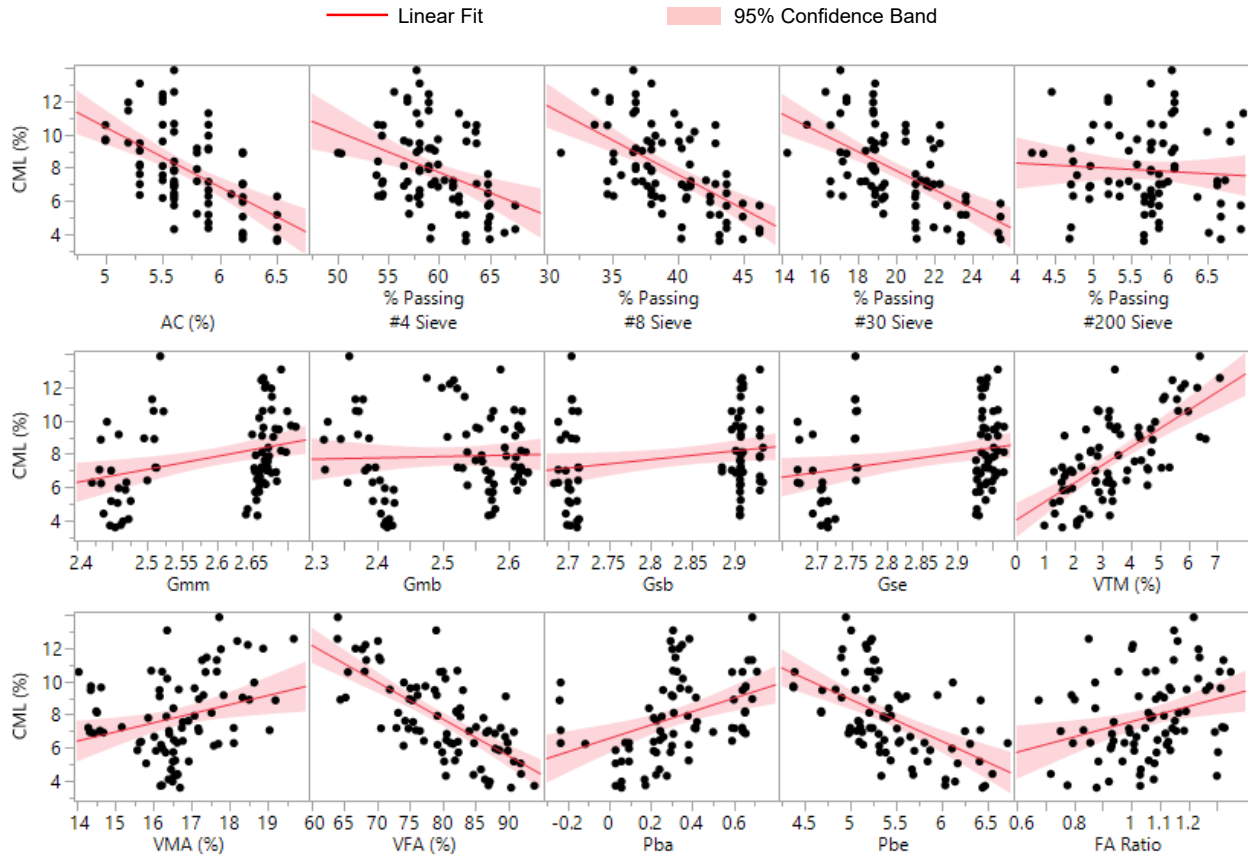


Figure 9. Correlation between CML and Individual Volumetrics and Gradation Parameters. AC = asphalt content; CML = Cantabro mass loss; FA = fines-to-asphalt; G_{mb} = bulk specific gravity; G_{mm} = maximum specific gravity; G_{sb} = aggregate bulk specific gravity; G_{se} = aggregate effective specific gravity; P_{ba} = absorbed asphalt content; P_{be} = effective asphalt content; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture.

For IDT-CT (Figure 10), a strong correlation was observed between CT_{Index} and AC ($r = 0.85$) and P_{be} ($r = 0.85$), as expected. Figure 10 shows two clusters of data between CT_{Index} and specific gravity values, which highlights the differences in aggregate source locations. Despite the clustering, moderate negative correlations were observed, with r values ranging from -0.54 for G_{mb} to -0.69 for G_{mm} . For gradation parameters, weak correlations were found, with r values ranging from -0.18 to 0.25 , suggesting only a marginal association. Nevertheless, it is known that CT_{Index} is affected by changes in more than one volumetric factor rather than an individual parameter.

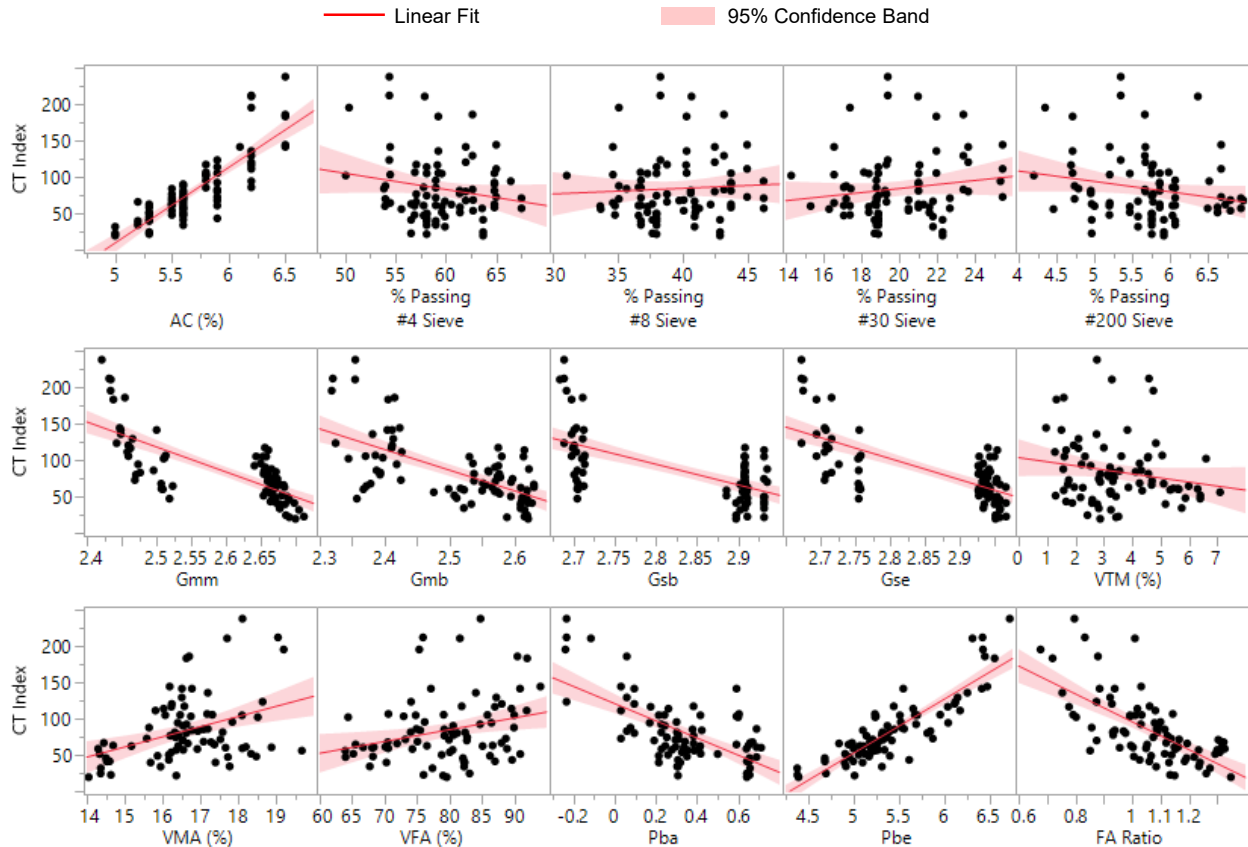


Figure 10. Correlation between CT_{Index} and Individual Volumetrics and Gradation Parameters. AC = asphalt content; CT = cracking tolerance; FA = fines-to-asphalt; G_{mb} = bulk specific gravity; G_{mm} = maximum specific gravity; G_{sb} = aggregate bulk specific gravity; G_{se} = aggregate effective specific gravity; P_{ba} = absorbed asphalt content; P_{be} = effective asphalt content; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture.

For APA (Figure 11), a strong positive correlation was found between rut depth and AC ($r = 0.70$), VFA ($r = 0.73$), and P_{be} ($r = 0.75$). This observation confirms the expected performance—rutting resistance decreases as AC increases if all other parameters are kept constant. Moderate positive correlations were observed with percent passing at sieves No. 8 ($r = 0.50$) and No. 30 ($r = 0.63$), suggesting that mixtures with finer gradations may be more prone to rutting. Specific gravity values showed moderate correlations between APA rut depth and G_{mm} ($r = -0.59$) and G_{se} ($r = -0.55$). VTM and P_{ba} were also found to be negatively correlated with APA rut depth, with r values of -0.67 and -0.60 , respectively. Increasing either VTM or P_{ba} reduces P_{be} leading to an increase in asphalt mixture rutting resistance if all other parameters are kept constant.

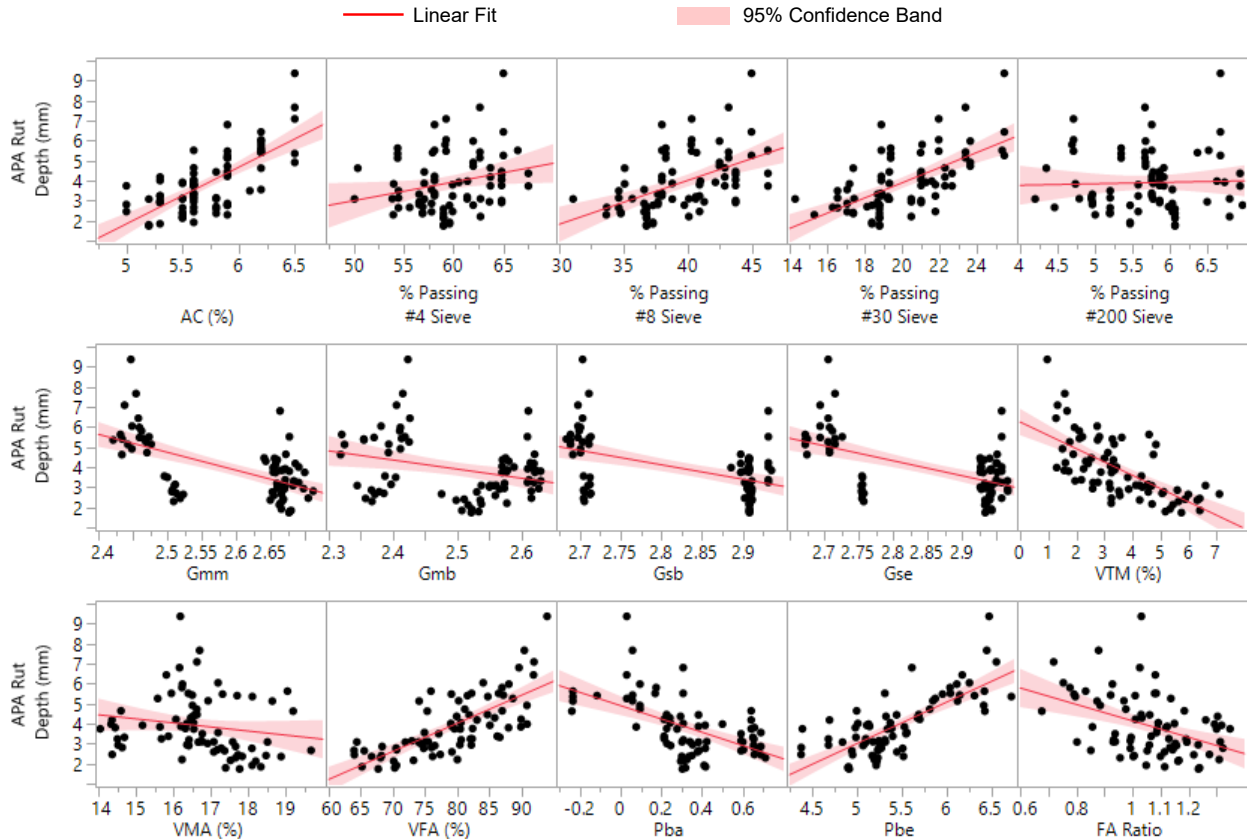


Figure 11. Correlation between APA Rut Depth and Individual Volumetrics and Gradation Parameters. AC = asphalt content; APA = Asphalt Pavement Analyzer; FA = fines-to-asphalt; G_{mb} = bulk specific gravity; G_{mm} = maximum specific gravity; G_{sb} = aggregate bulk specific gravity; G_{se} = aggregate effective specific gravity; P_{ba} = absorbed asphalt content; P_{be} = effective asphalt content; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture.

For IDT-HT (Figure 12), moderate to strong positive correlations were observed between strength and the specific gravities (i.e., G_{mm} , G_{mb} , G_{sb} , and G_{se}), with r values ranging from 0.59 to 0.73. Although a weak correlation was found with AC ($r = -0.46$), the negative trend indicates worse rutting resistance for increasing AC levels, as expected and in agreement with APA test analysis. In addition, weak correlations were observed between strength and gradation parameters.

Although Figure 9 to Figure 12 show the individual effect of volumetrics and gradation parameters on BMD test results, it is known that these parameters have a combined effect due to correlation among them. To truly assess their actual significance, these factors should not be evaluated individually. Instead, the collinearity among parameters needs to be evaluated and taken into consideration prior to conducting the statistical analysis. The collinearity analysis is presented in further details in the next section.

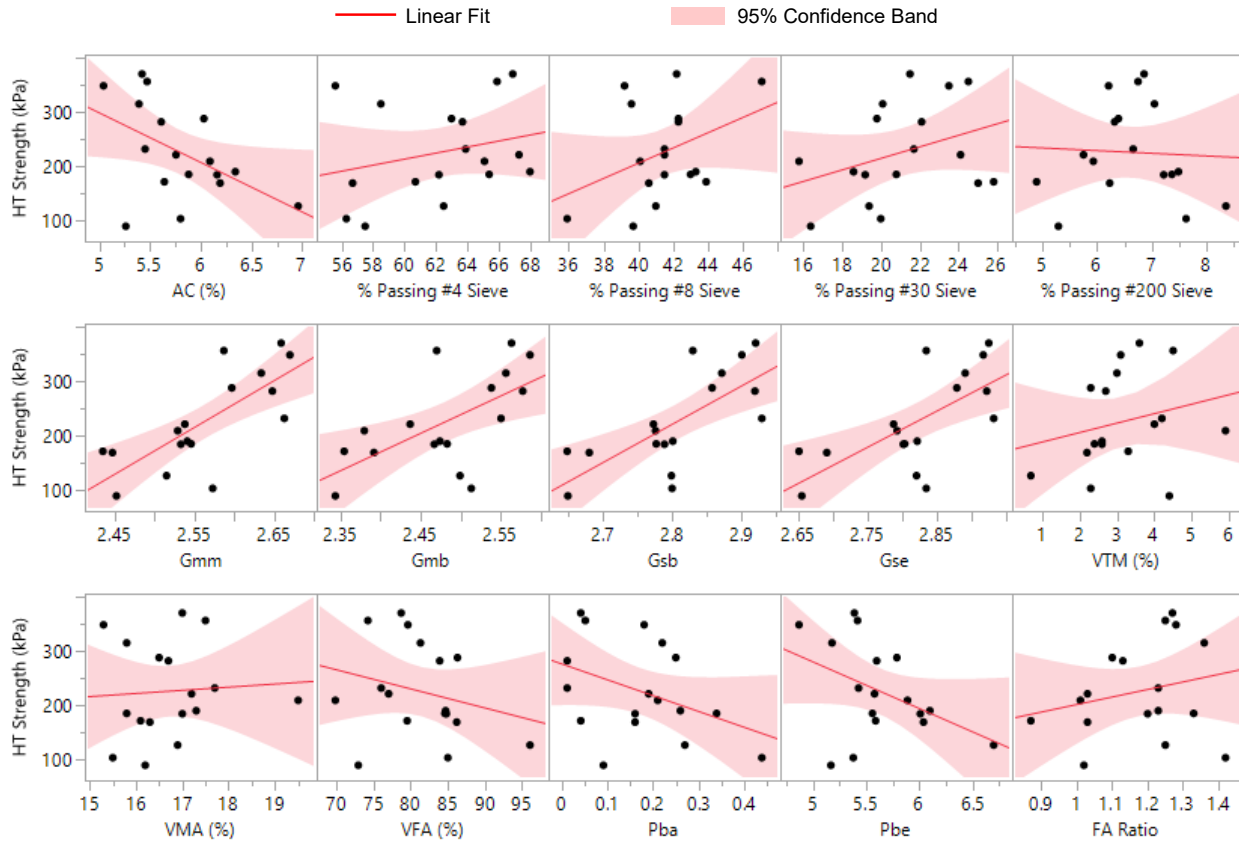


Figure 12. Correlation between HT Strength and Individual Volumetrics and Gradation Parameters. AC = asphalt content; FA = fines-to-asphalt; HT = high temperature; G_{mb} = bulk specific gravity; G_{mm} = maximum specific gravity; G_{sb} = aggregate bulk specific gravity; G_{se} = aggregate effective specific gravity; P_{ba} = absorbed asphalt content; P_{be} = effective asphalt content; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture.

