

Evaluation of Balanced Asphalt Surface Mixtures with Conventional and High RAP Contents Using Laboratory and Accelerated Pavement Testing

Bilin Tong

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY
IN
CIVIL ENGINEERING**

Gerardo W. Flintsch, Chair

Jhony Habbouche

Brian K. Diefenderfer

Samer Katicha

Alexander S. Brand

12/6/2024

Blacksburg, Virginia

Keywords: Sustainable Pavement, Accelerated Pavement Testing, High Reclaimed Asphalt Pavement, Balanced Mix Design, Performance Testing

Copyright © 2024, Bilin Tong

Evaluation of Balanced Asphalt Surface Mixtures with Conventional and High RAP Contents Using Laboratory and Accelerated Pavement Testing

Bilin Tong

Abstract

Balanced Mix Design (BMD) represents an asphalt mixture design methodology that replaces certain traditional volumetric parameters with performance-based testing to address predominant distresses such as rutting and cracking. This approach offers an avenue to properly design and produce engineered asphalt mixtures, including those with high reclaimed asphalt pavement (HRAP) contents, recycling agents (RAs), fibers, and polymer-modified binders. Laboratory performance tests are essential to the BMD process, as they ensure the production of durable, high-performance materials. Beyond laboratory performance evaluation, accelerated pavement testing (APT) plays a crucial role in bridging the gap between laboratory material characterization and field pavement performance.

This dissertation aimed to assess the BMD concept for designing durable, long-lasting surface mixtures in Virginia, with particular emphasis on higher RAP content mixtures (HRAP mixtures, i.e., exceeding 30% RAP). The study involved laboratory and APT testing of six surface mixtures featuring a range of RAP contents (both conventional and high), two binder grades (PG 64-22 and PG58-28), one RA, and one warm mix additive.

Findings indicated that dense-graded, unmodified surface mixtures with higher RAP contents can be successfully designed using the current Virginia Department of Transportation (VDOT) BMD special provision. These mixtures can be produced in the plant with no significant deviations in aggregate gradation and asphalt binder content from the design specifications. The combined effect of variations in different volumetric properties during production may influence the primary performance of the mixtures, potentially resulting in an imbalance. As a consequence, the produced BMD mixture may fail to meet one or more performance thresholds. Additionally, the results underscored the effectiveness of BMD concept with incorporating RAs and/or a softer binder when designing HRAP surface mixtures. Importantly, the current selected BMD tests characterized the laboratory performance of mixtures, aligning with the performance observed under APT. This research provided a steppingstone towards the examination and validation of the VDOT BMD thresholds, which ensures satisfactory field performance. The study also indicated that while current BMD thresholds provided sufficient margins for satisfactory field cracking performance, rutting resistance may become a concern for overly designed BMD HRAP mixtures. For instance, mixtures with excessively high asphalt binder content may exhibit compromised rutting resistance.

Furthermore, to address the challenges uncovered during BMD test analysis—issues like the constraints of traditional pair-wise comparisons, risks of repetitive design processes, and the

difficulty in pinpointing critical factors in mixture production—this dissertation proposed innovative solutions to enhance BMD application and streamline the evaluation process. First, a novel Composite Performance Index (CPI), visualized through a 3D plot, captured the “balance” status of various mixtures. Second, a machine learning-enhanced BMD framework was introduced, offering intelligent optimization throughout the design and production phases. The integration of these two tools offers significant potential for simultaneously improving multiple performance indices of asphalt mixtures.

Finally, this research demonstrated that the performance of higher RAP content mixtures can exceed that of lower RAP content mixtures through the application of BMD approaches. This dissertation not only advanced the implementation of BMD for surface mixtures but also contributed to the sustainable and performance-driven evolution of asphalt mix design. The insights gained from this study provided practical guidance and strategic recommendations for enhancing asphalt mixture design, production, and performance monitoring.

Evaluation of Balanced Asphalt Surface Mixtures with Conventional and High RAP Contents Using Laboratory and Accelerated Pavement Testing

Bilin Tong

General Audience Abstract

This dissertation evaluated a new way to design asphalt mixtures, called Balanced Mix Design (BMD). Unlike traditional methods that focus mostly on fundamental volumetric properties of materials, BMD uses testing to ensure the pavement will hold up against common issues like cracking and rutting. This approach allows for better use of materials, including high recycled asphalt pavement (HRAP), recycling agents (RAs), fibers, and modified binders.

Considering economic and environmental benefits, the study evaluated the application of BMD for designing durable asphalt surfaces for pavements, particularly mixtures with over 30% recycled asphalt pavement (RAP). Tests in the laboratory and with accelerated pavement testing (APT) showed that high RAP mixtures could be designed and produced reliably using current Virginia Department of Transportation (VDOT) guidelines. The addition of recycling agents or softer binders further supported these mixtures in meeting performance standards, ensuring satisfied field performance.

To address key challenges, such as the limitations of traditional pair-wise comparisons, risks of repetitive design processes, and the difficulty in finding key factors for production variability, this research introduces two innovative tools to enhance BMD. First, a Composite Performance Index (CPI) with a 3D visual tool offers a way to assess the balance status of mixtures. Second, a machine learning framework was developed to optimize BMD throughout the design and production phases. Together, these tools establish BMD as a valuable approach for designing and producing more sustainable, longer-lasting pavements.

Acknowledgements

As I reach the final chapter of my Ph.D. journey, I find myself flooded with gratitude—an overwhelming urge to thank the countless people who have stood by me over the past five years. The memories are too many to count, and though words feel like an inadequate currency to pay my debts of appreciation, please give me this shot. Brace yourselves.

First, to my advisor, Dr. Gerardo W. Flintsch—the keystone to my academic bridge. Your guidance has been my lighthouse, steering me through the foggy waters of research and endless data. From lab work to field experiments, you’ve opened doors that have profoundly shaped my understanding of materials’ properties. I still remember the first day of your asphalt technology class, hoping that one day I’d find myself in your lab. Thank you for making that hope a reality and for giving me a place to grow as a researcher. Without your opportunities, encouragement, and endless patience, I wouldn’t be here today. You have my gratitude—and my deepest respect.

To Dr. Jhony Habbouche—where do I even start? our first meeting at TRB 2023, helping me outline research ideas. Five papers in two years—it’s real now, right? I still remember staring at that first draft, which you returned covered in red notes. I might have needed smelling salts at that moment, but your meticulous feedback has made me a better researcher. You pushed me to step out of my comfort zone, guided me through reviewer comments, and even introduced me to the amazing friends at conferences. Thanks to you, I’ve learned to navigate the tricky terrain of academic writing—and maybe even enjoy it. Now, I hope my paper can go to your table with “reinforcement”.

Dr. Brian K. Diefenderfer, thank you for all those bi-weekly HVS meetings. They became my compass for navigating APT projects, discussing challenges, and ensuring everything stayed on track. I’m also grateful for your support with I-64 measurements and interdepartmental coordination. Additionally, your valuable insights on rutting and cracking studies helped shape the core of this dissertation.

Dr. Samer Katicha, you’re everyone’s the “go-to” for complex math and technical challenges. I’m deeply appreciative of each time you took to discuss my struggles, opening my mind to new perspectives that I couldn’t see alone. I’m excited about the possibility of collaborating on the TSD project—with machine learning and FE simulation.

To Dr. Alexander S. Brand, thank you for the chance to work with your student, Thien Tran. That collaboration broadened my understanding of the intersection between environmental considerations and traditional performance tests—I never expected to find so interesting. Your feedback on my dissertation helped shape my research and will stay with me as I continue in this field.

To Ernesto Urbaez Perez, collaborating with you has taught me the importance of systematic project management, research planning, and organization. Your depth of knowledge and passion for our field have been an inspiring reminder of what it means to be truly dedicated to one’s work.

I also owe thanks to many others who have shared in this journey. Dr. Stacey D. Diefenderfer and Dr. Ilker Boz, your expertise and constant support have been instrumental to the research. To Dr. Eugene Amarh, thank you for sharing your thoughts on project development, proposal writing, and career growth. My gratitude goes to Dr. Fabrizio Meroni for patiently teaching me the nuances of testing, Dr. Wenjing Xue for guiding me through HVS data processing, and to Billy Hobbs, Kenny Smith, Max Ratcliffe, David Songer, and Richard Mejia for their ongoing support. Billy, I can almost hear you now, “Finally, he’s done!” Special thanks to Carolina Benavides Ruiz for the countless I-64 measurements we did together and to Ahmed Nisar and Andrew Carella for their assistance in the lab. My appreciation extends to all the CSRI group members for the shared academic and cultural exchanges that have enriched my experience. Dr. Lingchen Kong, thank you for nudging me toward Virginia Tech. To Dr. Wenjiang Huang, I am grateful for the many experiences we shared during our Ph.D. journey, and for all your insights into life, the gym, the stock market, and research, And to Dr. Maryam Shakiba, my master’s advisor—you laid the foundation for my academic path, and for that, I am forever grateful.

To my family, words fall short. To my father, Yaole Tong, and my mother, Siyu Hou, I owe you everything. It’s been seven years since I last came home, and I’m so sorry for that. Thank you for supporting me without question, with unending patience and love. To my grandmother, Hengju Chen—your teachings on kindness and optimism have been the compass of my life. To my closest friend, my brother Junchen Ye—thank you for being there every time I felt overwhelmed by coding, and for offering support when I needed it most. Your friendship has been my lifeline in countless moments.

Finally, to my girlfriend, Jou-yin Irene Chen. You have been there for me through high and low, from the first paper of this journey to the very last sentence of this dissertation. You’ve supported me in moments of stress, and your encouragement has given me strength I didn’t know I had. In your presence, I’ve found balance—not just between work and life, but in my outlook on the world. I am constantly broadening my horizons, striving to become a better version of myself—changes that happen subtly, even if I don’t always realize it in the moment. Most importantly, through you, I’ve learned humility, understanding, and the value of true communication—qualities I’ll carry into every endeavor. And, of course, we’ll never forget Taco and Nacho, our two adopted kittens, who have brought so much chaos and joy to our home.

To those whose support can’t quite be captured with words—know that your impact is felt deeply, and your kindness has been an integral part of this journey. Thank you all for walking this path with me.

Table of Contents

CHAPTER 1 - Introduction.....	1
1.1. Background	1
1.2. Problem Statement	2
1.3. Research Objective and Scope	2
1.4. Significance.....	3
1.4.1. Intellectual Merit	3
1.4.2. Broader Impacts.....	3
1.5. Dissertation Organization.....	4
1.6. Reference.....	7
CHAPTER 2 - Literature Review.....	9
2.1. RAP Usage	9
2.2. RAP Mixture Design.....	10
2.3. The Definition of BMD.....	11
2.4. State BMD Efforts.....	12
2.5. Virginia BMD Efforts	15
2.6. Laboratory Performance Tests	16
2.6.1. Rutting Tests.....	16
2.6.2. Cracking Tests	19
2.6.3. Multi-level Performance Evaluation	21
2.7. Accelerated Pavement Testing.....	22
2.8. Machine Learning	24
2.8.1. Support Vector Machines	24
2.8.2. Artificial Neural Network (ANN)	25
2.8.3. Random Forest.....	26
2.8.4. Extreme Gradient Boosting	26
2.9. Summary of Literature Review	27
2.10. Reference.....	27
CHAPTER 3 - Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia.....	32
3.1. Abstract	32
3.2. Objective and Scope.....	33
3.3. Experimental Program.....	33

3.4. Durability Assessment	36
3.5. Cracking Performance Assessment.....	38
3.5.1. IDT Cracking Test and Results	38
3.5.2. Dynamic Modulus $ E^* $ Test and Results	39
3.5.3. Direct Tension Fatigue Test and Results	42
3.6. Rutting Performance Assessment.....	45
3.6.1. APA Rut Test and Results	45
3.6.2. Stress Sweep Rutting Test and Results.....	46
3.7. Correlations and Statistical Analyses	47
3.7.1. Performance Ranking	47
3.7.2. Correlation	48
3.7.3. Three-dimensional (3-D) Plot.....	50
3.7.4. Ternary Plots.....	54
3.8. Summary of Findings, Conclusions, and Ongoing Work	56
3.9. Reference.....	57
CHAPTER 4 - Assessment of Production Variability and Composite Performance Index for Conventional and High RAP BMD Mixtures	61
4.1. Abstract	61
4.2. Introduction	62
4.3. Objective and Scope.....	64
4.4. Methodology	65
4.4.1. Experimental Program.....	65
4.4.2. Mixture Designs	66
4.4.3. Production Sampling	66
4.4.4. BMD Tests.....	67
4.4.5. Statistical Analyses.....	67
4.5. Results	68
4.5.1. Mixture Volumetric Properties and Gradations.....	68
4.5.2. Durability Assessment.....	71
4.5.3. Cracking Performance Assessment	73
4.5.4. Rutting Performance Assessment.....	74
4.6. Additional Discussion and Analyses.....	76
4.6.1. Assessing the Effect of Reheating	76
4.6.2. Quantifying the Degree of Balance for the Produced Mixture.....	77
4.7. Summary of Findings and Future Work.....	81

4.8. Reference.....	82
CHAPTER 5 - Machine Learning-Based Prediction and Optimization of Balanced Mixture	
Design Performance Indices	85
5.1. Abstract	85
5.2. Introduction	86
5.3. Objective and Scope.....	87
5.4. Background and Literature Review.....	88
5.4.1. Production Variability	88
5.4.2. Machine Learning Approaches.....	90
5.5. Methodology	91
5.5.1. Experiments	91
5.5.2. BMD Testing	91
5.5.3. Production Sampling and Specimen Fabrication.....	92
5.5.4. Modeling Approach.....	92
5.5.5. Training Features, Data Processing, and Evaluation Metrics.....	93
5.5.6. Simulating the BMD Deployment.....	95
5.6. Results and Discussion.....	95
5.6.1 Feature Selection	95
5.6.2 Prediction Performance of ML Models	96
5.6.3 Feature Importance Analysis	100
5.6.4 Pseudo In-situ Deployment	101
5.6.5. Discussion.....	103
5.7. Conclusion.....	104
5.8. Reference.....	105
CHAPTER 6 - Rutting Performance Evaluation of BMD Surface Mixtures with Conventional	
and High RAP Contents Under Full-Scale Accelerated Testing	108
6.1. Abstract	108
6.2. Introduction	109
6.2.1. Rutting Performance Evaluation	109
6.2.2. Background on Virginia APT.....	110
6.3. Objective and Scope.....	110
6.4. APT Setup for 2020 BMD Experiment.....	110
6.5. Experimental Program.....	111
6.5.1. Materials Evaluated	111
6.5.2. Design of APT Test Cells.....	112

6.5.3. Loading Protocols.....	112
6.5.4. Experiment Timeline	113
6.5.5. Rut Depth Measurement.....	114
6.6. APT Results and Discussions.....	115
6.6.1. Environmental Aging Effect.....	116
6.6.2. Normalization	117
6.7. Correlation Between APT Measurements and Laboratory Test Results.....	121
6.7.1. APA Rut Test.....	121
6.7.2. SSR Test	122
6.8. Summary of Findings, Conclusions, and Future Work.....	125
6.9. Reference.....	125
CHAPTER 7 - Cracking Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: Insights from Accelerated Testing Program.....	128
7.1. Abstract	128
7.2. Objective and Scope.....	129
7.3. Experimental Design and Research Approach.....	129
7.3.1. Evaluated Materials	129
7.3.2. Virginia APT Facility	129
7.3.3. Test Section and Instrumentation	130
7.3.4. Loading Protocols.....	131
7.3.5. Data Collection	132
7.4. Results and Discussion.....	138
7.4.1. Longitudinal Strain Monitoring.....	138
7.4.2. Cracking Evaluation	143
7.5. Examining VDOT BMD Thresholds	146
7.5.1. Indirect Tensile Cracking Test (IDT-CT).....	146
7.5.2. Cantabro Test.....	147
7.5.3. Laboratory Test Comparisons and Correlations.....	147
7.6. Correlation with APT Rutting Studies	149
7.7. Summary of Findings, Conclusions, and Future Work.....	150
7.8. Reference.....	151
CHAPTER 8 - Summary, Conclusions, and Recommendations.....	153
8.1. Overview	153
8.2. Summary of Primary Findings	153
8.2.1. BMD Impact on Volumetric Properties and Gradations	153

8.2.2. Laboratory Evaluation	154
8.2.3. Machine Learning Application in BMD.....	154
8.2.4. APT Experiment	155
8.3. Conclusions	156
8.4. Significance.....	157
8.5. Recommendations	157
8.5.1. Implementation Recommendations	157
8.5.2. Recommendations for Future Research.....	158
Appendix: Chapter 3 – Statistic Analysis	159

List of Figures

Figure 1-1. Conceptual flowchart of each paper's relevance within the APT program data analysis process.....	5
Figure 2-1. Implementation Status of BMD Concept in the United States.	13
Figure 2-2. Loading in the flow number test	17
Figure 2-3. FN testing plot.....	17
Figure 2-4. Asphalt pavement analyzer.	18
Figure 2-5. Multi-level laboratory performance tests.	22
Figure 2-6. Virginia APT facility and field construction at VTTI.....	24
Figure 2-7: Separate 2-D data in 3-D dimension virtually	25
Figure 2-8. Example of three-layer ANN architecture	25
Figure 2-9. Random forest structure.	26
Figure 3-1. Performance Test Data for Cantabro Mass Loss. The box plot displays the five-number summary of a set of data... ..	37
Figure 3-2. Performance Test Data for IDT-CT Index: (a) Including Mixture 45_HR and (b) Excluding Mixture 45_HR.....	39
Figure 3-3. Dynamic Modulus Test Results: (a) $ E^* $ Master Curves in Log-log Scale and (b) δ Master Curves in Semi-log Scale.....	40
Figure 3-4. Glover-Rowe (G-Rm) Parameter at 20°C and 5 Hz	42
Figure 3-5. Performance Test Data for Direct Tension Cyclic Fatigue: (a) Damage Characteristic Curves and (b) DR Parameter.	44
Figure 3-6. <i>Sapp</i> Values.....	45
Figure 3-7. Performance Test Data for APA Rut Depth at 64°C and 8,000 Loading Cycles.	46
Figure 3-8. Performance Test Data for Stress Sweep Rutting Test..	47
Figure 3-9. (a) 3-D Plots of Performance Indices: CML vs. RSI vs. <i>Sapp</i> ; (b) Ternary Plots of Performance Indices: CML vs. RSI vs. <i>Sapp</i>	51
Figure 3-10. (a) 3-D Plots of Performance Indices CML vs. APA vs. CT index; (b) Ternary Plots of Performance Indices: CML vs. APA vs. CT index..	52
Figure 3-11. (a) 3-D Plots of Performance Indices: CML vs. APA vs. CT index Excluding Mixture 45_HR; (b) Ternary Plots of Performance Indices: CML vs. APA vs. CT index Excluding Mixture 45_HR.....	52
Figure 4-1. Experimental Plan for Laboratory- and Plant-Produced Mixtures.	66
Figure 4-2. Design Gradations for the Six Mixtures.....	70
Figure 4-3. 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.....	70
Figure 4-4. 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.....	72

Figure 4-5. 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.	74
Figure 4-6. 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.	75
Figure 4-7. Relationship Between Non-reheats and Reheats: (a) Cantabro Test; (b) IDT-CT Test; (c) APA Rutting Test.	76
Figure 4-8. 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.	78
Figure 4-9. 3-D Plots: (a) Non-reheated Specimens; (b) Reheated Specimens.	81
Figure 5-1. VDOT BMD Tests: (a) Cantabro Mass Loss Test; (b) IDT-CT; and (c) APA Rut Test. VDOT = Virginia Department of Transportation;	91
Figure 5-2. Experimental and Machine Learning (ML) Training Workflows. Left block: production sampling and specimen preparation. Middle block: balanced mix design (BMD) tests. Right block: ML models used in predicting three primary BMD tests.	92
Figure 5-3. Correlation Matrix Derived from Pearson Correlation Coefficients (r) for Selected Features Associated with the IDT-CT Index. The lower triangle displays linear regression plots, while the upper triangle shows r values between features. The diagonal plots illustrate the distribution of each feature or the target property (top left).	96
Figure 5-4. Prediction vs. Actual Cantabro mass loss (CML) using (a) LR, (b) RF, (c) XGB, and (d) SVR.	98
Figure 5-5. Prediction vs. Actual CT index using (a) LR, (b) RF, (c) XGB, and (d) SVR.	98
Figure 5-6. Prediction vs. Actual APA rut depth using (a) LR, (b) RF, (c) XGB, and (d) SVR.	99
Figure 5-7. Variability in ML model performances (R ² and accuracy) derived from 30 times random data splitting with different shuffle seeds.	99
Figure 5-8. Feature importance from the permutation importance method for: (a) CML from XGB, (b) IDT-CT index from RF, and (c) APA RD from RF model.	101
Figure 5-9. The schematic of pseudo in-situ deployment flowchart.	102
Figure 5-10. Uniform Manifold Approximation and Projection (UMAP) reduces the feature space to two dimensions for (a) CML, (b) CT index, and (c) APA RD datasets. The number of neighbors used in UMAP is 10, and it is asserted that the choice of hyperparameters has a negligible impact on the conclusions drawn.	103
Figure 5-11. Explore the Use of ML Models to Support the Implementation of BMD.	104
Figure 6-1. Photographs taken of: (a) Dynatest Mark VI Heavy Vehicle Simulator; (b) Laser Profiler Mounted on HVS Carriage.	111
Figure 6-2. (a) Site Layout for the 2020 APT project; (b) Pavement Structure.	112
Figure 6-3. Loading Protocols for Various Test Cells.	114

Figure 6-4. An Example of Measurement in the Experiment for Cell L1B on October 11, 2022: (a) 3-D Laser Profile; (b) Rut Depth and SVPD Illustration in a Specific Transverse Plane.....	115
Figure 6-5. Rutting Depth for 12 Rutting Test Cells. ESALs = equivalent single axle loads. ...	116
Figure 6-6. Fitting Result for Testing Lanes of (a) Mixture 30_C; (b) Mixture 45_HR; (c) Mixture 45_HR_RA; (d) Mixture 45_HR_L.....	119
Figure 6-7. Fitting Result for Testing Cells: (a) Cell 30_O_R1; (b) Cell 30_O_R2; (c) Cell 60_HR_L_RA_R1; (d) Cell 60_HR_L_RA_R1.....	120
Figure 6-8. Normalized AASHTO Rut Depth Versus ESALs..	121
Figure 6-9. Rutting Testing Results Plot: (a) APA vs. RSI; (b); APA vs. HVS (c) RSI vs. HVS; (d) Correlation Analyses.	123
Figure 7-1. Photographs taken of: (a) Virginia APT facility; (b) Laser profiler mounted on HVS carriage.....	130
Figure 7-2. (a) Site layout for the 2020 APT project; (b) Test section instrumentation typical layout. APT = accelerated pavement testing.....	131
Figure 7-3. (a) Installing longitudinal strain gauge KM-100 HAS; (b) Strain gauge in construction.....	131
Figure 7-4. Typical response of mixture 45_HR_RA during initial testing (a) Strain response under one loading pass for three strain gauges; (b) Strain gauge 1 example.....	134
Figure 7-5. Typical 4-min data collection window for mixture 30_O at phase I: Stable reference table at initial stage of testing..	135
Figure 7-6. Typical 4-min data collection window for mixture 30_O at (a) phase II: Fluctuated reference table occurs; (b) phase III: Increased reference table occurs..	135
Figure 7-7. Longitudinal strain response of mixture 30_C: (a) failed to extract tensile and compressive strain gauge 1 @ 495,505 ESALs; (b) crack initiation point at gauge 2 @ 932,282 ESALs.	136
Figure 7-8. Examples of cracking sections after testing with cracks marked for mixtures: (a) 30_C; (b) 45_HR_RA; (c) 60_HR_L_RA.....	137
Figure 7-9. An example of the location of field cores within the test section in Lane 1.	138
Figure 7-10. (a) Permeability test; (b) A portion of field cores for DMT; (c) Small-scale specimen DMT. DMT = dynamic modulus test.	138
Figure 7-11. Longitudinal strain in the center line monitored along with testing for different mixtures.....	139
Figure 7-12. Critical residual strain monitored along with testing for different mixtures.....	141
Figure 7-13. Deformation uniformity for (a) two-signal sections (b) three-signal sections.	142
Figure 7-14. Field survey on cracking length..	144
Figure 7-15. Permeability test results.	145
Figure 7-16. Dynamic modulus test results at 20°C and 10Hz..	146
Figure 7-17. Correlation between CT index and APT crack lengths (a) for all evaluated mixture; (b) for mixtures excluding mixture 45_HR and mixture 60_HR_L_RA.....	148

Figure 7-18. Correlation between Cantabro mass loss and APT crack lengths (a) for all evaluated mixture; (b) for mixtures excluding mixture 45_HR and mixture 60_HR_L_RA.	149
Figure 7-19. Correlation between APT crack lengths and APT rut depth.....	150

List of Tables

Table 3-1. Volumetric Properties and Gradations for APT Mixtures - Design	34
Table 3-2. Volumetric Properties and Gradations for APT Mixtures During Production.....	36
Table 3-3. Performance Test Results and Corresponding Statistical Parameters of Evaluated Mixtures Collected During Production	37
Table 3-4. Ranking of Mixtures based on Various Evaluated Performance Indices	48
Table 3-5. Correlation Matrix for All Evaluated Performance Indices	49
Table 3-6. Correlation Matrix for All Evaluated Performance Indices Excluding CT index for Mixture 45_HR	50
Table 3-7. Normalize Advanced Performance Tests (CML vs. RSI vs. <i>Sapp</i>) for 3-D Plot and Ternary Plot	55
Table 4-1. Virginia Performance Testing Criteria	67
Table 4-2. Summary of Primary Volumetric Properties Determined on Mixtures During Production	69
Table 4-3. Summary of Descriptive Statistics for Performance Tests of Plant-Produced Mixtures	72
Table 4-4. Summary of Normalized Performance indices for 3-D Plots.....	80
Table 5-1. Data Sets/Specimens Used in ML for Three Balanced Mix Design Tests.....	94
Table 5-2. Mean and standard deviation in <i>R</i> ² and accuracy from three BMD tests in Figure 5-7.	99
Table 6-1. Summary of Rutting Testing Cells and Timeline Within the APT BMD Experiment.	113
Table 6-2. Rutting Testing Results	122
Table 7-1. Summary of cracking performance measures with rankings for the evaluated mixtures.	141
Table 7-2. Summary of identified surface crack lengths at different severity per mixture	143
Table 7-3. Summary and ranking of different tests for the evaluated mixtures.	147
Appendix	
Table A- 1. Normality (Shapiro-Wilk Test) for Performance Tests of Mixtures	159
Table A- 2. Levene's Test of Equality of Error Variances.....	159
Table A- 3. ANOVA with Post-hoc Analysis (Tukey's HSD Test).....	159
Table A- 4. Nonparametric (Kruskal-Wallis) Test Results	159
Table A- 5. ANOVA with Post-hoc Analysis (Tukey's HSD Test) Results for Produced Samples	160
Table A- 6. Nonparametric (Kruskal-Wallis) Test with Pairwise Comparisons for Produced Samples	160

List of Acronyms

3-D: Three-Dimensional
AASHTO: American Association of State Highway and Transportation Officials
AC: Asphalt Content
ALF: Accelerated Loading Facility
AMPT: Asphalt Mixture Performance Tester
ANN: Artificial Neural Network
APA: Asphalt Pavement Analyzer
APT: Accelerated Pavement Testing
ASTM: American Society for Testing and Materials
BBF: Bending Beam Fatigue
BMD: Balanced Mix Design
BMs: Base Mixtures
C: The pseudo stiffness
CAYD: Charge As You Drive
CCPR: Cold Central Plant Recycling
CF: Cyclic Fatigue
CML: Cantabro Mass Loss
COV: Coefficient of Variance
CPI: Composite Performance Index
CPI_R: Revised Composite Performance Index
CT index: Cracking Tolerance Index
DAS: Data Acquisition System
DCT: Disk-Shaped Compact Tension
DMR: Dynamic Modulus Ratio
ESALs: Equivalent Single Axle Loads
FA: Fine Aggregate to Asphalt Ratio
F_{be}: Effective Asphalt Film Thickness
FDOT: Florida Department of Transportation
FHWA: Federal Highway Administration's
FN: Flow Number
G*: Dynamic Shear Modulus
G_{mb}: Bulk Specific Gravity
G_{mm}: Theoretical Maximum Specific Gravity
G-Rm: Glover-Rowe
G_{sb}: Bulk Aggregate Specific Gravity
G_{se}: Effective Aggregate Specific Gravity
HLWT: Hamburg Loaded Wheel Test

HMA: Hot Mix Asphalt
HRAP: High Reclaimed Asphalt Pavement
HVS: Heavy Vehicle Simulator
HWTT: Hamburg Wheel-Track Test
IDOT: Illinois Department of Transportation
IDT-CT: Indirect Tensile Cracking Test
I-FIT: Illinois Flexibility Index Test
IRI: International Roughness Index
JMF: Job Mix Formula
LR: Linear Regression
LTPP: Long-Term Pavement Performance
LTRC: Louisiana Transportation Research Center
LVDT: Linear Variable Displacement Transducer
LWFT: Loaded Wheel Fatigue Test
LWT: Loaded Wheel Tracking
MASE: Mean Absolute Percent Error
ME: Mechanistic-Empirical
ML: Machine Learning
NAPA: National Asphalt Pavement Association
NCAT: National Center for Asphalt Technology
NCHRP: National Cooperative Highway Research Program
 N_f : The number of cycles to failure
NJDOT: New Jersey Department of Transportation
OBC: Optimum Binder Content
OT: Overlay Test
 P_{ba} : Percent Binder Absorbed
 P_{be} : Effective Binder Content
PG: Performance Grade
QA: Quality Assurance
QC: Quality Control
RAP: Reclaimed Asphalt Pavement
RAs: Recycling Agents
RD: Rut Depth
RF: Random Forest
RSI: Rutting Strain Index
RSST: Repeated Simple Shear Test
SCB: Semi-Circular Bend
SHAs: State Highway Administrations
SMs: Surface Mixtures

SSR: Stress Sweep Rutting
S-VECD: Simplified Viscoelastic Continuum Damage
SVM: Support Vector Machines
SVPD: Surface Vertical Permanent Deformation
SVPD: Surface Vertical Permanent Deformation
SVR: Support Vector Regression
TSS: Triaxial Stress Sweep
TxDOT: Texas Department of Transportation
U.S.: United States
VDOT: Virginia Department of Transportation
VFA: Voids Filled with Asphalt
VMA: Voids in Mineral Aggregate
VTM: Voids In Total Mixture
VTRC: Virginia Transportation Research Council
VTTI: Virginia Tech Transportation Institute
WMA: Warm Mix Asphalt
XGB: Extreme Gradient Boosting

1.1. Background

In the United States (U.S.), the use of reclaimed asphalt pavement (RAP) in asphalt mixtures has yielded significant environmental and economic benefits, prompting state agencies to implement special provisions and specifications allowing for higher RAP contents (Zaumanis et al., 2013; Martin et al., 2020). The escalated use of RAP was anticipated to counterbalance the steadily rising costs of asphalt binders and fuel required for their transportation and placement (Zaumanis et al., 2013; Martin et al., 2020). In the U.S., the extent of available RAP in each state, its corresponding performance characteristics, diverse practices, standards in practice, and other contextual factors have driven the desire to characterize high reclaimed asphalt pavement (HRAP) asphalt mixtures as state-specific entities.

The principal concern associated with HRAP asphalt mixtures revolves around their potential for excessive stiffness, a consequence of the aged binder in the RAP, rendering them more susceptible to brittleness and premature cracking (Martin et al., 2020). The utilization of relatively HRAP proportions in mixtures poses an array of construction and performance challenges, encompassing difficulties related to compaction and workability in cool weather, susceptibility to thermal cracking, fatigue cracking, reflective cracking arising from repeated loading and daily/seasonal thermal stresses, as well as raveling accompanied by aging or moisture-induced damage (Martin et al., 2020). These challenges may be effectively addressed by adopting performance-based mix design strategies, such as the Balanced Mix Design (BMD) approach. BMD is defined as 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”* (AASHTO PP 105; West et al., 2018). Unlike conventional volumetric design, the BMD framework shifts emphasis to performance testing criteria, assessing mixtures for primary distresses, such as rutting and cracking. Approval of a mix is contingent upon meeting predefined performance benchmarks or thresholds, with potential additional testing during production for quality assurance (Habbouche et al., 2022; Diefenderfer et al., 2023). Within the BMD approach, specific techniques to improve mixture performance may include increasing binder content, incorporating recycling agents (RAs) and/or softer binders (West et al., 2018; Martin et al., 2020), and utilizing polymer-modified binders to improve the engineering properties of asphalt mixtures. The BMD approach has gained prominence in recent years, and in particular it has allowed the incorporation of HRAP contents in asphalt mixtures.

Within the BMD framework, laboratory performance testing serves a pivotal role to evaluate and screen mixtures, thereby mitigating the cracking-related and rutting-related risk in specific paving applications and design specifications (Meroni et al., 2021). In tandem with laboratory performance tests, the implementation of Accelerated Pavement Testing (APT) serves as a tool to bridge the gap between design models developed via laboratory material characterization and the long-term pavement performance observed in the actual field (Harvey and

Popescu, 2000; Steyn, 2012). In one of the recent experiments, the Virginia APT program (Tong et al., 2023) has provided a comprehensive and systematic opportunity to assess the application of the BMD concept in designing durable and long-lasting SMs in Virginia. This dissertation assessed the application of the BMD concept in designing HRAP SMs (RAP contents exceeding 30%) in Virginia, offering optimized methodologies and insights.

1.2. Problem Statement

The BMD approach has gained significant attention in recent years, shifting focus from traditional volumetric criteria to performance-based testing that evaluates mixtures for susceptibility to prevalent distresses, such as rutting and cracking. This approach not only opens the door for integrating innovation into mix designs, but also demonstrates its capacity to ensure satisfactory mixture performance. Such an approach is increasingly critical, as contemporary asphalt materials differ markedly from earlier generations due to heightened supply challenges and the growing demand for high-performance materials (Golalipour et al., 2021). Additionally, the environmental and economic advantages have encouraged transportation agencies to adopt HRAP in SMs, with BMD showing promising potential to address challenges associated with HRAP usage effectively. Despite these advancements, achieving a sustainable, durable, and cost-effective mix design requires further investigation. Key areas for exploration include evaluating representative laboratory performance tests and specimens to effectively screen and ensure the target mixture performance; subsequently examining the impact of production processes and associated variability on selected tests when using the BMD concept in HRAP SM design. Furthermore, investigating the relationships across multi-level laboratory tests is also essential to deepen understanding of the balance among the primary performance characteristics of mixtures and identify potential areas for optimization within the design process. Beyond the laboratory investigation, it is equally critical to understand long-term field performance, conduct in-depth data analysis, and develop correlations between laboratory performance tests and actual pavement performance. Addressing these aspects requires comprehensive analysis and exploration.

1.3. Research Objective and Scope

The primary objective of this dissertation was to assess the application of the BMD concept to design durable surface mixtures in Virginia, with a focus on relatively higher RAP contents (RAP contents exceeding 30%). This study:

1. Conducted multi-level laboratory performance tests on one control and five BMD SMs incorporating a range of RAP contents (conventional and high), two binder performance grades (PGs), one RA, and one warm mix asphalt (WMA) additive. The interconnections and distinctions among various laboratory tests were explored.
2. Assessed and examined the developed performance-based specifications (BMD special provisions) in Virginia in terms of design mechanism and selected tests. This evaluation was based on a detailed laboratory study comparing as-designed mixtures and as-produced mixtures.

3. Proposed optimization methods to address the challenges associated with the current BMD evaluation process.
4. Developed an APT experiment, to verify the mix design performance predicted by the laboratory tests. This included documenting paving operations and construction practices, conducting full-scale accelerated pavement testing on six surface mixtures (SMs), and performing comprehensive analyses of pavement responses and the collected data.
5. Documented observations and lessons learned.

1.4. Significance

This dissertation explored and evaluated the potential for increasing RAP content in SMs, driven by a growing interest in maximizing RAP use and building on Virginia's extensive experience with 100% RAP utilization in base mixtures (BMs). This study represented an application of VDOT's BMD special provisions, enabling a comprehensive evaluation of SM performance with 30%, 45%, and even 60% RAP. This study also generated insights and best practices for long-term pavement performance monitoring. Additionally, this study provided an opportunity to investigate the impact of BMD—along with the use of softer binders and recycling agents—on rutting and cracking performance. Ultimately, by bridging the gap between APT and laboratory evaluations, this research examined current VDOT test thresholds, advancing the implementation of BMD design and performance assessment in Virginia.

1.4.1. Intellectual Merit

This dissertation offers significant intellectual contributions through a comprehensive investigation of the balancing mechanisms in asphalt concrete design, particularly examining the trade-off between rutting and cracking performance. As durability becomes a priority for transportation agencies seeking longer-lasting pavements, this research provides valuable insights into the interplay among key performance characteristics, utilizing multi-level laboratory performance tests and APT. This dissertation introduces a Composite Performance Index (CPI) to evaluate the balance among three core indices, reducing the redundancy and complexity inherent in traditional comparative methods.

To address the challenges associated with applying the BMD concept throughout the design and production phases, this dissertation also presents optimization solutions for the BMD framework using machine learning (ML) models. Moreover, although APT has gained traction, there remains a scarcity of studies that interpret continuous data from embedded pavement instruments. This research fills that gap by tracking real pavement responses over time, offering insights into how various factors can impact pavement performance.

1.4.2. Broader Impacts

This dissertation makes a significant contribution to the field by strategically integrating high HRAP content with the BMD methodology. Its overarching objective is to develop pavements that are not only sustainable and durable but also economically viable and environmentally

responsible. A central focus of this research is the validation and refinement of BMD testing thresholds—an effort critical to strengthening the implementation and broader adoption of the BMD framework.

The findings address an urgent need within transportation agencies and the asphalt industry for performance-based tests in the mix design process. These tests demonstrate that BMD mixtures meet or exceed the performance of those designed using traditional volumetric methods. Additionally, the research facilitates the thoughtful incorporation of innovative approach into the mix design, promoting a balance between performance, cost, and environmental impact.

Another noteworthy contribution of this dissertation is its comprehensive analysis and synthesis of experiences from Virginia's APT operations. These insights provide an invaluable resource of practical knowledge and strategic guidance that can shape the direction of future APT initiatives. The lessons learned will enhance APT methodologies, enabling more effective evaluation of pavement performance under real-world conditions.

The outcomes of this research are expected to leave a lasting impact on both the academic and professional domains. By offering empirical evidence, innovative approaches, and strategic recommendations, this dissertation not only advances the state of knowledge but also provides actionable guidance for the design, evaluation, and implementation of HRAP asphalt mixtures. These contributions pave the way for sustainable and performance-driven practices in pavement engineering.

1.5. Dissertation Organization

The research objective was achieved through the development of five main research manuscripts, each addressing distinct and specific objectives. These manuscripts comprehensively covered the study's scope, offering an in-depth exploration of key research themes and questions. Alongside an **Introduction** and **Literature Review**, as well as a final **Findings, Conclusion, and Recommendations**, this dissertation will be structured into eight chapters. Figure 1-1 illustrates the organization of the dissertation, documenting and analyzing the entirety of the 2020 Virginia APT BMD program.

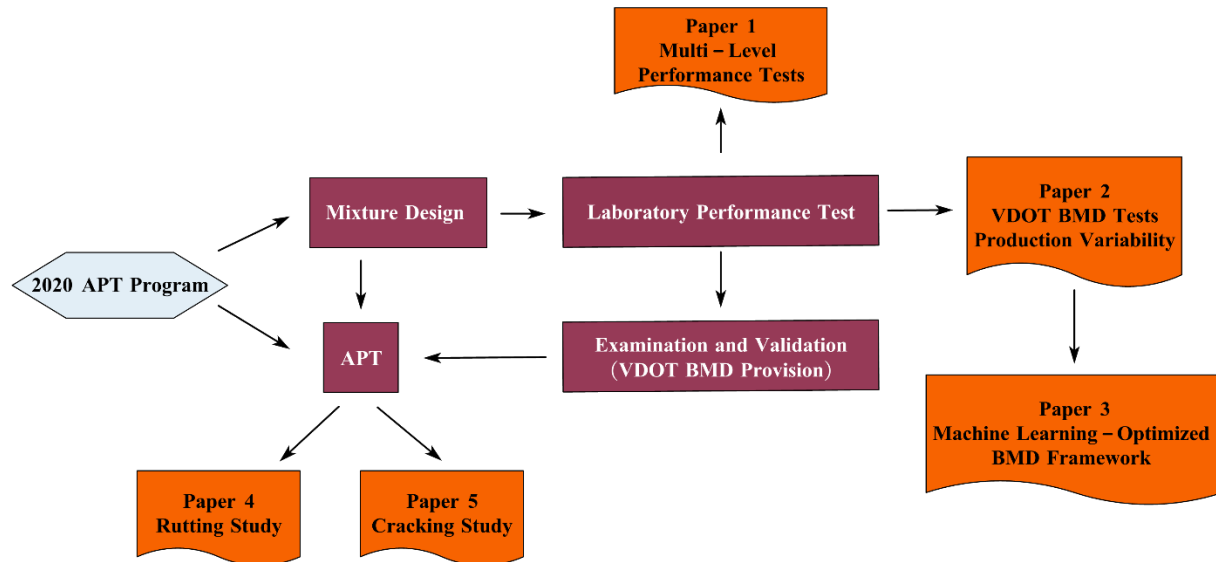


Figure 1-1. Conceptual flowchart of each paper's relevance within the APT program data analysis process.

A summary of each chapter's major contents follows below:

Chapter 1 – Introduction: This chapter presents an overview of the dissertation topic, articulated the problem statements, defined the research objectives and scope, highlighted the significance of the study, and outlined the organization of the dissertation.

Chapter 2 – Literature review: This chapter provides a comprehensive review of the literature, discussing the background and methodologies relevant to the dissertation. It comprised eight subsections, covering: RAP usage, HRAP mix design, definition of BMD, state-level BMD initiatives, Virginia's BMD efforts, laboratory performance evaluation, APT, and machine learning models.

Chapter 3 - Paper 1: Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia. This paper sets forth a series of specific objectives by employing performance testing across varying complexity levels—basic, intermediate, and advanced—this study offers a layered perspective on mixture performance. Following this, different performance test results are ranked to establish comparative insights. Diving deeper, an in-depth analysis was undertaken to unravel the interconnections and distinctions among various laboratory performance test outcomes, probing for inconsistencies that may surface when different performance test levels are applied. This exploration moved beyond isolated performance indices, delving into optimized BMD mixtures by examining relative correlations across primary performance metrics and proposed an innovative CPI. To visually capture the intricate "balance" among these indices, a 3-D plot, paired with a ternary plot, is introduced. This visualization tool

provides a dynamic and holistic view of BMD optimization impacts, offering a nuanced understanding of how the evaluated mixtures fare under comprehensive performance scrutiny. This paper has been published in the *International Journal of Pavement Engineering*.

Chapter 4 - Paper 2: Assessment of Production Variability and Composite Performance

Index for Conventional and High RAP BMD Mixtures. This paper sought to assess the impact of production and associated variability on a control mixture and five engineered mixtures designed using the BMD concept. Through VDOT's current suite of BMD tests, each mixture—sourced during production and paving of the APT facility—underwent evaluation in terms of durability, rutting and cracking performance. To advance and facilitate the analysis, a 3-Dimensional plot with a revised CPI (CPI_R) was introduced to optimize the process of evaluating the mixture “balance” status among multiple primary performances. The assessment offered comprehensive laboratory performance test references for quality measurement practices within BMD implementation. This paper has been published in the *Transportation Research Records*.

Chapter 5 - Paper 3: Machine Learning-Based Prediction and Optimization of Balanced

Mixture Design Performance Indices. In addressing the challenges highlighted in paper 1, this study introduces a ML-driven solution aimed at enhancing both the design and production phases within the current BMD framework. The primary aim of this research was to develop ML models capable of accurately predicting BMD performance indices based on mixture constituent and volumetric properties gathered from both design and field production stages. Beyond prediction, the study sought to leverage these ML models to explore the influence of individual constituent properties on BMD performance, pinpointing key factors that dominate in the preparation of BMD specimens. Traditional approaches often struggle with these complex, multi-property interactions, which ML is uniquely suited to unravel. To demonstrate the practical application of ML in BMD implementation, a simulated deployment was crafted, enabling simultaneous optimization across three primary BMD tests. Through this approach, top-performing ML-enhanced candidates were identified, presenting a compelling case for the strategic incorporation of ML in advancing the BMD process. This paper has been submitted for publication at the *Transportation Research Records* in October 2024.

Chapter 6 - Paper 4: Rutting Performance Evaluation of BMD Surface Mixtures with

Conventional and High RAP Contents Under Full-Scale Accelerated Testing. This paper embarks on an exploration into the permanent deformation (rutting) of six SMs incorporating both conventional and high RAP contents, all examined under full-scale accelerated testing with progressively increasing loading intensities. Central to this study are the carefully orchestrated design of pavement test cells, an intricately detailed traffic testing protocol, and

meticulous rutting performance measurements. This paper introduced an empirical rutting model to incorporate environmental aging effects. To deepen the analysis, a correlation study is undertaken, bridging the gap between APT-measured rut depths and outcomes from two laboratory-based rutting performance tests. As a final contribution, this work seeks to validate and refine the existing VDOT BMD thresholds, adding a layer of rigor to current standards. This paper has been published in the *Materials*.

