VTRC
Virginia Transportation Research Council

We Bring Innovation to Transportation

Mechanistic-Based Evaluation of Performance Thresholds for Balanced Mix Design Asphalt Surface Mixtures

https://vtrc.virginia.gov/media/vtrc/vtrc-pdf/vtrc-pdf/26-R08.pdf

JHONY HABBOUCHE, Ph.D., P.E.
Senior Research Scientist

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

STACEY D. DIEFENDERFER, Ph.D., P.E.
Associate Principal Research Scientist

Virginia Transportation Research Council

BENJAMIN SHANE UNDERWOOD, Ph.D.
Professor

OMAR OTHMAN
Graduate Research Assistant

North Carolina State University

Final Report VTRC 26-R08

VIRGINIA TRANSPORTATION RESEARCH COUNCIL
530 Edgemont Road, Charlottesville, VA 22903-2454
vtrc.virginia.gov

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 26-R08	2. Government Accession No.:	3. Recipient's Catalog No.:
4. Title and Subtitle: Mechanistic-Based Evaluation of Performance Thresholds for Balanced Mix Design Asphalt Surface Mixtures		5. Report Date: August 2025
		6. Performing Organization Code:
7. Author(s): Jhony Habbouche, Ph.D., P.E., Ilker Boz, Ph.D., P.E., Stacey Diefenderfer, Ph.D., P.E., Benjamin Shane Underwood, Ph.D., and Omar Othman.		8. Performing Organization Report No.: VTRC 26-R08
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.: 120747
12. Sponsoring Agencies' Name and Address: Virginia Department of Transportation, 1401 E. Broad Street, Richmond, VA 23219	Federal Highway Administration, 400 North 8th Street, Room 750, 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:

As part of the Virginia Department of Transportation (VDOT) balanced mix design (BMD) implementation process, VDOT selected three quick, simple, and practical empirical tests to evaluate asphalt mixtures against different modes of distress for design and acceptance. These tests include the Cantabro test, with a mass loss limit of 7.5%; the indirect tensile cracking test, requiring a minimum cracking tolerance index of 70; and the asphalt pavement analyzer (APA) rut test, with a maximum rut depth of 8.0 mm. These tests assess the potential performance in durability, cracking, and rutting of asphalt mixtures, respectively. In 2021, performance-based specifications were evaluated to fully implement BMD in Virginia. Additional mixture tests confirmed the reasonableness of the established performance criteria. However, further investigation was advised to assess correlations between the Cantabro, APA rut, and indirect tensile cracking, tests and fundamental tests, mechanistic-empirical (ME) simulations, and field performance, ensuring suitable threshold criteria for BMD implementation. This study's objectives were the following:

1. Compare the performance of VDOT's Superpave and BMD-designed mixtures through evaluation of empirical and advanced test results.
2. Assess the reliability of empirical tests with respect to advanced tests.
3. Validate and refine BMD performance thresholds by linking empirical and fundamental test results and incorporating asphalt mixture volumetrics into ME structural design, including establishing preliminary traffic-based performance thresholds for empirical tests.
4. Evaluate material properties' effect on long-term performance and the oxidative aging potential of Virginia surface mixtures with various reclaimed asphalt pavement contents.

The scope of the work included testing 18 surface mixtures—containing various reclaimed asphalt pavement contents, binder grades, recycling agents, fibers, and warm-mix additives—and ME simulations. This study highlights persistent knowledge and practical gaps in BMD despite significant development. Results indicate that VDOT's BMD tests effectively assess durability, rutting, and cracking performance, aligning with advanced tests measuring fundamental properties. The study showed that current BMD thresholds can be revised and refined, including incorporating traffic-based rutting and cracking thresholds. Specifically, ME simulations indicated that mixtures intended for moderate to heavy traffic pavement structures (D) should be designed to meet a maximum mass loss of 6.3% (instead of 7.5%), a minimum cracking tolerance index of 110 (instead of 70), while maintaining the same APA rut depth threshold of 8.0 mm. Mixtures intended for low-volume traffic pavement structures (A) should meet a maximum mass loss of 5.9% (instead of 7.5%), a minimum cracking tolerance index of 124, and a maximum APA rut depth of 10 mm.

The research team recommends that VDOT: (1) Continue using the Cantabro, APA rut, and indirect tensile cracking tests as part of the BMD framework for designing surface mixtures with A and D designations; (2) Continue efforts to validate and refine the BMD special provisions by incorporating traffic-based cracking and rutting performance thresholds; and (3) Consider benchmarking and establishing performance-based specifications specifically for mixtures intended for use on low-volume roads.

17. Key Words: BMD, Cantabro, IDT-CT, APA, dynamic modulus, CF, SSR, AASHTOWare Pavement ME Design, FlexPave™, correlation, traffic-based performance thresholds, specifications.	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: 74	22. Price:

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

FINAL REPORT

MECHANISTIC-BASED EVALUATION OF PERFORMANCE THRESHOLDS FOR BALANCED MIX DESIGN ASPHALT SURFACE MIXTURES

Jhony Habbouche, Ph.D., P.E.
Senior Research Scientist
Virginia Transportation Research Council

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

Stacey Diefenderfer, Ph.D., P.E.
Associate Principal Research Scientist
Virginia Transportation Research Council

Benjamin Shane Underwood, Ph.D.
Professor
North Carolina State University

Omar Othman
Graduate Research Assistant
North Carolina State University

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

Charlottesville, Virginia

August 2025
VTRC 26-R08

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.

ABSTRACT

As part of the Virginia Department of Transportation (VDOT) balanced mix design (BMD) implementation process, VDOT selected three quick, simple, and practical empirical tests to evaluate asphalt mixtures against different modes of distress for design and acceptance. These tests include the Cantabro test, with a mass loss limit of 7.5%; the indirect tensile cracking test, requiring a minimum cracking tolerance index of 70; and the asphalt pavement analyzer (APA) rut test, with a maximum rut depth of 8.0 mm. These tests assess the potential performance in durability, cracking, and rutting of asphalt mixtures, respectively. In 2021, performance-based specifications were evaluated to fully implement BMD in Virginia. Additional mixture tests confirmed the reasonableness of the established performance criteria. However, further investigation was advised to assess correlations between the Cantabro, APA rut, and indirect tensile cracking tests and fundamental tests, mechanistic-empirical (ME) simulations, and field performance, ensuring suitable threshold criteria for BMD implementation.

This study's objectives were the following:

1. Compare the performance of VDOT's Superpave and BMD-designed mixtures through evaluation of empirical and advanced test results.
2. Assess the reliability of empirical tests with respect to advanced tests.
3. Validate and refine BMD performance thresholds by linking empirical and fundamental test results and incorporating asphalt mixture volumetrics into ME structural design, including establishing preliminary traffic-based performance thresholds for empirical tests.
4. Evaluate material properties' effect on long-term performance and the oxidative aging potential of Virginia surface mixtures with various reclaimed asphalt pavement contents.

The scope of the work included testing 18 surface mixtures—containing various reclaimed asphalt pavement contents, binder grades, recycling agents, fibers, and warm-mix additives—and ME simulations.

This study highlights persistent knowledge and practical gaps in BMD despite significant development. Results indicate that VDOT's BMD tests effectively assess durability, rutting, and cracking performance, aligning with advanced tests measuring fundamental properties. The study showed that current BMD thresholds can be revised and refined, including incorporating traffic-based rutting and cracking thresholds. Specifically, ME simulations indicated that mixtures intended for moderate to heavy traffic pavement structures (D) should be designed to meet a maximum mass loss of 6.3% (instead of 7.5%), a minimum cracking tolerance index of 110 (instead of 70), while maintaining the same APA rut depth threshold of 8.0 mm. Mixtures intended for low-volume traffic pavement structures (A) should meet a maximum mass loss of 5.9% (instead of 7.5%), a minimum cracking tolerance index of 124, and a maximum APA rut depth of 10 mm.

The research team recommends that VDOT: (1) Continue using the Cantabro, APA rut, and indirect tensile cracking tests as part of the BMD framework for designing surface mixtures

with A and D designations; (2) Continue efforts to validate and refine the BMD special provisions by incorporating traffic-based cracking and rutting performance thresholds; and (3) Consider benchmarking and establishing performance-based specifications specifically for mixtures intended for use on low-volume roads.

TABLE OF CONTENTS

FINAL REPORT

MECHANISTIC-BASED EVALUATION OF PERFORMANCE THRESHOLDS FOR BALANCED MIX DESIGN ASPHALT SURFACE MIXTURES

Jhony Habbouche, Ph.D., P.E.
Senior Research Scientist
Virginia Transportation Research Council

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

Stacey Diefenderfer, Ph.D., P.E.
Associate Principal Research Scientist
Virginia Transportation Research Council

Benjamin Shane Underwood, Ph.D.
Professor
North Carolina State University

Omar Othman
Graduate Research Assistant
North Carolina State University

INTRODUCTION

Overview

Asphalt mixtures have evolved and increased in complexity because of the incorporation of unconventional materials such as reclaimed asphalt pavement (RAP), ground tire rubber, fibers, warm mix additives, recycling agents (RAs), polymers, and other additives. This evolution has led asphalt researchers and agencies to recognize that volumetric properties alone are incapable of fully characterizing the performance of asphalt mixtures in terms of cracking and rutting. Therefore, incorporating laboratory experimental assessments beyond recipe-type properties has become essential to achieve adequate mixture performance (West et al., 2018). Another constraint of the conventional volumetric design approach is a heavy reliance on the aggregate bulk specific gravity (G_{sb}). This property is not reliable because it can significantly change depending on the aggregate source, site geology, and mining procedures. Such variability in G_{sb} leads to conventional mixture designs having inaccurate volumetric properties and, consequently, imprecise performance characterization (Yin and West, 2021). Because of these concerns, among others, many state highway agencies opted to complement volumetric properties with performance testing through the process of balanced mix design (BMD). BMD is defined as "asphalt mixture design using performance tests on appropriately conditioned specimens that address multiple modes of distress taking into consideration mixture aging (shortand long-term), traffic, climate, and location within the pavement structure" in accordance with American Association of State Highway and Transportation Officials (AASHTO) PP 105 (2020c).

In 2017, researchers at the Virginia Transportation Research Council (VTRC) undertook an initial effort to benchmark the performance of several typical asphalt surface mixtures (SMs), designated as A and D mixtures, for traffic loads of 0 to 3 million equivalent single axle loads (MESALs) and 3 to 10 MESALs, respectively (Bowers and Diefenderfer, 2018; Bowers et al., 2022). These mixtures were produced and sampled in 2015. Three fast, simple, and practical but empirical performance-indicative tests addressing different modes of distress were selected for use as part of the BMD method. The Cantabro, indirect tensile cracking test (IDT-CT), and asphalt pavement analyzer (APA) rut tests were selected to assess the potential performance in durability, cracking, and rutting of asphalt mixtures, respectively. In addition, the Virginia Department of Transportation (VDOT) has long used the indirect tensile test (tensile strength ratio test), performed in accordance with AASHTO T 283, to evaluate the resistance of asphalt mixtures to moisture-induced damage (AASHTO, 2022c; Boz et al., 2021). Moreover, initial performance threshold criteria were developed for the selected tests in which the Cantabro mass loss (ML) was limited to a maximum of 7.5%, APA rut depth (RD) to a maximum of 8.0 mm, and the IDT-CT cracking tolerance (CT) index to a minimum of 70 (Bowers and Diefenderfer, 2018; Bowers et al., 2022). Following these findings, VDOT developed special provisions to design SMs with BMD (Diefenderfer et al., 2021a; 2021b). In 2021, the performance-based specifications were assessed and verified as part of several other concurrent efforts toward a successful full implementation of BMD in Virginia. Based on results from the additional mixtures tested in this study, the research team concluded that the performance criteria previously established were reasonable. However, further investigation is recommended to evaluate the relationship and correlations between the results of the Cantabro, IDT-CT, and APA tests and the outcomes of fundamental tests, mechanistic-empirical (ME) simulations and analyses, and field performance. This further investigation will help ensure application of the most appropriate threshold criteria for successful implementation of the BMD method (Diefenderfer et al., 2021a; 2021b).

VTRC recently completed an effort that proposed adopting the indirect tensile at high temperature (IDT-HT) test as a simpler and more practical substitute for the APA test (Boz et al., 2023). The proposed performance criterion for this test is 133 kPa for specimens conditioned in an environmental chamber or 100 kPa for specimens conditioned in a water bath without sealing specimens. The IDT-HT test has been introduced as part of the mix design in the 2023 BMD special provision solely for reporting purposes and will be included as a testing requirement in the 2024 BMD special provision. In 2024, the IDT-HT test will be used only for reporting purposes during production testing. The IDT-HT was not evaluated as part of this study.

Background

In general, the following three approaches can be considered to develop performance-based threshold criteria for BMD asphalt mixtures:

- Approach I subjects a pool of asphalt mixtures, typically designed and produced in a given region or climate, to a suite of practical but empirical tests considering various modes of distress. The results from these tests are then evaluated based on descriptive statistical parameters (e.g., minimum, maximum, or average value of the mixtures from a given test) or simple statistical tests (e.g., one-way t-test with a certain confidence level) to establish thresholds. A portion of asphalt mixtures designed and produced with respect to these performance thresholds are expected to outperform those designed and produced using conventional practices. VDOT adopted this approach in the initial establishment of BMD performance threshold criteria (Bowers and Diefenderfer, 2018; Bowers et al., 2022).
- Approach II uses empirical test results that are compared with results from fundamental tests to establish performance criteria. Based on this comparison, the empirical test results can be adopted and used as surrogate indices in lieu of the results from fundamental tests. Subsequently, the empirical test results can be considered into ME design, alongside the volumetric properties of asphalt mixtures. VDOT has been extensively building on the BMD initiative based on Approach I. The practical but empirical tests for durability, rutting, and cracking, along with their associated thresholds, have never been verified using Approach II. This project primarily focuses on this approach.
- Approach III directly compares empirical test results with the in-service performance of asphalt mixtures to establish performance thresholds. Although this approach is more definitive for establishing performance thresholds, it cannot always be easily considered because of the requirement to collect multiple years of in-service distress data, frequently test cores sampled from the field in the laboratory, and test pavement structures in-situ—both destructive and nondestructive—for years. However, this approach remains indispensable and will be used to verify and refine the established thresholds of the tests after numerous years of in-service use. For instance, VTRC Project 124332, Field Validation of Balanced Mix Design Initial Criteria, is ongoing and primarily focuses on this approach (Diefenderfer and Boz, 2025).

PURPOSE AND SCOPE

This project's first objective was to present a comparative analysis between the performance properties of VDOT mixtures designed solely based on Superpave volumetrics (referred to as control mixtures) and their corresponding BMD mixtures designed in accordance with the current BMD performance tests (referred to as BMD mixtures) using empirical and fundamental tests.

The second objective was to investigate whether fundamental and empirical tests yield consistent characterization of the evaluated mixtures in terms of performance ranking. This objective supports identifying potential discrepancies and establishing the reliability of empirical testing in predicting performance outcomes.

The third objective was to further validate or refine the preestablished BMD performance thresholds. This objective included establishing links between the results from empirical and fundamental tests and providing the incorporation of the volumetric properties of asphalt mixtures into ME structural pavement design. Finally, this objective supports establishing preliminary traffic-based performance thresholds for the empirical tests.

The fourth objective was to evaluate the implications of the material properties on long-term performance while also characterizing the oxidative aging potential of Virginia SMs with various RAP contents in accordance with the National Cooperative Highway Research Program (NCHRP) Project 09-54 method (Kim et al., 2018, 2021).

The scope of work consisted of laboratory testing and ME evaluation of 18 plant-produced SMs incorporating a range of RAP contents (conventional and high), binder grades, RAs, fibers, and warm mix additives.

METHODS

This effort included a literature review, selection and sampling of mixtures of interest, VDOT-selected BMD performance tests, identification and execution of fundamental tests, analysis of the volumetric and performance properties of mixtures and corresponding extracted and recovered binders, and ME simulations and analysis of the collected data. Figure 1 shows Part I of the experimental program, which primarily addresses the first objective of this study.

Literature Review

A comprehensive literature search was conducted to gather state-of-the-art information relevant to the objectives of this study. Various databases and search engines related to transportation engineering—such as TRID, Transportation Research Information Services, Scopus, the Catalog of Worldwide Libraries, Google® Scholar, ProQuest®, and Web of Science—were searched for relevant literature. The research team summarized and synthesized findings regarding the different test protocols proposed for designing asphalt mixtures and those mixtures' ability to resist distresses such as cracking and rutting based on BMD principles. The review also discussed test costs and state efforts to integrate performance tests into mixture design. Insights regarding the integration of performance tests into quality assurance practices were highlighted. In addition, a review of major outcomes from NCHRP Project 09-54, *Long-Term Aging of Asphalt Mixtures for Performance Testing and Prediction*, was provided (Kim et al., 2018, 2021). Furthermore, the review discussed the mechanistic frameworks and corresponding software available on the market, including AASHTOWare® Pavement ME Design and FlexPAVE™ (AASHTO, 2004; FHWA, n.d.). Finally, the review identified the knowledge gaps existing in the field at the time, which this report summarizes.

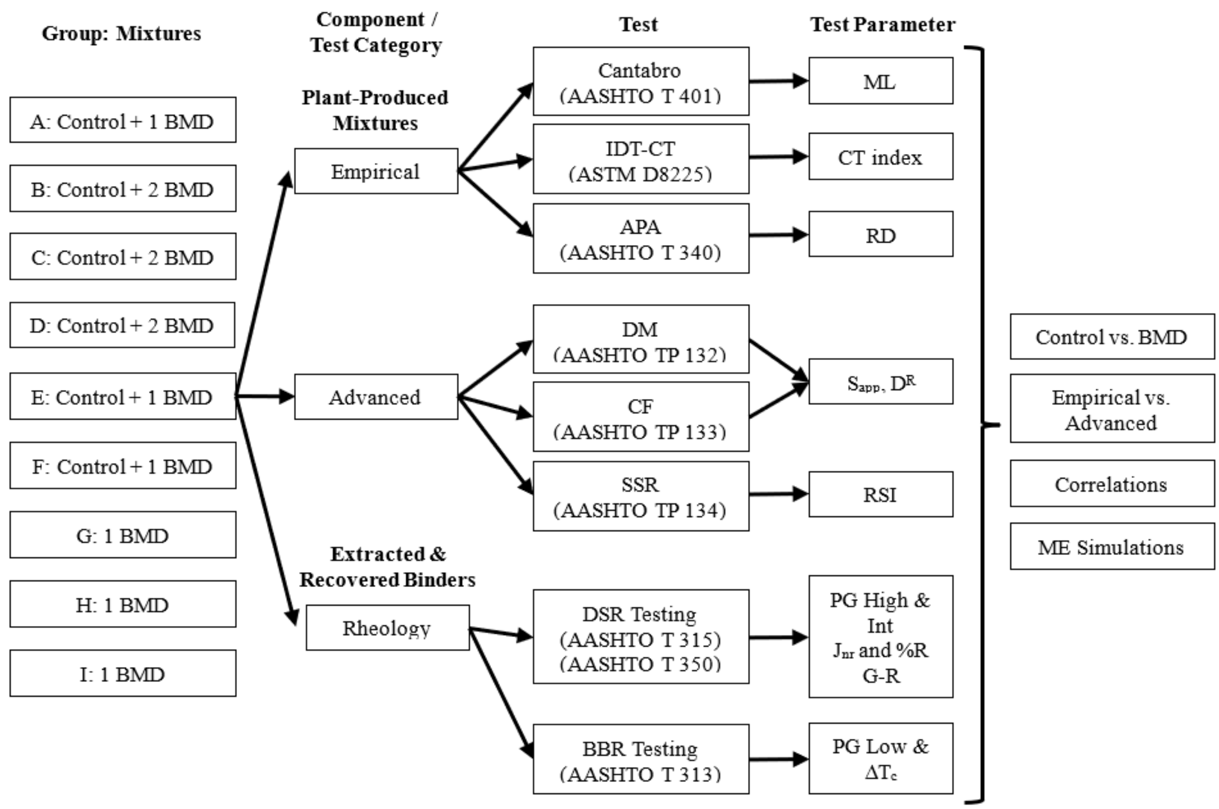


Figure 1. Experimental Program—Part I. ΔT_c = difference in critical low temperature; %R = percent recovery; AASHTO = American Association of State Highway and Transportation Officials; APA = asphalt pavement analyzer; ASTM = American Society for Testing Materials; BBR = bending beam rheometer; BMD = balanced mix design; CF = cyclic fatigue; CT = cracking tolerance; DM = dynamic modulus; D^R = pseudo-energy-based failure criterion; DSR = dynamic shear rheometer; G-R = Glover-Rowe parameter; IDT-CT = indirect tensile cracking test; Int = intermediate; J_{nr} = nonrecoverable creep compliance; ME = mechanistic-empirical; ML = mass loss; PG = performance grade; RD = rut depth; RSI = rutting strain index; S_{app} = cyclic fatigue index parameter; SSR = stress sweep rutting.

Mixtures

This study sampled 19 BMD and control asphalt SMs for evaluation that were produced and placed in Virginia during the 2019, 2020, 2021, and 2022 paving seasons. Table 1 shows a complete list of these mixtures. As mentioned previously, the mixtures incorporated various combinations of RAP contents, two binder performance grades (PGs), six RAs, one type of fiber, and two types of warm mix additives. These mixtures included the following:

- Six control dense-graded (DG) mixtures (Control).
- One BMD dense-graded mixture (BMD DG).
- One BMD dense-graded mixture with RA (BMD DG + RA).
- One BMD dense-graded mixture with a softer asphalt binder (LPG) (BMD DG + LPG).
- Two BMD high-RAP (HR) mixtures with RA (BMD HR + RA).
- Five BMD high-RAP mixtures with a softer asphalt binder (BMD HR + LPG).
- One BMD high-RAP mixture with a softer asphalt binder and RA (BMD HR + LPG + RA).

- One BMD high-RAP mixture with a softer asphalt binder, RA, and fiber (BMD HR + LPG + RA + F).

Dense-graded SMs contain less than or equal to 30% RAP content. HR SMs contain more than 30% RAP content, both by total weight of mixture. Out of the 5 BMD HR + LPG mixtures, this study uses only 4, bringing down the total of asphalt SMs to 18.

The 18 mixtures evaluated in this study were divided into nine major groups, each reflecting nine different field trials or scheduled pilot projects. Each group featured component materials from a single location with one nominal maximum aggregate size (NMAS), either 9.5 or 12.5 mm, and other compositional factors, including RAP contents, warm mix additives, asphalt binder PG, RAs (labeled here as RA1 through RA5), and aramid fibers.

The naming of the mixtures in Table 1 consists of a letter corresponding to the group (A through I). For the mixtures pertaining to groups A through F, a number indicating whether the mixture is a control or a BMD mixture was added. An identification (ID) containing the number 1 represents a control mixture, and an ID containing the numbers 2 or 3 represent a BMD mixture. Groups G, H, and I contain only one BMD mixture each. Therefore, the ID for those groups does not include any numbering. Because of limitations in available materials, mixture 14 was not used as part of this study.

Laboratory Evaluation of Asphalt Mixtures

Plant-produced loose mixture was collected from each mixture. Loose mixture was sampled from an approximately 3- to 5-ton quantity of mixture dumped on the ground at the plant and struck off using a loader. Loose mixture samples were placed into boxes, taken to VTRC and North Carolina State University, and stored in a climate-controlled area until further evaluated.

Specimens were fabricated from reheated loose mixture sampled in boxes during production. Reheated specimens were fabricated by reheating the loose mixture in boxes until workable, splitting the material into specimen quantities, heating to the appropriate compaction temperature, and compacting.

Table 1. Control and BMD SMs Considered for Evaluation

ID1	ID2	Mixture Type	Paving Season	District	Route	Details
1	A1	SM-9.5D 26% RAP PG 64S-22	2019	Salem	Route 460	Control
2	A2	SM-9.5D 26% RAP PG 64S-22 + RA1	2019	Salem	Route 460	BMD DG + RA
3	B1	SM-9.5D 30% RAP PG64S-22	2020	NOVA	US 50	Control
4	B2	SM-9.5D 40% RAP PG64S-22 with RA2	2020	NOVA	US 50	BMD HR + RA
5	B3	SM-9.5D 40% RAP PG58-28	2020	NOVA	US 50	BMD HR + LPG
6	C1	SM-12.5A 30% RAP PG64S-22	2020	Fredericksburg	SR 628	Control
7	C2	SM-12.5A 40% RAP PG64S-22 with RA3	2020	Fredericksburg	SR 628	BMD HR + RA
8	C3	SM-12.5A 40% RAP PG58-28	2020	Fredericksburg	SR 628	BMD HR + LPG
9	D1	SM-12.5A 30% RAP PG64S-22	2020	Richmond	SR 903	Control
10	D2	SM-12.5A 35% RAP PG58-28 with RA4	2020	Richmond	SR 903	BMD HR + LPG + RA
11	D3	SM-12.5A 35% RAP PG58-28 with RA5 + fiber	2020	Richmond	SR 903	BMD HR + LPG+ RA + F
12	E1	SM-12.5A 30% RAP PG64S-22	2020	Richmond	US 360	Control
13	E2	SM-12.5A 40% RAP PG58-28	2020	Richmond	US 360	BMD HR + LPG
14	NA	SM-12.5A 35% RAP PG58-28	2020	Richmond	SR 623	BMD HR + LPG
15	F1	SM-9.5D 30% RAP PG64S-22	2021	Salem	US 11	Control
16	F2	SM-9.5D 30% RAP PG64S-22	2021	Salem	US 11	BMD DG
17	G	SM-12.5D 30% RAP PG58-28	2021	Richmond	SR 627	BMD DG + LPG
18	H	SM-9.5D 40% RAP PG58-28	2021	NOVA	SR 287	BMD HR + LPG
19	I	SM-9.5A 40% RAP PG 64S-22 with RA 6	2022	NOVA	Route 2401	BMD HR + RA

A and D = mixture designations; BMD = balanced mix design; DG = dense-graded; F = fiber; HR = high RAP; ID = identification; LPG = softer asphalt binder; NA = not applicable, mixture was sampled but was not carried out in the study; NOVA = Northern Virginia; PG = performance grade; RA = recycling agent; RAP = reclaimed asphalt pavement; S = standard; SM = surface mixture; SR = state route.

Volumetric Properties and Aggregate Gradations of Mixtures

The theoretical maximum specific gravity of each mixture was determined in accordance with AASHTO T 209, *Standard Method of Test for Theoretical Maximum Specific Gravity (G_{mm}) and Density of Hot-Mix Asphalt (HMA)* (AASHTO, 2020b). The asphalt binder content of each mixture was determined by the ignition method in accordance with Virginia Test Method 102, *Determination of Asphalt Content from Asphalt Paving Mixtures by the Ignition Method* (Virginia Test Methods, 2013). The size distribution (gradation) of the recovered aggregate was determined in accordance with AASHTO T 11, *Standard Method of Test for Materials Finer Than 75-µm (No. 200) Sieve in Mineral Aggregates by Washing* (AASHTO, 2020e) and AASHTO T 27, *Standard Method of Test for Sieve Analysis of Fine and Coarse Aggregates* (AASHTO, 2020a). Loose mixtures were conditioned at the compaction temperature and then compacted to N_{design} gyrations (i.e., 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* (AASHTO, 2019d). Basic physical characteristics and volumetric parameters in terms of bulk specific gravity (G_{mb}), voids in total mixture, voids in mineral aggregate, voids filled with asphalt, fines to aggregate ratio, aggregate effective specific gravity, aggregate bulk specific gravity, absorbed asphalt binder content, effective asphalt binder content, and effective film thickness were determined.

Balanced Mix Design Performance Tests

Cantabro Test

The Cantabro ML was determined at room temperature (i.e., approximately 25°C) to evaluate the durability of asphalt mixtures in accordance with AASHTO T 401, *Standard Method of Test for Cantabro Abrasion Loss of Asphalt Mixture Specimens* (AASHTO, 2022b). Volumetric pills, 150 mm in diameter by 115 ± 5 mm in height compacted to N_{design} gyrations were used in this test. A lower ML indicates increased durability.

Indirect Tensile Cracking Test

The IDT-CT was conducted at 25°C in accordance with 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). All test specimens were 150 mm in diameter by 62 mm in height and were tested at 7.0 ± 0.5% air voids. The CT index was calculated from the test load-displacement curve collected during testing. A higher CT index value indicates greater resistance to cracking.

Asphalt Pavement Analyzer Rut Test

The APA rut test was performed 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, 2019a). All test specimens were 150 mm in diameter by 75 mm in height and tested at 7.0 ± 0.5% air voids. After 8,000 cycles were applied at a temperature of 64°C, the deformation of the specimen was measured. A lower APA RD indicates greater resistance to rutting.