Effect of Binder-related Properties

Asphalt binder plays a key role in asphalt mixture performance, with two factors being the most relevant: the quantity and the quality of the asphalt binder. The quantity is a function of the total AC and P_{be} (i.e., VBE). The quality depends on the virgin binder grade and source, as well as the properties of the recycled binder and its interaction with RAs and other additives. It is worth noting that a similar binder PG does not warrant similar performance because of differences in the virgin binder crude source or refining processes. Nonetheless, because no data were available for the extracted and recovered asphalt binder from mixtures in Dataset 2, information regarding RAP content, virgin binder PG, and RA was used in the analysis presented herein as an indirect measure of binder quality. In this section, the effect of binder quantity (e.g., VBE) and binder quality (e.g., RAP content, virgin binder PG, and RA) on BMD tests was assessed.

Volume of Effective Binder

The effect of asphalt binder quantity on BMD results was quantified using the VBE parameter. VBE corresponds to the volume of asphalt binder not absorbed by the aggregates (i.e.,

effective) and calculated as the difference between VMA and VTM. Higher VBE is often associated with improved cracking and durability and can be obtained by increasing VMA (e.g., coarser mixtures) or decreasing VTM (e.g., increasing AC and increasing number of design gyrations). On the other hand, lower VBE is often associated with improved rutting resistance and can be obtained by decreasing VMA (e.g., finer mixtures) or increasing VTM (e.g., reducing AC and reducing the number of design gyrations). The relationship between VMA, VTM, and VBE is illustrated in the interaction diagrams in Figure 13 for both design and production volumetric requirements.

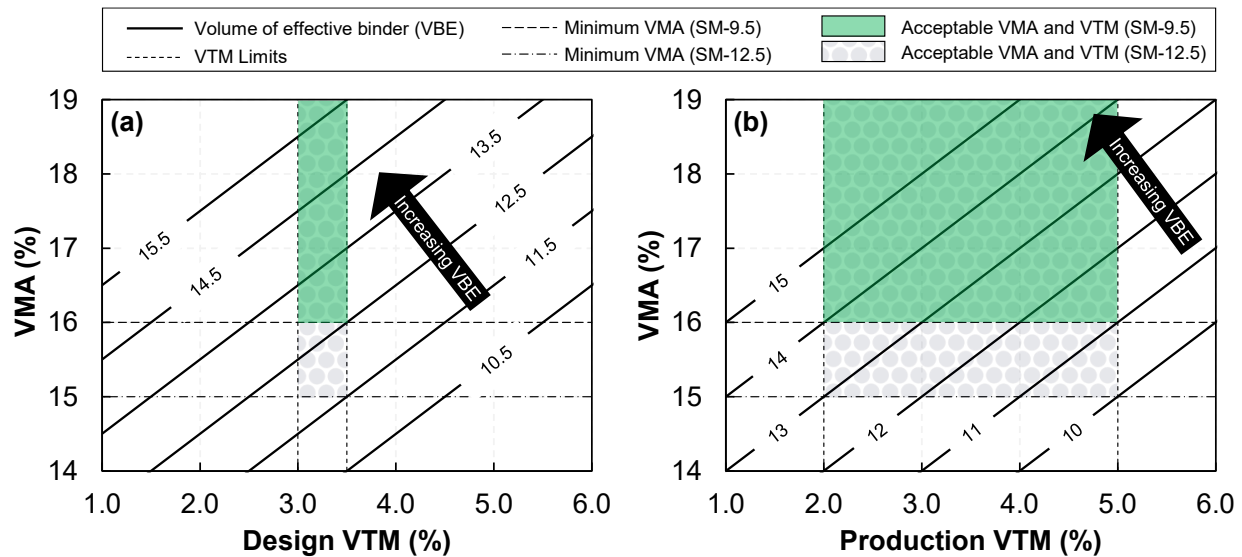


Figure 13. Interaction Diagrams for VBE for (a) VMA and Design VTM and (b) VMA and Production VTM based on VDOT’s Volumetrics Requirements. SM = surface mixture; VBE = volume of effective binder; VMA = voids in mineral aggregate; VTM = voids in total mixture.

Figure 13 highlights the areas corresponding to acceptable VMA and VTM for both SM-9.5 (solid green) and SM-12.5 mixtures (gray-circle pattern), along with the VBE isochrones. For design, the lowest VBE values that yield acceptable VMA and VTM are 12.5% and 11.5% for SM-9.5 and SM-12.5, respectively. For production, the corresponding VBE values are 11% and 10% for SM-9.5 and SM-12.5, respectively. Despite the determination of minimum VBE values, establishing minimum VBE limits would not necessarily ensure compliance with both VMA and VTM requirements. The correlation between VBE and BMD test results was explored mainly to verify its potential as a tool for mix adjustments during design or production.

A subset of Dataset 2 was used to explore the potential relationship between Cantabro, IDT-CT, and APA results with VBE. The subset comprised only mixtures with 12.5-mm NMAS at three AC: 5.6%, 5.9%, and 6.2%. Although different virgin binder PG were included, they are not expected to affect the volumetric properties of asphalt mixtures.

Figure 14 shows a trend between BMD tests and VBE. At a constant AC, mixtures with higher VBE showed improved cracking resistance and durability but lower rutting resistance across all three AC levels. On average, an increase of 1% in VBE corresponded to a decrease of 2% in CML, an increase of 16 units in CT_{Index} , and an increase in rut depth of 1 mm. This observation indicates that mixtures with the same AC may perform differently because of

variations in gradation, aggregate absorption, and other volumetric parameters influencing VMA, VTM, and, ultimately, VBE. Adjusting gradation and other volumetric parameters at a fixed AC to modify VBE can help meet BMD performance requirements. Figure 14 also shows that the effect of VBE in BMD test results seems to be more pronounced at higher AC for both CT_{Index} and APA rut depth, and less pronounced for CML. Since AC drives the volumetric balance in the mix design, the closer the mix is to instability due to potential excess of asphalt, the greater is the AC effect, suggesting that increasing VBE values at higher AC levels may have a greater effect on cracking and rutting resistances than at lower AC levels.

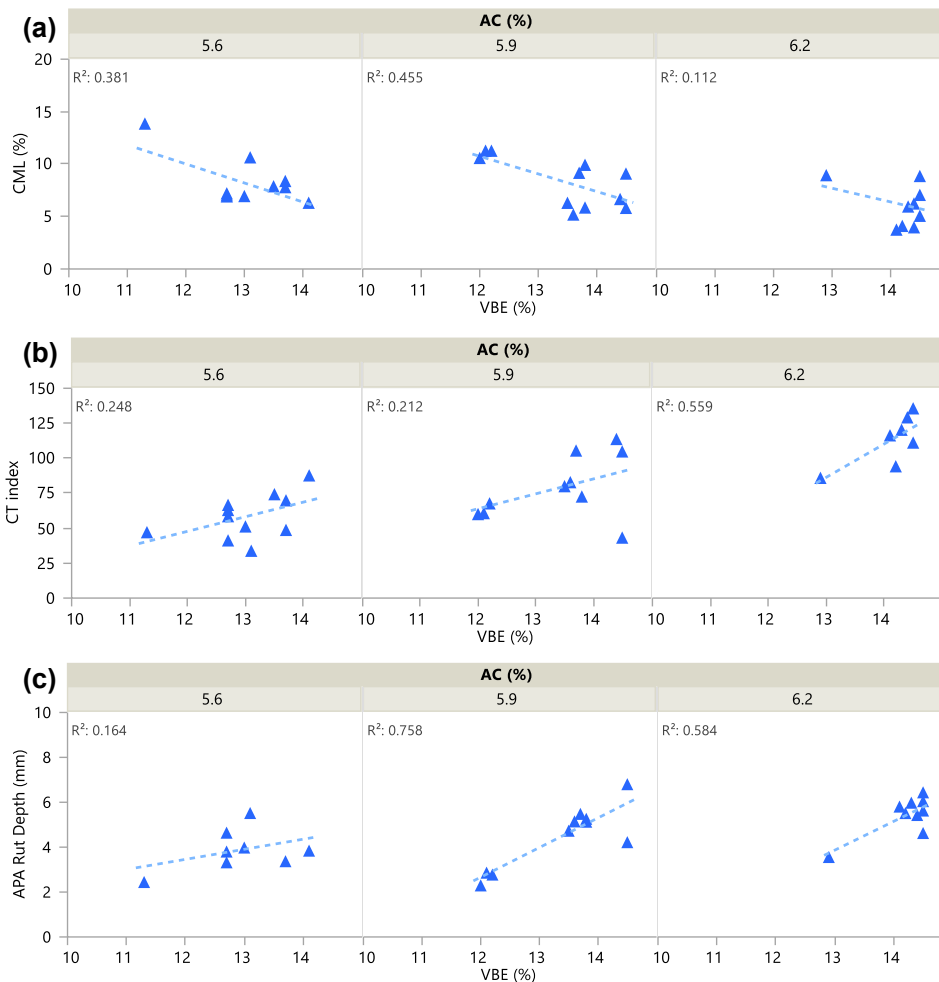


Figure 14. Relationship between Balanced Mix Design Test Results, AC, and Effective Volume of Binder for (a) Cantabro, (b) Indirect Tensile Cracking Test, and (c) APA Results. AC = asphalt content; APA = Asphalt Pavement Analyzer; CML = Cantabro mass loss; CT = cracking tolerance; VBE = volume of effective binder. Dashed blue line corresponds to the linear regression fit.

Overall, the findings presented herein are in accordance with other studies in the literature, which also reported a strong correlation between VBE and asphalt mixture performance (Christensen and Bonaquist, 2006; Bonaquist, 2014; Cox et al., 2017; Yin et al., 2023, Rivera-Perez et al., 2023; Bennert; 2026). Although establishing VBE limits would not necessarily guarantee compliance with both VMA and VTM requirements, it could be used as an

indicator to help guide mix adjustments during design or production, or both, when trying to meet BMD performance criteria.

RAP Content, Virgin Binder Performance Grade, and Recycling Agent

The effect of asphalt binder quality on BMD results was indirectly quantified using RAP content, virgin binder PG, and RA data in Dataset 2. Recall that these parameters were selected because Dataset 2 did not include actual binder properties and other rheological parameters from the extracted and recovered binder that are typically used to correlate with asphalt mixture performance. In this sense, RAP content, virgin binder PG, and the use of RAs were adopted as categorical factors to indirectly infer about the binder quality and further explore their correlation with mixture performance. Figure 15 summarizes the relationship between Cantabro, IDT-CT, and APA results with AC for different asphalt plants and clustered in five groups: (1) 30% RAP and PG 64S-22; (2) 35% RAP and PG 58-28; (3) 35% RAP, PG 58-28, and RA; (4) 40% RAP and PG 58-28; and (5) 40% RAP, PG 64S-22, and RA. Note that the 30% RAP and PG 64S-22 variation corresponds to the typical BMD asphalt surface mixture produced in Virginia.

Figure 15 shows a similar trend between BMD test results and AC for all plants—higher AC resulted in improved durability and cracking resistance but reduced rutting resistance. Similar trends were observed between BMD test results and VBE but were not included for brevity reasons. The 35% RAP and PG 58-28 (Plant B) and 35% RAP, PG 58-28, and RA (Plant E) mixtures showed the lowest CML values and highest CT_{Index} results, with moderate to high APA rut depths compared with other mixtures. This observation suggests that the moderate RAP content combined with a softer binder grade and RA reduces the overall binder stiffness, improving both durability and cracking resistance estimated by Cantabro and IDT-CT, respectively. On the other hand, the lower binder stiffness yielded higher rut depths and, therefore, resulted in more rut-susceptible mixtures.

For Plants A and C, no substantial differences in terms of durability, cracking, and rutting resistance were observed between the 30% RAP and PG 64S-22, 40% RAP and PG 58-28, and 40% RAP, PG 64S-22, and RA mixtures. For Plant D, these differences were more pronounced, with the 40% RAP and PG 58-28 mixture showing better durability, cracking resistance, and similar rutting resistance than the 30% RAP and PG 64S-22. Different trends were observed for Plant F for the same mixtures. In fact, the 40% RAP and PG 58-28 mixture exhibited worse durability and rutting resistance, but comparable cracking performance compared with the 30% RAP and PG 64S-22 mix. Moreover, the 40% RAP, PG 64S-22, and RA mixture showed similar performance compared with the 30% RAP and PG 64S-22 mix.

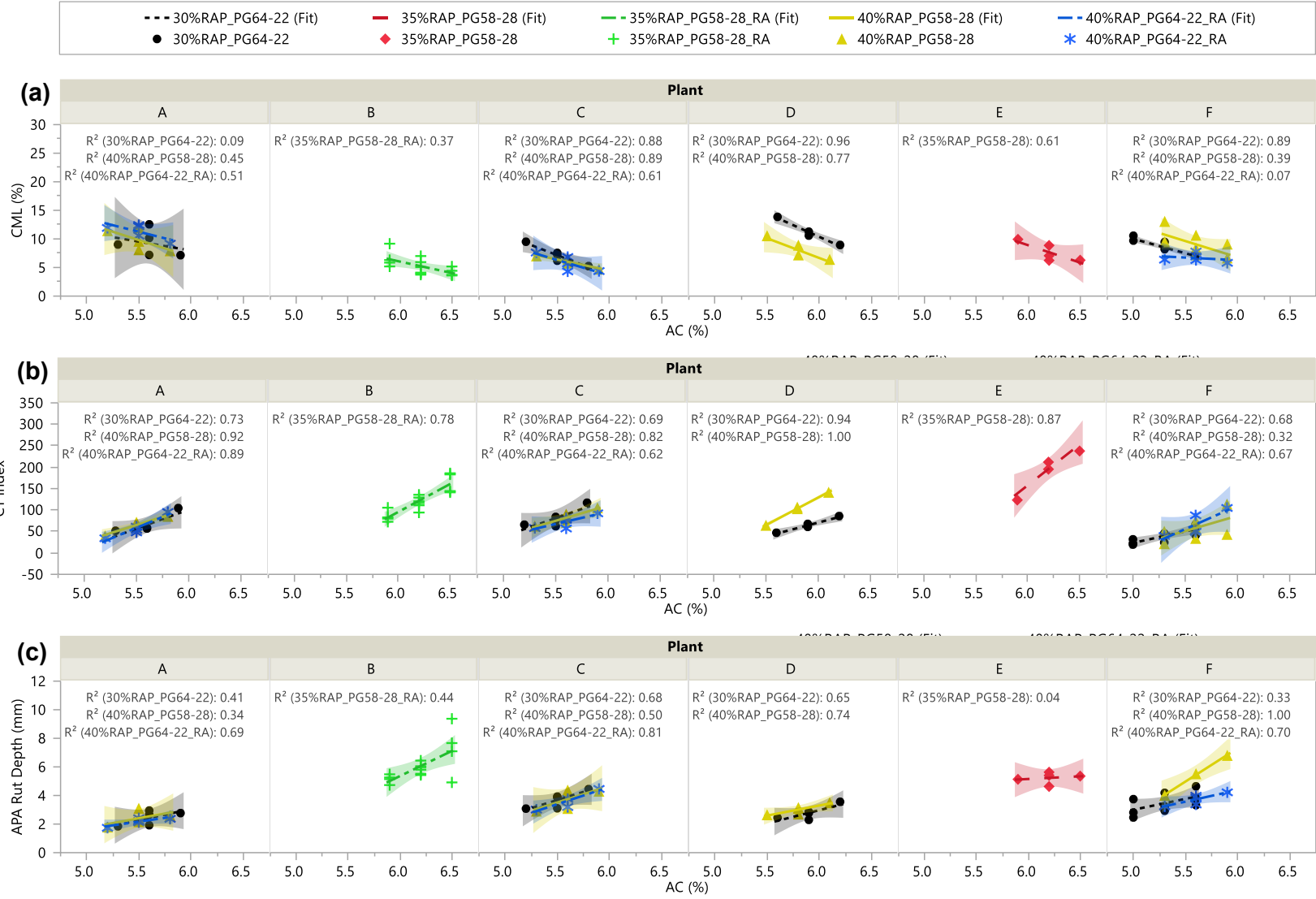


Figure 15. Relationship between RAP Content and Virgin Binder PG on (a) Cantabro, (b) Indirect Tensile Cracking Test, and (c) APA Test Results and Grouped by Asphalt Plants. AC = asphalt content; APA = Asphalt Pavement Analyzer; CML = Cantabro mass loss; CT = cracking tolerance; PG = performance grade; RA = recycling agent; RAP = reclaimed asphalt pavement. Shaded areas correspond to the 95% confidence band.

