Assessment of Production Variability and Composite Performance Index for Conventional and High RAP BMD Mixtures

Bilin Tong, Corresponding Author
Graduate Research Assistant
Virginia Tech Transportation Institute, Blacksburg, VA, 24060
Email: bilin@vt.edu, ORCID: 0000-0002-2831-8008

Jhony Habbouche, Ph.D., P.E.
Western US Regional Engineer
Asphalt Institute, Gilbert, AZ, 85295
Email: jhabbouche@asphaltinstitute.org, ORCID: 0000-0002-6216-3134

Stacey D. Diefenderfer, Ph.D., P.E.
Associate Principal Research Scientist
Virginia Transportation Research Council, Charlottesville, VA, 22903
Email: stacey.diefenderfer@vdot.virginia.gov, ORCID: 0000-0002-6570-2121

Gerardo W. Flintsch, Ph.D., P.E.
Dan Pletta Professor
Virginia Tech Transportation Institute, Blacksburg, VA, 24060
Email: gflintsch@vtti.vt.edu, ORCID: 0000-0003-4440-0711

Ilker Boz, Ph.D.
Research Scientist
Virginia Transportation Research Council, Charlottesville, VA, 22903
Email: Ilker.Boz@VDOT.Virginia.gov

ABSTRACT

The Balanced Mix Design (BMD) concept enables the design of engineered mixtures containing conventional and high reclaimed asphalt pavement (HRAP) contents, moving beyond the constraints of traditional volumetric design methodologies. During production, the designed mixture undergoes verification and potential modifications at the plant to accommodate actual production and field circumstances, irrespective of the mix design method. This study assessed the impact of production and associated performance variability on a volumetrically designed control mixture and five mixtures designed with BMD concept. This investigation showed relatively precise gradation control, but exceedances of volumetric property tolerances were observed in BMD-optimized mixtures during production. Performance, including durability, cracking, and rutting susceptibility, was evaluated using the Cantabro test, indirect tensile cracking test (IDT-CT) and Asphalt Pavement Analyzer (APA) test, respectively. Test results uncovered that produced mixtures may become unbalanced. Observations from Cantabro test and IDT-CT highlighted the necessity and effectiveness of employing BMD for HRAP mixtures. The potential aging effect introduced during the reheating process may compromise durability and cracking resistance. In addition, a 3-Dimensional plot with a revised Composite Performance Index (CPI_R) was used to optimize the process of evaluating the mixture "balance" status among multiple primary performances. It revealed almost all produced HRAP mixtures demonstrated a well-balanced status.. Finally, agencies can use the CPI_R as part of their acceptance program for BMD mixtures to determine a pay factor for possible bonuses or penalties.

Keywords: RAP, BMD, production variability, durability, cracking, rutting

1. INTRODUCTION

It is widely recognized that contemporary asphalt materials are considerably different from their predecessors due to the escalating challenges of supply and the amplified demand for high-performance materials *(1)*. The abundant use of additives, extensive application of reclaimed asphalt pavement (RAP), and the dynamic nature of the modern asphalt supply necessitate a fundamental shift in mixture design philosophy. Instead of focusing primarily on volumetric properties, the emphasis in design is pivoting toward properties that indicate the performance of mixtures *(1)*.

The emergence of the balanced mix design (BMD) methodology is a direct response to this necessary paradigm shift in asphalt mixture design. Defined by the former Federal Highway Administration Expert Task Group on mixtures and construction, BMD is an "*Asphalt mix design using performance tests on appropriately conditioned specimens that address multiple modes of distress taking into consideration, mix aging, traffic, climate, and location within the pavement structure*" *(2, 3)*. In contrast to traditional volumetric design, BMD replaces certain aspects with performance testing criteria aimed at assessing mixture susceptibility to common distresses such as rutting and cracking. The methodology requires that a mixture design must meet specified performance criteria for approval, with additional performance testing possibly required during the production phase for mixture acceptance. While BMD mixtures cannot rectify unsound underlying pavement structures or inappropriate maintenance treatments, the introduction of BMD nonetheless represents a significant stride towards enhancing asphalt mixture performance.

In the United States (U.S.), the recognized environmental and economic benefits of using RAP in asphalt mixtures have prompted state agencies to implement special provisions allowing for its increased use in asphalt mixtures *(4, 5)*. The strategic incorporation of RAP not only aims to offset the constantly inflating costs of asphalt binders and fuel for transport and placement but also reduces carbon dioxide emissions *(4-6)*. However, drawbacks to their use exist. High RAP (HRAP) asphalt mixtures risk becoming overly stiff due to the increased RAP usage, which in turn makes them more brittle and susceptible to cracking *(5, 7)*. Designing HRAP mixtures presents challenges, mainly in meeting volumetric mix design criteria. The substantial fine aggregate content in RAP, a result of crushing and milling the material, can lead to aggregate degradation, complicating the achievement of optimal volumetric properties (*1, 8*). Moreover, the high RAP content can trigger a range of construction and performance issues, including compaction difficulty in cool weather, potential for thermal cracking, fatigue, and reflective cracking from repeated loading and daily or seasonal thermal stresses, and raveling due to aging or moisture damage *(5)*. Therefore, one practical application of the BMD concept lies in the strategic design and production of HRAP asphalt mixtures. The BMD concept, in conjunction with the use of recycling agents (RAs) and/or softer binders, offers a solution to the complexities inherent in the design and production of HRAP mixtures, providing an innovative and practical response to the demands of contemporary pavement construction.

Regardless of the mix design method employed, the fundamental purpose is to ascertain an optimal combination of aggregates and asphalt binder that provides acceptable mixture performance. This derived blend is referred to as the job mix formula (JMF). During the production process, the JMF is subjected to verification and potential modifications at the plant level to account for actual production and field circumstances *(9)*. This becomes particularly challenging when using the BMD concept, as it deviates from traditional specification

restrictions on volumetric properties. Factors such as equipment calibration, personnel training, and the use of RAP from various sources could potentially complicate the production process for contractors. Furthermore, replicating the large-scale mixture production executed in the plant under laboratory design protocols can be an intricate endeavor *(9)*. Differences in asphalt mixture properties from the original design, which can occur during production, frequently instigate premature manifestations of pavement distress or can potentially lead to outright structural failure *(10, 11)*. Consequently, the implementation of quality control (QC) measures, coupled with quality acceptance (QA) testing, becomes important to ascertain that the produced mixture meets the design specifications. These scenarios further highlight the necessity of validating and documenting the impact of production and associated variability on the performance of asphalt mixtures designed with BMD.

To identify and assess the sources, causes, and degrees of variability in gradation, volumetric, and mechanical properties, the evaluation of various asphalt mixture specimen fabrication scenarios is imperative. For instance, National Cooperative Highway Research Program (NCHRP) Project 09-48 (Report 818) examined three core scenarios: (1) laboratory-mixed, laboratory-compacted specimens produced during the design process; (2) plant-mixed, laboratory-compacted specimens used in volumetric acceptance testing of plant-produced mix; and (3) plant-mixed, field-compacted specimens employed during density acceptance testing and forensic evaluation of in-place pavement. The goal of these assessments was to ascertain the sources of variability and to evaluate the precision and bias related to the volumetric and mechanical properties of dense-graded asphalt mixtures *(11)*. In addition to these scenarios, other alternatives include non-reheated and reheated plant-mixed, laboratory-compacted specimens, scenarios found in the Virginia Department of Transportation's (VDOT's) practice for contractors' quality control (*12*).

Numerous state highway agencies are actively exploring the design and acceptance of asphalt mixtures based on performance using the BMD concept *(2, 3)*. Virginia is among these agencies and has been endeavoring to implement BMD for a substantial period. In 2007, VDOT instituted specifications allowing up to 30% RAP in several specific surface mixtures (SMs). By 2013, VDOT was examining the feasibility of SMs containing up to 45% RAP *(13)*, leading to field trials of mixtures containing 40-45% RAP contents *(14)*. In 2017, VDOT began to consider the inclusion of performance requirements for mix designs using the BMD concept. In 2018, preliminary efforts offered performance benchmarks for dense-graded unmodified surface mixtures, in support of potential pilot projects with HRAP mixtures. Three fast, simple, and practical performance-indicative tests addressing varying distress modes were selected for use with the BMD method. These were the Cantabro test, the Indirect Tensile Cracking Test (IDT-CT) at intermediate temperature, and the Asphalt Pavement Analyzer (APA) test, assessing durability, cracking, and rutting, respectively. Initial performance thresholds were set *(15)* and subsequently the developed performance-based specifications were validated *(12, 16)*.

Two VDOT BMD special provisions, one for conventional SMs and one for HRAP SMs were drafted and revised for use in pilot projects (*12*). In 2019, two field trials were conducted in Virginia, marking the first field applications of the BMD specifications. These trials entailed the design, production, and application of nine asphalt mixtures which incorporated varying combinations of RAP contents, two binder grades, RAs, and two warm mix asphalt (WMA) additives. Three performance-indicative tests were conducted on laboratory-produced design specimens, as well as non-reheated and reheated plant-produced, laboratory-compacted specimens. The variability of these test results was documented in detail *(13)*. Subsequent field

trials were planned and implemented in 2020. These trials featured asphalt mixtures with high RAP contents, softer binders, RAs, and a variety of additional additives such as fibers and softening oils *(12)*. Based on the data derived from the 2019 and 2020 field trials, there is a need to evaluate a broader spectrum of mixtures with varied performance responses. This would facilitate the collection of a robust dataset, necessary for either confirming or debunking the observed lack of correlation and significant difference in results for the Cantabro, IDT-CT, and APA tests.

In 2020, a joint research initiative was undertaken by the Virginia Transportation Research Council (VTRC), VDOT, and Virginia Tech, involving the design, placement, and accelerated testing of a control mixture and five BMD mixtures at the Virginia Accelerated Pavement Testing (APT) facility, owned by VDOT and located at the Virginia Tech Transportation Institute (VTTI). This effort served, in part, as a highly controlled application of the BMD special provisions to production, offering an opportunity to scrutinize the implications of these specifications on various aspects such as the design, production, quality control and assurance practices, and construction of HRAP SMs *(17, 18)*.

2. OBJECTIVE AND SCOPE

The primary objective of this paper is to assess the impact of production and associated variability on a control mixture and five engineered mixtures designed using the BMD concept. This responds to the need for the recommendation of VDOT field trials studies to evaluate a broader spectrum of mixtures with varied performance responses. The assessment provided herein will offer comprehensive laboratory performance test references for quality measurement practices within BMD implementation. In addition, this study lays the foundation for future work aimed at bridging the gap between laboratory testing and APT field performance measurement. To accomplish this, in addition to testing performance during the design process, a comprehensive laboratory assessment focused on durability, cracking resistance, and rutting performance was undertaken on the mixtures collected during production and paving of the APT facility. Statistical analyses were carried out to discern performance differences among different specimens including laboratory-mixed laboratory-compacted specimens (designs), plant-mixed laboratory-compacted specimens (non-reheats), and plant-mixed laboratory-compacted specimens fabricated after a reheating process (reheats). A total of 648 specimens were evaluated in this study. In essence, the paper is structured to methodically address four fundamental inquiries:

1. Comparing and evaluating the three primary BMD test results including Cantabro test, IDT-CT, and APA rutting test across six mixtures, and demonstrating the variations in different performances.
2. Identifying statistical significance between different specimens (designs vs. non-reheats vs. reheats) and among different samples collected during production for each BMD tests.
3. Assessing the influence of the reheating process on the three primary performance indices of the plant-produced mixtures.
4. Exploring a more effective method for simultaneously evaluating multiple BMD primary performance metrics.