Chapter 7 - Paper 5: Cracking Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: Insights from Full-Scale Accelerated Testing Program. This paper embarked on a thorough investigation into the cracking performance of the six SMs with both conventional and high RAP contents through full-scale accelerated testing. This study encompasses the setup of APT pavement test cells for cracking study, and the traffic testing plan. A multi-layered analysis unfolds, integrating longitudinal strain monitoring, comprehensive field surveys, and an array of laboratory performance tests, each contributing to an in-depth ranking of the mixtures under evaluation. Further, correlation analyses bridge APT field data with laboratory-based BMD performance tests currently utilized by VDOT, specifically examining IDT-CT and Cantabro test results. Lastly, this study juxtaposes the cracking performance insights with rutting performance results, facilitating a balanced discussion on the primary performance attributes of the evaluated mixtures, ultimately advancing the understanding of the impact of BMD in surface mixture design. This paper is to be submitted for publication in January 2025.

Paper 8 - Findings, Conclusion, and Recommendations: This concluding chapter wraps up the dissertation by distilling the essence of the research undertaken, spotlighting key findings, and charting out insightful pathways for future exploration.

1.6. Reference

- 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.
- Diefenderfer, S. D., Habbouche, J., & Boz, I. (2023). Balanced Mix Design for Surface Mixtures: 2020 Field Trials (No. FHWA/VTRC 23-R13). Virginia Transportation Research Council (VTRC).
- Habbouche, J., Boz, I., & Diefenderfer, S. D. (2022). Validation of Performance-Based Specifications for Surface Asphalt Mixtures in Virginia. *Transportation Research Record*, 2676(5), 277-296.

- Martin, A. E., Kaseer, F., Arámbula-Mercado, E., Bajaj, A., Cucalon, L. G., Yin, F., ... & King, G. (2020). Evaluating the Effects of Recycling Agents on Asphalt Mixtures with High RAS and RAP Binder Ratios (No. Project 09-58).
- Golalipour, A., Veginati, V., & Mensching, D. J. (2021). Evaluation of Asphalt Mixture Performance Using Cracking and Durability Tests at a Full-Scale Pavement Facility. *Transportation Research Record*, 2675(11), 226-236.
- Harvey, J., & Popescu, L. (2000). Accelerated Pavement Testing of Rutting Performance of Two Caltrans Overlay Strategies. *Transportation Research Record*, 1716(1), 116-125.
- Meroni, F., Flintsch, G. W., Habbouche, J., Diefenderfer, B. K., & Giustozzi, F. (2021). Three-Level Performance Evaluation of High RAP Asphalt Surface Mixes. *Construction and Building Materials*, 309, 125164.
- Steyn, W. J. (2012). Significant Findings from Full-Scale Accelerated Pavement Testing (Vol. 433). Transportation Research Board.
- Tong, B., Habbouche, J., Flintsch, G. W., & Diefenderfer, B. K. (2023). Rutting Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents under Full-Scale Accelerated Testing. *Materials*, 16(24), 7611.
- West, R., Rodezno, C., Leiva, F., & Yin, F. (2018). Development of a Framework for Balanced Mix Design. *Project NCHRP*, 20-07.
- Zaumanis, M., Mallick, R. B., & Frank, R. (2013). Evaluation of Rejuvenator's Effectiveness with Conventional Mix Testing for 100% Reclaimed Asphalt Pavement Mixtures. *Transportation Research Record*, 2370(1), 17-25.

2.1. RAP Usage

The use of RAP is greatly influencing asphalt mixture production and development worldwide. While RAP usage has increased throughout the years, its growth plateaued at a national average of around 20% in 2014 (West and Copeland, 2015; Williams et al., 2020). According to a survey conducted by the National Asphalt Pavement Association in 2019, the average RAP content in asphalt mixtures was 21.1%, which remained the same as the preceding year (Williams et al., 2020). Strict requirements have been placed on the use of RAP in asphalt mixtures by state agencies; this is mainly due to perceptions of inferior quality and hurdles related to the source and quality of RAP materials along with processing during production (Diefenderfer et al., 2021a). However, some countries have been pushing the boundaries of recycling and have been leading the way in using relatively higher RAP contents; the Netherlands, for example, allows the inclusion of up to 50% RAP in most asphalt mixtures (Mohajeri, 2015). With the implementation of a new specification system in line with European standards (EN13108 series), Dutch contractors are able to design and select their own composition as long as the performance requirements for the resultant mixture are fulfilled. In Japan, RAP content averages 47% in asphalt mixtures with RAP fractionation, moisture control, frequent testing of RAP materials, and the use of RAs being among the measures adopted by the Japanese industry to ensure that performance requirements are met (West and Copeland, 2015).

In the U.S., the environmental and economic benefits gained by the use of RAP in asphalt mixtures have encouraged state agencies to introduce special provisions and specifications allowing its use at relatively higher contents in mixtures (Zaumanis et al., 2013; Martin et al., 2020). The increased use of RAP was expected to offset the continuously rising cost of asphalt binders used to produce mixtures and fuel needed to transport and place them (Zaumanis et al., 2013; Martin et al., 2020). In the U.S., the amount of RAP available in each state, its corresponding performance, the various practices and standards being implemented, and other factors have led the definition of HRAP asphalt mixtures to become state specific. In 2020, Habbouche et al. performed a review from multiple perspectives of the state of the practice on the use of recycled asphalt materials and RAs in surface mixtures (Habbouche et al., 2021a). The review focused on the permissibility and usage of RAP in mixtures alongside various other aspects. According to the responses received from state agencies and according to an online search of specifications, special provisions, and field trial / pilot projects, the allowable RAP content in mixtures varied from 15% to 50% (Habbouche et al., 2021a). For instance, California defines HRAP asphalt mixtures as those containing 25% to 40% RAP and limits their production to special pilots. Kansas allows up to 40% RAP content if the millings used correspond to the same field site where the resultant mixture is to be placed. In Nebraska and Vermont, the RAP content in HRAP asphalt mixtures ranges between 25 and 50%; however, it was reported that the majority of asphalt mixtures use no more than 20% RAP. In Georgia, mixtures that contain 30% and higher RAP content are considered

HRAP asphalt mixtures. A similar definition is being adopted by Missouri, North Carolina, South Carolina, South Dakota, and Virginia. States such as Missouri and New Jersey do not specify an upper limit on the RAP content as long as the resultant mixtures meet pre-established specific properties and performance requirements. Other states, such as Pennsylvania, require additional testing and processing for mixtures with RAP content greater than 15% (Habbouche et al., 2021a).

The primary concern with HRAP asphalt mixtures is that they can become overly stiff due to the increased amount of aged binder used, making them more brittle and prone to premature cracking (Martin et al., 2020). Numerous construction and performance issues can result from the use of relatively higher RAP contents in mixtures such as challenging compactability and workability in cool weather, thermal cracking, fatigue cracking, reflective cracking induced due to repeated loading and daily / seasonal thermal stresses, and raveling with subsequent aging or moisture damage (Martin et al., 2020). The challenges arising from the use of HRAP asphalt mixtures can be addressed through the use of softer binders and/or additives such as RAs.

2.2. RAP Mixture Design

The formal integration of RAP in hot mix asphalt (HMA) following the Superpave design procedure was established through the NCHRP Project 9–12. This project introduced a three-tier system based on the characteristics of the hardened RAP binder, and blending charts were developed for high RAP contents (McDaniel and Anderson, 2000, West et al., 2013). Subsequently, NCHRP Project 9–46, which focused on enhancing recycling practices, guided the design of high RAP mixes using the RAP binder ratio as a critical parameter. The recovery and grading of the aged binder were key steps in the design process. To facilitate the inclusion of higher RAP quantities, the use of recycling agents like rejuvenators and softening agents has been adopted. Rejuvenators, distinct from softening agents, are chemical agents that aim to restore the physical and chemical properties of the old binder, while softening agents primarily reduce the viscosity of the aged binder (Roberts et al., 1991). The selection of an appropriate rejuvenating agent, ensuring compatibility with the aged binder, is crucial (Al-Qadi et al., 2007).

However, uncertainties persist regarding the properties of the rejuvenated RAP binder. As a result, RAP has often been relegated to lower-level applications, not fully utilizing the potential value of the asphalt binder present in the RAP (Zaumanis et al., 2015). Despite advancements in this field, there remains a lack of consensus among researchers regarding the impact of RAP on field performance, highlighting a continuing area of debate and investigation. Huang et al. conducted a laboratory study and observed that incorporating up to 20% RAP in asphalt mixtures improved tensile strength and fracture resistance (Huang et al., 2004). However, increasing RAP content to 30% significantly altered the mix's performance. Kim et al. found that mixes with higher RAP content demonstrated enhanced rutting resistance but exhibited reduced fracture energy. They noted that the use of softer binders in high-RAP mixes could diminish rutting performance (Kim et al., 2007). Researchers also reported that mixtures containing RAP generally exhibit lower stiffness, reduced resistance to permanent deformation, and improved fatigue resistance compared

to their RAP-free counterparts. This behavior is likely attributed to the use of softer binders and rejuvenating agents (Widyatmoko, 2008). Apeagyei et al. observed that mixtures with both low (0%) and high (25%) RAP contents exhibited similar rutting performances, with the optimal performance seen in mixtures containing intermediate RAP levels (10% and 15%) (Apeagyei et al., 2011). Meanwhile, others found that increasing RAP content led to an increase in complex modulus, tensile strength, moisture damage resistance, fatigue life, and rutting resistance, while thermal cracking resistance decreased (Al-Qadi et al., 2012). Additionally, McDaniel et al. discovered that RAP addition increased the stiffness of the mixes, enhanced fatigue life, and did not significantly affect thermal cracking resistance (McDaniel et al., 2012). Izaks et al. observed that high RAP content mixtures had higher rutting resistance compared to traditional mixes, with similar mechanical properties such as fatigue resistance and stiffness (Izaks et al., 2015).

Numerous variables influence the characteristics of asphalt mixes, thereby precluding broad generalizations regarding the effects of RAP. Factors such as the inherent variability of RAP, the complexities in determining the interaction levels among virgin binder, recycled binder, and additives, as well as the unresolved questions surrounding rejuvenators, have necessitated the development of novel mix design methodologies (West et al., 2018; Meroni et al., 2020). In response to these challenges, the BMD concept was introduced. This approach aims to effectively design asphalt mixes with HRAP contents, ensuring favorable performance outcomes (Zhou et al., 2006; Meroni et al., 2020).

2.3. The Definition of BMD

BMD is defined as “Asphalt mix design using performance tests on appropriately conditioned specimens that address multiple modes of distress taking into consideration, mix aging, traffic, climate, and location within the pavement structure” according to the former Federal Highway Administration Expert Task Group on mixtures and construction (West et al., 2018; Yin and West, 2021). The BMD method replaces some aspects of traditional volumetric design with performance testing criteria for the most common distresses such as rutting and cracking. The approach requires that a mixture design pass performance criterion for approval. Additional performance testing during production and construction may be required for mixture acceptance. Although BMD mixtures cannot compensate for an unsound underlying pavement structure or the selection of inappropriate maintenance treatments, BMD constitutes a significant step forward in the pursuit of better-performing asphalt mixtures. Numerous state highway agencies are looking into designing and accepting asphalt mixtures using the BMD concept (West et al., 2018; Yin and West, 2021). AASHTO PP 105, Standard Practice for Balanced Design of Asphalt Mixtures, proposes four primary approaches to BMD for mixture design, labeled A through D, having increased reliance on performance properties over volumetric requirements. The potential of innovation increases as compliance with the existing volumetric requirements decreases from Approach A (volumetric design with BMD verification) to Approach D (BMD design only) (West et al., 2018; AASHTO, 2020; Yin and West, 2021; E-C280, 2022).

- ***Approach A – Volumetric Design with BMD Verification.*** The current volumetric mix design method (i.e., Superpave, Marshall, or Hveem) is utilized as a starting point for determining the optimum binder content (OBC). The AC mixture is then tested with selected mechanical tests correlated to field performance to assess its resistance to distresses of interest at the OBC. If the mix design meets the performance test criteria, the job mix formula (JMF) is established and production begins. Otherwise, it is necessary to repeat the entire design process.
- ***Approach B – Volumetric Design with BMD Optimization.*** This approach is an expanded version of Approach A. It also starts with the conventional in-practice mix design method for determining a preliminary OBC based on volumetric properties. The AC mixtures with the preliminary OBC and two or more additional binder contents are then tested with selected performance tests. The asphalt binder content at which the corresponding AC mixture meets all performance criteria is then identified as the final or target OBC. Otherwise, the entire mix design process needs to be repeated until all the BMD test criteria are satisfied.
- ***Approach C – BMD-Modified Volumetric Design.*** In this approach, initial component material properties, proportions, and OBC are established using the current volumetric mix design method. Following the performance test, binder content or mix component properties are adjusted (for example, aggregates, asphalt binder, recycled materials, and additives) until the performance criteria are satisfied. The idea of this approach is to primary focus on meeting performance test criteria and not necessarily all the volumetric criteria.
- ***Approach D – BMD Performance Only.*** This approach establishes and adjusts mixture components and proportions based on performance analysis with limited or no agency requirements for volumetric properties. Upon meeting the performance criteria in the laboratory, the mixture volumetrics may be checked for use in production.

2.4. State BMD Efforts

The National Center for Asphalt Technology (NCAT) conducted a survey of State Highway Administrations (SHAs) and the asphalt pavement industry. As of 2020, the implementation status of BMD in the U.S. is presented illustrated in Figure 1-1 Figure 2-1. Numerous state highway agencies are looking into designing and accepting asphalt mixtures using the BMD concept.

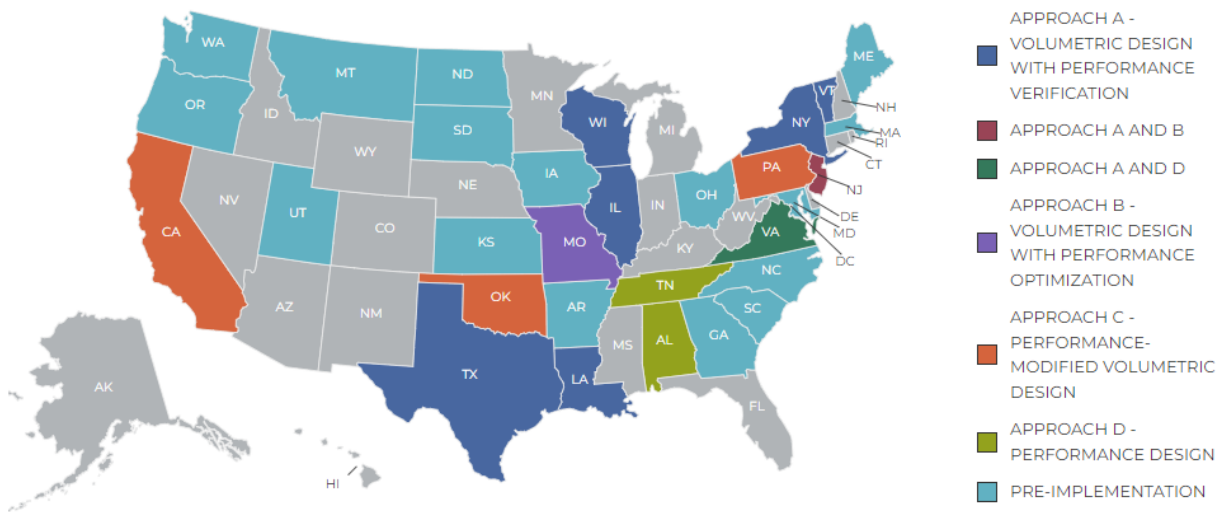


Figure 2-1. Implementation Status of BMD Concept in the United States (NAPA 2020).

California has implemented a comprehensive framework for BMD mixtures, encompassing performance-based specifications and its mechanistic empirical design program, CalME. These mixtures are predominantly designed for and applied to pavements with very high traffic volumes. In this framework, BMD specifications are specifically directed at plant-produced mixtures. The performance testing protocol within this system includes the Repeated Simple Shear Test (RSST), Bending Beam Fatigue (BBF) test, and the Hamburg Wheel-Track Test (HWTT) (Harvey et al., 2014).

The Illinois Department of Transportation (IDOT) is currently adopting the implementation of BMD Approach A. As part of this process, IDOT mandates the execution of the HWTT, Illinois Flexibility Index Test (I-FIT), and a modified version of the Tensile Strength Ratio test. These tests are crucial for assessing pavement susceptibility to rutting, fatigue, and moisture damage, respectively. A publication by Al-Qadi et al. (2019) introduced a long-term aging protocol for the I-FIT, proposing the use of forced-draft ovens for aging compacted specimens.

The Iowa DOT predominantly designs its asphalt mixtures using the conventional Superpave volumetric approach. However, for mixtures intended for very-high-volume traffic or those produced with specific aggregate mineralogy, an assessment for rutting resistance is mandatory. This assessment, conducted by either the contractor or an external mix design laboratory, involves the HWTT. The testing temperature for the HWTT is determined based on the Performance Grade (PG) high temperature of the asphalt binder. Furthermore, for specialized asphalt mixtures, additional performance testing and acceptance criteria may be necessary. In addition to these existing protocols, the Iowa DOT is contemplating the inclusion of the Disc-Shaped Compact Tension Test in their BMD framework. This test would be used to evaluate the mixture's resistance to thermal cracking, as suggested by West et al. (2018).

The Louisiana DOT has adopted the BMD Approach A, which integrates conventional volumetric criteria with loaded wheel tracking (LWT) and Semi-Circular Bend (SCB) tests. These tests are utilized to assess rutting and intermediate temperature cracking, respectively, as part of their streamlined performance-based specifications. The process for roadway acceptance testing involves the collection of 25 random cores from five sublots, with each subplot providing five random cores. Some of these cores are used for density measurement and verifications, while others are subjected to LWT and SCB testing, as outlined by Mohammad et al. (2016). Ongoing efforts are focused on evaluating variations in test parameters across different specimen types (such as those from mix design, plant-produced, and field cores), developing an accelerated aging protocol, and integrating the SCB test into the quality control process.

The New Jersey DOT (NJDOT) is currently employing the BMD Approach A for a variety of asphalt mixtures. These include RAP mixtures, high-performance thin overlays, binder-rich intermediate courses, bottom-rich base courses, and bridge deck waterproofing surface courses. The performance-testing protocol, applied during both the mix design and plant-production stages, encompasses APA testing at 64°C, Tensile Strength Ratio and Overlay Test (OT) at 25°C, and BBF testing at 15°C. For high RAP mixtures, the maximum APA rut depths after 8,000 cycles at 64°C are set at 4.0 mm for modified asphalt binders and 7.0 mm for unmodified binders. Furthermore, the minimum number of cycles to failure using the OT for HRAP SMs is established at 275 cycles for modified mixtures and 200 cycles for unmodified mixtures. For high RAP intermediate and base mixtures, the required number of overlay test cycles is reduced to 150 for modified mixtures and 100 for unmodified mixtures. Binder-rich intermediate courses and high-performance thin overlays are mandated to have APA rut depths of no more than 6.0 mm and 4.0 mm, respectively, and a minimum of 700 and 600 overlay test cycles. In a move towards simpler and faster testing methods, such as the IDT at intermediate and high temperatures to evaluate cracking and rutting, preliminary thresholds for IDT strength at high temperature and the Cracking Tolerance Index (CT index) at intermediate temperature have been identified. These are currently under further investigation for potential implementation, as reported by Bennert et al. (2020). This approach signifies NJDOT's commitment to refining and enhancing testing procedures for optimal asphalt mixture performance.

The Texas Department of Transportation (TxDOT) currently employs the BMD Approach A for specialized asphalt mixtures, including porous friction courses, Stone Matrix Asphalt, thin overlay mixtures, and hot in-place recycling of asphalt concrete surfaces. This approach necessitates the use of the HWTT and OT to assess the mixture's resistance to rutting/moisture damage and reflection/bottom-up cracking, respectively. The Superpave volumetric mix design criteria are utilized to establish an OBC, after which the HWTT and OT are applied to evaluate specimens at three different binder contents: OBC, OBC + 0.5%, and OBC + 1.0%. The final OBC selection is based on fulfilling the requirements of both tests. Additionally, Zhou et al. (2020) have proposed a quality control/quality assurance (QC/QA) acceptance protocol incorporating practical performance-related tests suitable for production QC, such as the IDT at intermediate and high

temperatures. This protocol involves sampling and conditioning produced loose mixtures at 135°C for 2 hours, followed by compacting performance test specimens to an air-void level of $7 \pm 0.5\%$. The proposed QC production acceptance criteria include a minimum CT index at 25°C of 105 and a minimum IDT shear strength of 1.02 MPa at 50°C (Zhou et al., 2020).

An increasing number of states are actively engaged in initiatives related to the Balanced Mix BMD approach. This list includes, but is not limited to, Florida, Georgia, Indiana, Minnesota, Nebraska, New Hampshire, New Mexico, Ohio, Oklahoma, Oregon, South Dakota, Utah, and Wisconsin. These states are part of a growing trend, recognizing the importance of BMD in advancing pavement design and performance (Diefenderfer et al., 2021a).

2.5. Virginia BMD Efforts

As one of these agencies, the VDOT has been working towards the implementation of BMD for several years. VDOT had been experiencing increased surface cracking over several years, and BMD was considered as a means to address cracking concerns. Another benefit of moving toward BMD that was attractive to VDOT was the opportunity to consider innovations in mixture designs. The efforts began during a time of interest in the use of high RAP contents and stemmed from the concern about the potential performance of these mixture and knowledge that volumetric design could not account for the performance of mixtures with higher RAP contents and/or other additives (Diefenderfer et al., 2023).

In 2007, VDOT introduced specifications to allow higher percentages of RAP (i.e., up to 30%) in SMs without adjustment of the virgin binder grade. By 2013, VDOT had begun to consider the feasibility of allowing the use of SMs containing up to 45% of RAP material (Diefenderfer et al., 2021b). Field trial projects featuring the use of 40-45% RAP contents in surface mixtures were constructed (Nair et al., 2019). In 2017, VDOT began to evaluate the feasibility of introducing performance requirements into mix design using the BMD method. In January 2018, an initial effort was undertaken to provide benchmark indications of performance for several conventional unmodified mixtures typically used in Virginia to support proposed pilot projects containing higher RAP contents. Three fast, simple, and practical performance-indicative tests addressing different modes of distress were selected for use as part of the BMD method. The selected tests were Cantabro test, the indirect tensile cracking test (IDT-CT) at intermediate temperature, and the asphalt pavement analyzer (APA) rut test for assessing the overall potential for durability, cracking, and rutting of asphalt mixtures, respectively. Initial performance threshold criteria were then developed for the selected tests: a maximum of 7.5% mass loss for the Cantabro test, a maximum of 8.0 mm of rut depth at 64°C for the APA rut test, and a minimum CT index of 70 for the IDT-CT at 25°C. Moreover, Diefenderfer et al. assessed and validated the developed performance-based specifications for surface asphalt mixtures produced using unmodified asphalt binders prior to full implementation in Virginia (Diefenderfer et al., 2021b; Habbouche et al., 2022).

Two VDOT BMD special provisions were then drafted and revised for use in pilot projects: Special Provision for HRAP Content Surface Mixtures Designed Using Performance Criteria and Special Provision for Dense Graded Surface Mixtures Designed Using Performance Criteria. In 2019, two field trials were planned and executed to design, produce, and place BMD mixtures in Virginia. These field trials constituted the first applications of these BMD specifications. Nine mixtures were evaluated from the two field trials; the mixtures incorporated combinations of different RAP contents, two binder grades, two RAs, and two warm mix asphalt (WMA) additives (Diefenderfer et al., 2021b). In 2020, additional field trials featuring the use of higher RAP content, softer binder, RAs, and various other additives such as fibers and softening oils were planned and constructed. Based on the test results, mixtures containing a softer binder, 40% RAP, and RAs may be designed and produced to meet current BMD performance thresholds and current volumetric properties, gradation, and asphalt content requirements. However, the long-term field and laboratory performance of all these mixtures needs to be evaluated to verify these early findings (Diefenderfer et al., 2023). Currently, VDOT allows the incorporation of RAP contents up to 30% in dense-graded SMs with A and D. The "A" designation corresponds to a PG 64-22 asphalt binder, while "D" corresponds to a PG 70-22 binder. These mixtures are intended for traffic loads of 0 to 3 million equivalent single axle loads (ESALs) and 3 to 10 million ESALs, respectively.

2.6. Laboratory Performance Tests

Laboratory performance tests play an essential role in the BMD process, as they serve not only to guarantee the production of sustainable and cost-effective materials but also to validate the long-term performance of HRAP asphalt mixtures. Various performance tests may be conducted to support the mix design and quality control. It is essential to undertake a comprehensive evaluation of the performance of mixtures by analyzing the results of various performance tests for rutting and cracking—the primary distresses of concern.

2.6.1. Rutting Tests

2.6.1.1. Flow Number Test

By applying haversine compressive-load pulses (represented in Figure 2-2) and recording the cumulative deformation as a function of a number of load cycles, the repeated load test was initially developed to identify the permanent deformation characteristics of HMA mixtures. A certain testing stress level must be selected in order to achieve tertiary flow in a reasonable number of cycles not exceeding 10,000. According to AASHTO T378, the use of confining pressure is optional but, if used, it must be maintained at a constant level throughout the test. The flow number (FN) is defined as the number of load cycles corresponding to the minimum rate of change in permanent axial strain, which marks the onset of the tertiary region (as depicted in Figure 2-3). In other words, FN indicates the initiation of shear deformation in asphalt mixtures, serving as a meaningful indicator of rutting resistance. Researchers also used this repeated load test for cold

recycling materials, applying a loading stress level of 138 kPa (0.1 s loading with 0.9 s rest) at 40 °C and found a higher flow number was obtained with the decrease of foam asphalt content (Kim et al., 2009).

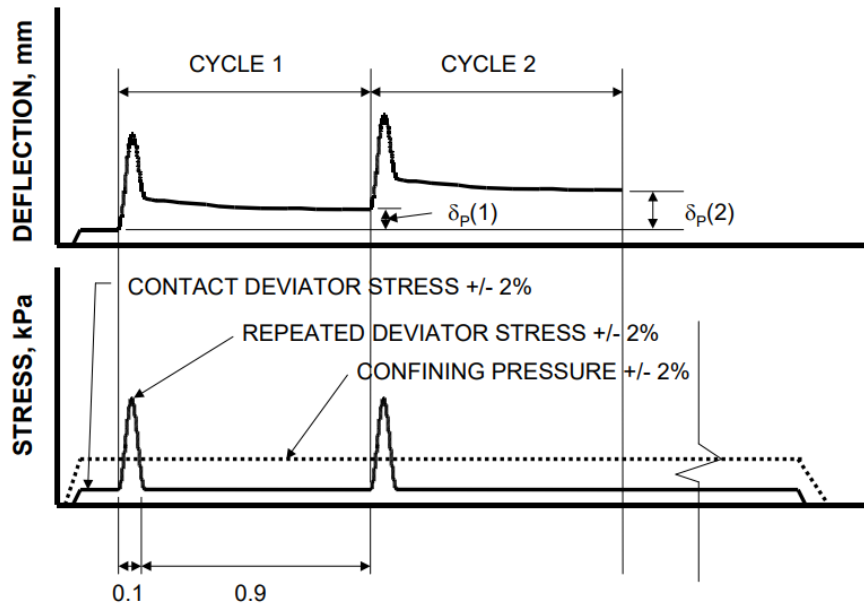


Figure 2-2. Loading in the flow number test (Bonaquist, 2012).

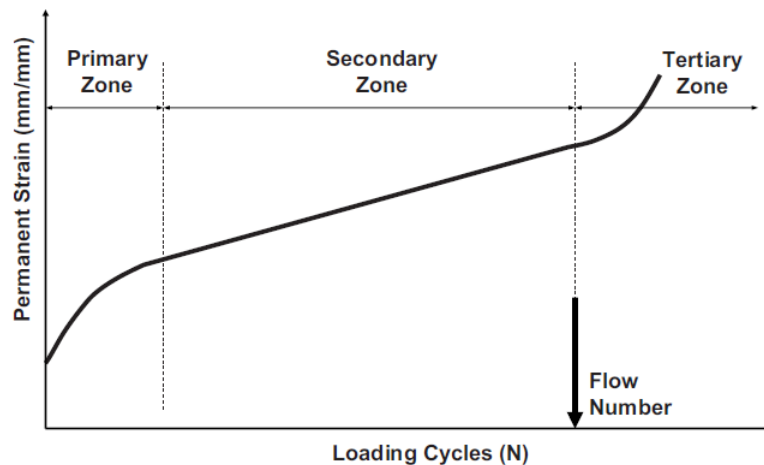


Figure 2-3. FN testing plot (Kim et al., 2009).

2.6.1.2. Asphalt Pavement Analyzer (APA) Rut Test

The Asphalt Pavement Analyzer (APA) (shown in Figure 2-4) is a loaded wheel tester (LWT) that checks for permanent deformation (rutting), fatigue cracking, and moisture susceptibility in both HMA and CR mixtures. A number of recycling mixtures containing hot lime slurry have been tested by APA for rutting resistance and moisture damage (Cross, 1999). Another feature of APA is that it can be used to perform Hamburg Loaded Wheel Test (HLWT), Loaded

Wheel Fatigue Test (LWFT) and Asphalt Pavement Analyzer Testing (Cox and Howard, 2015). The APA rut test was performed on specimens prepared from loose mixture collected during production in accordance with AASHTO T 340. The APA rut test was performed on specimens 150 mm in diameter by 75 ± 2 mm in height compacted using a gyratory compactor to in-place and $7 \pm 0.5\%$ air voids. This test simulates rutting in the laboratory by applying a loaded wheel back and forth over a pressurized rubber tube located along the surface of the test specimen at a temperature of 64°C which is equal to the high PG temperature of the employed asphalt binder. After 8,000 cycles were applied, the deformation of the specimen was measured. For each mixture, two replicates were tested, and an average was reported.

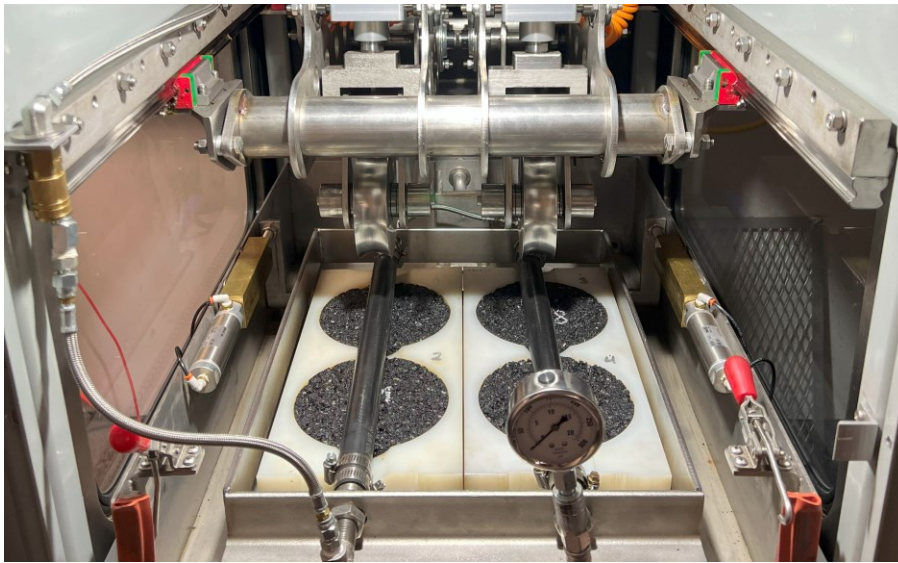


Figure 2-4. Asphalt Pavement Analyzer.

2.6.1.3. Hamburg Wheel Tracking (HWT)

To determine the rutting and stripping characteristics of asphalt mixtures, the HWT simulates traffic loading on specimens. In the HWT test, a steel wheel moves (around 52 passes/min) across a pair of asphalt mixing asphalt mixture specimens submerged in water. Through the use of a linear variable displacement transducer (LVDT), the rut depth is measured at 11 points in the wheel's passing direction (Bairgi et al., 2022).

2.6.1.4. Stress Sweep Rutting (SSR) Test

SSR test was performed in accordance with AASHTO TP 134 to assess the resistance of mixtures to rutting under repetitive loading, simulating the conditions experienced by pavements under repeated traffic loads. SSR testing was originally motivated by NCHRP Project 09-19, which showed that mixtures are sensitive to load level, loading history, and temperature under deviatoric axial loads. Four specimens were subjected to a confined cyclic compression testing at two different temperatures (T_H and T_L). In both cases, the specimens were subjected to three 200-cycle loading blocks of three deviatoric stress levels, and constant confining pressure of 69 kPa. In the

testing, the permanent deformation of specimens is measured with its deviatoric stress, loading history, and temperature. Once a shift model is established, the permanent deformation of asphalt layers as a function of pavement depth and time can then be predicted by the pavement performance prediction program, FlexPAVE™ (Ghanbari et al., 2020). The rutting strain index (RSI) represented by FHWA, with the support of the shift model built by SSR, can be used as a rutting indicator (Ghanbari et al., 2020). The RSI is the percent average permanent strain calculated by repeating 30 million 18-kip equivalent single-axle load (ESAL) over 20 years on a standard pavement structure (Ghanbari et al., 2020). A lower RSI value is an indicator of a greater resistance to rutting for a given mixture.

2.6.2. Cracking Tests

2.6.2.1. Indirect Tensile Cracking Test (IDT-CT)

IDT-CT is a basic test (also known as IDEAL-CT) commonly used by state agencies and was incorporated into the BMD approach for VDOT to evaluate mixture resistance to cracking. The test was conducted at 25°C on SGC specimens 150 mm in diameter and 62 mm in height compacted to $7 \pm 0.5\%$ air voids in accordance with ASTM D8225. During the test, specimens were subjected to a constant load-line displacement rate of 50 ± 2 mm/min. The cracking tolerance index (CT index) was then calculated from the test load-displacement curve / data. Five specimens were tested for each mixture and an average CT index was reported.

2.6.2.2. Texas Overlay Tester (OT testing)

To evaluate cracking susceptibility, the OT testing simulates the reflective cracking mechanism of opening and closing joints and/or cracks (Bairgi et al., 2022). A cylindrical specimen of 150mm in diameter and 100mm in height is cut into a specimen of 50mm in height and 76mm in width. With epoxy, the specimens were attached to the testing plate and cured for 24 hours. The test specimen is subjected to a triangular cyclic displacement with an amplitude of 0.025 inches, and the required load is measured. The test ends up with a target percent reduction (93%) of maximum load. Critical fracture energy (G_c) and crack propagation rate are determined, describing the fatigue and flexibility properties of the specimen during crack propagation.

2.6.2.3. Semi-circular Bending (SCB)

The SCB method is a three-point bending test performed on a semicircular specimen with a notch. The tension produced at the bottom of the sample leads to crack propagation. An evaluation of strain energy (U) and J-integral was completed in this study. The area under the load-deformation curve is defined as the strain energy. The linear elastic fracture mechanics describes the strain energy, which is the energy required to create a unit area of a crack. On the other hand, J-integrals indicate critical fracture energies or critical fracture resistances cracks (Bairgi et al., 2022).

2.6.2.4. Direct Tension Fatigue Test

Direct tension cyclic fatigue tests were performed using the AMPT machine in accordance with AASHTO TP 107. The simplified viscoelastic continuum damage (S-VECD) model was used to assess the fatigue characteristics of the six mixtures. The cyclic fatigue test was performed at 24°C on specimens 100 mm in diameter by 130 mm in height cored from samples 150 mm in diameter by 175 mm in height compacted from loose mixtures collected from the plant during production. First, dynamic footprint testing was performed to calculate a dynamic modulus ratio (DMR). The number of cycles to failure (N_f) of the test specimens was determined when the phase angle or modulus value multiplied by the number of loading cycles ($|E^*| \times N$) reaches a maximum value followed by a sharp drop (Underwood et al., 2010).

The S-VECD model developed by Underwood et al. was used to predict the cumulative damage for different loading histories under different temperatures as it incorporates time-temperature superposition principle and the pseudo strain energy concept (Underwood et al., 2010). The pseudo strain energy is dependent on both pseudo strain and pseudo stiffness; the pseudo strain was determined using Equation 2-1.

$$\varepsilon^R = \frac{1}{E^R} \int_0^t E(t - \tau) \frac{d\varepsilon}{d\tau} d\tau \quad (1) \quad \text{Equation 2-1}$$

Where ε^R = pseudo strain; t = time; τ = reduced time; $E(t)$ = relaxation modulus; ε = strain; and E^R = the reference modulus.

Based on assumptions from S-VECD theory, in the quasi-static cyclic status, the simplified ε^R is shown in Equation 2-2. The pseudo stiffness (C), presented in Equation 2-3, gradually decreases from 1 to nearly 0 and indicates the material condition from its intact state to failure. With the pseudo energy concept, the relationship between C and S , known as damage characteristic curve (C-S curve), can be determined using in Equation 2-4.

$$\varepsilon^R = |E^*| \cdot \varepsilon \quad \text{Equation 2-2}$$

$$C = \frac{\sigma}{\varepsilon^R} \quad \text{Equation 2-3}$$

$$C = 1 - C_{11} S^{C_{12}} \quad \text{Equation 2-4}$$

Where C_{11} and C_{12} are the model coefficients.

The D^R failure criterion, determined using Equation 2-5 can be applied to the S-VECD model. D^R can serve as an indicator of toughness for the evaluated material, as demonstrated in Figure 5b, where a higher value signifies greater resistance of the material to damage (Wang et al., 2020).

$$D^R = \frac{\int_0^{N_f} (1-C) dN}{N_f} = \frac{\text{sum}(1-C)}{N_f} \quad \text{Equation 2-5}$$

To define the mixtures' fatigue performance, a fatigue cracking index parameter (S_{app}), considered both the material's modulus and toughness, is used. The calculation of S_{app} can be

found in Equation 9 (Wang et al., 2020) and conducted with FlexMAT™ for Cracking, an Excel-based tool provided by the Federal Highway Administration (FHWA, 2019).

2.6.3. Multi-level Performance Evaluation

Meroni et al. defined three levels of performance tests (basic, intermediate, and advanced) based on degree of complexity, cost, time, and training requirements (Meroni et al., 2021).

- **The basic level** included tests characterized with a short time for the specimen preparation and testing without requiring any specific cutting, coring, and gluing. Simple performance tests are noteworthy due to their practical features such as easy execution, reduced expenses, and time saving. For instance, the Cantabro mass loss test and IDT-CT at intermediate temperature were selected by VDOT for BMD implementation.
- **The intermediate level** included tests that require longer times for specimens' preparation and testing. The APA rut test was also selected by VDOT for BMD implementation.
- **The advanced level** features tests that require operations of cutting and/or coring to prepare the specimens as well as multiple days to be able to complete and analyze the test results. These tests usually require the use of the Asphalt Mixture Performance Tester (AMPT). The development of AMPT was initiated through National Cooperative Highway Research Program (NCHRP) Project 09-19 (Witczak et al., 2011) and NCHRP Project 09-29 (Bonaquist et al., 2003). The AMPT-based tests provide fundamental properties that are used in a pavement mechanistic-empirical (ME) design framework (e.g., AASHTOWare® Pavement ME and FlexPave™) for structural pavement design and assessment. Examples of advanced tests include the dynamic modulus ($|E^*|$) test, the direct tension cyclic fatigue test (CF), the stress sweep rutting (SSR) test.

A summary of the tests across different categories is presented in Figure 2-5. The tests highlighted in red are those utilized in the evaluation within this dissertation.

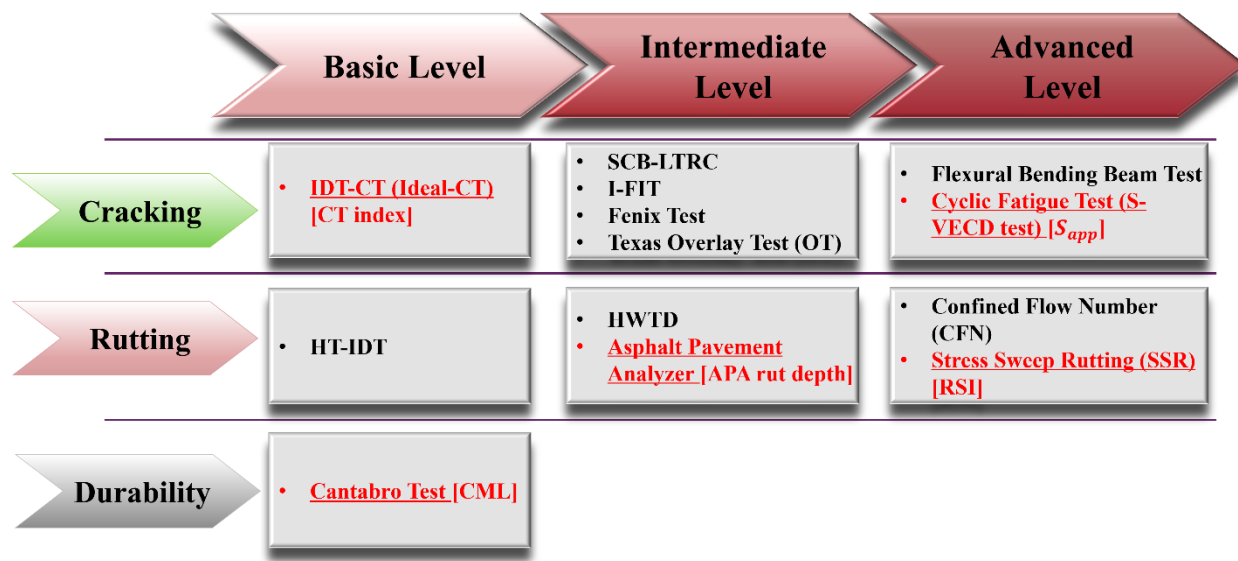


Figure 2-5. Multi-level laboratory performance tests.

2.7. Accelerated Pavement Testing

In addition to performance testing and evaluation in the laboratory, APT serves as a valuable tool to bridge the significant gap between ME design models developed using laboratory material testing characterization and the actual long-term pavement performance monitoring. APT offers a means to assess the performance of full-scale constructed pavements in an accelerated manner. This involves deliberately applying wheel loading, typically surpassing the de-sign load limit, to a pavement system to assess its response within a significantly condensed timeframe. Therefore, it provides insights that complement long-term pavement performance monitoring and analysis (Harvey and Popescu, 2000). This accelerated evaluation is accomplished through several means, including intensifying the number of load repetitions, adjusting loading conditions, introducing specific climatic conditions in terms of temperature and moisture, implementing thinner pavements with reduced structural capacity, or employing a combination of these factors (Steyn, 2012).

Numerous research institutions worldwide have developed and utilized APT facilities to investigate a vast array of topics. In South Africa, studies have delved into large aggregate mix bases, granular emulsion mixtures, rehabilitation strategies for cement-stabilized base pavements, and porous asphalt (Du Plessis et al., 2006). In 2020, Rust et al. evaluated two nano-silane-modified, emulsion-stabilized pavements under APT (Rust et al., 2022). In the Nordic region, institutes such as VTT in Finland and VTI in Sweden have examined common Finnish pavement structures, assessing the efficacy of mill-and-fill maintenance treatments aimed at increasing load-bearing capacities (Du Plessis et al., 2006). In France, pioneering research on electric road systems is gaining momentum, with the Charge As You Drive (CAYD) project conducting APT test on inductive charging system (Online resource: CAYD project). In the U.S., California's APT efforts have explored a diverse suite of materials—dense-graded asphalt concrete, asphalt rubber,

aggregate base, cement-treated base, and asphalt-treated permeable base (Plessis et al., 2006)—culminating in refined damage models that were calibrated using data from the Heavy Vehicle Simulator (HVS) for the California Mechanistic-Empirical software package (CalME) (Ullidtz et al., 2005). The Florida Department of Transportation (FDOT) has conducted assessments on the effects of polymer modification of Superpave mixtures, rutting performance of coarse and fine-grained mixtures, the implementation of PG 76-22 Asphalt Rubber Binder in Florida (Du Plessis et al., 2006, Greene et al., 2015). Furthermore, rutting performance evaluations of hot mix asphalt (HMA) and warm mix asphalt (WMA) concrete under high aircraft tire pressures and elevated temperatures were conducted at the National Airport Pavement and Material Research Center (Garg et al., 2018). Across the Pacific in China, top-down cracking has been a focal point, with researchers deploying a portable seismic property analyzer to monitor modulus reduction, offering a novel approach to evaluating pavement fatigue under APT simulations (Ma et al., 2018). Each of these initiatives underscores the dynamic and region-specific applications of APT, as institutions worldwide work to unravel the complexities of pavement performance under accelerated conditions.

In the U.S., Virginia is among the states that have implemented the use of APT facilities for pavement evaluation purposes (Figure 2-6). In 2015, a collaborative experiment between the VTRC, VDOT, and Virginia Tech was initiated and executed at the Virginia APT facility located at the VTTI. The experiment encompassed three major studies, including the investigation of the impact of different overlays on cold central plant recycling (CCPR) material (Xue et al., 2017; Xue et al., 2020; Flintsch et al., 2020), the assessment of field performance of two distinct dense-graded SMs designed with varying design gyration levels (50 and 65) (Meroni et al., 2020), and the examination of reflection cracking. In 2020, another effort was planned and initiated to assess the application of the BMD concept in designing durable and longer-lasting mixtures in Virginia with a particular emphasis on mixtures incorporating relatively higher RAP contents. The overarching scope of the effort included the development of the APT experiment, verification of mix design performance, documentation of paving operations and construction practices, and mixture sampling during production. Moreover, it included coring of as-placed material, testing and analysis of the volumetric and performance properties of mixtures, actual testing using a HVS for rutting and cracking studies, analyses of collected pavement responses and generated data, and documentation of observations and lessons learned. A total of six experimental-testing lanes were constructed. These lanes featured the use of typical and high RAP contents, RA, softer binder, and WMA additive.