Collectively, these observations highlight that RAP content, virgin binder PG, and RA use do not indicate binder quality, given the different outcomes obtained when comparing the durability, cracking, and rutting performance of these mixtures. Therefore, the need to explore rheological properties of the extracted and recovered asphalt binder is evident, especially when trying to correlate asphalt binder quality and asphalt mixture performance.

Identifying Significant Factors

The analysis presented in the previous sections provided a quantitative evaluation of the effect of different volumetrics, gradation, and binder-related properties on BMD test results. In this section, additional analysis was conducted to identify which volumetrics and gradation parameters significantly affect BMD results. The analysis focused on parameters measured during production and used for mixture acceptance and quality control practices in Virginia. The intent was to identify which parameters should be controlled more carefully during production to minimize the potential oscillation of BMD test results throughout different production lots. Despite the well-known effect of binder properties on BMD test results, they were not included in this analysis because they are not measured during production.

Prior to statistical analysis, the collinearity among volumetrics and gradation parameters was verified. Not accounting for collinearity may hinder the potential significance of other independent variables that would otherwise be deemed significant if collinearity were considered. A pairwise correlation analysis using the volumetrics and gradation parameters obtained from Dataset 2 was conducted and is summarized in Figure 16. Although the analysis for IDT-HT relied on data different from Dataset 2, similar observations regarding the volumetrics and gradation parameters pairwise correlation evaluation were found but are not presented herein for brevity.

Figure 16 highlights pairwise combinations with strong ($|r| \geq 0.7$), moderate ($0.5 \leq |r| < 0.7$), and weak ($|r| < 0.5$) correlations. The r values can be found in the lower triangle of the correlation matrix, and the correlation plots with a linear regression fit are presented at the upper triangle. Figure 16 shows that gradation parameters are strongly correlated with one another, especially for subsequent sieve sizes. The percent passing at sieve No. 4 (PP#4) has a high r value of 0.91 with respect to the percent passing at sieve No. 8 (PP#8), which, in turn, has also a high r value of 0.92 with respect to the percent passing at sieve No. 30 (PP#30). These strong correlations are expected because of the dense-graded characteristic of the mixtures included in Dataset 2 and highlight that changes in gradation at a given sieve size are likely to affect the subsequent sizes. This observation suggests that including multiple gradation data from subsequent sieve sizes in the statistical analysis can lead to collinearity effects.

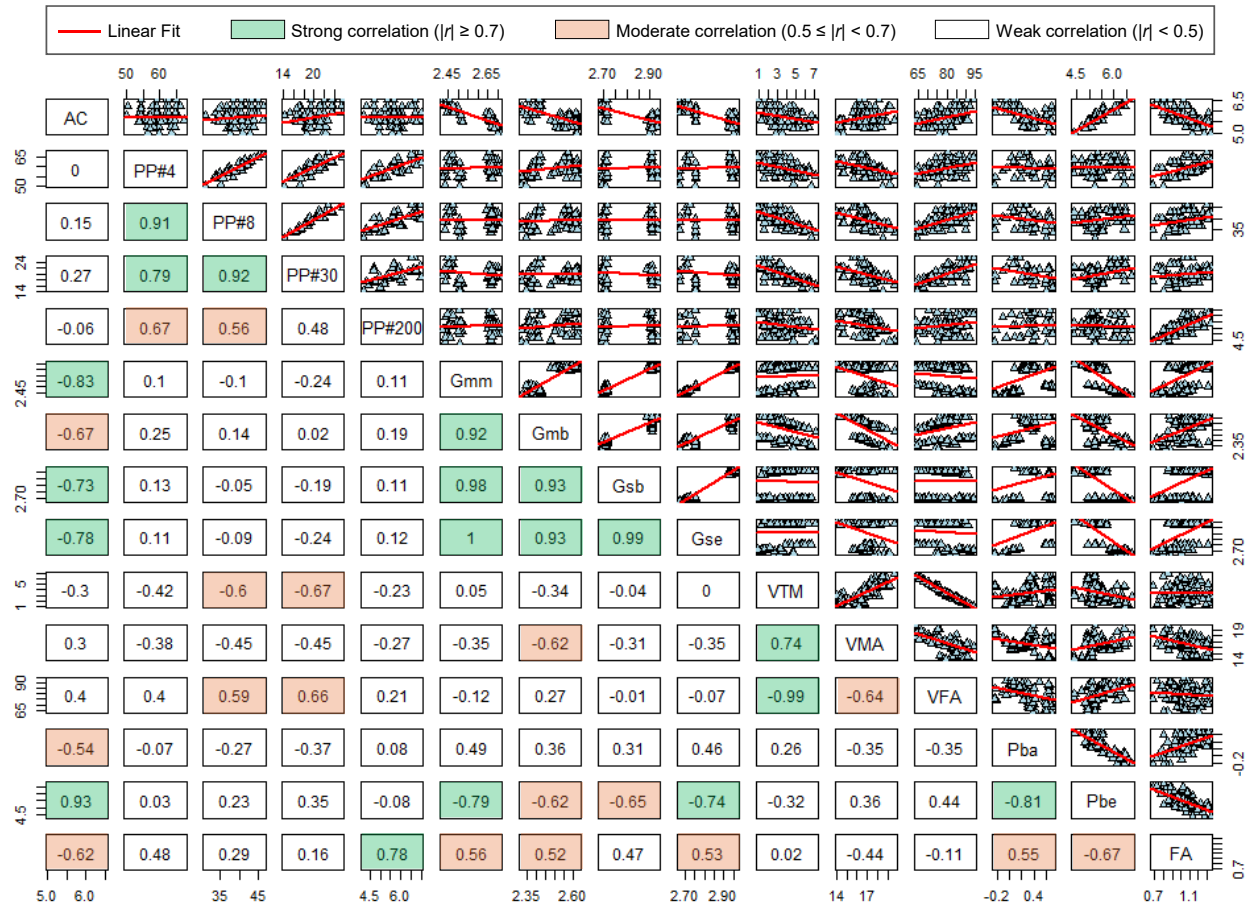


Figure 16. Summary of r Values for the Pairwise Correlation Analysis. AC = asphalt content; FA = fines-to-asphalt; G_{mb} = mixture bulk specific gravity; G_{mm} = mixture maximum specific gravity; G_{sb} = aggregate bulk specific gravity; G_{se} = aggregate effective specific gravity; P_{ba} = absorbed AC; P_{be} = effective AC; PP = percent passing; VFA = voids filled with asphalt; VMA = voids in the mineral aggregate; VTM = voids in total mixture.

Strong correlations were also observed among the specific gravity measurements, including G_{mm} , G_{mb} , G_{sb} , and G_{se} , with r values ranging from 0.92 to 1.00 among the pairwise combinations. These correlations are associated with the fact that specific gravity measurements are often used as inputs to calculate other specific gravity values. For instance, G_{se} measurements are dependent on the mix G_{mm} , and G_{sb} measurements are typically estimated from G_{se} values after a correction factor deduction under production conditions. Similar observations can be drawn from other relationships, including VTM, VMA, and VFA values that showed absolute r values ranging from 0.64 to 0.99.

To mitigate collinearity, the factor retained among two factors that showed moderate to strong correlation was the one exhibiting the stronger association with the corresponding BMD test result. After removing the collinear factors, a forward stepwise multiple linear regression approach was adopted to identify the significant volumetric and gradation factors that affect the different BMD test results. The p-value of 0.05 was selected as the stopping criterion for factors to both enter or leave the model when conducting the stepwise regression. Table 7 summarizes the significant factors for each BMD test, including both their estimate (or regression coefficients) and corresponding p-values.

Table 7. Summary of Forward Stepwise Multiple Linear Regression Analysis

Factor	CML		CT_{Index}		APA Rut Depth		IDT-HT Strength	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Intercept	27.27	1.0000	-295.32	1.0000	-7.00	1.0000	-1701.98	1.0000
Asphalt Content	-2.50	< 0.0001	75.21	< 0.0001	1.77	< 0.0001	-59.77	0.0402
PP at Sieve No. 8	-0.36	< 0.0001	--	--	--	--	15.37	0.0095
PP at Sieve No. 30	--	--	--	--	0.11	0.0257	--	--
PP at Sieve No. 200	1.23	0.0004	-8.14	0.0142	--	--	--	--
G_{sb}	--	--	--	--	--	--	582.12	0.0011
VTM	0.56	0.0002	--	--	-0.34	< 0.0001	--	--
P_{ba}	--	--	-27.13	0.0097	-1.11	0.0116	--	--
No. of Observations	78		74		76		16	
R ²	0.68		0.74		0.77		0.77	
Adjusted R ²	0.66		0.73		0.76		0.71	

-- = parameter was not significant at a 0.05 significance level; APA = Asphalt Pavement Analyzer; CML = Cantabro mass loss; G_{sb} = aggregate bulk specific gravity; IDT-HT = indirect tensile at high temperature; No. = number; P_{ba} = absorbed asphalt content; PP = percent passing; VTM = voids in total mixture. Cells highlighted in green represent factors positively correlated with balanced mix design test. Cells highlighted in orange represent factors negatively correlated with balanced mix design test.

The sign of the regression coefficients (i.e., positive or negative) indicates whether the correlation between the volumetric or gradation parameter and the BMD test result is positive or negative. Overall, the sign of the coefficients in Table 7 were in agreement with the correlation direction previously presented in Table 6. In addition, the VIF values across all regression model coefficients ranged from 1.0 to 2.2 and were below 5.0, suggesting that the significant factors identified in Table 7 are not collinear with each other.

For Cantabro, AC, percent passing sieve No. 8, percent passing sieve No. 200, and VTM were found to be significant factors affecting CML results at a 0.05 significance level. For AC, mixtures with more AC tend to have higher cohesion and, thus, lower mass loss, justifying the negative coefficient (-2.50). In terms of gradation, the coefficients for percent passing sieves No. 8 (-0.36) and No. 200 (1.23) showed contradicting outcomes with respect to CML, suggesting that the effect of gradation on the Cantabro test may be gradation or mixture dependent, or both. Lastly, VTM showed a positive coefficient (0.56), which suggests that higher air void content increases mass loss potential during the Cantabro test. This observation agrees with findings from previous studies and highlights its statistical significance (Cox et al., 2017, Bowers et al., 2023).

For IDT-CT, AC, percent passing at sieve No. 200, and P_{ba} were found to be significant factors affecting CT_{Index} results at a 0.05 significance level. For AC, the positive coefficient (75.21) suggests that adding asphalt leads to higher CT_{Index} values or vice versa. For gradation, the percent passing at sieve No. 200 negative regression coefficient (-8.14) suggests that finer mixtures may lead to a decrease in CT_{Index} . P_{ba} also showed a negative coefficient (-27.13), indicating that less absorptive aggregates may also result in higher CT_{Index} values. Collectively, these observations highlight the potential role of asphalt film thickness on CT_{Index} . Notably, increasing AC, coarsening the mix gradation (e.g., reducing the percent passing at sieve No. 200), or selecting less absorptive aggregates (i.e., reducing P_{ba}) may lead to an increase in the asphalt film thickness and, thus, higher CT_{Index} values.

For APA, AC, percent passing sieve No. 30, VTM, and P_{ba} were found to be significant factors affecting APA rut depth at a significance level of 0.05. For AC, the positive coefficient (1.77) shows that mixtures with more asphalt tend to be more rut susceptible, as expected, if all other factors are kept constant. Percent passing at sieve No. 30 also showed a positive coefficient (0.11), suggesting that finer mixtures may be more prone to rutting, potentially caused by a reduction in the larger aggregates' interlocking and load transfer ability through the aggregate structure. Lastly, VTM and P_{ba} had negative coefficients, indicating that higher air void contents and more absorptive aggregates (i.e., lower P_{be}) lead to lower rutting.

For IDT-HT, AC, percent passing at sieve No. 8, and G_{sb} were found to be significant factors affecting high temperature strength at a 0.05 significance level. Overall, these factors agree with the ones identified as significant in a previous study (Boz et al., 2023), which adopted a different analysis methodology using the same dataset but including asphalt binder rheological parameters. Table 7 shows that AC has a negative coefficient (-59.77), indicating lower rutting resistance (i.e., lower high temperature strength values) for higher AC values. G_{sb} showed a positive coefficient (582.12), indicating that aggregates with higher density may help improve rutting resistance. Percent passing at sieve No. 8 showed a positive coefficient (15.37), suggesting that finer mixtures may exhibit higher strength values, contradicting observations from APA analysis that indicated finer mixtures may rutting. This difference may be due to the different failure mechanisms that govern these tests. Although the APA test measures the densification and shear deformation of test specimens, the IDT-HT test measures the cohesion component of asphalt mixtures' shear strength to estimate their rutting potential (Zhou et al., 2021). Thus, increasing the amount of fines may have two effects: (1) reducing the larger aggregates' interlocking, leading to a more rut-susceptible mixture from the APA perspective, and (2) enhancing mixture cohesion, given that finer mixtures often require higher AC to meet volumetrics requirements, leading to an improved rutting performance from an IDT-HT perspective. Nonetheless, several studies have been conducted to underscore the strong correlation between APA and IDT-HT tests (Bennert et al., 2018; Boz et al., 2023; Hajj et al., 2025; Taylor et al., 2026). It is worth noting that further analysis is warranted because of inherent test variability and potential changes to the significant factors identified when including binder quality parameters in the statistical analysis.

From the collective observations, it can be seen that AC, percent passing at sieve No. 8, percent passing at sieve No. 30, percent passing at sieve No. 200, G_{sb} , VTM, and P_{ba} are critical factors that can significantly affect BMD test results. In fact, AC and gradation show as significant in all tests, suggesting that tight control of such parameters during production is crucial, given their potential effect on BMD test results. For instance, Diefenderfer et al. (2024) observed that AC is the most influential factor for potential failures on IDT-CT results during production.

Table 7 demonstrates that the regression models showed moderate predictability, with adjusted R^2 ranging from 0.66 to 0.76. Figure 17 shows the circular validation for each BMD test comparing the predicted test results using the multiple linear regression equations in Table 7 with the measured data. Overall, Figure 17 shows that the predicted values fall along the line of equality for all tests. For IDT-CT and IDT-HT, 93% and 88% of the samples fell within the repeatability limits, respectively. This observation highlights the similarities between predicted and measured CT_{Index} and high temperature strength values from the single-operator test

precision perspective. Similar analyses were not conducted for Cantabro and APA because precision limits have not been developed yet.