Advanced Performance Tests

Dynamic Modulus |E| Test*

The dynamic modulus (E*) test measures the stiffness of asphalt mixtures. Testing was conducted using the Asphalt Mixture Performance Tester (AMPT) in accordance with AASHTO TP 132, *Standard Method of Test for Determining Dynamic Modulus for Asphalt Mixtures Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT)* (AASHTO, 2019b). The test specimens were 38 mm in diameter by 110 mm in height and cored from a Superpave gyratory compacted specimen 150 mm in diameter by 180 mm in height. All tests were conducted in the uniaxial mode without confinement. All test specimens were compacted to 5.0 ± 0.5% air voids.

Stress Sweep Rutting Test

The stress sweep rutting (SSR) test assesses the rutting susceptibility of asphalt mixtures by applying repeated cyclic loading to confined cylindrical test specimens at two location-specific temperatures, low and high, (i.e., 23°C and 48°C) in accordance with AASHTO TP 134, *Standard Method of Test for Stress Sweep Rutting (SSR) Test Using Asphalt Mixture Performance Tester (AMPT)* (AASHTO, 2021). The SSR test specimens were 100 mm in diameter by 150 mm in height and cored from the center of a Superpave gyratory compacted specimen 150 mm in diameter by 180 mm in height. All test specimens were compacted to an air void level of 7.0 ± 0.5% air voids. The SSR test results were used to calculate the average permanent strain (in percent) and produce the rutting strain index (RSI) parameter calculated by the FlexMAT™ Rutting analysis (Version 2.1.4.3) (FHWA, 2021b, 2019c). Test results were also used to generate a permanent strain shift model that can be used with the FlexPAVE analysis to model rutting in the pavement layer. A lower RSI indicates relatively more resistance to rutting (FHWA, n.d., 2021b).

Repeated Load Permanent Deformation Test—Confined Flow Number Test

The rutting characteristics of specimens prepared from 11 selected loose mixtures (i.e., B1, B2, B3, C1, C2, C3, D1, D2, F1, H, and I) collected during construction were evaluated using the repeated load permanent deformation (RLPD) test in accordance with NCHRP Report 719, *Calibration of Rutting Models for Structural and Mix Designs* (Von Quintus et al., 2012). The repeated load triaxial test specimens were 100 mm in diameter by 150 mm in height and cored from the center of a Superpave gyratory compacted specimen 150 mm in diameter by 180 mm in height. All test specimens were compacted to an air void level of 7.0 ± 0.5% air voids. A repeated haversine axial compressive load pulse of 0.1 second every 1.0 second was applied to the specimens. The tests were performed in the confined mode using a deviator stress of 482.6 kPa. Air is used to supply the confining pressure, and the pressure was constant throughout the test at 68.9 kPa. The tests were continued for 10,000 cycles or a permanent strain of 10%, whichever came first. The accumulated permanent deformation was recorded from the actuator displacement at the end of each loading cycle. The RLPD test included a specimen load conditioning sequence before collecting the data for the permanent strain curve (100 load cycles using confining pressure of 68.9 kPa, repeated deviator stress of 48.3 kPa, and contact deviator stress of 2.4 kPa). The axial deformation after each pulse was measured, and the axial resilient strain (ε_r) was calculated. In addition, the cumulative permanent strain (ε_p) was calculated.

Direct Tension Cyclic Fatigue Test

The simplified viscoelastic continuum damage test, known as the direct tension cyclic fatigue (CF) test, was performed at 21°C using AMPT in accordance with AASHTO T 411, *Standard Method of test for Determining the Damage Characteristic Curve and Failure Criterion Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test* (AASHTO, 2023). The CF test was performed on specimens 38 mm in diameter by 110 mm in height cored from Superpave gyratory samples 150 mm in diameter by 180 mm in height compacted from loose mixtures collected during construction. All test specimens were compacted to 5.0 ± 0.5% air voids. The developed damage characteristic curves were then used with the viscoelastic material properties (i.e., |E*|) to obtain the fatigue behavior of the asphalt mixtures. Three key test outcomes were obtained from the CF tests: (1) the damage characteristic curve, also referred to as the material integrity (C) versus damage (S) curve; (2) the pseudo-energy-based failure criterion (D^R); and (3) the apparent damage capacity, CF index parameter (S_{app}). The last parameter measures the amount of fatigue damage the material can tolerate considering the effect of the material's toughness and modulus. A higher S_{app} indicates higher fatigue cracking resistance. The calculation was conducted with FlexMAT Cracking (Version 2.1.3b), a Microsoft® Excel®-based tool provided by FHWA (2019a, 2019b).

Testing and Characterization of Extracted and Recovered Asphalt Binders

Extraction and Recovery

Testing was performed on extracted and recovered asphalt binders from the mixtures collected at the plant. Extraction of asphalt binder from collected mixtures was performed in accordance with AASHTO T 164, *Standard Method of Test for Quantitative Extraction of Asphalt Binder from Hot Mix Asphalt (HMA)*, Method A (AASHTO, 2018) using n-propyl bromide as the solvent. The asphalt binder was then recovered from the solvent using the Rotavap recovery procedure specified in AASHTO T 319, *Standard Method of Test for Quantitative Extraction and Recovery of Asphalt Binder from Asphalt Mixtures* (AASHTO, 2019e).

Performance Grading and Multiple Stress Creep Recovery Testing

Asphalt binder performance grading was performed in accordance with AASHTO M 320, *Standard Specification for Performance-Graded Asphalt Binder* (AASHTO, 2017), and AASHTO M 332, *Standard Specification for Performance-Graded Asphalt Binder Using Multiple Stress Creep Recovery (MSCR) Test* (AASHTO, 2019e).

Difference in Critical Low-Temperature Performance Grade Limiting Temperatures

The difference in critical low-temperature PG limiting temperatures, commonly referred to as ΔT_c, was calculated by subtracting the m-critical low temperature ($T_{c,m}$) from the S-critical low temperature ($T_{c,S}$), as Equation 1 shows (FHWA, 2021a). Both temperatures were determined using the bending beam rheometer data in accordance with AASHTO T 313 (AASHTO, 2019c). The m-critical low temperature ($T_{c,m}$) is the resulting low temperature at which the creep relaxation m-value at 60 seconds of loading is exactly equal to the specification value of 0.300. The S-critical low temperature ($T_{c,S}$) is the resulting low temperature at which the creep stiffness S-value at 60 seconds of loading is exactly equal to the specification value of 300 MPa.

$$\Delta T_c = T_{c,S} - T_{c,m} \qquad \text{(Eq. 1)}$$

Frequency Sweep

Frequency sweep tests were conducted to evaluate the extracted and recovered asphalt binders from the collected mixtures at the plant during production. No standard currently exists for the construction of a binder master curve. In this study, the rheological Rheology Analysis Software package was used to perform the shifting of the G* master curves to a reference temperature of 45°C (Abatech, 2022). The software adopts the method of free shifting to fit the frequency sweep measured data into a smooth master curve. The term “free shifting” indicates that the measured data are shifted to a master curve without a predefined shape function (Habbouche et al., 2022).

Glover-Rowe Parameter

The Glover-Rowe (G-R) parameter was originally defined by Glover et al. (2005) and reformulated for greater practical use by Rowe et al. (2011) in a discussion (Anderson et al., 2011). The G-R parameter is determined at a temperature of 15°C and a frequency of 0.005 rad/s and is expressed using Equation 2. The G-R parameter captures both rheological parameters needed to characterize binder viscoelastic behavior: stiffness (represented by the complex shear dynamic modulus G*) and relaxation (represented by the phase angle δ). The G-R parameter refers to nonload cracking at intermediate temperature, and its limits relate to specific environmental conditions. The universal limits of the G-R parameter, 180 kPa and 600 kPa, are usually used as a reference to track the effect of aging and relative performance evaluation of binders.

$$G - R = \frac{G^*(cos\delta)^2}{sin\delta} \qquad \text{(Eq. 2)}$$

Analysis Approaches for Performance Test Results

Performance of Balanced Mix Design Versus Control Mixtures

Mixtures were compared based on average performance test results of evaluated asphalt mixtures, and in addition, statistical analyses at 95% confidence interval were performed to verify the presence of statistically significant differences between BMD and their corresponding control mixtures.

Performance Discrimination Potential of Tests

The ability of the test methods to discriminate or differentiate the performance of the mixtures in terms of durability and resistance to rutting and cracking was assessed through statistical analysis. The one-way analysis of variance with Tukey's multiple comparison method at a 95% confidence interval was used for that purpose. Before the analysis of variance, the data for each test were checked for the assumptions of normality and equal variances at a 95% confidence interval.

Correlation Analysis

The results obtained from the BMD tests were compared with those obtained from the fundamental tests to investigate the existence of potential correlations among them. Simple regression analyses and scatter plots were used to quantify and visualize any correlations among the test results. In addition, the results obtained from the binder tests, conducted on the binders extracted and recovered from the mixtures, were compared with both BMD and fundamental test results to investigate any correlations between binder properties and mixture performance. Similar statistical analyses were performed to assess the strength and significance of any observed correlations.

Mechanistic-Empirical Simulations

Laboratory-measured performance metrics and mixture volumetric properties can be coupled in an ME analysis framework, which is a vital step to quantify and evaluate the effect of using conventional and BMD mixtures on the overall performance of pavements. Two well-known frameworks were considered in this study: AASHTOWare Pavement ME Design and FlexPAVE software (AASHTO, 2022a; FHWA, n.d.).

AASHTOWare Pavement ME Design

Background

The *Guide for the Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures* (MEPDG) (AASHTO, 2015; Oldis et al., 2004) was written as part of NCHRP Project 1-37A to provide the highway community with a cutting-edge tool for designing new and rehabilitated pavement structures. MEPDG employs a calculated mechanistic response, combined with empirical results from pavement test sections in the Long-Term Pavement Performance Program, to predict the performance of pavement structures. MEPDG determines pavement responses, such as stresses, strains, and deflections, based on inputs such as traffic, climate, and material parameters, to forecast the pavement damage over time for asphalt pavements. Subsequently, transfer functions relate computed pavement responses, such as pavement damage, to observed pavement distresses at high reliability values (90–100%). MEPDG principles were integrated into AASHTOWare Pavement ME Design (AASHTO, 2022a).

Pavement Structures

In this study, two main pavement structures were considered—Structure for A traffic subjected to 3 MESALs (Figures 2a and 2b) and Structure for D traffic subjected to 10 MESALs of traffic. Structure A, shown in Figures 2a and 2b, consisted of three asphalt layers: a 1.5-inch SM layer for 9.5-mm NMAS mixtures (Figure 2a) or a 2.0-inch SM layer for 12.5-mm NMAS mixtures (Figure 2b), a 2-inch intermediate mixture layer, and a 4-inch base mixture layer. Structure D, shown in Figures 2c and 2d, also consisted of three asphalt layers: a 1.5-inch SM layer for 9.5-mm NMAS mixtures (Figure 2c) or 2.0-inch SM layer for 12.5-mm NMAS mixtures (Figure 2d), a 2-inch intermediate mixture layer, and a 6-inch base mixture layer. Both structures were designed to top a 6-inch 21-B base layer and an infinite subgrade.

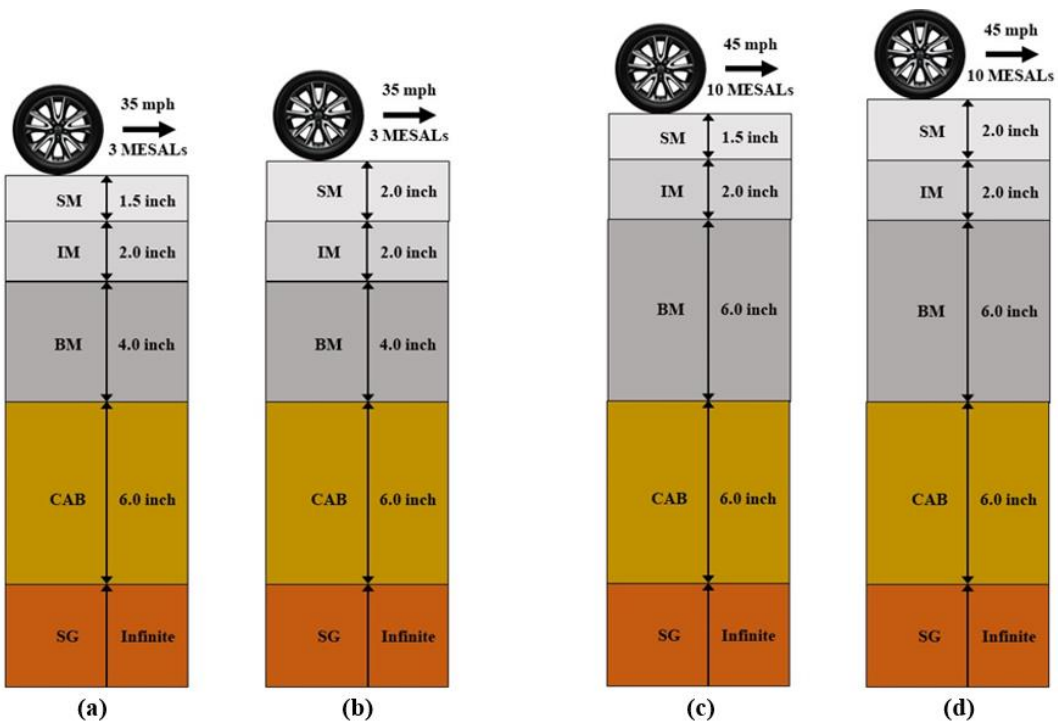


Figure 2. Pavement Structures for Mechanistic-Empirical Simulations: (a) Structure A with 1.5-Inch SM; (b) Structure A with 2-Inch SM; (c) Structure D with 1.5-Inch SM; (d) Structure D with 2-Inch SM. BM = base mixture; CAB = crushed aggregate base; IM = intermediate mixture; MESALs = million equivalent single-axle loads; SG = subgrade; SM = surface mixture.

Traffic

Traffic simulations for both Structures A and D were conducted for all evaluated mixtures, irrespective of their labeling according to the Job-Mix Formula.

Speed

Speeds of 25, 35, and 45 mph were considered for Structure A, and speeds of 35, 45, and 55 mph were considered for Structure D.

Climate

Simulations used data from climatic stations closest to the locations where the evaluated mixtures were placed. In addition, simulations were conducted using data from climatic stations with the warmest and coldest data in Virginia.

Density

A typical 93% in-place density (7.0% in-place air voids) was assumed for the SM layer. However, acknowledging the effect of using BMD in increasing in-place density, simulations were conducted at 95% in-place density (5.0% in-place air voids), which involved adjusting all volumetric properties and dynamic modulus (DM) of the SM layer to be suitable for either 5 or 7% in-place air voids using the DM close-form model developed by Sadat Sakhaei Far (2011).

Performance Properties of Evaluated Surface Mixtures

VDOT has adopted AASHTOWare Pavement ME Design for routine pavement design of new construction for interstate and primary routes (AASHTO, 2022a). Part of VDOT's implementation strategy was performing local calibration of the design software (AASHTOWare Pavement ME Design Version V2.2.6), primarily focusing on fatigue cracking and rutting distress modes. The local calibration factors (k_{r1}, k_{r2}, and k_{r3}) were identical to the MEPDG default global field calibration parameters for rutting included in the MEPDG Manual of Practice (AASHTO, 2015). These factors are referred to as VDOT_Default and were developed using unconfined RLPD testing.

However, concerns arose that the developed calibration factors could be insensitive when mixture types (e.g., BMD) that were not part of the calibration pool were used. Given these concerns, AASHTOWare Pavement ME Design (AASHTO, 2022a) analyses for the selected mixtures were performed using Level 1 inputs (dynamic modulus |E*|, mixture volumetric properties [i.e., asphalt content, air voids, and unit weight]) and (1) VDOT's recommended local calibration coefficients and (2) material-specific calibration coefficients.

In the case of material-specific calibration coefficients, the confined RLPD test was conducted at three different temperatures: 86, 104, and 122°F (30, 40, and 50°C) for all evaluated mixtures. A rutting laboratory model for each mixture was developed following Equation 3 based on the approach recommended in MEPDG (AASHTO, 2015).

$$\frac{\varepsilon_p}{\varepsilon_r} = 10^{k_{r1}} * (T)^{k_{r2}} * (N)^{k_{r3}} \qquad \text{(Eq. 3)}$$

Where:
ε_p = permanent axial strain, inch/inch (or mm/mm).
ε_r = resilient axial strain, inch/inch (or mm/mm).
N = number of loading cycles.
T = temperature of the asphalt mixture in °F.
k_{r1}, k_{r2}, and k_{r3} = experimentally determined coefficients.

Moreover, the coefficients (k_{f1}, k_{f2}, and k_{f3}) of the empirical fatigue law model (Equation 4) are historically calibrated from the bending beam fatigue test. However, because of material and time limitations, the bending beam fatigue test was not conducted. Therefore, the research team had to seek alternative methods to determinate an estimate of these model coefficients using some of the data collected as part of this project. Figure 3 outlines the 7-step procedure to derive the material-specific fatigue coefficients from the results of the CF test. A detailed explanation of this procedure is available in Appendix A.

$$N_f = k_{f1} * \left(\frac{1}{\varepsilon_t}\right)^{k_{f2}} * \left(\frac{1}{E_{AC}}\right)^{k_{f3}} \quad \text{(Eq. 4)}$$

Where:
N_f: fatigue life, number of load repetitions to fatigue damage.
ε_t: applied tensile strain, inch/inch (or mm/mm).
E_{AC}: dynamic modulus of the asphalt mixture, psi.
k_{f1}, k_{f2}, and k_{f3}: experimentally determined coefficients.

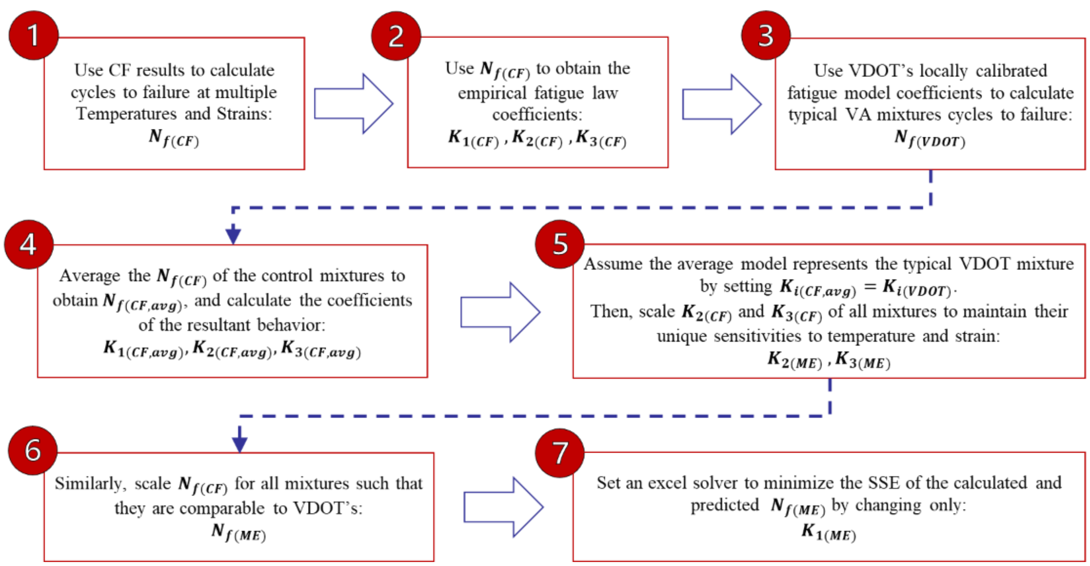


Figure 3. Summary Procedure to Derive Bending Beam Fatigue Model Coefficients from CF Data. avg = average; CF = cyclic fatigue; ME = mechanistic-empirical.

Correlation Analysis—AASHTOWare Pavement ME Design

APA RD values were compared with the total RD from AASHTOWare Pavement ME Design after 15 years of analysis (AASHTO, 2022a). VDOT's *AASHTOWare Pavement ME User Manual* sets a limit on the total rutting of 0.26 inch after 15 years (VDOT, 2017). In contrast, CT index values from IDT-CTs were compared with bottom-up fatigue cracking (BUFC) from AASHTOWare Pavement ME Design after 30 years. Although this comparison corresponds mainly to the damage the base mixture incurs, the cracking analysis was presented only for the sake of comparison. VDOT's *AASHTOWare Pavement ME User Manual* sets a limit on BUFC of 6% after 30 years.

FlexPAVE

Background

FlexPAVE is a specialized finite element program designed to predict the performance of asphalt pavement throughout its lifespan, specifically regarding fatigue and rutting (FHWA, n.d.). To achieve accurate predictions, the program considers various factors, including the three-dimensional response of moving loads, viscoelastic material properties, and climate data. In addition, the program imports data from FlexMAT Cracking and FlexMAT Rutting (FHWA, 2019b, 2019c), volumetric information, and details on unbound materials, enabling the program to establish materials equivalent to those used on typical pavement structures in Virginia. FlexPAVE also integrates a pavement temperature database, which accommodates the weather conditions on the site.

Factors Considered

FlexPAVE simulations were conducted using a subset of conditions used for the AASHTOWare Pavement ME Design simulations (AASHTO, 2022a; FHWA, n.d.). In this case, all the mixtures were analyzed assuming that each could be used on A or D structures. Factors considered in the simulations included the structure type (A or D), with A structures subjected to 3 MESALs and a 35 mph speed, and D structures subjected to 10 MESALs and a 45 mph speed. The same pavement structures shown in Figure 2 were used for the FlexPAVE simulations. The 18 tested SMs were characterized using the results from DM and CF testing analyzed with FlexMAT Cracking (Version 2.1.3b) and SSR testing analyzed with FlexMAT Rutting (Version 2.1.4.3) (FHWA, 2019b, 2019c). The Virginia-specific, Level 3 aging model was employed to obtain the aging parameters required for FlexPAVE aging analysis. Appendix C presents more information regarding this model.

Other than SMs, the characteristics of all layers were representative of typical VDOT materials and kept the same for all simulations. For climate selection, the closest North American Regional Reanalysis (NARR) station was used for each mixture. FlexPAVE analysis results in Total %Damage, Top %Damage, Bottom %Damage, Total Rut Depth, and the Rut Depth for each layer separately (FHWA, n.d.). The Top %Damage is confined to the damage only in the top one-third of the asphalt layer structure, whereas Bottom %Damage focuses on the bottom two-thirds of the asphalt layer structure.

Correlation Analysis—FlexPAVE

CT index values from the IDT-CTs were compared with the Top %Damage from FlexPAVE because it mostly corresponds to the damage incurred by SM (FHWA, n.d.). In contrast, the RD from the APA tests were compared with the Total Rut Depth from FlexPAVE to remain consistent with VDOT's *AASHTOWare Pavement ME User Manual*, which places a limit on the total rutting (VDOT, 2017). The lifespan was set at 10 years, because VDOT typically replaces asphalt roadway SMs (mill and fill maintenance) on an 8- to 12-year schedule.

RESULTS AND DISCUSSION

Literature Review—Knowledge Gaps

In the early 2000s, evidence of inadequate pavement performance, the incorporation of recycled asphalt materials into asphalt mixtures, and other factors renewed the interest in integrating performance tests in mixture design protocols, also known as performance mixture design. Multiple research efforts have been conducted to allow for successful implementation. In NCHRP Project 09-19 (Witczak et al., 2002), simplified performance test methods addressing permanent deformation, fatigue cracking, and low-temperature cracking were proposed as a final stage in the Superpave volumetric mixture design. However, the test protocols and attention to specific distress have varied across state agencies throughout recent years.

In the early implementation of the Superpave mixture design, more attention was given to improving rutting resistance, and many state agencies incorporated rutting test requirements. The adjustments made to the mixture designs to meet rutting specifications affected the durability of the mixture. Furthermore, many agencies indicated that cracking and raveling have become the main distresses controlling the service lives of asphalt pavements (West et al., 2018). This observation has driven national attention to seek new approaches that balance multiple modes of distress—BMD.

The literature review indicates many efforts to study different rutting and fatigue cracking protocols for integration into BMD have been made. However, aspects related to the selection of test methods and BMD implementation are still not fully elucidated, including the following aspects:

- No clear indication exists regarding which rutting and cracking test is most adequate for adoption by state agencies, considering the tradeoffs between simplicity, cost, degree of fundamental characterization, and any unique aspects in a particular state.
- Studies evaluating different test methods mainly focus on one category at a time. However, a need exists to simultaneously evaluate cracking and rutting test protocols to identify compatible methods and potential optimization of the BMD process when similar specimen geometries and test equipment are used for both cracking and rutting characterization.
- BMD test methods can provide an indication of the material's performance. However, most proposed protocols for rutting and cracking lack a mechanically sound methodology to link test results with fundamental engineering properties, which is essential for the practical integration of mixture design and pavement design.
- No explicit indication exists determining which of the AASHTO PP 105 (AASHTO, 2020c) approaches is most efficient at improving the overall performance of asphalt concrete mixtures (West et al., 2018; Yin and West, 2021). Approaches A and B are more conservative, retaining volumetric-based design and adding performance verification or optimization. On the other hand, Approaches C and D can provide more freedom to achieve the best performance of a selected test and consistency over production.
- As of yet, no agreed-on best practice exists for integrating performance testing into quality assurance. However, ongoing efforts are exploring some key issues. In general, all

insights agree that the simplicity of testing and turnaround time are critical considerations.

- No comprehensive studies have been completed relating index tests results to performance predictions from mechanistic-based analysis and the field.
- No universally set performance thresholds exist for BMD tests because the interpretation of results may vary for different locations and mixtures. Thus, determining and verifying adequate thresholds for each agency is necessary.

Testing and Characterization of Asphalt Mixtures

Volumetric Properties and Aggregate Gradations of Mixtures

Tables 2 and 3 summarize the volumetric properties and aggregate gradations determined for the evaluated mixtures. Nine mixtures with NMAS of 9.5 mm were evaluated, and the remaining nine had an NMAS of 12.5 mm. The binder contents ranged from 5.13 to 6.32% across the evaluated mixtures, with an average of 5.75%. All control mixtures (A1, B1, C1, D1, E1, and F1) contained either 26 or 30% RAP and were produced using PG 64S-22 asphalt binder. These control mixtures were designed following conventional Superpave methodology. The BMD mixtures contained varying percentages of RAP (26%, 30%, 35%, or 40%) and were produced with either PG 64S-22 or PG 58-28 binder. Five different RAs and one combination of fibers and a softening oil were used in these mixtures. Eight of the 12 BMD mixtures (A2, B2, B3, C2, C3, D2, D3, and E2) were designed following BMD Approach D (performance only). The remaining four BMD mixtures (F2, G, H, and I) were designed following a modified version of BMD Approach A, which allows for wider bands in terms of aggregate gradation and volumetric properties.

The gradations, reported in Tables 2 and 3, were measured during production, and although some samples were not within the design limits, most were within VDOT's four samples' acceptable range for production. The majority of deviations from production tolerances were for BMD mixtures designed following Approach D, which grants the most flexibility on volumetrics and gradations as long as performance tests' criteria are met. A significant decrease in binder content was noted for mixture A2 compared with control mixture A1. Aggregate gradations showed no significant differences between both mixtures. Slightly higher binder contents were observed for mixtures B2 and B3 compared with B1, possibly because of mixtures B2 and B3's higher RAP content and the design method used (Superpave versus BMD). Mixtures C2 and C3 exhibited a higher percent passing the No. 200 sieve compared with C1. Mixture D2 showed a notably higher binder content and percent passing the No. 200 sieve compared with D1 and D3, which had similar binder contents. BMD mixture E2 demonstrated lower binder content compared with control mixture E1. Conversely, BMD mixture F2 exhibited a higher binder content compared with control mixture F1.