Figure 2-6¹. Virginia APT facility and field construction at VTTI.

2.8. Machine Learning

In the current landscape, characterized by the exponential growth of data and the continuous advancement of iterative algorithms, the reemergence of ML approaches presents a promising avenue for uncovering nonlinear interactions among constituent factors within large datasets, thereby enhancing predictive accuracy across various domains including pavement engineering. Some commonly used ML models are introduced below:

2.8.1. Support Vector Machines

Support Vector Machines (SVM) is first developed by Cortes and Vapnik (1995). It can virtually transform inherent non-linearly related data in higher dimension space and separate them. (Nonlinear relation hopefully tends to be separable in different dimensions from the original one). A quick illustration is shown in the following figure. This aim is achieved by incorporating kernel functions such as radial kernel and polynomial kernel functions. The kernel function will help to calculate the high dimensional relationship between each pair of nodes (data) so as to find the soft margins (Figure 2-7). It is noting the accuracy of the SVM is strongly affected by the type of kernel function you selected. Gungor and Al-Qadi already utilized this algorithm to develop ML model for airfield pavement response prediction (Gungor and Al-Qadi, 2018). Support Vector Regression (SVR) extends the concepts of SVM to solve regression problems, aiming to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data. Introduced by Vapnik et al. (1996), SVR is grounded in the principles of statistical learning theory and structural risk minimization, which seeks to minimize an upper bound of the generalization error rather than the training error. This approach allows SVR to achieve high generalization performance, making it robust to overfitting, especially in high-dimensional spaces.

¹ This figure is stored in VTTI CSRI project folder.

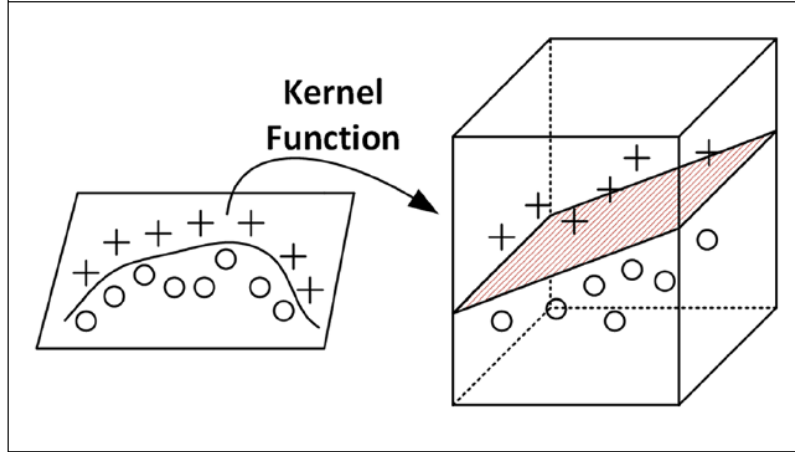


Figure 2-7: Separate 2-D data in 3-D dimension virtually (Gungor and Al-Qadi, 2018)

2.8.2. Artificial Neural Network (ANN)

The ANN algorithm has its roots in biological neuron systems. ANN algorithms are efficient at detecting nonlinear relationships between inputs and outputs. The architecture of an ANN consists of three layers: an input layer, a hidden layer, and an output layer, and each layer consists of neurons (nodes). Each node is connected to another neuron and has a weight and activation threshold associated with it. Training an ANN model involves adjusting the weights of all nodes by minimizing the error between outputs and actual targets. The activation threshold is described as an activation function. How well an ANN learns a dataset depends on the choice of the activation function. The ANN's architecture is mainly associated with the specific problem to be solved. A quick look at the specific architecture is shown below. Kim et al (2019) built the ANN-based predictive model for dynamic modulus. And ANN was used by Xiao et al., 2009 to predict pavement fatigue life (Figure 2-8).

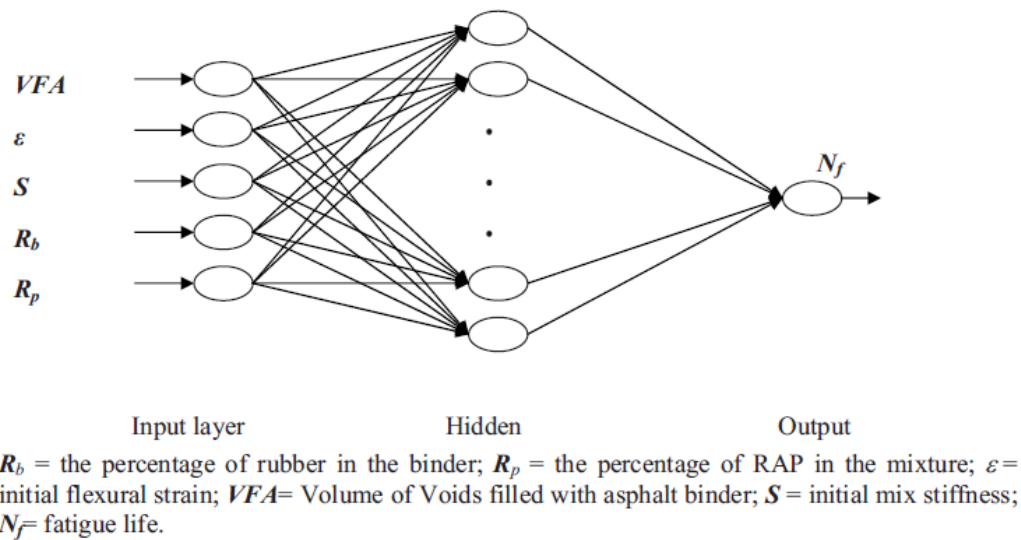


Figure 2-8. Example of three-layer ANN architecture (Xiao et al., 2009)

2.8.3. Random Forest

Random Forest (RF) stands out as a versatile and powerful ensemble learning method. It operates by constructing a multitude of decision trees during the training phase. The idea is to create an ensemble (forest) of trees where each tree is slightly different from the others. This diversity is achieved by using a random subset of the training data to build each tree and by randomly selecting a subset of features at each split in the construction of the trees (Breiman, 2001). During predictions, RF aggregates results by taking the majority vote from classification trees or averaging predictions from regression trees. This aggregation process helps in improving the predictive accuracy and mitigating over-fitting—a common problem in decision tree models. RF's versatility is a key strength, as it adeptly handles both categorical and continuous data while remaining robust to outliers and nonlinear data (Breiman, 2001). Furthermore, RF provides valuable insights into feature importance, facilitating the straightforward identification of the most influential variables contributing to the prediction process. Guo and Hao (2021) took RF to predict the location of potential damage on asphalt pavement. Rahman et al (2021) explored the RF to assess the indirect tensile strength. RF structure are represented in Figure 2-9.

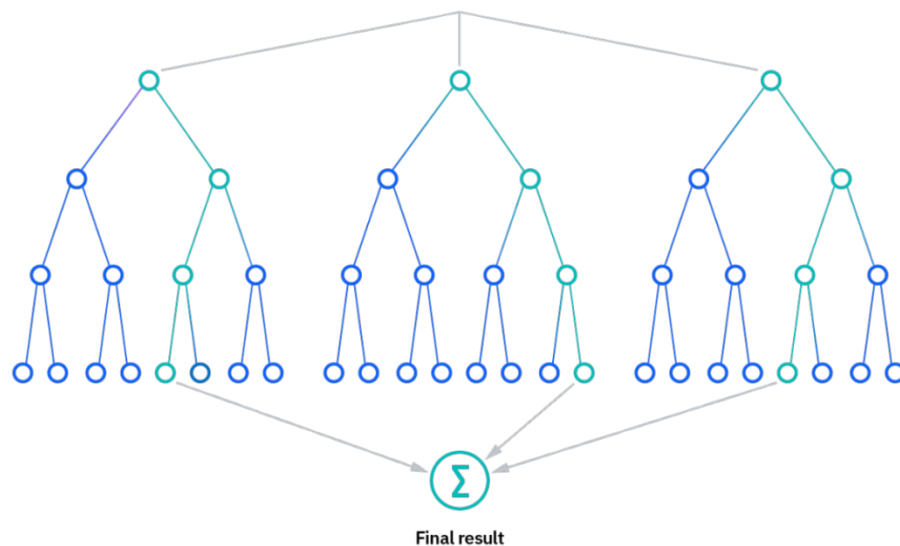


Figure 2-9. Random forest structure (Online resource).²

2.8.4. Extreme Gradient Boosting

XGBoost, short for Extreme Gradient Boosting, is an advanced and highly efficient implementation of gradient boosting machines (Chen and Guestrin, 2016). XGBoost has currently gained popularity in the machine learning community, emerging as a highly competitive algorithm. This acclaim stems from its adeptness in managing large-scale datasets, its exceptional predictive accuracy, and its fast, efficient, and open-source implementation. Moreover, XGBoost offers a commendable selection of default splitting criteria, as well as well-considered criteria for

² Online resource: <https://images.app.goo.gl/Fx86R7gCP5UK1Kch8>

determining when to stop splitting. Furthermore, it incorporates built-in regularization mechanisms designed to mitigate the risk of overfitting (Chen and Guestrin, 2016).

2.9. Summary of Literature Review

The literature review provided a comprehensive understanding of the current state of RAP usage across different regions and the implementation status of the BMD approach by various transportation agencies. The diverse range of performance tests, employing varying loading methods and mechanical principles, offers valuable tools for predicting the performance of designed asphalt mixtures in terms of rutting and cracking at the laboratory scale. The well-established application of APT further supports the validation of laboratory tests and field performance evaluations. Moreover, the widespread use of machine learning techniques and models in engineering has been recognized as a potential solution to address limitations in traditional methods and real-world challenges. With their powerful ability to explore complex nonlinear relationships and make accurate predictions, these techniques demonstrate significant potential for optimizing asphalt mixture design. Overall, the literature review offers detailed background, challenges, and methodological support for the research.

2.10. Reference

- Al-Qadi, I. L., Elseifi, M., & Carpenter, S. H. (2007). Reclaimed Asphalt Pavement—a Literature Review. *FHWA-ICT-07-001*.
- Al-Qadi, I. L., Ozer, H., Zhu, Z., Singhvi, P., Mohamed Ali, U., Sawalha, M., ... & Zehr, T. G. (2019). Development of Long-Term Aging Protocol for Implementation of the Illinois Flexibility Index Test (I-FIT). *FHWA-ICT-19-009*.
- American Association of State Highway and Transportation Officials. (2020). *Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards*. Washington, DC,
- Bairgi, B. K., Tarefder, R. A., Amanul Hasan, M., Mehedi Hasan, M., Gburi, M.A., & Khan, Z. H. (2022). Evaluation of Full-Depth Reclamation and Cold Central-Plant Recycling Mixtures for Laboratory Compaction, Mechanistic Response, and Performance Properties. *Transportation Research Record*.
- Bennert, T., Haas, E., Wass, E., & Berger, B. (2021). Indirect Tensile Testing for Balanced Mixture Design and Quality Control Performance Testing. In *Asphalt paving technology: association of asphalt paving technologists-proceedings of the technical sessions* (Vol. 89, pp. 363-389). Association of Asphalt Paving Technologist.
- Bonaquist, R. (2012). Evaluation of Flow Number (FN) as a Discriminating HMA Mixture Property.
- Bonaquist, R. F., Christensen, D. W. and Stump, W. (2003). NCHRP Report 513: Simple Performance Tester for Superpave Mix Design: First-Article Development and Evaluation. Transportation Research Board, Washington, D.C..
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

- Chen, T., & Guestrin, C. (2016). Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- Cortes, C., & Vapnik, V. N. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297.
- Cox, B. C., & Howard, I. L. (2015). Cold In-Place Recycling Characterization Framework and Design Guidance for Single or Multiple Component Binder Systems.
- Cross, S. A. (1999). Experimental Cold In-Place Recycling with Hydrated Lime. *Transportation Research Record*, 1684, 186 - 193.
- Diefenderfer, S. D., Habbouche, J., & Boz, I. (2023). *Balanced Mix Design for Surface Mixtures: 2020 Field Trials* (No. FHWA/VTRC 23-R13). Virginia Transportation Research Council (VTRC).
- Diefenderfer, S. D., Boz, I., & Habbouche, J. (2021a). Balanced Mix Design for Surface Asphalt Mixtures: Phase I: Initial Roadmap Development and Specification Verification. *Publication VTRC 21-R15*. Virginia Transportation Research Council, Charlottesville.
- Diefenderfer, S. D., Boz, I., and Habbouche, J. (2021b). Balanced Mix Design for Surface Asphalt Mixtures: 2019 Field Trials. *Publication VTRC 21-R21*. Virginia Transportation Research Council, Charlottesville.
- Du Plessis, L., Coetzee, N. F., Hoover, T. P., Harvey, J. T., & Monismith, C. L. (2006). Three Decades of Development and Achievements: The Heavy Vehicle Simulator in Accelerated Pavement Testing. In *Pavement Mechanics and Performance* (pp. 45-54).
- E-C280. (2022). Glossary of Terms for Balanced Mix Design of Asphalt Mixtures. Transportation Research Board. <https://onlinepubs.trb.org/onlinepubs/circulars/ec280.pdf>.
- Federal Highway Administration. (2019). *Cyclic Fatigue Index Parameter (Sapp) for Asphalt Performance Engineered Mixture Design*. FHWA-HIF-19-091. Washington, DC, 2019.
- Flintsch, G., Xue, W., Diefenderfer, B., & Meroni, F. (2020). Evaluation of Cold Central-Plant Recycling (CCPR) Technique Using Full-Scale Accelerated Pavement Testing. In *Accelerated Pavement Testing to Transport Infrastructure Innovation: Proceedings of 6th APT Conference* (pp. 270-279). Springer International Publishing.
- Garg, N., Kazmee, H., Ricalde, L., & Parsons, T. (2018). Rutting Evaluation of Hot and Warm Mix Asphalt Concrete under High Aircraft Tire Pressure and Temperature at National Airport Pavement and Materials Research Center. *Transportation Research Record*, 2672(23), 117-127.
- Ghanbari, A., Underwood, B. S., & Kim, Y. R. (2022). Development of a Rutting Index Parameter Based on the Stress Sweep Rutting Test and Permanent Deformation Shift Model. *International Journal of Pavement Engineering*, 23(2), 387-399.
- Greene, J., Chun, S., Nash, T., & Choubane, B. (2015). Evaluation and Implementation of PG 76-22 Asphalt Rubber Binder in Florida. *Transportation Research Record*, 2524(1), 3-10.
- Gungor, O. E., & Al-Qadi, I. L. (2018). Developing Machine-Learning Models to Predict Airfield Pavement Responses. *Transportation Research Record*, 2672, 23 - 34.

- Guo, X., & Hao, P. (2021). Using a Random Forest Model to Predict the Location of Potential Damage on Asphalt Pavement. *Applied Sciences*, 11(21), 10396.
- Habbouche, J., Boz, I., & Diefenderfer, S. D. (2022). Validation of Performance-Based Specifications for Surface Asphalt Mixtures in Virginia. *Transportation Research Record*, 2676(5), 277-296.
- Habbouche, J., Boz, I., Underwood, B.S., Castorena, C., Gulzar, S., Fried, A., & Preciado, J. A. (2021a). Review from Multiple Perspectives for the State of the Practice on the Use of Recycled Asphalt Materials and Recycling Agents in Asphalt Concrete Surface Mixtures. *Transportation Research Record*, 2676 (4): 407 – 420.
- Harvey, J., & Popescu, L. (2000). Accelerated Pavement Testing of Rutting Performance of Two Caltrans Overlay Strategies. *Transportation Research Record*, 1716(1), 116-125.
- Harvey, J., Wu, R., Signore, J., Basheer, I., Holikatti, S., Vacura, P., & HOLLAND, T. (2014). Performance-Based Specifications. *Application of Asphalt Mix Performance-Based Specifications*, 1.
- Huang, B., Zhang, Z., Kingery, W., & Zuo, G. (2004, May). Fatigue Crack Characteristics of HMA Mixtures Containing RAP. In *Proceedings, 5th International RILEM Conference on Cracking in Pavements, Limoges, France*.
- Izaks, R., Haritonovs, V., Klasa, I., & Zaumanis, M. J. P. E. (2015). Hot Mix Asphalt with High RAP Content. *Procedia Engineering*, 114, 676-684.
- Kim, C., Lee, J. Y., & Kim, M. (2019). Prediction of the Dynamic Stiffness of Resilient Materials Using Artificial Neural Network (ANN) Technique. *Applied Sciences*, 9(6), 1088.
- Kim, S., Byron, T., Sholar, G. A., & Kim, J. (2007). Evaluation of Use of High Percentage of Reclaimed Asphalt Pavement (RAP) for Superpave Mixtures. *Florida University*.
- Kim, Y., Lee, H. D., & Heitzman, M. (2009). Dynamic Modulus and Repeated Load Tests of Cold In-Place Recycling Mixtures Using Foamed Asphalt. *Journal of Materials in Civil Engineering*, 21(6), 279-285.
- Ma, Z., Liu, L., & Sun, L. (2018). Investigation of Top-Down Cracking Performance of In-Situ Asphalt Mixtures Based on Accelerated Pavement Testing and Laboratory Tests. *Construction and Building Materials*, 179, 277-284.
- Martin, A. E., Kaseer, F., Arámbula-Mercado, E., Bajaj, A., Cucalon, L. G., Yin, F., ... & King, G. (2020). Evaluating the Effects of Recycling Agents on Asphalt Mixtures with High RAS and RAP Binder Ratios (No. Project 09-58).
- McDaniel, R. S., & Anderson, R. M. (2001). Recommended Use of Reclaimed Asphalt Pavement in the Superpave Mix Design Method: Technician's Manual (No. Project D9-12 FY'97). National Research Council (US). Transportation Research Board.
- McDaniel, R. S., Shah, A., Huber, G. A., & Copeland, A. (2012). Effects of Reclaimed Asphalt Pavement Content and Virgin Binder Grade on Properties of Plant Produced Mixtures. *Road Materials and Pavement Design*, 13(sup1), 161-182.
- Meroni, F., Flintsch, G. W., Habbouche, J., Diefenderfer, B. K., & Giustozzi, F. (2021). Three-

- Level Performance Evaluation of High RAP Asphalt Surface Mixes. *Construction and Building Materials*, 309, 125164.
- Meroni, F., Xue, W., Flintsch, G. W., & Diefenderfer, B. K. (2020). Impact of Mix Design Optimization on HMA Rutting Performance under Accelerated Pavement Testing. In *Accelerated Pavement Testing To Transport Infrastructure Innovation: Proceedings of 6th APT Conference* (pp. 108-117). Springer International Publishing.
- Mohajeri, M. (2015). Hot Mix Asphalt Recycling: Practices and Principles. *Doctoral Thesis, TU Delft*, Netherland.
- Mohammad, L. N., Kim, M., & Challa, H. (2016). Development of Performance-Based Specifications for Louisiana Asphalt Mixtures (No. FHWA/LA. 14/558). Louisiana. Department of Transportation and Development.
- Nair, H., Diefenderfer, S. D., & Bowers, B. F. (2019). Assessing Increased Use of Reclaimed Asphalt Pavement in Asphalt Mixtures (No. FHWA/VTRC 20-R9). Virginia Transportation Research Council (VTRC).
- National Asphalt Pavement Association. (2020). Online Resource: Implementation Effort. Website: <https://www.asphaltpavement.org/expertise/engineering/resources/bmd-resource-guide/implementation-efforts>.
- Online resource: CAYD project - APT Test on Inductive Charging System. <https://ease.univ-gustave-eiffel.fr/en/research/research-projects/current-projects>. University of Gustave Eiffel.
- Rahman, S., Bhasin, A., & Smit, A.D. (2021). Exploring the Use of Machine Learning to Predict Metrics Related to Asphalt Mixture Performance. *Construction and Building Materials*, 295, 123585.
- Roberts, F. L., Kandhal, P. S., Brown, E. R., Lee, D. Y., & Kennedy, T. W. (1996). Hot Mix Asphalt Materials, Mixture Design and Construction.
- Rust, F. C., Smit, M. A., Akhalwaya, I., Jordaan, G. J., & Du Plessis, L. (2022). Evaluation of Two Nano-Silane-Modified Emulsion Stabilised Pavements Using Accelerated Pavement Testing. *International Journal of Pavement Engineering*, 23(5), 1339-1352.
- Steyn, W. J. (2012). Significant Findings from Full-Scale Accelerated Pavement Testing (Vol. 433). Transportation Research Board.
- Ullidtz, P., Harvey, J., Tsai, B. W., & Monismith, C. (2005). Calibration of Incremental-Recursive Flexible Damage Models in CalME Using HVS Experiments.
- Underwood, B. S., Kim, Y. R., & Guddati, M. N. (2010). Improved Calculation Method of Damage Parameter in Viscoelastic Continuum Damage Model. *International Journal of Pavement Engineering*, 11(6), 459-476.
- Vapnik, V.N., Golowich, S.E., & Smola, A. (1996). Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. Neural Information Processing Systems.
- Wang, Y. D., Underwood, B. S., & Kim, Y. R. (2022). Development of a Fatigue Index Parameter, Sapp, for Asphalt Mixes Using Viscoelastic Continuum Damage

- theory. *International Journal of Pavement Engineering*, 23(2), 438-452.
- West, R. C., Willis, J. R., & Marasteanu, M. O. (2013). Improved Mix Design, Evaluation, and Materials Management Practices for Hot Mix Asphalt with High Reclaimed Asphalt Pavement Content (Vol. 752). Transportation Research Board.
- West, R., Rodezno, C., Leiva, F., & Yin, F. (2018). Development of A Framework for Balanced Mix Design. *Project NCHRP*, 20-07.
- West, R.C., & Copeland, A. (2015). High RAP Asphalt Pavements: Japan Practice – Lessons Learned. Information Series 139, *National Asphalt Pavement Association*, Lanham, MD.
- Widyatmoko, I. (2008). Mechanistic-Empirical Mixture Design for Hot Mix Asphalt Pavement Recycling. *Construction and Building materials*, 22(2), 77-87.
- Williams, B. A., Willis, J. R., & Shacat, J. (2020). Asphalt Pavement Industry Survey on Recycled Materials and Warm-Mix Asphalt Usage: 2019. *National Asphalt Pavement Association*, Greenbelt, Md.
- Witczak, M., Schwartz, C. and Von Quintus, H. (2011). NCHRP Project 9-19: Superpave Support and Performance Models Management. Transportation Research Board, Washington, D.C..
- Xiao, F., Amirkhanian, S., & Juang, C. H. (2009). Prediction of Fatigue Life of Rubberized Asphalt Concrete Mixtures Containing Reclaimed Asphalt Pavement Using Artificial Neural Networks. *Journal of Materials in Civil Engineering*, 21(6), 253-261.
- Xue, W., Flintsch, G. W., & Diefenderfer, B. K. (2020). Measuring Pavement Permanent Deformation in Accelerated Pavement Testing. *Transportation Research Record*, 2674(5), 340-348.
- Xue, W., Flintsch, G. W., Diefenderfer, B. K., & Bowers, B. F. (2017). Characterization and Modeling of Recycled Pavement Sections (No. CAIT-UTC-NC29). Rutgers University. Center for Advanced Infrastructure and Transportation.
- Yin, F., & West, R. C. (2021). Balanced Mix Design Resource Guide (No. IS-143).
- Zaumanis, M., Mallick, R. B., & Frank, R. (2013). Evaluation of Rejuvenator's Effectiveness with Conventional Mix Testing for 100% Reclaimed Asphalt Pavement Mixtures. *Transportation Research Record*, 2370(1), 17-25.
- Zhou, F., Hu, S., & Newcomb, D. (2020). Development of a Performance-Related Framework for Production Quality Control with Ideal Cracking and Rutting Tests. *Construction and Building Materials*, 261, 120549.
- Zhou, F., Hu, S., & Scullion, T. (2006). *Integrated Asphalt (Overlay) Mixture Design, Balancing Rutting and Cracking Requirements* (No. FHWA/TX-06/0-5123-1). Texas Transportation Institute, Texas A & M University System.

CHAPTER 3 - Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia³

3.1. Abstract

This study investigated one control and five BMD optimized asphalt surface mixtures, four of which had high RAP contents (HRAP mixtures), using laboratory performance tests characterized with different levels of complexity. The performance of the evaluated mixtures was assessed based on durability, rutting resistance, and cracking resistance as emphasized by BMD. The study explored the ranking of a single index and correlations among various indices. Assisted by 3-D and ternary plots, this study also proposed a novel CPI that combines major indices (durability, cracking, and rutting) to evaluate the performance of BMD optimized mixtures. The results revealed discrepancies between basic/intermediate performance test results and advanced performance test results. The comparisons conducted also underscored the beneficial impacts derived from using softer binders and/or recycling agents in HRAP mixtures. Furthermore, the findings indicated that the BMD approach can serve as an effective framework for designing asphalt mixtures that simultaneously enhance both fatigue cracking and rutting performance. Moreover, the study revealed HRAP BMD surface mixtures can exhibit superior overall performance when compared to conventionally designed control mixtures.

Keywords: reclaimed asphalt pavement, high RAP, asphalt concrete, balanced mix design, performance, implementation, accelerated pavement testing, correlations, recycling agents.

³ This paper has been published in the *International Journal of Pavement Engineering*. Citation: Tong, B., Habbouche, J., Diefenderfer, S. D., & Flintsch, G. W. (2024). Multi-level performance evaluation of BMD surface mixtures with conventional and high RAP contents: a case study in Virginia. *International Journal of Pavement Engineering*, 25(1), 2325609. Contribution: Conceptualization: Lead; Methodology: Lead; Resources: Equal; Data Processing and Analysis: Lead; Laboratory testing: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead.

3.2. Objective and Scope

The 2020 APT testing lanes were among the trials that served as the first application of the BMD special provisions to production. They provided an opportunity to evaluate the impact of the specifications on the design, production, quality control and assurance practices, and construction of SMs. The specific objectives of this paper are to document the process of developing a control mixture and five engineered mix designs using the BMD concept and to provide a comprehensive laboratory evaluation of mixtures sampled during production using multiple levels of performance testing complexity (basic, intermediate, and advanced). Subsequently, this study presents the ranking results based on singular performance tests. Additionally, an in-depth analysis is conducted to investigate the relationships and distinctions among various laboratory performance test results, aiming to uncover potential discrepancies that may arise when different levels of performance tests are used. This investigation extends beyond a singular performance index and involves the exploration of optimized mixtures based on the BMD design by examining the correlations between primary performance tests and recommending a novel CPI. As a visual aid, a 3-D plot, complemented by a ternary plot, is proposed to effectively depict the internal "balance" state among the three primary performance indices. The utilization of this visualization tool allows for a thorough understanding of the impact of BMD optimization efforts on the evaluated mixtures.

3.3. Experimental Program

A total of one control and five BMD 9.5mm dense-graded SMs were designed, produced, and sampled for evaluation in this study. The mixtures incorporated various combinations of RAP contents, two binder PGs, one RA, and one WMA additive. 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) (West et al., 2018; Yin and West, 2021). The design volumetric properties and gradations of all mixtures are shown in Table 3-1. The BMD performance properties of the six mixtures at OBC determined during design are also shown in Table 3-1. 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 3-1. 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, 25°C, %	2.9	3.2	4.1	2.8	2.7	4.0
APA Rut Depth, 64°C, mm	5.4	4.2	5.6	7.2	4.9	3.7
CT index, 25°C	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.

The volumetric properties and gradations of all mixtures collected during one sampling during plant production are shown in Table 3-2. Note that corresponding plant-produced specimens were fabricated from reheated loose mixture sampled in boxes during production. The process of preparing reheated specimens involved reheating the loose mixture in boxes until workable, splitting the material into specimen quantities, and heating to the appropriate compaction temperature and compacting. Mixtures were not subjected to any additional oven aging.

According to Table 3-1, the major change noted among mixtures was that the selected OBCs for the BMD mixtures were higher when compared to the non-BMD control mixture 30_C, which could be attributed to the need for increased asphalt binder to meet the cracking performance criterion, a CT index of 70 for short-term aged mixtures. The use of softer binder and/or RA (mixtures 45_HR_RA, 45_HR_L, and 60_HR_L_RA) resulted in a lower OBC when compared with mixture 45_HR in which none of these alternatives were employed to address the use of high RAP content. Finally, mixture 45_HR exhibited the highest sensitivity to asphalt binder content among all evaluated mixtures. During the development of the mixture design, it was observed that a slight change in the binder content (e.g., $\pm 0.2\%$) may result in passing or failing CT index values.

During production (refer to Table 3-2), all evaluated mixtures except mixture 45_HR had binder contents similar to or slightly deviated from the design and within the allowable tolerance according to VDOT specifications. However, an increase of almost 0.4% in binder content from the design phase was observed for mixture 45_HR. No major changes in aggregate gradations between design and production were observed for any evaluated mixture regardless of the RAP content. Finally, the extracted and recovered asphalt binder from the produced mixtures (in accordance with AASHTO T 164 and AASHTO T 319) exhibited similar low and high PG temperatures except for mixture 45_HR; this mixture showed a much lower high temperature PG which could be attributed to the dominance of the virgin binder due to the relatively higher binder content. The subsequent sections primarily focus on the assessment and discussion of the performance of one set of reheated specimens taken from the APT production. In total, there were four sets of materials obtained from the production process. One of these sets (Table 3-2) was employed for conducting multi-level laboratory tests, while the remaining sets were only used for VDOT BMD tests (Tong *et al.* 2023). A comprehensive analysis of the reheating effect and the

variability in volumetric properties from design to production will be addressed in future publications.

Table 3-2. Volumetric Properties and Gradations for APT Mixtures During Production

Mixture ID	30_C	30_O	45_HR	45_HR_RA	45_HR_L	60_HR_L_RA
Description	Non-BMD	BMD	BMD	BMD	BMD	BMD HRAP
Composition						
RAP Content, %	30	30	45	45	45	60
Asphalt Binder	PG64S-22	PG64S-22	PG64S-22	PG64S-22	PG58-28	PG58-28
Additives	WMA	WMA	WMA	WMA + RA	WMA	WMA + RA
Property						
Ndesign, gyrations	50	50	50	50	50	50
NMAS, mm	9.5	9.5	9.5	9.5	9.5	9.5
Asphalt Content, %	5.44	6.14	7.22	6.15	6.00	6.14
Rice SG (G_{mm})	2.604	2.515	2.505	2.546	2.544	2.548
VTM, %	6.0	4.6	0.6	2.6	2.4	1.7
VMA, %	18.6	18.4	17.5	16.8	16.5	15.8
VFA, %	67.5	75.2	96.5	84.6	85.7	89.5
FA Ratio	1.12	1.07	1.21	1.39	1.21	1.35
Mixture Bulk SG (G_{mb})	2.448	2.400	2.490	2.480	2.482	2.506
Aggregate Effective SG	2.855	2.777	2.820	2.817	2.807	2.820
Aggregate Bulk SG	2.841	2.761	2.799	2.797	2.795	2.794
Absorbed Asphalt	0.18	0.21	0.27	0.26	0.16	0.34
Effective Asphalt	5.27	5.94	6.97	5.90	5.85	5.82
Effective Film Thickness	9.2	10.9	10.8	9.1	9.7	8.7
PG of E&R Binder	76.9-18.3	77.8-20.6	70.8-21.4	76.9-21.3	79.3-18.8	76.4-20.6
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)	99.3	99.5	98.0	98.5	98.7	98.4
3/8 in (9.5 mm)	93.7	94.3	91.9	94.2	93.8	93.4
No. 4 (4.75 mm)	64.5	65.4	61.1	66.8	63.3	66.5
No. 8 (2.36 mm)	39.9	41.1	39.0	42.4	41.9	43.6
No. 16 (1.18 mm)	29.4	25.4	25.7	27.0	27.2	29.9
No. 30 (600 µm)	23.5	16.6	18.4	18.8	19.1	21.8
No. 50 (300 µm)	13.9	11.2	13.6	13.4	13.0	15.2
No. 100 (150 µm)	8.2	8.2	10.6	10.2	9.2	10.5
No. 200 (75 µm)	5.9	6.4	8.4	8.2	7.1	7.9

PG = performance grade; S = standard traffic; BMD = Balanced Mix Design; HRAP = high reclaimed asphalt pavement content mixture; RAP = reclaimed asphalt pavement; WMA = warm mix additives; RA = recycling agent; NMAS = nominal maximum aggregate size; SG = specific gravity; VTM = voids in total mixture; VMA = voids in mineral aggregate; VFA = voids filled with asphalt; FA = fines to aggregate; PG = performance grade; E&R = extracted and recovered.

3.4. Durability Assessment

The Cantabro test is a basic test that was performed to assess the durability of asphalt mixtures (Cox *et al.* 2017) in accordance with AASHTO T 401. Three replicates were tested for each mixture, and an average mass loss was reported. The CML for each specimen was calculated

as the difference between the initial mass and the mass after testing divided by the initial mass (Cox *et al.* 2017).

Figure 3-1 and Table 3-3 shows the CML data for all evaluated mixtures. The box plot displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum. The mean CML ranged from 3.6% to 9.5% with a COV ranging from 1.9 to 9.8%. All mixtures except mixtures 30_C and 30_O met the 7.5% VDOT BMD mass loss criterion. Overall, all BMD mixtures exhibited lower CML values when compared with mixture 30_C. This can be attributed to the relatively higher binder content resulting from the application of BMD concept. Moreover, HRAP mixtures (mixtures 45_HR, 45_HR_RA, 45_HR_L, and 60_HR_L_RA) exhibited a higher variability when compared with mixtures 30_C and 30_O. This could be attributed to the variability induced with the use of higher RAP contents. Based on the results from the mixtures tested in this study, the CML test was able to capture the impact of BMD on the durability performance of asphalt mixtures.

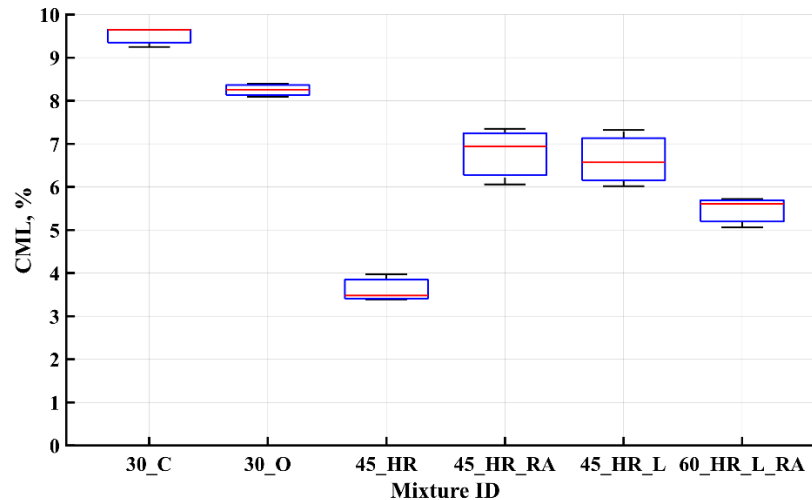


Figure 3-1. Performance Test Data for Cantabro Mass Loss. The box plot displays the five-number summary of a set of data. The red horizontal lines in the box are the median values. CML = Cantabro mass loss.

Table 3-3. Performance Test Results and Corresponding Statistical Parameters of Evaluated Mixtures Collected During Production

Mixture ID / Test Parameters	30_C	30_O	45_HR	45_HR_RA	45_HR_L	60_HR_L_RA
<i>Cantabro Mass Loss Test at 25°C - CML</i>						
Average, %	9.5	8.3	3.6	6.8	6.6	5.5
Stdv, %	0.2	0.2	0.3	0.7	0.7	0.4
COV, %	2.4	1.9	8.6	9.7	9.8	6.5
<i>IDT-CT at 25°C – CT index</i>						
Average	54	46	513	90	82	84
Stdv	8.0	8.3	38.5	14.7	11.0	12.4
COV, %	14.8	18.1	7.5	16.3	13.4	14.7
<i>APA at 64°C and after 8,000 loading cycles – Rut Depth</i>						

Average, mm	4.0	4.6	6.4	5.7	4.2	5.9
Stdv, mm	0.6	1.1	0.5	1.4	1.0	1.3
COV, %	13.9	24.6	7.9	24.7	24.5	22.7

Stdv = standard deviation; COV = coefficient of variation; IDT-CT = indirect tensile cracking test; CT = cracking tolerance; and APA = asphalt pavement analyzer.

3.5. Cracking Performance Assessment

Three performance tests were considered to assess the resistance of mixtures to cracking. These tests belong to two levels of testing: basic (IDT-CT) and advanced (dynamic modulus $|E^*|$ test and the direct tension CF) (Meroni *et al.* 2021).

3.5.1. IDT Cracking Test and Results

IDT-CT is a basic test commonly used by state agencies and was incorporated into the BMD approach for VDOT to evaluate mixture resistance to cracking. The test was conducted at 25°C accordance with ASTM D8225. The CT index was then calculated from the test load-displacement curve / data using Equation 3-1. Five specimens were tested for each mixture and an average CT index was reported.

$$CT\ index = \frac{G_f}{|m_{75}|} \cdot \frac{l_{75}}{D} \cdot \frac{t}{62} \quad \text{Equation 3-1}$$

$$m_{75} = \left| \frac{p_{85} - p_{65}}{l_{85} - l_{65}} \right| \quad \text{Equation 3-2}$$

Where G_f = total area under the load-displacement curve divided by the product of the specimen thickness $[t]$ and diameter $[D]$ in kN/mm, m_{75} = slope of interest expressed in Equation 3-2, p_{85} = 85% of the peak load (P_{max}) at the post-peak stage in kN, p_{75} = 75% of P_{max} at the post-peak stage in kN, p_{65} = 65% of P_{max} at the post-peak stage in kN, l_{85} = displacement corresponding to p_{85} in mm, l_{75} = displacement corresponding to p_{75} in mm, l_{65} = displacement corresponding to p_{65} in mm, D = specimen diameter in mm, and t = specimen thickness in mm.

Figure 3-2a shows the CT index values for the six evaluated mixtures. Higher CT index values generally indicate a better cracking resistance of the evaluated mixtures and, hence, less cracking potential in the field. Mixture 45_HR exhibited a significantly higher CT index average (513) when compared with the other mixtures. This could be attributed to the significantly higher flexibility induced with the high binder content used in the design and production of mixture 45_HR. Consequently, the specimens of mixture 45_HR fabricated for IDT-CT testing exhibited a wet nature. For a clearer comparison, Figure 3-2b shows the CT index values of the mixtures after excluding mixture 45_HR. The mean CT index ranged from 46 to 90, with a COV ranging from 7.5% to 18.1%. In fact, all evaluated mixtures exhibited COV values lower than the maximum allowable single operator COV of 18.3%, indicating an acceptable repeatability of the measured data (Habbouche *et al.* 2021b, Habbouche *et al.* 2022a). Mixtures 30_C and 30_O exhibited statistically similar CT index data indicating a possible similar cracking performance

regardless of the mix design methodology (i.e., Superpave vs. BMD). Mixtures 45_HR_RA, 45_HR_L, and 60_HR_L_RA exhibited higher CT index values when compared with mixtures 30_C and 30_O. This indicated that HRAP mixtures containing softer binders and/or an RA can provide performance that is equal to or better than the performance of their counterpart mixtures with 30% RAP. Moreover, upon comparing the CT index values of specimens fabricated during the design phase (Table 3-1) and the reheated specimens fabricated during the production phase (Table 3-3), it was evident that the reheating process possesses the potential to decrease the CT index, with the exception of mixture 45_HR. This reduction can be attributed to the aging effect introduced by the reheating process. Nevertheless, a more detailed discussion regarding the impact of the reheating effect, as well as the variability in volumetric properties from design to production, is reserved for future work.

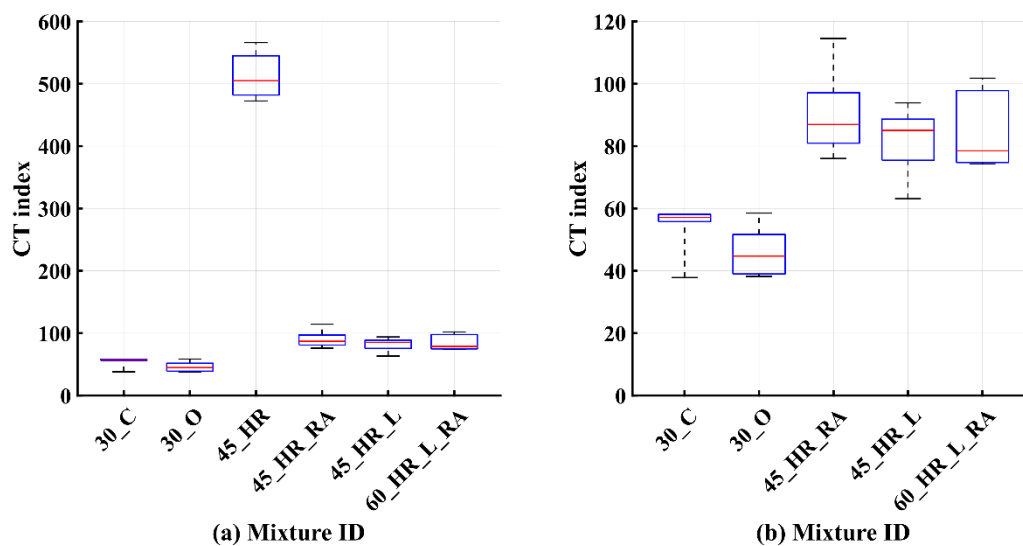


Figure 3-2. Performance Test Data for IDT-CT Index: (a) Including Mixture 45_HR and (b) Excluding Mixture 45_HR. The box plot displays the five-number summary of a set of data. The red horizontal lines in the box are the median values. CT = cracking tolerance.

3.5.2. Dynamic Modulus $|E^*|$ Test and Results

Dynamic modulus tests were performed using an AMPT machine in accordance with AASHTO T 378. Three test temperatures (4.4, 21.1, and 37.8°C) and six test frequencies ranging from 0.1 to 25 Hz were used. All tests were conducted in the uniaxial mode without confinement. Three replicate specimens were tested for each mixture, and results at each temperature-frequency combination were reported.

The dynamic modulus ($|E^*|$) and phase angle (δ) of the six mixtures are presented in Figure 3-3a and Figure 3-3b, respectively. The data in these figures were constructed at a reference temperature of 21.1°C using the 2S2P1D model (for both $|E^*|$ and δ data). The general model, known as 2S2P1D, was initially introduced in 2003 (Olard and Benedetto 2003). After a comprehensive comparison with various existing methods for fitting the measured dynamic

modulus, it was observed that the 2S2P1D model, when combined with a second-order polynomial shift factor function, proved capable of accurately predicting the measured data at low temperature (Keshavarzi 2019). A higher $|E^*|$ value at higher temperatures (and lower frequencies) is often attributed to a potential higher rutting resistance of asphalt mixtures. The data in Figure 3-3a shows that mixture 45_HR had significantly lower $|E^*|$ values when compared with the remaining mixtures regardless of the frequency of interest. This is attributed to the high binder content and wet nature of mixture 45_HR and could indicate a potential higher rutting susceptibility of the mixture. All remaining mixtures had similar $|E^*|$ values at intermediate frequencies. Among the evaluated mixtures, except for mixture 45_HR, mixture 60_HR_L_RA exhibited the lowest stiffness at lower frequencies, which can be attributed to the incorporation of a softer binder and RA. Figure 3-3b shows significantly higher δ values for mixture 45_HR at relatively higher reduced frequencies indicating a potential higher cracking resistance among the six mixtures. For the remaining mixtures, visual inspection of Figure 3-3b shows that mixture 60_HR_L_RA and mixture 30_O have the highest and lowest δ values at relatively lower frequencies, respectively. A higher $|E^*|$ value and lower δ at lower temperatures (and higher frequencies) are often associated with the potential for higher cracking susceptibility (Diefenderfer *et al.* 2021).

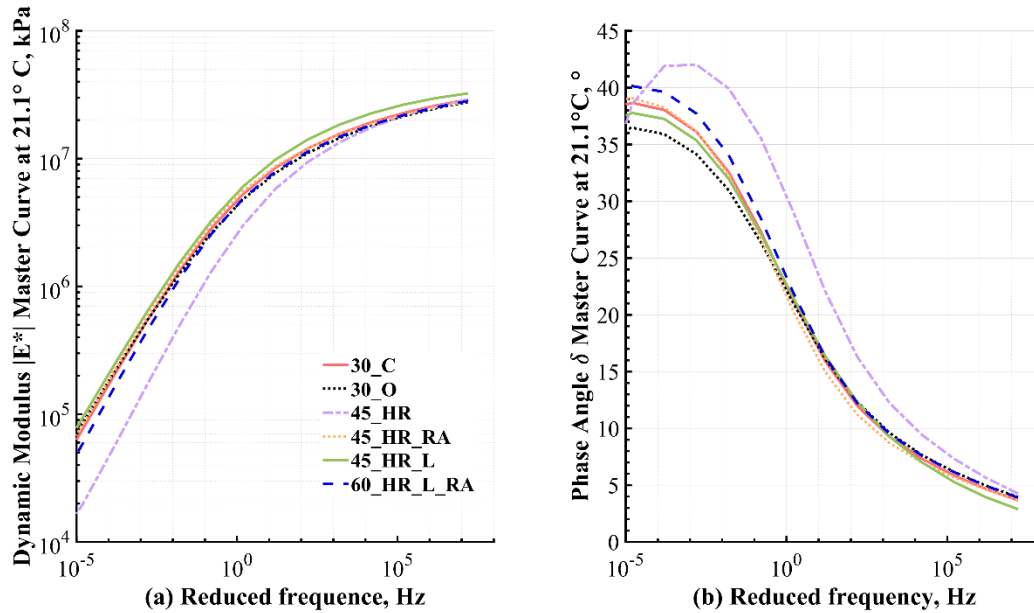


Figure 3-3. Dynamic Modulus Test Results: (a) $|E^*|$ Master Curves in Log-log Scale and (b) δ Master Curves in Semi-log Scale. $|E^*|$ = dynamic modulus; δ = phase angle.