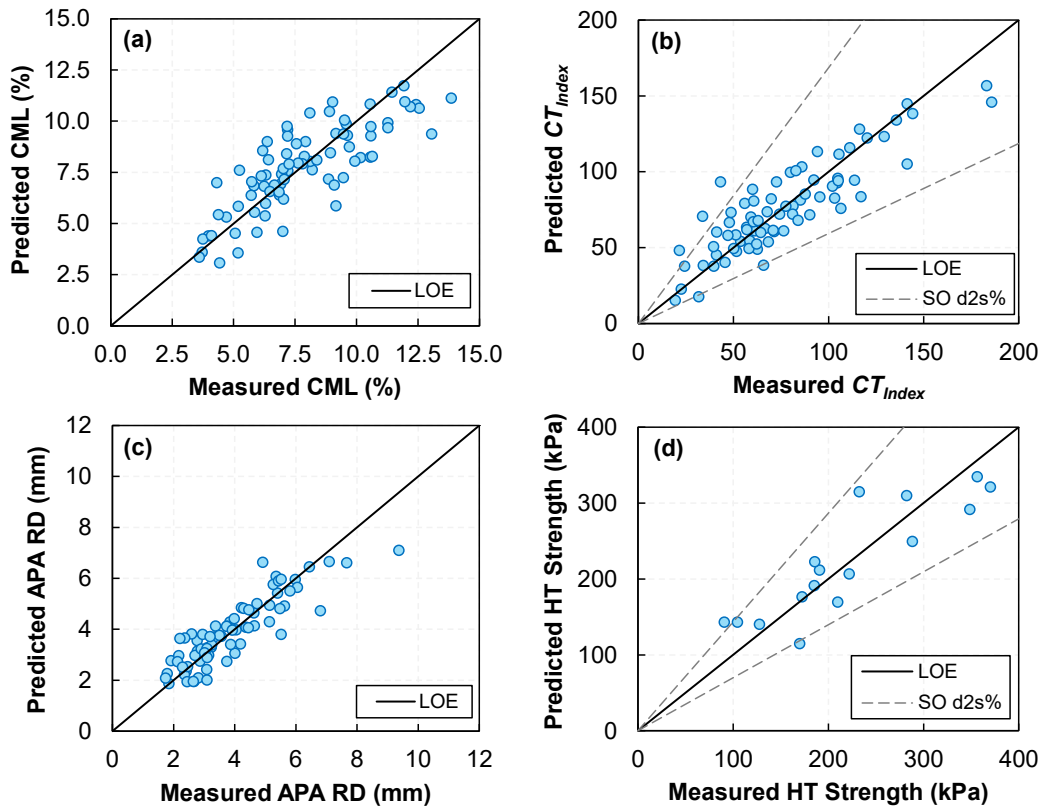


Figure 17. Circular Validation of the Volumetrics and Gradation Significant Factors and Multiple Linear Regression Functions for (a) Cantabro, (b) Indirect Tensile Cracking Test, (c) APA, and (d) Indirect Tensile at High Temperature. APA = Asphalt Pavement Analyzer; CT = cracking tolerance; HT = high temperature; LOE = line of equality; RD = rut depth; SO $d2s\%$ = single-operator precision limit.

Predictive Equation as Decision-Support Tool During Production: Case Study

In Virginia, mixtures with D-type designation are subjected to BMD tests during production for acceptance, including IDT-CT and Cantabro for every 2,000 tons and IDT-HT for every 4,000 tons of produced mix. Although these tests provide an indication of asphalt mixture performance, the testing frequency impairs the ability to take corrective actions in a timely manner. This impairment affects producers when addressing mixture composition issues during production and affects districts when evaluating the potential mixture acceptance.

This challenge may be more concerning when considering the changes in the 2026 construction season specifications that reduce the IDT-CT and Cantabro testing frequency during production from every 2,000 tons to every 4,000 tons, provided the first three consecutive lots meet all performance criteria. No changes are expected in terms of testing frequency for the IDT-HT (every 4,000 tons) and APA tests (only on request). In this scenario, understanding the effect of asphalt mixture composition on its performance is critical to increase the confidence of producers to make quick mix adjustments before or during production and for districts to determine mix acceptance.

A case study was conducted to evaluate the potential of using performance predictive equations as a support tool to increase the confidence in the decision-making process for requesting or waiving additional testing during production. Although Cantabro, IDT-HT, and IDT-CT are performed during production, the analysis focused solely on cracking performance (i.e., CT_{Index}) for the reasons stated in the Methodology section. CT_{Index} predictive equations were developed using a tiered approach (Tier 1 through Tier 3) to reflect varying testing frequencies for volumetrics and gradation data during production. Lastly, a risk assessment analysis was conducted to quantify the potential risk of adopting such predictive equations to support the decision to request or waive additional testing during production. Analysis was focused on test results obtained from reheated asphalt mixtures because the reheated condition is the most typical conditioning practice reported.

Development of CT_{Index} Predictive Equations

A tiered approach was proposed to consider the readily available parameters based on their testing frequency during production. Three tiers were proposed, with increasing numbers of factors when moving from Tier 3 to Tier 1. Table 8 summarizes the independent variables considered in the analysis for each tier. Only a subset was included in the predictive equations in each analysis tier. These variables were selected based on the outcomes of the collinearity analysis, with collinear factors being removed from the finalized equation. All analysis was conducted using Dataset 2.

Table 8. Factors Considered in Each Tier for CT_{Index} Predictive Equation

Independent Variable	Tier 3	Tier 2	Tier 1
Asphalt Content (AC)	X	X	X
Percent Passing at Sieve No. 4 (PP#4)	X	X	X
Percent Passing at Sieve No. 8 (PP#8)	X	X	X
Percent Passing at Sieve No. 30 (PP#30)	X	X	X
Percent Passing at Sieve No. 200 (PP#200)	X	X	X
Aggregate Bulk Specific Gravity (G_{sb})	X	X	X
Aggregate Effective Specific Gravity (G_{se})	X	X	X
Maximum Specific Gravity (G_{mm})	-	X	X
Bulk Specific Gravity (G_{mb})	-	X	X
Voids in Total Mixture (VTM)	-	X	X
Voids in the Mineral Aggregate (VMA)	-	X	X
Voids Filled with Asphalt (VFA)	-	X	X
Absorbed Asphalt Content (P_{ba})	-	X	X
Effective Asphalt Content (P_{be})	-	X	X
Fine-to-Asphalt Ratio (FA)	-	X	X
Cantabro Mass Loss (CML)	-	-	X

X = parameter considered in the statistical analysis; - = parameter not considered in the statistical analysis.

Table 8 shows that Tier 3 includes AC, gradation parameters, and aggregate specific gravities, parameters that are more frequently tested during production (i.e., every 500 tons) or that are available prior to the beginning of production. Tier 2 includes additional volumetric parameters that are obtained every 1,000 tons (i.e., VTM, VMA, VFA, and FA). Tier 1 includes CML data as a potential performance indicator to improve CT_{Index} predictability. It is worth noting that the Tier 1 approach is motivated by the correlation between CML and CT_{Index} , as previously shown in Figure 5 and Figure 7. Recent studies have shown that asphalt binder

properties, such as the Glover-Rowe parameter (*GRP*), have a strong correlation with asphalt mixtures' cracking resistance (Christensen, 2023; Boz et al., 2025d) and can be leveraged as input data into predictive equations to predict asphalt mixtures' CT_{Index} during production (Boz et al., 2026). Nonetheless, conducting frequent testing on the asphalt mixture extracted and recovered binder during production is neither feasible nor practical. To use *GRP* as input, it must be backcalculated from measured CT_{Index} (Boz et al., 2026). In this context, the use of CML as a potential surrogate to *GRP* values is beneficial because Cantabro tests can be more frequently performed using Superpave gyratory compacted specimens, which are fabricated every 1,000 tons for volumetric verification purposes during production. Thus, potential errors introduced by the uncertainty in the *GRP* backcalculation process might be minimized by using actual CML measured values.

To develop the CT_{Index} predictive equations, the stepwise forward multiple linear regression approach was selected, and collinearity among variables was verified using both pairwise correlation and VIF values. Recall that the stepwise regression approach starts with a null model and adds independent variables one at a time until a stopping criterion is met. The stepwise regression analysis was conducted using the p-value, selecting only the significant factors at a 0.05 significance level, and the minimum AIC stopping criteria for each predictive equation tier. In all cases, the AIC stopping criterion resulted in equations with improved predictability (i.e., higher adjusted R^2 and lower root mean squared error). The final equations for Tier 1 to Tier 3 are presented in Equations 1 to 3, respectively. The adjusted R^2 values were 0.79, 0.73, and 0.75 for Tier 1 to Tier 3 equations, respectively. In all cases, the factors were found to be statistically significant (p-value < 0.05), with VIF values below 5 (maximum of 1.98), indicating low collinearity. Note that the Tier 2 equation is the same as the equation presented in Table 7.

$$(CT\ index)_{Tier\ 1} = -269.96 + 56.24 \times AC - 5.61 \times PP\#200 + 6.02 \times VMA - 5.23 \times CML \quad [Eq. 1]$$

$$(CT\ index)_{Tier\ 2} = -295.32 + 75.21 \times AC - 8.14 \times PP\#200 - 27.13 \times P_{ba} \quad [Eq. 2]$$

$$(CT\ index)_{Tier\ 3} = -356.38 + 85.14 \times AC - 8.91 \times PP\#200 \quad [Eq. 3]$$

It is worth noting that these equations are preliminary and are shown as examples, with the goal of evaluating their potential as tools to support the decision of requesting or waiving additional testing, especially when reduced testing frequency is in place during production. These equations should not be used in lieu of any VDOT's BMD tests for mixture acceptance purposes.

Verification of CT_{Index} Predictive Equations

Verification of Equations 1 to 3 predictability was conducted using the volumetrics, gradation, and CT_{Index} data in Dataset 3. The analysis was conducted using both producers (Figure 18) and districts (Figure 19) data. Figure 18a and Figure 19a present the comparison between predicted and measured CT_{Index} for producers and districts, respectively. Figure 18b and Figure 19b show the percentage of data within the single-operator precision limit for producers and districts, respectively. Collectively, the CT_{Index} predictive power increases when moving from Tier 3 to Tier 1, with data points falling closer to the line of equality and within the

repeatability limits. Notably, improved predictions were obtained when the Tier 1 equation was used, suggesting the potential of leveraging CML as a performance indicator and an aid in the decision-making process during mixture acceptance and quality control practices.

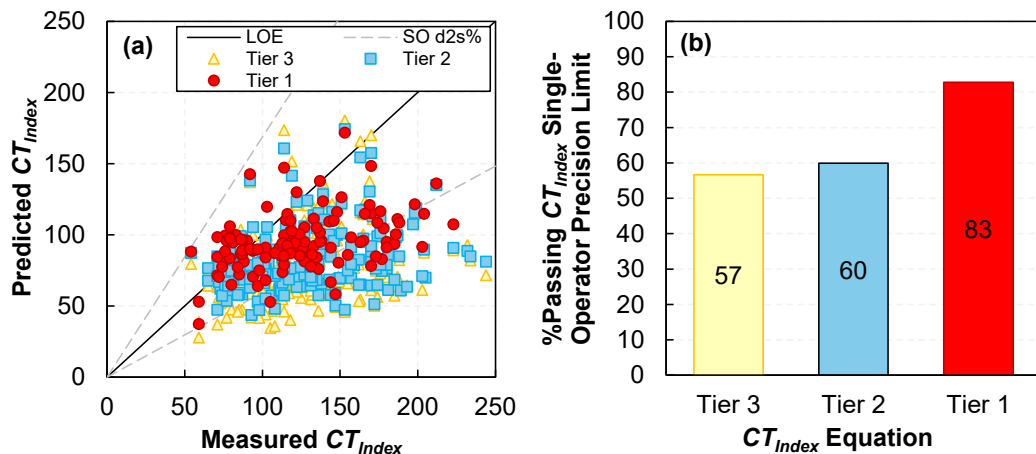


Figure 18. Evaluating Tiers 1 to 3 CT_{Index} Predictive Equations Using Producers' Data by (a) Comparing Measured and Predicted Values in an Equality Plot and (b) Quantifying Percentage of Samples Passing CT_{Index} Single-Operator Precision Limit. CT = cracking tolerance; LOE = line of equality; SO $d2s\%$ = single-operator precision limit.

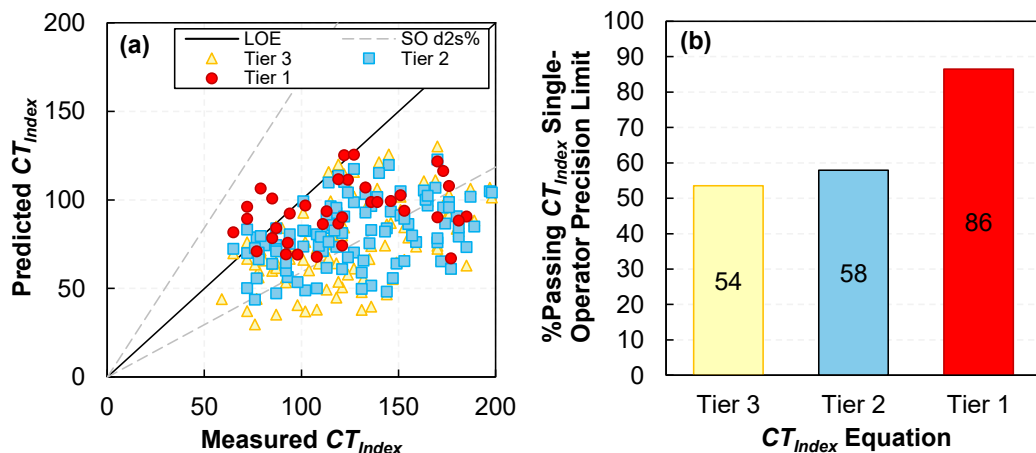


Figure 19. Evaluating Tiers 1 to 3 CT_{Index} Predictive Equations Using VDOT Districts' Data by (a) Comparing Measured and Predicted Values in an Equality Plot and (b) Quantifying Percentage of Samples Passing CT_{Index} Single-Operator Precision Limit. CT = cracking tolerance; LOE = line of equality; SO $d2s\%$ = single-operator precision limit.

Additional analysis is needed to quantify the potential risk of using such equations to determine additional testing needs during production. Risks would be associated with scenarios in which the predictive equation results in failing CT_{Index} values (below threshold), whereas actual tests result in passing CT_{Index} values (above threshold), or vice versa. In these scenarios, the need for additional testing from predicted results would be opposite to the outcome obtained from measured data. The risk assessment analysis is presented in more detail in the next section.

Risk Assessment Evaluation

The analysis and findings presented so far suggest that the proposed CT_{Index} predictive equations could be used in conjunction with readily available volumetric, gradation, and CML data to predict the cracking potential of asphalt mixtures and support the decision of requesting or waiving additional testing from upcoming production lots. In a hypothetical scenario, if consistent volumetrics are obtained and passing CT_{Index} values are predicted, practitioners could leverage this information to increase their confidence in waiving additional testing from the upcoming lots in support of the reduction in the production-level testing frequency. On the other hand, if inconsistent volumetrics are obtained and failing CT_{Index} values are predicted, these models could serve as a decision-support tool to request additional testing from upcoming lots and complement currently existing quality assurance protocols.

Although these models provide additional confidence to both producers and districts in their decision-making processes, they should not be used as a replacement for the BMD tests. In this sense, quantifying the potential risks associated with the adoption of such equations is needed, especially in cases in which the predicted and measured CT_{Index} values fall on different sides of the performance criterion. Figure 20 illustrates the different scenarios and potential risks associated in each of them. The following presents a brief description of each region.

- **Region 1:** *Passing predicted CT_{Index} and failing measured CT_{Index} .*
Passing predicted CT_{Index} values suggest that no additional testing would be needed for an upcoming production lot. However, the measured CT_{Index} showed failing results. In this scenario, the predicted CT_{Index} values would have incorrectly suggested that additional testing could be waived. Further analysis for this particular region is desired to quantify the risk of making the wrong decision with respect to additional testing needs during production.
- **Region 2:** *Passing predicted CT_{Index} and passing measured CT_{Index} .*
Both measured and predicted values would have resulted in passing CT_{Index} results, suggesting that if the predicted equations were used, additional IDT-CT could have been waived from the upcoming production lot, agreeing with the measured CT_{Index} results that also showed passing values. Therefore, no risks are associated with the use of CT_{Index} predictive equations within this region.
- **Region 3:** *Failing predicted CT_{Index} and passing measured CT_{Index} .*
Failing predicted CT_{Index} values suggest that additional testing would be needed for the upcoming production lot. However, the measured CT_{Index} showed passing results. In this scenario, the predicted CT_{Index} values would have incorrectly suggested that additional testing should be required. Further analysis for this particular region is desired to quantify the risk of making the wrong decision with respect to additional testing needs during production.
- **Region 4:** *Failing predicted CT_{Index} and failing measured CT_{Index} .*
Both measured and predicted values would have resulted in failing CT_{Index} results, suggesting that if the predicted equations were used, additional IDT-CT could have been requested from the upcoming production lot, agreeing with the measured CT_{Index} results that also showed failing values. Because the outcomes are the same, no risks are associated with the use of CT_{Index} predictive equations within this region.