3. METHODOLOGY

3.1. Experimental Program

A total of six mixtures encompassing one control and five BMD 9.5mm dense-graded SMs were designed, produced, and sampled for evaluation in this study. The experimental protocol included conducting volumetric and BMD laboratory testing on designs, non-reheats and reheats, followed by comprehensive analyses to evaluate the performance characteristics of these mixtures, as illustrated in Figure 1.

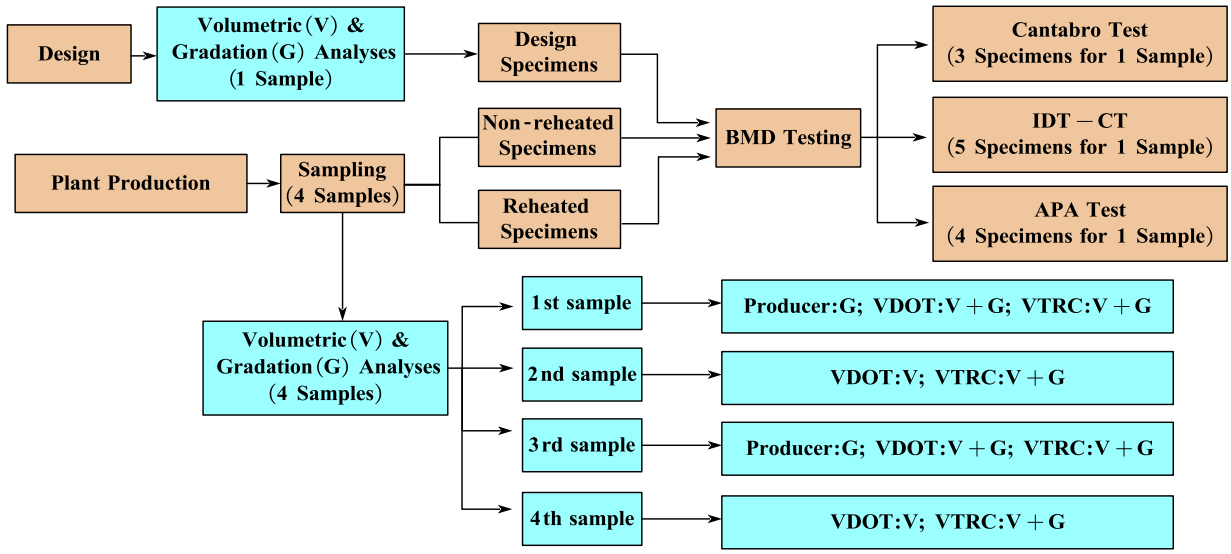

FIGURE 1. Experimental Plan for Laboratory- and Plant-Produced Mixtures. BMD = balanced mix design; IDT-CT = indirect tensile cracking test; APA = asphalt pavement analyzer; G = gradation analysis; V = volumetric analysis; VDOT = Virginia Department of Transportation; VTRC = Virginia Transportation Research Council.

3.2. Mixture Designs

The mixtures incorporated various combinations of RAP contents, two binder performance grades (PGs), one RA, and one WMA additive. Design specimens were laboratory-produced laboratory-compacted specimens, fabricated for mixture design. The RAP content in the mixtures was determined based on the total weight of the mixtures. Considering the scale of the plant production for Virginia APT project, the materials were well-controlled and the RAP use for all mixtures was the same. The RA, a product derived from tall oils and fatty acids, was added to the mixtures in addition to the virgin and RAP binders so that there was no partial replacement of either the virgin or RAP binders with the RA. A 100% blending ratio was assumed, and the design asphalt contents accounted for the contribution of asphalt binder from RAP materials. The WMA additive used is a bio-based product that is also used as an antistrip agent. The mixtures are defined as follows:

- 30_C: a non-BMD mixture serving as a control with 30% RAP content and PG 64S-22 binder (S denotes standard traffic). This mixture was designed following VDOT's Superpave mix design methodology with the optimum binder content (OBC) selected for voids in total mixture (VTM) of 4%.
- 30_O: a BMD optimized version of mixture 30_C featuring the use of 30% RAP content and a PG 64S-22 at a relatively higher OBC when compared to mixture 30_C. The

optimization process involved adjusting the aggregate gradation and increasing the asphalt binder content.

- 45_HR: a BMD optimized HRAP mixture with 45% RAP and PG 64S-22 binder. This mixture design resulted in a much higher OBC when compared with all other evaluated mixtures. No softer binder or RA were used during the design and production of this mixture.
- 45_HR_RA: a BMD optimized HRAP mixture with 45% RAP, PG 64S-22 binder, and an RA.
- 45_HR_L: a BMD optimized HRAP mixture with 45% RAP and a softer binder, PG 58-28.
- 60_HR_L_RA: a BMD HRAP mixture with 60% RAP, a softer binder (PG 58-28), and an RA.

The BMD optimized mixtures were designed following VDOT BMD specifications using Approach D (Performance Only) *(2, 3)*. The design volumetric properties and gradations of all mixtures are shown in Table 1. The BMD performance properties of the six mixtures at OBC determined during design are also shown in Table 1. During the design stage, the short-term oven aging (STOA) protocol consisted of 2 hours at compaction temperature for Cantabro and APA tests, and 4 hours at compaction temperature for the IDT-CT.

TABLE 1. Volumetric Properties and Gradations for APT Mixtures - Design

Mixture ID	**30_C**	**30_O**	**45_HR**	**45_HR_RA**	**45_HR_L**	**60_HR_L_RA**
Description	Non-BMD	BMD	BMD HRAP	BMD HRAP	BMD HRAP	BMD HRAP
Composition						
RAP Content, %	30	30	45	45	45	60
Asphalt Binder	PG 64S-22	PG 64S-22	PG 64S-22	PG 64S-22	PG 58-28	PG 58-28
Additives	WMA	WMA	WMA	WMA + RA	WMA	WMA + RA
Property						
N_{design}, gyrations	50	50	50	50	50	50
NMAS, mm	9.5	9.5	9.5	9.5	9.5	9.5
Asphalt Content, %	5.6	6.0	6.8	6.2	6.0	6.0
RBR	0.24	0.22	0.29	0.32	0.33	0.44
Rice SG (G_{mm})	2.531	2.517	2.497	2.519	2.521	2.538
VTM, %	4.0	2.1	5.2	2.7	4.2	1.5
VMA, %	16.8	15.9	20.2	16.9	17.9	15.3
VFA, %	76.2	87.0	74.3	84.0	76.6	90.0
FA Ratio	1.0	1.1	1.2	1.2	1.3	1.5
Mixture Bulk SG (G_{mb})	2.429	2.465	2.367	2.452	2.415	2.500
Performance Properties at Optimum Asphalt Binder Content						
Cantabro Mass Loss, 25ºC, %	2.9	3.2	4.1	2.8	2.7	4.0
APA Rut Depth, 64ºC, mm	5.4	4.2	5.6	7.2	4.9	3.7
CT index, 25ºC	58	112	299	96	141	128
Gradation / Sieve Size	**% Passing**					
¾ in (19.0 mm)	100.0	100.0	100.0	100.0	100.0	100.0
½ in (12.5 mm)	100.0	100.0	100.0	100.0	100.0	100.0
3/8 in (9.5 mm)	94.0	96.0	93.0	93.0	92.0	92.0
No. 4 (4.75 mm)	66.0	66.0	63.0	63.0	63.0	62.0
No. 8 (2.36 mm)	39.0	41.0	38.0	38.0	40.0	39.0
No. 16 (1.18 mm)	--	--	--	--	--	--
No. 30 (600 µm)	23.0	17.0	18.0	18.0	19.0	20.0

No. 50 (300 µm)	--	--	--	--	--	--
No. 100 (150 µm)	--	--	--	--	--	--
No. 200 (75 µm)	5.8	6.6	7.4	7.4	7.5	8.3

S = standard traffic; BMD = balanced mix design; HRAP = high reclaimed asphalt pavement content mixture; RAP = reclaimed asphalt pavement; WMA = warm mix additive; RA = recycling agent; NMAS = nominal maximum aggregate size; RBR = recycled binder ratio, which is the ratio of RAP binder content to the mixture total binder content; SG = specific gravity; VTM = voids in total mixture; VMA = voids in mineral aggregate; VFA = voids filled with asphalt; FA = fines to aggregate; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

3.3. Production Sampling

Sampling of the loose mixture was conducted four times throughout each day of production. The production sampling plan for each mixture is outlined in Figure 1. Samples were sequentially labelled from 1st to 4th for each mixture and samples of test specimens were correspondingly labelled based on the sample from which they were fabricated (for instance Cantabro specimen sample A was fabricated from sample A for the mixture under evaluation). The producer generally performed a gradation analysis on samples A and C. The VDOT district laboratory determined the volumetric properties for all four samples, in addition to determining the gradation for samples A and C. The VTRC laboratory determined volumetric properties and gradation for all four samples of each mixture. The three entities collected the samples at the same time from the same production lot.

The loose mixture samples collected during production for BMD testing underwent one of two procedures: they were either transported to the producer's laboratory and immediately compacted into test specimens—these were referred to as non-reheats - or they were placed in boxes, delivered to VTRC, and stored in a climate-controlled area for subsequent evaluation after reheating. Reheats were prepared by VTRC by reheating the loose mixture stored in boxes until the material was workable. After this, the material was divided into quantities sufficient for specimen fabrication. The material was then heated to the appropriate compaction temperature and compacted *(13)*. Mixtures were not subjected to any additional oven aging. A total of 648 specimens were fabricated for the three BMD tests considered in this study.

3.4. BMD Tests

BMD testing for this study included three tests adopted in Virginia: Cantabro test, IDT-CT, and APA test, which assess durability, cracking, and rutting, respectively *(12, 15).* The corresponding thresholds/criteria for these tests are shown in Table 2. Cantabro test has been recognized as a promising tool to assess the durability of dense-graded asphalt mixtures, as supported by several previous studies and efforts *(19)*. The Cantabro mass loss (CML) was determined in accordance with AASHTO T 401 on the three gyratory specimens used to determine volumetric properties. The IDT-CT was conducted at a temperature of 25°C in accordance with ASTM D8225-19 *(20)* on five replicate test specimens compacted at 7±0.5% air voids. The APA test was performed in accordance with AASHTO T 340 at a temperature of 64ºC; four replicate test specimens were tested in an APA Junior machine. The APA specimens were compacted at 7±0.5% air voids. In addition, VDOT requires that the moisture susceptibility of asphalt mixtures be evaluated by means of AASHTO T 283. However, this evaluation is only required for the first production lot (*21*) and is not discussed further in this study. The contractor performed the moisture damage test, and all mixtures met the TSR requirement of 0.80.

TABLE 2. Virginia Performance Testing Criteria

Test	Test Method	Specimens	Criteria
Cantabro test	AASHTO T401	3 replicates	Mass loss $\leqslant$ 7.5%
APA test	AASHTO T 340	4 replicates	Rutting $\leqslant$ 8.0 mm
IDT-CT test	ASTM D 8225	5 replicates	CT index $\geqslant$ 70
Moisture damage test	AASHTO T 283	6 replicates	TSR $\geqslant$ 80%

APA = asphalt pavement analyzer; IDT-CT = indirect tensile cracking test; TSR = tensile strength ratio. The corresponding data collected by the producer for TSR was not discussed in this manuscript.

3.4. Statistical Analyses

To determine whether there are statistically significant differences between various specimens (designs, non-reheats, and reheats) and among the four produced samples for different mixtures, statistical analysis using various available tests was conducted using SPSS Statistics (*22*). The Shapiro-Wilk (SW) test was used to evaluate the normality of the performance indices for each mixture. The null hypothesis posited by the SW test was that the data conformed to a normal distribution. Should the p-value fall below 0.05, the null hypothesis would be rejected, leading to the conclusion that the data does not follow a normal distribution. Subsequently, the Levene's Test was employed to assess the equality of variances. The analysis of variance (ANOVA) along with the Tukey's pairwise comparison test was performed at a significance level of 5% (*22*). The ANOVA results are presented with the Tukey's pairwise comparison pertaining to the comparison of different treatments (designs vs. non-reheats vs. reheats) and different samples during production. However, for those data not following the assumptions of normality or homogeneity of variance, a nonparametric test (at a significance level of 5%), the Kruskal-Wallis (KW) test was employed, as an alternative to the more traditional pair-wise comparison test (*23*).