Table 2. Volumetric Properties and Gradation for Evaluated Mixtures—Group A through C

Mixture Identification	A1	A2	B1	B2	B3	C1	C2	C3
Mixture Type	D	D	D	D	D	A	A	A
Description	Control	BMD DG + RA	Control	BMD HR + RA	BMD HR + LPG	Control	BMD HR + RA	BMD HR + LPG
Composition								
RAP Content, %	26	26	30	40	40	30	40	40
Asphalt Binder	PG 64S-22	PG 64S-22	PG 64S-22	PG 64S-22	PG 58-28	PG 64S-22	PG 64S-22	PG 58-28
Additives	WMA	WMA + RA1	WMA	WMA + RA2	WMA	WMA	WMA + RA3	WMA
Property								
N_{design}, gyrations	50	50	50	50	50	50	50	50
NMAS, mm	9.5	9.5	9.5	9.5	9.5	12.5	12.5	12.5
Asphalt Content, %	6.03	5.13	5.29	5.43	5.61	5.20	5.27	5.43
Rice SG (G_{mm})	2.597	2.631	2.688	2.689	2.681	2.688	2.694	2.677
VTM, %	2.3	4.3	3.6	3.9	3.0	4.3	3.1	1.6
VMA, %	16.5	16.3	16.8	17.2	16.9	15.0	15.9	15.0
VFA, %	86.3	73.5	78.7	77.5	82.3	71.4	80.6	89.3
FA Ratio	1.10	1.28	1.11	1.18	1.07	1.19	1.21	1.12
Mixture Bulk SG (G_{mb})	2.538	2.518	2.591	2.585	2.601	2.573	2.611	2.634
Aggregate Effective SG (G_{se})	2.878	2.873	2.953	2.963	2.963	2.948	2.961	2.948
Aggregate Bulk SG (G_{sb})	2.858	2.853	2.950	2.953	2.953	2.868	2.943	2.930
Absorbed Asphalt Content (P_{ba}), %	0.25	0.25	0.04	0.12	0.12	0.97	0.21	0.21
Effective Asphalt Content (P_{be}), %	5.79	4.89	5.26	5.32	5.49	4.28	5.07	5.22
Gradation / Sieve Size	**% Passing**							
¾ in (19.0 mm)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
½ in (12.5 mm)	98.9	99.8	100.0	99.8	100.0	97.3	96.7	95.7
3/8 in (9.5 mm)	93.4	93.8	93.0	93.4	93.1	86.9	87.8	83.9
No. 4 (4.75 mm)	63.0	63.8	59.4	70.0	62.9	56.5	60.3	50.5
No. 8 (2.36 mm)	42.3	42.8	39.3	48.0	40.8	36.9	41.5	33.5
No. 16 (1.18 mm)	28.9	30.2	27.8	33.8	28.6	26.4	29.7	24.6
No. 30 (600 µm)	19.8	20.8	19.6	23.6	20.3	19.5	21.6	18.4
No. 50 (300 µm)	13.0	13.4	13.5	16.0	13.9	13.9	15.3	13.5
No. 100 (150 µm)	8.8	8.9	9.0	10.0	9.0	8.7	10.1	9.2
No. 200 (75 µm)	6.4	6.3	5.8	6.3	5.9	5.1	6.1	5.9

A and D = mixture designation; BMD = balanced mix design; DG = dense-graded; FA = fines to aggregate; HR = high RAP content mixture; LPG = softer asphalt binder; NMAS = nominal maximum aggregate size; PG = performance grade; RA = recycling agent; RAP = reclaimed asphalt pavement; S = standard traffic; SG = specific gravity; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture; WMA = warm mix additive.

Table 3. Volumetric Properties and Gradation for Evaluated Mixtures—Group D through I

Mixture Identification	D1	D2	D3	E1	E2	F1	F2	G	H	I
Mixture Type	A	A	A	A	A	D	D	D	D	A
Description	Control	BMD HR + LPG + RA	BMD HR + LPG + RA + F	Control	BMD HR + LPG	Control	BMD DG	BMD DG + LPG	BMD HR + LPG	BMD HR + RA
Composition										
RAP Content, %	30	35	35	30	40	30	30	30	40	40
Asphalt Binder	PG 64S-22	PG 58-22	PG 58-22	PG 64S-22	PG 58-28	PG 64S-22	PG 64S-22	PG 58-28	PG 58-28	PG 64S-22
Additives	WMA	WMA+ RA4	WMA + RA 5 + F	WMA	WMA	WMA	WMA	WMA	WMA	WMA + RA6
Property										
N_{design}, gyrations	50	50	50	50	50	50	50	50	50	50
NMAS, mm	12.5	12.5	12.5	12.5	12.5	9.5	9.5	12.5	9.5	9.5
Asphalt Content, %	5.88	6.20	5.88	5.90	5.48	5.89	6.52	5.64	6.11	5.53
Rice SG (G_{mm})	2.472	2.462	2.477	2.520	2.516	2.469	2.464	2.473	2.620	2.718
VTM, %	2.3	1.3	1.7	1.6	3.0	2.0	1.8	5.2	2.1	3.5
VMA, %	15.8	15.1	14.9	15.0	15.7	15.7	17.0	17.3	17.1	16.9
VFA, %	85.4	91.5	88.7	89.6	81.1	87.4	89.5	70.0	87.9	79.6
FA Ratio	1.02	1.11	1.11	1.32	1.09	1.05	1.01	1.02	1.12	1.21
Mixture Bulk SG (G_{mb})	2.415	2.430	2.436	2.481	2.441	2.420	2.420	2.345	2.566	2.624
Aggregate Effective SG (G_{se})	2.709	2.711	2.716	2.771	2.746	2.706	2.729	2.699	2.912	3.006
Aggregate Bulk SG (G_{sb})	2.701	2.686	2.694	2.745	2.738	2.703	2.726	2.674	2.906	2.984
Absorbed Asphalt Content (P_{ba}), %	0.11	0.35	0.31	0.35	0.11	0.04	0.04	0.36	0.07	0.25
Effective Asphalt Content (P_{be}), %	5.77	5.87	5.59	5.56	5.38	5.85	6.48	5.30	6.04	5.29
Gradation / Sieve Size	**% Passing**									
¾ in (19.0 mm)	100.0	100.0	100.0	100.0	99.7	100.0	100.0	100.0	100.0	100.0
½ in (12.5 mm)	98.2	98.9	99.4	93.3	93.7	99.8	99.6	97.6	99.1	100.0
3/8 in (9.5 mm)	87.8	90.0	90.6	84.6	83.2	88.6	88.8	88.2	91.2	92.2
No. 4 (4.75 mm)	60.5	60.8	60.5	55.4	50.8	52.0	53.9	55.9	59.8	61.3
No. 8 (2.36 mm)	42.9	42.3	42.0	38.0	34.7	39.8	38.2	37.3	37.4	43.1
No. 16 (1.18 mm)	32.0	32.8	32.2	27.8	25.5	33.8	30.5	26.1	25.5	30.4
No. 30 (600 µm)	23.9	25.7	25.2	20.4	18.5	26.1	23.2	18.0	18.7	21.4
No. 50 (300 µm)	16.8	18.7	18.4	14.8	12.8	13.7	13.5	12.1	13.9	14.7
No. 100 (150 µm)	10.1	11.5	11.2	10.6	8.5	8.2	8.8	8.0	10.1	9.8
No. 200 (75 µm)	5.9	6.5	6.2	7.3	5.8	6.2	6.5	5.4	6.8	6.4

A and D = mixture designation; BMD = balanced mix design; DG = dense-grade; F = fibers; FA = fines to aggregate; HR = high RAP content mixture; LPG = softer asphalt binder; NMAS = nominal maximum aggregate size; PG = performance grade; RA = recycling agent; RAP = reclaimed asphalt pavement; S = standard traffic; SG = specific gravity; VFA = voids filled with asphalt; VMA = voids in mineral aggregate; VTM = voids in total mixture; WMA = warm mix additive.

Balanced Mix Design Test Results

Durability Assessment—Cantabro Mass Loss

The Cantabro test was conducted on specimens compacted from reheated mixtures collected at the plant. Figure 4 presents the Cantabro test results, with error bars indicating the range of minimum and maximum values. VDOT's maximum ML threshold of 7.5% is denoted by a dashed red line. According to the ML results, BMD mixtures in Groups B, C, D, and F demonstrated better durability on average compared with their respective control mixtures, despite their higher RAP content. Typically, the use of high-RAP contents in asphalt mixtures leads to greater ML values by means of Cantabro test. The reduced ML, indicating increased durability, highlights the effectiveness of using RAs and softer binders, such as PG 58-28 instead of PG 64S-22, to counteract the increased stiffness of mixtures within the BMD framework.

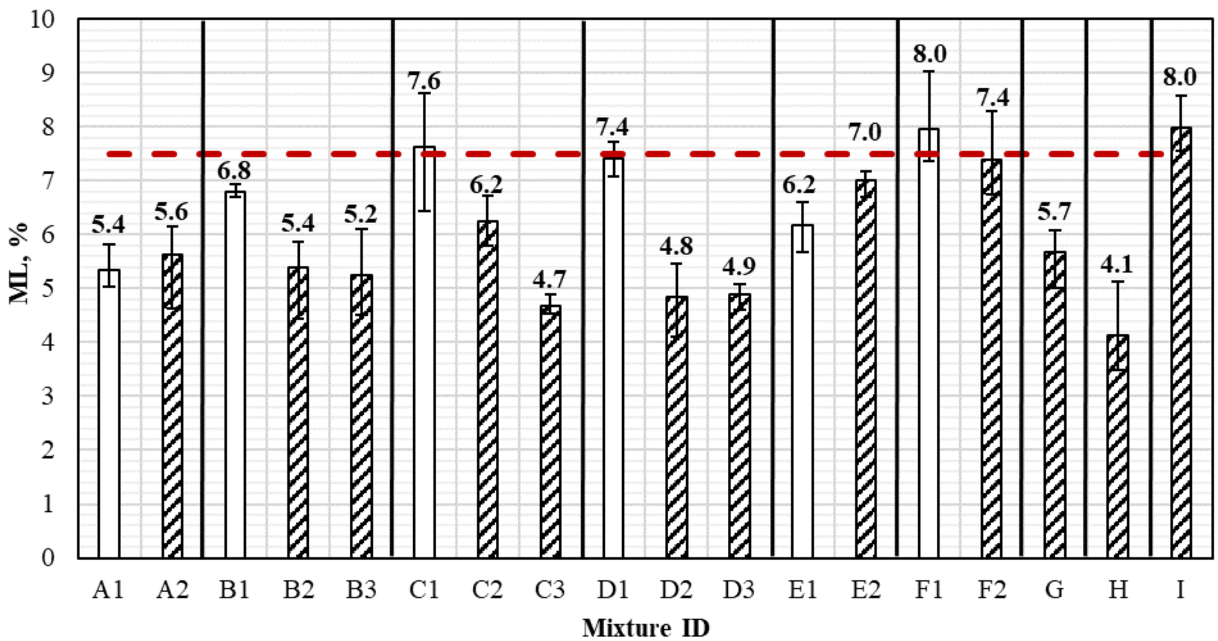


Figure 4. Performance Test Data for Mean Cantabro Mass Loss of Reheat Specimens for All Evaluated Mixtures. Error bars indicate the minimum and maximum value range. Solid white bars designate control mixtures; hashed bars indicate BMD mixtures. Dashed red line indicates VDOT's balanced mix design limit for asphalt surface mixtures with A and D designations. A1 through I = mixture identification as defined in Table 1; ML = mass loss.

In Groups A and E, BMD mixtures exhibited slightly higher ML on average compared with the control mixture, although the difference was not statistically significant. This increase in ML could be attributed to the lower asphalt content determined in these BMD mixtures after production. Although both the BMD and control mixtures in these groups were designed to have nearly the same asphalt content, variations occurred during production. Notably, despite A2 having 0.90% lower asphalt content compared with A1, the difference in ML was not statistically significant, highlighting the benefits of using RA as an additive.

Based on the statistical analysis, all BMD mixtures performed similarly to or better than the control mixtures in terms of durability. Overall, none of the BMD mixtures exceeded VDOT's maximum ML limit of 7.5%, except for mixture I (BMD HR + RA), whereas two of the control mixtures (C1 and F1) failed to meet this criterion.

Cracking Assessment—Cracking Tolerance Index

Figure 5 presents the IDT-CT results in terms of CT index, with error bars indicating the range of minimum and maximum values. A dashed red line denotes VDOT's minimum threshold of 70. The statistical analysis of this test indicates that all BMD mixtures, except for A2, performed similar to or better than their respective control mixtures concerning cracking, even with higher RAP content, highlighting the effectiveness of using RAs and softer binder to alleviate the increased stiffness of mixtures within the BMD framework.

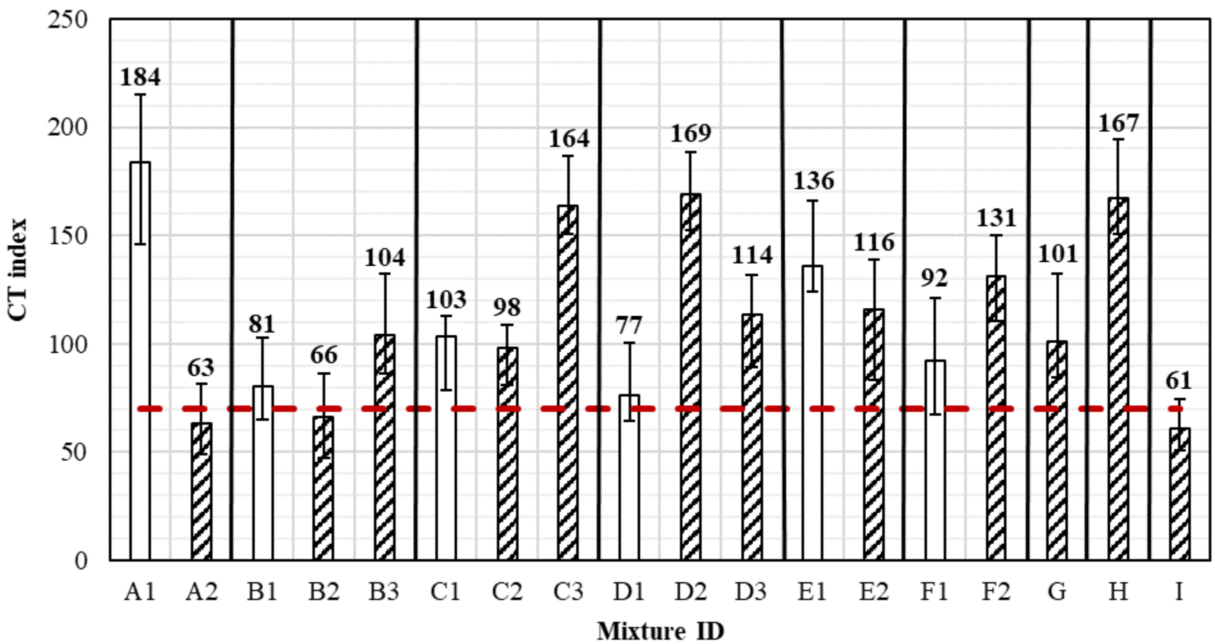


Figure 5. Performance Test Data for Mean CT Index of Reheat Specimens for All Evaluated Mixtures. Error bars indicate the minimum and maximum value range. Solid white bars designate control mixtures; hashed bars indicate BMD mixtures. Dashed red line indicates VDOT's balanced mix design limit for asphalt surface mixtures with A and D designations. A1 through I = mixture identification as defined in Table 1; CT = cracking tolerance.

In Group A, mixture A2 exhibited a statistically significant CT index reduction compared with A1, which had the highest value among all mixtures. Notably, mixtures A2, B2, and I were the only mixtures with a CT index less than the acceptable limit of 70 that VDOT uses for BMD mixtures. These mixtures contained either high-RAP contents or RAs, or both. Therefore, they may require further assessment in terms of resistance to cracking. As previously mentioned, the variation in asphalt content during production could explain the observed results, particularly for Group A, because the mixtures were designed with the same asphalt content but were produced with a 0.90% difference. The BMD mixture had 0.90% less asphalt content than the control mixture. A similar explanation applies to mixture B2, which ended up with 0.17% lower asphalt content than B3 during production.

Rutting Assessment—Asphalt Pavement Analyzer Rut Depth

Figure 6 shows the APA RD results. The error bars represent the range of minimum and maximum values. The dashed red line denotes VDOT's maximum APA RD threshold of 8.0 mm. All mixtures met VDOT's RD limit of 8.0 mm. Surprisingly, the obtained results were better than expected considering the IDT-CT results and revealed that BMD mixtures that performed similarly or better in cracking compared with control mixtures showed no statistically significant difference in rutting with the exception of F2. Mixture F2 ended up with 0.63% more asphalt content compared with F1, despite being designed to have a lower asphalt content than F1. This increase was previously observed in the CT index, for which F2 had a statistically significantly higher value than F1. However, the APA results demonstrate that F2 also performs better in regard to rutting. In addition, the statistical analysis indicates that A2 exhibited improved rutting performance, whereas B3 is the only mixture with significantly worse performance, possibly because of B3's 0.32% higher asphalt content compared with B1.

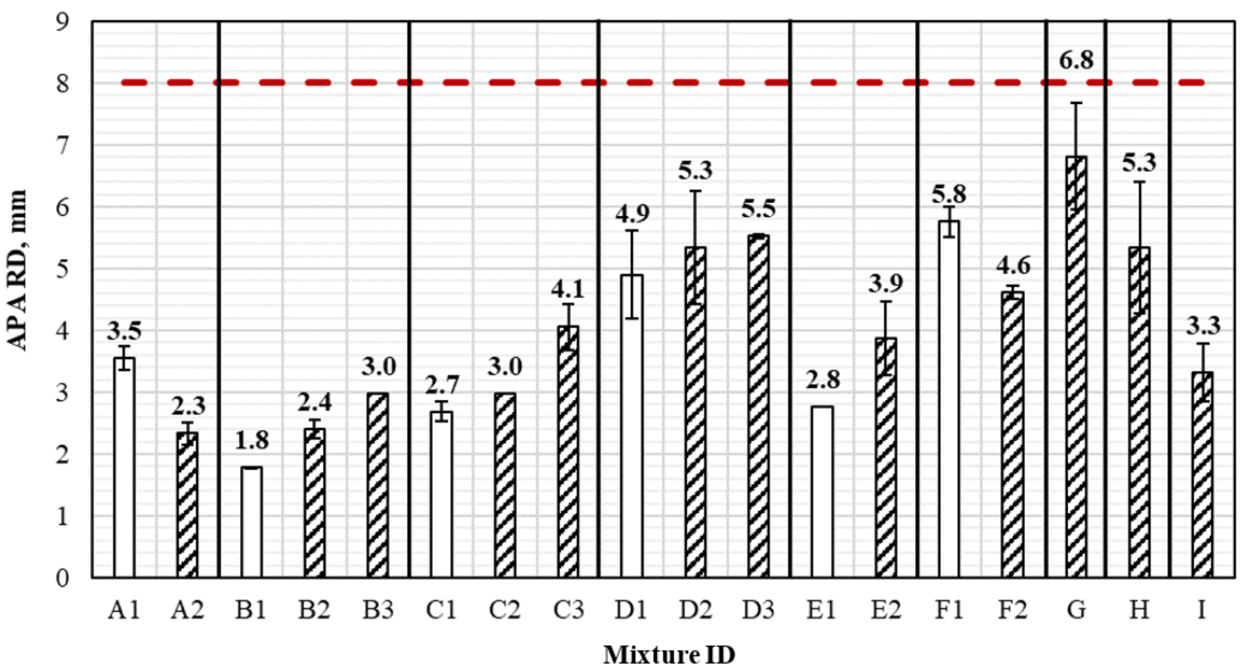


Figure 6. Performance Test Data for Mean APA RD of Reheat Specimens for All Evaluated Mixtures. Error bars indicate the minimum and maximum value range. Solid white bars designate control mixtures; hashed bars indicate BMD mixtures. Dashed red line indicates VDOT's balanced mix design limit for asphalt surface mixtures with A and D designations. A1 through I = mixture identification as defined in Table 1; APA = asphalt pavement analyzer; RD = rut depth.

Advanced Performance Tests

Stiffness Properties—Dynamic Modulus |E|*

Figure 7 shows the DM master curves for all tested mixtures at a reference temperature of 21.1°C. In this multipart figure, a separate plot shows each group.

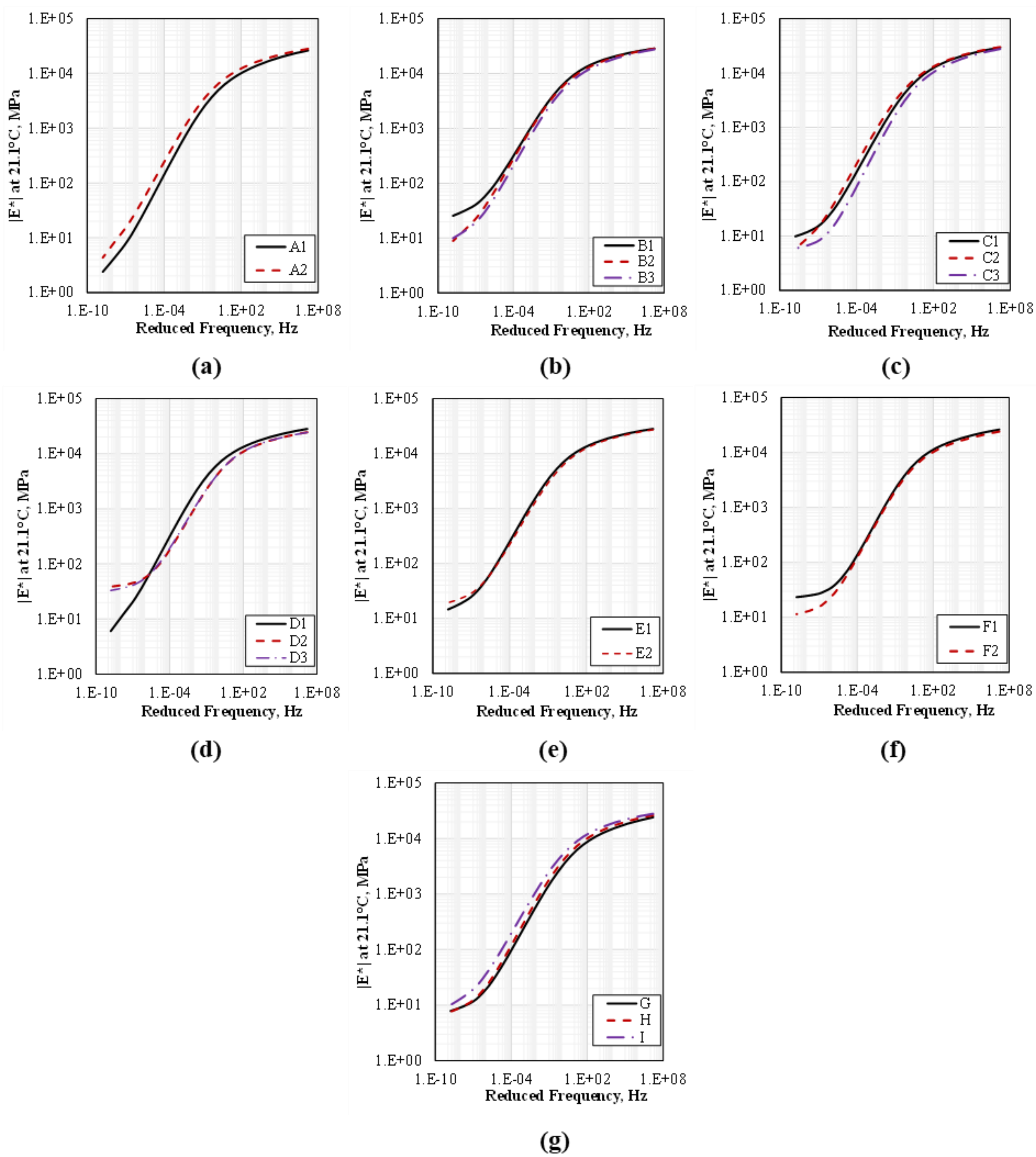


Figure 7. Dynamic Modulus (|E*|) Master Curves at 21.1°C: (a) Group A; (b) Group B; (c) Group C; (d) Group D; (e) Group E; (f) Group F; (g) Groups G, H, and I. A1 through I = mixture identification as defined in Table 1.

Among all 18 mixtures, B1 is the stiffest, and C3 is the softest. Note that A2 is stiffer than A1, even though A2 contains RA. However, as mentioned before, the statistically significant difference in binder content may have contributed to this outcome. On the other hand, B2 and B3, which are high-RAP mixtures, are softer than B1, likely because of the effects of the RA and softer binder. Moving to Group C, mixture C3, which had higher RAP and a lower PG binder, emerged as the softest, whereas C2, with higher RAP and RA, was the stiffest. In Group D, BMD mixtures D2 and D3 exhibited similar results and were both softer than D1. Reduced frequencies below approximately 1 x 10^{-6} Hz are extrapolated and should not be considered in the same way as data at reduced frequencies above 1 x 10^{-6} Hz. Notably, Groups E and F exhibited comparable DM master curves for both control and BMD mixtures. These DM test results are also essential for the subsequent analysis of the CF test, presented in the following section.

Cracking Assessment—Cyclic Fatigue Test Results and Analyses

Figure 8 shows the C versus S curves for each group. The mixture modulus influences the position of the C versus S curve, with higher modulus generally resulting in a higher positioned curve. When mixtures have the same toughness (i.e., same D^R), the mixture with a higher modulus would experience lower strains under the same loading conditions, leading to an extended fatigue life. Similarly, a mixture with higher D^R can resist cracking for a longer time, resulting in better fatigue life compared with another mixture with a similar modulus but lower D^R. However, because the modulus and D^R vary, ranking the cracking resistance of the mixtures becomes more challenging. As mentioned previously, both effects are accounted for in S_{app}, which is why S_{app} was used for ranking purposes.

Figures 9 and 10 show the results of the CF test in terms of D^R and S_{app}. Figure 8 shows that A2 is stiffer than A1, and A2's C versus S curve is positioned higher, which aligns with the general trends seen in other cases. However, A2 has a lower D^R, resulting in overall worse fatigue performance according to S_{app}, matching the IDT-CT results. The C versus S curve of B1 is higher than B2 and B3. However, the BMD mixtures have higher D^R, leading to higher S_{app} values. Although IDT-CT results showed that mixture B2 failed to meet the VDOT minimum CT index requirement of 70, the CF results indicated that B2 had one of the higher S_{app} values. In Group C, the C versus S curve was the highest for C2, followed closely by C1, and C3 had the lowest. For D^R, C1 had the lowest value, and C3 had the highest. The resultant S_{app} values indicate that C1 has the worst fatigue performance, and C2 and C3 have similar performance. Notably, from Groups B and C, using RA leads to a higher C versus S curve with lower D^R compared with using a softer asphalt binder. The combination of a higher curve with lower D^R and higher D^R with a lower curve resulted in equivalent S_{app} values and, consequently, equivalent and improved fatigue performance. In Group D, D1 had the highest C versus S curve but the lowest D^R, and D2 had the lowest curve but with a higher D^R, leading to equivalent S_{app} values. D3 had a slightly higher curve than D2, but D3 had higher D^R, resulting in a higher S_{app} value. The improved fatigue performance observed in mixture D3 may be attributed to the use of a softener RA, fibers, and other influencing variables during production. Mixtures in Groups E and F exhibited similar C versus S curves and D^R, resulting in similar S_{app} values. Overall, concerning S_{app}, mixtures B3, C2, C3, and D3 exhibited statistically better fatigue performance compared with their control mixtures, whereas all remaining BMD mixtures performed statistically the same. In addition, considering that the tested mixtures are all SMs with A and D designations used for 0–3 and 3–10 MESALs, respectively, the S_{app} values support the current nationally calibrated guidance of using S_{app} values greater than 8 for traffic levels less than 10 MESALs (Wang et al., 2020).