Another performance indicator parameter that can be calculated from the $|E^*|$ and δ master curves is the mixture G-Rm parameter, which is an adapted form of the binder Glover-Rowe (G-R) parameter (Mensching *et al.* 2017, Ogbo *et al.* 2019, Zhang *et al.* 2020). The G-Rm parameter is simply the same as the binder G-R parameter, but it uses mixture dynamic modulus ($|E^*|$) and phase angle (δ) in lieu of the binder complex shear modulus and phase angle as expressed in Equation 3-3. This parameter, calculated at 20°C and 5 Hz, has been used to evaluate the resistance

of mixtures to cracking and their susceptibility to aging, especially with changes in recycled content (e.g., RAP) and RA type and dosage (Martin *et al.* 2020). In a draft AASHTO specification proposed in NCHRP Project 09-58, this parameter was recommended as a mixture performance evaluation parameter at intermediate temperatures to minimize the cracking potential of asphalt mixtures (Martin *et al.* 2020). A maximum G-Rm parameter value of 8,000 MPa was recommended for laboratory-made mixtures subjected to short-term oven-aging for 2 hours at the specified compaction temperature. A maximum G-Rm parameter value of 19,000 MPa was recommended for long-term oven-aged laboratory-made mixtures in accordance with either AASHTO R 30 or the proposed standard method for long-term conditioning of hot mix asphalt for performance testing in NCHRP Project 09-54 (Kim *et al.* 2018).

$$G - Rm = \frac{|E^*| \times (\cos \delta)^2}{\sin \delta} \quad \text{Equation 3-3}$$

Figure 3-4 shows the G-Rm parameter calculated for the mixtures used in this study. The G-Rm parameter values for all mixtures except mixture 45_HR are more comparable to the recommended threshold value (19,000 MPa) for long-term oven-aged laboratory-made mixtures than that for short-term oven-aged laboratory-produced mixtures (8,000 MPa). Since these mixtures were plant-produced and subjected to reheating, a process known to result in higher stiffness, neither threshold is directly applicable to the mixtures as no recommendations have been made for plant-produced mixtures. A relative comparison among the mixtures indicated that mixtures 45_HR_L and 45_HR ranked as the most and least crack-susceptible mixtures, respectively. Nevertheless, through the application of one-sample t-test, it was determined that there was no statistically significant difference in the G-RM value between the control mixture (30_C) and BMD mixtures (p-value of 0.4379). Similarly, no significant difference in the G-Rm value was observed between 30% RAP mixtures (30_C and 30_O) and HRAP mixtures (p-value of 0.7116).

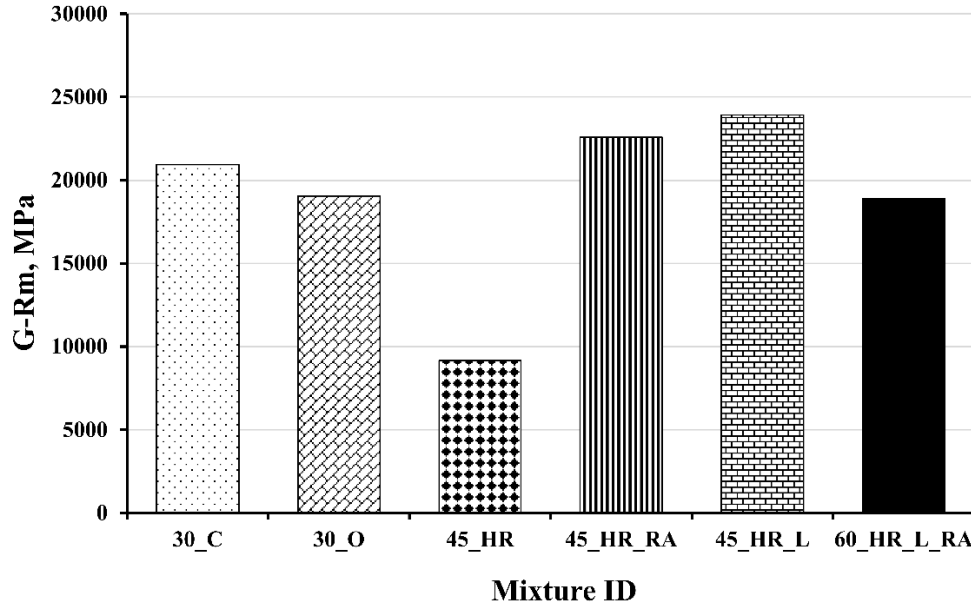


Figure 3-4. Glover-Rowe (G-Rm) Parameter at 20°C and 5 Hz

3.5.3. Direct Tension Fatigue Test and Results

Direct tension cyclic fatigue tests were performed using the AMPT machine in accordance with AASHTO TP 107. The S-VECD model was used to assess the fatigue characteristics of the six mixtures. The cyclic fatigue test was performed at 24°C. The S-VECD model developed by Underwood et al. was used to predict the cumulative damage for different loading histories under different temperatures as it incorporates time-temperature superposition principle and the pseudo strain energy concept (Underwood *et al.* 2010). The pseudo strain energy is dependent on both pseudo strain and pseudo stiffness; the pseudo strain was determined using Equation 3-4.

$$\varepsilon^R = \frac{1}{E^R} \int_0^t E(t - \tau) \frac{d\varepsilon}{d\tau} d\tau \quad \text{Equation 3-4}$$

Where ε^R = pseudo strain; t = time; τ = reduced time; $E(t)$ = relaxation modulus; ε = strain; and E^R = the reference modulus.

Based on assumptions from S-VECD theory, in the quasi-static cyclic status, the simplified ε^R is shown in Equation 3-5. The pseudo stiffness (C), presented in Equation 3-6, gradually decreases from 1 to nearly 0 and indicates the material condition from its intact state to failure. With the pseudo energy concept, the relationship between C and S , known as damage characteristic curve (C-S curve), can be determined using in Equation 3-7.

$$\varepsilon^R = |E^*| \cdot \varepsilon \quad \text{Equation 3-5}$$

$$C = \frac{\sigma}{\varepsilon^R} \quad \text{Equation 3-6}$$

$$C = 1 - C_{11}S^{C_{12}} \quad \text{Equation 3-7}$$

Where C_{11} and C_{12} are the model coefficients.

Figure 3-5 shows the damage characteristic curves for the six mixtures. The damage characteristic curve is strain independent for a given mixture and its position captures the mixture's stiffness. Similar observations to the dynamic modulus $|E^*|$ can be derived with a visual observation of the damage characteristic curves. Mixtures 45_HR and 45_HR_L showed the lowest and highest damage curves indicating a relatively softer and stiffer behaviors, respectively; meanwhile the remaining mixtures exhibited similar curves. From the perspective of stiffness alone, a material that is stiffer is expected to generate a lower strain under the same load conditions. Assuming other properties such as toughness are equal, a smaller strain in the stiffer material will ultimately lead to a longer fatigue life (Wang *et al.* 2020).

However, relying solely on stiffness evaluation may often lack persuasiveness when considering the fatigue life of materials. Consequently, the inclusion of material toughness as a complementary factor becomes necessary. The toughness is the ability of a material to absorb energy without fracturing. I.e, a tougher material is expected to exhibit a longer fatigue life under the same applied load. The D^R failure criterion, determined using Equation 3-8 can be applied to the S-VECD model. D^R can serve as an indicator of toughness for the evaluated material, as demonstrated in Figure 3-5b, where a higher value signifies greater resistance of the material to damage (Wang *et al.* 2020).

$$D^R = \frac{\int_0^{N_f} (1-C)dN}{N_f} = \frac{\text{sum}(1-C)}{N_f} \quad \text{Equation 3-8}$$

To define the mixtures' fatigue performance, a fatigue cracking index parameter (S_{app}), considered both the material's modulus and toughness, is used. The calculation of S_{app} can be found in Equation 3-9 (Wang *et al.* 2020) and conducted with FlexMAT™ for Cracking, an Excel-based tool provided by the FHWA (FHWA-HIF-19-091). Figure 3-6 shows the S_{app} values for the six mixtures. S_{app} was calculated at the average temperature of the high and low PG of the conventional asphalt binder typically used in the area (21°C for PG 64S-22). The six mixtures fell into the same category when compared to the recommended S_{app} threshold values and the associated traffic level and ranges (i.e., Standard traffic). All mixtures displayed S_{app} values ranging from 11.9 for mixture 45_HR to 19.8 for mixture 45_HR_L. Furthermore, upon closer examination, it becomes apparent that S_{app} accounts for considerations from both the stiffness and toughness perspectives. Notably, mixture 30_C exhibited higher stiffness than 30_O; however, due to its lower toughness compared to 30_O, its overall fatigue resistance was lower than that of

30_O. Similarly, although 30_O demonstrated the highest toughness, its stiffness fell short of that exhibited by 45_HR_L, resulting in slightly lower cracking resistance in the S_{app} assessment in comparison. Additionally, mixture 60_HR_L_RA showcased the lowest toughness, nonetheless, owing to its higher stiffness relative to 45_HR, it attained a greater S_{app} value than 45_HR. In this context, it is noteworthy that the evaluation of S_{app} has diverged from the assessment based on the CT index. Specifically, the HRAP mixtures did not exhibit a significant increase in fatigue resistance compared to the control mixture. Further details regarding this observation will be discussed in subsequent correlation analyses.

$$S_{app} = 1000^{\frac{\alpha}{2}-1} \frac{a_T^{1/(\alpha+1)} \left(\frac{D^R}{C_{11}} \right)^{1/C_{12}}}{|E^*|^{\alpha/4}} \quad \text{Equation 3-9}$$

Where α = damage growth rate, and a_T = time-temperature shift factor value at the target temperature.

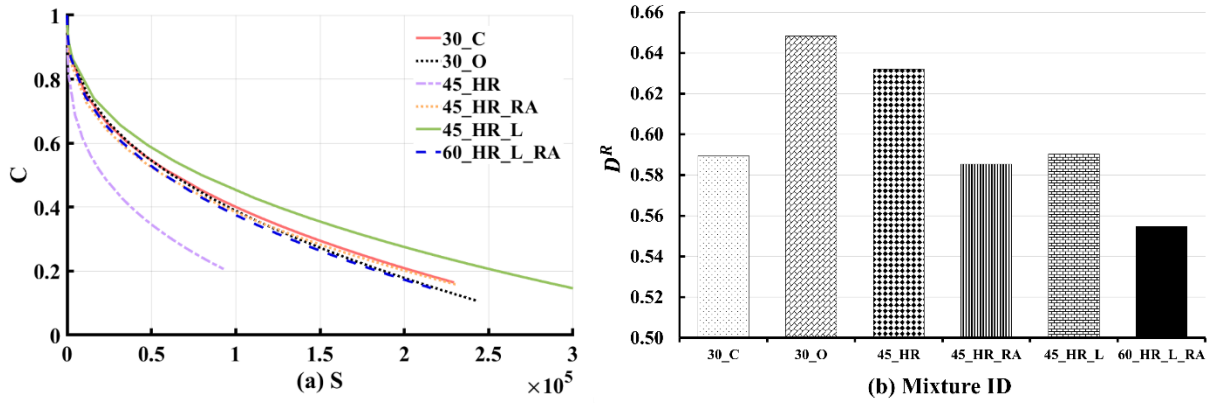


Figure 3-5. Performance Test Data for Direct Tension Cyclic Fatigue: (a) Damage Characteristic Curves and (b) DR Parameter.

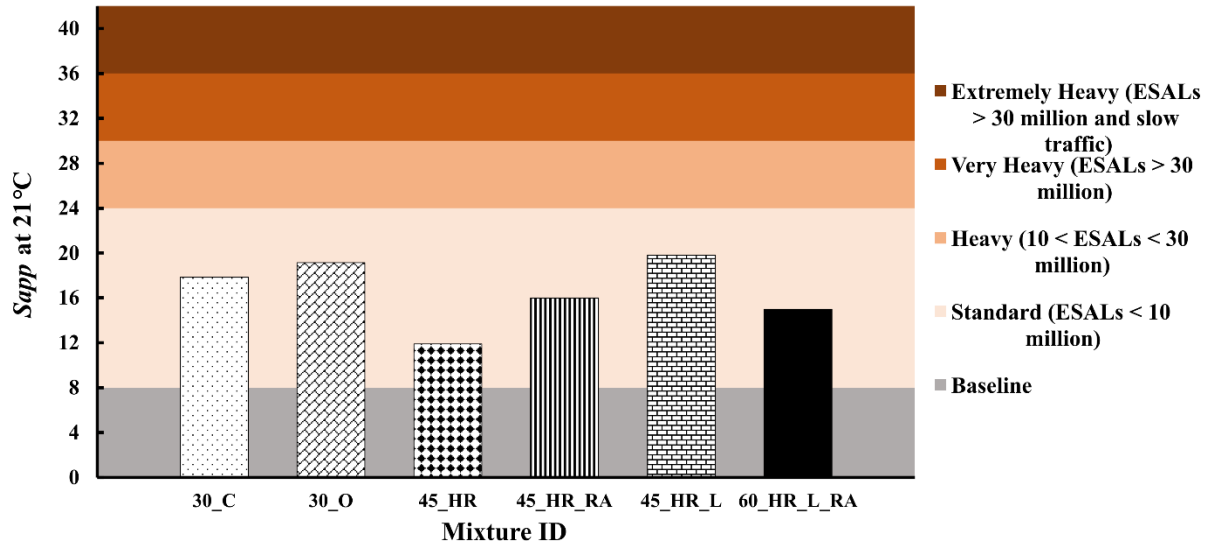


Figure 3-6. S_{app} Values. ESAL = equivalent single axle load; S_{app} = cyclic fatigue index parameter.

3.6. Rutting Performance Assessment

Two performance tests were considered to assess the rutting resistance of the six mixtures. These tests belong to two levels of testing: *intermediate* (APA rut test) and *advanced* (SSR test).

3.6.1. APA Rut Test and Results

The APA rut test was performed on specimens prepared from loose mixture collected during production in accordance with AASHTO T 340. This test simulates rutting in the laboratory by applying a loaded wheel back and forth over a pressurized rubber tube located along the surface of the test specimen at a temperature of 64°C which is equal to the high PG temperature of the employed asphalt binder. After 8,000 cycles were applied, the deformation of the specimen was measured. For each mixture, two replicates were tested, and an average was reported. Figure 3-7 presents the APA rut depth data at 64°C after 8000 loading cycles for the six mixtures. A higher APA rut depth value is an indication of an increased susceptibility to rutting damage. Mixture 45_HR exhibited the highest APA rut depths followed by Mixtures 60_HR_L_RA and 45_HR_RA, then followed by Mixtures 30_O, 45_HR_L, and 30_C. All evaluated mixtures exhibited an APA rut depth lower than VDOT BMD threshold of 8.0 mm. The APA rut test seemed to be insensitive to the RAP content and the use of softer binder and RA for the evaluated mixtures.

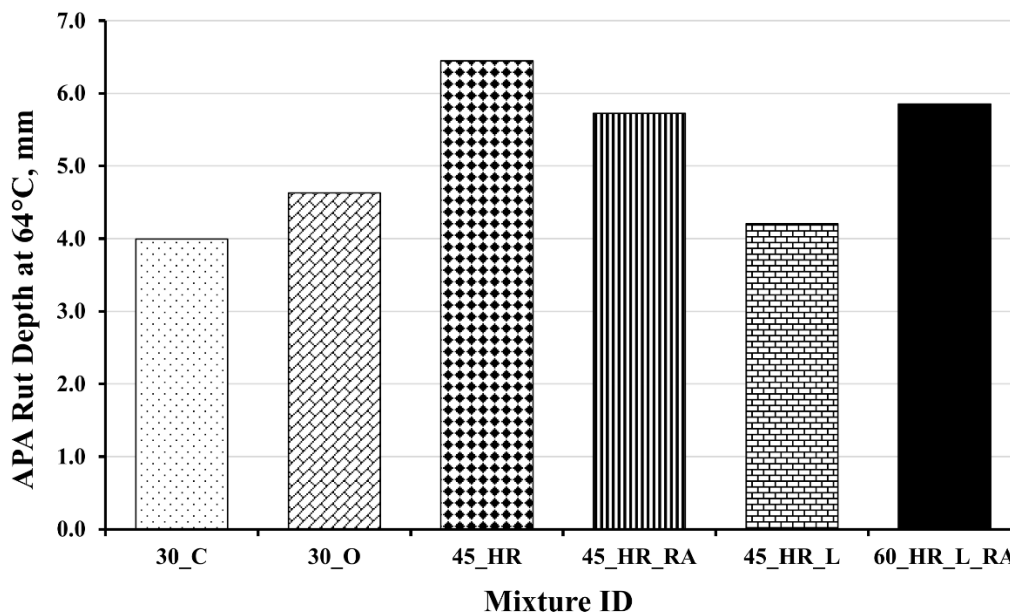


Figure 3-7. Performance Test Data for APA Rut Depth at 64°C and 8,000 Loading Cycles. APA = asphalt pavement analyzer.

3.6.2. Stress Sweep Rutting Test and Results

SSR test was performed in accordance with AASHTO TP 134 to assess the resistance of mixtures to rutting under repetitive loading, simulating the conditions experienced by pavements under repeated traffic loads. SSR testing was originally motivated by NCHRP Project 09-19, which showed that mixtures are sensitive to load level, loading history, and temperature under deviatoric axial loads. Four specimens were subjected to a confined cyclic compression testing at two different temperatures (T_H and T_L). In both cases, the specimens were subjected to three 200-cycle loading blocks of three deviatoric stress levels, and constant confining pressure of 69 kPa. In the testing, the permanent deformation of specimens is measured with its deviatoric stress, loading history, and temperature. Once a shift model is established, the permanent deformation of asphalt layers as a function of pavement depth and time can then be predicted by the pavement performance prediction program, FlexPAVE™ (Ghanbari *et al.* 2020).

The RSI recommended by FHWA, with the support of the shift model built by SSR, can be used as a rutting indicator (Ghanbari *et al.* 2020). The RSI is an average permanent strain in percentage calculated by repeating 30 million 18-kip equivalent single-axle load (ESAL) over 20 years on a standard pavement structure (Ghanbari *et al.* 2020). A lower RSI value is an indicator of a greater resistance to rutting for a given mixture. Figure 3-8 shows the RSI values for the six mixtures alongside the recommended threshold values as function of traffic. Mixture 30_O had the lowest RSI (1.85%) and fell in the very heavy traffic category. Mixtures 30_C, 45_HR_RA, and 45_HR_L all fell in the heavy traffic category with RSI values of 2.33%, 3.28%, and 2.72%,

respectively. Mixtures 45_HR and 60_HR_L_RA showed the highest RSI values (6.39% and 5.77%, respectively) and fell into the standard traffic category.

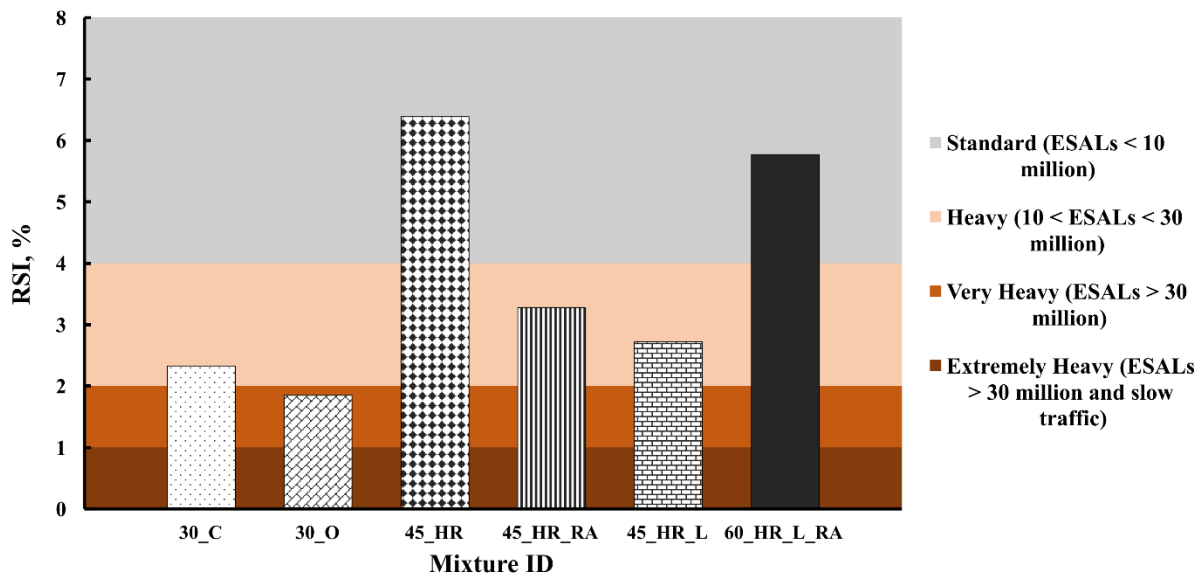


Figure 3-8. Performance Test Data for Stress Sweep Rutting Test. RSI = rutting strain index.

3.7. Correlations and Statistical Analyses

Various performance tests were conducted on the six mixtures and corresponding test results were presented and discussed in the previous sections. The tests considered and presented in this section were evaluated in terms of performance ranking and correlations among the indices. Finally, 3-D, ternary plots, and a novel CPI showcasing combinations of three primary performance indices.

3.7.1. Performance Ranking

The indices were compared in terms of their potential to rank the performance of the mixtures towards various distresses (i.e., durability, cracking, and rutting). This was done by sorting the average mean value of each index from the highest to the lowest for CT index and S_{app} parameter and from lowest to highest for CML, G-Rm, APA rut depth, and RSI. Table 3-4 shows the order ranking of the mixtures from the best (A) to the least (F) performing for each index. The selection of an optimal or suboptimal mixture from a ranking table established based on various material performance remained highly challenging which could be attributed to mechanics, comprehensiveness, and the level of complexity behind each selected performance test. As observed in Table 3-4, CML, G-Rm and CT index indicate that mixture 45_HR had the highest durability and resistance to cracking; meanwhile S_{app} showed that this mixture was the most susceptible to cracking. This could be attributed to a disagreement between the basic performance tests and the advanced performance tests, especially when these mixtures did not meet VDOT

mixture volumetric and gradation requirements (BMD Approach D was adopted for designing these mixtures). S_{app} ranked mixture 45_HR_L as the most resistant to cracking. APA rut depth and RSI ranked Mixture 45_HR as the most susceptible to rutting which again could be attributed to the relatively higher OBC used to design and produce this mixture. Finally, APA and RSI ranked Mixtures 30_C and 30_O as the most resistant to rutting, respectively.

Overall, relying on a singular ranking to design a BMD mixture in general and a HRAP mixture in particular that satisfied all performance requirements was not feasible. Consequently, it was imperative to investigate the correlations and discrepancies between different test results to improve the exploration and prediction of HRAP mixture performance.

Table 3-4. Ranking of Mixtures based on Various Evaluated Performance Indices

Parameters / Mixture	30_C	30_O	45_HR	45_HR_RA	45_HR_L	60_HR_L_RA
CML	F	E	A	D	C	B
CT index	E	F	A	B	D	C
S_{app}	C	B	F	D	A	E
G-Rm	D	C	A	E	F	B
APA	A	C	F	D	B	E
RSI	B	A	F	D	C	E

RSI = rutting strain index; CT = cracking tolerance; APA = asphalt pavement analyzer; CML = Cantabro mass loss; G-Rm = Glover-Row parameter of the mixture; S_{app} = cyclic fatigue index parameter.

3.7.2. Correlation

The Pearson correlation coefficient, expressed in Equation 3-10, was used to study the strength and direction of linear relationship between two variables (Cohen 1969).

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad \text{Equation 3-10}$$

Where r is the Pearson correlation ranging from -1 (perfect negative linear correlation) to 1 (perfect positive linear correlation), with 0 indicating no correlation; x and y are two variables of interest.

A correlation matrix built among different tests is shown in Table 3-5. As the r value changed from -1 to 1, the filled color changed from dark blue to dark red. As expected, it was observed that a strong positive linear correlation existed between RSI and APA (0.863). Thus, indicating that both test methods (intermediate vs. advanced) would lead to similar rutting performance assessment. However, this was not the case for cracking, as the various indices resulted in different and random trends. A strong negative linear correlation was observed between S_{app} and CT index (-0.813) indicating that for a given mixture, if the CT index increased, the corresponding S_{app} parameter would decrease. The discrepancies between the S_{app} and CT indices in this study were also noted in a Maine DOT study that assessed asphalt mixture cracking (Veeraragavan *et al.* 2021). This could be attributed to the fact that the IDT-CT (indirect tensile cracking test) is primarily influenced by the stiffness of the evaluated asphalt mixture, while the

direct tension cyclic fatigue test is more of a strain-controlled test. However, cracking itself is a complex phenomenon, with various types such as bottom-up, top-down, longitudinal, and thermal cracking occurring based on underlying failure mechanisms (Underwood and Braham 2019). These tests (IDT-CT vs. direct tension fatigue test) may simulate different types and/or combinations of cracking, making it challenging to comprehensively assess the fatigue performance of asphalt pavement, especially when dealing with BMD mixtures that deviate from traditional volumetric requirements. Moreover, the inclusion of higher RAP contents and the use of various additives, such as recycling agents, further complicate the evaluation of BMD mixtures (Rocha *et al.* 2023). Strong negative linear relationships existed between S_{app} and RSI (-0.906) and S_{app} and APA (-0.907). A lower RSI and/or APA value indicates stronger resistance to rutting, while a higher value for S_{app} indicates higher resistance to cracking. This suggests that the application of BMD provided the possibility of designing a mixture that exhibited simultaneous improvement in both fatigue cracking and rutting performance. CT index showed moderate positive correlations with RSI (0.732) and APA (0.683). It should be mentioned that the high CT index for mixture 45_HR was observed to be outside the range of typical values observed for VDOT BMD mixtures and may have substantially impacted the resulting correlations. Therefore, a similar analysis was performed after excluding the IDT-CT results for mixture 45_HR and is presented in Table 3-6. It can be noted that the negative correlation between S_{app} and CT index gets weaker (with r dropping from -0.813 to -0.521). A similar trend but slightly weaker correlation could be observed between CT index and RSI and between CT index and APA. In this context, it is evident that the considerable higher CT index of mixture 45_HR had an influence on the correlation between S_{app} and the CT index, thereby highlighting S_{app} as a comprehensive indicator for assessing fatigue cracking performance. Hence, in the subsequent discourse, it is prudent to distinguish between the fatigue cracking performance, as indicated by S_{app} , and the stiffness-driven cracking, as indicated by the CT index. Furthermore, from a stiffness perspective, pronouncedly, the correlation between CT index and G-Rm changed from a significant strong relationship to a minor positive relationship, reflecting that a simple linear relationship cannot describe the correlation between different cracking indices.

For the comparison of various indices with CML (durability), it was found that both rutting indices (RSI and APA) showed strong negative correlations with CML (-0.896 and -0.853, respectively). These results indicated that rutting resistance was negatively associated with durability. Additionally, a strong negative correlation was observed between CT index and CML (-0.824), suggesting that mixture could be designed to possess higher cracking resistance and durability. However, a contrary trend was also observed between cracking resistance and durability, as reflected by the correlation between S_{app} and CML (0.779).

Table 3-5. Correlation Matrix for All Evaluated Performance Indices

	<i>CML</i>	<i>CT index</i>	<i>Sapp</i>	<i>G-Rm</i>	<i>APA</i>	<i>RSI</i>
<i>CML</i>		-0.785	0.779	0.664	-0.853	-0.896
<i>CT index</i>	-0.785		-0.813	-0.905	0.683	0.732
<i>Sapp</i>	0.779	-0.813		0.810	-0.907	-0.906
<i>G-Rm</i>	0.664	-0.905	0.810		-0.673	-0.720
<i>APA</i>	-0.853	0.683	-0.907	-0.673		0.863
<i>RSI</i>	-0.896	0.732	-0.906	-0.720	0.863	

RSI = rutting strain index; CT = cracking tolerance; APA = asphalt pavement analyzer; CML = Cantabro mass loss; G-Rm = Glover-Row parameter of the mixture; S_{app} = cyclic fatigue index parameter. As the r value changed from -1 to 1, the filled color changed from dark blue to dark red.

Table 3-6. Correlation Matrix for All Evaluated Performance Indices Excluding CT index for Mixture 45_HR

	<i>CML</i>	<i>CT index</i>	<i>Sapp</i>	<i>G-Rm</i>	<i>APA</i>	<i>RSI</i>
<i>CML</i>		-0.824	0.779	0.664	-0.853	-0.896
<i>CT index</i>	-0.824		-0.521	0.482	0.633	0.647
<i>Sapp</i>	0.779	-0.521		0.810	-0.907	-0.906
<i>G-Rm</i>	0.664	0.482	0.810		-0.673	-0.720
<i>APA</i>	-0.853	0.633	-0.907	-0.673		0.863
<i>RSI</i>	-0.896	0.647	-0.906	-0.720	0.863	

RSI = rutting strain index; CT = cracking tolerance; APA = asphalt pavement analyzer; CML = Cantabro mass loss; G-Rm = Glover-Row parameter of the mixture; S_{app} = cyclic fatigue index parameter. As the r value changed from -1 to 1, the filled color changed from dark blue to dark red.

3.7.3. Three-dimensional (3-D) Plot

In addition to accounting for the interdependence between the performance indices, this study employed 3-D plots as a visual tool to demonstrate the effort of BMD optimization in HRAP mixture. By investigating and presenting the combinations of three performance indices, namely durability, rutting, and cracking, within a single plot and through one single index “CPI”, the study offered a comprehensive evaluation that enabled designers and agencies to assess both the singular performance and integrated performance of mixtures. To enhance the comparability of indices in the 3-D plot, a common normalization method was utilized for the index of interest for a given mixture. The method involved expressing the index as a percentage of the maximum value achieved by that index across different mixtures. For instance, Table 3-3 and Table 3-7 present the average testing values of mixtures, with the maximum CML value attained in Mixture 30_C. In this case, the maximum value was set as 1 (100%), and the mean values of other mixtures were converted into their respective percentage values. As a result, the normalized values obtained for mixtures 30_C, 30_O, 45_HR, 45_HR_RA, 45_HR_L, and 60_HR_L were 100%, 86.7%, 38.0%, 71.3%, 69.7%, and 57.4%, respectively.

It is important to note that the CML, RSI, and APA can be regarded somehow as reverse indices, whereby larger values correspond to poorer performance. Thus, the CML, RSI, and APA values were converted into (100%-CML), (100%-RSI), and (100%-APA), respectively, to better depict their positive relative performance relationship in the 3-D plot. The normalization for the 3-D plot of advanced performance tests can be found in Table 3-7. Consequently, the longer the distance from the origin in the 3-D plot is, the better the CPI is considered (Equation 3-11). The

prototype of CPI here presumes that the three primary indices contribute equally to the overall composite performance, which implies that the variables A, B, and C in Equation 3-11 were all assigned a value of 1. However, taking into account the varying degrees of concern that local designers and agencies may have for three primary distresses, the values of A, B, and C can be adjusted. This flexibility allows for the BMD mixtures to satisfy local requirements.

$$CPI = \sqrt{Ax^2 + By^2 + Cz^2} \quad \text{Equation 3-11}$$

Where CPI is the composite performance index; x, y, x are rutting resistance indicator, cracking resistance indicator, and durability, respectively; A, B, C are adjustable weight factor.

The 3D plots are shown in Figure 3-9a, Figure 3-10a and Figure 3-11a, where the size of each point signifies the distance from the origin for a better visualization. The larger size of points implies that they possess a superior CPI for the mixture. Additionally, the color of each point was determined based on the CPI (Equation 3-11), indicated by the distance from origin, wherein the colormap depicted that the color of points changed from the bottom color to the top color as the exhibited CPI for different mixtures increased. Along the positive x (S_{app} , %), y (1-RSI, %), and z (1-CML, %) axes, the cracking resistance, rutting resistance, and durability of the mixture are expected to exhibit an upward trend. Upon comparative analysis, it was observed that each ranking utilizing the x, y, and z axes corresponded with the conclusions derived from the ranking table (refer to Table 3-4), thereby ensuring consistency.

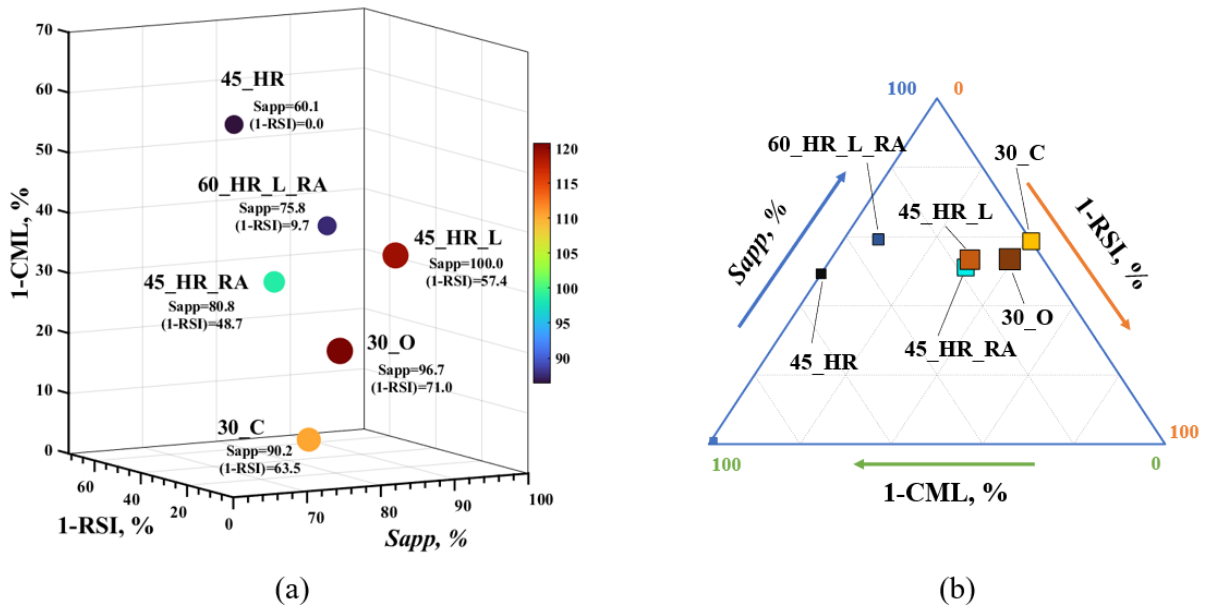


Figure 3-9. (a) 3-D Plots of Performance Indices: CML vs. RSI vs. S_{app} ; (b) Ternary Plots of Performance Indices: CML vs. RSI vs. S_{app} . 3-D = three dimensional; CML = Cantabro mass loss; S_{app} = cyclic fatigue index parameter; and RSI = rutting strain index.

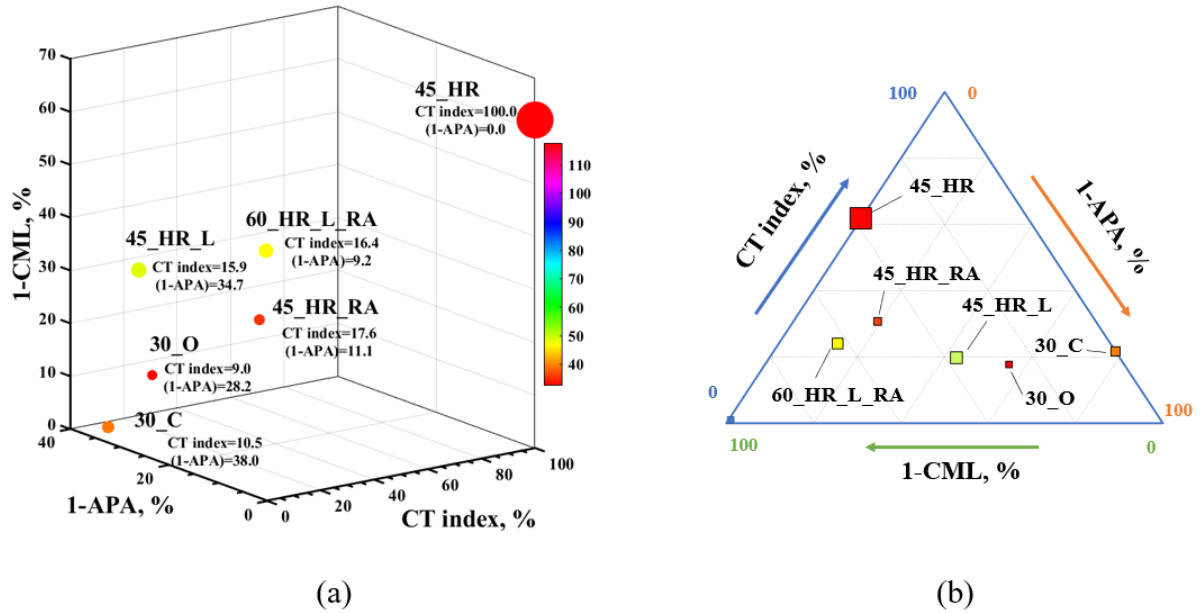


Figure 3-10. (a) 3-D Plots of Performance Indices CML vs. APA vs. CT index; (b) Ternary Plots of Performance Indices: CML vs. APA vs. CT index. 3-D = three dimensional; CML = Cantabro mass loss; APA = asphalt pavement analyzer; CT = cracking tolerance.

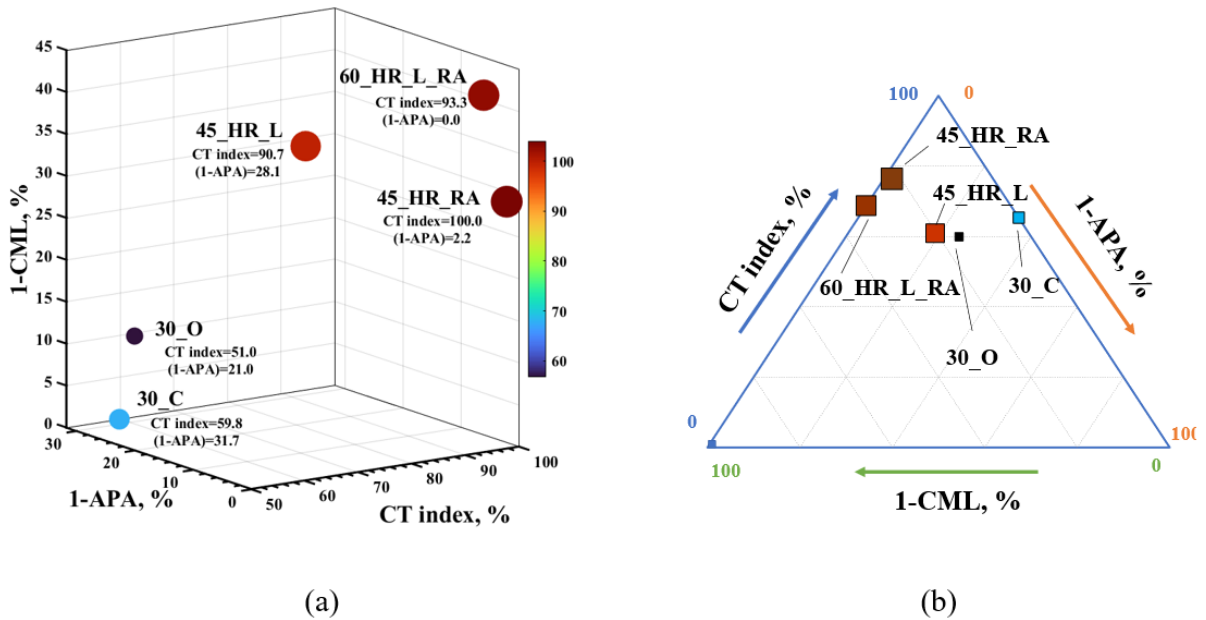


Figure 3-11. (a) 3-D Plots of Performance Indices: CML vs. APA vs. CT index Excluding Mixture 45_HR; (b) Ternary Plots of Performance Indices: CML vs. APA vs. CT index Excluding Mixture 45_HR. 3-D = three dimensional; CML = Cantabro mass loss; APA = asphalt pavement analyzer; CT = cracking tolerance.

The three images presented in Figure 3-9a, Figure 3-10a, and Figure 3-11a unequivocally demonstrate that HRAP mixtures exhibited superior durability, reflected by lower CML values (larger z-axis values), when compared to the 30% RAP mixture. By examining Figure 3-9a, it can

be observed that although mixture 45_HR had the lowest S_{app} and the highest RSI, the performance of mixture 45_HR_RA, mixture 45_HR_L, and mixture 60_HR_L_RA could confirm that the use of a softer binder, an RA, or a combination of both alternatives positively affected both fatigue cracking and rutting performance of HRAP mixtures. Upon comparing mixture 30_O with the control mixture 30_C, it became evident that BMD application has resulted in enhanced resistance to the three primary distresses. Further analysis of mixture 45_HR_L (compared with mixture 30_C and mixture 30_O) revealed that, contrary to expectations due to the introduction of HRAP, there was no significant increase in rutting resistance. Instead, an enhancement in fatigue cracking resistance and durability, along with a sacrifice in rutting performance, was observed as a result of BMD optimization. The CPI of these three mixtures demonstrated remarkable similarity, with mixture 30_O demonstrating the best composite performance. However, considering the RSI of mixture 45_HR_L, indicating its favorable performance under heavy traffic (Figure 3-8), and the CML less than 7.5% VDOT BMD mass loss criterion (Figure 3-1), it was deemed that mixture 45_HR_L outperformed all other mixtures and held potential for further BMD optimization to enhance fatigue cracking resistance. Nevertheless, other HRAP mixtures, including mixture 45_HR, mixture 45_HR_RA, and mixture 60_HR_L_RA, did not surpass the control mixture in performance, except for improvements in durability. Thereby, among the 45% BMD mixtures, the beneficial impact of utilizing a softer binder on CPI was notably superior to that of incorporating RAs or extra higher binder content.

Figure 3-10a reveals a notable shift in the observed scenario from advanced level tests (Figure 3-9a), where mixture 45_HR achieved a superior composite performance by leveraging its significantly higher CT index compared to other mixtures. In terms of stiffness-driven cracking resistance, all HRAP mixtures demonstrated higher CT indices than the 30% RAP mixtures, yet they also exhibited lower rutting resistance than the control mixture (30_C). Notably, mixture 45_HR_L demonstrated rutting resistance comparable to mixture 30_C and mixture 30_O, while exhibiting superior cracking resistance, durability, and CPI. Based on these findings, it was reasonable to conclude that the 3-D plots, employed in both advanced level tests (CML vs. RSI vs. S_{app}) and basic/intermediate level tests (CML vs. APA vs. CT), consistently demonstrated the superiority of HRAP BMD surface mixtures over the control mixture. This was exemplified by the outperformance of mixture 45_HR_L compared to the control mixture 30_C. Moreover, the 3-D plot of the basic/intermediate level tests further emphasized that all HRAP BMD surface mixtures possessed the capability to outperform the control mixture 30_C. This could be attributed to the adherence of all HRAP BMD surface mixtures to the APA rut depth requirements (Figure 3-7), ensuring the retention of sufficient rutting resistance reserved. Despite the lower rutting resistance displayed by all HRAP BMD surface mixtures in the 3-D plot in comparison to mixture 30_C, their advantages in terms of cracking resistance (indicated by the CT index) and durability and sufficient rutting resistance suggested that HRAP BMD surface mixtures surpassed the control mixture 30_C in overall performance.

The extremely high CT index of mixture 45_HR had an influence on the comparison of relative cracking resistance between HRAP and 30% RAP mixtures, which is addressed in Figure 3-11a. In this figure, a distinct "gap" is observed between HRAP mixtures and the control mixture, or rather, HRAP mixtures and 30% RAP mixtures. This gap is characterized by a significant increase in cracking resistance (indicated by the CT index) and durability in all HRAP mixtures. These findings provide strong evidence for the beneficial effects of utilizing a softer binder, RA, or a combination of both alternatives on the cracking and durability performance of HRAP mixtures. Furthermore, these results align with the insights revealed by the 3-D plot in the advanced tests.

In summary, although there were disparities in the evaluation of cracking resistance between Sapp and the CT index, the 3-D plots obtained from both advanced and basic/intermediate tests produced two key findings. Firstly, HRAP BMD surface mixtures can achieve superior CPI compared to the traditionally designed control mixture, albeit achieving this in the advanced tests proved to be more challenging than in the basic/intermediate tests. Secondly, the application of a softer binder, RA, or a combination of both alternatives demonstrated beneficial effects on the overall performance of HRAP BMD mixtures. More importantly, the 3D plot provided a superior visualization for comparing the complex array of performance indices. By departing from conventional methods that involve pairwise comparisons of performance indices, the implementation of a 3D plot associated with CPI significantly alleviates the burdensome nature of traditional comparison methods, as well as minimizes the risk of overlooking crucial factors that may result in errors or imbalances in decision-making. Consequently, designers are endowed with a more efficient and comprehensive reference framework that enables them to effectively balance the three primary performance indices from multiple perspectives.