4. RESULTS

4.1. Mixture Volumetric Properties and Gradations

Producer and VDOT district data for volumetric properties and gradations were available from VDOT's Materials Information Tracking System. Volumetric and gradation analyses for corresponding reheated samples were performed by VTRC. The data collected included gradation and asphalt binder content; theoretical maximum and bulk specific gravities (G_{mm} and G_{mb}); VTM; voids in mineral aggregate (VMA); voids filled with asphalt (VFA); bulk and effective aggregate specific gravities (G_{sb} and G_{se}); fine aggregate to asphalt ratio (FA); percent binder absorbed (P_{ba}); effective binder content (P_{be}); and effective asphalt film thickness (F_{be}). The allowable production tolerances for volumetric properties and gradation sieves can be found in the VDOT specifications (*24*).

Table 3 presents the results of production volumetric analyses performed by VDOT (for non-reheats) and VTRC (for reheats). Specimens prepared for the volumetric analyses were subsequently used for Cantabro testing. The production tolerance values were used for acceptance of the control mixture and used only as a relative reference for BMD mixtures, as only mixture 30_C was designed in accordance with volumetric requirements. The other five mixtures were designed using BMD approach D (performance only), and volumetric requirements were waived, except for the production tolerance requirement on asphalt content and gradation (*25*). The asphalt binder content in all produced mixtures fell within acceptable tolerance, with the exception of some samples from mixture 45_HR that exceeded this tolerance.

Notably, mixture 45_HR exhibited the highest binder content among all evaluated mixtures. This high binder content, combined with extremely lower air void observed during production, contributed to the "wet" nature of mixture 45_HR.

Mixture 30_C exhibited the fewest volumetric properties that exceeded tolerance ranges. For produced mixture 30_O specimens, higher than acceptable VTM were observed, along with lower than acceptable VFA. Regarding HRAP mixtures, deviations outside of typical allowable tolerance ranges were primarily evident in VTM, VFA, and FA ratio values. This deviation stemmed from the mixture design process, which resulted in higher asphalt contents. This, combined with the aggregate structure, led to mixtures with lower VTM and higher VFA and FA ratios compared to those seen under VDOT's current volumetric design process. These changes in the mixture structure, influenced by the high RAP content, introduced more fines to the mixture. Variability in volumetric properties was more pronounced in mixture 60_HR_L_RA featuring the highest RAP content.

TABLE 3. Summary of Primary Volumetric Properties Determined on Mixtures During Production

Property	JMF	Production tolerance	Sample A		Sample B		Sample C		Sample D	
			VDOT	VTRC	VDOT	VTRC	VDOT	VTRC	VDOT	VTRC
Mixture 30_C										
AC, %	5.6	±0.3	5.57	5.47	5.39	5.44	5.69	5.76	5.60	5.39
G_{mm}	2.531	-	2.536	2.545	2.545	2.604	2.543	2.538	2.542	2.546
VTM, %	4.0	2.0-5.0	4.4	4.6	4.0	6.0	4.2	4.0	3.8	4.4
VMA, %	16.8	Min. 16	17.1	17.1	16.4	18.6	17.2	17.2	16.7	16.7
VFA, %	76.2	70-85	74.2	72.9	75.5	67.5	75.8	77.0	77.1	73.6
FA Ratio	1	0.7-1.3	1.1	1.1	1.2	1.1	1.1	1.0	1.1	1.1
Mixture 30_O										
AC, %	6.0	±0.3	6.12	6.12	6.18	6.14	6.10	6.09	6.12	5.97
G_{mm}	2.517	-	2.520	2.524	2.523	2.515	2.521	2.529	2.523	2.523
VTM, %	2.1	2.0-5.0	5.8	5.4	6.1	4.6	6.5	5.9	6.1	5.4
VMA, %	15.9	Min. 16	19.5	19.1	19.9	18.4	20.0	19.5	19.9	18.8
VFA, %	87.0	70-85	70.0	71.7	69.1	75.2	67.5	69.8	69.0	71.1
FA Ratio	1.1	0.7-1.3	1.0	1.1	-	1.1	1.1	1.0	-	1.1
Mixture 45_HR										
AC, %	6.8	±0.3	6.81	6.72	6.87	7.22	6.67	7.11	6.82	6.97
G_{mm}	2.497	-	2.500	2.506	2.503	2.505	2.512	2.508	2.515	2.510
VTM, %	5.2	2.0-5.0	0.3	0.8	0.5	0.6	0.6	0.6	0.8	0.6
VMA, %	20.2	Min. 16	16.1	16.4	16.5	17.5	16.2	17.2	16.7	16.9
VFA, %	74.3	70-85	98.3	95.3	97.1	96.5	96.1	96.4	95.2	96.4
FA Ratio	1.2	0.7-1.3	1.2	1.1	-	1.2	1.2	1.2	-	1.2
Mixture 45_HR_RA										
AC, %	6.2	±0.3	6.10	6.19	6.24	6.15	6.30	6.34	6.28	6.21
G_{mm}	2.519	-	2.540	2.541	2.542	2.546	2.539	2.541	2.536	2.535
VTM, %	2.7	2.0-5.0	1.9	2.1	2.6	2.6	2.2	2.6	2.6	2.3
VMA, %	16.9	Min. 16	16.0	16.4	17.0	16.8	16.8	17.3	17.1	16.7
VFA, %	84.0	70-85	88.5	87.5	84.9	84.6	86.9	84.7	84.8	86.0
FA Ratio	1.2	0.7-1.3	1.4	1.5	-	1.4	1.2	1.2	-	1.3
Mixture 45_HR_L										
AC, %	6.0	±0.3	6.05	5.97	6.00	6.00	6.07	6.16	6.17	6.02
G_{mm}	2.521	-	2.522	2.536	2.547	2.544	2.536	2.533	2.534	2.537
VTM, %	4.2	2.0-5.0	1.9	2.0	3.2	3.2	2.4	2.6	3.2	3.0
VMA, %	17.9	Min. 16	16.1	16.0	17.2	17.2	16.7	17.0	17.5	17.0
VFA, %	76.6	70-85	88.2	87.7	81.5	81.6	85.3	84.8	81.8	82.3

FA Ratio	1.3	0.7-1.3	1.3	1.2	-	1.2	1.2	1.2	-	1.2
Mixture 60_HR_L_RA										
AC, %	6.0	±0.3	5.99	5.89	5.84	6.14	5.7	5.88	6.01	5.90
G_{mm}	2.538	-	2.523	2.550	2.548	2.548	2.547	2.546	2.537	2.542
VTM, %	1.5	2.0-5.0	0.2	1.4	1.3	1.7	1.9	2.4	2.1	2.2
VMA, %	15.3	Min. 16	14	15.0	14.8	15.8	15.0	15.8	15.8	15.7
VFA, %	90.0	70-85	98.9	90.4	91.2	89.5	87.1	84.6	87.8	85.9
FA Ratio	1.5	0.7-1.3	1.5	1.4	-	1.4	1.4	1.3	-	1.4

AC = asphalt binder content; G_{mm} = maximum specific gravity; VTM = voids in total mixture; VMA = voids in mineral aggregate; VFA = voids filled with asphalt; FA = fines to asphalt; VDOT = Virginia Department of Transportation; VTRC = Virginia Transportation Research Council; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

Generally, the gradations of different samples across each mixture, as determined by the producer, VDOT district, and VTRC, were within the allowable production tolerance. Table 1 details the gradations of the mixture designs, which are shown in Figure 2. Figure 2 indicates that the design gradations of the six mixtures closely resemble one another, with standard deviation in passing percentages at different sieve sizes ranging from 0 to 2.1%. The design gradations, alongside the average producer, VDOT district, and VTRC gradations, are shown in Figure 3. Overall, the gradation curves of the designs, non-reheats, and reheats are strongly congruent, with the curves of the non-reheats and reheats showing near-perfect overlap.

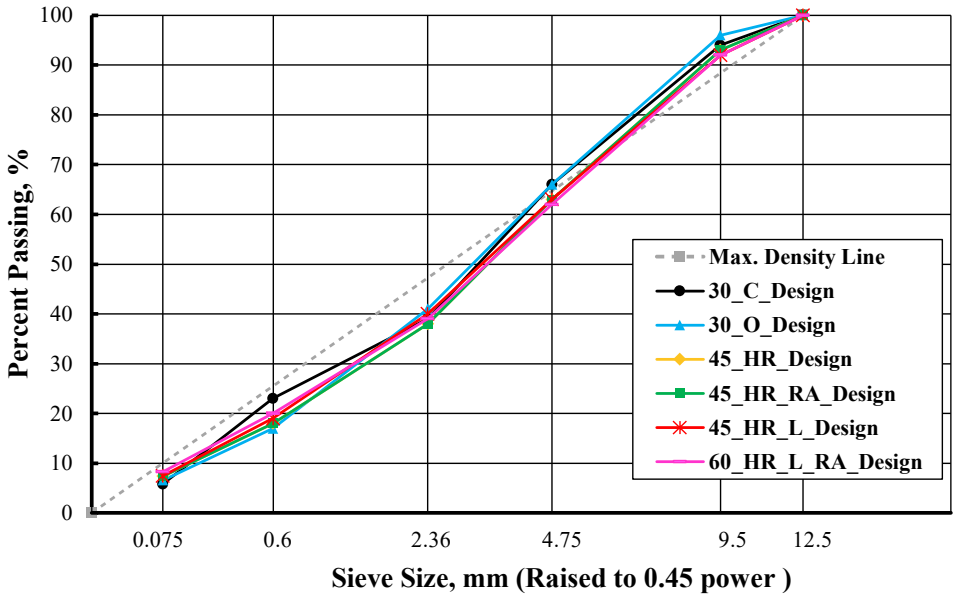

FIGURE 2. Design Gradations for the Six Mixtures. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

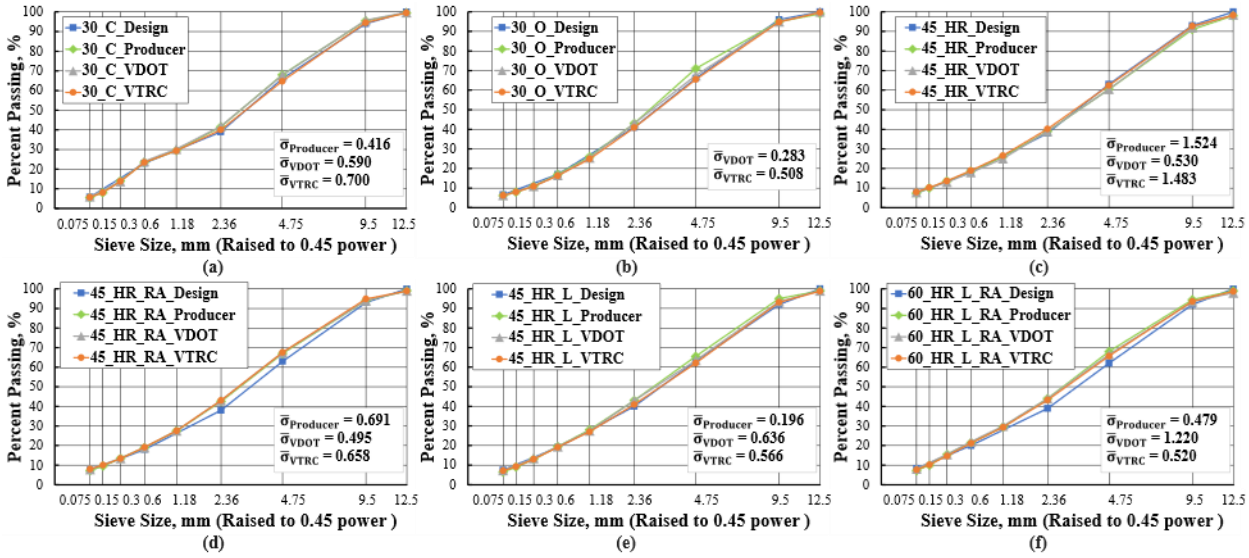

FIGURE 3. Average Aggregate Gradations Determined During Design by Producer and During Production by Producer, VDOT District, and VTRC for: (a) Mixture 30_C; (b) Mixture 30_O; (c) Mixture 45_HR; (d) Mixture 45_HR_RA; (e) Mixture 45_HR_L; and (f) Mixture 60_HR_L_RA. VDOT = Virginia Department of Transportation; VTRC = Virginia Transportation Research Council; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; $\bar{\sigma}$ = average standard deviation. $\bar{\sigma}$ from producer missed for mixture 30_O, because only the 1st sample was checked by producer.