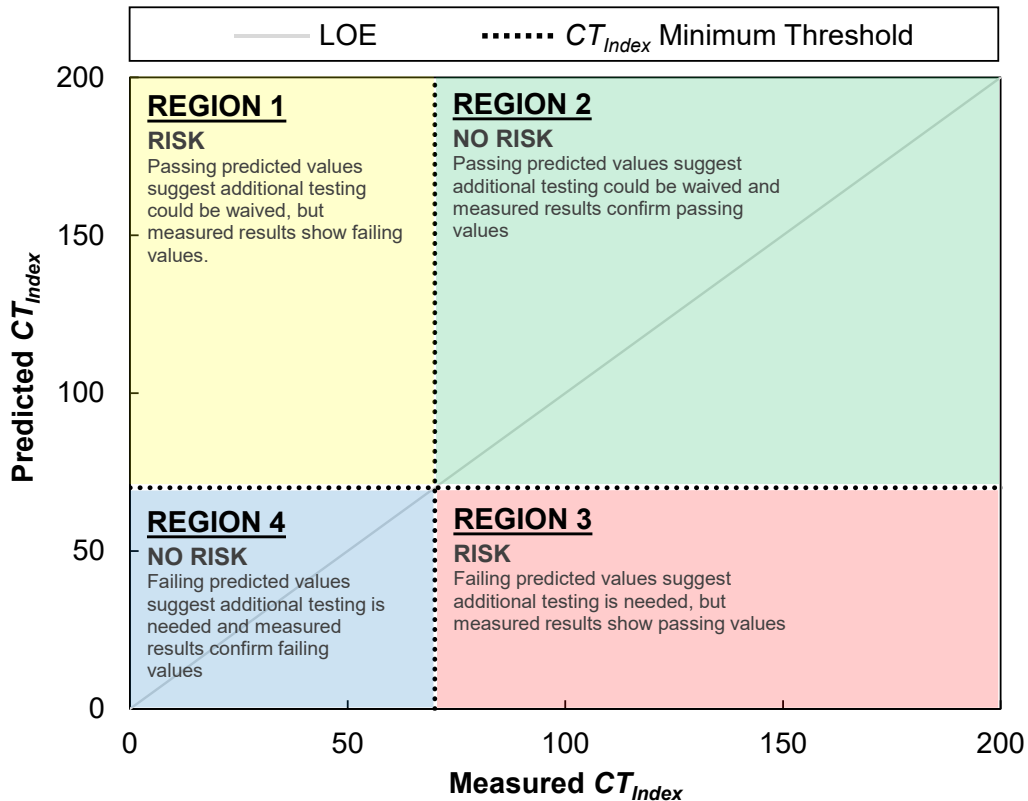


Figure 20. Illustration of Different Scenarios Related to the Potential Adoption of CT_{Index} Predictive Equations. CT = cracking tolerance; LOE = line of equality.

Analysis from Figure 20 suggests that the regions in which predicted CT_{Index} values would have led to the incorrect decision on requiring or waiving additional testing are Regions 1 and 3. Additional analysis was conducted using the CT_{Index} predictive equations for each tier to predict the CT_{Index} values from producers and districts using Dataset 3 and to quantify the risk based on the region in which the predicted values would fall. Recall that all volumetrics data that failed VDOT’s process tolerance limits, as well as multilaboratory precision limits, were removed prior to the analysis to avoid potential bias in the predicted CT_{Index} values mainly caused by inconsistencies in the asphalt mixture composition under production settings. For this analysis, it is assumed that the data in Dataset 3 correspond to the fourth lot production data to compare predicted and measured CT_{Index} values.

Figure 21 compares predicted and measured CT_{Index} values and shows the risk analysis by region for producers and districts. The CT_{Index} values were predicted using the corresponding production lot volumetrics and gradation test results and compared with their corresponding measured CT_{Index} . The comparison is presented separately for each of the predictive equations.

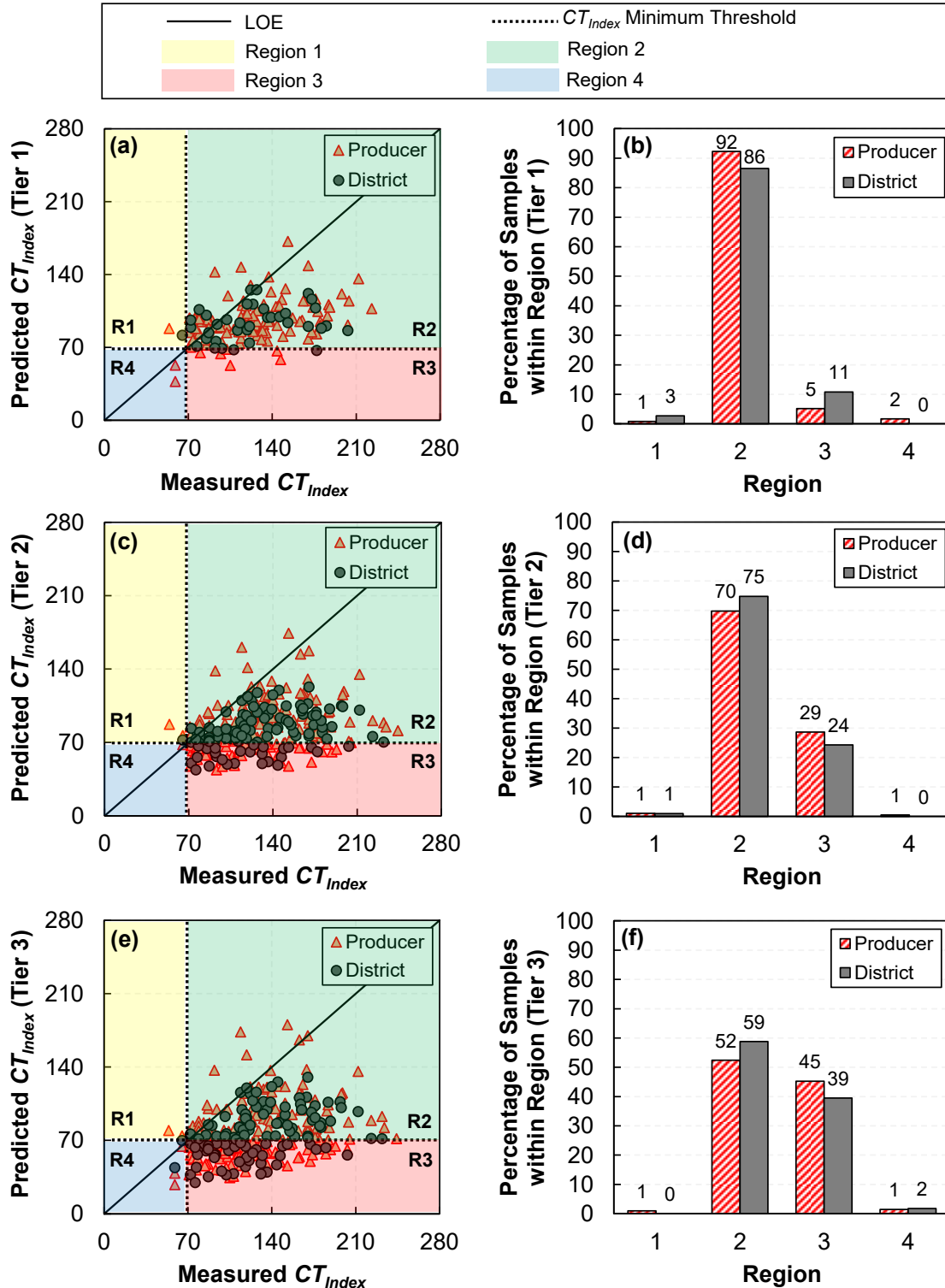


Figure 21. Evaluation of CT_{Index} Predictive Equations Adoption, Including (a) Predicted versus Measured CT_{Index} for Tier 3, (b) Risk Assessment for Tier 3, (c) Predicted versus Measured CT_{Index} for Tier 2, (d) Risk Assessment for Tier 2, (e) Predicted versus Measured CT_{Index} for Tier 1, (f) Risk Assessment for Tier 1 Using Both Producers and Districts Data. CT = cracking tolerance; LOE = line of equality; R1 = Region 1; R2 = Region 2; R3 = Region 3; R4 = Region 4.

For Tier 3, several samples fell within Region 3 (Figure 21a), suggesting that the predictive equation results would have incorrectly led to additional testing despite passing measured results. In fact, 45% of the producer samples and 39% of the district samples fell within Region 3, representing a high risk if the Tier 3 equation is used for performance testing decisions. Only 1% of producer samples fell within Region 1, where the equation results would have led to incorrectly waived testing despite failing measured results. Overall, 53% of the producer samples and 61% of the district samples align with correct decisions: 52% and 59% in Region 2 (both passing for producers and districts, respectively) and 1% and 2% in Region 4 (both failing for producers and districts, respectively). The Tier 3 equation uses AC and percent passing at sieve No. 200 as predictors because they are tested frequently during production (every 500 tons), thereby potentially improving confidence in CT_{Index} estimates. Nevertheless, the substantial risk of incorrect decisions when using Tier 3 offsets this higher testing frequency.

Tier 2 provided more accurate CT_{Index} predictions compared with Tier 3, as Figure 21c shows, where samples align more closely to the line of equality. Figure 21d presents the risk associated with the adoption of Tier 2 predictive equation: 29% of producer samples and 24% of district samples fall in Region 3 and 1% in Region 1 for both groups, indicating a relatively high chance of incorrect decisions regarding additional testing. Although 71% of producer samples and 75% of district samples fell in no-risk regions, the overall risk remains high—30% for producers and 25% for districts—compared with the typical 5% risk under standard mixture acceptance and quality control protocols. The improved predictions from Tier 2 stem from adding P_{ba} to the AC and percent passing No. 200 sieve variables used in Tier 3, but this enhancement does not sufficiently reduce decision-making risk.

The Tier 1 equation yielded the most accurate CT_{Index} predictions compared with Tiers 2 and 3. As shown in Figure 21e, only a few samples fell in Regions 1 and 3. For instance, Figure 21f indicates that Region 3 contains 5% of producer samples and 11% of district samples—similar to the acceptable error in typical mixture acceptance and quality control protocols. Region 1 includes only 1% of producer samples and 3% of district samples. Overall, the Tier 1 predictions equation matched outcomes from measured IDT-CT results in 94% of producer cases and 86% of district cases: 92% and 86% in Region 2 (both passing for producers and districts, respectively) and 2% in Region 4 (both failing for producers). Figure 22 summarizes these results, showing that most samples—94% for producers and 86% for districts—fall in the “No Risk” region when the Tier 1 equation was applied.

These findings underscore the importance of the volumetrics and gradation parameters identified previously as significant factors affecting BMD test results. Moreover, they show that using CML as a performance indicator improves CT_{Index} predictions and supports the potential of the Tier 1 predictive equation as a tool to increase confidence in the decision-making process of requesting or waiving additional testing during production. Although promising, further analysis with local datasets is needed to confirm the equations’ applicability to specific districts, asphalt plants, and material conditions.

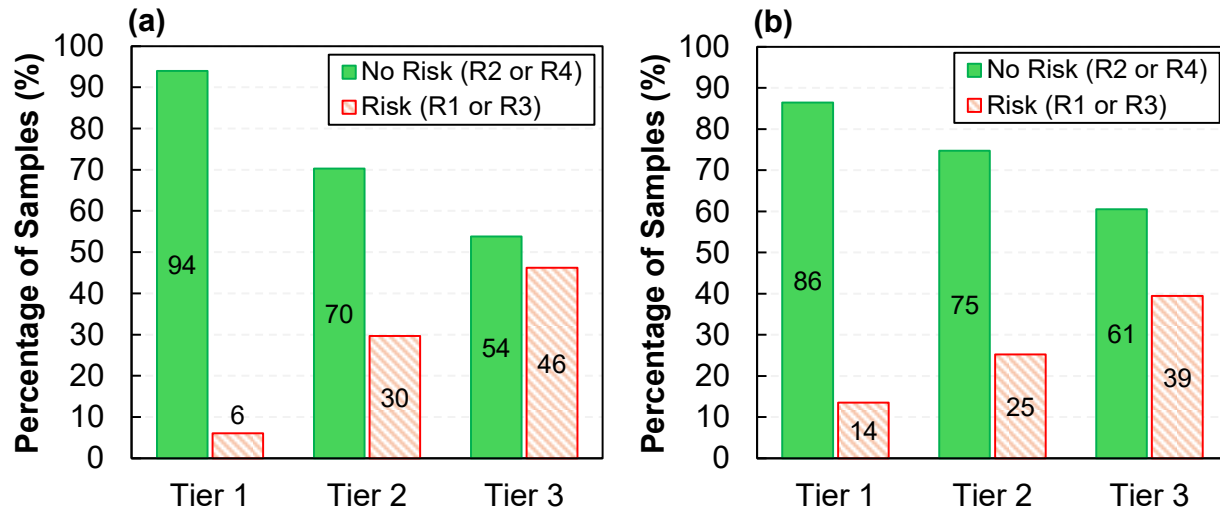


Figure 22. Risk Assessment Quantifying the Percentage of Samples with Different Outcomes between Measured and Predicted CT_{Index} Using (a) Producer's Data and (b) District's Data. R1 = Region 1; R2 = Region 2; R3 = Region 3; R4 = Region 4.

Laboratory Evaluation of Storage and Mixture Conditioning Protocols

Survey

A survey was conducted to gather information regarding the practices of producers and VDOT districts in Virginia related to sampling of plant-produced asphalt surface mixtures, material conditioning (reheat or non-reheat), and specimen fabrication practices, such as lag time for BMD tests during production. Table 9 summarizes the number of responses by district and organization. Although the survey results provide a general overview of current state practices involved in the sampling, material conditioning, and specimen fabrication for BMD testing, they are not comprehensive across all districts and producers in Virginia. Moreover, the responses reflect typical practices rather than all potential variations that may arise under production conditions. For brevity, information regarding lag time only is presented herein. This information was used to refine the experimental plan for the laboratory evaluation. Although not included in the report, information related to reheated or non-reheated testing was also used to supplement Dataset 3 prior to evaluating the effect of reheating in BMD test results.

Table 9. Number of Survey Respondents

District	Producer	VDOT District	Total by District
Bristol	0	0	0
Culpeper	3	1	4
Fredericksburg	5	1	6
Hampton Roads	2	1	3
Lynchburg	4	1	5
Northern Virginia ^a	8	0	8
Richmond	3	1	4
Salem	10	1	11
Staunton	4	1	5
Total by Organization	39	7	N/A

N/A = not applicable. ^a The Northern Virginia district does not have a laboratory; tests are performed at the Culpeper district laboratory.

Figure 23 presents the percentage of responses relative to the total number of producers or total number of districts regarding typical interval lag times and dwell times (i.e., time interval between specimen fabrication and BMD testing). Based on the collective responses (producers and districts), 67% reported a lag time of less than 1 day, and 26% reported a lag time of 1 to 2 days. Only 8% reported a lag time greater than 2 days. In terms of dwell time, the collective responses (producers and districts) showed that 17% reported a dwell time of less than 1 day, and 80% reported a dwell time of 1 to 2 days. Only one producer reported a dwell time of 2 to 4 days. Figure 23 also shows an overall trend among the producers' responses. Individuals who reported less than 1 day of lag time also reported 1 to 2 days of dwell time and vice versa. This observation suggests that longer lag times are compensated by shorter dwell times, and shorter lag times are balanced by longer dwell times. This trend was consistent across different producers and highlights overall adherence to the 48-hour requirement to submit BMD performance test results as specified in VDOT's BMD special provision. For VDOT districts, the lag time ranged from less than 1 day to 4 to 7 days, with most producers (63%) reporting 1 to 2 days of lag time. All districts reported a dwell time of 1 to 2 days.