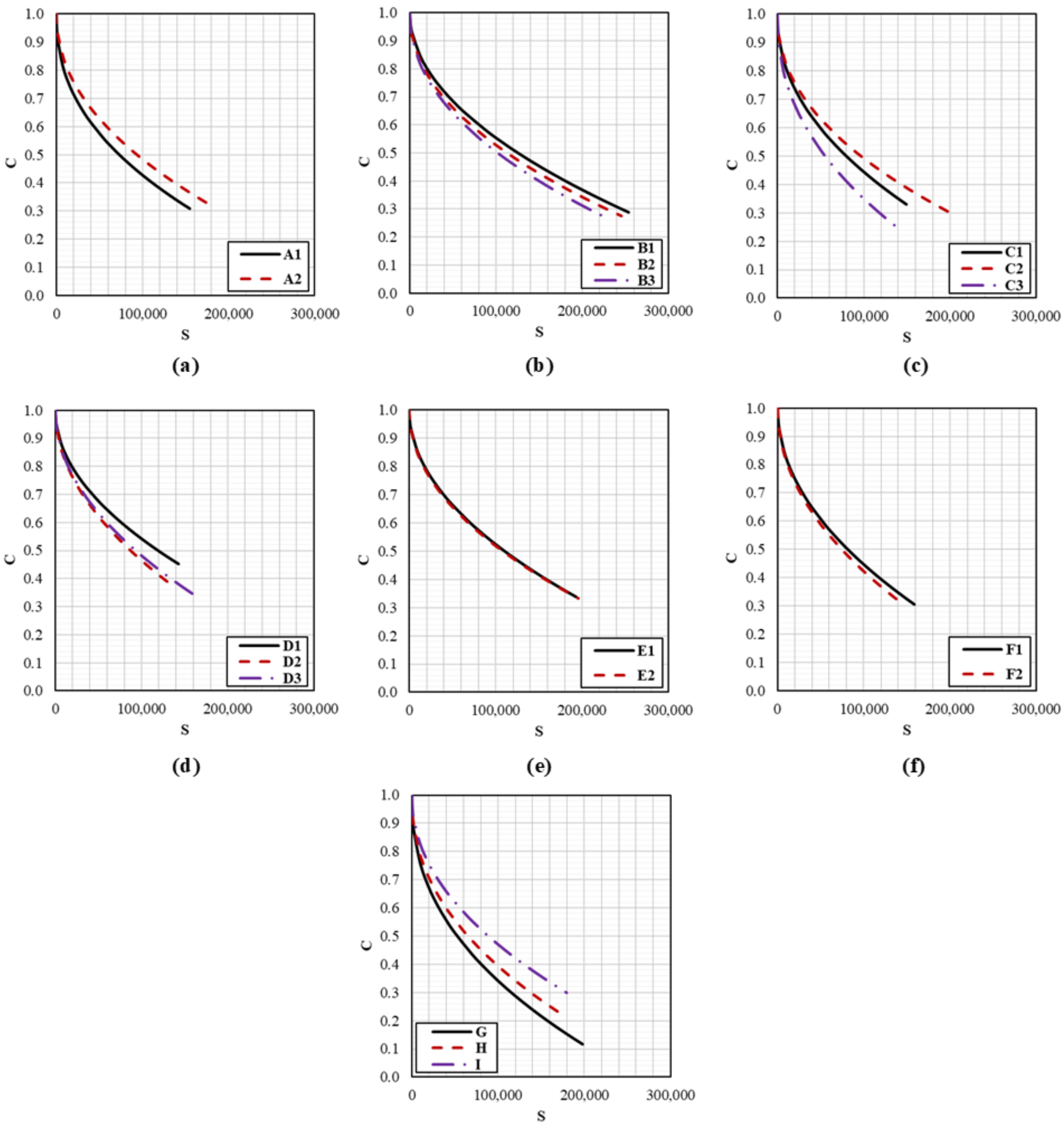


Figure 8. Material Integrity (C) versus Damage (S) Curves: (a) Group A; (b) Group B; (c) Group C; (d) Group D; (e) Group E; (f) Group F; (g) Groups G, H, and I. A1 through I = mixture identification as defined in Table 1.

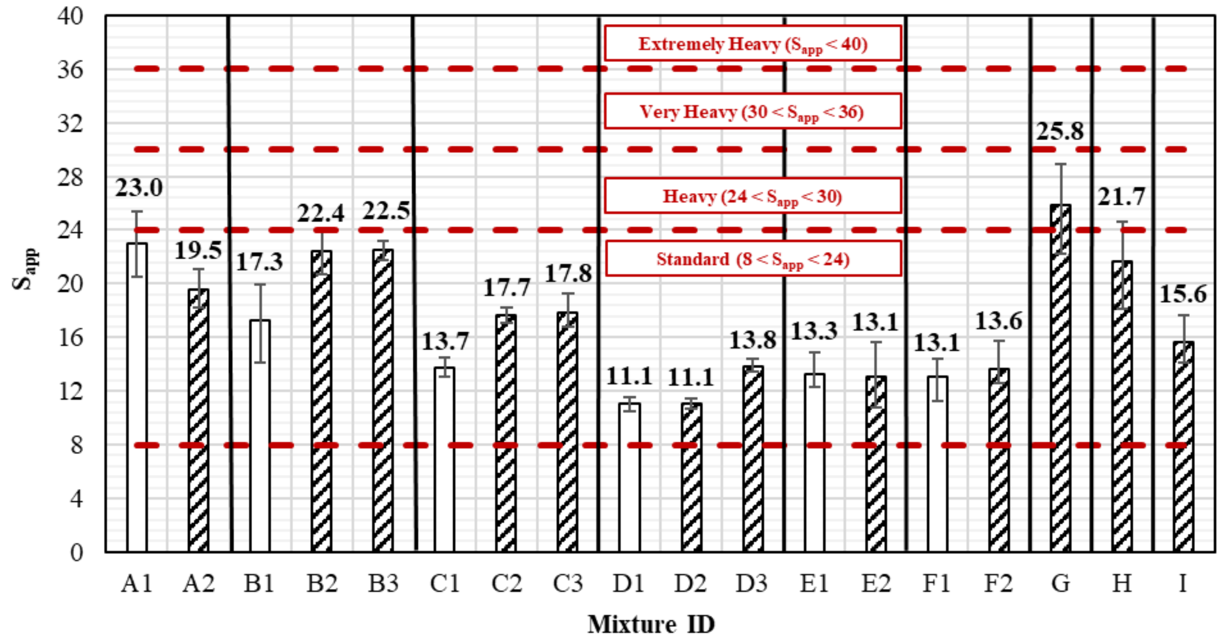


Figure 9. Cyclic Fatigue Performance Test Data for Mean S_{app} of Reheat Specimens for All Evaluated Mixtures. Error bars indicate the minimum and maximum value range. Solid white bars designate control mixtures; hashed bars indicate balanced mix design mixtures. Dashed red line indicates recommended threshold values for S_{app} index parameter as a function of traffic tier. A1 through I = mixture identification as defined in Table 1; S_{app} = cyclic fatigue index parameter.

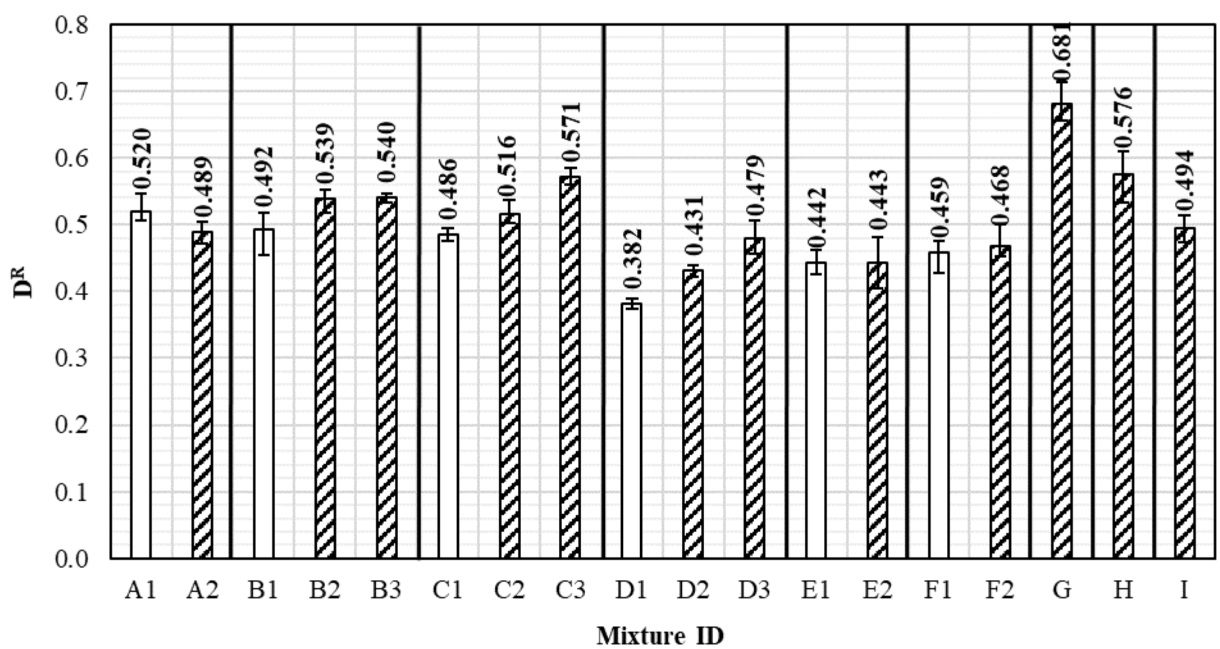


Figure 10. Cyclic Fatigue Performance Test Data for Mean D^R of Reheat Specimens for All Evaluated Mixtures. Error bars indicate the minimum and maximum value range. Solid white bars designate control mixtures; hashed bars indicate balanced mix design mixtures. A1 through I = mixture identification as defined in Table 1; D^R = pseudo-energy-based failure criterion.

Rutting Assessment—Rutting Strain Index

Figure 11 shows the RSI values for all mixtures. In Group A, A1 outperformed A2 in both the IDT-CT and CF test because of the higher asphalt content during production. However, this increase in asphalt content led to worse rutting performance, as indicated by the higher RSI value. In Groups B and C, the BMD mixtures with a softer binder, which were found to be softer in the DM and CF results, exhibited a more noticeable reduction in rutting performance compared with the stiffer BMD mixtures. In Group D, D2 and D3 have significantly higher RSI values compared with D1. D2 and D3 contain 35% RAP compared with 40% RAP, but both include RA and softer binder. Therefore, in these mixtures, the RAP mitigation strategies appear to have caused the rutting potential to increase. In Group E, the addition of 10% RAP and use of a lower PG in E2 led to improved rutting performance while maintaining similar cracking, as observed in the CF test results previously. Group F shows that F1 and F2 perform similarly with regard to rutting, although F2 did not have a higher RAP content and used the same PG without including RA. The only difference was a higher production asphalt content, and otherwise F2 performed similarly according to all advanced performance tests. Overall, all mixtures had RSI values less than 12, which is the recommended value for standard traffic of less than 10 MESALs (Ghanbari et al., 2020).

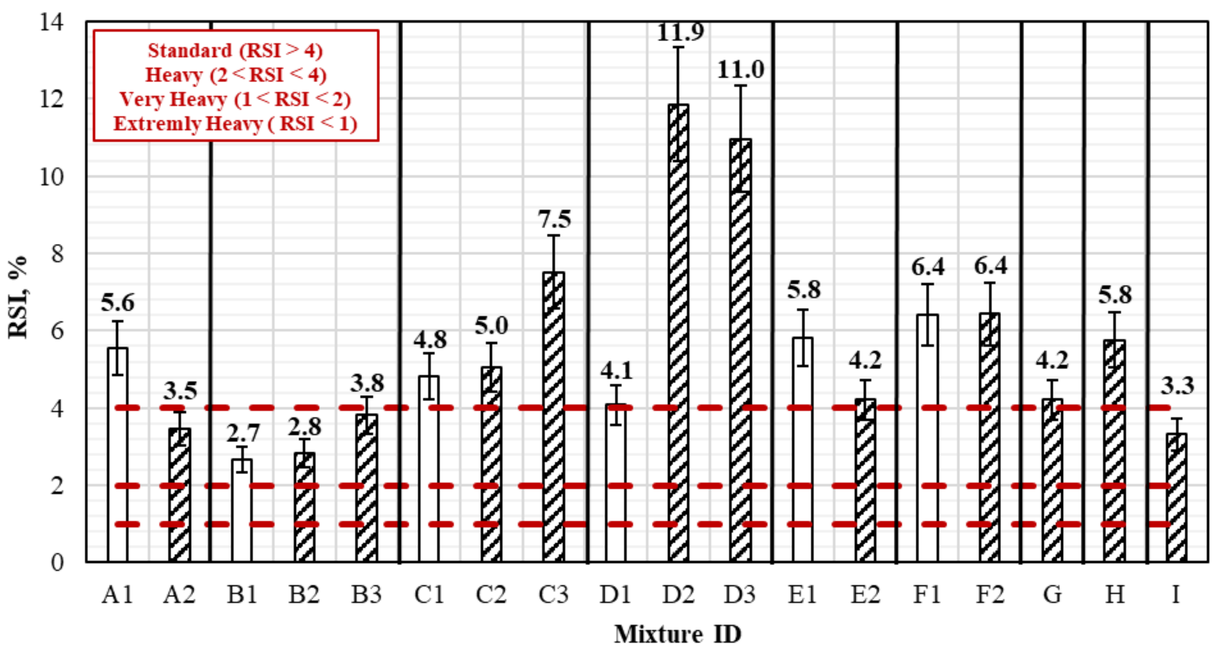


Figure 11. Stress Sweep Rutting Performance Test Data for Mean RSI of Reheat Specimens for All Evaluated Mixtures. Error bars indicate the minimum and maximum value range. Solid white bars designate control mixtures; hashed bars indicate balanced mix design mixtures. A1 through I = mixture identification as defined in Table 1; RSI = rutting strain index.

Rutting Assessment—Confined Flow Number Test Results

Figures 12a and 12b show the rutting relationship at 50°C for all 9.5- and 12.5-mm selected asphalt mixtures, respectively. A lower rutting characteristic indicates a lower accumulated permanent strain with loading, thus indicating a better resistance to rutting.

Furthermore, a flatter curve indicates a lower susceptibility of the asphalt mixtures to rutting by repeated loading. Overall, all evaluated mixtures exhibited a lower rutting relationship in terms of intercept with the y-axis and were flatter compared with the mixture using locally calibrated factors, referred to as VDOT_Default. Appendix B provides the regression coefficients of the MEPDG rutting model. Note that these coefficients were developed using data generated from confined RLPD testing. Meanwhile, the VDOT_Default values were identical to those of the MEPDG Manual of Practice that were determined using unconfined RLPD testing (AASHTO, 2015).

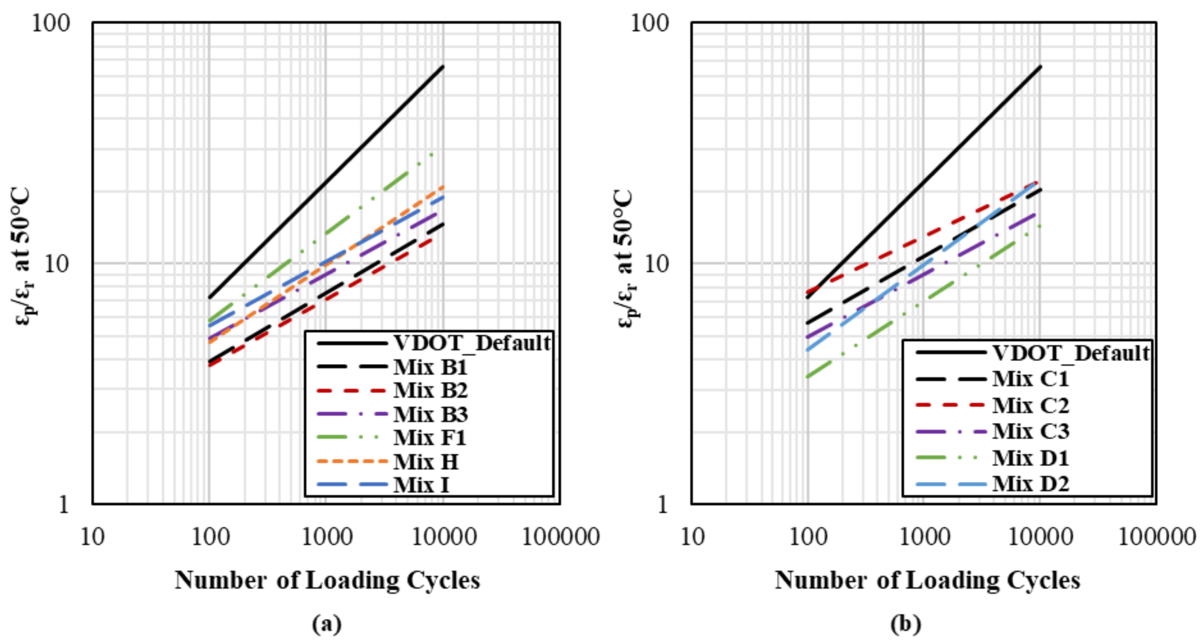


Figure 12. Flow Number Performance Test Data in Terms of Rutting Characteristics at 50°C for All Evaluated Mixtures: (a) 9.5 mm Nominal Maximum Aggregate Size; (b) 12.5 mm Nominal Maximum Aggregate Size. B1 through I = mixture identification as defined in Table 1.

Overview of Mixture Performance Test Results—Discrimination Potential

After confirming that the datasets were normally distributed and exhibited equal variances across all tests, both assessed at a 95% confidence level, the analysis of variance results showed statistically significant differences for the test results among the mixtures. Subsequently, the Tukey test was performed to identify the differences or similarities between specific pairs of the mixture groups for a given test. The results delineated that ML produced six distinct groupings, APA RD also resulted in six groupings, CT index formed eight groupings, S_{app} parameter created seven groupings, and the D^R index led to nine groupings.

The outcome of the analyses highlighted the following two important aspects:

- The mixtures selected for this study do not conform to a singular performance level category. Rather, the experimental design encompassed a relatively broad spectrum of mixtures showcasing a diverse range of performance potentials.
- Although the statistical categorization of mixtures was similar for indices indicating cracking performance in general, the performance rankings were not always identical. In addition, some mixtures were grouped differently across the cracking-related tests.

Testing and Characterization of Extracted and Recovered Asphalt Binders

Performance Grading

Table 4 presents PG of extracted and recovered asphalt binders for all evaluated mixtures. Mixtures within the same group exhibited similar high, intermediate, and low PG temperatures, except for Groups C (difference in high PG temperature) and D (difference in high and low PG temperature). Mixtures in Groups A, B, E, and F exhibited PG grades of 70-22, 76-22, 70-22, and 70-22, respectively. In Group C, mixture C2 (PG 76-22) exhibited a higher high PG temperature compared with mixture C1, whereas mixture C3 (PG 64-22) showed a lower high PG temperature compared with C1 (PG 70-22). This variance could be attributed to changes in the properties of the RAP stockpile used in the mixture, in addition to the use of RA and softer binders. Similar observations can be made for Group D mixtures. Mixture D1 (PG 76-22) exhibited a notably higher high PG temperature compared with mixtures D2 (PG 70-22) and D3 (PG 70-28). However, D2 and D3 demonstrated substantial improvements in terms of low PG temperature compared with the control mixture D1.

Table 4. Properties of Extracted and Recovered Asphalt Binders for All Evaluated Mixtures

Mix ID	Continuous Grade			MSCR Test Results		Other Parameters	
	High[a], °C	Int[b], °C	Low, °C	$J_{nr\,3.2}$, kPa^{-1}	%R, %	ΔT_c, °C	G-R, kPa
A1	73.4	23.3	– 23.8	1.1314	5.2499	– 3.3	434.36
A2	73.3	24.7	– 23.4	1.2039	3.9640	– 2.1	508.08
B1	80.3	25.1	– 24.2	0.3937	27.8624	– 1.1	378.31
B2	78.4	24.1	– 24.2	0.5010	14.2146	– 3.0	346.00
B3	76.8	23.6	– 24.1	0.6642	9.8403	– 2.5	412.88
C1	73.1	23.3	– 24.8	1.2569	3.6560	– 1.5	183.67
C2	77.0	25.5	– 23.1	0.6593	9.1005	– 2.5	130.99
C3	69.1	20.9	– 26.4	2.2392	1.3832	– 1.7	308.96
D1	76.9	27.2	– 22.5	0.7777	7.0417	– 0.8	452.81
D2	71.1	21.2	– 27.8	1.7465	2.3246	+ 1.2	88.52
D3	71.0	21.1	– 28.6	1.5993	3.4225	+ 0.4	133.79
E1	70.9	24.0	– 24.2	1.7846	2.6442	– 2.2	250.78
E2	74.6	23.6	– 26.2	1.0565	5.5756	+ 0.4	178.17
F1	74.6	25.5	– 23.1	0.9415	5.0625	– 1.1	228.92
F2	75.5	26.3	– 22.7	0.9141	5.3564	– 1.7	285.69
G	73.2	20.6	– 25.3	1.1881	5.9624	– 4.5	221.95
H	75.7	24.6	– 22.6	0.7525	9.3275	– 4.3	405.49
I	70.8	24.1	– 20.7	1.7150	3.6820	– 9.5	370.33

ΔT_c = difference in critical low temperature performance grade limiting temperatures; %R = percent recovery; A1 through I = mixture identification as defined in Table 1; G-R = Glover-Rowe parameter; Int = intermediate; $J_{nr\,3.2}$ = nonrecoverable creep compliance; MSCR = Multiple Stress Creep Recovery. [a] Temperature corresponds to $G^*/\sin\delta$ = 2.2 kPa. [b] Temperature corresponds to $G^*\sin\delta$ = 5,000 kPa.

Multiple Stress Creep Recovery Test Results

Binder grading was also performed in accordance with AASHTO M 322 (AASHTO, 2022d), which incorporates the nonrecoverable creep compliance at 3.2 kPa ($J_{nr,\,3.2}$ kPa) and percent recovery at 3.2 kPa (%R) determined using the Multiple Stress Creep Recovery (MSCR) test. The MSCR test was conducted at 64°C, the average 7-day maximum pavement design temperature for Virginia. AASHTO M 332 (AASHTO, 2019f) specifies a maximum $J_{nr,\,3.2}$ kPa requirement for standard (S), heavy (H), very heavy (V), and extremely heavy (E) traffic of 4.5 kPa^{-1}, 2.0 kPa^{-1}, 1.0 kPa^{-1}, and 0.5 kPa^{-1}, respectively. Table 4 shows the MSCR testing data for all evaluated extracted and recovered binders. VDOT specifications call for a minimum of PG 64S-16 and PG 64H-16 virgin asphalt binders for SMs with A and D designations, respectively (VDOT, 2020). The data in Table 4 indicate that all extracted and recovered binders from the mixtures evaluated in this study met or exceeded the VDOT specification criteria in terms of asphalt binder properties. For example, binders A1 and A2 were in category H. Binder B1 was in category E, and binders B2 and B3 were in category V. Although binder B1 met the E traffic requirement, this result could be attributed to some elastomeric polymers present in the corresponding RAP stockpile, given that %R for binder B1 was low compared with typical polymer-modified asphalt binders. Binders C1, D2, D3, E1, E2, G, and I were in category H. Binders C2, D1, F1, F2, and H were in category V. Binder C3 was the only binder in category S.

Difference in Critical Low Temperature Performance Grade—ΔT_c Results

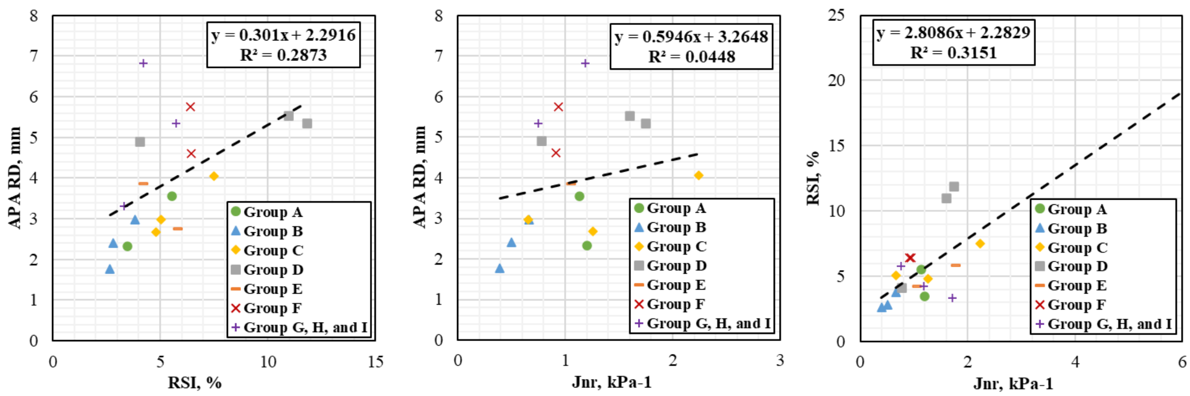
Table 4 presents the ΔT_c values for all evaluated extracted and recovered binders. All binders had ΔT_c values ranging from -4.5 to 1.2°C, with none exceeding the traditional cracking zone of -5.0°C except for binder I, which had a ΔT_c value of -9.5°C.

Cracking Assessment by Means of the Glover-Rowe Parameter

Table 4 presents the G-R parameter values for all evaluated binders after 20-hour Pressure Aging Vessel aging. A G-R value between 180 (onset cracking) and 600 kPa (significant cracking) defines the damage zone and range where cracking is likely to begin because of brittle rheological behavior. Three binders (D2, D3, and E2) had G-R values less than the 180 kPa onset cracking value, with the G-R of binder E2 being very close to 180, which indicates a potential high resistance of these binders to nonload cracking. The remaining binders had G-R values spreading between the 180 onset of cracking value and 600 kPa significant cracking value. None of the evaluated binders had a G-R value greater than 600 kPa.

Correlation Analysis Among Mixture and Binder Performance Tests Parameters

Figures 13 and 14 present scatter plots illustrating the correlations between the BMD tests and fundamental tests, as well as binder tests and mixture tests. The correlations between the BMD tests and fundamental tests ranged from poor to weak. Particularly, the correlations involving APA RD and RSI, S_{app} and ML, and D^R and ML were relatively stronger compared with those correlations between the S_{app} and CT index and the D^R and CT index. The observed lack of strong correlation among the tests can be attributed to potential influencing factors such as differences in test conditions—such as temperature, air voids content, loading configuration and rate, and failure mechanisms—and the sensitivity of test methods to materials properties. Despite the lack of strong correlations between the tests, the overall correlation trends among the tests pointed in the right direction, with tests of interest either indicating performance improvement or deterioration.



(a) (b) (c)

Figure 13. Correlations among Mixture and Binder Rutting Performance Indices and Paramaters: (a) APA RD versus RSI; (b) APA versus J_{nr}, (c) RSI versus J_{nr}. APA = asphalt pavement analyzer, J_{nr} = noncreep compliance; RD = rut depth; RSI = rutting strain index.

Similar observations were also made for the correlations between the binder tests and mixture tests, which ranged from poor to weak correlations. Although overall trends were in the expected directions, the following indices did not demonstrate the anticipated trend with respect to G-R parameter: ML, S_{app}, and D^R. Confounding factors such as the quantity of component materials and volumetric properties could have influenced these trends. For instance, differences in asphalt content of the same binder grade can result in varied mixture performance while the binder property remains constant, thereby hindering correlation.

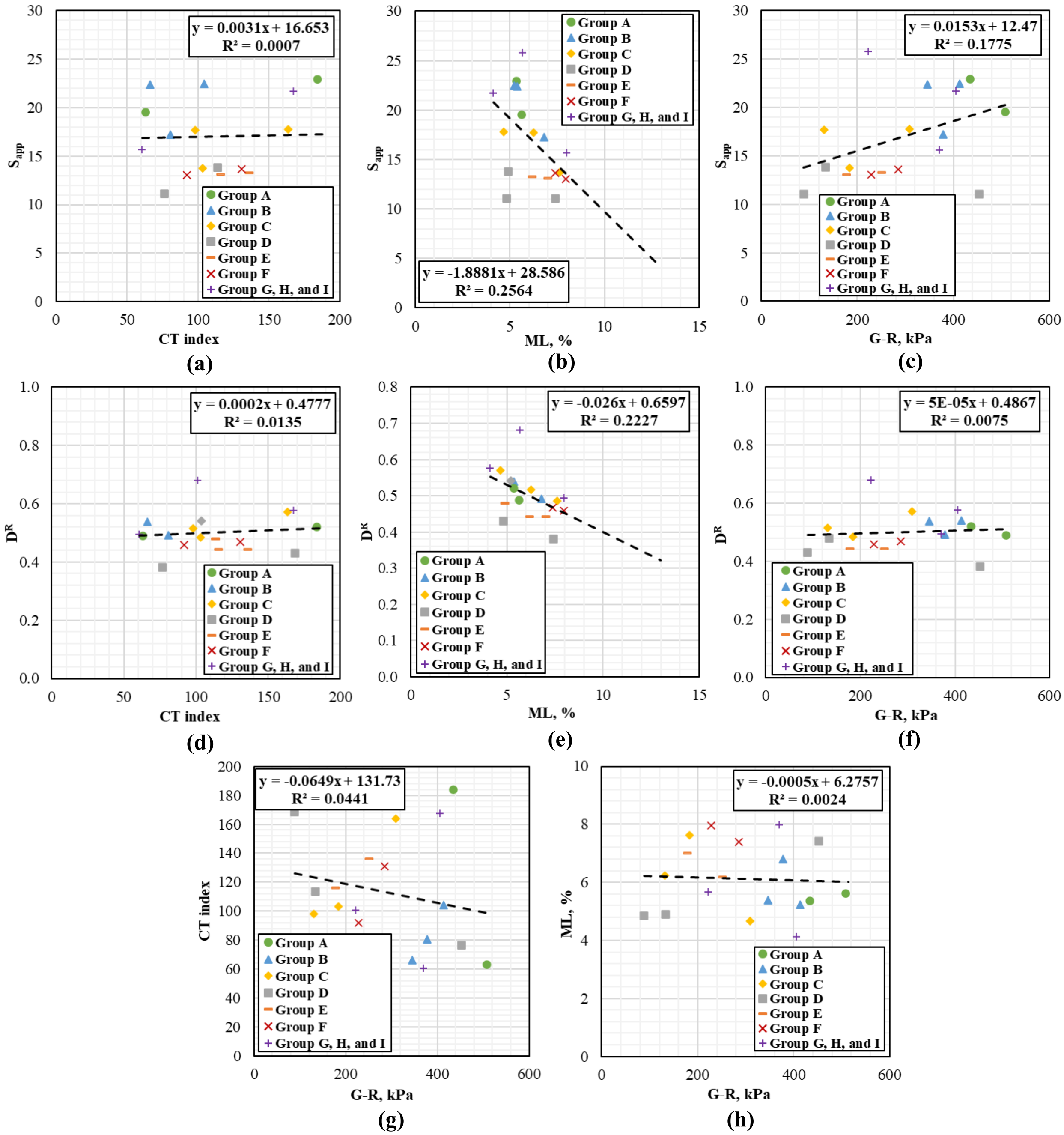


Figure 14. Correlations among Mixture and Binder Durability and Cracking Performance Indices and Paramaters: (a) S_{app} versus CT Index; (b) S_{app} versus ML; (c) S_{app} versus G-R; (d) D^R versus CT Index; (e) D^R versus ML; (f) D^R versus G-R; (g) CT Index versus G-R; (h) ML versus G-R. CT = cracking tolerance; D^R = pseudo-energy-based failure criterion; G-R = Glover-Rowe parameter; ML = mass loss; S_{app} = cyclic fatigue index parameter.