4.2. Durability Assessment

The Cantabro test was used to evaluate the durability of AC mixtures *(19)* and was performed on designs, non-reheats, and reheats. Figure 4 shows the CML for all mixtures with its sample standard deviation. The lower CML indicated better durability performance. The maximum limit of 7.5% as specified by the VDOT BMD special provision is represented by a red dashed line. All plant-produced HRAP BMD mixture samples complied with the maximum 7.5% CML VDOT BMD limit requirement. Meanwhile, all plant-produced BMD mixtures exhibited superior durability compared to that of mixture 30_C, as indicated by their lower mean ML values in Table 3. These observations highlighted the enhancement of durability in BMD mixtures attributed to higher binder content. In comparison to HRAP mixtures, mixtures 30_C and 30_O exhibited a more pronounced discrepancy between plant-produced mixtures and laboratory-produced mixtures in terms of CML. In the case of mixture 30_O, such discrepancies could be attributed to higher VTM and lower VFA compared to the designs. Furthermore, the control mixture 30_C showed increased sensitivity to plant production variations compared to BMD mixtures.

The CML values of the plant-produced mixtures (both non-reheats and reheats) were significantly higher compared to the corresponding designs. However, exceptions are observed in Table A-3 of the Appendix: in the comparison between the designs and reheats for mixture 45_HR, the plant-produced mixture, with its higher asphalt content and lower VTM, exhibited a reduced CML. Additionally, in the comparison between designs and non-reheats of mixture 60_HR_L_RA, there was no significant difference detected. Gyratory pills were used for the Cantabro test. The plant-production variability and the mass-controlled compaction may be factors impacting the CML. Additionally, a uniform pattern was observed across all mixtures, wherein the reheats significantly presented higher CML values than the non-reheated

counterparts. This suggested that the reheating process may generally result in lower durability, which aligns with previous studies (*13*). A more detailed comparison of reheats and non-reheats can be found in Table A-4-1 of the Appendix. Statistical significance is generally not found among samples in both non-reheats and reheats. However, exceptions were observed between non-reheats_3 and non-reheats_4 of mixture 45_HR_RA, and between reheats_2 and reheats_3 of mixture 60_HR_L_RA. These discrepancies can be attributed to variations in VTM and VFA, as shown in Table 3.

Table 4 provides a summary of the descriptive statistics employed in the CML of plant-produced mixtures. For non-reheats, the coefficient of variance (COV) exhibited a range between 9.6% and 20.5% with the lowest COV was observed in the 30_C mixture. The COV values for reheats consistently remained below their corresponding non-reheated counterparts, varying between 8.8% and 12.9%. An important factor to consider is that the preparation of non-reheated specimens must be completed shortly after sampling during production. However, there is typically less time pressure associated with preparing reheated specimens.

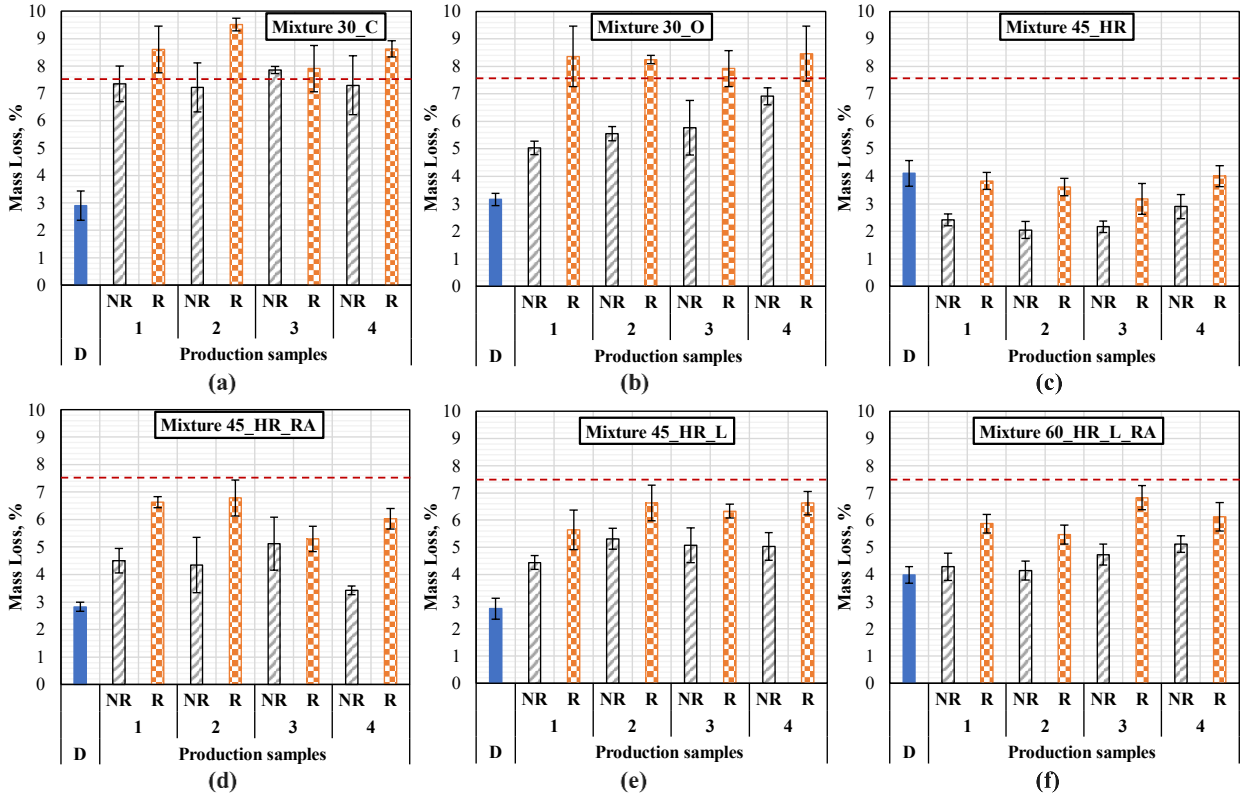

FIGURE 4. Cantabro Mass Loss Results for Mixtures: (a) Mixture 30_C; (b) Mixture 30_O; (c) Mixture 45_HR; (d) Mixture 45_HR_RA; (e) Mixture 45_HR_L; and (f) Mixture 60_HR_L_RA. Values are the average of three replicates; I-bars indicate ± one standard deviation; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; NR = Non-reheats; R = Reheats; Red dashed line = BMD special provision limit.

Table 4. Summary of Descriptive Statistics for Performance Tests of Plant-Produced Mixtures

Mass loss, %	Mean		COV	
	Non-reheats	Reheats	Non-reheats	Reheats
30_C	7.4	8.7	9.6	9.3
30_O	5.8	8.2	14.7	8.8
45_HR	2.4	3.7	18.1	12.9
45_HR_RA	4.3	6.2	20.5	11.7
45_HR_L	5.0	6.3	10.6	10.0
60_HR_L_RA	4.6	6.1	11.5	10.4
CT index	**Mean**		**COV**	
	Non-reheats	Reheats	Non-reheats	Reheats
30_C	138.8	54.1	20.7	9.4
30_O	242.2	75.9	24.0	25.3
45_HR	943.7	405.2	20.4	28.3
45_HR_RA	342.2	158.0	24.7	49.6
45_HR_L	220.0	94.3	33.7	20.4
60_HR_L_RA	219.7	101.2	22.0	21.1
APA rut depth, mm	**Mean**		**COV**	
	Non-reheats	Reheats	Non-reheats	Reheats
30_C	4.2	4.3	26.6	26.8
30_O	6.1	4.3	22.4	20.6
45_HR	11.9	6.7	12.2	15.2
45_HR_RA	5.3	5.1	20.7	16.6
45_HR_L	4.4	4.5	30.0	25.4
60_HR_L_RA	3.9	5.2	33.3	21.2

COV = coefficient of variation; CT = cracking tolerance; APA = asphalt pavement analyzer; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

4.3. Cracking Performance Assessment

The IDT-CT (also known as IDEAL-CT) is being used by VDOT to assess mixture resistance to cracking. Figure 5 displays the mean IDT-CT test results along with the respective sample standard deviations. The higher CT index indicated the higher cracking resistance. The minimum limit of 70 as specified by the VDOT BMD special provision was represented by a red dashed line. In general, all produced HRAP BMD mixtures met the VDOT BMD CT index threshold of 70, indicating satisfactory cracking resistance. Nevertheless, lower CT indices were observed in the reheats of mixtures 30_C and 30_O, with this phenomenon being particularly pronounced in mixture 30_C. These observations were similar to findings from Cantabro tests, highlighting the enhancement of cracking resistance in BMD mixtures attributed to higher binder content. Additionally, all BMD mixtures exhibited better cracking resistance compared to mixture 30_C, as indicated by their higher mean CT index values (the same pattern found in the durability evaluation), shown in Table 4.

Reheats across different mixtures demonstrated CT index values closer to the designs for all mixtures. Non-reheats generally exhibited significantly higher CT index values compared to designs excluding mixture 45_HR_L and 60_HR_L_RA. The disparity between reheats and non-reheats across different mixtures proved to be statistically significant, as illustrated in Table A-3-2 of the Appendix. This observation underscored the potential aging effect introduced during the reheating process, which may potentially compromise the cracking resistance of the mixtures, aligning with findings from previous studies (*13*). The sensitivity to reheating varied among

different mixtures. In the pairwise comparisons using the KW test between samples presented in Table A-4-2 of the Appendix, the absence of statistically significant differences among four samples in both non-reheats and reheats indicates a consistent plant-production and specimens fabrication process. Note that the IDT-CT specimens were compacted at VTM of 7±0.5%. Compared with Cantabro specimens (volumetric pills), the controlled air voids may have contributed to less statistical significance/variance in the CT index between design and production specimens.

In the case of non-reheats, the COV ranged from 20.4% to 33.7%. while for reheats, COV values ranged from 9.3% to 49.6%. In most instances, the COV of the CT index exceeded VDOT's single-operator precision estimate (1s) of 18.3% (*26*). Furthermore, a consistent trend of higher CT index COV for all BMD mixtures was observed compared to mixture 30_C, except for the non-reheats of mixture 45_HR. This could be attributed to higher RAP contents, as RAP stockpiles are known to be variable and require stricter control and better management during production.