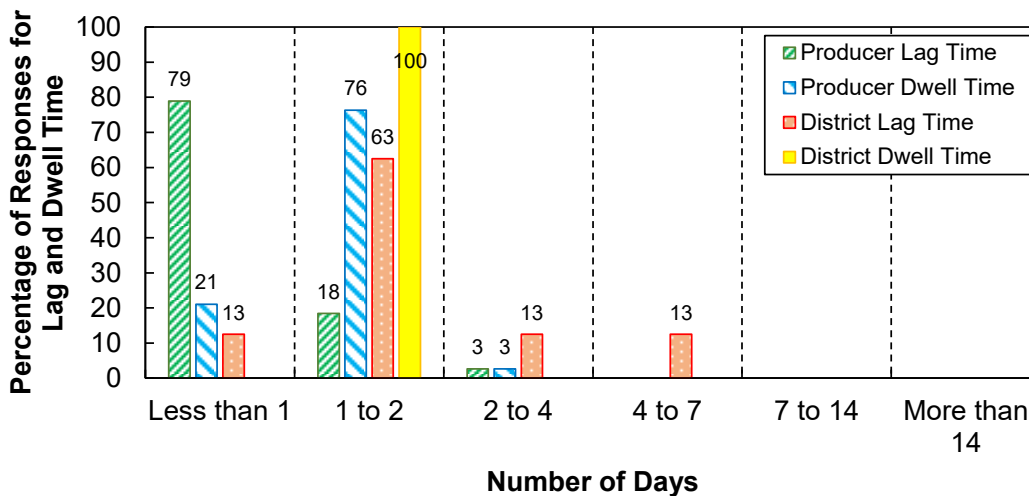


Figure 23. Percentage of Responses Regarding Lag and Dwell Times. Dwell time = time interval between specimen fabrication and balanced mix design testing; Lag time = time interval between asphalt sampling and specimen fabrication.

Based on the survey responses, 1- and 4-day lag times were selected to represent the range of typical lag times reported by producers and districts for the laboratory evaluation.

Lag Time and Reheating Evaluation

Figure 24 presents a summary of the test results at the NR, R1d, and R4d conditions for Cantabro, IDT-CT, APA, and IDT-HT. Because of challenges with equipment and specimen handling with mixture M4 at the NR condition for Cantabro and R1d condition for APA, data for this mixture were not included in the analysis for these tests. Statistical analysis was also conducted to verify whether the differences in test results between different conditions were significant. Although not included in the report for brevity reasons, PG of the virgin and

extracted recovered asphalt binders met the specified PG in the JMF and the requirements in VDOT's specification, respectively (VDOT, 2020).

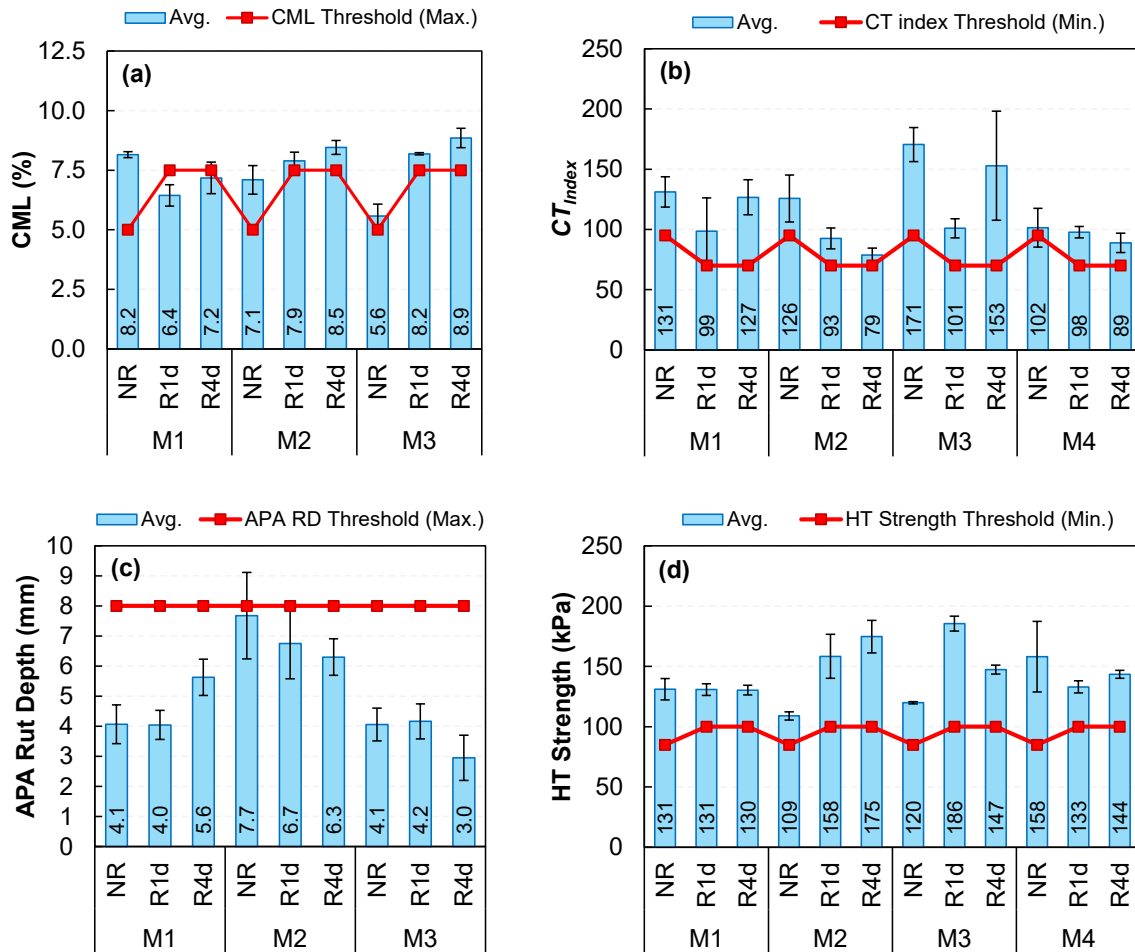


Figure 24. Comparison between Test Results Obtained at Non-reheated, Reheated 1 Days after Sampling, and Reheated 4 Days after Sampling for (a) Cantabro, (b) Indirect Tensile Cracking Test, (c) APA, and (d) Indirect Tensile At High Temperature. APA = Asphalt Pavement Analyzer; Avg. = average; CML = Cantabro mass loss; CT = cracking tolerance; HT = high temperature; M = mixture; Max. = maximum; Min. = minimum; NR = non-reheated; R1d = reheated after 1-day lag time; R4d = reheated after 4-day lag time; RD = rut depth. Error bars indicate ± 1 standard deviation.

Figure 24a shows the Cantabro results for M1 to M3. M1 showed a decrease in CML between NR and R1d but an increase in CML values from the R1d to R4d conditions. M2 and M3 showed an increase in CML between NR and R1d and a further increase from R1d and R4d. Figure 9a also shows that applying the 7.5% mass loss criterion to both NR and R1d or R4d conditions would yield conflicting outcomes for M2 and M3, with passing CML values for the non-reheat case but failing values for the reheat conditions. Differentiating the non-reheated performance criterion proposed in this study (i.e., 5.0%) from the reheated criterion (i.e., 7.5%) would result in consistent outcomes, with both M2 and M3 failing the thresholds regardless of the conditioning method.

Figure 24b presents the IDT-CT results for M1 to M4. In all cases, the NR condition yielded the higher CT_{Index} compared with R1d and R4d, as expected, confirming the trends in Figure 1 and past VTRC studies. For M1 and M3, a CT_{Index} increase was observed from the R1d to R4d conditions, which could be associated with IDT-CT variability. No correlation was observed between CT_{Index} and volumetrics or gradation parameters that justifies the counterintuitive trend at these conditions. In addition, it is unlikely that higher oxidative aging levels induced by the increase in lag time from 1 to 4 days would yield such an increase in CT_{Index} (51.5% increase). As part of a Consortium for Asphalt Pavement Research and Implementation (CAPRI) extended study, it was observed that CT_{Index} was not significantly different for lag times ranging from 2 days to 2 months for 90% of the evaluated mixtures (Taylor, 2025). The results from this study corroborate the findings from the CAPRI study.

Figure 24c shows the APA results for M1 to M3. For M1, rut depth remained relatively constant from NR to R1d but increased from R1d to R4d. For M2 and M3, a gradual decrease in rut depth values was observed from NR to R4d, suggesting that longer lag times may reduce the asphalt mixture rutting susceptibility. For all mixtures, the measured rut depth at the end of the 8,000 cycles was below the 8.0 mm criterion, regardless of the mixture conditioning method.

Figure 24d presents the IDT-HT results for M1 to M4. M1 and M4 showed relatively constant high temperature strength values among different reheating conditions. On the other hand, M2 and M3 showed an increase in strength values from the NR to the R1d or R4d conditions. An increase in strength values is expected from the non-reheated to reheated condition because of the additional oxidative aging induced onto the asphalt mixture during reheating, which increases binder stiffness. This increase may provide a false sense of improved rutting resistance. Similar to the APA test, all mixtures showed test results passing the performance criteria, with strength values above both the preliminary non-reheated threshold (85 kPa) and the reheated one (100 kPa). Additional mixture testing is warranted to verify the proposed non-reheated criterion for the IDT-HT test.

Additional analysis was conducted to verify whether the differences in test results among NR, R1d, and R4d conditions were statistically significant. The analyses included the BMD test results as the response variable and an interaction term as a factor combining “mixture” (M1 to M4) and “condition” (NR, R1d, and R4d), termed “Mixture Type*Condition.” For IDT-CT, a Welch Analysis of Variance (ANOVA) was conducted at a 0.05 significance level, because the data met the normality assumption but failed the equal variance assumption. For Cantabro, APA, and IDT-HT, a Welch Analysis of Variance (ANOVA) was conducted at a 0.05 significance level, because both normality and equal variance assumptions were met.

Table 10 shows the summary of the statistical analysis. The p-values suggest that at least one test condition (NR, R1d, or R4d) is significantly different from the others at a 0.05 significance level for Cantabro, IDT-CT, and IDT-HT. Additional analysis was conducted to identify the pairwise comparisons that are statistically different and is presented next. For APA, the p-value is greater than the test significance level, suggesting that rut depths obtained at NR, R1d, or R4d conditions are statistically similar. Moreover, recall that the preliminary non-reheated APA rut depth performance criterion proposed was 7.5 mm, which is similar to the current 8.0 mm. Collectively, these observations suggest that APA is not as sensitive in capturing

differences between reheated and non-reheated tests as the other BMD tests. In addition, APA is not often conducted during production and is required only on request. In this context, developing a separate criterion for the APA non-reheated condition may not be practically or significantly effective in screening out rut-susceptible mixtures.

Table 10. Summary of ANOVA Statistical Analysis for the Mixture Type*Condition Interaction Factor for Cantabro, IDT-CT, APA, and IDT-HT

Test	Number of Observations	Degrees of Freedom	p-value
Cantabro	27	4	< 0.0001
IDT-CT	60	6	< 0.0001
APA	18	4	0.9936
IDT-HT	36	6	< 0.0001

ANOVA = Welch Analysis of Variance; APA = Asphalt Pavement Analyzer; IDT-CT = indirect tensile cracking tolerance; IDT-HT = indirect tensile at high temperature.

A pairwise comparison was further conducted to identify the significant differences between different Mixture Type*Condition interaction factors for Cantabro, IDT-CT, and IDT-HT results. For the Cantabro and IDT-HT tests, a Tukey’s Honestly Significant Difference (HSD) post-hoc test was conducted at a 0.05 significance level, given that the data met the equal variance assumption. For IDT-CT, a Games-Howell post-hoc test was conducted instead at a 0.05 significance level, because data met the normality assumption but failed the equal variance assumption.

Table 101 presents the statistical groupings based on Tukey’s HSD or Games-Howell post-hoc tests. Interaction factors that share the same letter indicate that the difference between their mean values is not significantly different at a 0.05 significance level. Conversely, interaction factors with different letters are considered statistically different at a 0.05 significance level. The interaction factors with the highest mean value have the “A” designation, and the factor with the lowest mean has the highest letter (in alphabetical order) among the tested conditions.

Table 11. Pairwise Statistical Comparison of Lag Times for Cantabro, IDT-CT, and IDT-HT

Mixture	Condition	Interaction Factor	Statistical Groupings		
			Cantabro	IDT-CT ^a	IDT-HT
M1	Non-reheat	M1*NR	A B	B	C D E
	Reheat 1-day lag time	M1*R1d	C D	B C D E	C D E
	Reheat 4-day lag time	M1*R4d	B C	B C	C D E
M2	Non-reheat	M2*NR	B C	A B C D E	E
	Reheat 1-day lag time	M2*R1d	A B	C D E	A B C
	Reheat 4-day lag time	M2*R4d	A	E	A B
M3	Non-reheat	M3*NR	D	A	D E
	Reheat 1-day lag time	M3*R1d	A B	B C D	A
	Reheat 4-day lag time	M3*R4d	A	A B C D E	B C D
M4	Non-reheat	M4*NR	--	B C D E	A B C
	Reheat 1-day lag time	M4*R1d	--	C D	C D E
	Reheat 4-day lag time	M4*R4d	--	D E	B C D

-- = interaction factor not considered in the analysis; IDT-CT = indirect tensile cracking tolerance; IDT-HT = indirect tensile at high temperature; M = mixture; NR = non-reheated; R1d = reheated after 1-day lag time; R4d = reheated after 4-day lag time. Interaction factors sharing the same letter are statistically similar at a significance

level of 0.05. ^a IDT-CT statistical grouping was determined based on Games-Howell analysis, and Cantabro and IDT-HT were determined from Tukey's Honestly Significant Difference.

For the Cantabro test, different lag times did not yield significantly different CML results for all three mixtures. Similar observations were made for IDT-CT, with different lag times not resulting in significantly different CT_{Index} values for all four mixtures. It is worth noting that, despite the unexpected increase in CT_{Index} from R1d to R4d for M1 and M3 (Figure 24b), the CT_{Index} difference between these two conditions is below the single-operator precision of 51.1% of their average (Habbouche et al., 2022) for M1 and just marginally failing the limit for M3. For the IDT-HT test, three out of four mixtures (M1, M2, and M4) showed no significant difference between 1 and 4 days of lag time. The only exception was M3, which showed a significant decrease in strength values from the R1d to R4d conditions. Although unexpected, this decrease is below the single-operator precision limit of 35.6% of their average, suggesting the strength results between R1d and R4d are considered similar and within the test precision (Boz et al., 2025a).

Based on the results from M1 to M4 testing, no significant effect on Cantabro, IDT-CT, and IDT-HT results was observed between lag times of 1 to 4 days, which represent the range of typical lag times reported by producers and districts in Virginia. The insignificant effect of varying lag times is potentially masked by the inherent material and test variability.

Verification of Mixture Composition Effect on Balanced Mix Design Test Results

To further verify the significance of the volumetrics and gradation parameters on BMD test results, the predictive equations presented in Table 7 were used along with the measured volumetrics and gradation parameters to estimate the CML, CT_{Index} , APA rut depth, and high temperature strength values for mixtures M1 to M4. Data from the R1d condition were used because they correspond to the most typical lag time reported statewide. For APA analysis, no data were available for M4, given challenges with the test equipment. Figure 25 summarizes the measured and estimated BMD test results along with the performance criteria and precision limits calculated based on the measured values, applicable to IDT-CT and IDT-HT only.

For Cantabro (Figure 25a), the predictive equation did not result in consistent CML estimates. In fact, diverging practical outcomes were obtained in terms of CML acceptance, with the predictive equation indicating a passing mix and the measured CML indicating a failing mix, or vice versa. Overall, an overestimation of CML values was observed for mixtures M1, M3, and M4, whereas M2 showed an estimated CML value substantially less than the measured one.

For IDT-CT (Figure 25b), an overall good agreement was obtained between measured and estimated CT_{Index} . For two out of four mixtures (i.e., M2 and M4), the estimated CT_{Index} was within the single-operator coefficient of variation limit ($1s\%$) of 18.3%, with a maximum difference of 5 CT_{Index} units between measured and estimated. In terms of single-operator acceptable range ($d2s\%$), three out of four mixtures (i.e., M2, M3, and M4) were within the limits, indicating that the estimated values are similar to the measured values. Notably, the estimated and measured values were in agreement, with both results above the performance criterion of 70 for all mixtures.