Mechanistic-Empirical Simulations and Analyses

AASHTOWare Pavement ME Design Simulations

Overview

AASHTOWare Pavement ME Design simulations were conducted for the selected SMs B1, B2, B3, C1, C2, C3, D1, D2, F1, H, and I (AASHTO, 2022a). Each mixture was assessed for both pavement structures—A and D—encompassing 3 and 10 MESALs of traffic, respectively. A brief analysis was performed to evaluate the effect of climate station selection on the AASHTOWare Pavement ME Design outcomes for mixture C2. Simulations using data from all climatic stations in Virginia were executed. Overall, no significant effect was observed on the outcome of AASHTOWare Pavement ME Design in terms of total and asphalt concrete rutting at 15 years and BUFC at 30 years. Consequently, simulations for this study were conducted using stations closest to where each mixture was placed in the field.

Initially, three speeds (25, 35, and 45 mph) were considered for Structure A, and three speeds (35, 45, and 55 mph) were considered for Structure D. A preliminary analysis considering the three speeds for both structures was conducted for mixtures C2, D1, and D2. Again, no notable effect of speed was observed on the outcome of AASHTOWare Pavement ME Design in terms of total and asphalt concrete rutting at 15 years and BUFC at 30 years. Therefore, simulations for this study were conducted using a speed of 35 mph for Structure A and 45 mph for Structure D.

Two densities (93 and 95%) were considered for all SM layers in all simulations. The 11 mixtures were evaluated using VDOT defaults and material-specific rutting and fatigue calibrated models. Note that in both cases, material-specific DM values were used for the simulations.

Results when Using Default Parameters

The collected data were categorized into eight segments reflecting the structure (A or D), rutting and fatigue model (VDOT default versus calibrated), and density (93 versus 95%). Figure 15a illustrates APA RD of the evaluated mixtures as a function of AASHTOWare Pavement ME Design total RD at 15 years for segment Structure-D-Default-45mph-93%. A moderate correlation was observed. VDOT's AASHTOWare Pavement ME Design limit of 0.26 inches after 15 years corresponds to approximately 4.0-mm APA RD for all mixtures with 3.7-mm APA RD for 9.5-mm NMAS and 4.4 mm for 12.5-mm NMAS. Although the APA RD values are considerably lower compared with the VDOT BMD criteria of 8 mm, this discrepancy could be attributed to the use of default material properties and the fundamental difference between the AASHTOWare Pavement ME Design-predicted rutting and laboratory APA RD. Furthermore, no rutting issues have been reported in Virginia since transitioning to Superpave. Figures 15b and 15c show the CT index and ML of the evaluated mixtures as a function of AASHTOWare Pavement ME Design BUFC at 30 years, respectively, for segment Structure-D-Default-45mph-93%. Overall, no correlations exist. VDOT's AASHTOWare Pavement ME Design limit of 6% after 30 years corresponds to approximately 122 CT index and 6.3% ML for all mixtures.

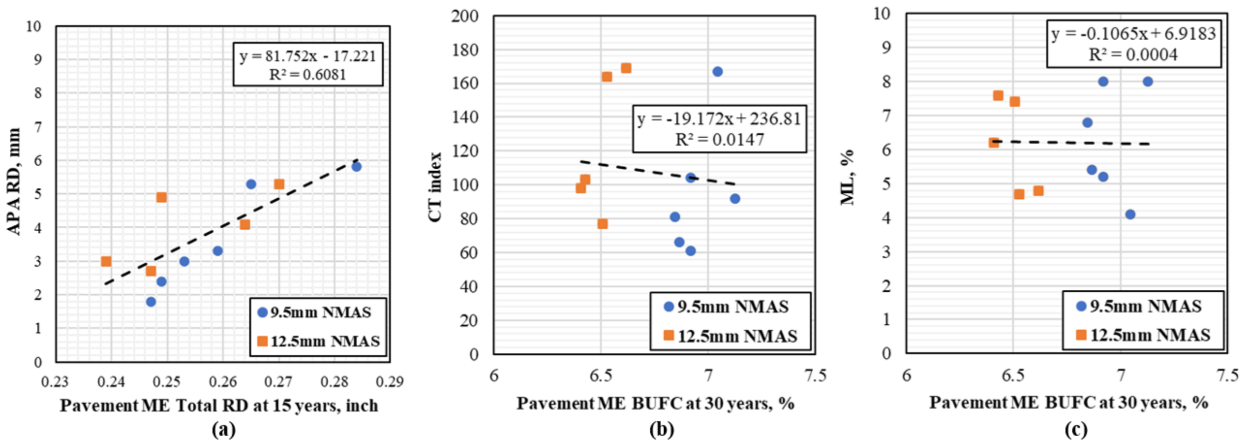
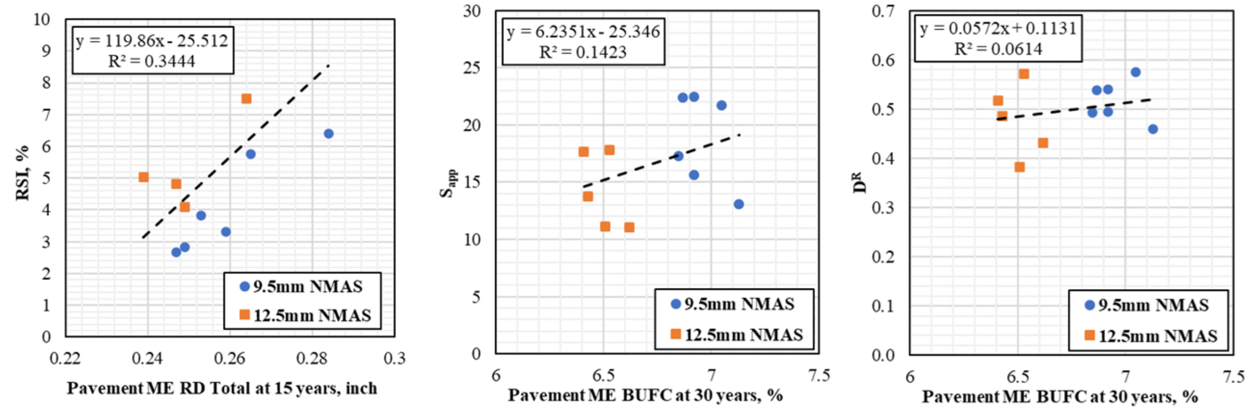


Figure 15. Correlations between Balanced Mix Design Indices and Pavement ME Outputs for Segment Structure-D-Default-45mph-93%: (a) APA RD versus Total RD at 15 Years; (b) CT Index versus BUFC at 30 Years; (c) ML versus BUFC at 30 Years. APA = asphalt pavement analyzer; BUFC = bottom-up fatigue cracking; CT = cracking tolerance; ME = mechanistic-empirical; ML = mass loss; NMAS = nominal maximum aggregate size; RD = rut depth.

Figure 16a shows RSI of the evaluated mixtures as a function of AASHTOWare Pavement ME Design total RD at 15 years for segment Structure-D-Default-45mph-93% (AASHTO, 2022a). A poor correlation was observed with the combined 9.5- and 12.5-mm NMAS mixtures. VDOT's AASHTOWare Pavement ME Design limit of 0.26 inches after 15 years corresponds to an approximate RSI of 5.6%, corresponding to the standard traffic category. Figures 16b and 16c show the S_{app} and D^R of the evaluated mixtures as a function of AASHTOWare Pavement ME Design BUFC at 30 years, respectively, for segment Structure-D-Default-45mph-93%. The observed correlations did not align as expected. In other words, higher S_{app} and D^R values were expected to result in lower BUFC, which was not the case here. Better correlations were observed with rutting because of the software's capability to decouple the rutting models for each layer independently. However, a single model is considered for BUFC, mainly affected by the base layer properties.



(a) (b) (c)

Figure 16. Correlations between Parameters from Advanced Testing and Pavement ME Outputs for Segment Structure-D-Default-45mph-93%: (a) RSI versus Total RD at 15 Years; (b) S_{app} versus BUFC at 30 Years; (c) D^R versus BUFC at 30 Years. BUFC = bottom-up fatigue cracking; D^R = pseudo-energy-based failure criterion; ME = mechanistic-empirical; NMAS = nominal maximum aggregate size; RD = rut depth; RSI = rutting strain index; S_{app} = cyclic fatigue index parameter.

Results when Using Materials-Specific Parameters

Figure 17a illustrates the APA RD of the evaluated mixtures as a function of AASHTOWare Pavement ME Design total RD at 15 years for the segment Structure-D-Calibrated-45mph-93% (AASHTO, 2022a). In contrast to VDOT_Default calibration models, no correlation exists. VDOT's AASHTOWare Pavement ME Design limit of 0.26 inches after 15 years corresponds to an extrapolated value of 4.6-mm APA RD for all mixtures, with 11.8-mm APA RD for 9.5-mm NMAS and 17.1 mm for 12.5-mm NMAS. Figures 17b and 17c show the CT index and ML of the evaluated mixtures as a function of AASHTOWare Pavement ME Design BUFC at 30 years, respectively, for segment Structure-D-Calibrated-45mph-93%. Currently, no correlation exists for the CT index. However, a moderate correlation exists between ML and BUFC. VDOT's AASHTOWare Pavement ME Design limit of 6% after 30 years corresponds to approximately 107 CT index and 6.3% ML for all mixtures.

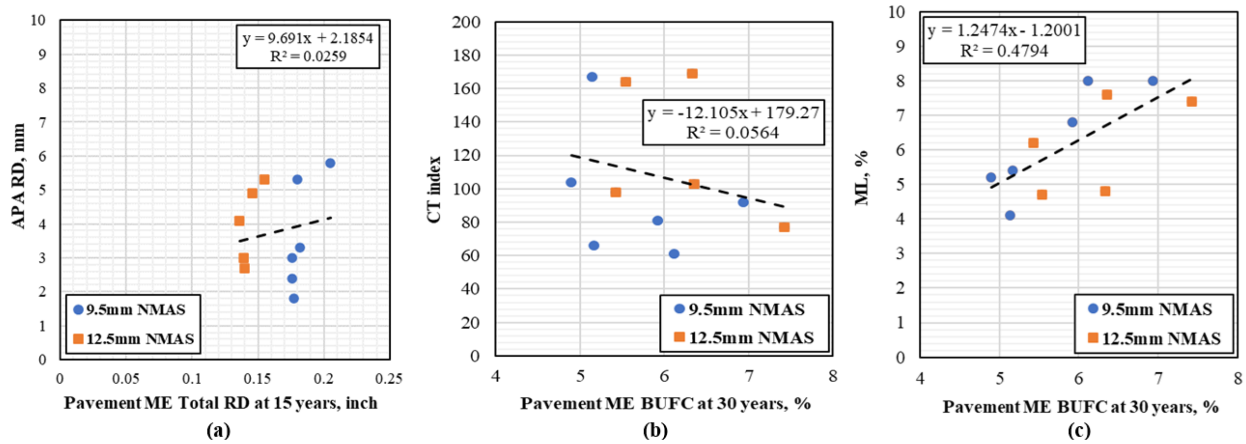


Figure 17. Correlations Between Balanced Mix Design Indices and Pavement ME Outputs for Segment Structure-D-Calibrated-45mph-93%: (a) APA RD versus Total RD at 15 Years; (b) CT Index versus BUFC at 30 Years; (c) Mass Loss versus BUFC at 30 Years. APA = asphalt pavement analyzer; BUFC = bottom-up fatigue cracking; CT = cracking tolerance; ME = mechanistic-empirical; ML = mass loss; NMAS = nominal maximum aggregate size; RD = rut depth.

Figure 18a presents RSI of the evaluated mixtures as a function of AASHTOWare Pavement ME Design total RD at 15 years for segment Structure-D-Calibrated-45mph-93% (AASHTO, 2022a). A poor correlation was observed. VDOT's AASHTOWare Pavement ME Design limit of 0.26 inches after 15 years corresponds to an approximate RSI of 2.6%, categorizing it under heavy traffic. Figures 18b and 18c display the S_{app} and D^R of the evaluated mixtures as a function of AASHTOWare Pavement ME Design BUFC at 30 years, respectively, for segment Structure-D-Calibrated-45mph-93%. Very good correlations were observed. VDOT's AASHTOWare Pavement ME Design limit of 6% after 30 years corresponds to a S_{app} value of 16.4, aligning with the standard traffic category.

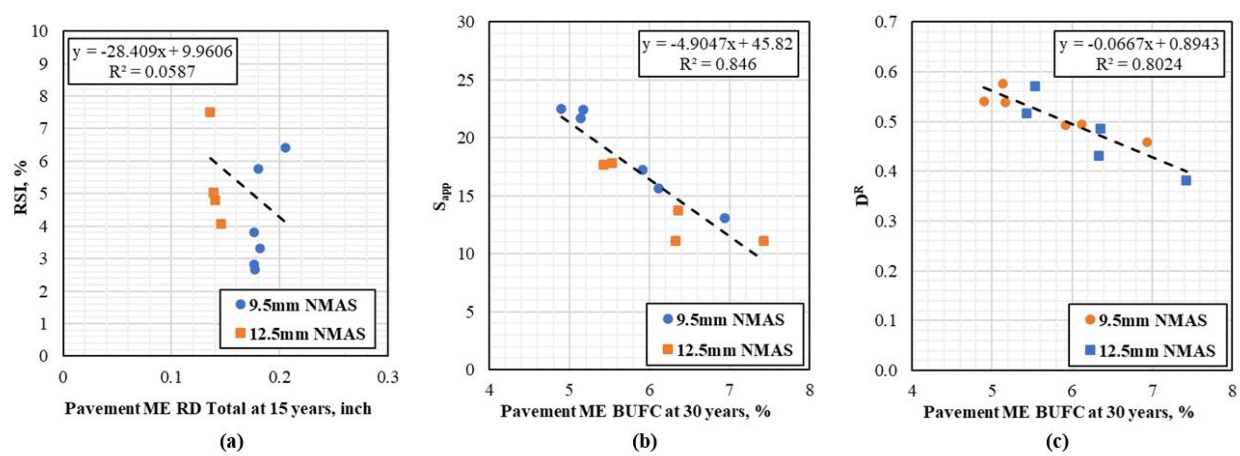


Figure 18. Correlations between Parameters from Advanced Testing and Pavement ME Outputs for Segment Structure-D-Calibrated-45mph-93%: (a) RSI versus Total RD at 15 Years; (b) S_{app} versus BUFC at 30 Years; (c) D^R versus BUFC at 30 Years. BUFC = bottom-up fatigue cracking; D^R = pseudo-energy-based failure criterion; ME = mechanistic-empirical; NMAS = nominal maximum aggregate size; RSI = rutting strain index; S_{app} = cyclic fatigue index parameter.

Table 5 summarizes APA RD corresponding to AASHTOWare Pavement ME Design Total RD of 0.26 inches, and the CT index and ML corresponding to AASHTOWare Pavement

ME Design BUFC of 6% for the eight segments (AASHTO, 2022a). For Structure A, a significant effect on total rutting appears to occur when using material-specific calibration models rather than the VDOT default models. However, no significant effect was observed for the CT index and ML. The effect of density was minor for APA RD, CT index, and ML.

Table 5. Summary of Balanced Mix Design Index Parameters Determined Based on Pavement Mechanistic-Empirical Simulations by Extrapolation or Interpolation

Segment Identification	Balanced Mix Design Index		
	APA RD (mm)	CT Index	Mass Loss (%)
Structure-A-Default-35mph-93%	14.8	110	6.8
Structure-A-Calibrated-35mph-93%	5.8	109	6.0
Structure-A-Default-35mph-95%	13.4	117	6.5
Structure-A-Calibrated-35mph-95%	5.5	108	6.2
Structure-D-Default-45mph-93%	4.0	122	6.3
Structure-D-Calibrated-45mph-93%	4.7	107	6.3
Structure-D-Default-45mph-95%	4.6	119	6.3
Structure-D-Calibrated-45mph-95%	4.5	106	6.4

APA = asphalt pavement analyzer;; CT = cracking tolerance; RD = rut depth.

FlexPAVE Simulations

Total Rut Depth Versus Asphalt Pavement Analyzer

Figure 19 shows APA RD versus Total Rut Depth at 10 years from the FlexPAVE simulations and the results for all 18 mixtures and both A and D structures (FHWA, n.d.) Data points are shown distinctly for different structures and NMAS mixtures. However, the equation is fit using all data per structure. Figure 19 shows that APA RD exhibits the expected increasing trend with the Total Rut Depth, with R^2 of 0.5276 and 0.5197 for A and D structures, respectively. A 0.26-inch threshold of total rutting, as VDOT's *AASHTOWare Pavement ME User Manual* recommends, was used to obtain the threshold APA RD leading to 10.3- and 7.6-mm APA RD thresholds for structures A and D, respectively (VDOT, 2017).

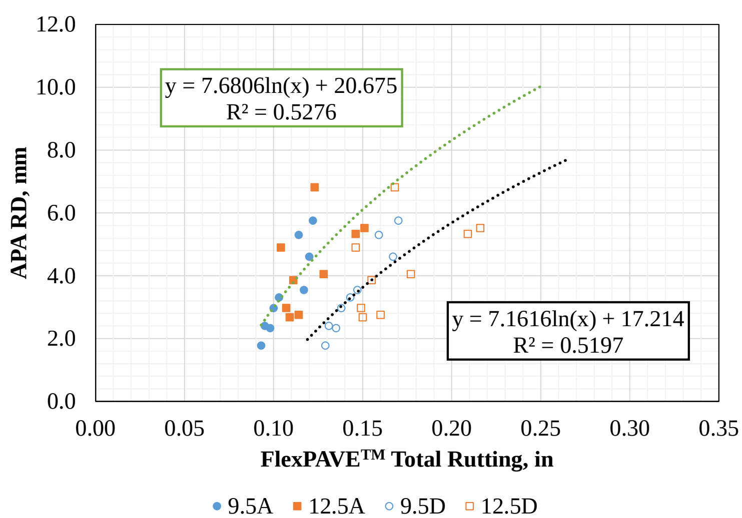


Figure 19. Asphalt Pavement Analyzer (APA) Rut Depth (RD) versus Total Rutting from FlexPave Simulations (FHWA, n.d.)

Top %Damage Versus Mass Loss

The Top %Damage was used for the remaining analyses because Top %Damage mostly represents the SM-incurred damage. In terms of damage, a different method was used to obtain candidate thresholds for ML and the CT index because of the absence of calibration between FlexPAVE damage and field cracking (FHWA, n.d.). The analysis for ML assumes the Top %Damage at year 10 is observed when ML is set to VDOT's 7.5% threshold. The ML value is then calculated to have the same Top %Damage but for year 17 for A structures and year 15 for D structures. The idea is to find a new threshold that would result in a pavement life extension of 7 years for A structures and 5 years for D structures. The research team decided to consider a 7-year life extension for A structures to avoid more frequent maintenance activities on secondary roads and to account for the fact that the foundation of A-level pavements may be more uncertain or in a poorer condition than the D-level structures.

Figure 20a shows a lack of correlation between ML and Top %Damage at year 10 for both structures. However, although this relationship does not exist for all mixtures per structure, 12.5-mm NMAS mixtures exhibit this relationship. Figure 20b shows the 12.5-mm NMAS mixtures relation between ML and Top %Damage at 10 and 17 years for A structures and at 10 and 15 years for D structures. Assuming 7.5% ML at year 10, the Top %Damage is 14.88 and 15.96% for 12.5A and 12.5D mixtures, respectively. The thresholds to have the same damage at the chosen later years are 5.9 and 6.3% for 12.5A and 12.5D mixtures, respectively.

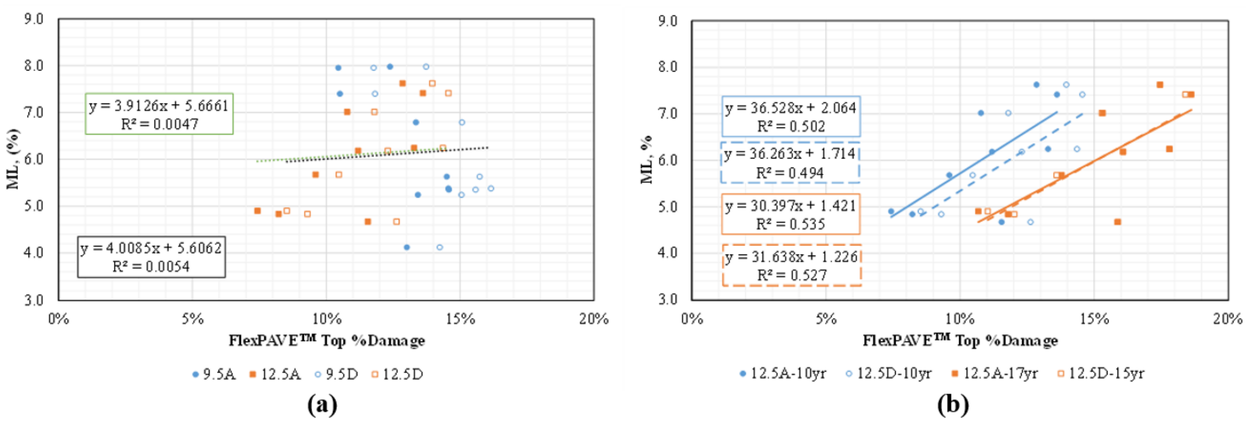


Figure 20. ML versus Top %Damage (a) at Year 10; (b) for 12.5-mm Nominal Maximum Aggregate Size Mixtures Excluding Outliers. ML = mass loss.

Top %Damage Versus Cracking Tolerance Index

The analysis for the CT index began similarly to ML analysis, in which the Top %Damage at year 10 is assumed to occur when the CT index is set to VDOT's threshold of 70. The CT index value is then calculated to have the same Top %Damage but for year 17 for A structures and year 15 for D structures based on the same rationale used for the ML analysis. Figure 21a shows the lack of correlation between the CT index and Top %Damage, when all the mixtures are used in both structures.

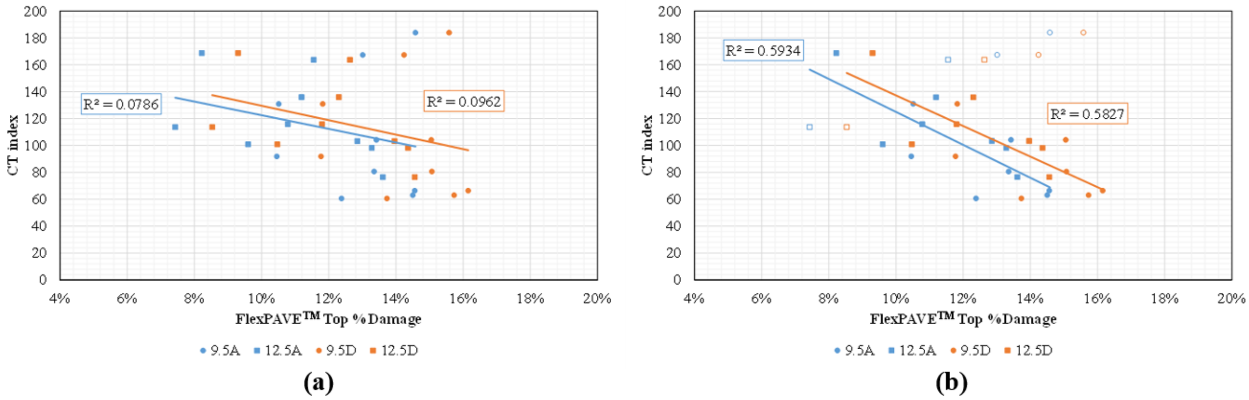


Figure 21. CT Index versus Top %Damage at (a) Year 10; (b) Year 10 Excluding Mixtures A1, C3, D3, and H. CT = cracking tolerance.

In an attempt to find a better correlation, Figure 21b shows the relationship between the CT index and Top %Damage after excluding four mixtures, two from each NMAS, namely mixtures A1, C3, D3, and H. In Figures 21a and 21b, the general trend is as expected, in which a higher CT index results in a lower Top %Damage, indicating better durability. Assuming a CT index of 70 at year 10, the Top %Damage is 14.48 and 15.91% for 12.5A and 12.5D mixtures, respectively. The thresholds to have the same damage at the chosen later years are 123.7 and 109.5 for A and D structures, respectively.

Predicted Life Extension as a Function of Increase in Cracking Tolerance Index

The analysis was repeated, such that the CT index needed to incrementally extend the pavement life was obtained. Figure 22 shows the procedure for D structures, in which the CT index versus Top %Damage is plotted at 10, 15, 20, 25, and 30 years. After conducting this procedure with 1-year increments for all A and D mixtures, the entire analysis was repeated in reference to VDOT's threshold being years 8 and 12, instead of year 10, to study the effect of conducting maintenance procedures in the range of 8 to 12 years. Figures 23a and 23b show the change of the CT index with time, in which every curve corresponds to the needed CT index to maintain the same damage as the reference year at which the curve starts (8, 10, or 12) for A and D structures, respectively. Figures 21a and 21b also show the required CT index to extend the pavement life by 7 and 5 years for A and D structures, respectively. The results show that in terms of life extension, changing the reference year does not significantly affect the needed CT index for either structure.

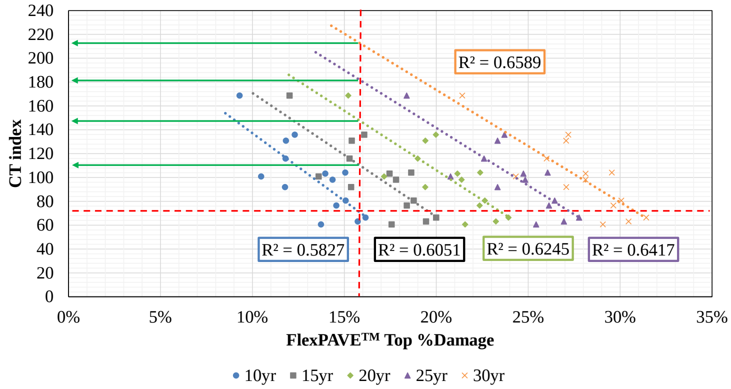
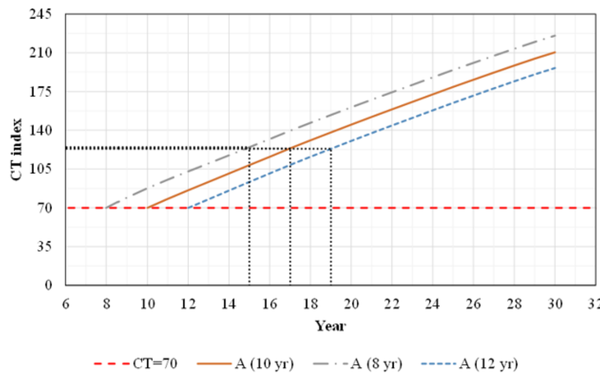


Figure 22. CT Index versus Top %Damage for D Structures at 10, 15, 20, 25, and 30 Years. CT = crakcing tolerance.



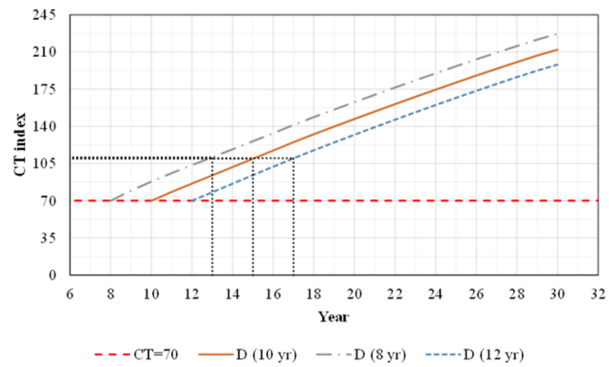


Figure 23. Change of Required CT Index with Time for (a) A Structures; (b) D Structures. CT = crakcing tolerance.