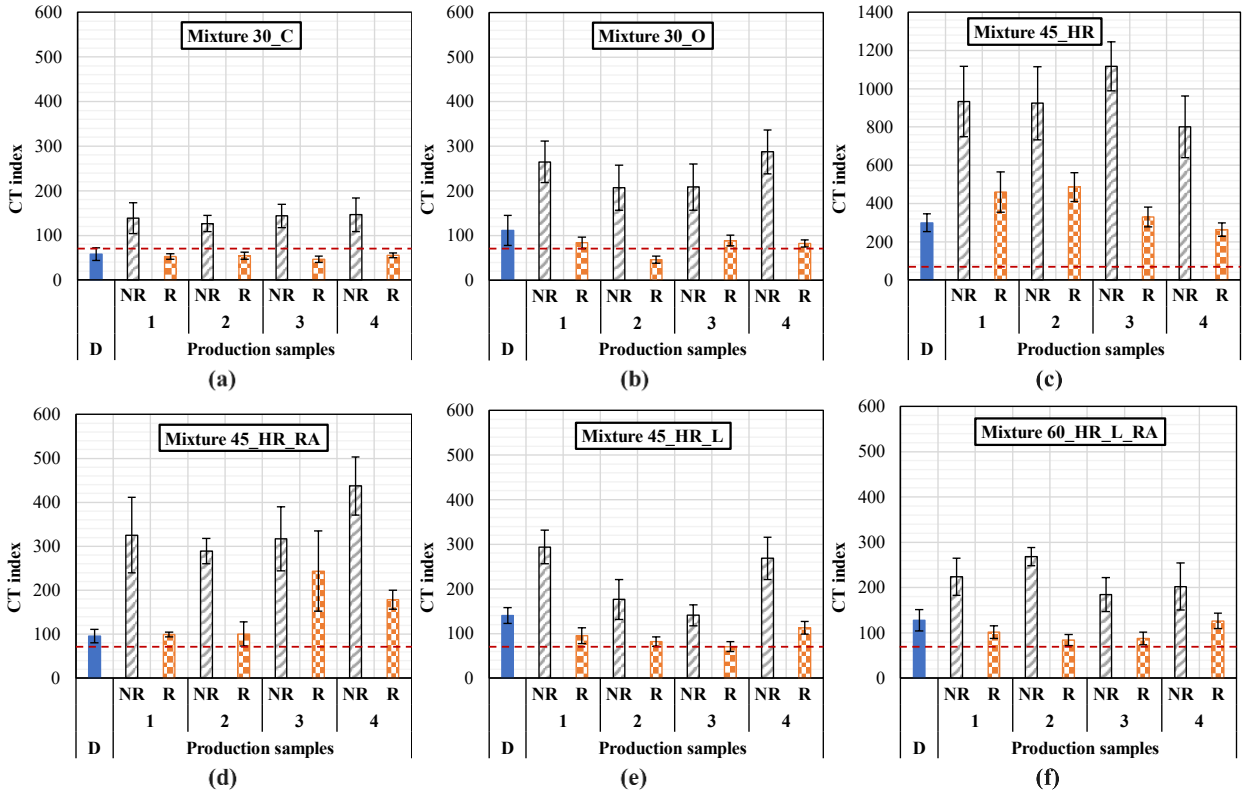

FIGURE 5. Performance Test Data for IDT-CT: (a) Mixture 30_C; (b) Mixture 30_O; (c) Mixture 45_HR; (d) Mixture 45_HR_RA; (e) Mixture 45_HR_L; and (f) Mixture 60_HR_L_RA. Values are the average of five replicates. I-bars indicate ± one standard deviation; CT = cracking tolerance; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; NR = Non-reheats; R = Reheats; Red dashed line = BMD special provision limit.

4.4. Rutting Performance Assessment

The APA rut test was conducted in accordance with AASHTO T 340 *(27)*. A higher APA rut depth (RD) value indicated an increased susceptibility to rutting damage. The APA test results for the respective mixtures are illustrated in Figure 6. Within this figure, each bar symbolizes the

mean values derived from 2 molds (4 specimens), accompanied by its corresponding error bar denoting the range of plus or minus one sample standard deviation. Note that the non-reheats of both sample A and C in mixture 45_HR only have a final mean value documented. The maximum 8 mm threshold is represented by a red dashed line.

All mixtures except the non-reheats of mixture 45_HR, met the VDOT BMD recommended threshold of 8 mm. This could be attributed to the incorporation of an excessive asphalt binder content during the production. Such observations indicate that for BMD mixtures, rutting performance can achieve a well-balanced status during both the design and production phases. Furthermore, there are no significant differences between the designs and production samples (both reheats and non-reheats), except for the designs vs. non-reheats of mixture 45_HR, where a significant difference was detected as shown in Table A-3-2 of the Appendix. A more detailed comparison, presented in Table A-4 of the Appendix, reveals the absence of statistically significant differences among the four samples in both non-reheats and reheats, indicating a consistent plant-production and specimen fabrication process. Additionally, the reheating process introduced a potential aging effect, resulting in a decreased APA RD for mixtures 30_O and 45_HR. However, the effect of the reheating process on rutting performance varied when RA or soft binder conditions were applied to mixtures, including 45_HR_RA, 45_HR_L, and 60_HR_L_RA. Similarly, likely due to air void requirements, statistical differences in the APA RD test between produced specimens and designs were less significant than that of the Cantabro test. In Table 3, for non-reheats, COV values exhibited a range spanning from 12.2% to 33.3%. For reheats, COV values ranged from 15.2% to 26.8%.

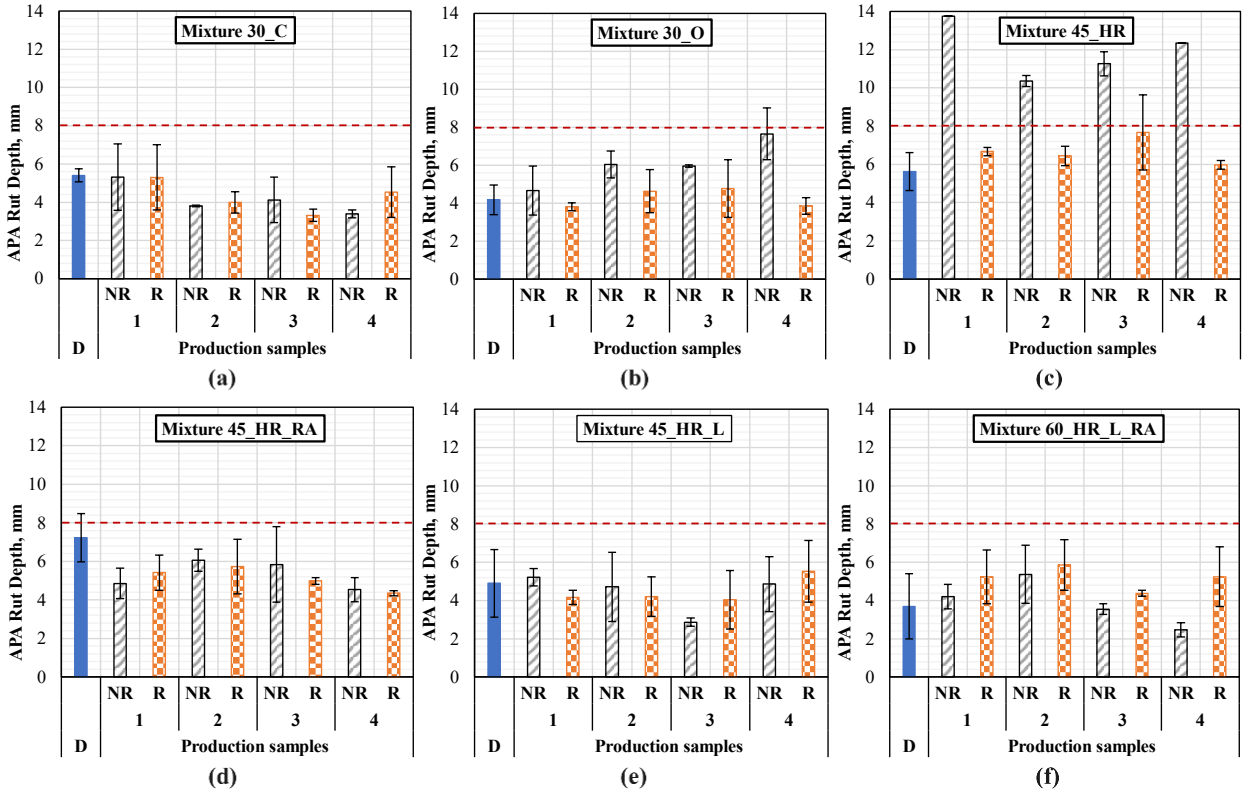

FIGURE 6. Performance Test Data for APA Rut Depth at 64°C and 8,000 Loading Cycles: (a) Mixture 30_C; (b) Mixture 30_O; (c) Mixture 45_HR; (d) Mixture 45_HR_RA; (e) Mixture 45_HR_L; and (f) Mixture 60_HR_L_RA. Values are the average of three replicates. I-bars indicate ± one standard deviation; APA = asphalt pavement

analyzer; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; NR = Non-reheats; R = Reheats; Red dashed line = BMD special provision limit.

5. ADDITIONAL DISCUSSION AND ANALYSES

5.1. Assessing the Effect of Reheating

Based on pair-wise comparisons (Table A-4), there were no significant differences in performance indices among the four produced samples for both non-reheats and reheats, except for two cases (non-reheat-3 vs.non-reheat-4 of mixture 45_HR_RA and reheat-2 vs. reheat-3 of mixture 60_HR_L_RA) from Cantabro specimens. The relationship between the mean performance indices of non-reheats and reheats for the six evaluated mixtures is illustrated in Figure 7 using mean sample values. Figures 7a and 7b reveal that the reheating process resulted in higher CML values and lower CT index, suggesting a potential decrease in the durability and cracking resistance of the produced mixtures for both control and BMD mixtures. The current CML requirement established by VDOT is based on reheated specimens. This analysis indicated that a mean CML value of 7.5% for reheats corresponded to a mean CML value of 5.8% for non-reheats, as indicated by the regression equation presented in Figure 7a. Notably, nineteen out of twenty-four (79.16%) mean CML values for non-reheats fell below the 5.8% threshold. In contrast, all reheats from 30% RAP mixtures exceeded the 7.5% limit. Regarding the CT index, an independent analysis conducted by VTRC examined the potential correlation between CT index values for reheats and non-reheats using data from field trials spanning 2019, 2020, and 2021 (*28*) and included both control and BMD mixtures. The study recommended a CT value of 95 as a criterion for non-reheats. Applying this criterion, it was observed that all non-reheats (100%) evaluated in this study exhibited mean CT index values exceeding 95. For reheats, five out of twenty-four (20.83%) mean CT index values fell below 70, as shown in Figure 7b. Figure 7c indicated that the impact of the reheating process on rutting performance of evaluated mixtures, excluding mixture 45_HR, was not significant. A total of twenty out of twenty-four (83.33%) mean RD values for non-reheats were less than 8.0 mm, while all reheats (100%) evaluated in this study exhibited mean APA RD below 8 mm.