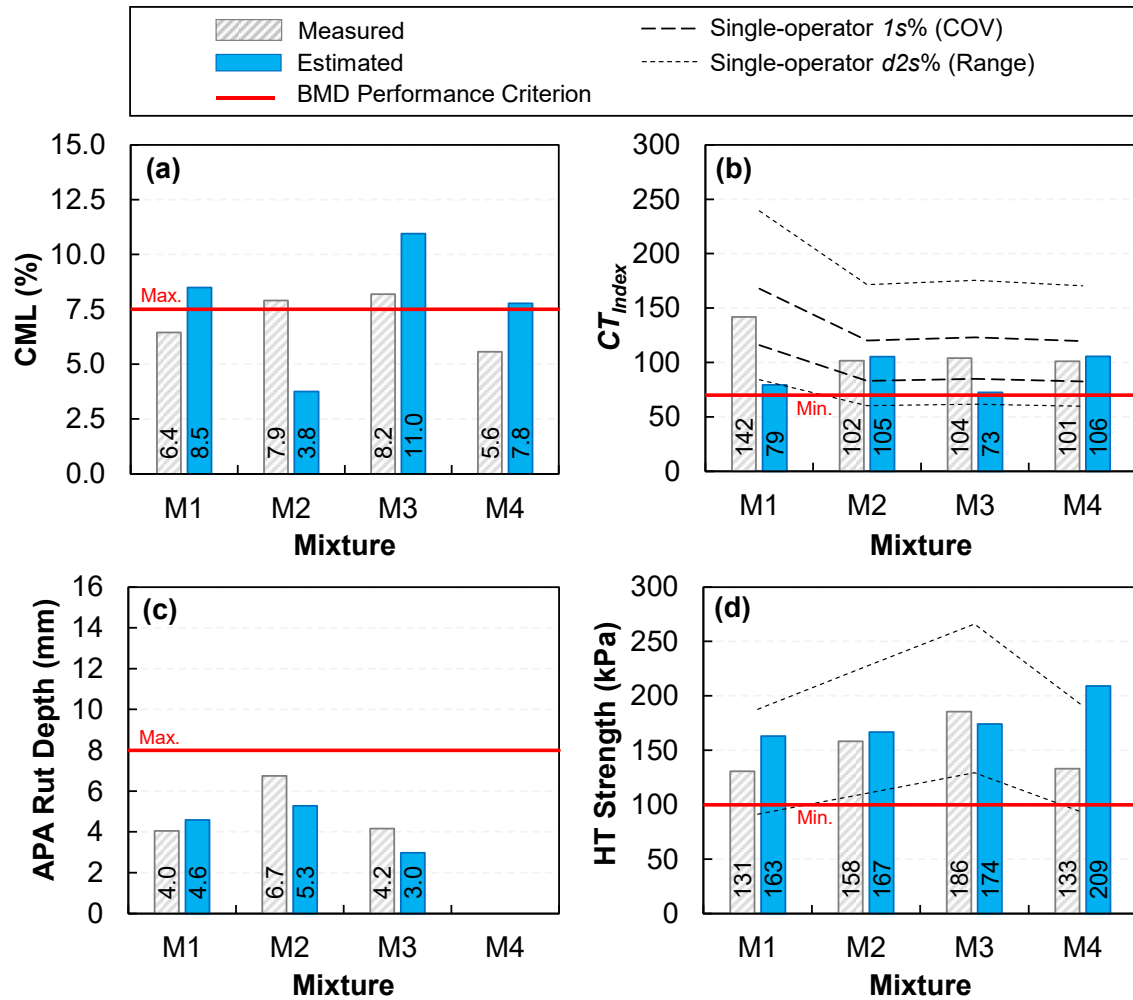


Figure 25. Comparison between Measured and Estimated Values for (a) Cantabro, (b) Indirect Tensile Cracking Test, (c) APA, and (d) Indirect Tensile at High Temperature Using Predictive Equations. APA = Asphalt Pavement Analyzer; BMD = balanced mix design; CML = Cantabro mass loss; COV = coefficient of variation; HT = high temperature; M = mixture; Max. = maximum; Min. = minimum; $1s\%$ = coefficient of variation; $d2s\%$ = acceptable range.

For APA (Figure 25c), reasonable rut depth estimates were obtained, with the difference between measured and estimated rut depth ranging from 0.6 to 1.4 mm. In fact, a similar ranking in terms of APA rutting was observed between measured and estimated results, with M3 being the most rut-resistant mixture and M2 the least rut-resistant one. Lastly, the estimated and measured results were in agreement, with both results falling below the performance criterion of 8.0 mm for all mixtures.

For IDT-HT (Figure 25d), an additional step was needed to shift the estimated strength values prior to comparing them with the measured results. Although the predictive equation was developed using data obtained from dry-conditioned test specimens, M1 to M4 were tested after wet-conditioning, which is the currently specified protocol in Virginia Test Method 145 (VDOT, 2025). Past VTRC research project showed that the dry strength values are 1.33 times greater than the wet ones (Boz et al., 2025a). Therefore, the estimated strength values in Figure 25d were obtained by dividing the estimated high temperature strength values by a factor of 1.33. Figure

25d shows that, overall, good strength predictions were obtained. In three out of four mixtures (i.e., M1, M2, and M3), the estimated strength value was within the single-operator precision limit of 35.6%, with a maximum range of 21.8% for M1. Moreover, both estimated and measured strength values were above the performance criterion of 100 kPa for all mixtures.

Overall, given the good agreement between estimated and measured BMD results, the analysis presented herein validates the volumetrics and gradation parameters identified as significant in Table 7. Although promising, further testing on additional mixtures should be performed to confirm these findings.

Summary of Findings

- The ratio in CT_{Index} performance criteria between 100 and 130 for the reheated and non-reheated conditions during production, respectively, was validated.
- Although the current performance criteria for Cantabro, APA, and IDT-HT tests were developed based on test results from reheated asphalt mixtures during production, they are applied to both reheated and non-reheated conditions. Using the same criteria can bias the acceptance rates, especially if the differences between reheated and non-reheated test results are significant.
- A preliminary non-reheated threshold of 5.0% maximum mass loss was proposed for the Cantabro test. Statistical analysis of 64 PMLC mixtures showed that the difference in CML values between reheated and non-reheated samples was significant.
- A preliminary non-reheated threshold of 85 kPa minimum strength was proposed for the IDT-HT test. Statistical analysis of six PMLC mixtures showed that the difference in strength values between reheated and non-reheated samples was significant.
- A preliminary non-reheated threshold of 7.5 mm maximum rut depth was proposed for the APA test. However, given the similarity with the reheated criterion (i.e., 8.0 mm), the marginally significant difference between reheated and non-reheated results for 35 PMLC mixtures, and the fact that the APA is not required during production, specifying a separate criterion for the non-reheated condition may not be practical.
- The correlation between asphalt binder quantity and BMD test results was consistent and confirmed observations from the literature. The VBE parameter provided insights beyond what total AC alone can offer. At a constant AC, mixtures with higher VBE showed lower CML, higher CT_{Index} , and higher rut depths. Therefore, although establishing VBE limits would not necessarily guarantee compliance with both VMA and VTM requirements, it could be used as an indicator to help guide mix adjustments during design and/or production when trying to meet BMD performance criteria.
- The correlation between asphalt binder quality and BMD test results was not consistent across different material source locations. The binder quality was indirectly estimated using RAP content, virgin binder PG, and the use of RAs as categorical factors in the correlation

analysis with BMD. The analysis highlighted the need to explore alternative rheological properties of the extracted and recovered asphalt to correlate binder quality to mixture performance.

- Factors such as AC, percent passing sieve No. 8, percent passing sieve No. 30, percent passing sieve No. 200, G_{sb} , VTM, and P_{ba} , in the absence of binder quality parameters, were found to be significant in influencing Cantabro, IDT-CT, APA, and IDT-HT results, with specific subsets affecting individual tests.
- A strong correspondence was identified between the reheated performance criteria obtained from Cantabro (7.5% mass loss) and IDT-CT (CT_{Index} of 70). Similarly, the proposed non-reheated Cantabro criterion (5.0% mass loss) aligned with the corresponding non-reheated IDT-CT performance criterion (CT_{Index} of 95). Nonetheless, it is worth noting that the Cantabro test should not be used as a replacement for IDT-CT.
- A case study demonstrated the potential of the CT_{Index} predictive equation incorporating CML (Tier 1) to be used as a support tool to increase confidence in the decision-making process of requesting or waiving additional testing during production, especially under reduced testing frequency. The case study showed a substantial improvement in CT_{Index} prediction accuracy when CML was included as a predictor, with similar outcomes between predicted and measured CT_{Index} values (i.e., both passing or both failing), ranging from 86% to 94% of the production samples.
- A VTRC survey conducted as part of this study showed that the range of typical lag times reported by producers and districts in Virginia is 1 to 4 days.
- Laboratory testing of four PMLC mixtures showed no significant differences in Cantabro, IDT-CT, APA, and IDT-HT results between 1- and 4-day lag times for reheated mixtures.
- Statistical analysis on laboratory test results showed no significant differences in APA test results between non-reheated and reheated conditions, confirming the findings presented previously in this study. Significant differences were observed for Cantabro, IDT-CT, and IDT-HT results between non-reheated and reheated conditions.
- Laboratory tests on four PMLC mixtures indicated that applying the 5.0% mass loss criterion to Cantabro non-reheated test results yielded mixture acceptance decisions consistent with those of the reheated case. Using 7.5% for both non-reheated and reheated conditions yielded divergent outcomes, with passing results for the non-reheated condition and failing results for the reheated condition for the same mixture.
- Overall, predictive equations showed a good agreement between estimated and measured IDT-CT, APA, and IDT-HT laboratory results for the four PMLC mixtures. This analysis helped validate the volumetrics and gradation parameters identified as significant for these tests. Although promising, further evaluation is warranted to confirm these findings along with the binder quality parameters.

CONCLUSIONS

- *The ratio in CT_{Index} performance criteria between 100 for reheated mixtures and 130 for non-reheated mixtures was confirmed to be appropriate for production conditions.*
- *Preliminary non-reheated performance criteria were proposed, including a maximum of 5.0% mass loss for the Cantabro test and a minimum 85 kPa for the IDT-HT test. The 5.0% mass loss non-reheated criterion aligned well with the CT_{Index} non-reheated criterion of 95, and the 7.5% mass loss reheated criterion aligned well with the IDT-CT reheated criterion of 70. For the IDT-HT test, an apparent correspondence between mixtures that failed the 85 kPa non-reheated criterion and mixtures with low binder stiffness was observed, which would agree with findings from the National Cooperative Highway Research Program IDEA Project 245 that concluded that the IDT-HT test is sensitive to changes in binder stiffness and correlates with the asphalt mixture extracted binder PG (Boz et al., 2025b). Given the limited number of mixtures that failed the 85 kPa threshold, further analysis is warranted to confirm the findings.*
- *The CT_{Index} predictive equation incorporating CML (Tier 1) can be used as a support tool to increase confidence in the decision-making process of requesting or waiving additional testing during production, especially under reduced testing frequency. Including CML as a parameter substantially improved CT_{Index} predictability and underscored the potential of using CML as a cracking performance indicator during production. In fact, similar outcomes between predicted and measured CT_{Index} values (i.e., both passing or both failing) were obtained for 86% to 94% of the production samples. When used as a support tool, this framework may increase confidence from the producer perspective by allowing timely mix adjustments, and from the districts' perspective, when evaluating mix acceptance and the need to request or waive additional testing. The predicted CT_{Index} should not be used to fully or partially replace VDOT's BMD tests for mixture acceptance purposes.*
- *Multiple mixture compositional and volumetric parameters, such as AC, percent passing sieve No. 8, percent passing sieve No. 30, percent passing sieve No. 200, G_{sb} , VTM, and P_{ba} , in the absence of binder quality parameters, are significant factors influencing Cantabro, IDT-CT, APA, and IDT-HT results, with specific subsets affecting individual tests. AC and gradation parameters were significant for all BMD tests, suggesting that tight control of such parameters is crucial during production. Higher AC led to improved durability and cracking but worse rutting resistance. Finer mixtures resulted in worse cracking resistance, and higher VTM showed a reduction in durability. Mixtures with more absorptive aggregates showed reduced cracking resistance but improved rutting resistance, highlighting the role of the effective asphalt (content or volume) in mixture performance.*

RECOMMENDATIONS

1. *VDOT should continue specifying the CT_{Index} performance criteria of 100 and 130 for reheated and non-reheated asphalt mixtures during production, respectively.* This study validated the ratio in CT_{Index} values between reheated and non-reheated conditions.
2. *VDOT should continue benchmarking the Cantabro, IDT-CT, and IDT-HT production results for future construction seasons to evaluate the feasibility of the performance criteria under non-reheated conditions proposed in this study of maximum 5.0% mass loss and minimum 85 kPa for the Cantabro and IDT-HT tests, respectively.* The continued monitoring effort aligns with the outcomes from VTRC Report No. 26-R22, which recommends benchmarking the variability of IDT-CT and IDT-HT results following the adoption of the Virginia Test Method for BMD specimen preparation.
3. *VDOT should consider incorporating the proposed CT_{Index} predictive equation framework based on the Cantabro test results (Tier 1) into a guidance document to support the current BMD specification when reduced testing frequency is applied during production, with refinement as additional data become available.* Analysis showed that Cantabro performance criteria (both non-reheated and reheated conditions) align with the CT_{Index} criteria and demonstrate potential for use as a cracking performance indicator during production to support informed decisions regarding additional cracking tests. Adoption of such a framework would warrant an interlaboratory study for the Cantabro test to develop precision limits for data quality assessment. This framework should not be used to fully or partially replace any of VDOT's BMD tests for mixture acceptance purposes.
4. *VDOT should continue monitoring and evaluating the effect of mixture composition on BMD test results.* In addition to AC and gradation parameters, the statistical analysis presented in this study indicated that G_{sb} , VTM, and P_{ba} significantly influence BMD test results related to durability, cracking, and rutting resistance. Moreover, the quality of the asphalt binder plays a key role in asphalt mixture performance, highlighting the need to explore correlations between BMD results and other rheological parameters other than PG.

IMPLEMENTATION AND BENEFITS

The researcher and the technical review panel (listed in the Acknowledgments) for the project collaborate to craft a plan to implement the study recommendations and determine the benefits of doing so. This process is to ensure that the implementation plan is developed and approved with the participation and support of those involved with VDOT operations. The implementation plan and the accompanying benefits are provided here.

Implementation

Regarding Recommendation 1, VDOT's Materials Division agrees to continue specifying the CT_{Index} thresholds of 100 for reheated mixtures and 130 for non-reheated mixtures during production.

Regarding Recommendation 2, VDOT's Materials Division will continue to support the development of a research needs statement (RNS) to be submitted as part of the implementation action item from VTRC Report No. 26-R22. The proposed RNS will focus on benchmarking efforts to evaluate the variability of BMD test results following implementation of the proposed Virginia Test Method for specimen preparation in BMD testing. This RNS will incorporate elements outlined in Recommendation 2 of this study. The anticipated research needs statement is expected to be submitted to the VTRC Pavement Research Advisory Subcommittee B by the fall of 2027.

*Regarding Recommendations 3 and 4, VTRC, with the help of VDOT's Materials Division and districts, will continue evaluating the proposed CT_{Index} predictive equation framework based on the Cantabro test results (Tier 1) and the effect of mixture composition and asphalt binder quality (using alternative rheological parameters) on BMD results using 2025 and 2026 production data. A sensitivity analysis will also be conducted to quantify the effect of varying individual mixture composition factors on BMD results. Once the validation analysis is completed, VDOT's Materials Division will consider including the proposed framework as part of a guidance document. This effort will be undertaken as part of the ongoing VTRC Project No. 128620, entitled *Evaluating the Impact of Volumetric Properties and Reheating on the Balanced Mix Design Test Results - Phase II*, that is expected to be completed by January 2027.*

Benefits

Implementing the recommendations of this study has four main benefits. First, it provides confidence to VDOT's Materials Division on adopting the revised IDT-CT performance criteria (100 for reheated and 130 for non-reheated samples), which became effective in 2026. Second, benchmarking the production test results in the 2026 construction season will help validate and assess the feasibility of the proposed preliminary non-reheated thresholds for Cantabro and IDT-HT tests, coupled with testing additional PMLC mixtures under VTRC Project No. 128620. Third, the critical volumetrics and gradation parameters that can significantly affect BMD test results were identified, suggesting that tight control of such parameters during production is crucial. Lastly, the proposed framework, including CT_{Index} predictive equations, can be used in a guidance document by districts as a tool to support the decision of requesting or waiving additional testing, especially when reduced testing frequency is in place during production. Collectively, this study supports VDOT's Materials Division by ensuring BMD implementation is performed in an efficient yet responsible manner in Virginia.