Furthermore, the analysis was also repeated separately for 9.5- and 12.5-mm NMAS to explore possible discrepancies in the resulting thresholds for mixtures with different NMAS. Figure 24 shows the results for A structures with year 10 as the reference. The results showed that the different analysis considerations did not affect the results until 12 years of extended life. After 12 years, 12.5-mm mixtures deviate from the 9.5-mm mixtures and the overall trend. However, this deviation fell within the repeatability limit of ASTM D8225, in which the standard deviation was below 13.5 when comparing the three curves at 20 years extended life, when the difference is the highest (ASTM International, 2019). In addition, because BMD is not used with the intention of extending life by 10 or more years, looking beyond that period is not practical. Similar observations were found when comparing the remaining 2 reference years for A structures and for the analysis using D structures. Thus, within the range of the intended 5 to 7 years of added life, the mixtures with different NMAS do not require a different threshold for the CT index.

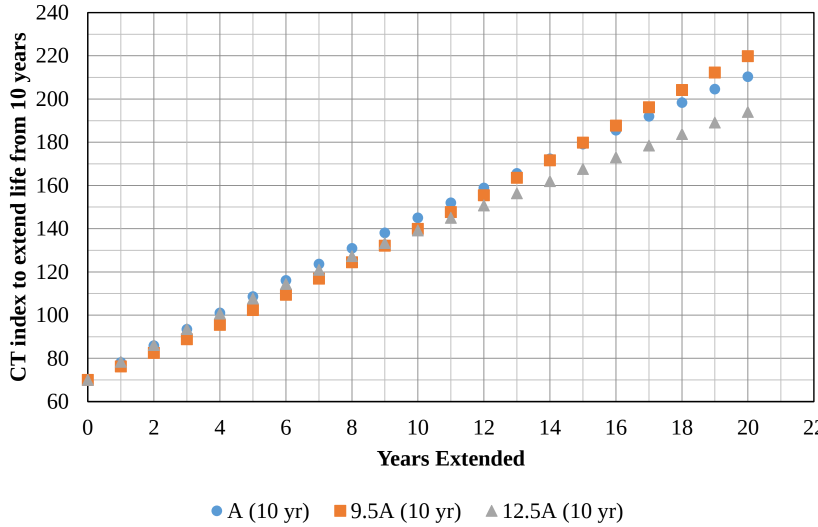
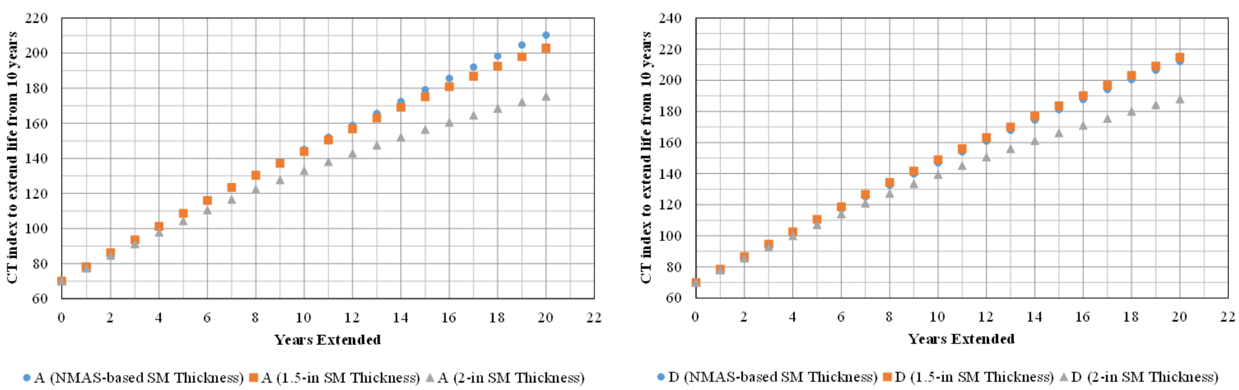


Figure 24. Effect of Nominal Maximum Aggregate Size on Change of CT Index with Time for A Structures. CT = cracking tolerance.

Finally, the effect of the SM thickness was studied by repeating all FlexPAVE simulations while maintaining the same SM thickness (FHWA, n.d.). The original runs had NMAS-based SM thickness such that A structures would have a 1.5-inch thick SM, and D structures would have a 2-inch thick SM. The additional simulations were conducted such that both A and D structures would have 1.5-inch thick SMs, or both would have 2-inch thick SMs. Figure 25a shows the results for A structures with year 10 as the reference for life extension. Results show that increasing the SM thickness leads to a lower required CT index, which explains the deviation observed with 12.5-inch mixtures previously. Although the deviation is more pronounced when all SMs are set to a 2-inch thickness, the deviation is still not significant in the range of 5 to 7 years of added life. For D structures, the observed deviation was less pronounced compared with the results of A structures (Figure 25b). Thus, the thresholds obtained previously from the original simulations' analysis remain as the candidate suggested CT index thresholds, with a 123.7 for A structures and 109.5 for D structures.



(a) **(b)**

Figure 25. Effect of SM Thickness on Change of CT Index with Time for (a) A Structures; (b) D Sturctures. CT = cracking tolerance; NMAS = nominal maximum aggregate size; SM = surface mixture.

Resulting Recommended Thresholds

Table 6 shows the resulting recommended thresholds for ML (12.5-mm NMAS), APA RD, and the CT index for both structures. Figures 26 through 28 show the performance maps for the 18 tested mixtures, with APA and the CT index thresholds from current VDOT specifications recommended for A structures and D structures, respectively (VDOT, 2020). Overall, results for both A and D structures suggest a need to increase the CT index threshold. Thresholds for A structures require the higher CT index of 123.7 but also allow more rutting in APA, with a 10.4-mm maximum RD. This increase in allowable rutting is beneficial to ensure a balanced performance because improving cracking performance often leads to more rutting. Although the CT index threshold increase for D structures to 109.5 is less than A structures, the APA RD threshold of 7.6 mm is slightly less than VDOT's 8-mm limit. This change is harder to achieve because it requires mixtures that perform simultaneously better in cracking and rutting. However, maximum APA RD from the 18 tested mixtures was 6.8 mm, suggesting that having a tighter APA threshold is possible. Comparing VDOT's thresholds in Figure 26 and the recommended thresholds in Figures 27 and 28, none of the mixtures fail the APA thresholds for either structure. However, for the CT index, increasing the threshold to 123.7 leads to 12 failed mixtures, and a threshold of 109.5 leads to 10 failed mixtures compared with only 3 failed mixtures when using VDOT's current threshold of 70.

Table 6. APA RD and CT Index Recommended Thresholds

Structure	A	D	Average
Mass Loss (%)	5.9	6.3	6.1
APA RD (mm)	10.4	7.6	9.0
CT Index	123.7	109.5	117.8

APA = asphalt pavement analyzer; CT = cracking tolerance; RD = rut depth.

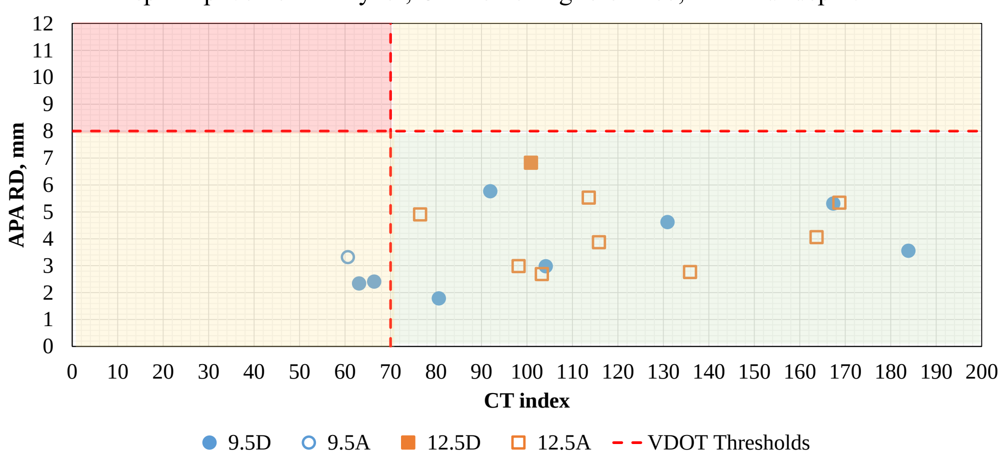


Figure 26. Mixtures Performance Map with VDOT Current Thresholds. APA = asphalt pavement analyzer; CT = cracking tolerance; RD = rut depth.

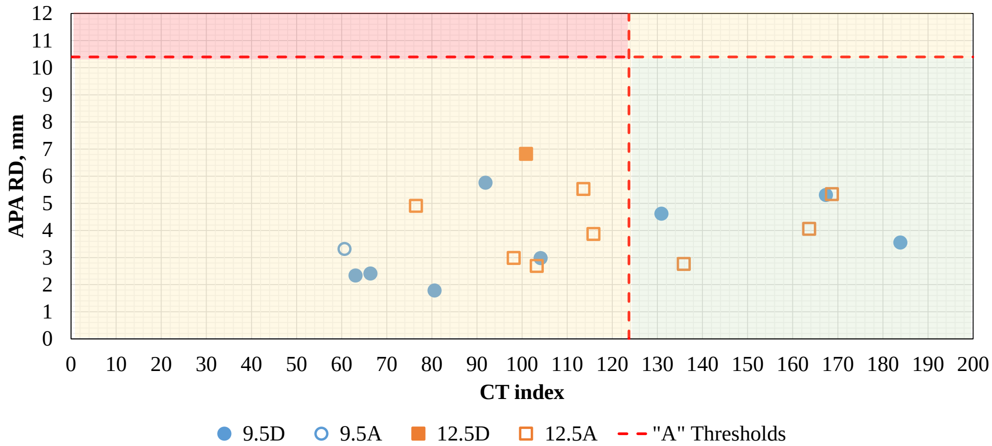


Figure 27. Mixtures Performance Map with A Structures Recommended Thresholds. APA = asphalt pavement analyzer; CT = cracking tolerance; RD = rut depth.

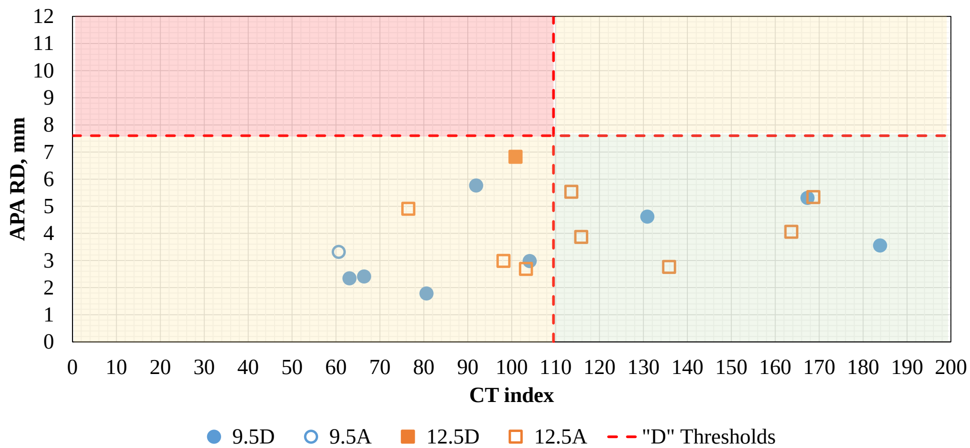


Figure 28. Mixtures Performance Map with D Structures Recommended Thresholds. APA = asphalt pavement analyzer; CT = cracking tolerance; RD = rut depth.

CONCLUSIONS

- *Despite the extensive work and development undertaken in the field of BMD, knowledge and practical gaps still persist.*

- *The results of the mixtures tested in this study indicate that VDOT's BMD tests demonstrate durability, rutting, and cracking performance discrimination potential and trends consistent with advanced tests measuring fundamental properties.*

- *Material-specific coefficients for the AASHTOWare Pavement ME Design fatigue model can be obtained from conversion of CF test results* (AASHTO, 2022a).

- *Current BMD thresholds need to be revised and refined. Traffic-based rutting and cracking thresholds should be incorporated.* In this study, the ME simulations indicated that mixtures intended for placement on moderate to heavy traffic pavement structures (D) should be designed to meet a maximum ML of 6.3% (instead of 7.5%, based on 12.5-mm NMAS mixtures), a minimum CT index of 110 (instead of 70), while maintaining the same APA RD threshold of a maximum 8.0 mm. Moreover, the ME simulations showed that mixtures intended for placement on low-volume traffic pavement structures (A) should be designed to meet a maximum ML of 5.9% (instead of 7.5%), a minimum CT index of 124, and a maximum APA RD of 10 mm.

RECOMMENDATIONS

1. *VDOT's Materials Division should continue to use the Cantabro test, APA rut test, and IDT-CT as part of the BMD framework for designing SMs with A and D designations.* Overall, the data presented and discussed in this report demonstrated that these tests exhibited performance trends pointing in the right direction, indicating either performance improvement or deterioration compared with advanced tests measuring fundamental properties.

2. *VDOT's Materials Division and VTRC should continue efforts to validate and refine the BMD special provisions by incorporating traffic-based cracking and rutting performance thresholds.* In this study, the ME simulations indicated that mixtures intended for placement on moderate to heavy traffic pavement structures (D) should be designed to meet a maximum ML of 6.3% (instead of 7.5%, based on 12.5-mm NMAS mixtures), a minimum CT index of 110 (instead of 70), while maintaining the same APA RD threshold of a maximum 8.0 mm. Moreover, the ME simulations showed that mixtures intended for placement on low-volume traffic pavement structures (A) should be designed to meet a maximum ML of 5.9% (instead of 7.5%), a minimum CT index of 124, and a maximum APA RD of 10 mm.

3. *VDOT's Materials Division and VTRC should consider benchmarking and establishing performance-based specifications specifically for mixtures intended for use on low-volume roads.* The mechanistic simulations and analysis conducted in this study provided insights into SMs' performance when placed on low-volume traffic pavement structures. However, all mixtures currently available and evaluated in this study were D mixtures designed to be placed on moderate-volume routes.

IMPLEMENTATION AND BENEFITS

The researchers and the technical review panel (listed in the Acknowledgments) for the project collaborated to craft a plan to implement the study recommendations and to 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 the process of BMD implementation using the currently selected performance tests for all SM-9.5

and SM-12.5 mixtures with A and D designations to be produced during the 2024 and 2025 paving seasons.

Regarding Recommendation 2, VDOT's Materials Division agrees to continue to collaborate with VTRC on ongoing research related to validation and refinement of BMD special provisions. VTRC Project 124200 (Completed, VTRC Report 25-R16), *Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents Under Laboratory-Scale and Full-Scale Accelerated Testing*, revealed that the selected BMD tests (Cantabro test, APA rut test, and IDT-CT) characterized the laboratory performance of a suite of mixtures similarly to their performance observed under accelerated pavement testing (Habbouche et. al., 2025). Moreover, VTRC Project 124332, Field Validation of Balanced Mix Design Initial Criteria, is ongoing. This project aims to validate the initial BMD test criteria determined from the laboratory testing of 11 50-gyration SMs placed in 2015. Analyses will be performed to evaluate the performance of the mixtures during 8 years of service and determine the appropriateness of the current BMD test thresholds.

Regarding Recommendation 3, a research needs statement addressing the evaluation of BMD for low-volume roads has been drafted and will be revised and submitted to the appropriate Pavement Research Advisory Subcommittee for funding consideration. If selected, VTRC will undertake a project to benchmark and establish performance-based specifications specifically for mixtures intended for use on low-volume roads. The research needs statement is expected to be revised and submitted for consideration by the Advisory Subcommittee no later than the end of calendar year 2027.

Benefits

This study revealed several potential benefits, established links between BMD and fundamental tests, and validated the use of VDOT-selected BMD performance tests using these connections. Furthermore, this study verified and refined performance thresholds using an ME approach (Approach II), rather than relying solely on empirical methods and statistical analyses (Approach I). This ME approach uses a deeper understanding of pavement behavior and facilitates the development of more accurate performance criteria until a full field performance validation can be completed. Finally, the research resulted in initial traffic-based BMD cracking and rutting performance thresholds.

Ensuring greater confidence in the performance-based thresholds currently in use and continuously validating and refining them will contribute to sustained progress in BMD implementation and the production of longer lasting asphalt mixtures.

As mentioned previously, Approach I, initially used to define VDOT BMD thresholds, was based on descriptive statistical parameters from an experimental plan consisting of 13 mixtures. Using these thresholds was expected to improve the performance of at least one-half (50%) of the VDOT SM D asphalt mixtures paved (non-BMD). The following paragraph and table present a ***hypothetical*** example of the cost avoidance resulting from using the BMD approach successfully.

Table 7 summarizes the total tonnage of BMD SM D mixes and tonnage of those mixes by district for the 2024 pavement maintenance contracts, along with the average associated price per ton. According to Approach I, if BMD successfully extends the performance life of pavements by at least 1 year, a total amount of about $82.3 million per year (calculated as 50% of approximately 1.5 million tons multiplied by $108.52) of paving will be deferred. This number is expected to increase significantly given that BMD is anticipated to extend the performance life of pavements by more than 1 year. Table 7 presents a breakdown of deferred costs per district.

Table 7. Summary Table of Tonnage, Price, and Deferred Costs of Balanced Mix Design Mixture D SMs Provided as a Hypothetical Example Based on Approach I

District	SM D Mixes, Tons	Price per Ton, $	Deferred Cost, Million $
Bristol	NA	NA	NA
Culpeper	179,579	94.02	~8.4
Fredericksburg	263,436	112.23	~14.8
Hampton Roads	204,691	126.91	~13.0
Lynchburg	137,565	106.06	~7.3
Northern Virginia	250,159	101.68	~12.7
Richmond	156,717	100.02	~7.8
Salem	188,124	118.16	~11.1
Staunton	135,796	112.28	~7.6
Total	**1,516,067**	**108.52**	**~82.3**

D = mixture designation; NA = not applicable; SM = surface mixture.

Approach II uses empirical test results compared with results from fundamental tests to establish performance criteria. Subsequently, these criteria can be integrated into ME design alongside the volumetric properties of asphalt mixtures. For example, in this study, ME simulations indicated that mixtures intended for placement on moderate to heavy traffic pavement structures (D) should be designed to meet a maximum ML of 6.3% (instead of 7.5%, based on 12.5-mm NMAS mixtures), an increased minimum CT index of 110 (instead of 70), while maintaining the same APA RD threshold of a maximum 8.0 mm. After assessing the various BMD mixtures designed and placed in Virginia since 2019, these more conservative BMD thresholds are expected to result in the improvement of approximately 80% of the non-BMD mixtures, addressing an additional 30% of the mixtures designed using the current BMD specifications. The following paragraph and table present a hypothetical example of the cost avoidance resulting from using the BMD approach successfully with the refined thresholds for D mixtures.

Table 8 summarizes the total tonnage of BMD SM D mixes and those by district for the 2024 pavement maintenance contracts, along with the augmented associated price per ton. **For hypothetical calculation purposes**, these prices are assumed to be 10% higher compared with those in Table 7. This assumption, valid only for the hypothetical example presented here, encompasses inflation, the supply chain, and additional effort that contractors might need to undertake to design mixtures that perform better to meet the revised or refined thresholds. After applying the Approach II thresholds, if BMD successfully extends the performance life of pavements by at least 1 year, an additional total amount of about $54.3 million per year (calculated as 30% of approximately 1.5 million tons multiplied by $119.37) will be deferred. This number is expected to increase significantly given that BMD is anticipated to extend the performance life of pavements by more than 1 year. Table 8 presents a breakdown of these additional deferred costs per district.

Table 8. Summary Table of Tonnage, Price, and Additional Deferred Costs of Balanced Mix Design Mixture D SMs Based on Application of Approach II

District	SM D Mixes, Tons	Augmented Price per Ton, $	Additional Deferred Cost, Million $
Bristol	NA	NA	NA
Culpeper	179,579	103.42	~5.6
Fredericksburg	263,436	123.45	~9.8
Hampton Roads	204,691	139.60	~8.6
Lynchburg	137,565	116.67	~4.8
Northern Virginia	250,159	111.85	~8.4
Richmond	156,717	110.02	~5.2
Salem	188,124	129.98	~7.3
Staunton	135,796	123.51	~5.0
Total	**1,516,067**	**119.37**	**~54.3**

D = mixture designation; NA = not applicable; SM = surface mixture.

The calculations in the previous paragraph, based on observations from Approaches I and II, are hypothetical and assume that the underlying pavement conditions are either in good or equal condition compared with the structures on which non-BMD and BMD mixes are designed using the current specifications. Moreover, the estimated cost avoidance may fluctuate with additional data collected as the agency progresses with BMD implementation.

Finally, Approach III, in which empirical test results are directly correlated to the in-service performance of asphalt mixtures, is the baseline practice for establishing performance thresholds. This approach is essential and will be used to verify and refine the established thresholds of the tests once mixtures are in service long enough to begin to deteriorate. A more accurate estimate of the quantitative benefits of implementing BMD is expected upon completing this approach.

ACKNOWLEDGMENTS

The authors are grateful to the following individuals who served on the technical review panel for this study: Girum Merine (Project Champion and Pavement Program Manager, VDOT Materials Division), Angela Beyke (Assistant State Materials Engineer, VDOT Materials Division), Bryan C. Smith (Senior Asphalt Pavement Field Engineer, VDOT Materials Division), Donald (Clyde) Landreth (Materials Pavement Engineer, Salem Districts Materials), Jason T. Provines (Senior Research Scientist, VTRC), Robert Crandol (Assistant State Maintenance Engineer, VDOT Maintenance Division), and Thomas Schinkel (District Materials Engineer, VDOT Richmond District).

The authors thank Andrew Barbour, Brian Powell, Derek Lister, David (Zack) McKee, Jacob Oliver, Jennifer Samuels, and Troy Deeds of VTRC and Danny Martinez Rodriguez, formerly of VTRC, for their assistance in sampling and testing the collected mixtures and extracted and recovered asphalt binders.

REFERENCES

Abatech, Inc. *Rheology Analysis* (software). Version 2.0. Blooming Glen, PA, 2022.

American Association of State Highway and Transportation Officials. *AASHTOWare Pavement* (software), 2004. https://www.aashtoware.org/products/pavement/pavement-overview/. Accessed April 4, 2025.

American Association of State Highway and Transportation Officials. *Mechanistic-Empirical Pavement Design Guide, A Manual of Practice*. Second Edition. Washington, DC, 2015.

American Association of State Highway and Transportation Officials. *Standard Specification for Performance-Graded Asphalt Binder*. M 320. Washington, DC, 2017.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Quantitative Extraction of Asphalt Binder from Hot Mix Asphalt (HMA)*. T 164-14. Washington, DC, 2018.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Determining Rutting Susceptibility of Hot Mix Asphalt (HMA) Using the Asphalt Pavement Analyzer (APA)*. T 340. Washington, DC, 2019a.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Determining the Flexural Creep Stiffness of Asphalt Binder Using the Bending Beam Rheometer (BBR)*. T 313. Washington, DC, 2019b.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Preparing and Determining the Density of Asphalt Mixture Specimens by Means of the Superpave Gyratory Compactor*. T 312. Washington, DC, 2019c.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Quantitative Extraction and Recovery of Asphalt Binder from Asphalt Mixtures*. T 319. Washington, DC, 2019d.

American Association of State Highway and Transportation Officials. *Standard Specification for Performance-Graded Asphalt Binder Using Multiple Stress Creep Recovery (MSCR) Test*. M 332. Washington, DC, 2019e.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Sieve Analysis of Fine and Coarse Aggregates*. T 27. Washington, DC, 2020a.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Theoretical Maximum Specific Gravity (G_{mm}) and Density of Hot-Mix Asphalt (HMA)*. T 209. Washington, DC, 2020b.

American Association of State Highway and Transportation Officials. *Standard Practice for Balanced Design of Asphalt Mixtures*. PP 105. Washington, DC, 2020c.

American Association of State Highway and Transportation Officials. *Standard Test Method for Materials Finer than 75-µm (No. 200) Sieve in Mineral Aggregates by Washing*. T 11. Washington, DC, 2020e.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Stress Sweep Rutting (SSR) Test Using Asphalt Mixture Performance Tester (AMPT)*. TP 134. Washington, DC, 2021.

American Association of State Highway and Transportation Officials. AASHTOWare Pavement ME Design (software), 2022a. https://www.aashtoware.org/products/pavement/pavement-overview/. Accessed April 4, 2025.

American Association of State Highway and Transportation Officials. Standard Method of Test for Cantabro Abrasion Loss of Asphalt Mixture Specimens. T 401. Washington, DC, 2022b.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Resistance of Compacted Asphalt Mixtures to Moisture-Induced Damage*. T 283. Washington, DC, 2022c.

American Association of State Highway and Transportation Officials. *Standard Specification for Rail-Steel and Axle-Steel Deformed Bars for Concrete Reinforcement*. M 322M/M 322. Washington, DC, 2022d.

American Association of State Highway and Transportation Officials. *Standard Method of Test for Determining the Damage Characteristic Curve and Failure Criterion Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test*. AASHTO T 411. Washington, DC, 2023.

Anderson, R.M., King, G.N., Douglas, H., and Blankenship, P.B. Evaluation of the Relationship Between Asphalt Binder Properties and Non–Load Related Cracking. *Journal of the Association of Asphalt Paving Technologist*, Vol. 80, 2011, pp. 615–649.

ASTM International. *Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature.* ASTM D8225-19. West Conshohocken, PA, 2019.

Bowers, B.F., and Diefenderfer, S.D. *Technical Memorandum: Benchmarking of Asphalt Mixtures – Summary of Findings and Preliminary Recommendations*. Virginia Transportation Research Council, Charlottesville, VA, 2018.

Bowers, B.F., Diefenderfer, S.D., Moore, N. and Lynn, T. Balanced Mix Design and Benchmarking: A Case Study in Establishing Performance Test Thresholds.

Transportation Research Record: Journal of the Transportation Research Board, Vol. 2676, No. 2676, 2022, pp. 586–598.

Boz, I., Habbouche, J., Diefenderfer, S.D., Coffey, G.P., Ozbulut, O.E., and Seitllary, A. Simple and Practical Tests for Rutting Evaluation of Asphalt Mixtures in the Balanced Mix Design Process. VTRC 23-R11. Virginia Transportation Research Council, Charlottesville, VA, 2023.

Boz, I., Habbouche, J., and Diefenderfer, S.D. The Use of Indirect Tensile Test to Evaluate the Resistance of Asphalt Mixtures to Cracking and Moisture-Induced Damage. In *Proceedings of the International Airfield and Highway Pavements Conference*. American Society of Civil Engineers, Reston, VA, 2021, pp. 104–114.

Diefenderfer, S.D., and Boz, I. *Field Validation of Balanced Mix Design Initial Criteria.* Project No: 124332. Virginia Transportation Research Council, Charlottesville, VA, 2025. https://vtrc.virginia.gov/projects/all-projects/124332/. Accessed April 16, 2024.

Diefenderfer, S.D., Boz, I., and Habbouche, J. *Balanced Mix Design for Surface Asphalt Mixtures: Phase I: Initial Roadmap Development and Specification Verification.* VTRC 21-R15. Virginia Transportation Research Council, Charlottesville, VA, 2021a.

Diefenderfer, S.D., Boz, I., and Habbouche, J. *Balanced Mix Design for Surface Asphalt Mixtures: 2019 Field Trials.* VTRC 21-R21. Virginia Transportation Research Council, Charlottesville, VA, 2021b.

FHWA. *FlexPAVE* (software), n.d.

Federal Highway Administration. *Cyclic Fatigue Index Parameter (S_{app}) for Asphalt Performance Engineered Mixture Design*. FHWA-HIF-19-091. Washington, DC, 2019a.

Federal Highway Administration. *FlexMAT Cracking* (software). Version 2.1.3b, 2019b.

Federal Highway Administration. *FlexMAT Rutting* (software). 2019c.

Federal Highway Administration. *Delta Tc Binder Specification Parameter*. FHWA-HIF-21-042. Washington, DC, 2021a.

Federal Highway Administration. *Rutting Strain Index (RSI) Parameter for Asphalt Balanced Mixture Design*. FHWA-HRT-21-044. Washington, DC, 2021b.

Ghanbari, A., Underwood, B.S., and Kim, Y. K. Development of a Rutting Index Parameter Based on the Stress Sweep Rutting Test and Permanent Deformation Shift Model. *International Journal of Pavement Engineering*, Vol. 23, No. 2, 2020, pp. 387–399.

Glover, C.J., Davison, R.R., Domke, C.H., Ruan, Y., Juristyarini, P., Knorr, D.B., and Jung, S.H. *Development of a New Method for Assessing Asphalt Binder Durability With Field*

Validation. FHWA/TX-05/1872-2. Texas Transportation Institute, College Station, TX, 2005.