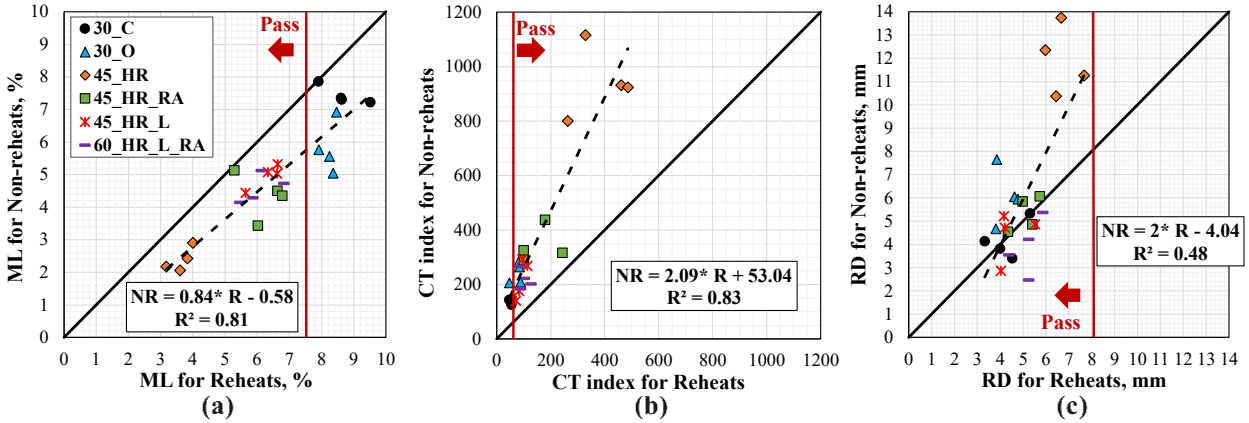

FIGURE 7. Relationship Between Non-reheats and Reheats: (a) Cantabro Test; (b) IDT-CT Test; (c) APA Rutting Test. RD = rut depth; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; NR = non-reheats; R = reheats; Solid black line = equality line. Solid red line = VDOT's balanced mix design limit for reheated asphalt surface mixtures.

5.2. Quantifying the Degree of Balance for the Produced Mixture

In the preceding analysis, the results of the BMD performance tests were discussed individually, employing the VDOT BMD thresholds established based on reheats. Additionally, it determined the limits applicable for the analysis of non-reheats. The following section was primarily devoted to conducting a comprehensive assessment of whether the produced mixtures maintain balance among different performance indices during production. To facilitate the assessment, traditional comparative methods necessitated separate sets of pairwise comparisons of primary performances (*29*). Figures 8a, 8b, 8c respectively illustrate the comparisons between the IDT-CT and Cantabro test; the APA rut test and Cantabro test; and the IDT-CT and APA rutting test, respectively, for both designed and produced mixtures. However, the CT index of non-reheated specimens of mixture 45_HR was excessively high. For a more effective visualization, Figure 8d displays the IDT-CT vs. APA rut test comparison, excluding non-reheats of mixture 45_HR. In addition to presenting the mean values of different performance indices, VDOT BMD limits for reheats and applicable limits for non-reheats were represented by solid and dashed lines of different colors, respectively. Currently, VDOT has not specified a limit for CML when evaluating non-reheats; therefore, the limit of 5.8% derived from last section is used solely for the purposes of this study. Two rectangular green blocks indicate the regions where mixtures maintain balance, with the lighter region designated for reheats and the darker region for non-reheats.

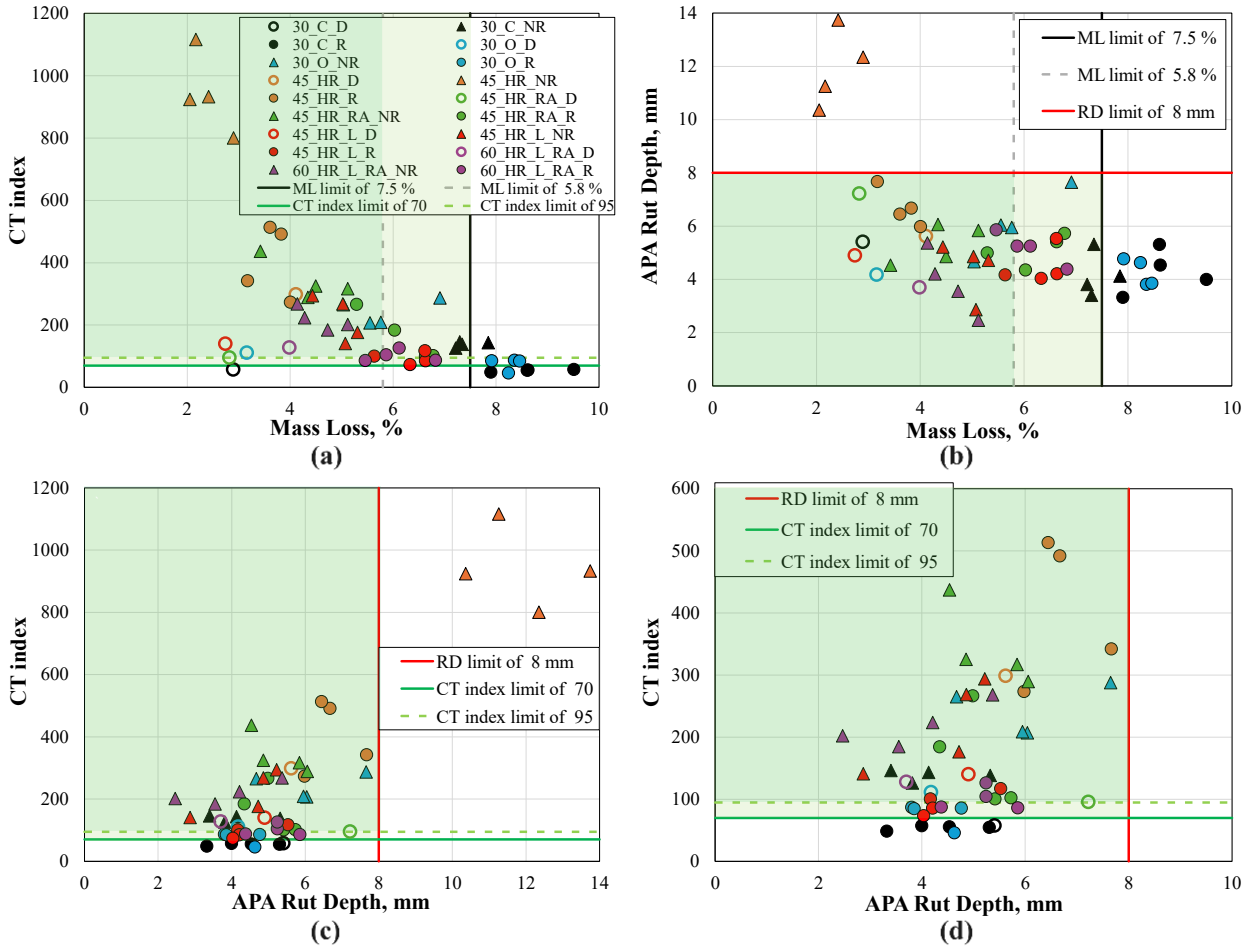

FIGURE 8. Comparison of Mean Performance for Each Mixture Variation: (a) IDT-CT vs. Cantabro test; (b) APA Rutting Test vs. Cantabro test; (c) IDT-CT vs. APA Rutting Test; (d) IDT-CT vs. APA Rutting Test without non-reheats of mixture 45_HR. CT = cracking tolerance; RD = rut depth; D = design; NR = non-reheats; R = reheats; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder. Solid lines = VDOT's balanced mix design limit for reheated asphalt surface mixtures; Dashed lines = VDOT's balanced mix design limit for non-reheated asphalt surface mixtures.

In the analysis, it was evident that while pairwise comparisons were effective in evaluating two primary indices, assessing the balance status of three primary indices in this paper through this comparison method can be cumbersome and prone to oversight. Tong et al. *(17)* employed a Three-Dimensional (3-D) plot to comprehensively analyze the relative relationships among the three primary performance indices of six mixtures and proposed a Composite Performance Index (CPI) to evaluate mixture overall performances. Inspired by this approach, this paper also aimed to utilize 3-D plots to examine the balance status of produced mixtures among three primary performances and propose a revised CPI, denoted as CPI_R, to evaluate the composite performance for produced mixtures. This composite performance encompasses considerations of durability, cracking resistance, and rutting resistance concurrently. For the 3-D plot, each performance index should be normalized by its corresponding threshold using Equations 1 through 3. The purpose of obtaining the normalized indices is to evaluate the mixtures' performances using a unified manner. Specifically, a smaller normalized index indicates better performance. For the CT index, a larger value signifies better cracking resistance, which is opposite to the trends of other metrics such as CML and RD. Therefore, Equation (3) is formulated differently compared to Equations (1) and (2). Table 5 presents the normalized performance indices of the evaluated produced mixtures. If a value exceeds 1, it indicates that the mixture's corresponding performance is not balanced and is marked in red. This outcome is succinctly depicted by the 3-D plot, with CML, RD, and CT index corresponding to the x, y, and z coordinates, respectively.

Figure 9 displays the 3-D plots for all produced mixtures, wherein points within a red unit cube reflect mixtures maintaining a well-balanced status across three primary performances. Moreover, mixtures closer to the origin denote better composite performance. CPI_R , representing the distance from the origin to the point on the 3-D plot, can be computed using Equation 4. The 'R' in CPI_R stands for "revised" and "reversed," indicating that a lower CPI_R value, approaching zero, denotes superior composite performance of the respective mixture.

$$x = Normalized\ ML = \frac{ML}{ML\ threshold} \qquad \text{Equation 1}$$

$$y = Normalized\ RD = \frac{RD}{RD\ threshold} \qquad \text{Equation 2}$$

$$z = Normalized\ CT\ index = \frac{CT\ index\ threshold}{CT\ index} \qquad \text{Equation 3}$$

$$CPI_R = \sqrt{Ax^2 + By^2 + Cz^2} \qquad \text{Equation 4}$$

Where CML thresholds for non-reheats and reheats are 5.8% and 7.5%, respectively; RD thresholds for non-reheats and reheats are both 8 mm; *CT index thresholds* for non-reheats and reheats are 95 and 70, respectively; A, B, C are weight factors, in this case, A = B = C =1.

Addressing the constraints of conventional pairwise comparisons, the utilization of 3-D plots offers a more intuitive approach to assessing the balanced status of produced mixtures. Additionally, 3-D plots incorporate a scaling function that can mitigate the impact of extreme test results, such as the CT index of 45_HR (refer to Figures 8c and 8d), thereby reducing interference in image-based analyses. Owing to distinct thresholds for non-reheats and reheats, separate 3-D plots for each are depicted in Figures 9a and 9b. For HRAP mixtures, among non-reheats, 75% (12 out of 16) of mixtures exhibited balanced performance across three primary metrics, albeit with non-reheats of 45_HR falling short of the 8 mm limit. Conversely, all reheats of HRAP mixtures achieved a balanced performance across all three primary metrics, attaining a 100% balance rate. Instances of unbalanced performance were predominantly observed in 30% RAP mixtures, particularly evident in 30_C. Despite the normalized RD metric for all such cases meeting the prescribed limit, both non-reheats and reheats of mixture 30_C failed to meet the CML limit. Additionally, reheats of 30_C also failed to meet the CT index limit. As for mixture 30_O, the primary issue arose from reheats failing to meet the CML limit.

Table 5 also provides a summary of all CPI_R values, with those exceeding 1.73 (the length of the diagonal of a unit cube [$\sqrt{3}$] in 3-D plot) highlighted in yellow. Notably, the CPI_R values for all HRAP mixtures remained below 1.73. It was also evident that exceeding a single prescribed limit might not necessarily result in a CPI exceeding 1.73. However, a significant increase in CPI_R values surpassing 1.73 was primarily observed in the reheats of mixture 30_C, attributed to both its CML and CT index exceeding their respective limits of 1. In this study, the CPI_R calculation used identical weight factors (A=B=C=1) for the three distinct metrics, with 1.73 serving as an initial limit for CPI_R. As additional laboratory experimental data and further field performance evaluations collected and compiled together, transportation agencies can undertake a thorough assessment of the composite performance of BMD mixtures and, if necessary, establish corresponding CPI_R thresholds. Furthermore, agencies can more effectively use the CPI_R as part of their acceptance program to determine appropriate pay factors, determining contractor bonuses or penalties based on a single/unified index that reflects the composite performance in terms of durability, cracking, and rutting. However, it is crucial to clarify that the weight factors (A, B, C) in our study are applied to the normalized values of the CML, RD, and CT indices (as per Equations 1-3). These normalized values have already taken into account the varying levels of concern different states have for different types of distress, which is reflected in the selection of different prescribed test thresholds. For instance, Virginia has chosen thresholds of 7.5%, 8 mm, and 70 for the CML, RD, and CT indices, respectively. In states where cracking resistance is a higher priority, transportation agencies can adjust the corresponding thresholds, such as increasing the CT index threshold from 70 to 90.