3.7.4. Ternary Plots

The utilization of a 3-D plot offered one of viable approaches to compare three primary performance indices among different AC mixtures within a single visualization. However, it is important to acknowledge the inherent limitations of this approach, as it does not fully capture the intricate trade-offs or balanced relationships among the three primary performance indices within an individual mixture. Hence, an auxiliary methodology is required to offer additional guidance and different perspective for designers. The ternary plot can be utilized as a complementary tool to the 3-D plot, enabling the display of each performance metric's weight in the composite performance of a mix, thereby assisting the BMD. In this case, an additional normalization process was applied that transforms each calculated percentage value for the 3-D plot into a percentage of the sum of the percentages resulting from different indices for each AC mixture, which allowed for the comparison of the contribution of each index to the composite performance of the evaluated mixture. For instance, considering mixture 30_C, after the first normalization in the 3-D plot, the values for S_{app} , (1-RSI), and (1-CML) were 90.15%, 63.54%, and 0%, respectively. Subsequently, by adding them up and converting each index to a percentage of the total sum (153.69), the

normalized values were determined to be 58.7%, 41.3%, and 0%, respectively. Table 3-7 illustrates more details. The normalized values were then plotted in the equilateral triangular coordinate system.

Table 3-7. Normalize Advanced Performance Tests (CML vs. RSI vs. S_{app}) for 3-D Plot and Ternary Plot

Mix ID	Raw data			3-D plot normalization			SUM	Ternary plot normalization		
	S_{app}	RSI	CML	S_{app}	1-RSI	1-CML		S_{app}	1-RSI	1-CML
30_C	17.9	2.3	9.5	90.15	63.54	0	153.69	58.7	41.3	0.0
30_O	19.2	1.9	8.3	96.72	71.05	13.31	181.08	53.4	39.2	7.4
45_HR	11.9	6.4	3.6	60.10	0	62.03	122.13	49.2	0.0	50.8
45_HR_RA	16.0	3.3	6.8	80.81	48.67	28.72	158.20	51.1	30.8	18.2
45_HR_L	19.8	2.7	6.6	100	57.43	30.26	187.70	53.3	30.6	16.1
60_HR_L_RA	15.0	5.8	5.5	75.76	9.70	42.59	128.05	59.2	7.6	33.3

RSI = rutting strain index; CT = cracking tolerance; APA = asphalt pavement analyzer; CML = Cantabro mass loss; G-Rm = Glover-Row parameter of the mixture; S_{app} = cyclic fatigue index parameter.

A ternary plot is shown in Figure 3-9b, Figure 3-10b, and Figure 3-11b to assess the contribution between durability (1-CML), rutting resistance (1-RSI) or (1-APA), and fatigue cracking resistance (S_{app}) or basic cracking resistance (CT index) within each mixture. In the ternary plot, the size and the color of a point correspond to the points determined in the 3-D plot, which reflect the CPI. The calculated normalized values were transformed into a coordinate system of an equilateral triangle, where each value represented a sub-axis coordinate. Each sub-axis was obtained by rotating clockwise 60 degrees in the direction of increasing index values.

The exploration of ternary plots in conjunction with corresponding 3-D plots allows for a comprehensive understanding of the intricate "balance" within mixtures, illustrating the complexity of their internal equilibrium that cannot be simplified as a mere trade-off. In Figure 3-9b, it is observed that mixture 30_C exhibited inadequate durability (as depicted in Figure 3-1), whereas mixture 30_O addressed this issue by incorporating BMD to enhance its durability while simultaneously contributing to its CPI. Although this contribution diminished the rutting and cracking resistance of mixture 30_O relative to its CPI proportionally compared to mixture 30_C, the CPI of mixture 30_O was elevated, resulting in higher values for its three primary distress indices (or superior singular rankings) compared to mixture 30_C. Additionally, as a formidable competitor to mixture 30_O, mixture 45_HR_L endeavored to further improve durability. However, it was observed that the CPI of mixture 45_HR_L no longer continued to increase relative to mixture 30_O. The increased proportion of durability within its CPI restricted rutting performance. However, mixture 45_HR_L obtained higher durability and fatigue cracking resistance than mixture 30_O. Further examination of mixture 45_HR_RA, mixture 60_HR_L_RA, and mixture 45_HR revealed that the continued emphasis on enhancing their own durability proportion led to a decline in CPI, consequently sacrificing their cracking resistance and

rutting resistance. This effect was particularly pronounced in mixture 60_HR_L_RA and mixture 45_HR.

Figure 3-10b represents the ternary plot for basic/intermediate performance tests. Similarly, starting from mixture 30_C, mixture 30_O here aimed to improve its durability. However, unlike the ternary plot in the advanced tests (Figure 3-9b), the CPI of mixture 30_O was directly reduced in the BMD optimization process. Additionally, the contributions of the rutting indicator and cracking indicator to the CPI were constrained by the increased proportion of durability, resulting in inferior rutting resistance and stiffness-driven cracking resistance compared to mixture 30_C. Nonetheless, within the HRAP BMD mixtures, it can be observed that further enhancing durability led to an increase in CPI, except for mixture 45_HR_RA. Mixture 45_HR_RA exhibited higher cracking resistance contributions, resulting in a higher cracking resistance compared to mixture 30_C. Simultaneously, it sacrifices the proportion of rutting resistance compared to the control mixture 30_C. As customary, Figure 3-11b provides a better comparison of the contribution of each performance index for the other five mixes. The observation from Figure 3-11b is consistent with that of Figure 3-10b, with the exception of mixture 45_HR_RA. Upon further examination, it was notable that mixture 45_HR_RA had the highest CPI, despite being the inferior-performing mix initially from Figure 3-10b. This was because, after removing mixture 45_HR from the analysis, mixture 45_HR_RA had the highest CT index value, yet this did contribute significantly to its cracking resistance in the ternary plot. Additionally, it was evident that for HRAP mixtures, compared to 30% RAP mixtures, rutting resistance had a relatively minor contribution to their CPI, but durability and cracking resistance were particularly important for HRAP mixtures and strengthened the CPI, overshadowing the 30% RAP mixture. It further demonstrated the positive effects of using softer binders and RA in HRAP mixtures.

3.8. Summary of Findings, Conclusions, and Ongoing Work

This research evaluated one control and five BMD optimized SMs through laboratory performance tests at various levels of complexity. The mixtures incorporated various combinations of RAP contents, two binder PGs, one RA, and one WMA additive. The tests performed included the Cantabro, IDT-CT, $|E^*|$, S-VECD fatigue, APA rutting, and SSR test. The primary focus of the assessment was on the durability, rutting resistance, and cracking resistance as envisioned by the BMD concept/methodology being applied in Virginia.

This study investigated the challenge of finding a consistent ranking order based on a single performance index for six mixtures evaluated using performance tests at various levels of complexity. The results indicated that the ranking order differed across different tests. For instance, the APA (intermediate level) and RSI (advanced level) tests did not produce the same ranking order for rutting evaluation. Similarly, the CT index (basic level) and S_{app} (advanced level) tests did not yield the same ranking order in cracking evaluation, even resulting in reversed outcomes.

Additional investigations were conducted by comparing the Pearson correlation coefficient between two given performance indices. Comparisons revealed that different levels of rutting performance tests can produce a relatively consistent results, as reflected by the strong positive linear correlation found between RSI and APA. Additionally, this study clarified the disparities between Sapp and the CT index, providing explanations for their divergent outcomes. It is recommended to separately discuss their evaluations pertaining to cracking. Furthermore, it was found that BMD has great potential for the design of HRAP mixtures that exhibit simultaneous improvement in both fatigue cracking and rutting performance. Rutting resistance showed a negative correlation with durability. Based on S_{app} , fatigue cracking resistance showed a negative correlation with durability, while based on CT index and G-Rm, cracking resistance showed a positive correlation with durability.

Furthermore, this study proposed a novel index CPI associated with a 3D-plot to evaluate the composite performance of different mixtures and a ternary plot to assist in evaluating the contribution of individual major performance results in BMD optimized mixtures. The findings of this study indicated disparities between the results of basic/intermediate performance tests and advanced performance tests, as corroborated by the correlation analysis. Nonetheless, the overall comparisons consistently highlighted that HRAP BMD surface mixtures can exhibit superior overall performance as compared to conventionally designed control mixtures. However, achieving this enhanced performance in the advanced tests posed greater challenges compared to the basic/intermediate tests. Furthermore, the comparisons emphasized the beneficial effects arising from the utilization of softer binders, RA, or their combinations in HRAP mixtures.

3.9. Reference

- Advanced Asphalt Technologies, LLC. (2011). NCHRP Report 673: A Manual for Design of Hot Mix Asphalt with Commentary. Transportation Research Board, Washington, D.C.
- Al-Qadi, I. L., Ozer, H., & Golias, M. M. (2018). Review of Asphalt Mixture Performance Tests. *Transportation Research Record*. 2672(12): 128-140.
- American Association of State Highway and Transportation Officials. (2017). Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards. Washington, DC.
- ASTM International. (2019). 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.
- Bonaquist, R. F., Christensen, D. W., and Stump, W. (2003). NCHRP Report 513: Simple Performance Tester for Superpave Mix Design: First-Article Development and Evaluation. Transportation Research Board, Washington, D.C.
- Cohen, J. (1969). Statistical Power Analysis for the Behavioral Sciences. The SAGE Encyclopedia of Research Design.
- Cox, B. C., Smith, B. T., Howard, I. L., & James, R. S. (2017). State of knowledge for Cantabro testing of dense graded asphalt. *Journal of Materials in Civil Engineering*, 29(10),

04017174.

- Diefenderfer, S. D., Boz, I., & Habbouche, J. (2021). Balanced Mix Design for Surface Asphalt Mixtures: Phase I: Initial Roadmap Development and Specification Verification. Publication VTRC 21-R15. Virginia Transportation Research Council, Charlottesville.
- Diefenderfer, S. D., Habbouche, J., & Boz, I. (2023). Balanced Mix Design for Surface Asphalt Mixtures: 2020 Field Trials. Publication VTRC 23-R13. Virginia Transportation Research Council, Charlottesville.
- E-C280. (2022). Glossary of Terms for Balanced Mix Design of Asphalt Mixtures. Transportation Research Board, <https://onlinepubs.trb.org/onlinepubs/circulars/ec280.pdf>.
- Martin, A. E., Kaseer, F., Arámbula-Mercado, E., Bajaj, A., Cucalon, L. G., Yin, F., ... & King, G. (2020). Evaluating the effects of recycling agents on asphalt mixtures with high RAS and RAP binder ratios (No. Project 09-58).
- Federal Highway Administration. (2019). Cyclic Fatigue Index Parameter (Sapp) for Asphalt Performance Engineered Mixture Design. FHWA-HIF-19-091. Washington, D.C.
- Ghanbari, A., Underwood, B. S., & Kim, Y. R. (2020). Development of a Rutting Index Parameter Based on the Stress Sweep Rutting Test and Permanent Deformation Shift Model. *International Journal of Pavement Engineering*, 23(2): 387-399, <https://doi.org/10.1080/10298436.2020.1748190>.
- Habbouche, J., Boz, I. & Diefenderfer, S. D. (2021a). Validation of Performance-Based Specifications for Surface Asphalt Mixtures in Virginia. *Transportation Research Record*, 2676 (5): 277 – 296.
- Habbouche, J., Boz, I., & Diefenderfer, S. D. (2022a). Interlaboratory Study for the Indirect Tensile Cracking Test at Intermediate Temperature: Phase II. VTRC Report 23-R3. Virginia Transportation Research Council, Charlottesville.
- Habbouche, J., Boz, I., Underwood, B. S., Castorena, C., Gulzar, S., Fried, A., & Preciado, J. (2022b). Review from Multiple Perspectives for the State of the Practice on the Use of Recycled Asphalt Materials and Recycling Agents In Asphalt Concrete Surface Mixtures. *Transportation Research Record*, 2676(4), 407-420.
- Habbouche, J., Boz, I., Diefenderfer, S. D., & Bilgic, Y. K. (2021b). Round Robin Testing Program for the indirect tensile cracking test at intermediate temperature: Phase I (No. FHWA/VTRC 22-R3). Virginia Transportation Research Council, Charlottesville.
- Habbouche, J., Boz, I., Underwood, B. S., Castorena, C., Preciado, J., Gulzar, S., & Fried, A. (2023). Engineered Frameworks for Evaluating the Use of Recycling Agents in Surface Asphalt Mixtures for Virginia. Virginia Transportation Research Council, Charlottesville.
- Harvey, J., & Popescu, L. (2000). Accelerated Pavement Testing of Rutting Performance of Two Caltrans Overlay Strategies. *Transportation Research Record*, 1716(1), 116-125.
- Keshavarzi, B. (2019). Prediction of Thermal Cracking in Asphalt Pavements Using Simplified Viscoelastic Continuum Damage Theory. North Carolina State University.
- Kim, Y. R., Castorena, C., Elwardany, M., Rad, F. Y., Underwood, B. S., Gundla, A., ... & Glaser, R. R. (2018). Long-Term Aging of Asphalt Mixtures for Performance Testing and

Prediction.

- Mensching, D. J., Rowe, G. M., & Sias Daniel, J. (2017). A Mixture-Based Black Space Parameter for Low-Temperature Performance of Hot Mix Asphalt. *Road Materials and Pavement Design*, 18(sup1), 404-425.
- Meroni, F., Flintsch, G. W., Habbouche, J., Diefenderfer, B. K., & Giustozzi, F. (2021). Three-Level Performance Evaluation of High RAP Asphalt Surface Mixes. *Construction and Building Materials*, 309, 125164.
- Mohajeri, M. (2015). Hot Mix Asphalt Recycling: Practices and Principles.
- Nair, H., Diefenderfer, S. D., & Bowers, B. F. (2019). Assessing Increased Use of Reclaimed Asphalt Pavement in Asphalt Mixtures (No. FHWA/VTTC 20-R9). Virginia Transportation Research Council (VTRC).
- Ogbo, C., Kaseer, F., Oshone, M., Sias, J. E., & Martin, A. E. (2019). Mixture-Based Rheological Evaluation Tool for Cracking in Asphalt Pavements. *Road Materials and Pavement Design*, 20(sup1), S299-S314.
- Olard, F., & Di Benedetto, H. (2003). General “2S2P1D” Model and Relation between the Linear Viscoelastic Behaviours of Bituminous Binders and Mixes. *Road materials and pavement design*, 4(2), 185-224.
- Rocha, M. L., Aragão, F. T. S., Nascimento, L. A. H. D., & Underwood, B. S. (2023). Balanced Mixture Design Framework for Asphalt Mixtures Based on Index-and Performance-Volumetrics Relationships. *Transportation Research Record*, 2677(10), 233-245.
- Tong, B., Habbouche, J., Flintsch, G. W., & Diefenderfer, B. K. (2023). Rutting Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents under Full-Scale Accelerated Testing. *Materials*, 16(24), 7611.
- Underwood, B. S., & Braham, A. (2019). Fatigue or Not Fatigue, That is the Question - a White Paper.
- Underwood, B. S., Kim, Y. R., & Guddati, M. N. (2010). Improved Calculation Method of Damage Parameter in Viscoelastic Continuum Damage Model. *International Journal of Pavement Engineering*, 11(6), 459-476.
- Veeraragavan, R. K., Nener-Plante, D., Myers, L., Nash, C., & Tran, N. H. (2022). Balanced Mix Design Benchmarking of Field-Produced Asphalt Mixtures in Maine, US. *Transportation Research Record*, 2676(5), 263-276.
- Wang, Y. D., Underwood, B. S., & Kim, Y. R. (2022). Development of a Fatigue Index Parameter, Sapp, for Asphalt Mixes Using Viscoelastic Continuum Damage Theory. *International Journal of Pavement Engineering*, 23(2), 438-452.
- West, R. C., & Copeland, A. (2015). *High RAP Asphalt Pavements: Japan Practice-Lesson Learned* (No. IS 139).
- West, R., Rodezno, C., Leiva, F., & Yin, F. (2018). Development of a Framework for Balanced Mix Design. *Project NCHRP*, 20-07.
- Williams, B. A., Willis, J. R., & Shacat, J. (2020). Asphalt Pavement Industry Survey on Recycled Materials and Warm-Mix Asphalt Usage: 2019. National Asphalt Pavement

Association, Greenbelt, MD.

Witczak, M., Schwartz, C., & Von Quintus, H. L. (2011). NCHRP Project 9-19: Superpave Support and Performance Models Management. Transportation Research Board, Washington, D.C.

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

Zaumanis, M., Mallick, R. B., & Frank, R. (2013). Evaluation of Rejuvenator's Effectiveness with Conventional Mix Testing for 100% Reclaimed Asphalt Pavement Mixtures. *Transportation Research Record*, 2370(1): 17-25.

Zhang, R., Sias, J. E., & Dave, E. V. (2022). Evaluation of the Cracking and Aging Susceptibility of Asphalt Mixtures Using Viscoelastic Properties and Master Curve Parameters. *Journal of Traffic and Transportation Engineering (English Edition)*, 9(1), 106-119.

CHAPTER 4 - Assessment of Production Variability and Composite Performance Index for Conventional and High RAP BMD Mixtures⁴

4.1. 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

⁴ This paper has been published in the *Transportation Research Records*. <https://doi.org/10.1177/03611981241287785>. Citation: Tong, B., Habbouche, J., Diefenderfer, S. D., Flintsch, G. W., & Boz, I. (2024). Assessment of the Production Variability and Composite Performance Index for Conventional and High Reclaimed Asphalt Pavement Balanced Mix Design Mixtures. *Transportation Research Record*, 03611981241287785. Contribution: Conceptualization: Lead; Methodology: Lead; Resources: Equal; Data Processing and Analysis: Lead; Laboratory testing: Equal; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead.

4.2. 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 (Golalipour et al., 2021). The abundant use of additives, extensive application of 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 (Golalipour et al., 2021).

The emergence of the 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*” (West et al., 2018; Yin and West, 2021). 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 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 (Zaumanis et al., 2013; Martin et al., 2020). 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 (Zaumanis et al., 2013; Martin et al., 2020; Abu Saq et al., 2023). 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 (Martin et al., 2020; Sabouri, 2020). 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 (Golalipour et al., 2021; McDaniel, 2012). 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 (Martin et al., 2020). 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 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 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 (Brown et al., 2009). 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 (Brown et al., 2009). 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 (Hughes, 2005; Mohammad et al., 2016). Consequently, the implementation of QC measures, coupled with 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, 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 (Mohammad et al., 2016). In addition to these scenarios, other alternatives include non-reheated and reheated plant-mixed, laboratory-compacted specimens, scenarios found in the Virginia VDOT's practice for contractors' quality control (Diefenderfer et al., 2023).

Numerous state highway agencies are actively exploring the design and acceptance of asphalt mixtures based on performance using the BMD concept (West et al., 2018; Yin and West, 2021). 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 SMs. By 2013, VDOT was examining the feasibility of SMs containing up to 45% RAP (Diefenderfer et al., 2021), leading to field trials of mixtures containing 40-45% RAP contents (Nair et al., 2019). 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 IDT-CT at intermediate temperature, and the APA test, assessing durability, cracking, and rutting, respectively. Initial performance thresholds were set (Bowers et al., 2022) and subsequently the developed performance-based specifications were validated (Diefenderfer et al., 2023; Habbouche et al., 2021).

Two VDOT BMD special provisions, one for conventional SMs and one for HRAP SMs were drafted and revised for use in pilot projects (Diefenderfer et al., 2023). 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 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 (Diefenderfer et al., 2021). 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 (Diefenderfer et al., 2021). 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 VTRC, VDOT, and Virginia Tech, involving the design, placement, and accelerated testing of a control mixture and five BMD mixtures at the Virginia APT facility, owned by VDOT and located at the VTTI (Tong et al., 2023). 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 (Tong et al., 2024).

4.3. 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 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.

4.4. Methodology

4.4.1. 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 4-1.

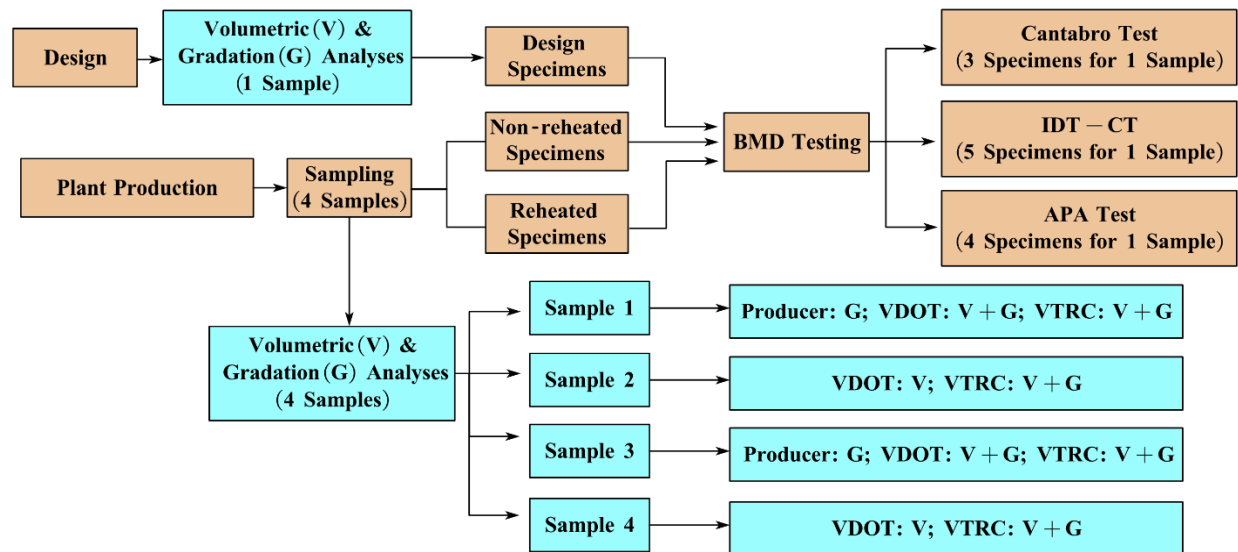


Figure 4-1. Experimental Plan for Laboratory- and Plant-Produced Mixtures. BMD = balanced mix design; IDT-CT = indirect tensile cracking test; APA = asphalt pavement analyzer; G = gradation analysis; V = volumetric analysis; VDOT = Virginia Department of Transportation; VTRC = Virginia Transportation Research Council.

4.4.2. Mixture Designs

The mixtures incorporated various combinations of RAP contents, two PGs, one RA, and one WMA additive. Design specimens were laboratory-produced laboratory-compacted specimens, fabricated for mixture design. The mixtures are defined in Table 3-1.

4.4.3. 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 4-1. 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.. Samples were sequentially labelled from 1 to 4 (A to D) 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 1 was fabricated from sample 1 for the mixture under evaluation). The producer generally performed a gradation analysis on samples 1 and 3. The VDOT district laboratory determined the volumetric properties for all four samples, in addition to determining the gradation for samples 1 and 3. 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 (Diefenderfer et al., 2021). 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.

4.4.4. 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 (Diefenderfer et al., 2023; Bowers et al., 2022). The corresponding thresholds/criteria for these tests are shown in Table 4-1. 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 (Cox et al., 2017). The 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 on five replicate test specimens compacted at $7\pm0.5\%$ air voids. The APA test was performed in accordance with AASHTO T 340 at a temperature of 64°C; four replicate test specimens were tested in an APA Junior machine. The APA specimens were compacted at $7\pm0.5\%$ 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 (Boz et al., 2021) 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 4-1. Virginia Performance Testing Criteria

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

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.

4.4.5. 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 (Braver et al., 2003). 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% (Braver et al., 2003). 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 pairwise comparison test (Ostertagová et al., 2014). Results are represented in Table A-1 and Table A-2 for different BMD tests.

4.5. Results

4.5.1. 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 (VDOT 2020).

Table 4-2 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 requirements on asphalt content and gradation (Diefenderfer et al., 2019). 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 4-2. Summary of Primary Volumetric Properties Determined on Mixtures During Production

Property	JMF	Production tolerance	Sample 1		Sample 2		Sample 3		Sample 4	
			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 4-2. Figure 4-2 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 4-3. 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.

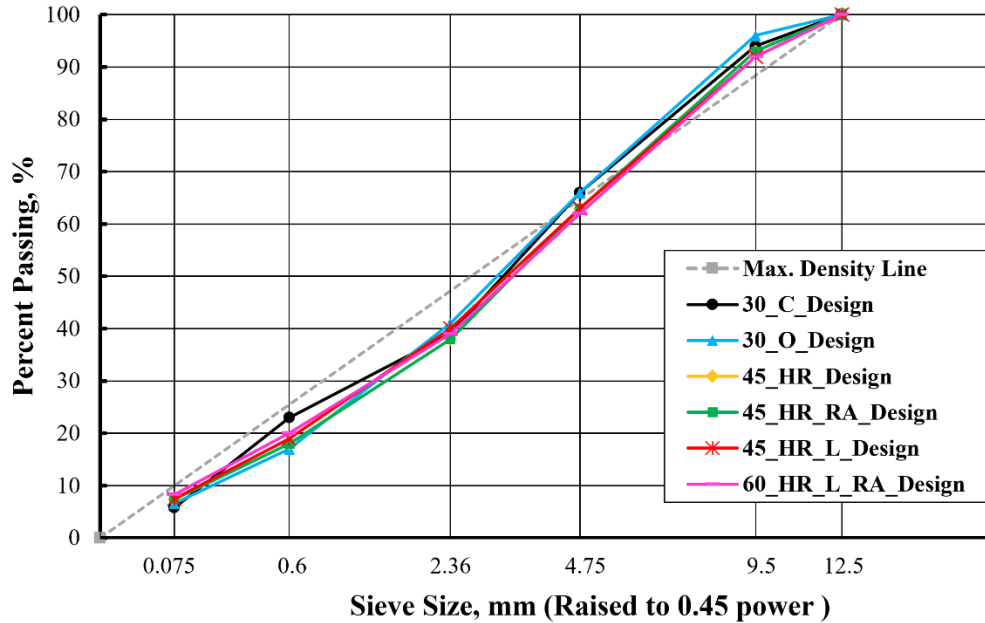


Figure 4-2. Design Gradations for the Six Mixtures. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

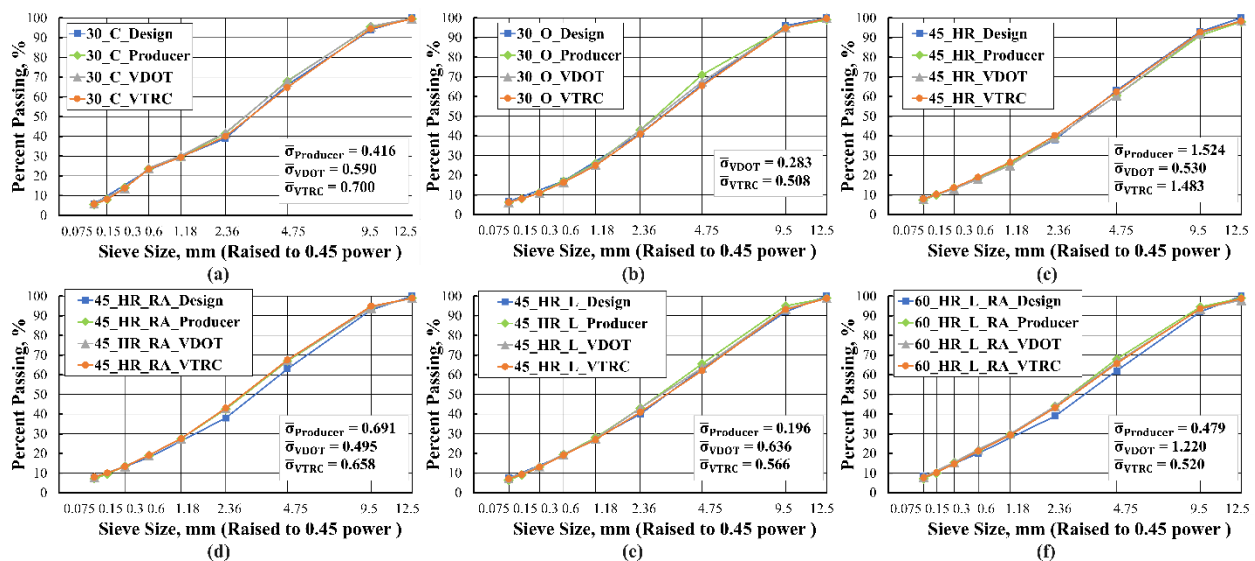


Figure 4-3. Average Aggregate Gradations Determined During Design by Producer and During Production by Producer, VDOT District, and VTRC for: (a) Mixture 30_C; (b) Mixture 30_O;

(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.

4.5.2. Durability Assessment

The Cantabro test was used to evaluate the durability of AC mixtures (Cox et al., 2017) and was performed on designs, non-reheats, and reheats. Figure 4-4 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 CML values in Table 4-2. 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. Volumetric 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 (Diefenderfer et al., 2021). A more detailed comparison of reheats and non-reheats can be found in Table A-5 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 4-2.

Table 4-3 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.

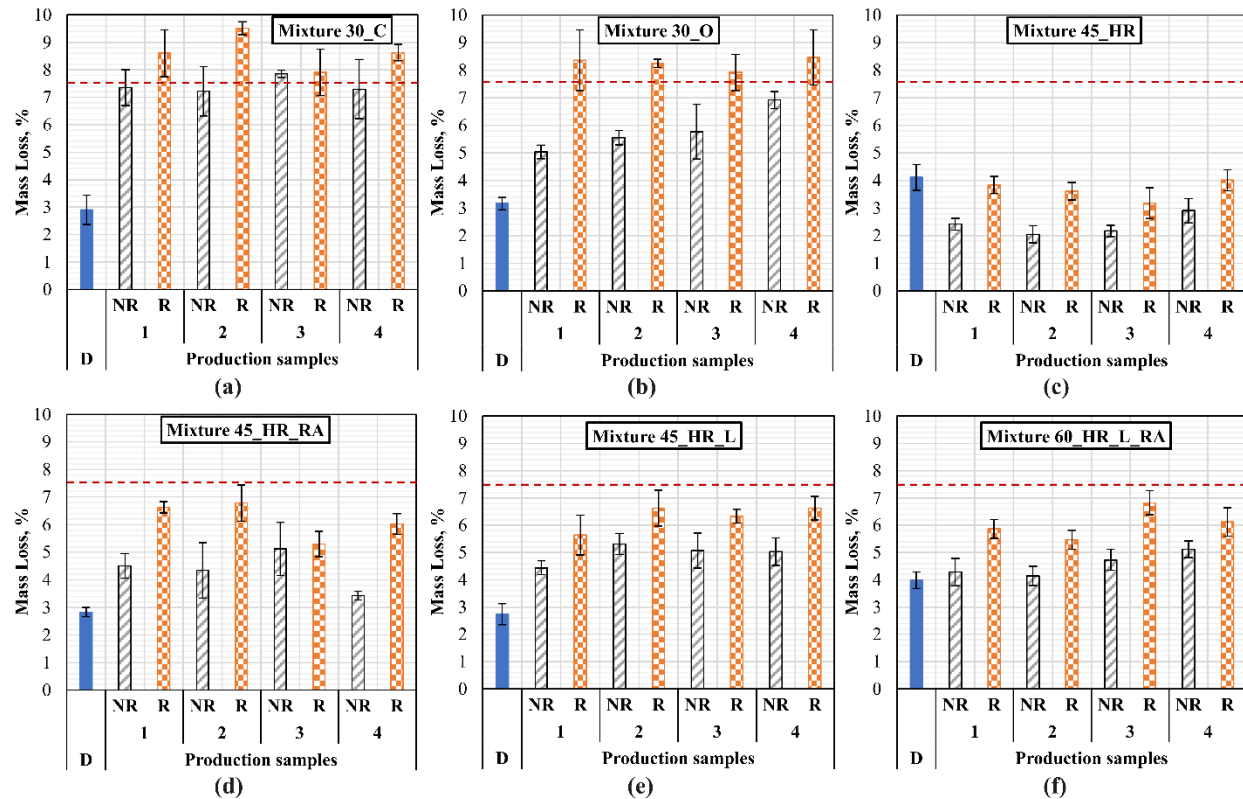


Figure 4-4. Cantabro Mass Loss Results for Mixtures: (a) Mixture 30_C; (b) Mixture 30_O; (c) Mixture 45_HR; (d) Mixture 45_HR_RA; (e) Mixture 45_HR_L; and (f) Mixture 60_HR_L_RA. Values are the average of three replicates; I-bars indicate $\pm$ one standard deviation; C = control; O = optimized; HR =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

Table 4-3. 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.

4.5.3. Cracking Performance Assessment

The IDT-CT (also known as IDEAL-CT) is being used by VDOT to assess mixture resistance to cracking. Figure 4-5 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 4-3.

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-4 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 (Diefenderfer et al., 2021). The sensitivity to reheating varied among different mixtures. In the pairwise comparisons using the KW test between samples presented in Table A-6 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 $7\pm0.5\%$. 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% (Habbouche et al., 2022). 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.

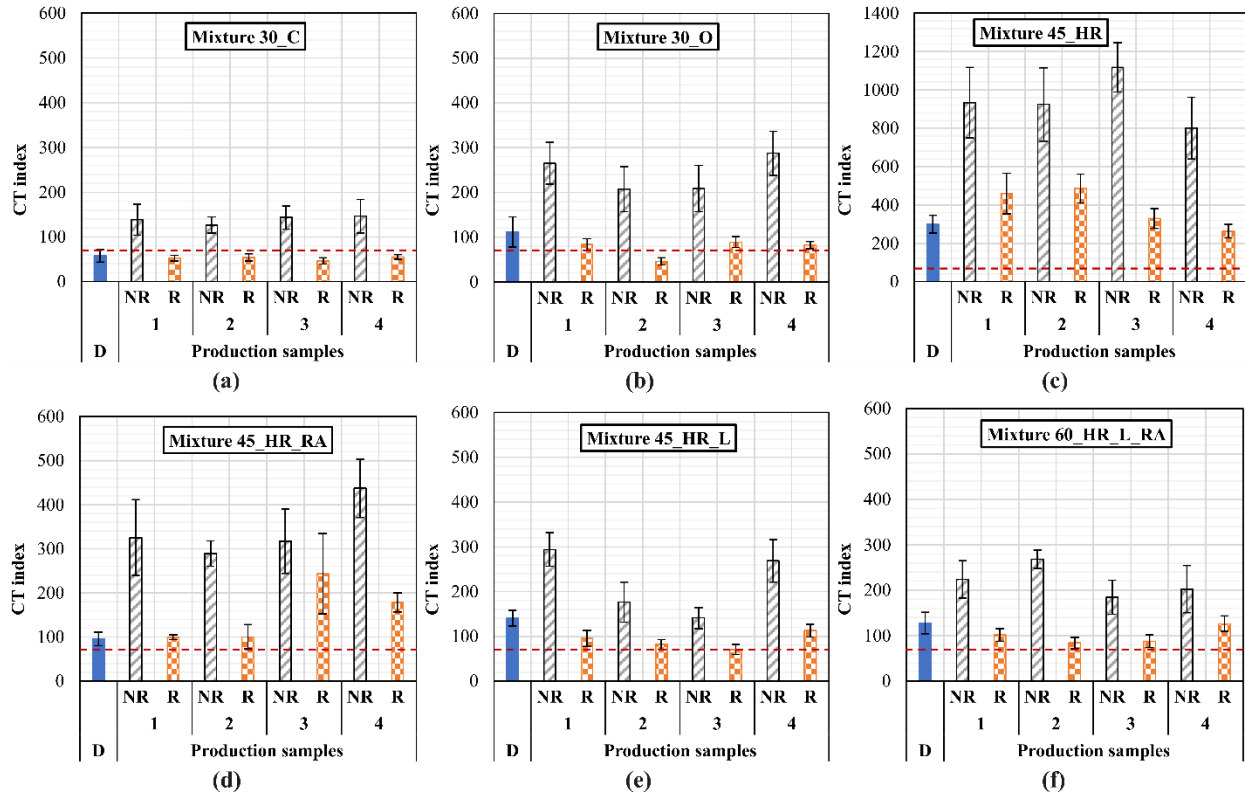


Figure 4-5. Performance Test Data for IDT-CT: (a) Mixture 30_C; (b) Mixture 30_O; (c) Mixture 45_HR; (d) Mixture 45_HR_RA; (e) Mixture 45_HR_L; and (f) Mixture 60_HR_L_RA. Values are the average of five replicates. I-bars indicate $\pm$ one standard deviation; CT = cracking tolerance; C = control; O = 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.

4.5.4. Rutting Performance Assessment

The APA rut test was conducted in accordance with AASHTO T 340. 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 4-6. 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 1 and 3 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-5 of the Appendix. A more detailed comparison, presented in Table A-6 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%.

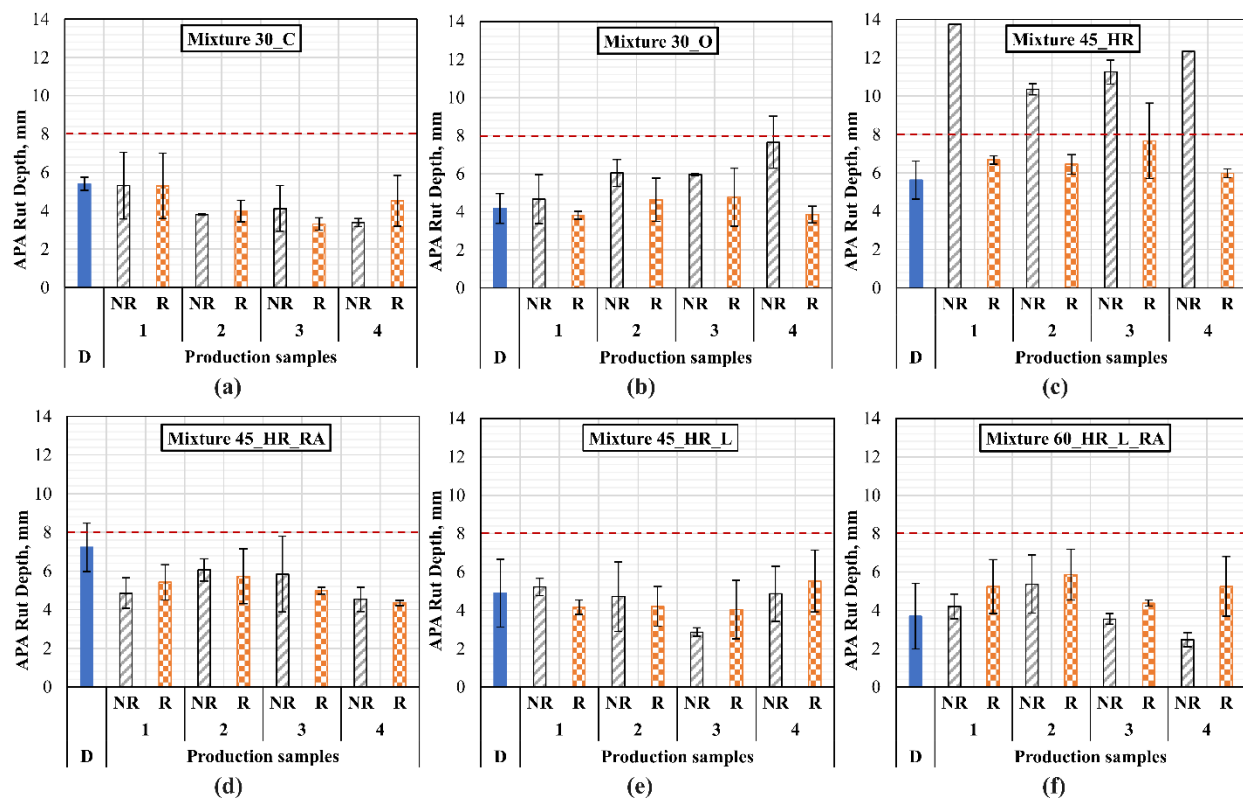


Figure 4-6. Performance Test Data for APA Rut Depth at 64°C and 8,000 Loading Cycles: (a) Mixture 30_C; (b) Mixture 30_O; (c) Mixture 45_HR; (d) Mixture 45_HR_RA; (e) Mixture 45_HR_L; and (f) Mixture 60_HR_L_RA. Values are the average of three replicates. I-bars indicate $\pm$ one standard deviation; 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.

4.6. Additional Discussion and Analyses

4.6.1. 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 4-7 using mean sample values. Figure 4-7a and Figure 4-7b 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 4-7a. 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 (Boz et al., 2023) 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 4-7b. Figure 4-7c 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.

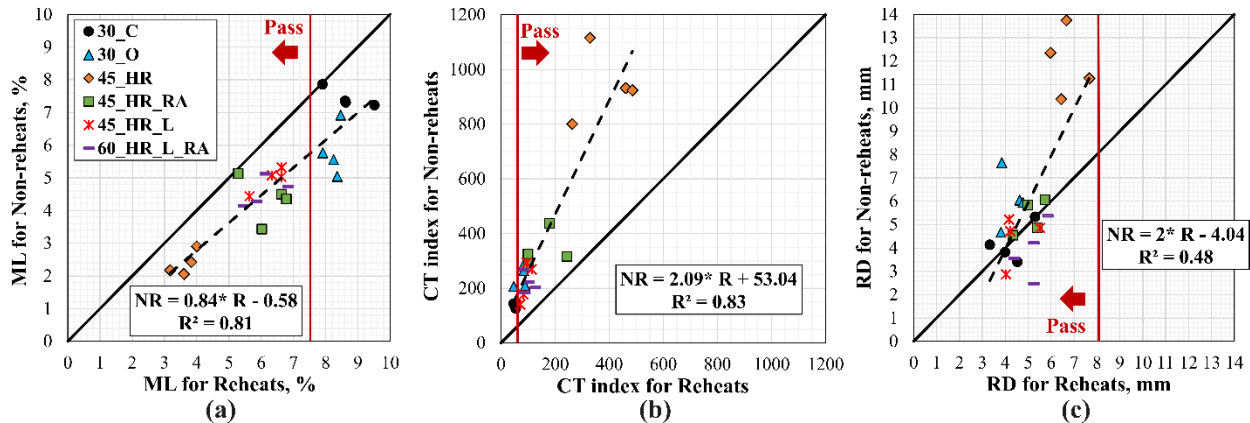


Figure 4-7. Relationship Between Non-reheats and Reheats: (a) Cantabro Test; (b) IDT-CT Test; (c) APA Rutting Test. RD = rut depth; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; NR = non-reheats; R = reheats; Solid

black line = equality line. Solid red line = VDOT's balanced mix design limit for reheated asphalt surface mixtures.

4.6.2. 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 (Bowers et al., 2023). Figure 4-8a, Figure 4-8b, Figure 4-8c 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 4-8d 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.

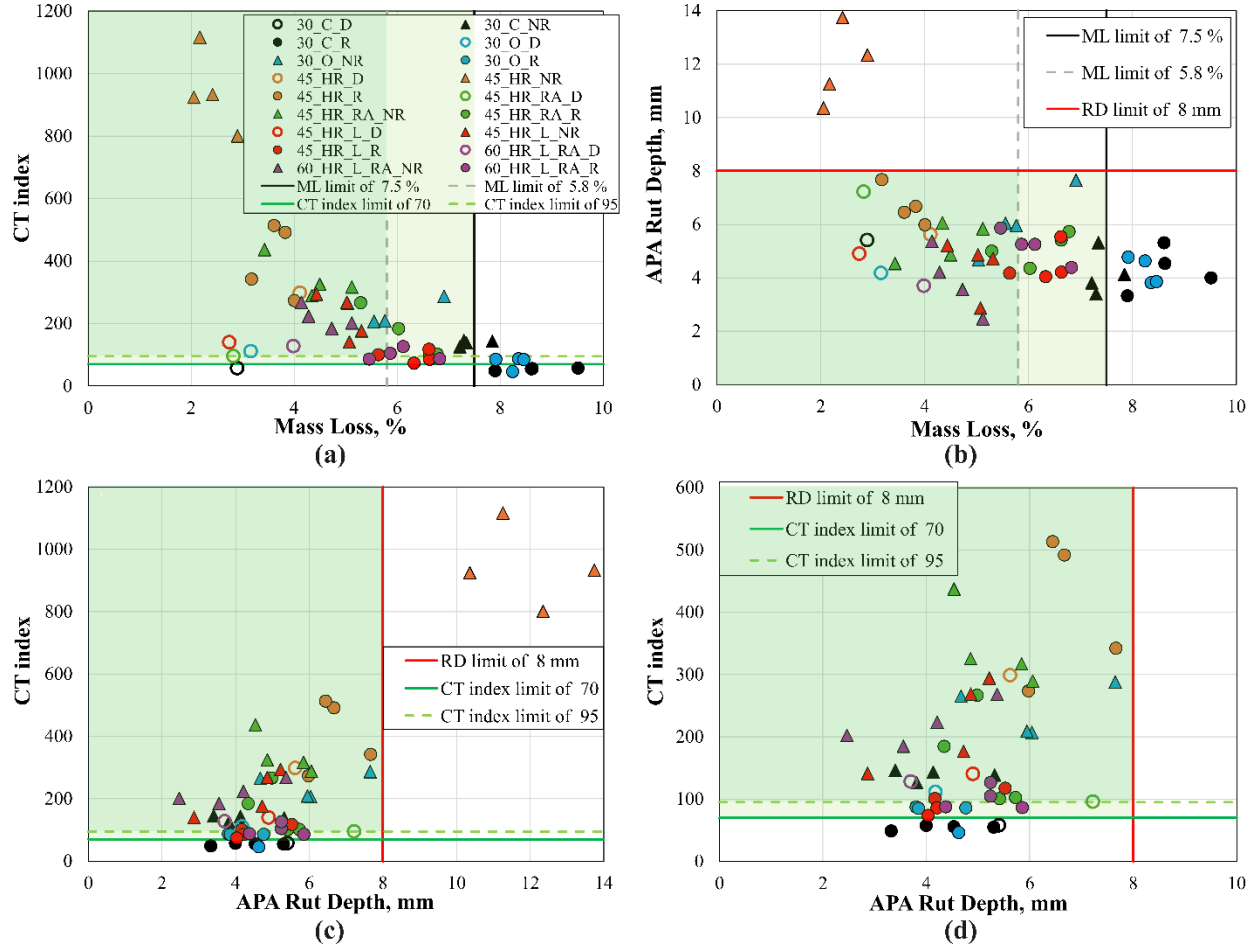


Figure 4-8. Comparison of Mean Performance for Each Mixture Variation: (a) IDT-CT vs. Cantabro test; (b) APA Rutting Test vs. Cantabro test; (c) IDT-CT vs. APA Rutting Test; (d) IDT-CT vs. APA Rutting Test without non-reheats of mixture 45_HR. CT = cracking tolerance; RD = rut depth; D = design; NR = 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. (2024) 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 4-1 through 4-3. 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 (4-3) is formulated differently compared to Equations (4-1) and (4-2). Table 4-4 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 4-9 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 4-4. 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.

$$x = \text{Normalized ML} = \frac{ML}{ML \text{ threshold}} \quad \text{Equation 4-1}$$

$$y = \text{Normalized RD} = \frac{RD}{RD \text{ threshold}} \quad \text{Equation 4-2}$$

$$z = \text{Normalized CT index} = \frac{CT \text{ index threshold}}{CT \text{ index}} \quad \text{Equation 4-3}$$

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

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 Figure 4-8c and Figure 4-8d), 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 Figure 4-9a and Figure 4-9b. 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 4-4 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 3-1 to 3-3). 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 4-4. 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 1	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 2	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 3	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 4	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 1	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 2	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 3	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 4	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 1	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 2	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 3	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 4	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 1	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 2	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 3	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 4	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.