Habbouche, J., Boz, I., Hajj, E.Y., and Morian, N.E. Influence of Aging on Rheology- and Chemistry-Based Properties of High Polymer-Modified Asphalt Binders. *International Journal of Pavement Engineering*, Vol. 23, 2022, pp. 3285–3303.

Habbouche, J., Diefenderfer, S.D., Diefenderfer, B.K., Flintsch, G., Tong, B., and Urbaez Perez, E. *Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents Under Laboratory-Scale and Full-Scale Accelerated Testing*. VTRC 25-R16. Virginia Transportation Research Council, Charlottesville, VA, 2025.

Kim, Y.R., Castorena, C., Elwardany, M., Rad, F.Y., Underwood, S., Gundla, A., Gudipudi, P., Farrar, M.J., and Glaser, R.R. *Long-Term Aging of Asphalt Mixtures for Performance Testing and Prediction*. NCHRP Report 871. National Academy of Sciences, Washington, DC, 2018.

Kim, Y.R., Castorena, C., Saheh, N.F., Braswell, E., Elwardany, M., and Rad, F.Y. *Long-Term Aging of Asphalt Mixtures for Performance Testing and Prediction: Phase III Results*. NCHRP Report 973. National Academy of Sciences, Washington, DC, 2021.

Olidis, C., Eng, P., and Hein, D. Guide for the Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures: Materials Characterization. Presented at the *2004 Annual Conference of the Transportation Association of Canada*, Quebec City, Canada, 2004.

Rowe, G.M., Sharrock, M.J., Bouldin, M.G., and Dongre, R.N. Advanced Techniques to Develop Asphalt Master Curves From the Bending Beam Rheometer. *Petroleum and Coal*, Vol. 43, 2011, pp. 54–59.

Sadat Sakhaei Far, M. *Development of New Dynamic Modulus (|E*|) Predictive Models for Hot Mix Asphalt Mixtures*. Doctoral Dissertation, North Carolina State University, Raleigh, NC, 2011.

Virginia Department of Transportation. *AASHTOWare Pavement ME User Manual*. Richmond, VA, 2017.

Virginia Department of Transportation. *Road and Bridge Specifications*. Richmond, VA, 2020.

Virginia Test Methods. *VTM-102: Determination of Asphalt Content From Asphalt Paving Mixtures by the Ignition Method*. Virginia Department of Transportation, Richmond, VA, 2013.

Von Quintus, H.L., Mallela, J., Bonaquist, R., Schwartz, C.W., and Carvalho, R.L. *Calibration of Rutting Models for Structural and Mix Design*. NCHRP Report 719. Transportation Research Board, Washington, DC, 2012.

Wang, Y.D., Underwood, B.S., and Kim, Y.K. 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.

West, R.C., Rodezno, C., Leiva, F., and Yin, F. *NCHRP Project 20-07/Task 406 Development of a Framework for Balanced Mix Design*. Transportation Research Board, Washington, DC, 2018.

Witczak, M.W., Kaloush, K., Pellinen, T., El-Basyouny, M., and Von-Quintus, H.L. *Simple Performance Test for Superpave Mix Design*. NCHRP Report 465. Transportation Research Board, Washington, DC, 2002.

Yin, F., and West, R. *Balanced Mix Design Resource Guide*. NAPA IS-143. National Asphalt Pavement Association, Greenbelt, MD, 2021.

APPENDIX A: USING CYCLIC FATIGUE RESULTS TO OBTAIN FATIGUE MODEL COEFFICIENTS

This appendix describes the 7-step procedure to obtain the material-specific fatigue coefficients from the results of the cyclic fatigue (CF) test.

Step 1: Cyclic Fatigue N_f

Calculated the number of cycles to failure (N_f) from the CF test results and Equation A1 for all the mixtures at multiple temperatures and strain levels. This calculation used the temperatures 10, 20, and 27°C. The N_f calculated in this step are referred to as $N_{f(CF)}$.

$$N_{f(CF)} = \frac{\left(D^R \times \frac{C_{12}+\rho}{C_{11}\times\rho}\right)^{\frac{\rho}{C_{12}}}(f_{red})(2^{\alpha})}{\rho(C_{11}C_{12})^{\alpha}\left(\frac{(\beta+1)}{2}(\varepsilon_{0,pp})(|E^*|_{LVE})\right)^{2\alpha} K_1} \qquad \text{(Eq. A1)}$$

Where:

D^R = *temperature-dependent failure criterion* = $0.1 + \frac{0.85}{1+e^{d+4.49\times log|E^*|}}$.

$d = ln\left[\left(\frac{0.85}{D^R-0.1}\right) - 1\right] - 4.49 \times log|E^*|$.

C_{11}, C_{12} = fitting coefficients for the power model of the damage characteristic curve.

$\rho = 1 - \alpha \times C_{12} + \alpha$.

α = continuum damage model power term.

f_{red} = reduced frequency (Hz).

$\varepsilon_{0,pp}$= peak-to-peak initial applied strain.

$|E^*|_{LVE}$= linear viscoelastic dynamic modulus (kPa).

$\beta = 1$ for direct tension.

$K_1 = 0.3005 - 0.0259\alpha$.

Step 2: Fatigue Model Coefficients from the Cyclic Fatigue Test

Used the $N_{f(CF)}$ from Step 1 and the empirical fatigue law in Equation A2 to backcalculate the fatigue model coefficients, denoted as $k_{f1(CF)}$, $k_{f2(CF)}$, and $k_{f3(CF)}$. Figure A1 shows the resultant N_f as a function of the strain level at each temperature for mixture C3. The data points in Figure A1 correspond to the $N_{f(CF)}$ measured in Step 1, and the fitted curves correspond to the N_f calculated from the empirical fatigue law using the backcalculated CF coefficients for mixture C3. Table A1 lists the CF coefficients for the 18 mixtures evaluated in this study and for the default fatigue coefficients calibrated for VDOT. Note that for the VDOT default calibrations, the coefficients are just k_{f1}, k_{f2}, and k_{f3} because they were not determined using CF tests. Table A1 shows that the coefficients from the CF test are very different from VDOT's default coefficients.

$$N_f = k_{f1(CF)} \left(\frac{1}{\varepsilon}\right)^{k_{f2(CF)}} \left(\frac{1}{|E^*|}\right)^{k_{f3(CF)}} \qquad \text{(Eq. A2)}$$

Where:
N_f = number of cycles to failure.
ε = peak-to-peak strain.
$|E^*|$= dynamic modulus, psi.
$k_{f1(CF)}, k_{f2(CF)}, k_{f3(CF)}$= fatigue model coefficients.

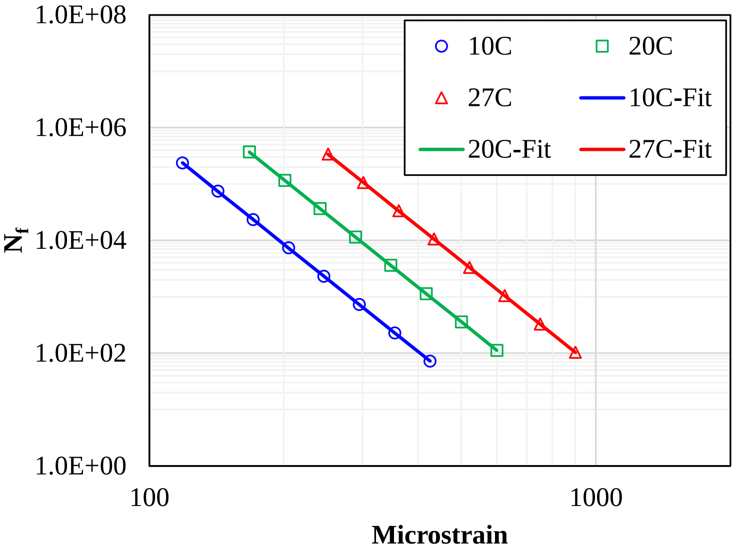


Figure A1. N_f Versus Microstrain for Mixture C3

Table A1. Fatigue Model Coefficients from the Cyclic Fatigue Test

Mixture Identification	$k_{f1(CF)}$	$k_{f2(CF)}$	$k_{f3(CF)}$
VDOT Default	*0.007566*	*3.9492*	*1.281*
A1	1.46E+12	6.465	5.100
A2	1.32E+07	6.704	4.432
B1	1.36E+07	7.292	4.757
B2	4.67E+05	6.972	4.288
B3	7.08E+08	6.934	4.789
C1	3.13E+13	6.672	5.488
C2	3.00E+09	6.681	4.767
C3	1.62E+13	6.342	5.243
D1	2.07E+04	6.763	4.058
D2	5.24E+13	7.221	5.911
D3	3.70E+13	7.381	5.932
E1	6.46E+06	6.872	4.500
E2	3.78E+11	7.034	5.370
F1	8.48E+11	6.730	5.289
F2	5.41E+11	6.715	5.255
G	9.11E+10	6.881	5.133

Mixture Identification	$k_{f1(CF)}$	$k_{f2(CF)}$	$k_{f3(CF)}$
H	7.07E+13	6.818	5.614
I	9.59E+12	7.043	5.616

Step 3: N_f Using Default VDOT Coefficients

VDOT's default locally calibrated fatigue model coefficients were used in the empirical fatigue law from Equation A2 to obtain the N_f for typical VDOT mixtures at strain levels from 200 to 900 microns and at the three temperatures selected in Step 1.

Step 4: Central Tendency of the Control Mixtures

The central tendency of the control mixtures was obtained by averaging their $N_{f(CF)}$ to obtain $N_{f(CF,avg)}$. All the control mixtures, excluding mixture A1, were used for this calculation because A1 had significantly different behavior compared with the other control mixtures, which can be attributed to A1's higher production asphalt content. Figure A2 shows the N_f as a function of strain level for all the control mixtures excluding A1, the averaged central tendency, and VDOT's default model for comparison. The fatigue coefficients were calculated for the central tendency, as in Step 2, resulting in $k_{f1(CF,avg)}$, $k_{f2(CF,avg)}$, and $k_{f3(CF,avg)}$.

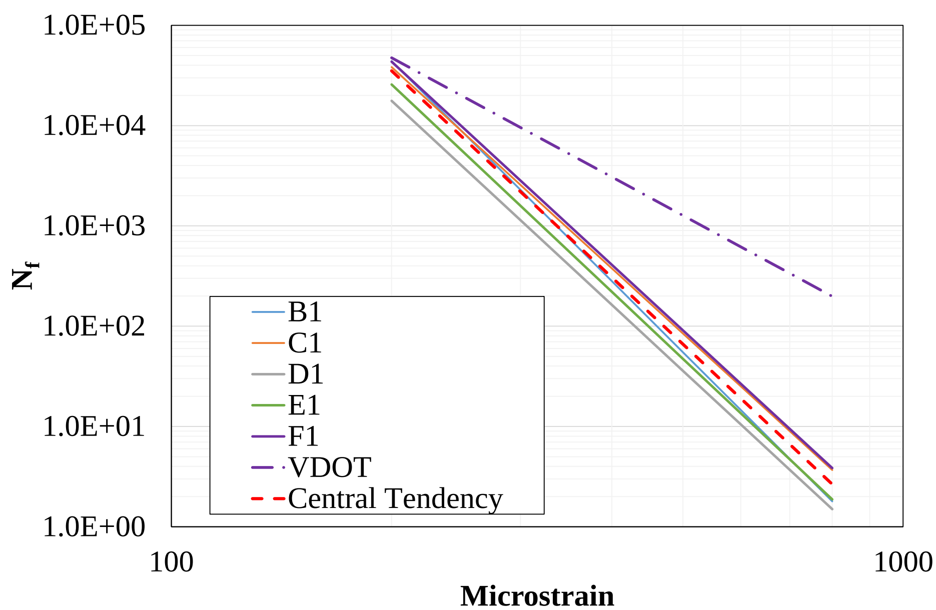


Figure A2. N_f of Control Mixtures, Central Tendency, and VDOT's Default Model at 20°C

Step 5: Scale the Central Tendency to Match VDOT's Default

The central tendency was assumed to represent the typical VDOT mixture having the default VDOT calibrated fatigue coefficients so that coefficients $k_{f2(CF)}$ and $k_{f3(CF)}$ can be scaled for all the mixtures, as Equations A3 and A4 show. The scaled coefficients denoted as $k_{f2(ME)}$ and $k_{f3(ME)}$ become more similar to VDOT's default values and, thus, are the final unique coefficients to be used in AASHTOWare Pavement ME Design (AASHTO, 2022a).

$$k_{f2(ME)} = k_{f2(CF)} \times \frac{k_{f2(VDOT)}}{k_{f2(CF,avg)}} \quad \text{(Eq. A3)}$$

$$k_{f3(ME)} = k_{f3(CF)} \times \frac{k_{f3(VDOT)}}{k_{f3(CF,avg)}} \quad \text{(Eq. A4)}$$

Step 6: Scale All Mixtures to Match VDOT's Default

Similar to Step 5, the $N_{f(CF)}$ for all mixtures are scaled such that the $N_{f(CF)}$ for all mixtures are comparable with VDOT's default. The scaled values are denoted as $N_{f(ME)}$ and are calculated using Equation A5. Figure A3 shows the N_f for the control mixtures, the central tendency, and the default VDOT after scaling. Figure A2 showed the difference between $N_{f(CF)}$ for the control mixtures and the N_f from VDOT's default model. Figure A3 demonstrates that after scaling the $N_{f(CF)}$, VDOT's model replaces the central tendency such that the position of all mixtures relative to the central tendency is now relative to VDOT's default model.

$$N_{f(ME)} = N_{f(CF)} \times \frac{N_{f(VDOT)}}{N_{f(CF,avg)}} \quad \text{(Eq. A5)}$$

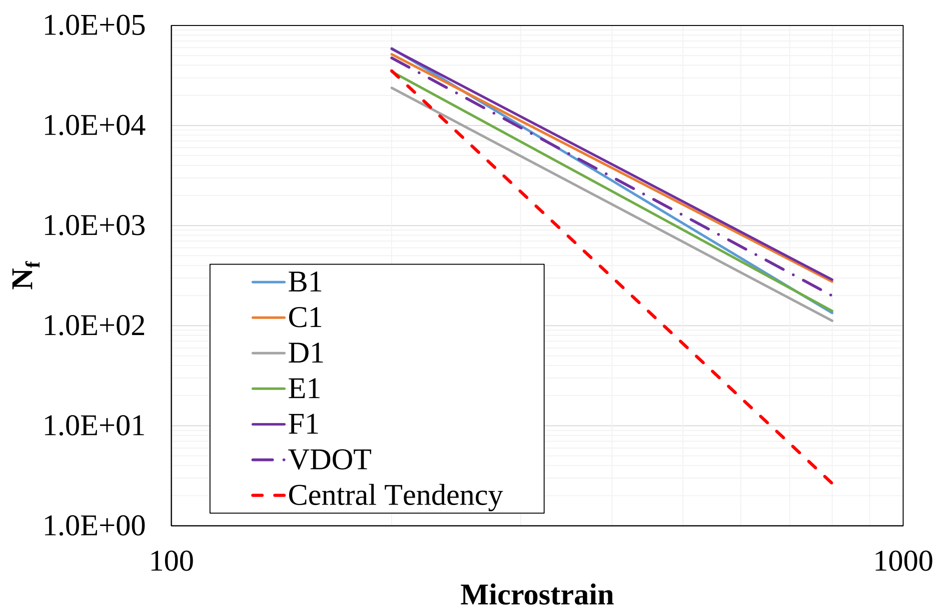


Figure A3. Scaled N_f of Control Mixtures, Relative to the Cyclic Fatigue N_f Central Tendency and to VDOT's Default Model at 20°C

Step 7: Finalize Model to Use in AASHTOWare Pavement ME Design

This final step used $k_{f2(ME)}$, $k_{f3(ME)}$, and $N_{f(ME)}$ in the empirical fatigue law from Equation A2 to obtain $k_{f1(ME)}$, finalizing the fatigue model coefficients for use in AASHTOWare Pavement ME Design (AASHTO, 2022a). Figures A4 and A5 show the $N_{f(ME)}$ for all the mixtures with the control mixtures as solid lines and the balanced mix design mixtures as dashed or dotted lines. Mixtures G, H, and I were plotted with VDOT's default model because those mixtures had no reference control mixture. The results showed that all balanced mix design mixtures had higher cycles to failure compared with their control mixtures except A1. Mixtures G, H, and I all had higher cycles to failure compared with VDOT's default model, which implies that those mixtures perform better than the average control mixture at the same strain level. Table A2 shows the final fatigue coefficients used in AASHTOWare Pavement ME Design.

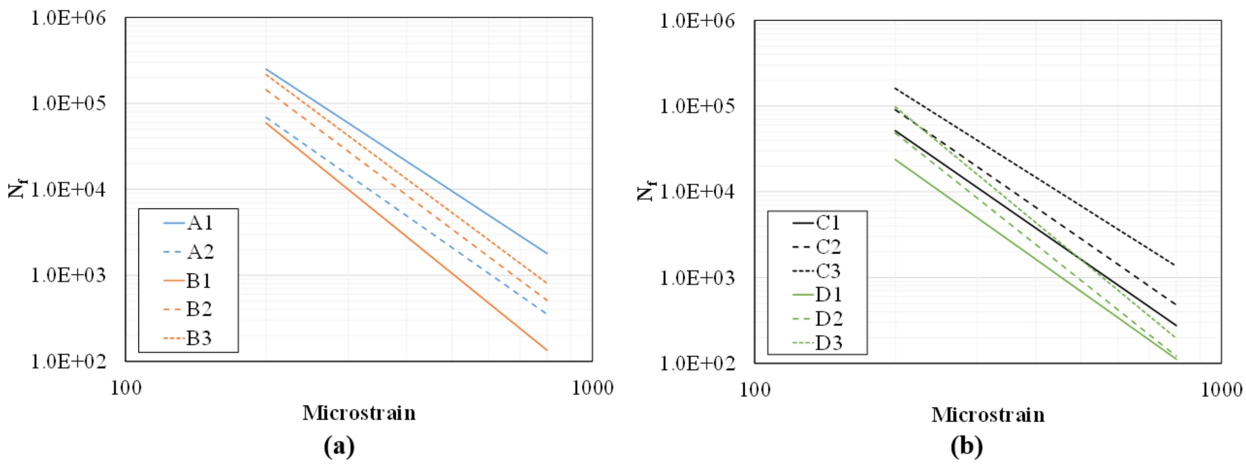


Figure A4. ***$N_{f(ME)}$*** **at 20°C for Mixtures: (a) A1, A2, B1, B2, and B3; (b) C1, C2, C3, D1, D2, and D3**

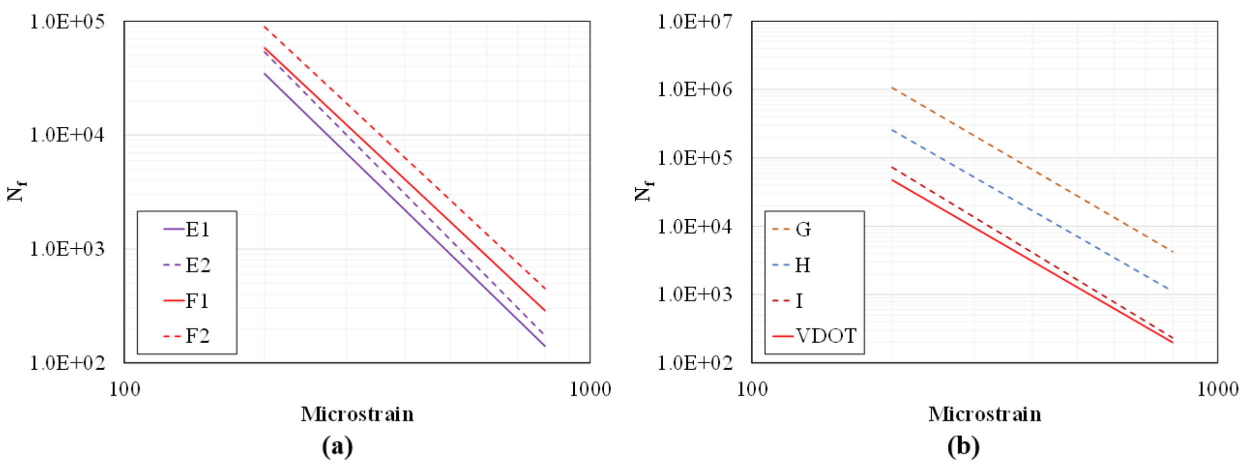


Figure A5. ***$N_{f(ME)}$*** **at 20°C for Mixtures: (a) E1, E2, F1, and F2; (b) G, H, I, and VDOT's Default**

Table A2. Fatigue Model Coefficients for AASHTOWare Pavement ME Design

Mixture	$k_{f1(ME)}$	$k_{f2(ME)}$	$k_{f3(ME)}$
VDOT_Default	*0.007566*	*3.9492*	*1.281*
A1	0.161387	3.7287	1.252
A2	0.001685	3.8667	1.088
B1	0.000236	4.2059	1.168
B2	0.000554	4.0211	1.052
B3	0.005103	3.9994	1.176
C1	0.054725	3.8482	1.347
C2	0.008434	3.8530	1.170
C3	0.315946	3.6579	1.287
D1	0.000129	3.9005	0.996
D2	0.010317	4.1649	1.451
D3	0.009669	4.2572	1.456
E1	0.000505	3.9636	1.105
E2	0.006066	4.0572	1.318
F1	0.020923	3.8816	1.298
F2	0.027221	3.8729	1.290
G	0.063018	3.9689	1.260
H	0.141057	3.9321	1.378
I	0.016115	4.0620	1.378

APPENDIX B: ADVANCED PERFORMANCE CHARACTERISTICS: RUTTING—REPEATED TRIAXIAL LOAD TEST

Table B1. Summary of MEPDG Rutting Model Cofficicnets for 11 Evaluated Mixtures

Mix Identification	Rutting Model Coefficients			R^2
	k_{r1}	k_{r2}	k_{r3}	
VDOT (Default)[a]	– 3.35412	1.560600	0.479100	NA
B1	– 3.901808	1.880801	0.285607	0.98
B2	– 3.299445	1.596693	0.273224	0.98
B3	– 3.897680	1.943945	0.266154	0.99
C1	– 3.689642	1.865789	0.276620	0.97
C2	– 2.513506	1.409024	0.229382	0.98
C3	– 5.835626	2.881373	0.260193	0.96
D1	– 2.526606	1.167820	0.311814	0.99
D2	– 4.107706	1.941102	0.351777	0.97
F1	– 6.940332	3.347773	0.360724	1.00
H	– 4.378998	2.112333	0.322362	1.00
I	– 2.930598	1.506011	0.266277	0.92

MEPDG = mechanistic-empirical pavement design guide; NA = not applicable.
[a] VDOT default values were identical to those of the MEPDG Manual of Practice, which were determined based on unconfined repeated load permanent deformation (RLPD) testing (AASHTO, 2015). The rutting model coefficients for the mixtures evaluated in this study were determined based on confined RLPD testing.

APPENDIX C: LONG-TERM OVEN AGING STUDY

Methodology

A long-term aging study was carried out on some of the mixtures following the National Cooperative Highway Research Program (NCHRP) Project 09-54 procedures. This study's purpose was principally to characterize the aging susceptibility of VDOT mixtures for simulations with FlexPAVE (FHWA, n.d.). However, the results were also used to verify the NCHRP Project 09-54 pavement aging model (PAM) and Asphalt Mixture Aging-Cracking (AMAC) model for Virginia mixtures and calibrate Virginia-specific aging models for use in Mechanistic-Empirical analysis using FlexPAVE.

Eight mixtures were chosen for this study: B1, B2, C2, E1, E2, F2, G, and H. These mixtures included four 9.5 and four 12.5 NMAS mixtures, three PG 64-22 binders, three PG 58-28 binders, and two PG 64-22 + recycling agent binders. Samples were first aged in accordance with the NCHRP Project 09-54 by spreading the mixture onto thin pans and placing the pans into a 95°C oven for 1, 3, 6, and 11 days. The exceptions to this procedure were mixtures G and H, which were not evaluated for 6 days of aging. The long-term oven aged (LTOA) mixtures were then tested for dynamic modulus (DM) and cyclic fatigue at the various aging durations. In addition, dynamic shear modulus testing at 64°C and 10 rad/sec and temperature-frequency sweep were performed on binder that was extracted and recovered from these mixtures.

Mixture Long-Term Oven Aged Test Results

The selected mixtures were tested for DM at 4, 20, and 40°C, and at frequencies of 10, 1, and 0.1 Hz. An additional frequency of 0.01 Hz was tested at 40°C to better capture the lower asymptote of the DM mastercurve. The mixtures were then tested for cyclic fatigue at 18°C, staying consistent with the testing on the reheated mixtures. Reheated mixtures results are used to identify the effects of aging and are referred to as LTOA-0d in this section to distinguish them from short-term oven aged materials, which are lab-mixed rather than plant-mixed and reheated. Figure C1 and C2 show the mastercurves from the DM tests and the damage characteristic curves from the cyclic fatigue tests for the eight tested mixtures, respectively. Figure C1 shows that DM increased with aging level for all mixtures as expected because of the increased stiffness from aging. The damage characteristic is expected to shift upward as the material ages, which was found to occur consistently for aging levels of 1, 3, and 11 days (Figure C2). However, for mixtures B1 and C2, the LTOA-6d curve went below the LTOA-3d curve, suggesting a similar fatigue behavior at these two aging levels. Figures C3 and C4 show the evolution of cyclic fatigue index parameter (S_{app}) and D^R as a function of aging level. The figures show that these parameters decrease with aging for most mixtures. The only exception, although not statistically significant, is between LTOA-3d and LTOA-6d, which sometimes do not follow the decreasing trend, as well as the LTOA-6d and LTOA-11d for the D^R of mixture E2. The figures also suggest that starting with a higher S_{app} or D^R value at the LTOA-0d level leads to a more pronounced decrease in these parameters with aging. The figures also suggest that the most change is noticed from 1 and 11 days of aging, corresponding to the initial drop because of aging and the harshest aging level, respectively.

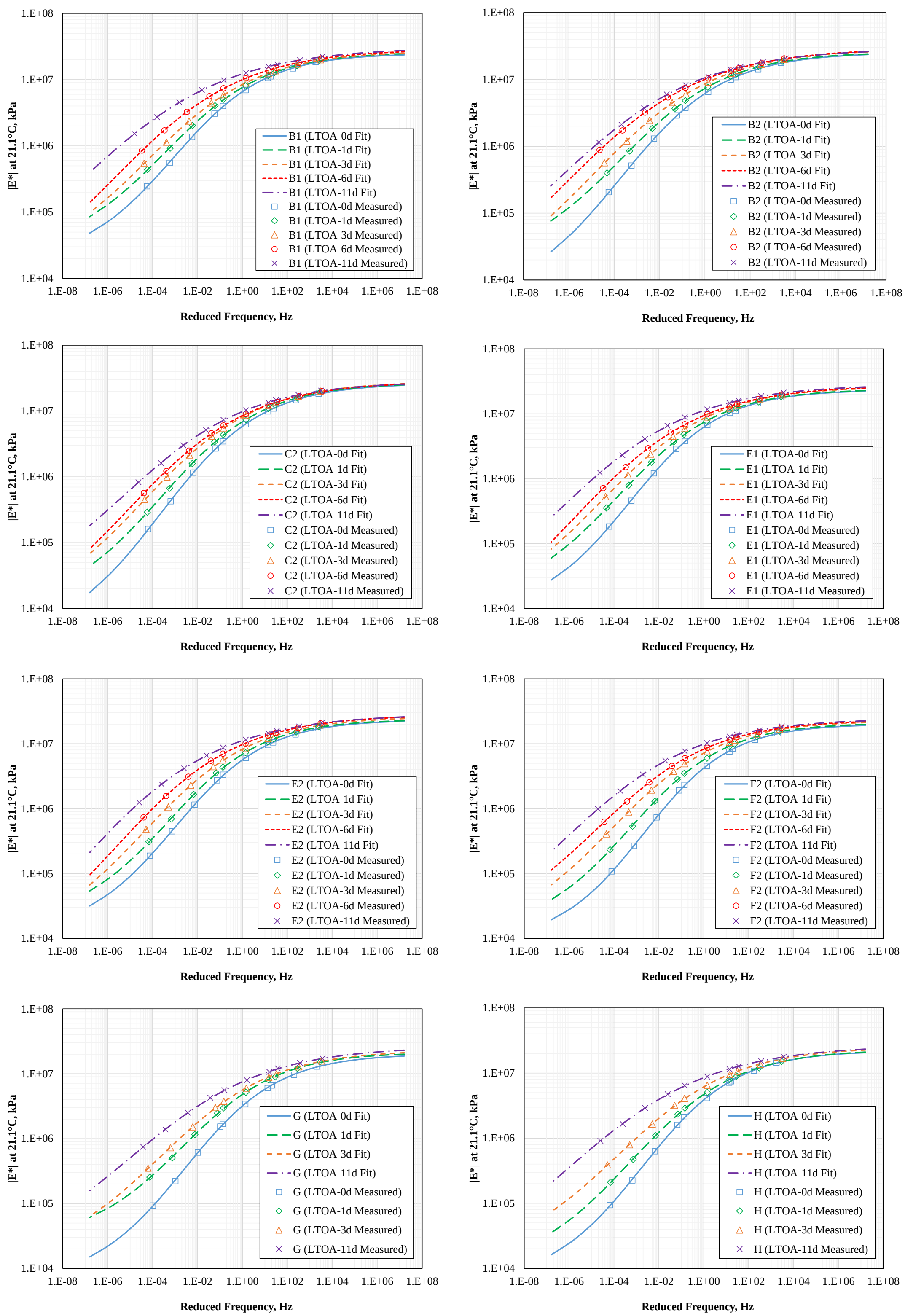


Figure C1. Dynamic Modulus Master Curves of Mixtures B1, B2, C2, E1, E2, F2, G, and H at Various Aging Levels. d = days; LTOA = long-term oven aging.