Table 5. Summary of Normalized Performance indices for 3-D Plots

Samples	30_C Normalized Values				30_O Normalized Values				45_HR Normalized Values			
	CML	*RD*	*CT index*	CPI_R	*CML*	*RD*	*CT index*	CPI_R	*CML*	*RD*	*CT index*	CPI_R
Non-reheats A	1.27	0.66	0.68	1.59	0.87	0.58	0.36	1.11	0.42	1.72	0.10	1.77
Non-reheats B	1.24	0.48	0.75	1.53	0.96	0.75	0.46	1.30	0.35	1.30	0.10	1.35
Non-reheats C	1.35	0.52	0.66	1.59	0.99	0.74	0.46	1.32	0.37	1.41	0.09	1.46
Non-reheats D	1.26	0.42	0.65	1.48	1.19	0.96	0.33	1.56	0.50	1.54	0.12	1.63

Reheats A	1.15	0.66	1.28	1.84	1.12	0.48	0.80	1.45	0.51	0.83	0.14	0.99
Reheats B	1.27	0.50	1.22	1.83	1.10	0.58	1.52	1.96	0.48	0.81	0.14	0.95
Reheats C	1.05	0.42	1.44	1.83	1.06	0.60	0.82	1.46	0.42	0.96	0.20	1.07
Reheats D	1.15	0.57	1.26	1.80	1.13	0.48	0.83	1.48	0.53	0.75	0.26	0.95
Samples	45 HR RA Normalized Values				45 HR L Normalized Values				60 HR L RA Normalized Values			
	CML	*RD*	*CT index*	CPI_R	*CML*	*RD*	*CT index*	CPI_R	*CML*	*RD*	*CT index*	CPI_R
Non-reheats A	0.78	0.61	0.29	1.03	0.77	0.65	0.32	1.06	0.74	0.53	0.43	1.00
Non-reheats B	0.75	0.76	0.33	1.11	0.92	0.59	0.54	1.21	0.71	0.67	0.35	1.04
Non-reheats C	0.88	0.73	0.30	1.18	0.87	0.36	0.67	1.16	0.82	0.44	0.51	1.06
Non-reheats D	0.59	0.57	0.22	0.85	0.87	0.61	0.35	1.12	0.88	0.31	0.47	1.05
Reheats A	0.88	0.68	0.70	1.31	0.75	0.52	0.70	1.15	0.78	0.66	0.67	1.22
Reheats B	0.90	0.72	0.69	1.34	0.88	0.53	0.82	1.31	0.73	0.73	0.81	1.31
Reheats C	0.71	0.62	0.26	0.98	0.84	0.50	0.95	1.37	0.91	0.55	0.80	1.33
Reheats D	0.80	0.54	0.38	1.04	0.88	0.69	0.60	1.27	0.82	0.66	0.55	1.18

ML = mass loss; RD = rutting depth; CT = cracking tolerance; CPI_R = revised composite performance index; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; A, B, C, and D = Sample ID.

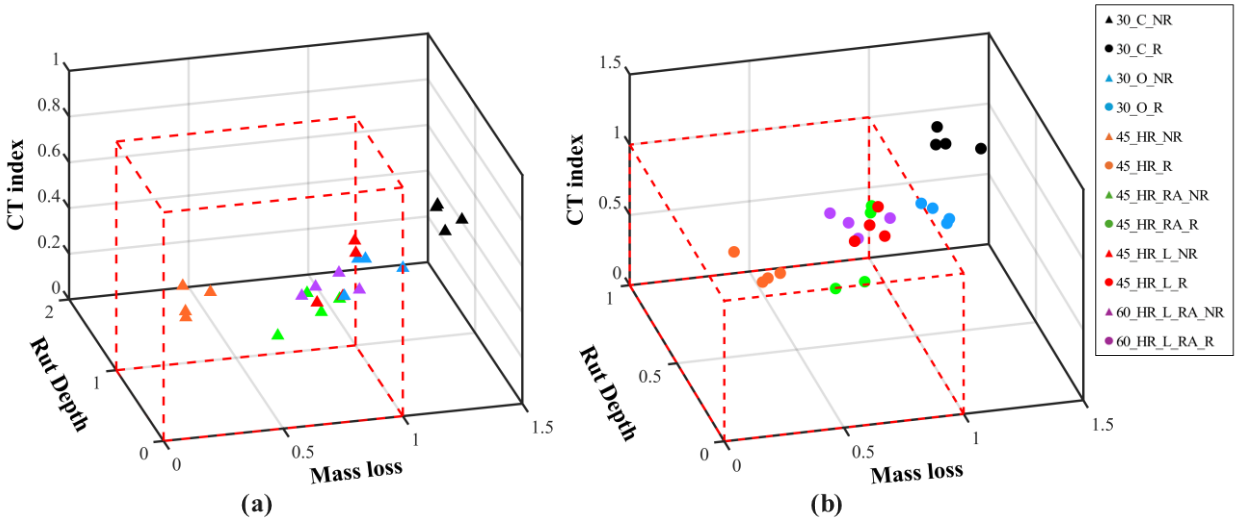

FIGURE 9. 3-D Plots: (a) Non-reheated Specimens; (b) Reheated Specimens. NR = non-reheats; R = reheats; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; the red unit cube indicates the well-balanced area for the three normalized primary performance indices.

6. SUMMARY OF FINDINGS AND FUTURE WORK

This research assessed the impact of production and associated variability on a control mixture and five BMD surface mixtures with conventional and high RAP contents. These mixtures were produced during design and collected at the plant during production and paving of the APT facility. This evaluation focused on durability, cracking resistance, and rutting resistance as primary facets of the BMD concept in Virginia. These specimens included laboratory-mixed laboratory-compacted specimens (designs), plant-mixed laboratory-compacted specimens at the plant (non-reheats), and plant-mixed laboratory-compacted specimens that underwent a reheating process (reheats). Statistical analyses were used to evaluate variations in performance metrics between specimens fabricated using different procedures and among various samples produced using mixtures collected from various lots. The main outcomes and key recommendations are summarized as follows:

1- Precision in plant control over the mixture's gradation was observed, resulting in minor differences between various samples. The averaged standard deviations of sieve sizes across all samples were remarkably low, suggesting that gradation variability may not substantially influence the results of performance tests.

2- The control mixture met the prescribed production tolerances for volumetric requirements. Deviations were mostly evident in the VTM and VFA variance of BMD optimized mixtures, which could be the key factors resulting from production and specimen fabrication affecting CML.

3- Compared with mass-controlled volumetric pills used for the Cantabro test, compacting IDT-CT and APA specimens at VTM of 7±0.5% exhibited less significant difference between designs and produced specimens and among the different samples during production.

4- The observations from both durability and cracking resistance assessments for the evaluated mixtures yielded similar outcomes. The produced HRAP BMD mixtures complied with the prescribed VDOT BMD special provisions. Instances failing to meet the corresponding specifications were exclusively identified within the 30% RAP mixtures, with the control mixture exhibiting the most pronounced deviation. This highlighted the ability of a higher OBC to concurrently enhance the durability and cracking performance of mixtures. It also emphasized the necessity and effectiveness of employing soft binder and RAs with the BMD concept for HRAP mixtures.

5- The aging effect introduced during the reheating process may compromise the durability and cracking resistance of the mixtures.

6- Different specimens showed more consistent APA RD results compared with Cantabro test and IDT-CT. Furthermore, the impact of the reheating process on rutting performance varied when a RA or softer binder were used in the mixtures.

7- The use of 3-D plots provided a more intuitive method for evaluating the balanced status among multiple performance indices of produced mixtures. It demonstrated that almost all produced HRAP mixtures demonstrate a well-balanced status. Additionally, A CPI_R was introduced to assess the composite performance of the produced mixture -durability, cracking, and rutting. CPI_R could be used as part of agencies' acceptance program for BMD mixtures to determine a pay factor for possible bonuses or penalties. This approach is particularly pertinent for BMD mixture production, moving beyond the exclusive focus on volumetric properties that predominates current practices.

As the primary laboratory testing component of the Virginia APT project, this study will establish a foundation for future work aimed at bridging the gap between laboratory testing and APT field performance measurement. The analyses presented in this paper highlight the advantages and disadvantages of using specimens fabricated with and without the reheating of loose mixture samples. The effectiveness of using reheated and non-reheated samples with different thresholds will be evaluated by comparing APT field performance with laboratory metrics.

ACKNOWLEDGMENTS

The authors thank Troy Deeds, Derek Lister, Jacob Oliver, Jennifer Samuels, and Andrew Barbour of VTRC; Danny Martinez Rodriguez and Donnie Dodds formerly of VTRC; and Billy Hobbs of VTTI, for their outstanding efforts in sample collection and testing. Appreciation is also extended to Travis Higgs and Clyde Landreth of VDOT for their assistance and support. Sincere appreciation is expressed to Boxley Asphalt Paving and Ingevity for their contributions to this effort.

FUNDING

This research was funded by the Virginia Department of Transportation (Project number 0000124200).

DISCLAIMER

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. 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 paper 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.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design, data collection, analysis and interpretation of result, draft manuscript preparation: B. Tong, J. Habbouche, S. D. Diefenderfer, G. W. Flintsch and I. Boz. All authors reviewed the results and approved the final version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw data was generated at VDOT/VTRC. Derived data supporting the findings of this study are available from the corresponding author [B.T.] on request.

REFERENCES

1. Golalipour, A., V. Veginati, and D. J. Mensching. Evaluation of Asphalt Mixture Performance Using Cracking and Durability Tests at a Full-Scale Pavement Facility. *Transportation Research Record: Journal of Transportation Research Record*, 2021. 2675: 226 - 236.
2. West, R., C. Rodezno, F. Leiva, and F. Yin. *NCHRP Project 20-07/Task 406 Development of a Framework for Balanced Mix Design*. Transportation Research Board, Washington, DC, 2018.
3. Yin, F., and R. West. *Balanced Mix Design Resource Guide*. NAPA IS-143, National Asphalt Pavement Association, Greenbelt, MD, 2021.