ACKNOWLEDGMENTS

The authors are grateful to the following members serving on this study's technical review panel: Angela Beyke, Project Champion and Assistant State Materials Engineer, VDOT's Materials Division; Sean Li, Asphalt Program Manager, VDOT's Materials Division; Clyde Landreth, Pavement Engineer, VDOT Salem District; Miranda Kidd, District Materials Engineer, VDOT Hampton Roads District; Travis Cable, Vice-President Quality Control, Colony

Construction, Inc.; Jason Provines, Senior Research Scientist, VTRC Structures Team; and Mourad Bouhajja, formerly of VDOT's Materials Division.

The authors thank Andrew Barbour, Derek Lister, Troy Deeds, Jennifer Samuels, Brian Powell, Ignacio Simon, and Jennifer Gonzalez from VTRC for their outstanding work and assistance with sampling and testing the materials from this study. The authors also thank Stacey Diefenderfer, VTRC, for her review and feedback.

REFERENCES

- Al-Qadi, I.L., Wang, H., Baek, J., Leng, Z., Doyen, M., and Gillen, S. Effects of curing time and reheating on performance of warm stone-matrix asphalt. *Journal of Materials in Civil Engineering*, Vol. 24, No. 11, 2012, pp. 1422–1428.
[https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0000513](https://doi.org/10.1061/(ASCE)MT.1943-5533.0000513).
- American Association of State Highway and Transportation Officials. *Standard Specifications for Transportation Materials and Methods of Sampling and Testing and AASHTO Provisional Standards*. Washington, DC, 2022.
- American Association of State Highway and Transportation Officials. *Standard Specifications for Transportation Materials and Methods of Sampling and Testing and AASHTO Provisional Standards*. Washington, DC, 2023.
- American Association of State Highway and Transportation Officials. *Standard Specifications for Transportation Materials and Methods of Sampling and Testing and AASHTO Provisional Standards*. Washington, DC, 2024.
- American Association of State Highway and Transportation Officials. *Standard Specifications for Transportation Materials and Methods of Sampling and Testing and AASHTO Provisional Standards*. Washington, DC, 2025.
- ASTM International. *ASTM D8225-19: Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature*. West Conshohocken, PA, 2019.
- Bennert, T. Role of Asphalt Binder in Balanced Mix Design Performance Testing. *Presented at 105th Transportation Research Board Annual Meeting*, Washington, D.C., 2026.
- Bennert, T., Haas, E., and Wass, E. Indirect Tensile Test (IDT) to Determine Asphalt Mixture Performance Indicators during Quality Control Testing in New Jersey. *Transportation Research Record*, Vol. 2672, No. 28, 2018, pp. 394–403.
<https://doi.org/10.1177/0361198118793276>.
- Bonaquist, R. *Impact of mix design on asphalt pavement durability*. Transportation Research Circular E-C186. Transportation Research Board, Washington, DC, 2014, pp. 1–17.

- Bowers, B., Lynn, T., Yin, F., Moore, N., Diefenderfer, S.D., and Boz, I. *Impact of Production Variability on Balanced Mix Designs in Virginia*. VTRC Report No. 23-R20. Virginia Transportation Research Council, Charlottesville, VA, 2023.
- Boz, I., Diefenderfer, S.D., and Habbouche, J. *Development of a Performance Threshold for CT_{Index} from Asphalt Mixtures Tested during Production (non-reheat)*. Presented at the 11th Balanced Mix Design Technical Committee Meeting, Virginia Transportation Research Council, Charlottesville, VA, 2022.
- Boz, I., Diefenderfer, S.D., Habbouche, J., and Seitllari, A. *Interlaboratory Study for the Indirect Tensile at High Temperature Test and Ideal Rutting Test*. VTRC Report No. 25-R14. Virginia Transportation Research Council, Charlottesville, VA, 2025a.
- Boz, I., Habbouche, J., Diefenderfer, S.D., Coffey, G.P., Ozbulut, O.E., and Seitllari, A. *Simple and Practical Tests for Rutting Evaluation of Asphalt Mixtures in the Balanced Mix Design Process*. VTRC Report No. 23-R11. Virginia Transportation Research Council, Charlottesville, VA, 2023.
- Boz, I., Habbouche, J., Taylor, A.J., and Moore, N. *Ruggedness Study of Specimen Preparation and Fine-Tuning of Test Methods for IDT-CT and IDT-HT Test*. VTRC Report No. 26-R22. Virginia Transportation Research Council, Charlottesville, VA, 2025c.
- Boz, I., Hajj, E.Y., Habbouche, J., Diefenderfer, S.D., Alam, I., and El Hajj, H. *A Practical Oven-Aging Method for Evaluating Long-Term Cracking Performance of Asphalt Mixtures in Virginia*. VTRC Report No. 26-R04. Virginia Transportation Research Council, Charlottesville, VA, 2025d.
- Boz, I., Seitllari, A., Kuchiishi, A.K., and Diefenderfer, S.D. A Model for Estimating Cracking Resistance Potential of Asphalt Mixtures in a Balanced Mix Design Framework. Accepted for Publication in the *Journal of Materials in Civil Engineering*. Under Review, 2026.
- Boz, I., Yin, F., Silva, L.P.V., and Chen, C. *Simple and Rapid Tests for Assessing Quality and Consistency of Reclaimed Asphalt Pavement (RAP) for Recycled Asphalt Mixture Applications*. National Cooperative Highway Research Program Report IDEA Project 245. Transportation Research Board of the National Academies, Washington, D.C., 2025b.
- Christensen, D.W. *Interlayer Mixture Design*. Publication WHRP 0092-21-04. Wisconsin Department of Transportation, Research and Library Unit, Wisconsin Highway Research Program, Madison, WI, 2023.
- Christensen, D.W., and Bonaquist, R.F. *Volumetric Requirements for Superpave Mix Design*. National Cooperative Highway Research Program Report 567. Transportation Research Board of the National Academies, Washington, D.C., 2006.

- Cox, B.C., Smith, B.T., Howard, I.L., and James, R.S. State of Knowledge for Cantabro Testing of Dense Graded Asphalt. *Journal of Materials in Civil Engineering*, Vol. 29, No. 10, 2017, 04017174. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002020](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002020).
- Daniel, J.S., Corrigan, M., Jacques, C., Nemati, R., Dave, E.V., and Congalton, A. Comparison of asphalt mixture specimen fabrication methods and binder tests for cracking evaluation of field mixtures. *Road Materials and Pavement Design*, Vol. 20, No. 5, 2019, pp. 1059–1075. <https://doi.org/10.1080/14680629.2018.1431148>.
- Daniel, J. S., Jacques, C., and Salehi, S. *Performance of high RAP pavement sections in NH*. FHWA-NH-RD-15680B. New Hampshire Department of Transportation, Bureau of Materials and Research, Concord, NH, 2015.
- Diefenderfer, S.D., and Boz, I. *Field Validation of Balanced Mix Design Initial Criteria*. VTRC Report No. 26-R20. Virginia Transportation Research Council, Charlottesville, VA, 2025.
- Diefenderfer, S.D., Boz, I., and Habbouche, J. *Balanced Mix Design for Asphalt Surface Mixtures: 2019 Field Trials*. VTRC Report No. 21-R21. Virginia Transportation Research Council, Charlottesville, VA, 2021.
- Diefenderfer, S.D., Boz, I., and Habbouche, J. *Balanced Mix Design for Surface Mixtures: 2021 and 2022 Plant Mix Schedule Pilots*. VTRC Report No. 23-R19. Virginia Transportation Research Council, Charlottesville, VA, 2023a.
- Diefenderfer, S.D., Boz, I., Bowers, B., Lynn, T., Yin, F., and Moore, N. *Asphalt Mixture Variability and Its Effect on Balanced Mix Design Testing from Design to Production*. VTRC Report No. 25-R1. Virginia Transportation Research Council, Charlottesville, VA, 2024.
- Diefenderfer, S.D., Habbouche, J., and Boz, I. *Balanced Mix Design for Surface Mixtures: 2020 Field Trials*. VTRC Report No. 23-R13. Virginia Transportation Research Council, Charlottesville, VA, 2023b.
- Diefenderfer, S.D., Smith, B.C., and Entwisle, C. *Virginia's 2024 Implementation of Balanced Mix Design*. VTRC Report No. 26-R28. Virginia Transportation Research Council, Charlottesville, VA, 2026.
- Etheridge, R.A., Wang, Y.D., Kim, S.S., and Kim, Y.R. Evaluation of Fatigue Cracking Resistance of Asphalt Mixtures Using Apparent Damage Capacity. *Journal of Materials in Civil Engineering*, Vol. 31, No. 11, 2019, 04019257. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002870](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002870).

- Habbouche, J., Boz, I., and Diefenderfer, S.D. *Interlaboratory Study for the Indirect Tensile Cracking Test at Intermediate Temperature: Phase II*. VTRC Report No. 23-R3. Virginia Transportation Research Council, Charlottesville, VA, 2022.
- Hajj, E.Y., Elias, N.G., Khanal, B., Alrajhi, A., Hand, A., Bennert, T., Zhou, F., Epps, J., Chan, K.M., Decker, C.S., Patel, H., and Duval, J. *Balanced Mix Design: Rutting Performance Tests*. Airport Asphalt Pavement Technology Program, Atlantic City, NJ, 2025.
- JMP Statistical Discovery LLC. 2015 JMP® version 12.2.0. Cary, NC, 2015.
- Kandhal, P.S., and Cooley, Jr., L.A. *Accelerated Laboratory Rutting Tests: Evaluation of the Asphalt Pavement Analyzer*. National Cooperative Highway Research Program Report 508. Transportation Research Board of the National Academies, Washington, D.C., 2003.
- Kidd, A., Stephenson, G., and White, G. Implications of reheating of asphalt mixes on performance testing. In *Proceedings of the 18th AAPA International Flexible Pavements Conference*, Sydney, Australia, 2019, pp. 1–15.
- Lemke, Z., Sadek, H., Swiertz, D., Reichelt, S., and Bahia, H.U. Effects of reheating procedure and oven type on performance testing results of asphalt mixtures. *Transportation Research Record*, Vol. 2672, No. 26, 2018, pp. 124–133.
<https://doi.org/10.1177/0361198118773886>.
- Leng, Z., Al-Qadi, I.L., Baek, J., Doyen, M., Wang, H., and Gillen, S. Effect of reheating plant warm SMA on its fracture potential. In *Proceedings of the 7th RILEM International Conference on Cracking in Pavements*, A. Scarpas et al. (Eds.), pp. 1341–1349, 2012.
- Mivehchi, M., Wen, H., Wen, Y., and Wang, L. Study of Measures to Design Asphalt Mixes Including High Percentages of Recycled Asphalt Pavement and Recycled Asphalt Shingles. *Transportation Research Record*, Vol. 2677, No. 1, 2023, pp. 869–879.
<https://doi.org/10.1177/03611981221108982>.
- Oshone, M., Pires, G.M., Jiménez del Barco Carrión, A., Rahbar-Rastegar, R., Airey, G.D., Daniel, J.S., Bailey, H., and Smith, D. Simulating Plant Produced Material in the Laboratory to Replicate Rheological and Fatigue Properties. *Road Materials and Pavement Design*, Vol. 21, No. 1, 2020, pp. 253–261.
<https://doi.org/10.1080/14680629.2018.1484384>.
- Prowell, B.D., Zhang, J., and Brown, E.R. *Aggregate Properties and the Performance of Superpave-Designed Hot Mix Asphalt*. National Cooperative Highway Research Program Report 539. Transportation Research Board of the National Academies, Washington, D.C., 2005.
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2025. <https://www.R-project.org/>.

- Rivera-Perez, J., Talebpour, A., and Al-Qadi, I.L. Prediction of asphalt concrete flexibility index and rut depth utilizing deep learning and Monte Carlo Dropout simulation. *International Journal of Pavement Engineering*, Vol. 24, No. 1, 2023, 2253964. <https://doi.org/10.1080/10298436.2023.2253964>.
- Rocha, M.L., Enriquez-Leon, A.J., Aragao, F.T.S., Espindola, R.P., and Nascimento, L.A.H. Exploring the effects of volumetrics and binder properties on the performance of asphalt mixtures. *Construction and Building Materials*, Vol. 438, 2024, 137054. <https://doi.org/10.1016/j.conbuildmat.2024.137054>.
- Tarefder, R.A., Zaman, M., and Hobson, K. A Laboratory and Statistical Evaluation of Factors Affecting Rutting. *International Journal of Pavement Engineering*, Vol. 4, No. 1, 2003, pp. 59–68. <https://doi.org/10.1080/10298430310001593263>.
- Taylor, A.J. *Effect of Dwell Time and Lag Time on IDEAL-CT and IDEAL-RT Results—Expanded Study*. Consortium for Asphalt Pavement Research and Implementation, 2025.
- Taylor, A.J., Chen, C., and Moore, N. *Validation of Rapid Rutting Test Procedures Using Plant Mixes from the 2021 NCAT Test Track*. NCAT Report 26-01. National Center for Asphalt Technology, Auburn, AL, 2026.
- Tong, B., Huang, W., Habbouche, J., Boz, I., Guo, Q., Diefenderfer, S.D., and Flintsch, G.W. Machine Learning-Based Prediction and Optimization of Balanced Mixture Design Performance Indices. *Transportation Research Record*, Vol. 2679, No. 7, 2025, pp. 156–173. <https://doi.org/10.1177/03611981251324209>.
- Veeraragavan, R.K., Nener-Plante, D., Myers, L., Nash, C., and Tran, N.H. Balanced Mix Design Benchmarking of Field-Produced Asphalt Mixtures in Maine, U.S. *Transportation Research Record*, Vol. 2676, No. 5, 2022, pp. 263–276. <https://doi.org/10.1177/03611981211061552>.
- Virginia Department of Transportation. *Road and Bridge Specifications*. Richmond, VA, 2020.
- Virginia Department of Transportation. *Virginia Test Methods*. Richmond, VA, 2025.
- Wang, H., Al-Qadi, I.L., Baek, J., and Leng, Z. Rutting potential of warm mix asphalt and possible misleading results due to reheating. In *Proceedings of the International Conferences on the Bearing Capacity of Roads, Railways and Airfields*, 2013, pp. 627–636.
- Wang, Y.D., Underwood, B.S., and Kim, Y.R. Development of a fatigue index parameter, S_{app} , for Asphalt mixes using viscoelastic continuum damage theory. *International Journal of Pavement Engineering*, Vol. 23, No. 2, 2020, pp. 438–452. <https://doi.org/10.1080/10298436.2020.1751844>.

- Wang, Y.D., Ghanbari, A., Underwood, B.S., and Kim, Y.R. Development of a Performance-Volumetric Relationship for Asphalt Mixtures. *Transportation Research Record*, Vol 2673, No. 6, 2019, pp. 416–430. <https://doi.org/10.1177/0361198119845364>.
- Xiao, F., Putman, B., and Amirkhani, S. Plant and laboratory compaction effects on performance properties of plant-foamed asphalt mixtures containing RAP. *Journal of Materials in Civil Engineering*, Vol. 27, No. 9, 2015, 04014240. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0001202](https://doi.org/10.1061/(ASCE)MT.1943-5533.0001202).
- Yin, F., Taylor, A.J., and Tran, N. *Performance Testing for Quality Control and Acceptance of Balanced Mix Design*. NCAT Report 20-02. National Center for Asphalt Technology, Auburn University, Auburn, AL, 2020.
- Yin, F., Moore, N., Chen, C., and Taylor, A. Case Study on Using Warm Mix Asphalt at Reduced Production Temperatures for Balanced Mix Design. *Transportation Research Record*, Vol. 2679, No. 1, 2023, pp. 1064–1077. <https://doi.org/10.1177/03611981231214230>.
- Yin, F., West, R.C., Xie, Z., Taylor, A., and Julian, G. *Effects of Loading Rate and Mix Reheating on Indirect Tensile N_{flex} Factor and Semi-Circular Bend J-Integral Test Results to Assess the Cracking Resistance of Asphalt Mixtures*. NCAT Report 17-09. National Center for Asphalt Technology, Auburn University, Auburn, AL, 2017.
- Zaumanis, M., Poulidakos, L.D., and Partl, M.N. Performance-based design of asphalt mixtures and review of key parameters. *Materials and Design*, Vol. 141, 2018, pp. 185–201. <https://doi.org/10.1016/j.matdes.2017.12.035>.
- Zhou, F., Steger, R., and Mogawer, W. Development of a coherent framework for balanced mix design and production quality control and quality acceptance. *Construction and Building Materials*, Vol. 287, 2021, 123020. <https://doi.org/10.1016/j.conbuildmat.2021.123020>.