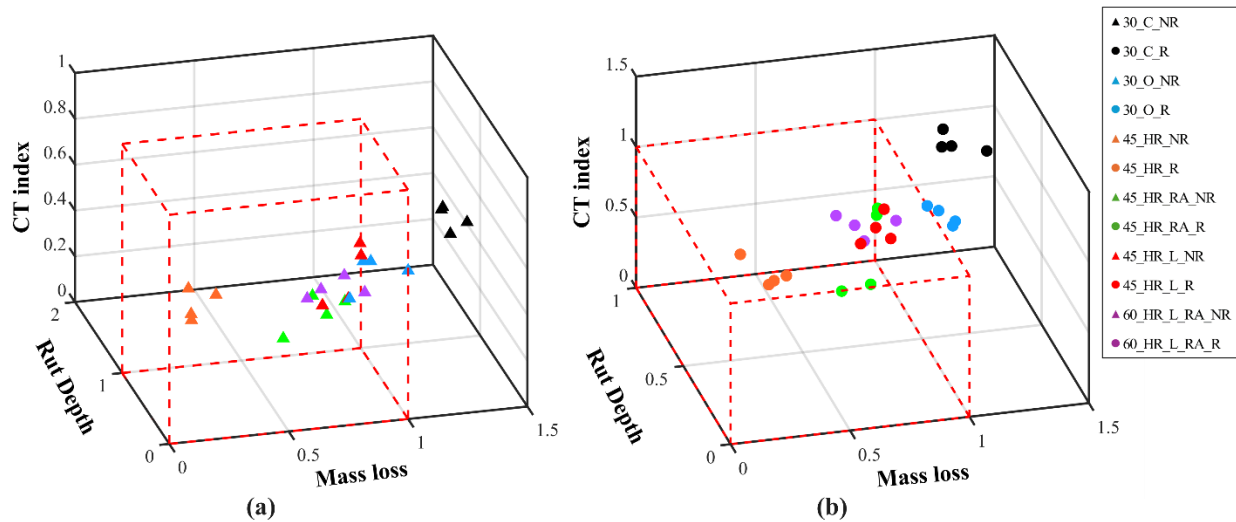


Figure 4-9. 3-D Plots: (a) Non-reheated Specimens; (b) Reheated Specimens. NR = non-reheats; R = reheats; C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; the red unit cube indicates the well-balanced area for the three normalized primary performance indices.

4.7. 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 $7\pm 0.5\%$ 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.

4.8. Reference

Abu Saq, M., Kassem, E., & Muftah, A. (2023). Cracking Performance Evaluation of RAP Mixes. *Journal of Transportation Engineering, Part B: Pavements*, 149(1), 04022072.

- American Association of State Highway and Transportation Officials. (2017). Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards. Washington, D.C..
- ASTM D8225. (2019). Standard Test Method for Determination of Cracking Tolerance Index of Asphalt Mixture Using the Indirect Tensile Cracking Test at Intermediate Temperature.
- Bowers, B. F., Diefenderfer, S. D., Moore, N., & Lynn, T. (2022). Balanced Mix Design and Benchmarking: A Case Study in Establishing Performance Test Thresholds. *Transportation Research Record*, 2676(12), 586-598.
- Bowers, B. F., Lynn, T., Yin, F., Moore, N., Diefenderfer, S. D., & Boz, I. (2023). Impact of Production Variability on Balanced Mix Designs in Virginia (No. FHWA/VTRC 23-R20). Virginia Transportation Research Council (VTRC).
- Boz, I., Habbouche, J., & Diefenderfer, S. (2021). 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).
- Boz, I., Diefenderfer, S. D., and Habbouche, J. (2023). 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.
- Braver, S. L., MacKinnon, D. P., & Page, M. (2003). *Levine's Guide to SPSS for Analysis of Variance*. Psychology Press.
- Brown, E. R., Kandhal, P. S., Roberts, F. L., Kim, Y. R., Lee, D. Y., & Kennedy, T. W. (2009). *Hot Mix Asphalt Materials, Mixture Design, and Construction*. NAPA Research and Education Foundation.
- Cox, B. C., Smith, B. T., Howard, I. L., & James, R. S. (2017). State of Knowledge for Cantabro Testing of Dense Graded Asphalt. *Journal of Materials in Civil Engineering*, 29(10), 04017174.
- Diefenderfer, S. D., & Bowers, B. F. (2019). Initial Approach to Performance (Balanced) Mix Design: The Virginia Experience. *Transportation Research Record*, 2673(2), 335-345.
- Diefenderfer, S. D., Habbouche, J., & Boz, I. (2023). Balanced Mix Design for Surface Mixtures: 2020 Field Trials (No. FHWA/VTRC 23-R13). Virginia Transportation Research Council, Charlottesville.
- Diefenderfer, S.D., Boz, I., and Habbouche, J. (2021b). Balanced Mix Design for Surface Asphalt Mixtures: 2019 Field Trials. Publication VTRC 21-R21. Virginia Transportation Research Council, Charlottesville.
- Martin, A. E., Kaseer, F., Arámbula-Mercado, E., Bajaj, A., Cucalon, L. G., Yin, F., ... & King, G. (2020). *Evaluating the Effects of Recycling Agents on Asphalt Mixtures with High RAS and RAP Binder Ratios* (No. Project 09-58).
- Golalipour, A., Veginati, V., & Mensching, D. J. (2021). Evaluation of Asphalt Mixture Performance Using Cracking and Durability Tests at a Full-Scale Pavement Facility. *Transportation Research Record*, 2675(11), 226-236.

- Habbouche, J., Boz, I., & Diefenderfer, S. D. (2022). Validation of Performance-Based Specifications for Surface Asphalt Mixtures in Virginia. *Transportation Research Record*, 2676(5), 277-296.
- Habbouche, J., Boz, I., & Diefenderfer, S. D. (2022). Interlaboratory Study for the Indirect Tensile Cracking Test at Intermediate Temperature: Phase II (No. FHWA/VTRC 23-R3). Virginia Transportation Research Council, Charlottesville.
- Hughes, C. S. (2005). NCHRP Synthesis 346: State Construction Quality Assurance Programs. Transportation Research Board, Washington DC.
- McDaniel, R. S., Shah, A., Huber, G. A., & Copeland, A. (2012). Effects of Reclaimed Asphalt Pavement Content and Virgin Binder Grade on Properties of Plant Produced Mixtures. *Road Materials and Pavement Design*, 13(sup1), 161-182.
- Mohammad, L. N., Elseifi, M. A., Cooper III, S. B., Hughes, C. S., Button, J. W., & Dukatz Jr, E. L. (2016). Comparing the Volumetric and Mechanical Properties of Laboratory and Field Specimens of Asphalt Concrete (No. Project 09-48).
- Nair, H., Diefenderfer, S. D., & Bowers, B. F. (2019). Assessing Increased Use of Reclaimed Asphalt Pavement in Asphalt Mixtures (No. FHWA/VTRC 20-R9). Virginia Transportation Research Council, Charlottesville.
- Ostertagova, E., Ostertag, O., & Kováč, J. (2014). Methodology and Application of The Kruskal-Wallis Test. *Applied mechanics and materials*, 611, 115-120.
- Sabouri, M. (2020). Evaluation of Performance-Based Mix Design for Asphalt Mixtures Containing Reclaimed Asphalt Pavement (RAP). *Construction and Building Materials*, 235, 117545.
- Tong, B., Habbouche, J., Diefenderfer, S. D., & Flintsch, G. W. (2024). Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia. *International Journal of Pavement Engineering*, 25(1), 2325609.
- Tong, B., Habbouche, J., Flintsch, G. W., & Diefenderfer, B. K. (2023). Rutting Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents Under Full-Scale Accelerated Testing. *Materials*, 16(24), 7611.
- Virginia Department of Transportation. (2020). Road and Bridge Specifications. Richmond.
- West, R., Rodezno, C., Leiva, F., & Yin, F. (2018). Development of a Framework for Balanced Mix Design. Project NCHRP, 20-07.
- Yin, F., & West, R. C. (2021). Balanced Mix Design Resource Guide (No. IS-143).
- Zaumanis, M., Mallick, R. B., & Frank, R. (2013). Evaluation of Rejuvenator's Effectiveness with Conventional Mix Testing for 100% Reclaimed Asphalt Pavement Mixtures. *Transportation Research Record*, 2370(1), 17-25.

CHAPTER 5 - Machine Learning-Based Prediction and Optimization of Balanced Mixture Design Performance Indices⁵

5.1. Abstract

The BMD concept is an emerging methodology that facilitates the design of engineered asphalt mixtures. This approach is particularly beneficial for mixtures containing conventional and high reclaimed asphalt pavement contents, for which the traditional volumetric design methods may fail to effectively address the performance characteristics. However, due to production variability, these engineered mixtures can still fail to meet the required thresholds. Additionally, identifying the cause of this imbalance is challenging. To maximize the benefits of BMD implementation, this study introduces machine learning (ML) algorithms including linear regression (LR), random forest (RF), extreme gradient boosting (XGB), and support vector regression (SVR) as strategic tools to predict mixtures' BMD performance indices. 648 specimens fabricated for quality acceptance as part of the 2020 Virginia Accelerated Pavement Testing Program is used for the modeling and analysis. Durability, cracking, and rutting susceptibility of the specimens were evaluated using the Cantabro test, indirect tensile cracking test (IDT-CT), and Asphalt Pavement Analyzer (APA) rut test, respectively. Key outcomes include: (a) ML models, including RF, XGB, and SVR, demonstrated superior performance compared to LR; (b) Feature importance analysis from ML models identified dominant factors for each BMD test. Additionally, the reheating process was highlighted; (c) A pseudo in-situ deployment was simulated to optimize BMD implementation. The dimensionality reduction analysis — uniform manifold approximation and projection — highlighted the practical challenges associated with concurrently improving multiple performance metrics. Ultimately, the pivotal role of ML in advancing both the design and production phases was emphasized.

Keywords: machine learning, random forest, XGBoost, SVR, BMD, RAP, production variability.

⁵ This paper, with Bilin Tong as the first author, has been submitted for publication at the *Transportation Research Records* in October 2024. It is co-authored by Wenjiang Huang, Ph.D., Jhony Habbouche, Ph.D., P.E., Ilker Boz, Ph.D., P.E., Qing Guo, Ph.D., Stacey D. Diefenderfer, Ph.D., P.E., Gerardo W. Flintsch, Ph.D., P.E.. Contribution: Conceptualization: Lead; Methodology: Lead; Resources: Lead; Data Processing and Analysis: Lead; Modelling: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead.

5.2. Introduction

In the ever-evolving landscape of asphalt mixture design, the industry faces significant challenges driven by the necessity for high-performance pavements, the increasing demand for sustainable practices, and the escalating costs of raw materials (Golalipour et al., 2021). Current applications of the Superpave methodology have predominantly focused on volumetric properties of asphalt mixtures in the U.S.. However, the dynamic nature of contemporary asphalt materials, characterized by the widespread use of the extensive application of RA and additives, necessitates a shift towards a design philosophy that emphasizes long-term performance.

The BMD methodology has emerged as a pivotal innovation in this context. As delineated by the National Asphalt Pavement Association (NAPA) and the American Association of State Highway and Transportation Officials (AASHTO), BMD is characterized as 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* (NCAT & NAPA 2023; AASHTO 2021)”. The BMD methodology introduces performance testing criteria as a replacement for certain traditional aspects, such as volumetric properties, focusing on evaluating the resistance of mixtures to common distresses, such as rutting and cracking. To gain acceptance, BMD mixtures are required to meet defined performance benchmarks during the design phase, with the possibility of additional performance tests during the production phase. Additionally, the integration of BMD methodology with the use of RAs and/or softer binders, presents a strategic approach to design and production of engineered HRAP mixtures, providing an innovative and practical response to the demands of contemporary pavement construction (Tong et al., 2024; West et al., 2018). In line with this advancement, a growing number of state highway agencies are actively investigating and implementing the BMD framework for the design and approval of asphalt mixtures (West et al., 2018; Yin and West, 2021; Mogawer et al., 2023).

Implementing the BMD framework typically involves two primary steps. The first step initiates with a design phase, wherein different state agencies may select varied foundational BMD tests and corresponding thresholds based on their specific requirements. For instance, in Virginia, three fast, simple, and practical performance-indicative tests were adopted for BMD SMs. These tests include the Cantabro test, capped at a maximum CML of 7.5%, to assess durability; the IDT-CT at intermediate temperatures, with a minimum CT index of 70, for cracking performance assessment with short-term aged conditions; and the APA rut test, with a maximum rut depth (RD) limit of 8 mm, for rutting resistance evaluation (Bowers et al., 2022; Habbouche et al., 2021). The second step aims to verify that the design is reproduced during production by ensuring that the performance anticipated from the laboratory-designed mixtures is realized in the field through a QA program. If performance criteria are not met (one or more indices exceed predetermined thresholds), the mixture is deemed unbalanced, thus diminishing the full potential benefits of BMD implementation (Lynn et al., 2024).

Variations in asphalt mixture properties from the original design may arise during production and construction phases, because of the inherent variability in environmental conditions, plant operations, and natural materials (Lynn et al., 2024). Additionally, the designed JMF is subjected to verification and potential plant modifications to account for actual production and on-site circumstances, irrespective of mixture design (Brown et al., 2009). This process poses significant challenges when applying the BMD methodology, which deviates from conventional volumetric property specification, thereby introducing potential difficulties for contractors. Furthermore, replicating the scale of mixture production that occurs in a plant setting, following laboratory design protocols, can be an intricate endeavor (Brown et al., 2009). It is important to note that differences in asphalt mixture properties from the original design often trigger premature manifestations of pavement distress or can potentially result in outright structural failure (Lynn et al., 2024, Mohammad et al., 2016). To identify and assess the sources, causes, and degrees of variability in gradation, volumetric, and mechanical properties, NCHRP Project 09-48 (Report 818) examined laboratory-mixed, laboratory-compacted specimens; plant-mixed, laboratory-compacted specimens used in volumetric acceptance testing; and plant-mixed, field-compacted specimens employed during density acceptance testing and forensic evaluation of in-place pavement. The aim of these assessments was to identify the sources of variability and to evaluate the precision and bias associated with the volumetric and mechanical properties of dense-graded asphalt mixtures (Mohammad et al., 2016). Alternatively, the quality assessment can be done using reheated plant-mixed, laboratory-compacted specimens (reheated specimens), a practice observed in the VDOT independent assurance practice (Diefenderfer et al., 2023).

In 2020, VTRC, VDOT, and Virginia Tech initiated collaborative research involving the design, placement, and laboratory and accelerated field testing of a control mixture and five BMD mixtures at the Virginia APT facility, owned by VDOT and situated at the VTTI. This initiative aimed, in part, to apply VDOT BMD special provisions to production under highly controlled condition, offering an opportunity to scrutinize the implications of these specifications on different facets, including design, production, quality control and assurance practices, and construction of HRAP SMs (Tong et al., 2023).

In the current landscape, characterized by the exponential growth of data and the continuous advancement of iterative algorithms, the reemergence of machine learning (ML) approaches presents a promising avenue for enhancing the efficiency of analyzing data and identifying key factors. In the Virginia 2020 APT program, extensive fundamental performance tests on BMD samples in the field production were carried out, alongside analyses of the associated variability (Tong et al., 2024). Specimens fabricated for quality acceptance as part of the Virginia APT program were used for the ML modeling and analysis in this study.

5.3. Objective and Scope

The primary objective of this research was to develop ML models capable of predicting BMD performance indices based on mixture constituent and volumetric properties collected during

both the design and field production phases at the project level. Four machine learning algorithms—linear regression (LR), random forest (RF), extreme gradient boosting (XGB), and support vector regression (SVR)—were employed to predict the performance of BMD mixtures. The performance of these different ML models was compared and analyzed. The secondary goal of this study was to use these developed ML models to assess the influence of constituent properties on the performance of BMD mixtures and to identify the dominant factors in BMD specimen preparation. This was particularly challenging for traditional approaches due to the complex interactions of multiple constituent properties on the performance indices of asphalt mixtures. Finally, this study proposed the incorporation of ML into the design phase of BMD. To further demonstrate the application of ML models in BMD implementation, a simulated BMD deployment was strategically designed. Through simultaneous optimization across three primary BMD tests, optimized or ML-enhanced candidates were selected.

5.4. Background and Literature Review

5.4.1. Production Variability

The variability of different performance tests has been reported in many studies. Gopalipour et al. (2021) evaluated the performance of reheated specimens as part of the 2013 Federal Highway Administration's (FHWA's) Accelerated Loading Facility (ALF) study. This study determined coefficient of variance (COV) of six performance tests including: the Cantabro mass loss test, the N_{flex} factor test, the overlay test, the I-FIT test, the IDT-CT test, and the cyclic fatigue test. The FI parameter from I-FIT presented the highest average COV (up to 34%) among all tests, while CT index achieved COV of 40% in Lane 7. Buttlar et al. (Buttlar et al., 2020) conducted cracking tests, including the Disk-Shaped Compact Tension Test (DCT), I-FIT test, and the IDEA-CT test, as part of the development of a local BMD method. Their findings indicated that while the average COV for DCT in field section evaluations was 19.5%, the I-FIT test showed a significantly higher average COV of 52.2%. In preparation for the full implementation of the BMD methodology, the VDOT initiated a round-robin study to assess testing repeatability across various laboratories. Forty-one laboratories, including those from the agencies, contractors, and independent testing facilities, participated in the initiative. Two distinct mix designs were utilized, with 46 sets of test specimens distributed for each design. Phase 1 of the study focused on evaluating IDEAL-CT test results from different labs, while Phase 2 commenced in 2021 (Habbouche et al., 2022). Additionally, currently accepted production tolerance of volumetric properties and gradation in Virginia can be found in (VDOT, 2020). Current VDOT specified COV value of 18.3% for the CT index for 5 replicates.

Concerns about unbalanced performance in the resultant mixture, due to high variability in performance tests as reported in previous studies, have led researchers to dedicate efforts towards establishing tolerance limits for constituent properties during loose mixture sampling and specimen fabrication. However, recent research has found that a BMD mixture can still become unbalanced during production, even within specified tolerance limits. In the early stages of the

BMD concept, Austerman et al. (2018) modified asphalt binder content, gradation passing the No.200 sieve, and asphalt binder source in mixture to represent production variations. The study revealed that mixture performances, such as moisture susceptibility, may become unbalanced during production. They further noted that specific production factors can impact volumetric properties and the performance of mixtures developed using the BMD concept. Subsequently, Mogawer et al. (2023) indicated that some mixtures produced at the lower limit of the OBC production tolerance, with or without alterations in aggregate gradation production tolerances, exhibited heightened susceptibility to fatigue cracking. Furthermore, it concluded that the binder source that can change during production results in unbalanced BMD mixtures. Finally, it was recommended that a mixture's resistance to rutting be assessed at the upper production tolerance limit for OBC and for cracking at the following conditions: (1) the lower production tolerance limit for OBC; (2) the lower production tolerance limit for OBC in conjunction with the upper aggregate gradation production tolerance limits; and (3) the lower production tolerance limit for OBC combined with the lower aggregate gradation production tolerance limits. Recently, Bowers et al. (2023) conducted an evaluation of fourteen accepted mixture designs, adjusting their coarse and fine gradation along with binder content to specified tolerance limits. This study found mixtures exhibited excellent rutting performance but displayed susceptibility to durability and cracking issues as the gradation and binder content change. It further underscored the potential for initially balanced mixtures to become unbalanced due to production variability. In addition to replicating production variability in the laboratory, Tong et al. (2024b) showcased field production variability for one Superpave control mixture and five BMD mixtures, designed using BMD Approach D (performance only), incorporating both conventional and high RAP content. Their investigation unveiled the possibility of encountering unbalanced BMD mixtures. It is important to note that in BMD Approach D, tolerance limits for volumetric properties are exempted except for binder content.

As a result, whether using BMD approaches with specified production tolerance or relying solely on performance-based BMD approach, concerns persist regarding the potential occurrence of unbalanced produced mixtures. Researchers may continue to propose refinements of BMD specifications, specifically, to determine an acceptable tolerance level (deviation from design) for gradation, asphalt content (AC), and volumetric properties, within which test values are still deemed to constitute a qualified or balanced mixture (Lynn et al., 2024). However, simulating production variability in the laboratory and evaluating modified mixtures is a highly complex process, imposing significant costs in terms of manpower and materials. Additionally, efforts are also underway to model how component materials or mixtures properties influence corresponding mixture performance, as emphasized in several VTRC reports (Diefenderfer et al., 2023, Bowers et al., 2023). Yet, in a practical project, identifying the critical factors is highly intricate, as simple linear relationships may not adequately interpret the specific causes of production variability. This complexity arises due to the multitude of variables involved in production, underscoring the importance of employing nonlinear regression analyses.

5.4.2. Machine Learning Approaches

Some common ML approaches have been used in related prior studies (Rahman et al., 2021). RF is a robust ensemble method that constructs multiple decision trees during the training phase. Each tree is developed using a randomly chosen subset of training data and features, enhancing model stability by mitigating overfitting and elevating prediction accuracy. Notably, RF effectively processes both categorical and continuous data and offers valuable insights into feature importance, aiding in the identification of critical predictors (Breiman., 2001). XGB, another tree-based model, refines gradient boosting techniques by incorporating features such as built-in regularization and dynamic stopping criteria to counteract overfitting (Chen and Guestrin, 2016). SVR, an extension of support vector machines (Vapnik et al., 1996), is designed to predict values within a specified deviation for all training data, with a focus on minimizing the upper bound of generalization errors, thereby enhancing model robustness. SVR utilizes kernel functions to map input vectors into a higher-dimensional space, facilitating linear regression analysis and ensuring robust generalization performance across diverse datasets (Smola et al., 2004). The emergence of ML applications across a broad spectrum of industries exemplifies its wide-ranging utility. For example, in the context of asphalt concrete, Zeiada et al. (Zeiada et al., 2020) employed an artificial neural network (ANN) in conjunction with a forward sequential feature selection algorithm to identify the most influential design factors prevailing in warm climate regions using data sourced from the Long-Term Pavement Performance (LTPP) database. The efficacy of these applications can be attributed to the integration of feature importance, a main component of various ML techniques. Additionally, ML plays an increasingly vital role in predicting numerous essential parameters in asphalt concrete or asphalt pavement such as dynamic modulus (Daneshvar and Behnood, 2022), air void (Zavrtanik et al., 2016), rutting performance (Liu et al., 2022), fatigue performance (Ahmed et al., 2017), international roughness index (IRI) (Zhou et al., 2021), and CT-index (Nguyen et al., 2023). Rahman et al. (2021) explored the use of ML with large, diverse datasets to predict asphalt concrete metrics, such as IDT strength and HWTT test rut depth, marking a pivotal discussion on ML's integration into performance frameworks. They noted biases from trial batch samples and a significant data reduction due to missing values, decreasing datasets from 17,920 to 2,408 for HWT and 2,189 for IDT. Moreover, training ML models emphasized data quality over quantity, acknowledging that data variability often affects prediction accuracy, challenging the robustness and generalizability of models for BMD mixtures. Throughout the specimen preparation process, considerable variability can emerge from variations in technician handling, experimental conditions, and testing apparatus. This variability proves challenging to accurately incorporate as a quantifiable variable in the training of ML models. Therefore, conducting performance testing and data collection under rigorously controlled conditions may significantly mitigate the adverse effects on data quality resulting from disparate sources. Once a reliable ML model is developed, it can bridge the gap between the initial mix design and actual field performance by predicting the effects of mixture components and production variables on final product. Leveraging data fabricated from a highly controlled experiment schedule, ML

algorithms can offer valuable insights into potential performance outcomes before field implementation, allowing for proactive adjustments and maximizing the benefits of BMD implementation.

5.5. Methodology

5.5.1. Experiments

A total of one control and five BMD 9.5mm dense-graded SMs were designed, produced, and sampled for evaluation in the 2020 Virginia APT program. The mixtures incorporated various combinations of three RAP contents, two binder PGs, one RA, and one WMA additive. The RAP content in the mixtures was determined based on the total weight of the resulting mixtures. A 100% recycled binder availability was assumed, and the design asphalt contents accounted for the contribution of asphalt binder from RAP materials. The six mixtures are defined in Table 3-1 of Chapter 3.

5.5.2. BMD Testing

Three BMD tests adopted in Virginia (shown in Figure 5-1) are: Cantabro test, IDT-CT, and APA test, which assess durability, cracking, and rutting, respectively. The CML was determined in accordance with AASHTO T 401 (AASHTO 2017) on three design specimens that were used to determine volumetric properties. The IDT-CT was conducted at a temperature of 25°C in accordance with ASTM D8225-19 (ASTM 2019) on five replicate test specimens compacted at $7 \pm 0.5\%$ air voids. The APA test was performed in accordance with AASHTO T 340 (AASHTO 2017) at a temperature of 64°C; four replicate test specimens were tested in an APA Junior machine. The APA specimens were compacted at $7 \pm 0.5\%$ air voids. VDOT BMD test thresholds were indicated in the **Introduction** section.

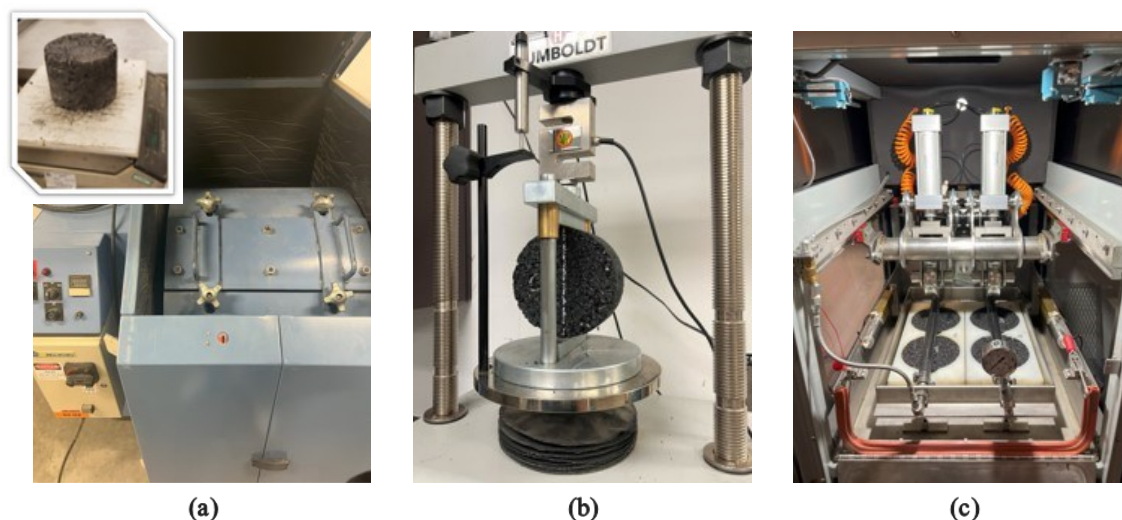


Figure 5-1. VDOT BMD Tests: (a) Cantabro Mass Loss Test; (b) IDT-CT; and (c) APA Rut Test. VDOT = Virginia Department of Transportation; BMD = balanced mix design; IDT-CT = indirect tensile cracking test; APA = asphalt pavement analyzer.

5.5.3. Production Sampling and Specimen Fabrication

A total of 648 specimens fabricated for quality acceptance as part of the 2020 Virginia APT experiment were used in this study. In addition to testing performance during the design process, a comprehensive laboratory assessment focusing on durability, cracking resistance, and rutting performance was conducted on mixtures collected during production. The test specimens included laboratory-mixed laboratory-compacted specimens (design), plant-mixed laboratory-compacted specimens (non-reheats), and plant-mixed laboratory-compacted specimens fabricated after a reheating process (reheats). Designs were laboratory-produced and laboratory-compacted, fabricated for mixture design purposes, while the produced mixtures were sampled to monitor their quality and associated production variability. Sampling of the loose mixture was conducted four times daily during production. The production sampling plan for each mixture is delineated in Figure 5-2. Samples were sequentially labeled from 1st to 4th (A through D) for each mixture. 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 promptly compacted into test specimens—these were referred to as plant specimens (non-reheats) - or they were placed in boxes, transported to VTRC, and stored in a climate-controlled area for the fabrication of reheated specimens at a later time (reheats) (Tong et al., 2024b).

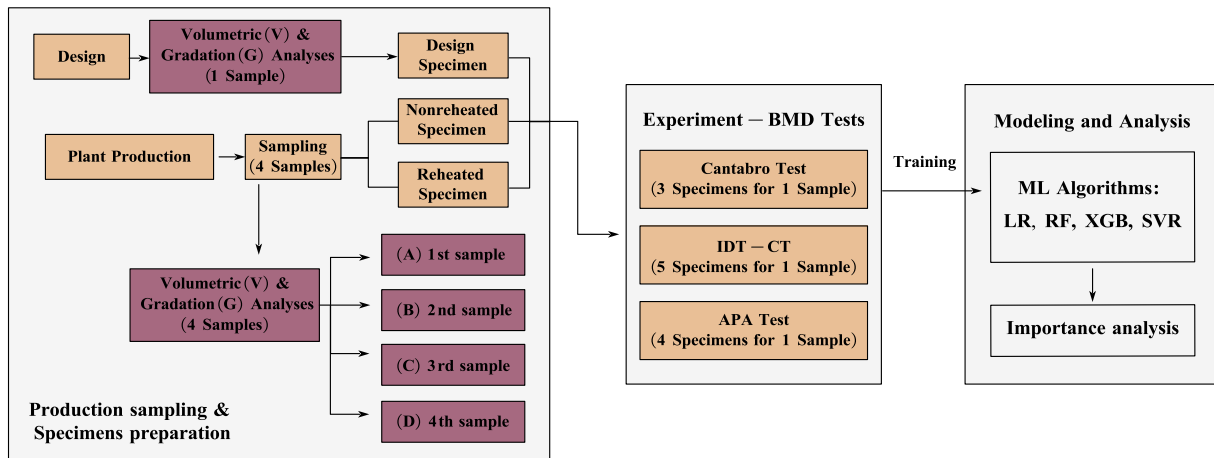


Figure 5-2. Experimental and Machine Learning (ML) Training Workflows. Left block: production sampling and specimen preparation. Middle block: balanced mix design (BMD) tests. Right block: ML models used in predicting three primary BMD tests. IDT-CT = indirect tensile cracking test; APA = asphalt pavement analyzer; G = aggregate gradation sieve analysis; V = volumetric properties analysis.

5.5.4. Modeling Approach

In this study, several ML algorithms were used to predict outcomes of three BMD tests in terms of CML, CT index, and APA rut depth. The employed models included LR (the simplest ML model), RF, XGB, and SVR. Figure 5-2 illustrates the experimental and ML training workflows. The left block displays the experimental procedures (detailed in the **Experiments** section), the middle block represents the three BMD tests (detailed in the **BMD testing** section),

and the right block shows the ML algorithms used to predict those properties in BMD tests. The targets of ML prediction were the current VDOT adopted BMD mixture performance indices, which include CML, CT index, and APA rut depth. In developing the ML models, the dataset was randomly divided into training and testing sets in an 80:20 ratio. Feature normalization was applied to the training set, with the same scaling coefficients subsequently used on the testing set. Hyperparameter optimization for each ML model was conducted via exhaustive grid-search method. For the RF model, hyperparameter variations included the number of trees (100 to 2000), tree depth (10 to 30), and minimum samples per leaf node (1 to 4). The XGB model parameters ranged from the number of trees (100 to 2000), learning rate (0.0001 to 0.1), tree depth (3 to 9), and subsample ratios of training instances and columns (0.7, 0.8, or 1). The SVR model's hyperparameters were the regularization parameter (0.001 to 100), epsilon (0.001 to 1), and kernel type (polynomial, radial basis, or sigmoid). The settings in an ML model that produced the highest mean R^2 values from the test set were selected as the optimal hyperparameters. The Scikit-learn library was utilized for constructing the ML models. Finally, to assess the influence of constituent properties on the performance of BMD mixtures and to identify the dominant factors in BMD specimen preparation, permutation importance, a model-agnostic method (Breiman, 2001, Kim et al., 2021), was utilized to interpret feature importance across all used ML models. This approach evaluates a feature's contribution to a model's predictive power by measuring performance changes when the feature's values are randomly shuffled, disrupting the correlation with the target variable. The comparative analysis of model performance, with and without the shuffled feature, quantifies each feature's impact.

The authors recognize that the data collected in this study was limited, but this method provided a highly controlled analysis environment and systematic data collection, reducing many avoidable issues such as data recording errors from different contractors or multiple laboratories. Predictions in a small dataset also demonstrate the effectiveness of ML and set an example for project-level analysis. Additionally, the approach and findings from this study can be adapted to accommodate other indices for mixture performance as different transportation agencies implement various BMD performance tests.

5.5.5. Training Features, Data Processing, and Evaluation Metrics

The input data collected for both designed and plant produced specimens contained 26 attributor/features, including categorical variables, volumetric properties, gradation parameters, and mechanical property. Categorical variables include design methods (encoded as 0 for conventional Superpave and 1 for BMD) and specimen types (encoded as 0 for design, 1 for non-reheats, and 2 for reheats). Volumetric properties consist of 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}), effective asphalt film thickness (F_{be}), and RAP content (100% blending ratio is assumed).

Gradation parameters are documented as the percentage passing through various sieve sizes: 12.5mm, 9.5mm, 4.75mm, 2.36mm, 1.18mm, 600μm, 300μm, 150μm, 75μm. The dynamic shear modulus (G^*) at 64°C (high performance grade temperature), obtained from binder extracted and recovered from loose mixtures of field production sampling, was used as a mechanical property in this study. These features incorporated in the modeling for design and non-reheated specimens were recorded by VTRC, while those for non-reheated specimens were documented by the VDOT district office. To address the challenges posed by limited data and remedy overfitting, Pearson correlation coefficients were calculated between each pair of features to determine the selection of eventual training features in ML models. This process, coupled with 5-fold cross-validation during training, ensures robust model performance, with further details provided in the **Results** section.

The datasets for each performance test comprised a different number of specimens: 162 for CML, 270 for IDT-CT, and 216 for APA RD tests. The APA tests were performed following AASHTO T-340 standard but only the average RD from two specimens on both left and right were reported, reducing the APA RD dataset to 108. Regarding missing data, the IDT-CT test had 3 missing entries, which were removed, while the APA test had two, which were filled using the median value of the corresponding RD. This approach to handling missing was empirically determined, considering the varying sizes of the dataset. Table 5-1 summarizes the details of each dataset.

Table 5-1. Data Sets/Specimens Used in ML for Three Balanced Mix Design Tests

Data sets	CML	IDT-CT	APA RD test
Data sets/Specimens fabricated	162/162	270/270	108/216
Missing data sets	0	3	2
Pre-reserved data for pseudo in-site deployment	10	10	10
Data sets in an ML model	152	257	98

CML = Cantabro mass loss; IDT-CT = indirect tensile cracking test; APA = asphalt pavement analyzer; RD = rut depth.

To evaluate model performance, two metrics—the coefficient of determination (R^2) (Equation 5-1) and the accuracy in terms of mean absolute percent error (MAPE) (Equation 5-2)—were used,

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad \text{Equation 5-1}$$

$$\text{Accuracy} = 1 - \text{MAPE} = 1 - \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad \text{Equation 5-2}$$

where y_i is the target values, $\bar{y}$ is the average of target values, $\hat{y}_i$ is the predicted value, and N is the number of data sets.

5.5.6. Simulating the BMD Deployment

Given the limited availability of materials for sampling during production and cost considerations, a pseudo in-situ deployment method is designed in this study to mimic real-world deployment scenarios for these ML models. For each dataset, 10 random specimens meeting specific criteria (CML < 7.5%, CT index > 70, APA < 8 mm) are pre-reserved and excluded from model development. Designing specimens in the laboratory that simultaneously satisfy these three primary criteria poses significant experimental challenges. This is because feature selection is highly empirical, and therefore the layered selection process based on various performance tests of interest often results in extensive repetitive work. Here, the aim of the ML pseudo in-situ deployment is to generate samples exhibiting all three balanced performances directly. ML models are trained with the remaining data (refer to Table 5-1) to predict each of these properties accurately (one ML model for one performance prediction). Once trained, samples are randomly generated based on the mean and standard deviation of current dataset's distribution to ensure they can be created in-situ, and their BMD performances are predicted by the corresponding ML models. Each models' confidence is assessed using the corresponding 10 pre-reserved specimens. For instance, if an ML model predicts that 9 out of 10 pre-reserved CML samples have values less than 7.5%, the confidence level of this model is determined to be 90%. By establishing different thresholds for each predicted performance metric, it is possible to selectively identify samples that meet desired properties and confidence levels. Note that the confidence levels from the 10 pre-reserved specimens measure model precision against specific property thresholds, differing from the test set accuracy in training an ML model, which assesses overall predictive effectiveness across diverse data.

5.6. Results and Discussion

5.6.1 Feature Selection

Pearson correlation coefficients (r) (ranging from -1 to 1 to indicate perfect negative and positive correlations respectively) were used for feature selection in this work. To mitigate multicollinearity, only features with an absolute r value less than 0.8 were chosen, resulting in 10 selected features. These include 2 categorical variables: "Design" and "Type", 4 constituent volumetric properties: AC, VTM, G_{se} , and the RAP content, 3 gradation parameters: sieve sizes of 75 μm , 600 μm and 2.36 mm, and 1 mechanical property: $\log |G^*|$. The $\log |G^*|$ was measured at 64°C and 10 radians per second (rad/s) to account for binder viscosity suggested by Kim et al. (2021) and NCHRP Report 871. The phase angle was excluded from the analysis for two key reasons. First, the phase angle δ exhibits a stronger linear relationship with G^* , which could introduce collinearity in the feature selection process for ML models. Second, the objective was to select a single predictive parameter, rather than a combined one such as $\log |G^*|/\sin \delta$. Therefore, the feature selection rules applied in this work present a blend of both machine learning principles and domain knowledge. Figure 5-3, taking CT index as an instance, presents a correlation matrix for these features, with the lower triangle displaying linear regression plots and the upper triangle

showing r values between features. The diagonal plots represent the distribution of each feature or the target property (the top left). Notably, a relatively high positive correlation (0.66) between the CT index and AC, and a negative correlation (-0.73) between the CT index and $\log|G^*|$ were observed. These findings align with expectations that cracking resistance increases with higher AC content, while increased binder stiffness due to aging negatively impacts cracking resistance (Fisher et al., 2018). While both AC and $\log|G^*|$ show relatively higher correlations with CT index compared to other features, it is evident that they do not sufficiently capture the full complexity of CT index predictions. This highlights the necessity of employing machine learning algorithms to uncover the nonlinear interactions among the mixture properties and to enhance prediction accuracy.

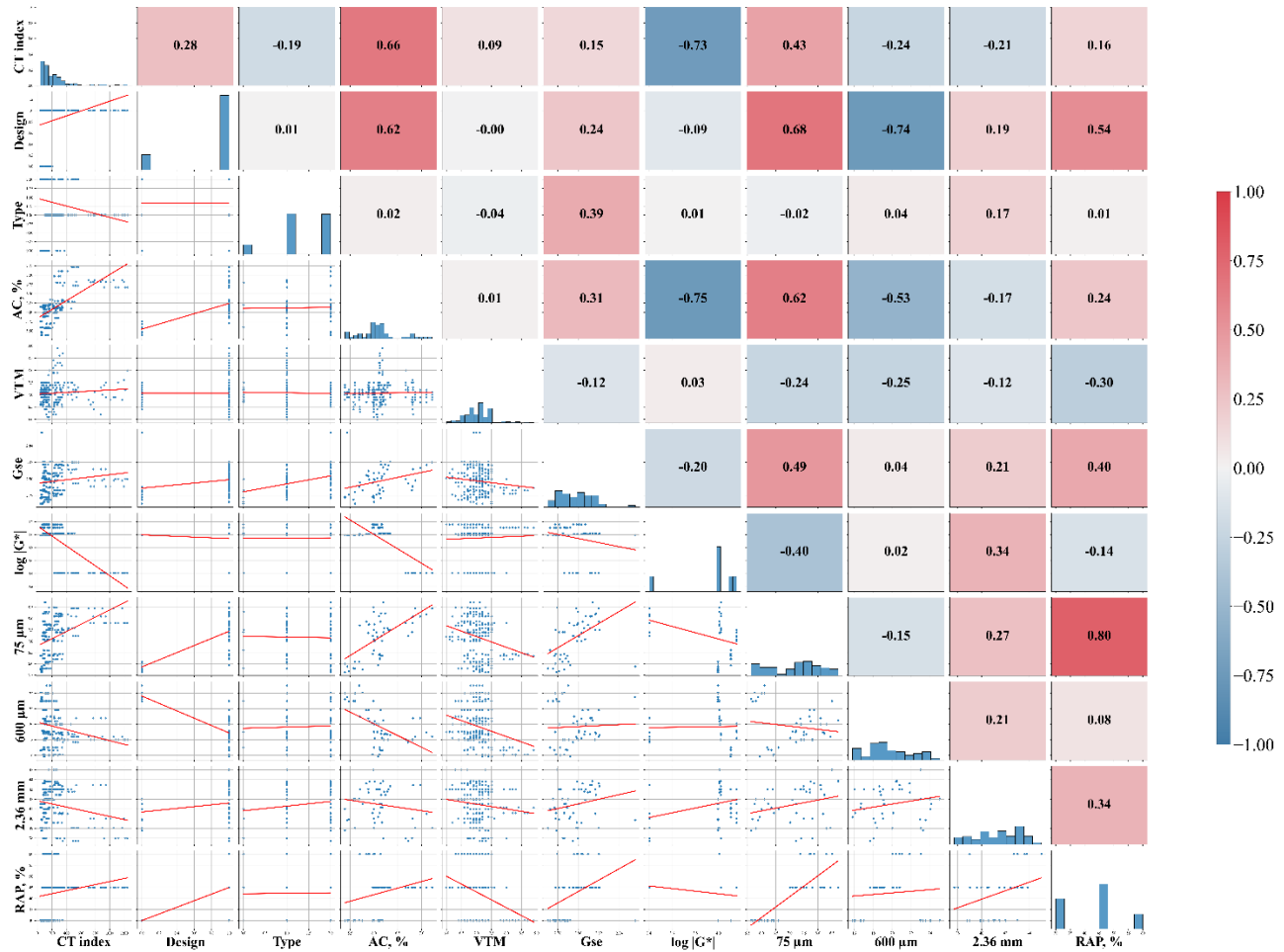


Figure 5-3. Correlation Matrix Derived from Pearson Correlation Coefficients (r) for Selected Features Associated with the IDT-CT Index. The lower triangle displays linear regression plots, while the upper triangle shows r values between features. The diagonal plots illustrate the distribution of each feature or the target property (top left).

5.6.2 Prediction Performance of ML Models

Figure 5-4, Figure 5-5, and Figure 5-6 display the predictive performance of the ML models for the CML, CT index, and APA RD parameters, respectively. After optimization of

hyperparameters, the R^2 (Equation 5-1) and accuracy (Equation 5-2) values are presented for each model. Specifically, Figure 5-4 illustrates the performance for CML predictions with R^2 values of 0.86 (LR), 0.92 (RF), 0.94 (XGB), and 0.91 (SVR) and accuracy values of 0.88, 0.92, 0.92, and 0.90, respectively. Figure 5-5 details the CT index predictions with R^2 values of 0.63 (LR), 0.88 (RF), 0.92 (XGB), and 0.88 (SVR) and accuracy values of 0.47, 0.83, 0.81, and 0.77, respectively. Figure 5-6 shows the APA RD predictions with R^2 values of -0.03 (LR), 0.42 (RF), 0.56 (XGB), and 0.53 (SVR) and accuracy values of 0.69, 0.76, 0.79, and 0.77, respectively. The LR model consistently underperforms compared to the other ML models, which demonstrate stronger predictive abilities. Models like XGB consistently delivered outstanding R^2 and accuracy values for the CML and IDT-CT index, demonstrating their precision in predicting these two BMD performances. For the APA RD predictions, the low R^2 values indicate that the models do not closely follow the trends in the limited dataset (98 data points). However, the relatively high accuracy scores suggest that the models effectively categorize most outcomes correctly.