4. Zaumanis, M., R. B. Mallick, and R. Frank. Evaluation of Rejuvenator's Effectiveness with Conventional Mix Testing for 100% Reclaimed Asphalt Pavement Mixtures. *Transportation Research Record: Journal of the Transportation Research Board*, 2013. 2370(1): 17-25.
5. Epps-Martin, A., F. Kaseer, E. Arámbula-Mercado, A. Bajaj, L. G. Cucalon, F. Yin, A. Chowdhury, J. Epps, C. Glover, E. Y. Hajj, N. Morian, J. S. Daniel, M. Oshone, R. Rahbar-Rastegar, C. Ogbo, and G. King. *NCHRP Report 927: Evaluating the Effects of Recycling Agents on Asphalt Mixtures with High RAS and RAP Binder Ratios.* Transportation Research Board, Washington, D.C., 2020.
6. Abu Saq, M., E. Kassem, and A. Muftah. Cracking Performance Evaluation of RAP Mixes. Journal of Transportation Engineering, Part B: Pavements. 2023.
7. Sabouri, M. Evaluation of performance-based mix design for asphalt mixtures containing Reclaimed Asphalt Pavement (RAP). Construction and Building Materials, 235, 117545. 2020.
8. McDaniel, R. S., A. Shah, G. A. Huber, and A. Copeland. "Effects of reclaimed asphalt pavement content and virgin binder grade on properties of plant produced mixtures." Supplement, Road Mater. Pavement Des. 13 (S1): 161–182. 2012. https://doi.org/10.1080/14680629.2012.657066.
9. Brown, E. R., P. S. Kandhal, F.L. Roberts, Y. R. Kim, D Lee, and T.W. Kennedy. Hot Mix Asphalt, Materials, Mixture Design and Construction, 3rd Edition, NAPA Research and Education Foundation. Lanham, Maryland, 2009
10. Hughes, C. S. NCHRP Synthesis 346: State Construction Quality Assurance Programs. TRB, Washington, DC, pp. 45. 2005.
11. Mohammad, L. N., M. A. Elseifi, S. B. Cooper, C. S. Hughes, J.W. Button, and E. Dukatz Comparing the Volumetric and Mechanical Properties of Laboratory and Field Specimens of Asphalt Concrete. NCHRP Report. 2016.
12. Diefenderfer, S. D., J. Habbouche, and I. Boz. *Balanced Mix Design for Surface Asphalt Mixtures: 2020 Field Trials*. Publication VTRC 23-R13. Virginia Transportation Research Council, Charlottesville, 2023.
13. Diefenderfer, S. D., I. Boz, and J. Habbouche. *Balanced Mix Design for Surface Asphalt Mixtures: 2019 Field Trials.* Publication VTRC 21-R21. Virginia Transportation Research Council, Charlottesville, 2021.
14. Nair, H., S. D. Diefenderfer, and B.F. Bowers. *Assessing Increased Use of Reclaimed Asphalt Pavements in Asphalt Mixtures.* Publication VTRC 20-R9. Virginia Transpiration Research Council, Charlottesville, 2019.
15. Bowers, B. F., Diefenderfer, S. D., Moore, N. and Lynn, T. (2022) Balanced Mix Design and Benchmarking: A Case Study in Establishing Performance Test Thresholds. Transportation Research Record: Journal of the Transportation Research Board, 2676 (12): 586–598.
16. Habbouche, J., I. Boz, and S. D. Diefenderfer. Validation of Performance-Based Specifications for Surface Asphalt Mixtures in Virginia. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2676 (5): 277 – 296.
17. Tong, B., J. Habbouche, S. D. Diefenderfer, and G.W. Flintsch. Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia. *International Journal of Pavement Engineering*, 2024.
18. Tong, B., J. Habbouche, B. K. Diefenderfer, and G. W. Flintsch. Rutting Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents under Full-Scale Accelerated Testing. Materials, 16, 7611. https://doi.org/10.3390/ma16247611.
19. Cox, B. C., B. T. Smith, I. L. Howard, and R. S. James. State of Knowledge for Cantabro Testing of Dense Graded Asphalt. *Journal of Materials in Civil Engineering*, Vol. 29, No. 10, 2017.
20. ASTM International. *ASTM D 8225-19: Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at* Intermediate Temperature. West Conshohocken, PA, 2019.

21. Boz, I., J. Habbouche, and S. Diefenderfer. The Use of the Indirect Tensile Test to Evaluate the Resistance of Asphalt Mixtures to Cracking and Moisture-Induced Damage. In *Airfield and Highway Pavements 2021* (pp. 104-114), 2021.
22. Braver, S.L., D.P. Mackinnon, and M.C. Page. Levine's Guide to Spss for Analysis of Variance (2003).
23. Ostertagová, E., O. Ostertag, and J. Kováč. Methodology and Application of the Kruskal-Wallis Test. Applied Mechanics and Materials, 611, 115 – 120, 2014.
24. Virginia Department of Transportation. Road and Bridge Specifications. Richmond, 2020.
25. Diefenderfer, S.D., and B.F. Bowers. Initial Approach to Performance (Balanced) Mix Design: The Virginia Experience. Transportation Research Record, 2673, 335 – 345, 2019.
26. Habbouche, J., I. Boz, and S. D. Diefenderfer. Interlaboratory Study for the Indirect Tensile Cracking Test at Intermediate Temperature: Phase II. VTRC Report 23-R3. Virginia Transportation Research Council, Charlottesville, 2022.
27. American Association of State Highway and Transportation Officials. *Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards.* Washington, DC, 2017.
28. Boz, I., S. D. Diefenderfer., and J. Habbouche. *Development of a Performance Threshold for CT Index from Asphalt Mixtures Tested During Production (Non-Reheat)*. Presented at the 11th Balanced Mix Design Technical Committee Meeting held at the Virginia Transportation Research Council on September 28, 2022, Charlottesville, 2023.
29. Bowers, B. F., T. Lynn., F. Yin., N. Moore., S.D. Diefenderfer., I. Boz. Impact of Production Variability on Balanced Mix Designs in Virginia. FHWA/VTRC 23-R20. 2023. URL : https://rosap.ntl.bts.gov/view/dot/67963.

APPENDIX

Table A-1. Normality (Shapiro-Wilk Test) for Performance Tests of Mixtures

Mix ID	P-value of Sampling for CML			P-value of Sampling for IDT-CT			P-value of Sampling for APA		
	Design	Non-reheats	Reheats	Design	Non-reheats	Reheats	Design	Non-reheats	Reheats
30_C	0.215	0.516	0.555	0.578	0.604	0.023	-	0.031	0.147
30_O	0.975	0.689	0.313	0.228	0.543	0.006	-	0.519	0.016
45_HR	0.765	0.424	0.643	0.201	0.768	0.045	-	0.602	0.007
45_HR_RA	0.810	0.518	0.995	0.780	0.068	0.004	-	0.565	0.129
45_HR_L	0.148	0.298	0.631	0.671	0.168	0.381	-	0.276	0.791
60_HR_L_RA	<0.01	0.927	0.487	0.611	0.383	0.318	-	0.681	0.056

CML = Cantabro mass loss; IDT = indirect tensile cracking test; CT = cracking tolerance; APA = asphalt pavement analyzer; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

Table A-2. Levene's Test of Equality of Error Variances

Mix ID	P-value for CML	P-value for IDT-CT	P-value for APA
30_C	0.700	0.04	0.495
30_O	0.114	<0.001	0.769
45_HR	0.926	0.011	0.586
45_HR_RA	0.155	0.036	0.662
45_HR_L	0.759	<0.01	0.447
60_HR_L_RA	0.325	<0.01	0.847

CML = Cantabro mass loss; IDT = indirect tensile cracking test; CT = cracking tolerance; APA = asphalt pavement analyzer; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

Table A-3-1. ANOVA with Post-hoc Analysis (Tukey's HSD Test)

Group-1	Group-2	P_value for 30_C	P_value for 30_O	P_value for 45_HR	P_value for 45_HR_RA	P_value for 45_HR_L	P_value for 60_HR_L_RA
		CML	CML	CML	CML	CML	CML
Design	Non-reheats	0.000	0.000	0.000	0.015	0.000	0.262
Design	Reheats	0.000	0.000	0.283	0.000	0.000	0.000
Non-reheats	Reheats	0.001	0.000	0.000	0.000	0.000	0.000

CML = Cantabro mass loss; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

Table A-3-2. Nonparametric (Kruskal-Wallis) Test Results

Group-1	Group-2	Adj. Sig. for 30_C		Adj. Sig. for 30_O		Adj. Sig. for 45_HR		Adj. Sig. for 45_HR_RA		Adj. Sig. for 45_HR_L		Adj. Sig. for 60_HR_L_RA		
		CT Index	APA	CT Index	APA	CT Index	APA	CT Index	APA	CT Index	APA	CML	CT Index	APA
Design	Non-reheats	0.007	0.349	0.048	0.174	0.000	0.030	0.000	0.310	0.731	1.000	1.000	0.108	1.000
Design	Reheats	1.000	0.519	0.644	1.000	0.779	1.000	0.653	0.199	0.128	1.000	0.004	0.740	0.781
Non-reheats	Reheats	0.000	1.000	0.000	0.026	0.000	0.110	0.000	1.000	0.000	1.000	0.001	0.000	0.183

APA = asphalt pavement analyzer; CT = cracking tolerance; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; Adj.Sig.= significance values have been adjusted by the Bonferroni correction for multiple tests. CML of 60% RAP mixture does not meet the assumptions of normality, as also presented in this table.

Table A-4-1. ANOVA with Post-hoc Analysis (Tukey's HSD Test) Results for Produced Samples

Group-1	Group-2	P_value for 30_C		P_value for 30_O		P_value for 45_HR		P_value for 45_HR_RA		P_value for 45_HR_L		P_value for 60_HR_L_RA	
		CML	APA	CML	APA	CML	APA	CML	APA	CML	APA	CML	APA
NR_1	NR_2	1.000	0.852	0.985	0.869	0.944	0.194	1.000	0.947	0.477	1.000	1.000	0.976
NR_1	NR_3	0.991	0.950	0.905	0.904	0.994	0.466	0.914	0.982	0.814	0.667	0.895	0.999
NR_1	NR_4	1.000	0.654	0.052	0.166	0.797	0.958	0.407	1.000	0.866	1.000	0.261	0.828
NR_2	NR_3	0.963	1.000	1.000	1.000	1.000	0.974	0.772	1.000	0.999	0.857	0.672	0.800
NR_2	NR_4	1.000	1.000	0.285	0.757	0.182	0.685	0.595	0.843	0.998	1.000	0.124	0.336
NR_3	NR_4	0.983	0.997	0.481	0.708	0.331	0.976	0.043	0.919	1.000	0.808	0.945	0.984
R_1	R_2	0.785	0.920	1.000	0.991	0.998	1.000	1.000	1.000	0.321	1.000	0.935	1.000
R_1	R_3	0.933	0.622	0.994	0.978	0.468	0.956	0.173	1.000	0.742	1.000	0.141	0.996
R_1	R_4	1.000	0.996	1.000	1.000	1.000	0.994	0.927	0.971	0.333	0.968	0.996	1.000
R_2	R_3	0.162	0.998	0.999	1.000	0.871	0.885	0.097	0.997	0.997	1.000	0.012	0.916
R_2	R_4	0.801	1.000	1.000	0.994	0.917	1.000	0.796	0.896	1.000	0.973	0.543	1.000
R_3	R_4	0.924	0.945	0.980	0.984	0.202	0.648	0.814	0.999	0.998	0.947	0.466	0.996

CML = Cantabro mass loss; APA = asphalt pavement analyzer; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

Table A-4-2. Nonparametric (Kruskal-Wallis) Test with Pairwise Comparisons for Produced Samples

Group-1	Group-2	Adj. Sig. for 30_3		Adj. Sig. for 30_O		Adj. Sig. for 45_HR		Adj. Sig. for 45_HR_RA	Adj. Sig. for 45_HR_L	Adj. Sig. for 60_HR_L_RA
		CT Index	APA	CT Index	APA	CT Index	APA	CT Index	CT Index	CT Index
NR_1	NR_2	1	1	1	1	1	1	1	1	1
NR_1	NR_3	1	1	1	1	1	1	1	1	1
NR_1	NR_4	1	1	1	1	1	1	1	1	1
NR_2	NR_3	1	1	1	1	1	1	1	1	1
NR_2	NR_4	1	1	1	1	1	1	1	1	1
NR_3	NR_4	1	1	1	1	1	1	1	1	1
R_1	R_2	1	1	1	1	1	1	1	1	1
R_1	R_3	1	1	1	1	1	1	1	1	1
R_1	R_4	1	1	1	1	1	1	1	1	1
R_2	R_3	1	1	1	1	1	1	1	1	1
R_2	R_4	1	1	1	1	1	1	1	1	1
R_3	R_4	1	1	1	1	1	1	1	1	1

APA = asphalt pavement analyzer; CT = cracking tolerance; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; Adj.Sig.= significance values have been adjusted by the Bonferroni correction for multiple tests. Some data from APA do not meet the assumptions of normality, as also presented in this table.