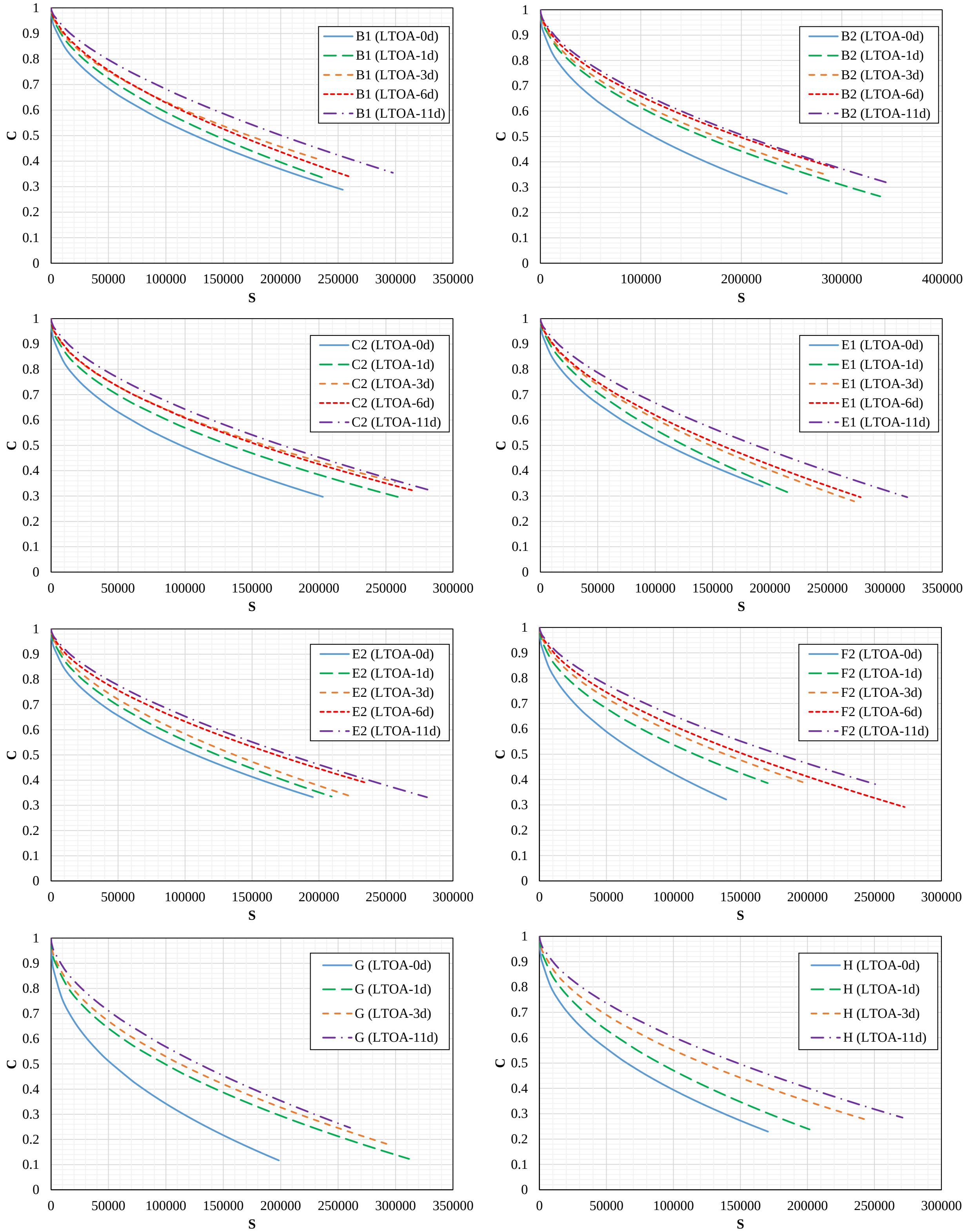


Figure C2. Damage Characteristic Curves for Mixtures B1, B2, C2, E1, E2, F2, G, and H at Various Aging Levels. d = days; LTOA = long-term oven aging.

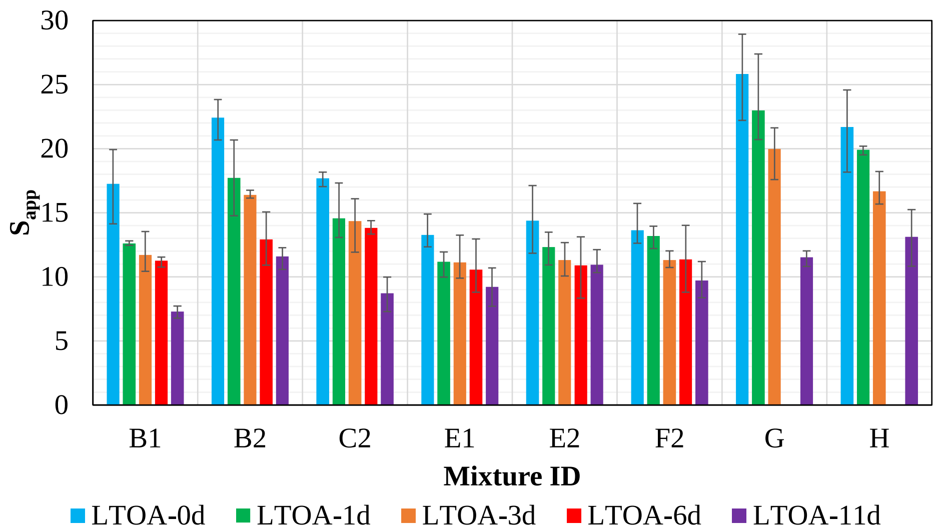


Figure C3. Effect of Aging Duration on S_{app} Values for Mixtures B1, B2, C2, E1, E2, F2, G, and H at Various Aging Levels. d = days; LTOA = long-term oven aging; S_{app} = cyclic fatigue index parameter.

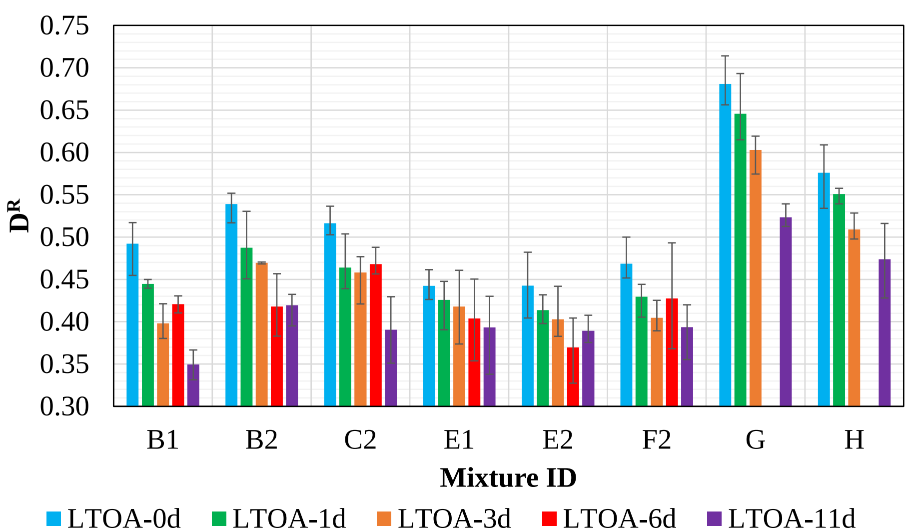


Figure C4. Effect of Aging Duration on D^R Values for Mixtures B1, B2, C2, E1, E2, F2, G, and H at Various Aging Levels. d = days; D^R = pseudo-energy-based failure criterion; LTOA = long-term oven aging.

Binder Long-Term Oven Aged Test Results

The binders from the eight mixtures at each age level were extracted and recovered, then conditioned in a vacuum chamber under a nitrogen blanket for 1-hour at 163°C to evaporate any remaining solvent while avoiding further aging the binder. The prepared binders were tested for $\log|G^*|$ at 64°C and 10 rad/sec, which is the chosen aging parameter in NCHRP Project 09-54. The results were used as Level 1 inputs to obtain the parameters for the PAM model ($\log|G^*|_0$ and aging susceptibility, M) used to predict $\log|G^*|$ at 64°C and 10 rad/s at any aging duration. The binders were also tested for temperature frequency sweep to obtain the binder dynamic shear modulus at the DM test temperatures to obtain the parameter c, which is the slope of time-temperature shift factor versus difference in $\log|G^*|$ used to calculate the time-aging shift factors for the AMAC model DM prediction. Figures C5a and C5b show the PAM model and the time-aging parameter fit for mixture F2, respectively. Figure C6 shows the binder dynamic shear modulus mastercurve for mixture F2 at the multiple aging levels. Table C1 shows the aging model parameters for all LTOA tested mixtures.

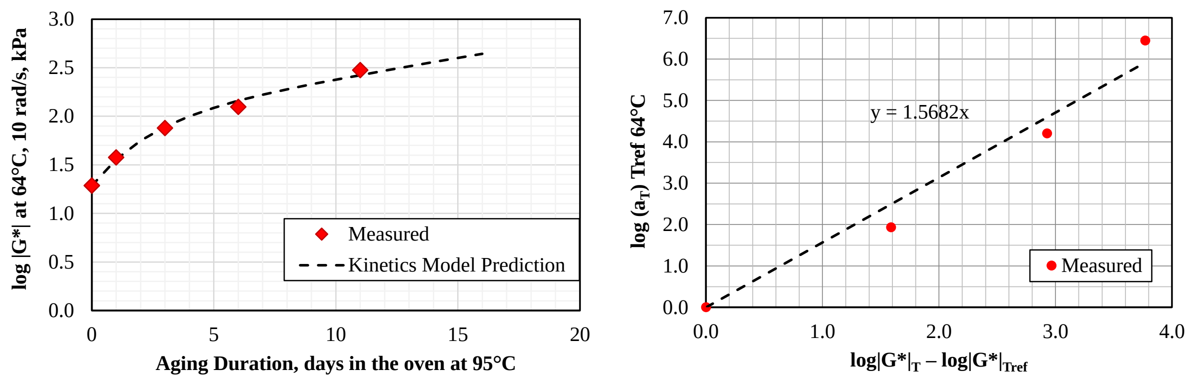


Figure C5. (a) Pavement Aging Model for Mixture F2; (b) Time-Aging Parameter Fit for Mixture F2

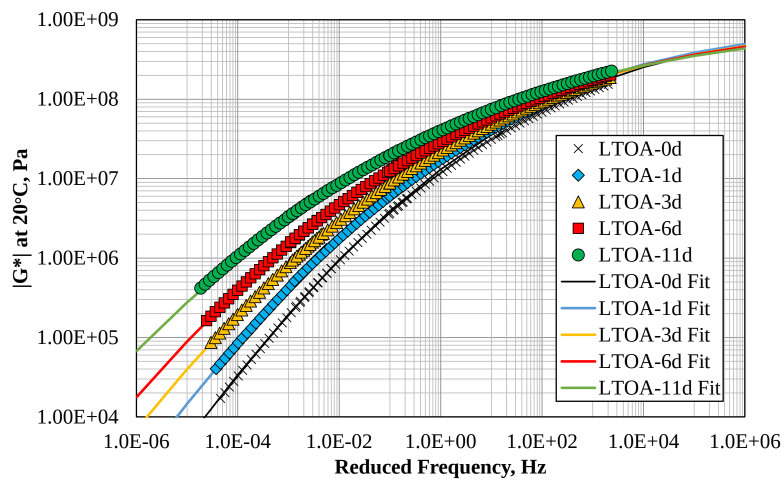


Figure C6. Dynamic Shear Mastercurve for Mixture F2. d = days; LTOA = long-term oven aging.

Table C1. Aging Parameters for LTOA Tested Mixtures

Mixture Identification	B1	B2	C2	E1	E2	F2	G	H
log|G^*|, LTOA-0d	1.413	1.427	1.250	1.356	1.304	1.285	1.134	1.282
M	0.688	0.665	0.687	0.751	0.777	0.781	0.691	0.733
c	1.84	1.79	1.67	1.80	1.60	1.57	1.50	1.78

d = days; LTOA = long-term oven aging.

Evaluating NCHRP 9-54 Level 1 Aging

The PAM model proved to be good for estimating the log|G^*| at 64°C and 10 rad/s at the different age levels for all eight mixtures. In terms of the AMAC model, the NCHRP Project 09-54 found that estimating *c* from mixture time-temperature superposition factors was sufficient for predicting mixture DM values at different aging levels. Yet, the applicability of this method needed to be evaluated for Virginia mixtures. Figure C7 compares the measured DM values and fitted DM mastecurves at the different aging levels for mixture F2 using AMAC M Level 1 aging model. Figure C7 shows that using the estimated *c* value of 1.568 underestimated the stiffness of mixture F2. Similar results were found for most of the tested mixtures; thus, an additional effort was undertaken to obtain the most accurate *c* parameter, referred to as "optimized *c*."

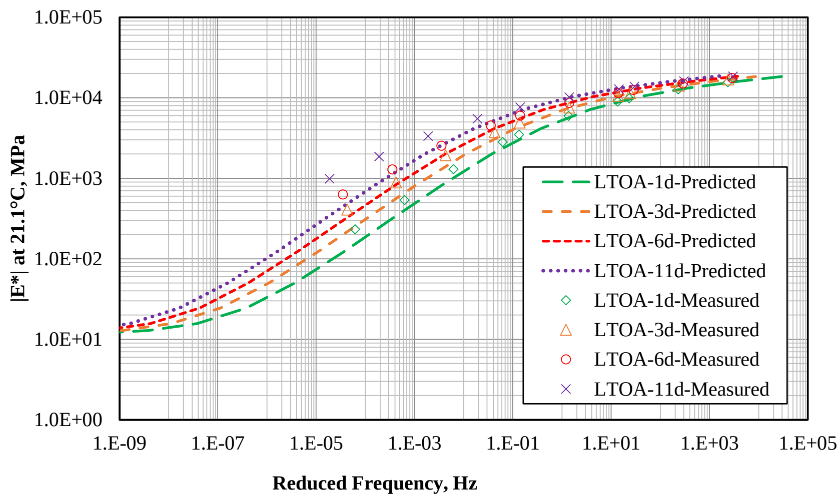


Figure C7. Application of Asphalt Mixture Aging-Cracking Dynamic Module Prediction for Mixture F2 Using Estimated *c*. d = days; LTOA = long-term oven aging.

The optimized *c* was obtained by fitting the actual measured time-aging shift factors from the test results as a function of the difference in log|G^*| at aging level *t* and LTOA-0d as shown in Figure C8a for mixture F2. The optimized *c* value of 2.205 was used to predict the DM mastercurve for mixture F2 to evaluate this method as shown in Figure C8b. Using the optimized *c* proved to significantly decrease the errors in predicting aged DM values, and thus the optimized *c* was determined for all eight mixtures. Table C2 shows the estimated and optimized *c* values for all mixtures. In NCHRP Project 09-54, Level 3 aging requires using a universal *c* value of 1.71 obtained by averaging all *c* values from the mixtures in the project. For the tested mixtures, using the NCHRP estimation method leads to a close average *c* of 1.69, whereas using the optimization method results in an average optimized value of 2.00. This comparison shows that estimating the *c* value using time-temperature shift factors led to the underestimation of the aging susceptibility for Virginia mixtures in terms of DM. Overall, the optimized *c* was considerably higher for most of the mixtures, except for mixture C2 and E1, where the estimated *c* was slightly higher. This information becomes important in the following section where NCHRP Level 3 is used to estimate the aging parameters for the remaining 10 mixtures that were only tested at reheated (LTOA-0d) conditions.

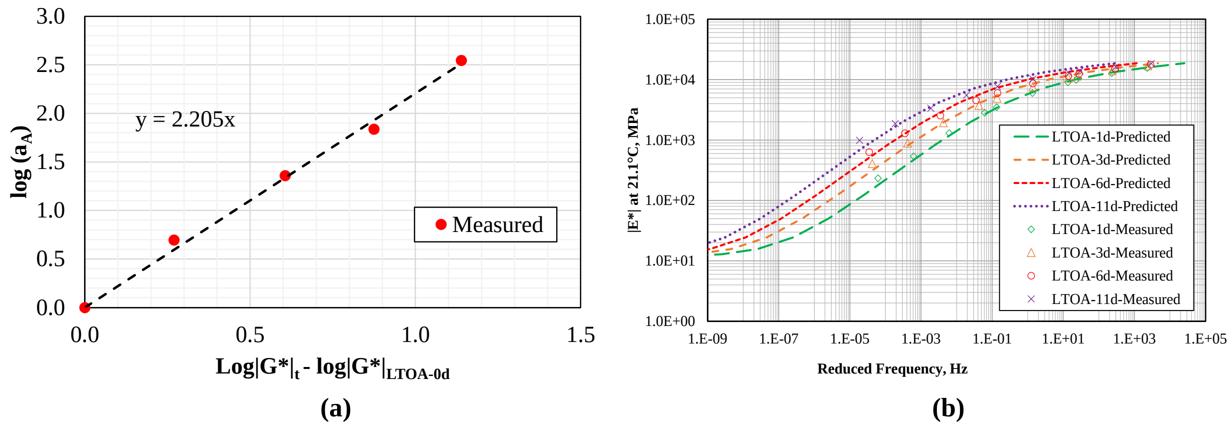


Figure C8. (a) Optimized Time-Aging Parameter Fit for Mixture F2; (b) Application of Asphalt Mixture Aging-Cracking Dynamic Modulus Prediction for Mixture F2 Using Optimized *c*. d = days; LTOA = long-term oven aging.

Table C2. Estimated and Optimized Time-Aging Shift Factors

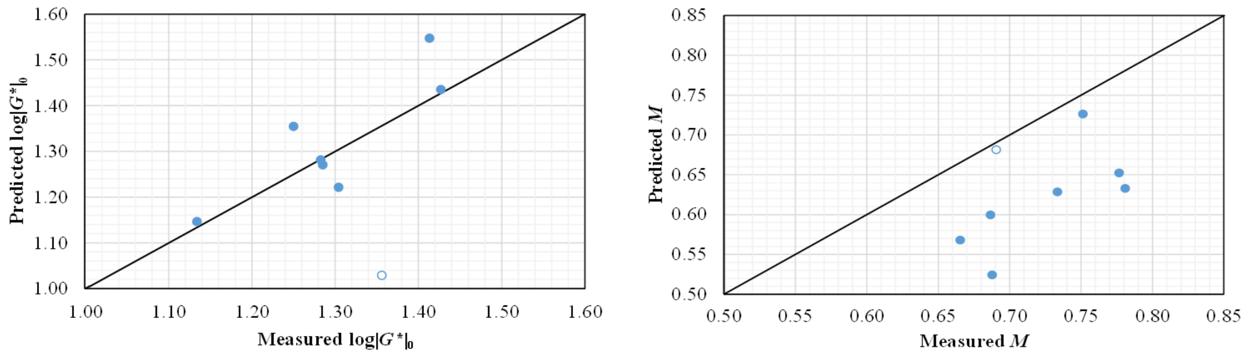
Mixture Identification	B1	B2	C2	E1	E2	F2	G	H	Average
Estimated *c*	1.84	1.79	1.67	1.80	1.60	1.57	1.50	1.78	1.69
Optimized *c*	2.03	2.07	1.64	1.75	1.85	2.21	2.36	2.16	2.00

Evaluating NCHRP 9-54 Level 3 Aging

NCHRP Project 09-54 Level 3 aging uses a universal c value of 1.71 for DM prediction with the AMAC model. Level 3 also uses equations to estimate the PAM parameters using the mixture's binder blended high performance grade (HPG) as Equation C1 shows for $\log|G^*|_0$ and Equation C2 for M. As mentioned previously, an average c value of 2.00 was found for the tested mixtures and thus was chosen for Level 3 aging analysis instead of 1.71. Equations C1 and C2 were used to predict the $\log|G^*|_0$ and M and compare the $\log|G^*|_0$ and M with the measured values as Figures C9a and C9b respectively show. As the figures show, the equations did not well estimate the PAM aging parameters. Thus these equations were adjusted to better fit the tested mixtures. The hollow points in Figures C9a and C9b were labeled as outliers and excluded from the calibration. Equations C3 and C4 represent the calibrated Level 3 equations for estimating PAM aging parameters for Virginia mixtures. Figures C10a and C10b show the measured and predicted $\log|G^*|_0$ and M using the calibrated Level 3 equations, respectively.

$$log|G^*|_0 = \frac{6.552}{1+\left(\frac{HPG}{107.167}\right)^{-4.068}} \quad \text{(Eq. C1)}$$

$$M = \frac{0.94}{1+\left(\frac{HPG}{82.68}\right)^{7.96}} \quad \text{(Eq. C2)}$$

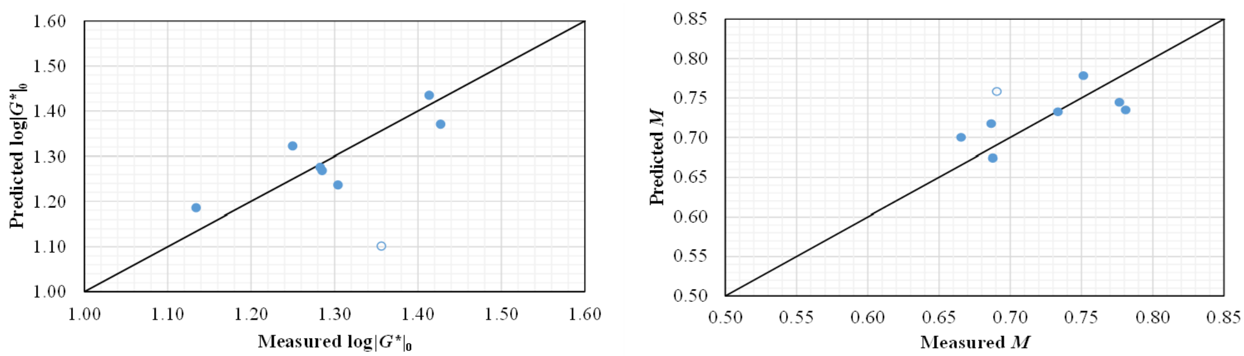


(a) **(b)**

Figure C9. (a) Measured versus Predicted log|*G|₀ Using NCHRP 9-54 Level 3 Equation; (b) Measured versus Predicted *M* Value Using NCHRP 9-54 Level 3 Equation. NCHRP = National Cooperative Highway Research Program.**

$$log|G^*|_0 = \frac{2.657}{1+\left(\frac{HPG}{77.186}\right)^{-4.068}} \tag{Eq. C3}$$

$$M = \frac{0.856}{1+\left(\frac{HPG}{94.667}\right)^{7.96}} \tag{Eq. C4}$$



(a) **(b)**

Figure C10. (a) Measured versus Predicted log|*G|₀ Using the Calibrated Level 3 Equation; (b) Measured versus Predicted *M* Value Using the Calibrated Level 3 Equation**

Developing Virginia-Specific Aging Levels

After finding some discrepancies between measured aging parameters and the NCHRP Project 09-54 method for calculating *c* and Level 3 equations, Virginia-specific hierarchical aging levels were proposed to ensure accurate aging analysis. Four potential pathways were identified as summarized in Table C3. The most advanced level, Level 1, is consistent with the NCHRP Project 09-54 model, except the c value is specified to be 2.00. This same *c* value is recommended for all pathways selected because the method proposed in NCHRP Project 09-54 was not found to be applicable for the tested mixtures and aging and testing mixtures at multiple aging levels to obtain the optimized *c* is not practical. Thus the difference in the three remaining pathways depends on how the PAM model parameters are obtained. These three pathways are all a subset of the NCHRP Project 09-54 Level 3 approach that relies on empirical equations to find the $\log|G^*|_0$ and M coefficients. These pathways are labeled as Level 3a, 3b, or 3c depending on the accuracy. In Level 3a, the binder in a mixture should be extracted and recovered from the mixture and the binder's HPG measured. Level 3b also uses the calibrated Level 3 equations, but with the HPG estimated using the blending theory knowing the virgin binder HPG (HPG_{Virgin}), the reclaimed asphalt pavement HPG (HPG_{RAP}), and the asphalt binder replacement. Level 3c includes assuming the inputs needed in Level 3b to predict the blended HPG in cases where some or no binder data are available. Table C3 summarizes the four levels.

Table C3. Proposed Hierarchical Aging Levels

Pathway	AMAC Level	Inputs Needed	Aging Parameters	
			$\log\|G^*\|_0$ and M	**c**
1	Level 1	$\|G^*\|$ at 64°C, 10 rad/s from E&R binder at multiple age levels	NCHRP Level 1 - Kinetics Model	2
2	Level 3a	HPG_{Blend} from E&R binder at LTOA-0d	Calibrated NCHRP Level 3 Equations	2
3	Level 3b	HPG_{Virgin}, HPG_{RAP}, and calculated ABR: HPG_{Blend} is predicted using the blending theory	Calibrated NCHRP Level 3 Equations	2
4	Level 3c	Assume the values needed in L4 to predict HPG_{Blend}	Calibrated NCHRP Level 3 Equations	2

ABR = asphalt binder replacement; AMAC = Asphalt Mixture Aging-Cracking model; HPG = high performance grade; LTOA = long-term oven aged; NCHRP = National Cooperative Highway Research Program; RAP = reclaimed asphalt pavement.

To maintain consistency, all mixtures aging parameters were obtained using the proposed Level 3a, including the tested mixtures. This step is needed for the mechanistic analysis using FlexPAVE (FHWA, n.d.). Figure C11 shows the PAM model parameters estimated with Level 3 for all mixtures, with the solid bars and data points representing the eight aged and tested mixtures.

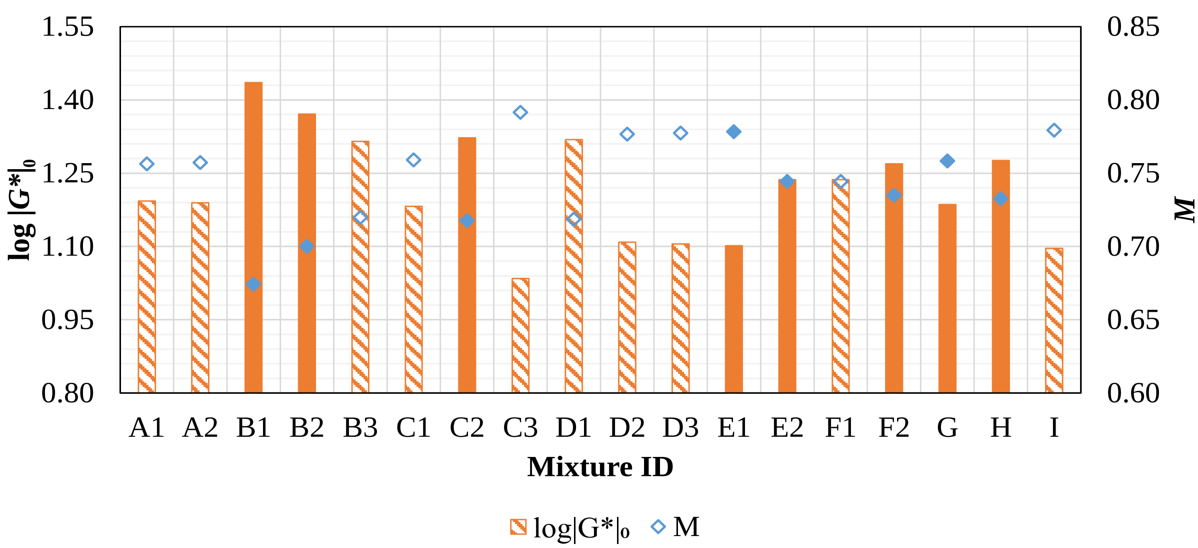


Figure C11. Level 3 Aging Pavement Aging Model Parameters for All Mixtures