Given the limited data available for each test, the division of data into training and testing sets can significantly affect the performance of the ML models. To evaluate the consistency of these models, a robustness check was conducted by randomly sampling the dataset using different seeds and training the models with the same optimized hyperparameters previously identified (i.e., from Figure 5-4, Figure 5-5, and Figure 5-6). This procedure was repeated 30 times to assess performance variability. Figure 5-7 illustrates the ML model capability (R^2 and accuracy) for BMD performance metrics prediction, and Table 5-2 summarizes the mean and standard deviation for each BMD performance metric. The LR model consistently underperforms relative to the other models, independently from data sampling. For CML and IDT-CT predictions, the models demonstrate stable prediction capability, as evidenced by the small standard deviation in R^2 (less than 0.2) and accuracy (less than 0.08). However, the APA RD predictions show significant fluctuations in R^2 values due to the limited dataset, though the accuracy fluctuation remains minimal, indicating that the models can reasonably categorize broader outcomes of the APA RD test.

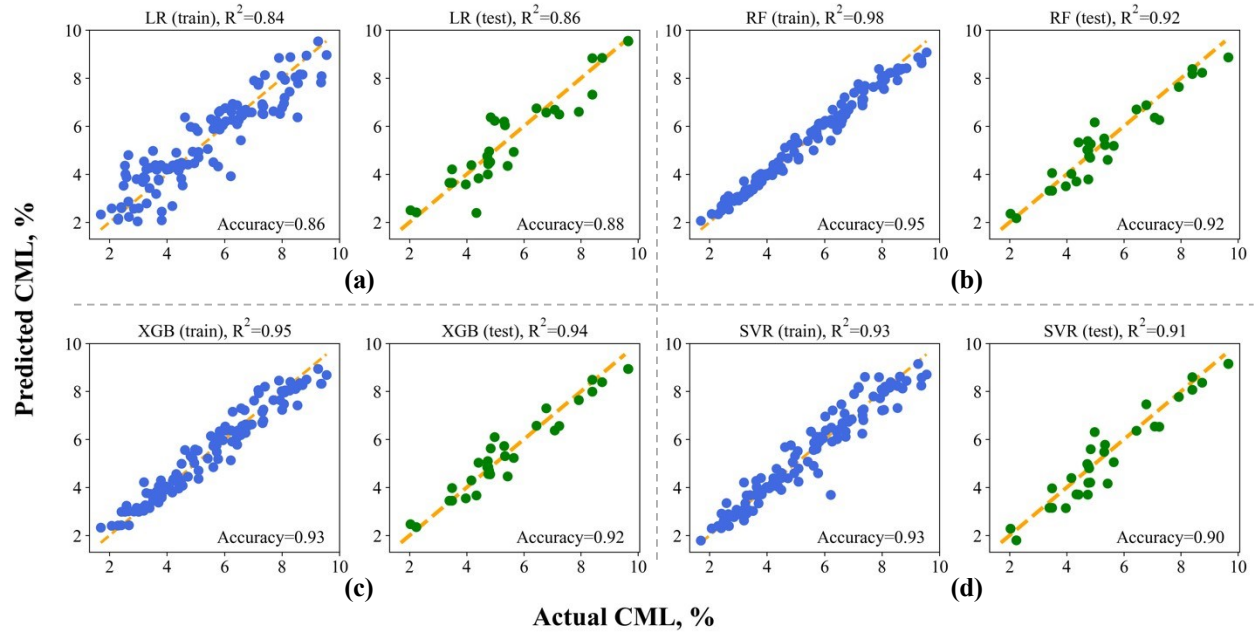


Figure 5-4. Prediction vs. Actual Cantabro mass loss (CML) using (a) LR, (b) RF, (c) XGB, and (d) SVR.

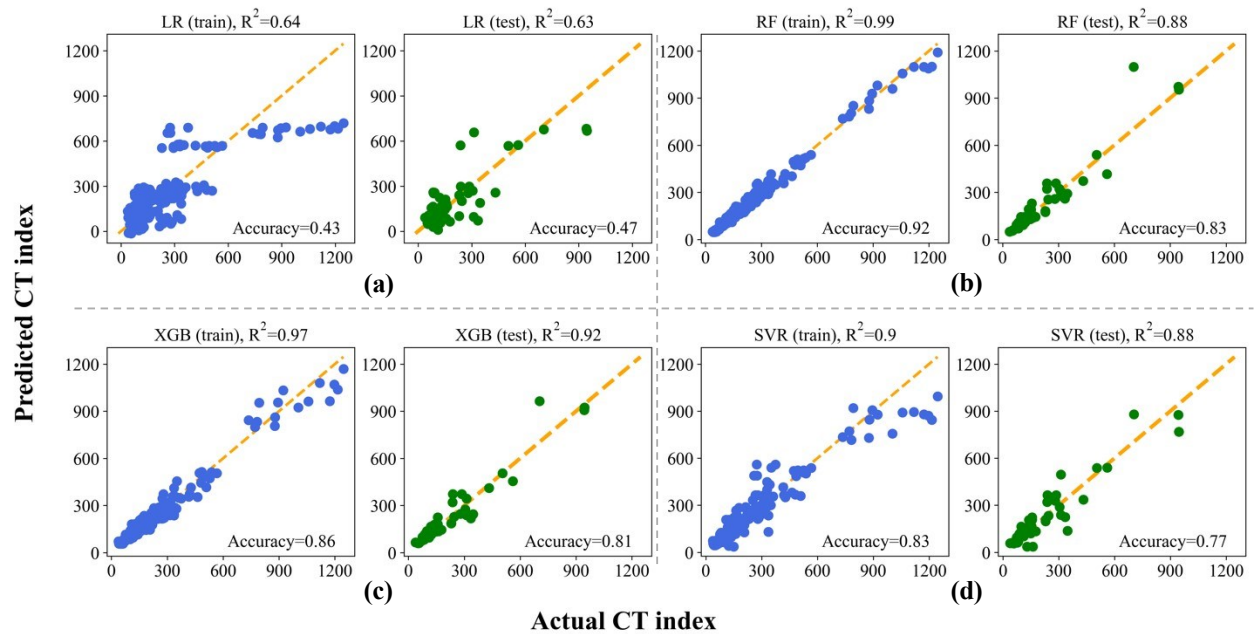


Figure 5-5. Prediction vs. Actual CT index using (a) LR, (b) RF, (c) XGB, and (d) SVR.

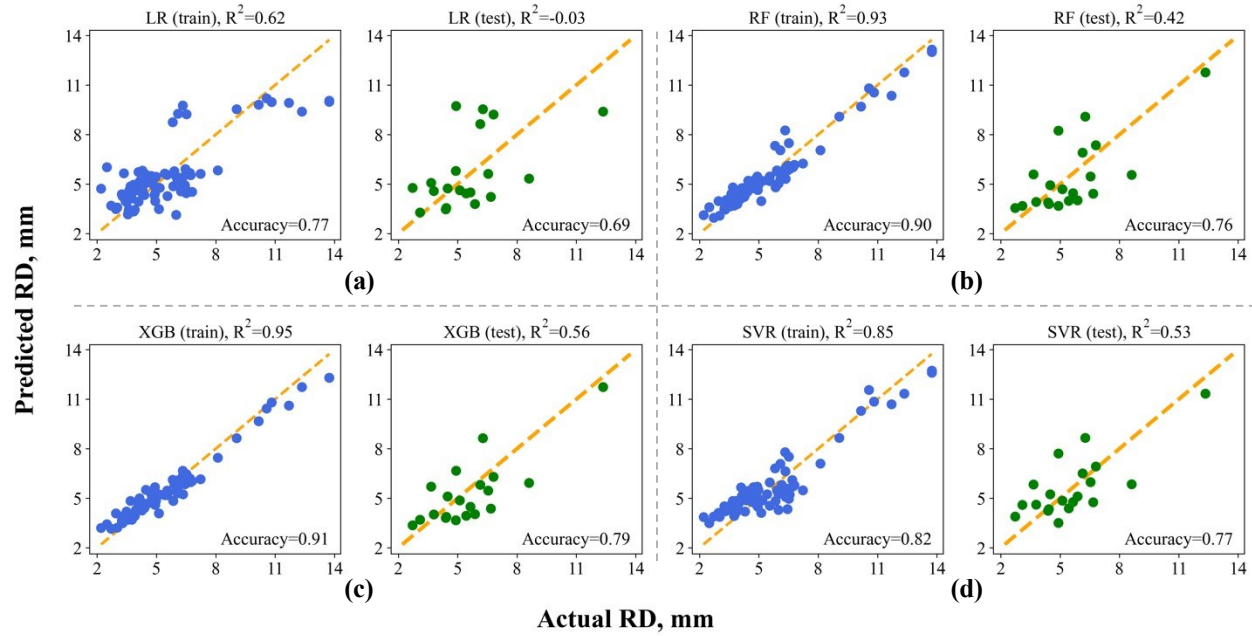


Figure 5-6. Prediction vs. Actual APA rut depth using (a) LR, (b) RF, (c) XGB, and (d) SVR.

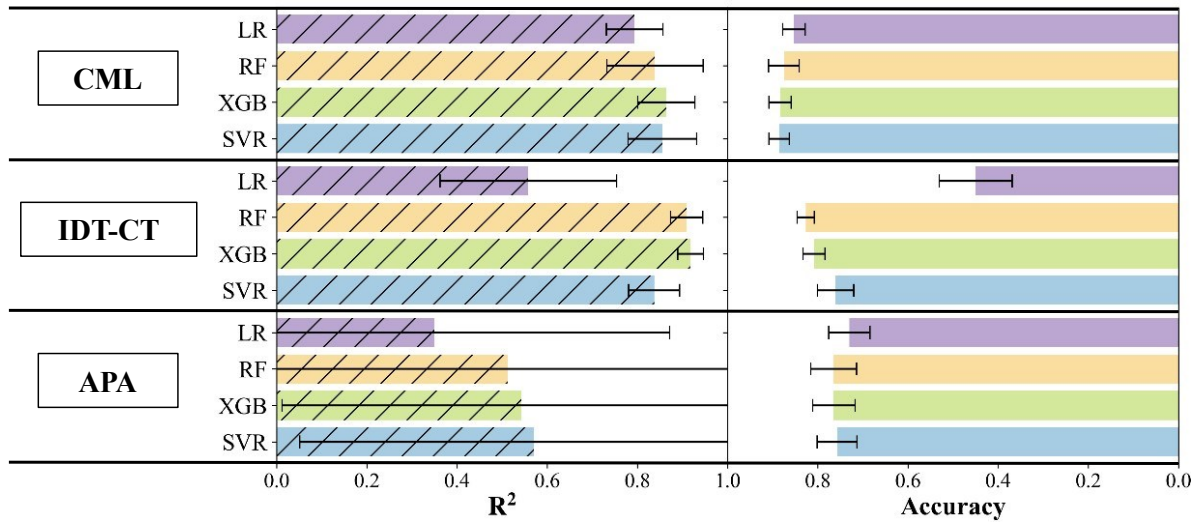


Figure 5-7. Variability in ML model performances (R^2 and accuracy) derived from 30 times random data splitting with different shuffle seeds.

Table 5-2. Mean and standard deviation in R^2 and accuracy from three BMD tests in Figure 5-7.

Model	Mean $R^2 \pm$ Std.			Mean accuracy $\pm$ Std.		
	CML	IDT	APA	CML	IDT	APA
LR	0.79 ± 0.06	0.56 ± 0.20	0.35 ± 0.52	0.85 ± 0.03	0.45 ± 0.08	0.73 ± 0.05
RF	0.84 ± 0.11	0.91 ± 0.04	0.51 ± 0.58	0.88 ± 0.03	0.83 ± 0.02	0.76 ± 0.05
XGB	0.86 ± 0.06	0.92 ± 0.03	0.54 ± 0.53	0.88 ± 0.02	0.81 ± 0.02	0.76 ± 0.05
SVR	0.86 ± 0.08	0.84 ± 0.06	0.57 ± 0.52	0.89 ± 0.02	0.76 ± 0.04	0.76 ± 0.04

CML = Cantabro mass loss; IDT-CT = indirect tensile cracking test; APA = asphalt pavement analyzer; RD = rut depth.

5.6.3 Feature Importance Analysis

Understanding the key factors influencing BMD tests allows engineers not only to focus on critical variables during the design phase but also to make proactive adjustments during the production phase. For this analysis, the models that demonstrated the highest accuracy across 30 instances of data sampling (Figure 5-7) were selected: XGB for CML (accuracy=0.93), RF for IDT-CT index (accuracy = 0.87), and RF for APA RD (accuracy = 0.83).

As shown in Figure 5-8, the feature importance analysis reveals distinct dominant factors influencing each BMD performance. For CML (Figure 5-8a), the key features include Type, 75 μm sieve size, AC and VTM. For IDT-CT index (Figure 5-8b), the primary features are AC, Type, 75 μm sieve size. For APA (Figure 5-8c), the dominant features are AC, $\log|G^*|$ and VTM. Notably, the Type (type of specimens: design, non-reheats, and reheats) emerges as a significant feature for predicting both CML and IDT-CT index, suggesting the durability and cracking resistance may significantly vary by specimen type and be impacted by reheating effect. This observation is consistent with our previous studies emphasizing the impact of reheating on BMD performances (Tong et al., 2024b). Furthermore, models on both the IDT-CT index and APA identified AC as a dominant factor, similarly highlighted in the previous study (Mogawer et al., 2023). The finding suggests that stricter control of asphalt binder content may be used as a proactive adjustment to reduce performance variability in mixture performances during plant production. The $\log|G^*|$ parameter, highlighted in (Kim et al., 2023) for reflecting binder viscosity—including contributions from virgin binder, RAP binder, and effects of aging—was also recognized by the ML models as a significant feature for the CT index and APA RD. This recognition further confirms the model's explanatory capability. Although it has been found that most gradation parameters do not significantly affect the performance of produced mixtures due to strict gradation controls at the plant (Tong et al., 2024b), Austerman et al. (2018) have suggested that greater attention should be paid to the percentage passing through the 75 μm sieve size during the production phase. Notably, 75 μm sieve size was identified as non-negligible in CML and CT index predictions, whereas the 600 μm and 2.36 mm sieve size showed limited impact on the model. These findings underscore the exceptional capability of the ML models to identify critical factors that align closely with prior observations and studies. Moreover, the VTM feature displays notable influences on both CML and APA RD predictions. While the impact of VTM deviation on CML tests has been reported in (Tong et al., 2024b), its effect on APA RD tests may require more data for validation due to the unstable model prediction caused by limited APA data.

Design methods (i.e., Superpave or BMD) and RAP content are commonly emphasized. However, feature importance analysis reveals that neither the design method nor RAP content significantly impacts prediction performance. This is likely due to the use of BMD with a soft binder, with or without a recycling agent, which effectively compensates for variations in RAP content within the mixture (Tong et al., 2024a, Yin and West, 2021). Furthermore, the selection of other constituent properties in this study may effectively captures the differences introduced by

BMD and Superpave design methods, clarifying the potential performance variations between HRAP mixtures and conventional RAP mixtures.

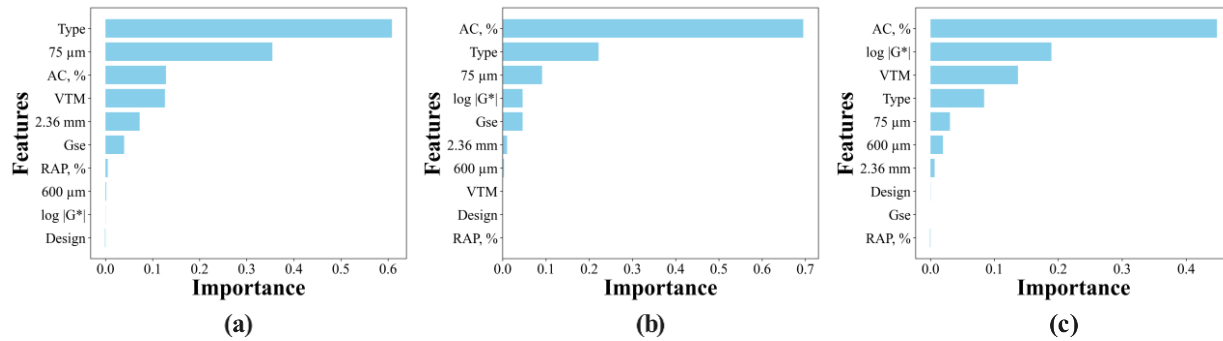


Figure 5-8. Feature importance from the permutation importance method for: (a) CML from XGB, (b) IDT-CT index from RF, and (c) APA RD from RF model.

5.6.4 Pseudo In-situ Deployment

With the ML models demonstrating satisfactory predictive performance on BMD test properties, pseudo in-situ deployment was initiated to generate samples exhibiting balanced three primary BMD performances (i.e. CML < 7.5%, CT index > 70, APA < 8 mm). Figure 5-9 illustrates the schematic of this deployment process. Initially, 1000 synthetic samples were generated based on the mean and standard deviation of each feature from the original data pool, consisting of 618 data points excluding the 30 pre-reserved for testing. These samples were first subjected to filter-I, using the XGB model for CML prediction, to retain samples with CML values below 7.5%. Meanwhile, the CML values of 10 pre-reserved data were predicted to assess the confidence level of filter-I. This approach was similarly applied in filter-II and filter-III for IDT-CT and APA models, with thresholds set at IDT-CT > 70 and APA RD < 8 mm, respectively. Samples passing all three filters were deemed balanced mixtures. In this evaluation, 812 samples passed filter-I, of which 811 proceeded past filter-II, and 771 passed filter-III. The confidence level for all filters reached 100%, indicating accurate predictions for all pre-reserved data.

It is commonly understood that simultaneously enhancing multiple primary BMD performances in mixture design is challenging. However, unlike earlier versions, contemporary asphalt materials are characterized by the widespread use of the extensive application of RAP and additives. This evolution complicates their performance and may challenge conventional wisdom (*Error! Reference source not found.*). The utilization of ML models holds great potential to provide an enhanced solution for the three primary performance metrics. In this case, filter-IV applied stricter thresholds (CML < 5.8%, CT index > 200, and APA RD < 6 mm) to identify samples with enhanced mixture performances, narrowing the field to 60 candidates from the 771 remaining after filter-III. The stricter thresholds are based on the relationship between non-reheated and reheated specimens determined by previous study (Tong et al., 2024b). Additionally, practical requirements for VTM content (7.0 ± 0.5) in IDT and APA specimens led to the implementation of filter-V. After all five filters, only 19 out of the initial 1000 samples met the stringent criteria. This pseudo in-situ deployment demonstrates the effectiveness of using ML

models to address the challenges of enhancing mixture properties, allowing for flexible adjustment of criteria in each filter to efficiently tailor the mixture design for specific practical applications.

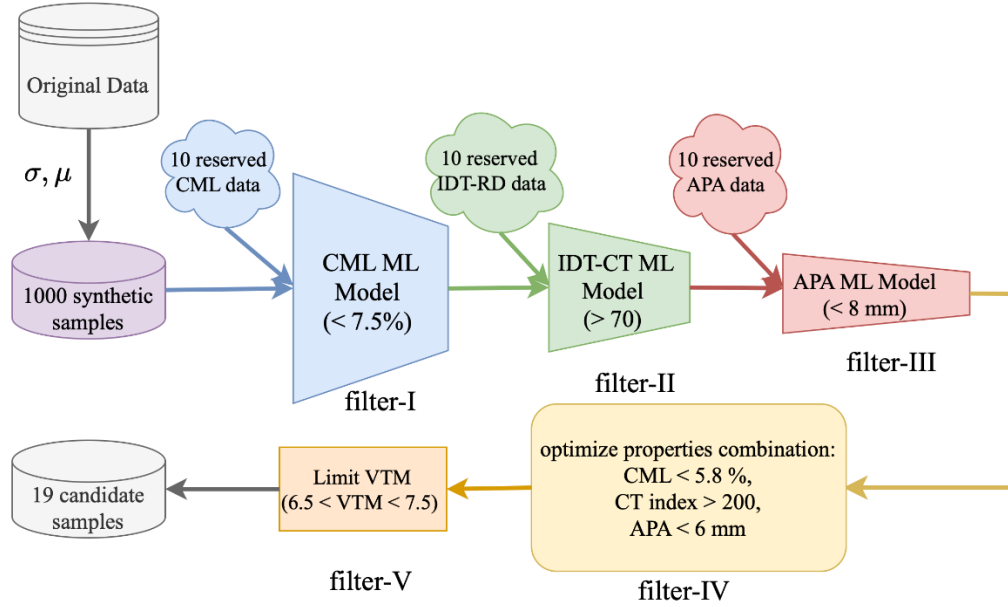


Figure 5-9. The schematic of pseudo in-situ deployment flowchart.

Experimentally designing a mixture that simultaneously enhances key properties—such as reducing CML and APA RD values while increasing the CT index—presents significant challenges due to the extensive feature space (10 features in this work). To understand this complexity, uniform manifold approximation and projection (UMAP) (McInnes et al., 2018) was employed to reduce the feature space into two dimensions for visualization and to compare the distribution of qualified samples from filter-IV (60 data points) against the original datasets. For each performance metric, such as CML, data were divided into two groups: those with CML values below 5.8 and those with higher values. This categorization was similarly applied to the CT and APA datasets. Note that the samples from filter-IV, which have all three primary performances enhanced, were plotted alongside the original datasets, where each dataset represents only one BMD performance. Figure 5-10 displays these distributions, with qualified samples marked by purple stars and the original dataset's qualified data points represented by red dots in Figure 5-10a, Figure 5-10b, and Figure 5-10c for CML, CT index, and APA RD test, respectively. The visualization revealed that samples meeting all enhanced criteria rarely overlapped with the distributions of the original datasets. This highlights the challenges engineers face in achieving all desired enhancements experimentally, as the target distribution with all enhanced performances is scarce within the broad feature space, making exhaustive searches in experiments typically impractical.

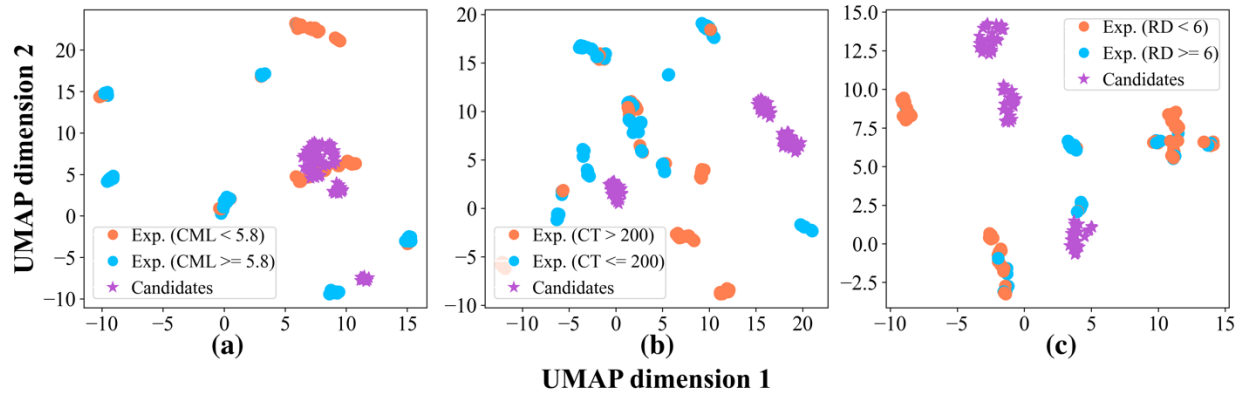


Figure 5-10. Uniform Manifold Approximation and Projection (UMAP) reduces the feature space to two dimensions for (a) CML, (b) CT index, and (c) APA RD datasets. The number of neighbors used in UMAP is 10, and it is asserted that the choice of hyperparameters has a negligible impact on the conclusions drawn.

5.6.5. Discussion

This section summarizes the potential applications of ML models within the BMD framework during the design and production phases, thereby showcasing the enhancements brought by ML models. Figure 5-11 illustrates the current BMD framework, using BMD approach D as an example. A pivotal loop in the BMD design phase involves the selection of feature combinations based on required performance tests. Each loop is not only subject to variability from specimen fabrication, testing, and device, but may also become cumbersome due to the need to restart the selection process if the results of a performance test fail to meet the specified BMD thresholds. With the assistance of developed ML models, this cumbersome design loop can potentially be bypassed, significantly enhancing design efficiency. Additionally, concerns about unbalanced cases resulting from mixture production can be addressed through project-level ML models, which identify dominant factors contributing to production variance. It is suggested that more stringent control of the constituent properties could reduce performance variability in produced mixtures. Furthermore, if ML models can be trained with more data in a highly controlled environment and further validation of models' predictions can be achieved, it is highly promising that the QA program in BMD implementation will soon significantly reduce the reliance on sampling mixtures for BMD performance tests during production, opting instead to predict performance based solely on component properties. Finally, this study establishes a framework for the application of ML in BMD and verifies its feasibility at the project level. State DOT agencies can reference the data collection methods presented in this study to build a comprehensive database, facilitating the development of reliable ML models that incorporate a wider variety of material sources, such as binder types, aggregate resources, and additives.

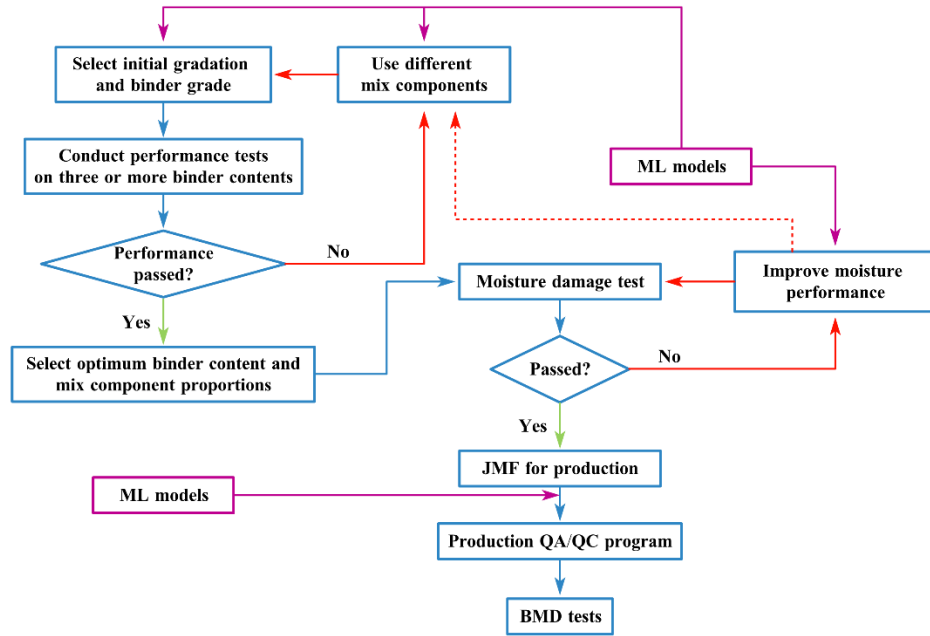


Figure 5-11. Explore the Use of ML Models to Support the Implementation of BMD.

5.7. Conclusion

This study leverages ML algorithms to help address some of the challenges prevalent in the existing BMD framework, thereby enhancing its implementation during both the design and production phases. Four prominent ML algorithms were employed to predict three primary performance metrics of BMD mixtures, namely, CML, CT index and APA RD. Hyperparameters for each ML model were optimized using an exhaustive grid search method with sufficient parameter spaces. Among the models, RF, XGB, and SVR exhibited superior performance compared to LR, effectively capturing the complex interactions within the mixture properties. The robustness of the ML models was evaluated and validated through 30 independent training/testing dataset splits in a random manner. ML models reliably delivered accurate predictions across all BMD tests, corroborated by the identification of significant influencing factors such as Type of specimens, AC, $\log|G^*|$, 75 μm sieve size passing percentage, and VTM. The impact of the reheating process on mixtures' durability and cracking performance was also identified, supported by previous studies. The study also recommended stricter control of identified features to minimize performance variability in mixtures. However, the models faced challenges in achieving stable R^2 values in the APA RD test, likely due to limited data availability and the use of average values from APA specimens.

To demonstrate the practical application of the developed ML models in BMD mixture design, a pseudo in-situ deployment (Figure 5-9) was strategically implemented to filter and select 60 out of 1,000 synthetic samples with optimal BMD performances through simultaneous optimization of three BMD performance metrics. By simply adjusting the filters in the pseudo in-situ deployment, one can effortlessly fine-tune the desired BMD performances. Dimensionality reduction via the UMAP method visualized minimal overlap in the distributions of these 60

candidates compared to the original datasets, underscoring the complexities of managing large feature spaces and improving multiple BMD performances simultaneously in experimental settings. Current study establishes a framework for implementing ML in BMD applications and demonstrates its feasibility at the project level. State DOT agencies can utilize the data collection methodologies outlined in this research to construct a comprehensive database. This database will support the development of robust ML models, incorporating a broader range of material sources such as binder types, aggregate resources, and additives.

5.8. Reference

- Golalipour, A., Veginati, V., & Mensching, D. J. (2021). Evaluation of Asphalt Mixture Performance Using Cracking and Durability Tests at a Full-Scale Pavement Facility. *Transportation Research Record*, 2675(11), 226-236.
- National Center for Asphalt Technology and National Asphalt Pavement Association. (2023). Online resource: Balanced Mix Design Resource Guide. <https://www.asphaltpavement.org/expertise/engineering/resources/bmd-resource-guide>. Accessed January 24, 2023.
- American Association of State Highway and Transportation Officials (AASHTO). (2021). Standard Specifications for Transportation Materials and Methods of Sampling and Testing. AASHTO, Washington, D.C., 2021.
- Tong, B., Habbouche, J., Diefenderfer, S. D., & Flintsch, G. W. (2024a). Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia. *International Journal of Pavement Engineering*, 25(1), 2325609.
- West, R., Rodezno, C., Leiva, F., & Yin, F. (2018). Development of a Framework for Balanced Mix Design. *Project NCHRP*, 20-07.
- Yin, F., & West, R. C. (2021). Balanced Mix Design Resource Guide (No. IS-143).
- Mogawer, W. S., Austerman, A. J., Stuart, K. D., Abdalfattah, I. A., & Zhou, F. (2023). Comprehensive Balanced Mixture Design Protocol: Design Phase through Production Quality Assurance. *Transportation Research Record*, 03611981231186422.
- Bowers, B. F., Diefenderfer, S. D., Moore, N., & Lynn, T. (2022). Balanced Mix Design and Benchmarking: A Case Study in Establishing Performance Test Thresholds. *Transportation Research Record*, 2676(12), 586-598.
- Habbouche, J., Boz, I., & Diefenderfer, S. D. (2022). Validation of Performance-Based Specifications for Surface Asphalt Mixtures in Virginia. *Transportation Research Record*, 2676(5), 277-296.
- Lynn Wright, T. Y., Bowers, B. F., Diefenderfer, S. D., & Boz, I. (2024). Evaluation of the Influence of Production Tolerance Limits on Virginia Balanced Mixtures. *Transportation Research Record*, 0361198123122248.
- Brown, E. R., Kandhal, P. S., Roberts, F. L., Kim, Y. R., Lee, D. Y., & Kennedy, T. W. (2009). Hot Mix Asphalt Materials, Mixture Design, And Construction. NAPA Research and Education Foundation.

- Mohammad, L. N., Elseifi, M. A., Cooper III, S. B., Hughes, C. S., Button, J. W., & Dukatz Jr, E. L. (2016). Comparing the Volumetric and Mechanical Properties of Laboratory and Field Specimens of Asphalt Concrete (No. Project 09-48). Transportation Research Board.
- Diefenderfer, S. D., Habbouche, J., & Boz, I. (2023). Balanced Mix Design for Surface Mixtures: 2020 Field Trials (No. FHWA/VTRC 23-R13). Virginia Transportation Research Council, Charlottesville.
- Tong, B., Habbouche, J., Flintsch, G. W., & Diefenderfer, B. K. (2023). Rutting Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents under Full-Scale Accelerated Testing. *Materials*, 16(24), 7611.
- Tong, B., Habbouche, J., Diefenderfer, S. D., & Flintsch, G.W. (2024b). Assessment of Production Variability and Composite Performance Index for Conventional and High RAP BMD Mixtures, *Transportation Research Record*, 2024 (in-production).
- Buttlar, W. G., Urrea-Contreras, L., Jahangiri, B., Rath, P., & Majidifard, H. (2020). Support for Balanced Asphalt Mixture Design Specification Development in Missouri. *Missouri Department of Transportation*, 4(September), 103.
- Habbouche, J., Boz, I., & Diefenderfer, S. D. (2022). Interlaboratory Study for the Indirect Tensile Cracking Test at Intermediate Temperature: Phase II (No. FHWA/VTRC 23-R3). Virginia Transportation Research Council, Charlottesville.
- VDOT. (2020). Road and Bridge Specification. Virginia Department of Transportation.
- Austerman, A. J., Mogawer, W. S., & Stuart, K. D. (2018). Influence of Production Considerations on Balanced Mixture Designs. *Transportation Research Record*, 2672(28), 426-437.
- Bowers, B. F., Lynn, T., Yin, F., Moore, N., Diefenderfer, S. D., & Boz, I. (2023). Impact of Production Variability on Balanced Mix Designs in Virginia (No. FHWA/VTRC 23-R20). Virginia Transportation Research Council, Charlottesville
- Rahman, S., Bhasin, A., & Smit, A. (2021). Exploring the Use of Machine Learning to Predict Metrics Related to Asphalt Mixture Performance. *Construction and Building Materials*, 295, 123585.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.
- Chen, T., & Guestrin, C. (2016). Xgboost: A Scalable Tree Boosting System. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Vapnik, V., Golowich, S., & Smola, A. (1996). Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. *Advances in neural information processing systems*, 9.
- Smola, A. J., & Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and computing*, 14, 199-222.
- Zeiyada, W., Dabous, S. A., Hamad, K., Al-Ruzouq, R., & Khalil, M. A. (2020). Machine Learning for Pavement Performance Modelling in Warm Climate Regions. *Arabian*

- Journal for Science and Engineering*, 45(5), 4091-4109.
- Daneshvar, D., & Behnood, A. (2022). Estimation of the Dynamic Modulus of Asphalt Concretes Using Random Forests Algorithm. *International Journal of Pavement Engineering*, 23(2), 250-260.
- Zavrtanik, N., Prosen, J., Tušar, M., & Turk, G. (2016). The Use of Artificial Neural Networks for Modeling Air Void Content in Aggregate Mixture. *Automation in Construction*, 63, 155-161.
- Liu, J., Liu, F., Zheng, C., Zhou, D., & Wang, L. (2022). Optimizing Asphalt Mix Design Through Predicting the Rut Depth of Asphalt Pavement Using Machine Learning. *Construction and Building Materials*, 356, 129211.
- Ahmed, T. M., Green, P. L., & Khalid, H. A. (2017). Predicting Fatigue Performance of Hot Mix Asphalt Using Artificial Neural Networks. *Road Materials and Pavement Design*, 18(sup2), 141-154.
- Zhou, Q., Okte, E., & Al-Qadi, I. L. (2021). Predicting Pavement Roughness Using Deep Learning Algorithms. *Transportation Research Record*, 2675(11), 1062-1072.
- Nguyen, L. N., Le, T. H., Nguyen, L. Q., & Tran, V. Q. (2023). Machine Learning Approaches for Predicting Cracking Tolerance Index (Ctindex) of Asphalt Concrete Containing Reclaimed Asphalt Pavement. *Plos One*, 18(10), e0287255.
- American Association of State Highway and Transportation Officials. (2017). Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards. Washington, DC.
- ASTM International. (2019). *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.
- Kim, Y. R., Castorena, C., Saleh, N. F., Braswell, E., Elwardany, M., & Rad, F. Y. (2021). Long-Term Aging of Asphalt Mixtures for Performance Testing and Prediction: Phase III results (No. Project 09-54).
- Fisher, A., Rudin, C., & Dominici, F. (2019). All Models Are Wrong, But Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, 20(177), 1-81.
- McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*.

CHAPTER 6 - Rutting Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents Under Full-Scale Accelerated Testing⁶

6.1. Abstract

Rutting is one of the major distresses that adversely affects the performance and service life of flexible pavement structures. Understanding the extent to which the rutting resistance of pavement mixtures can be compromised and finding methods to optimize their performance requires thorough investigation. BMD constitutes a significant step forward in the pursuit of better-performing asphalt mixtures. This framework / approach provides an opportunity to properly design and produce engineered asphalt mixtures. The primary objective of this paper is to conduct a comprehensive investigation of the permanent deformation (rutting) behavior of surface mixtures with conventional and HRAP contents through full-scale accelerated testing under incremental loading conditions, while accounting for the environmental aging effect. Results showed that all BMD HRAP mixtures exhibited higher rut depths compared to the control mixture, which can be attributed to the inclusion of a high binder content aimed at enhancing cracking resistance. The APA rut test and the stress sweep rutting tests were performed on asphalt mixtures sampled during production. Correlation analysis revealed significant and strong positive correlations between APT testing and the multilevel laboratory rutting performance tests considered in this study. Finally, while acknowledging the limitations and all the assumptions considered in this study, the correlation analysis recommended refining the BMD APA rut depth threshold by lowering the current limit of 8 mm to 7 mm to ensure good performing mixtures from rutting point of view.

Keywords: high RAP, asphalt concrete, balanced mix design, rutting performance, aging, accelerated pavement testing, heavy vehicle simulator, correlations, specifications, thresholds.

⁶ This paper has been published in the *Materials*. Citation: Tong, B., Habbouche, J., Flintsch, G. W., & Diefenderfer, B. K. (2023). Rutting performance evaluation of BMD surface mixtures with conventional and high RAP contents under full-scale accelerated testing. *Materials*, 16(24), 7611. Contribution: Conceptualization: Lead; Methodology: Lead; Resources: Equal; Data Processing and Analysis: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead.

6.2. Introduction

Distresses in flexible pavements refer to various types of damage or deterioration that can occur over time due to traffic loads, environmental factors, and material deficiencies. These distresses can significantly impact the structural integrity and serviceability of the road surface. Rutting stands as a significant distress that adversely affects the durability and service life of flexible pavement structures. This form of deterioration not only diminishes user comfort and compromises safety but also leads to escalated operational expenses for highway maintenance and rehabilitation. Understanding the extent to which the rutting resistance of pavement mixtures can be optimized requires thorough investigation. Given agencies' concerns about the long-term performance of pavements, research on rutting investigation has become an essential and indispensable area of study. Addressing this issue will contribute to the development of more robust and longer-lasting flexible pavements, benefiting both road users and maintenance authorities.

6.2.1. Rutting Performance Evaluation

Laboratory performance tests play a crucial role in the BMD process as they ensure the production of high-performing materials. Over the past few decades, several testing methods and analysis protocols have been proposed and adopted in the laboratory to characterize the rutting resistance of asphalt mixtures. These performance tests include the APA (Choubane et al. 2000), Hamburg wheel tracking (Yin et al., 2020), French simulator (Mota et al., 2021), and uniaxial repeated load test to determine the Flow Number (Faccin et al., 2021). However, these procedures typically employ pass-fail criteria based on fixed testing conditions, which limits their ability to capture the complex nature of permanent deformation under varying field conditions. For instance, as part of VDOT BMD specifications, the APA rut test is conducted at 64°C, with a threshold of a maximum rut depth of 8.0 mm after 8000 loading cycles. To address this limitation and achieve a more comprehensive characterization of resistance to permanent deformation, two testing protocols have emerged in recent years: the TSS and the SSR tests. The TSS test offers advantages over existing models by considering the influence of temperature, loading time, and deviatoric stress on permanent deformation behavior (Kim and Kim, 2017). Additionally, the SSR test has been introduced to further simplify the TSS test, aiming to reduce its complexity and cost while maintaining its effectiveness in evaluating the resistance to permanent deformation (Kim and Kim, 2017). Furthermore, Meroni et al. proposed a classification of rutting performance tests into three levels - basic, intermediate, and advanced - based on various factors including complexity, cost, time, and training requirements (Meroni et al., 2021). This classification offers flexibility in selecting the appropriate level of testing depending on the specific objectives and constraints of a given research or engineering project.

6.2.2. Background on Virginia APT

In the U.S., Virginia is among the states that have implemented the use of APT facilities for pavement evaluation purposes. In 2015, a collaborative experiment between the VTRC, VDOT, and Virginia Tech was initiated and executed at the Virginia APT facility located at the VTTI. The experiment encompassed three major studies, including the investigation of the impact of different overlays on CCPR material (Xue et al., 2017; Xue et al., 2020; Flintsch et al., 2020), the assessment of field performance of two distinct dense-graded SMs designed with varying design gyrations levels (50 and 65) (Meroni et al., 2020), and the examination of reflection cracking. In 2020, another effort was planned and initiated to assess the application of the BMD concept in designing durable and longer-lasting mixtures in Virginia with a particular emphasis on mixtures incorporating relatively higher RAP contents. The overarching scope of the effort included the development of the APT experiment, verification of mix design performance, documentation of paving operations and construction practices, and mixture sampling during production. Moreover, it included coring of as-placed material, testing and analysis of the volumetric and performance properties of mixtures, actual testing using a HVS for rutting and cracking studies, analyses of collected pavement responses and generated data, and documentation of observations and lessons learned. A total of six experimental-testing lanes were constructed. These lanes featured the use of typical and high RAP contents, RA, softer binder, and WMA additive.

6.3. Objective and Scope

The primary objective of this paper is to conduct a comprehensive investigation of the permanent deformation (rutting) behavior of SMs with conventional and high RAP contents through full-scale accelerated testing under incremental loading conditions. This paper specifically covers the design of pavement test cells, the traffic test plan, and the measurement of rutting performance. In addition, this paper proposes an empirical rutting model to account for the environmental aging effect observed in the APT measurements. Furthermore, correlation analysis is conducted between APT measured rut depth and two laboratory rutting performance tests. Finally, an attempt to validate and refine the current VDOT BMD threshold was performed.

6.4. APT Setup for 2020 BMD Experiment

Conventional laboratory test protocols primarily aim to simulate the stress and strain conditions experienced under typical traffic scenarios. However, it is important to acknowledge that the stress or strain conditions imposed on specimens in laboratory settings may differ from those encountered in the field. To bridge this gap between laboratory simulations and/or ME design models developed using laboratory material testing characterization and actual pavement stress conditions, APT tests can be conducted using representative truck wheel loads and tire pressures. HVS tests provide a valuable approach to overcome the limitations inherent in conventional laboratory tests and offer a more precise representation of the stress conditions experienced by pavements in actual field conditions.

The central component of the APT BMD experiment was the Dynatest HVS system, specifically the Dynatest Mark VI model shown in Figure 6-1a. This system featured an environmental chamber designed to maintain a consistent temperature at the loaded area. The HVS system automatically regulates the surface temperature to ensure that a depth of 2 inches from the surface is maintained at the temperature of interest. Inside the environmental chamber, a dual tire carriage was mounted, capable of achieving a speed up to 12.4 mph with a tolerance of ± 2 mph. The loading was applied using a dual tire assembly, utilizing 11.00R22.5 tires inflated to a pressure of 105 psi. Additionally, the system allowed for the completion of 12,000 uni-directional passes or 24,000 bi-directional passes within a 24-hour period. The dual tire carriage was capable of accommodating loads ranging from 6,750 lb to 22,500 lb. A laser profiler, installed on the HVS carriage and shown in Figure 6-1b, was used to conduct scans of the pavement surface and quantify the vertical permanent deformation present at the surface. A DAS was employed to capture signals from various instruments and monitor the conditions of the pavement. To facilitate this data acquisition process, input modules from National Instruments were installed in a chassis along with a controller. The LabVIEW software was utilized to develop an interactive interface for data management and analysis. The DAS system, along with its associated components, was housed in a weather-proof chamber located near the HVS apparatus.

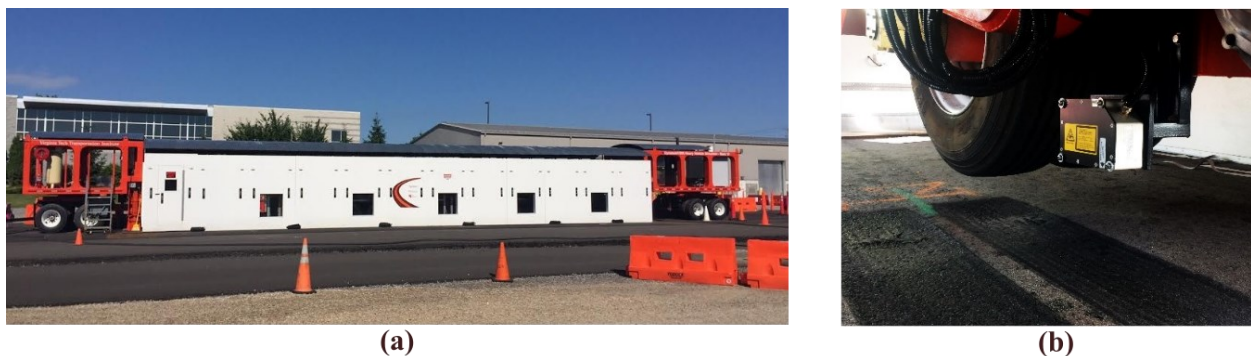


Figure 6-1. Photographs taken of: (a) Dynatest Mark VI Heavy Vehicle Simulator; (b) Laser Profiler Mounted on HVS Carriage.

6.5. Experimental Program

6.5.1. Materials Evaluated

A total of one control and five BMD 9.5mm dense-graded SMs were investigated in this study. The mixtures incorporated various combinations of RAP contents, two binders (PG 64S-22 and PG 58-28), one RA, and one WMA additive. The BMD optimized mixtures were designed following VDOT BMD special provisions using Approach D (Performance Only). The volumetric properties and gradations of all mixtures collected during design with various performance testing details can be in Table 3-1. Those collected for each sample / subplot during production can be found elsewhere (Tong et al., 2023).

6.5.2. Design of APT Test Cells

Six APT lanes were constructed, each lane was designated to evaluate one of the six mixtures under full-scale accelerated loading. Each lane had a length of 300 ft and a width of 10 ft. Within each lane, five test cells (labeled A through E) were established. Test cells A, C, and E were specifically designated as the test cells for comprehensive data collection. These test cells were instrumented with a variety of sensors, including strain gauges, pressure cells, moisture sensors, and thermocouples to collect pavement response data. Test cells B and D served as backup test cells to address unforeseen issues or unexpected damage during the experiment. The test cells were strategically positioned within each lane, spanning the center 200 ft of the 300 ft lane; each test cell had a length of 24 ft. Space between each test cell provided room for the HVS support structure. Figure 6-2a provides a visual representation of the site layout for the 2020 APT BMD experiment. The pavement structure, shown in Figure 6-2b, was consistent across all six constructed lanes. Each lane was built with a research mixture consisting of two 1.5-inch lifts, with a total thickness of 3 inches. Each research mixture was placed on top of a 12-inch VDOT 21B base layer, followed by an additional 26-inch layer of VDOT 21B as a subgrade layer placed over a reinforced foundation.

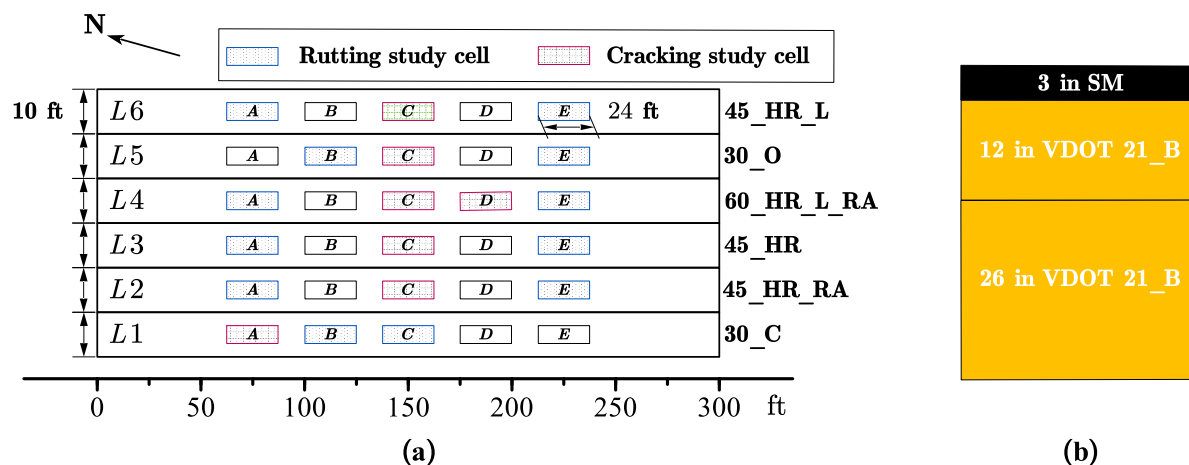


Figure 6-2. (a) Site Layout for the 2020 APT project; (b) Pavement Structure.

6.5.3. Loading Protocols

Two studies were conducted per lane, one for rutting evaluation and another for cracking evaluation. Specific cells were assigned for each experiment. The key distinguishing factors between the cracking and rutting studies were the variation in testing temperatures, loading protocols, and wandering settings. For this paper, the focus is the investigation of rutting performance within the 12 rutting test cells, two cells in each of the six lanes.

The loading protocol implemented within each test cell followed a specific pattern. Initially, a wheel load of 9,000 lbf was maintained for the first two weeks. Subsequently, the load was increased to 12,000 lbf for approximately one week, followed by a further increase to 15,000 lbf for approximately another week. This progressive loading protocol was designed to expedite

the development of pavement distresses. Wandering was set to a maximum of 2 inches on either side of the centerline. The internal pavement temperature was maintained at 40°C (104°F) measured by a thermocouple embedded at a depth of 5.1 cm (2 inch) beneath the pavement surface. The 9,000 lbf load level was selected to simulate half of an 18,000 lbf standard axle load. Throughout the test period, the HVS operated continuously, except for regular daily maintenance and occasional repairs. Additionally, temporary staff shortages during the COVID-19 period may have impacted the testing schedule. The loading and repetitions were converted into ESALs, with the target of achieving approximately 500,000 ESALs in each rutting test cell. The number of ESALs were calculated using Equation 6-1. The 18-kip single-axle equivalency concept was introduced based on the AASHO road accelerated testing completed in 1961 (Kawa et al., 1998). One pass at the 9,000 lbf, 12,000 lbf, and 15,000 lbf load levels corresponded to 1.00, 3.35, and 8.55 ESALs, respectively.

$$\text{ESALs} = \left(\frac{\text{wheel load in lb}}{9000} \right)^{4.2} \quad \text{Equation 6-1}$$

6.5.4. Experiment Timeline

The APT BMD experiment was conducted over a period of approximately three years. Detailed information regarding the loading timeline and applied loading conditions for the pavement during the testing period can be found in Table 6-1. Summary of Rutting Testing Cells and Timeline Within the APT BMD Experiment. The notation R1 is used to indicate rutting cell_1, which underwent testing prior to rutting cell_2 denoted by R2. This differentiation is made to facilitate the subsequent explanation of potential environmental aging effects. The loading protocols employed for the test cells were summarized in Figure 6-3. The slope of the curves depicted in the figure indicated the loading rate, which might be influenced to a minor extent by daily management practices but was primarily driven by the magnitude of the applied load. Specifically, as the loading magnitude increased, the slope of the curves became steeper. Besides, it was observed that the accumulated ESALs for cell 45_HR_R1 and 45_HR_R2 were relatively lower than other test cells. This can be attributed to the fact that these cells experienced rutting failure at an earlier stage compared to the remaining cells. It is important to note that, due to the arbitrary arrangement of lanes and material allocations, the test IDs have been redefined in Table 6-1 for clarity and consistency in subsequent discussions. Initially, the dual tire assembly ran unidirectionally at a constant speed of 4 mph. However, due to equipment aging issues, the loading speed was increased to 6 mph on May 6, 2021. Consequently, most of the test cells were subjected to loading at 6 mph, while 3 test cells continued to be subjected to loading at 4 mph.

Table 6-1. Summary of Rutting Testing Cells and Timeline Within the APT BMD Experiment.

Test ID	Cell	Paving Date	Test Period		Loading		
			Start	End	Speed	# of passes	# of ESALs
30_C_R1	L1C	10/20/2020	06/22/2021	08/13/2021	6 mph	182,605	479,170
30_C_R2	L1B	10/20/2020	09/13/2022	10/17/2022	6 mph	136,422	589,261

30_O_R1	L5B	04/29/2020	06/10/2020	07/30/2020	4 mph	116,199	476,737
30_O_R2	L5E	04/29/2020	08/03/2020	10/02/2020	4 mph	134,378	505,751
45_HR_R1	L3E	07/15/2020	10/29/2020	12/19/2020	4 mph	135,411	256,075
45_HR_R2	L3A	07/15/2020	04/18/2022	05/09/2022	6 mph	110,664	374,457
45_HR_RA_R1	L2E	07/17/2020	08/20/2021	09/30/2021	6 mph	136,912	590,851
45_HR_RA_R2	L2A	07/17/2020	05/16/2022	06/11/2022	6 mph	132,439	527,431
45_HR_L_R1	L6E	10/22/2020	11/08/2021	12/08/2021	6 mph	130,645	516,337
45_HR_L_R2	L6A	10/22/2020	07/28/2022	08/25/2022	6 mph	130,358	530,659
60_HR_L_RA_R1	L4E	10/23/2020	10/13/2021	11/03/2021	6 mph	122,395	483,630
60_HR_L_RA_R2	L4A	10/23/2020	06/21/2022	07/22/2022	6 mph	132,661	550,435

APT = accelerated pavement testing; BMD = balanced mix design; R1 = rutting cell_1; R2 = rutting cell_2; ESALs = equivalent single axle loads.

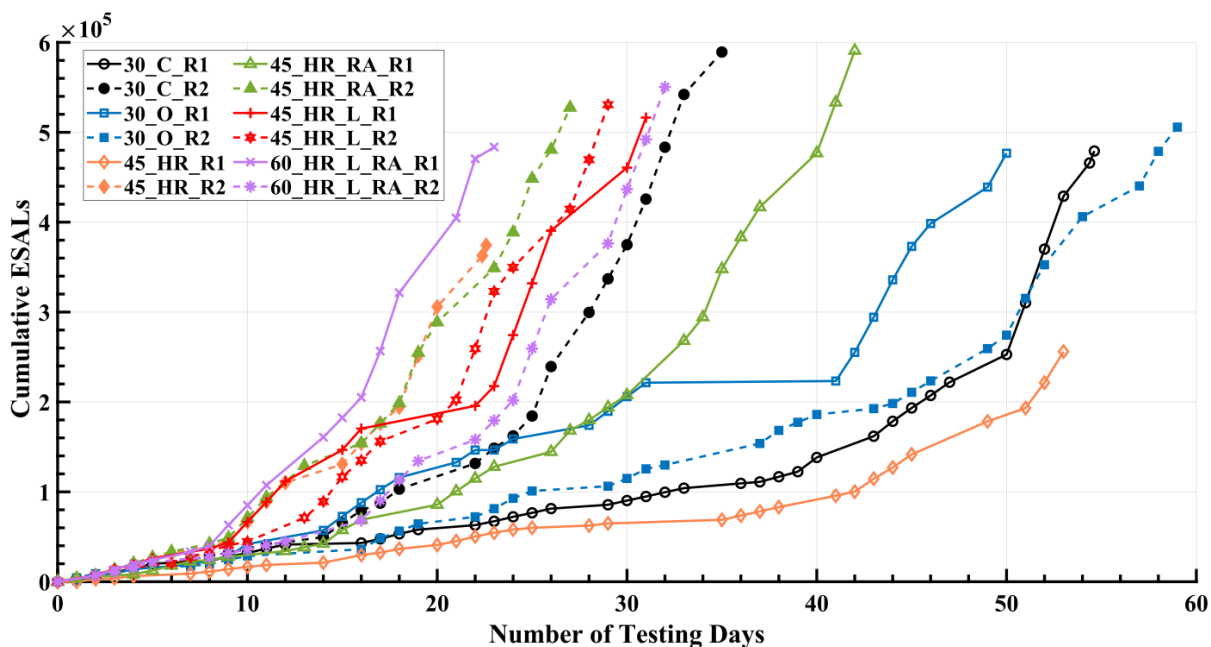


Figure 6-3. Loading Protocols for Various Test Cells. ESALs = equivalent single axle loads; # = number; R1 = rutting cell 1; R2 = rutting cell 2.

6.5.5. Rut Depth Measurement

The laser profiler, positioned on the carriage of the HVS as shown in Figure 6-1b, was used to conduct regular surface scans of the test bed. These scans were performed on a daily or bi-daily basis, excluding weekends and vacation periods. The laser profilers captured surface elevation measurements at 4-inch intervals across the longitudinal span of the HVS test cell (24 ft), resulting in the generation of a row of data containing 69 points. However, due to the location of the laser profiler at the head of the mounted carriage (Figure 6-1b), only the initial 53 points, equivalent to a distance of 5.385 m (17 ft. 8 in.), were deemed to provide effective measurement data. Subsequent to completing one set of longitudinal measurements, the laser profiler was shifted inward by 2.54 cm (1 inch) in the transverse direction. Within the transverse direction, a total of

70 measurement points were established, distributed symmetrically with 35 points on each side of the wheel path's center. A daily surface scan, presented as a three-dimensional (3-D) profile consisting of 53 transverse planes as shown in Figure 6-4a, effectively visualized the deformation of the pavement surface.

According to AASHTO glossary (2017) and ASTM E 1703/E 1703M, rut depth is defined as "the maximum measured perpendicular distance between the bottom surface of the straightedge and the contact area of the gage with the pavement surface at a specific location" (ASTM E 1703). This concept is visually represented in Figure 6-4b. Surface vertical permanent deformation (SVPD) serves as a quantitative measure of the maximum vertical permanent deformation experienced by the pavement surface within a specific transverse plane, as depicted in Figure 6-4b. In the subsequent analysis, the median value of rut depth within the 3-D profile was selected as a representative indicator of the magnitude of rutting, which was used to analyze and quantify the progression of rutting.

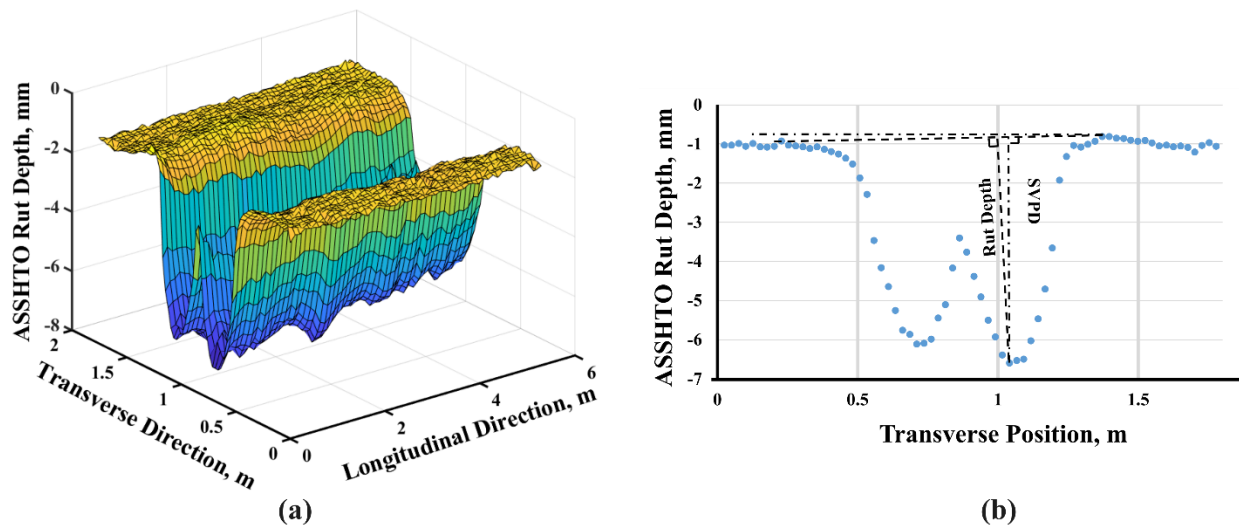


Figure 6-4. An Example of Measurement in the Experiment for Cell L1B on October 11, 2022: (a) 3-D Laser Profile; (b) Rut Depth and SVPD Illustration in a Specific Transverse Plane. SVPD = surface vertical permanent deformation.

6.6. APT Results and Discussions

Figure 6-5 shows the rut depth measurements collected within the 12 rutting test cells, with rutting cell_R1 and cell_R2 represented by solid triangles and circles, respectively. It is evident that mixture 45_HR showed the poorest rutting resistance compared to the other five mixtures, as indicated by its rapid accumulation of rut depth, reaching 12.5 mm, which is a typical threshold for rutting failure in pavements (Greene et al., 2015). This performance can be attributed to the significantly higher flexibility resulting from the high OBC used in the design and production of mixture 45_HR. On the other hand, the rutting performance of the remaining five mixtures appeared comparable, except for 60_HR_L_RA_R2, which exhibited a significant difference from 60_HR_L_RA_R1 after increasing the load magnitude from 9,000 lbf to 12,000 lbf, with its rut depth approaching 12.5 mm at the end of loading cycles. Nevertheless, despite a thorough

examination of the five mixtures, determining their rutting performance ranking proved to be difficult. This challenge arises from the variance in test results observed between the corresponding test cells, R1 and R2, which can confound a simple ranking approach. Consequently, further investigation was warranted to elucidate the underlying factors that contributed to the observed differences in performance between test cells R1 and R2, thereby facilitating a comprehensive evaluation and comparison of the various mixtures.

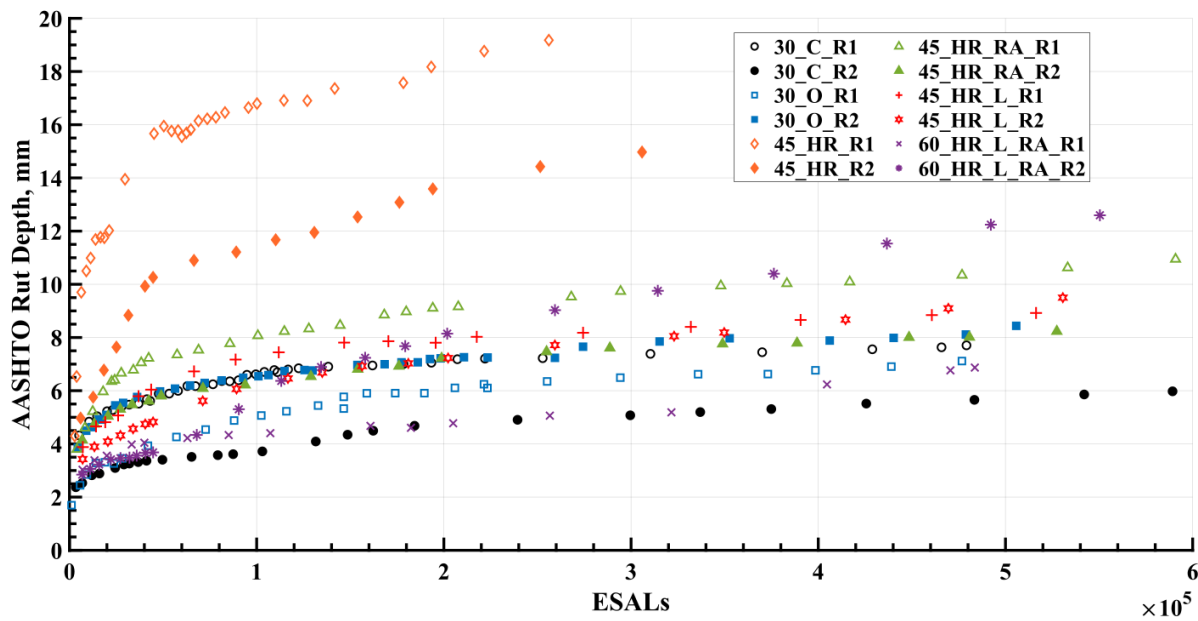


Figure 6-5. Rutting Depth for 12 Rutting Test Cells. ESALs = equivalent single axle loads.

6.6.1. Environmental Aging Effect

Upon placement and exposure to the environment, asphalt mixtures undergo an aging process characterized by the oxidation of the asphalt binder, leading to changes in the viscous properties of the binder and subsequently impacting the overall properties of the mixture (Cong et al., 2016). The measured rut depth results in Figure 6-5 clearly demonstrate the influence of aging effects, as the later-tested test cells (R2) outperformed their corresponding earlier-tested test cells (R1) in terms of rutting performance. In Figure 6-5, the solid circles, typically associated with larger values, surpassed the corresponding solid triangles, representing mixtures 30_C, mixture 45_HR, mixture 45_HR_RA, and mixture 45_HR_L. This observation aligned with the experiment timeline presented in Table 6-1, where test cells R2 were tested 448, 536, 269, and 262 days later than their corresponding test cells R1 for mixture 30_C, mixture 45_HR, mixture 45_HR_RA, and mixture 45_HR_L, respectively. Instances of such aging effects were commonly observed in the APT BMD experiment. For instance, in an APT testing conducted at the National Airport Pavement and Materials Research Center, it was found that the outdoor test sections exhibited superior rutting performance compared to their corresponding indoor sections, attributed to the environmental aging experienced by the outdoor sections (Garg et al., 2018). Furthermore,

aging effects were considered and investigated in APT testing conducted by FDOT (Zou et al., 2012), Indiana DOT (Tian et al., 2018), and University of California Pavement Research Center in Davis (Jones et al., 2011).

However, it was not only the aging effect but also the construction variability that played a significant role in investigating the resultant rut depth, which was believed to be a major contributing factor for the exception observed in mixture 30_O and mixture 60_HR_L_RA from the aging effect. For mixture 30_O, the test cells 30_O_R1 and 30_O_R2 were tested back-to-back and considering the relatively small difference in rut depth between them, it can be inferred that the construction variability of 30_O_R1 and 30_O_R2 was within a reasonable range and the influence of aging was minimal. As for mixture 60_HR_L_RA, test cell 60_HR_L_RA_R2 was tested 251 days later than 60_HR_L_RA_R1, and the scenario where the rut depth of 60_HR_L_RA_R1 (solid circle) exceeded that of 60_HR_L_RA_R2 (solid triangle) was observed only under the 9,000 lbf loading condition. However, as the applied loading magnitude increased to 12,000 lbf, the rut depth of 60_HR_L_RA_R2 immediately surpassed that of 60_HR_L_RA_R1. Within 500,000 ESALs, 60_HR_L_RA_R2 closely approached the failure threshold of 12.5 mm. It is believed that test cell 60_HR_L_RA_R1 and 60_HR_L_RA_R2 exhibited significant construction variability, surpassing the influence of aging. Notably, in the earlier laboratory performance tests, both the APA test and the SSR test, to be discussed in detail in the following sections, indicated relatively inferior rutting resistance for mixture 60_HR_L_RA compared with the other five mixtures, only slightly better than that of mixture 45_HR (Tong et al., 2024). Considering this, the performance of test cell 60_HR_L_RA_R2 aligned more closely with expectations, as it had substantial rut depth at the end of testing life.

6.6.2. Normalization

The aging process significantly impacted the rutting performance of the respective mixtures. Prior to comparing the rutting performance across distinct mixtures and establishing a correlation with results derived from laboratory testing, it was imperative to normalize them to the same set of initial conditions. To further this understanding, pavement age should be integrated as a significant variable within rutting model. Subsequently, an empirical model expressed in Equation 6-2 (Xue et al., 2017) was formulated to simulate the evolution of rut depth for each lane (two test cells).

$$RD = \alpha \cdot ESAL^{\beta_1} \cdot Age^{\beta_2} \quad \text{Equation 6-2}$$

Where RD is the AASHTO rut depth in mm; Age is the number of days passed from pavement paving date to the rutting measurement; and α , β_1 , β_2 are regression coefficients.

For the sake of brevity, the regression coefficients were not included in this paper. A positive coefficient ($\beta_1 > 0$) indicated that rut depth increases with an increase in ESALs. Conversely, a negative coefficient ($\beta_2 < 0$) suggested that aging had a negative impact on rut depth,

meaning that as the pavement aged, it contributed positively to the resistance against rutting. The high values of R^2 , ranging from 0.95 to 0.99, indicated a strong fit of the proposed rutting model to the measured data. Moreover, the high value of R^2_{adjust} highlighted the significance of both ESALs and age as influential factors in determining the rutting depth of the test cells. The combined effect of these two parameters explains a minimum of 94.78% (obtained in mixture 45_HR) of the variation in the measured rutting data during the experiment. Notably, loading speed was extended into Equation 2 to account for its potential influence on rutting performance, particularly for mixtures 30_O and 45_HR. However, the results of the regression and statistical analyses revealed that loading speed was not a significant predictor in these cases. This outcome could be attributed to the controlled nature of loading speed, which was implemented using the half-sine function. The rutting measurements were found to be less sensitive to small variations in loading speed compared to the effects of ESALs and pavement age. Therefore, Equation 2 was deemed sufficient for regression analysis across all test lanes.

In Table 6-1, it is noteworthy that the rutting model was intended to capture the measured data for two test cells within a single lane. However, as discussed earlier, deviations from the expected behavior were observed in the cases of mixture 30_O and mixture 60_HR_L_RA. For mixture 30_O, the two rutting cells exhibited similar outcomes due to their close testing times, while for mixture 60_HR_L_RA, the influence of construction variability surpassed the effects of aging. Consequently, separate curve fitting analyses were conducted for their corresponding rutting cells, namely R1 and R2, in these exceptional scenarios. The fitted curves depicting the rutting depth were then juxtaposed against the corresponding experimental measurements, as shown in Figure 6-6. Owing to the disparities in testing periods (Table 6-1), the measured rut depth in test cell R2 (blue point) frequently exhibited a smaller magnitude compared to that in test cell R1 (orange point), thereby highlighting the presence of an aging effect. This phenomenon is effectively showcased in Figure 6-6 through the effective utilization of the rutting model. Additionally, Figure 6-7 specifically represents the exceptional cases. The visual representations presented in these figures affirm that the proposed rutting model effectively captures the diverse rutting behaviors exhibited by the different test cells.

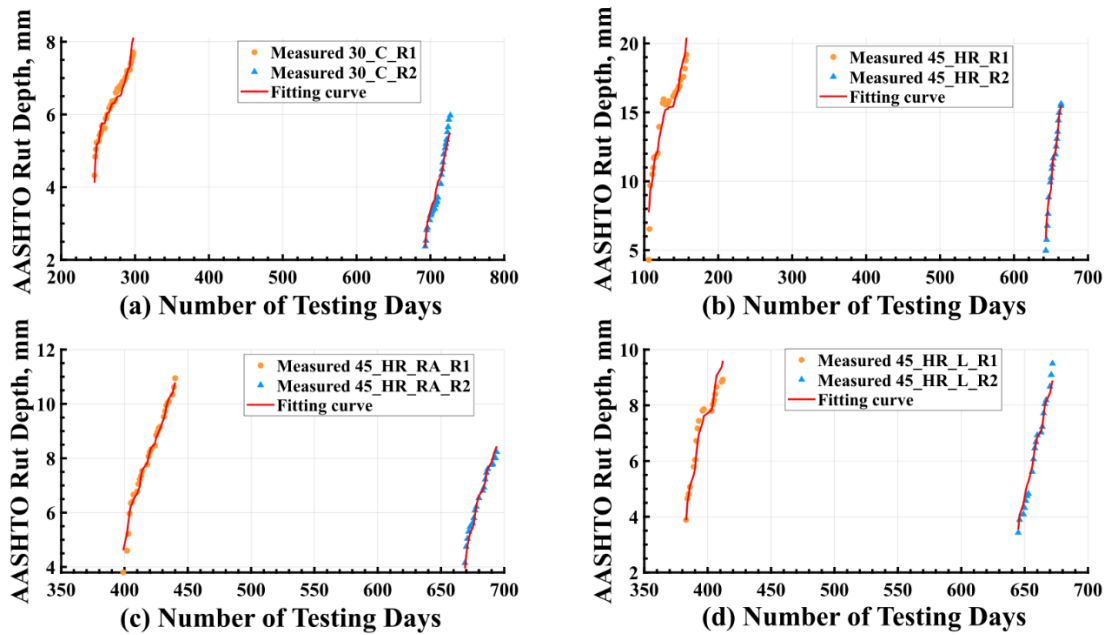


Figure 6-6. Fitting Result for Testing Lanes of (a) Mixture 30_C; (b) Mixture 45_HR; (c) Mixture 45_HR_RA; (d) Mixture 45_HR_L. ESALs = equivalent single axle loads; R1 = rutting Cell 1; R2 = rutting Cell 2.

Through the application of regression analysis, the rutting model facilitated the prediction of rutting development for each mixture under a well-defined loading protocol. This methodology ensured the comparability of different mixtures, allowing for more controlled analysis of variables. In addition to normalizing testing periods or environment aging effect, the rutting model effectively addressed the disparities that arise from the use of APT loading protocols under realistic management conditions (refer to Figure 6-3). Assuming identical testing start time was applied to all lanes, following the same regimen as the control mixture 30_C (245 days after paving). Adhering to the predetermined schedule of loading protocols for all test cells, a load of 9,000 lbf was applied for the initial 14 days, with 5,000 passes per day, resulting in a cumulative total of 70,000 ESALs during this period. Subsequently, a load of 12,000 lbf was applied for 7 days, followed by a final load of 15,000 lbf for a duration of one week (7.32 days), resulting in a total of 500,000 ESALs. The entire loading process spanned a duration of 28.32 days. The progression of rutting depth for each lane is shown in Figure 6-8.

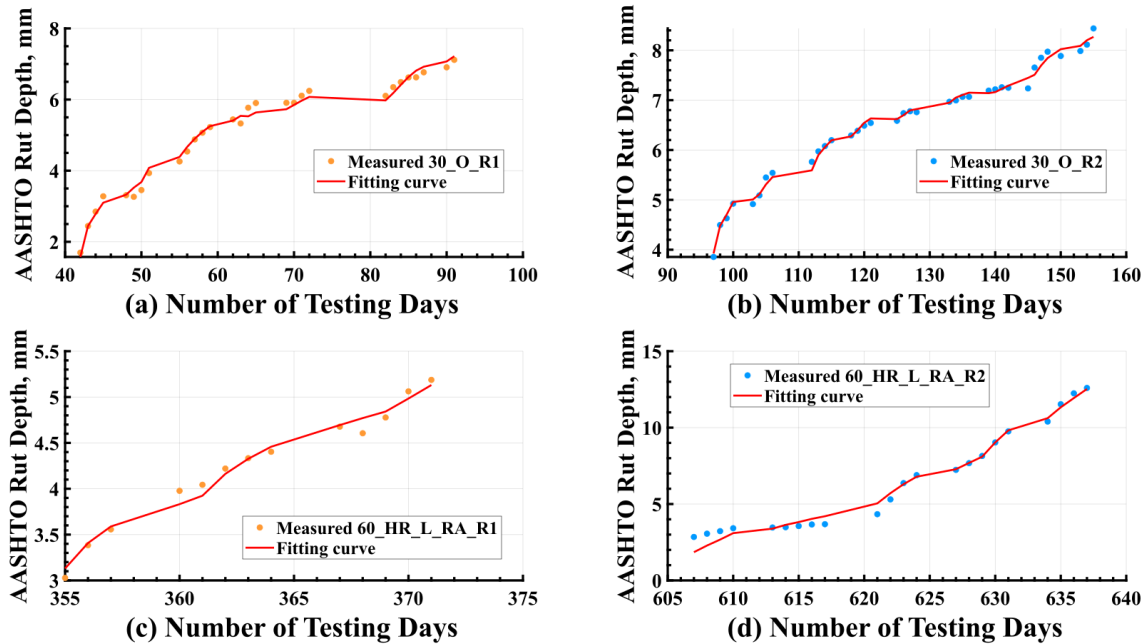


Figure 6-7. Fitting Result for Testing Cells: (a) Cell 30_O_R1; (b) Cell 30_O_R2; (c) Cell 60_HR_L_RA_R1; (d) Cell 60_HR_L_RA_R2. ESALs = equivalent single axle loads; R1 = rutting Cell 1; R2 = rutting Cell 2.

Upon visual inspection, the normalized results demonstrated comparability with the measured results (refer to Figure 6-5) among different mixtures, attesting to the consistency between the measured data and normalization approach. Notably, the exceptional case mixtures, 30_O and 60_HR_L_RA, were separately presented for their respective test cells R1 and R2. Among these, test cell 30_O_R1 and test cell 30_O_R2 demonstrated remarkably similar rutting performance, supporting the use of their average value for subsequent comparisons with other mixtures. Conversely, test cell 60_HR_L_RA_R1 displayed augmented sensitivity to increasing load magnitude when contrasted with test cell 60_HR_L_RA_R2. Drawing upon the insights garnered from the laboratory rutting performance tests conducted on mixture 60_HR_L_RA at varying complexity levels, the rut depth development of test cell 60_HR_L_RA_R2 was deemed representative and employed for subsequent comparisons with different mixtures. In terms of rutting performance, mixture 45_HR exhibited the highest rut depths followed by mixtures 60_HR_L_RA and 45_HR_RA, and then mixtures 45_HR_L, 30_C, and 30_O. It was observed that all BMD HRAP mixtures showed higher rut depths compared to the control mixture 30_C. This can be attributed to the inclusion of a high binder content in BMD HRAP mixtures to enhance cracking resistance, which may have compromised rutting resistance. It is worth noting that, despite the similarity in rutting performance between mixture 30_O and mixture 30_C as shown in Figure 6-5, the testing for cells containing mixture 30_O was conducted 203 days earlier than that for mixture 30_C. Considering the aging effect, after normalization, mixture 30_O exhibited better rutting performance compared to mixture 30_C. Furthermore, at the end of 500,000 ESALs,

certain mixtures, namely 45_HR, 45_HR_L, and 60_HR_L_RA, exceeded the failure limitation of 12.5 mm in terms of rut depths. It should be noted that the rutting failure threshold in APT may require further consideration due to the accelerated manners of APT, which involves incremental loading protocols and high testing temperatures.

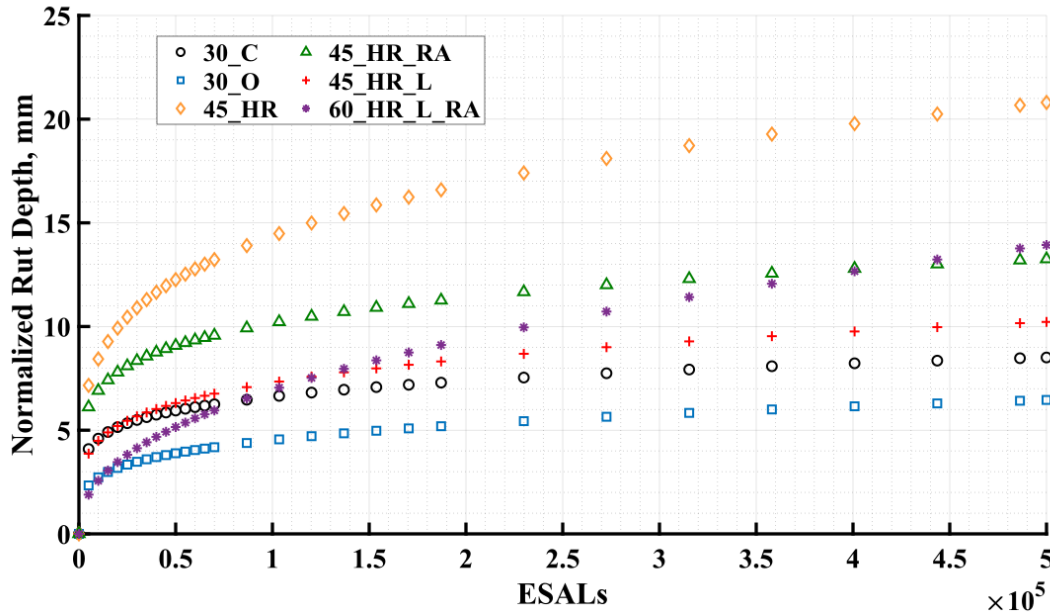


Figure 6-8. Normalized AASHTO Rut Depth Versus ESALs. ESALs = equivalent single axle loads.

6.7. Correlation Between APT Measurements and Laboratory Test Results

Two performance tests were used to evaluate the rutting resistance of the six mixtures. These tests encompassed two levels of testing: intermediate (APA rut test) and advanced (SSR test). Further information regarding the test results can be found elsewhere (Tong et al., 2024). Subsequent sections primarily focus on establishing and assessing the correlation between rut depth measurements obtained through the HVS and laboratory rutting performance tests.

6.7.1. APA Rut Test

The APA rut test was performed on specimens prepared from loose mixture obtained during production in accordance with AASHTO T 340. The test specimens had a diameter of 150 mm and a height of 75 ± 2 mm. The specimens were compacted using an Superpave gyratory compactor to achieve an air void content of $7 \pm 0.5\%$. This laboratory test aimed to simulate rutting by subjecting a loaded wheel in cyclic motion over a pressurized rubber tube positioned along the surface of the specimens. The test was performed at a temperature of 64°C , which corresponds to the high-temperature PG of asphalt binders typically used in Virginia (i.e., PG 64S-22). The deformation of the specimen, measured after applying 8,000 loading cycles, is referred to as the APA rut depth. Two replicate tests (four specimens in total) were conducted for each mixture, and the average results were reported (Tong et al., 2024).

6.7.2. SSR Test

The SSR test, conducted in accordance with AASHTO TP 134 (0), was employed to evaluate the rutting resistance of mixtures under repetitive loading, simulating the traffic-induced conditions experienced by pavements. Four specimens were subjected to a confined cyclic compression testing at two different temperatures, T_H (55°C) and T_L (26°C). In both cases, the specimens were subjected to three sets of 200-cycle loading blocks, each at three deviatoric stress levels, while maintaining a constant confining pressure of 69 kPa. During the testing, the permanent deformation of specimens was measured along with its deviatoric stress, loading history, and temperature. The rutting strain index (RSI), a metric adopted by the FHWA and supported by the SSR shift model, served as an indicator of rutting. The RSI represents the average permanent strain, expressed as a percentage, resulting from the repetition of 30 million ESALs over a period of 20 years on a standard pavement structure (Ghanbari et al., 2020). A lower RSI value indicates a greater resistance to rutting for a given mixture. It is noteworthy that only the second set of reheated mixtures sampled from plant production, compacted in the laboratory, underwent this advanced-level rutting performance evaluation.

6.7.3. Correlation

The Pearson correlation coefficient (Cohen, 1969) was used to examine the strength and direction of linear relationship between HVS testing results and laboratory rutting performance tests (APA and SSR tests). The results are summarized in Table 6-2. Rutting Testing Results. Among these, the HVS testing results were selected based on the normalized AASHTO rut depth for each lane at the culmination of 500,000 ESALs, as shown in Figure 6-8. For mixture 30_O, the average value of 6.46 mm was obtained by considering the mean rut depth measurements derived from test cell 30_O_R1 (6.25 mm) and test cell 30_O_R2 (6.67 mm). Conversely, for mixture 60_HR_L_RA, the rut depth of 13.93 mm at the termination of 500,000 ESALs was recorded from test cell 60_HR_L_RA_R1 and used for subsequent analysis. In the study, the SSR test results derived from the second set of reheated mixtures were employed as the representative advanced level indicator for rutting performance, as displayed in Table 6-2. Correspondingly, the APA test results from this identical set of reheated mixtures were also represented. Finally, the average values computed from four sets of non-reheated specimens and four sets of reheated specimens, alongside their standard deviations, were included in Table 6-2 for comprehensive comparison.

Table 6-2. Rutting Testing Results

Rutting Performance	Mixture ID					
	30_C	30_O	45_HR	45_HR_RA	45_HR_L	60_HR_L_RA
HVS Measured Rut Depth, mm	8.507	6.460	20.805	13.256	10.221	13.931
SSR RSI (for Set 2), %	2.33	1.85	6.39	3.28	2.72	5.77
APA Rut Depth (for Set 2), %	3.997	4.627	6.445	5.727	4.207	5.855
Average APA Rut Depth for Non-Reheats, mm	4.164	6.076	11.926	5.320	4.414	3.896
Stdev APA Rut Depth for Non-Reheats, mm	0.825	1.222	1.456	0.741	1.055	1.215

Average APA Rut Depth for Reheats, mm	4.288	4.262	6.689	5.117	4.484	5.180
Stdev APA Rut Depth for Reheats, mm	0.838	0.504	0.712	0.598	0.699	0.607

HVS = heavy vehicle simulator; SSR = stress sweep rutting; RSI = rutting strain index; APA = asphalt pavement analyzer; Stdev = standard deviation.

The rutting testing results, obtained from pairwise comparisons between APA_Set2 and RSI_Set2, between RSI_Set2 and HVS, and between APA_Set2 and HVS are presented in Figure 6-9a, Figure 6-9b, and Figure 6-9c, respectively. Figure 6-9d combines the comparison between the APA average value of four sets of non-reheated specimens and HVS, and between the APA average value of 4 sets of reheated specimens and HVS. The error bar indicates their corresponding ± 1 standard deviation. The use of linear regression curves with its R-square provided a characterization of the relationships between any two testing methods. Furthermore, the assessment of correlation was quantified through the use of correlation coefficients (r values). The r values of correlations between APA_Set2 and RSI_Set2, RSI_Set2 and HVS, and APA_Set2 and HVS were 0.863, 0.914, and 0.871, respectively. These elevated r values indicated a significant positive correlation among the rutting performance tests considered in this study.

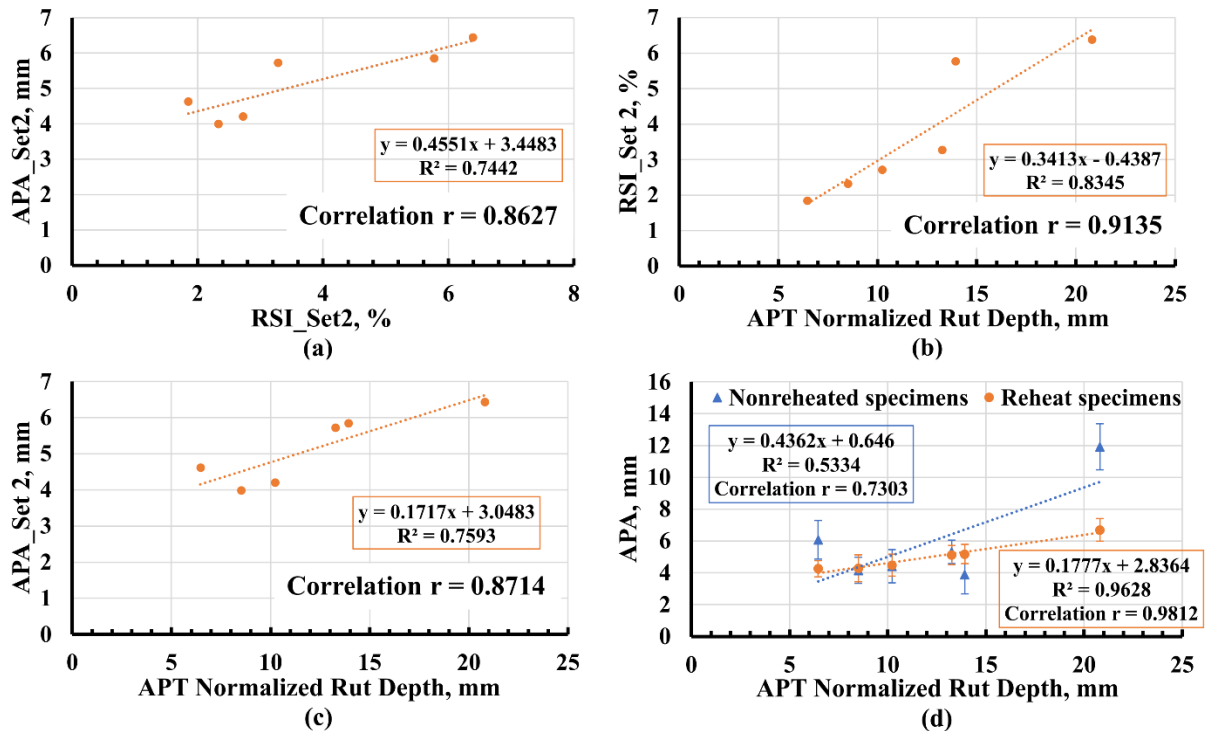


Figure 6-9. Rutting Testing Results Plot: (a) APA vs. RSI; (b); APA vs. HVS (c) RSI vs. HVS; (d) Correlation Analyses. HVS = Heavy vehicle simulator; APA = asphalt pavement analyzer; RSI = rutting strain index.

The highest positive correlation ($r = 0.9812$) was observed between the APA average value of four sets of reheated specimens and the HVS test. This aligned with the anticipated outcome, as averages derived from samples collected at various times during production or HVS construction

provided a more representative measure. Conversely, the weakest correlation ($r=0.73$) was seen between the APA average value of four sets of non-reheated specimens and the HVS test. The inconsistency in results can primarily be attributed to production variability and contractor-induced variability when rushing to prepare non-reheated specimens at the right air void during production. For example, this resulted in significant differences between the non-reheated and reheated sets of mixtures 30_O and 45_HR. Basically, the non-reheated specimens from mixtures 30_O and 45_HR had considerably more air voids than their reheated and designed counterparts, leading to greater APA rut depth. More details related to this analysis can be found elsewhere (Ghanbari et al., 2020).

Assuming the influence of production is discounted, the rutting performance of non-reheated sets could be comparable to that of the reheated sets. Consequently, a stronger correlation between the performance of non-reheated specimens and the HVS test results could possibly be extrapolated. Moreover, a linear relationship aptly describes the connection between different test results, particularly evident in the R-square value of 0.9628 for the APA average value of four sets of reheated specimens and the HVS test, highlighting a robust linear relationship between APA and HVS.

Aside from examining the correlation between laboratory-scale and full-scale testing, another crucial consideration of this APT experiment was the validation of the VDOT rutting threshold of 8 mm in terms of APA rut depth (determined on reheated specimens). For instance, FHWA recommended an RSI value ranging between 4% to 12% for standard traffic (ESALs < 10 million) (Ghanbari et al., 2020). The corresponding APA rut depth determined using the linear regression shown in Figure 6-9a, ranged between 5.269 mm and 8.910 mm with an average of 7.090 mm. Likewise, for the recommended RSI of 4% to 12%, the corresponding HVS rut depth ranged between 13.005 mm and 36.445 mm, as derived from Figure 6-9b. By using the regression analysis curve for reheated sets, the corresponding APA rut depth ranged between 5.15 mm and 9.31 mm with an average value of 7.23 mm. This value is lower than the VDOT current APA rut depth threshold of 8 mm. It should be reminded that the allowable HVS rut depth of 12.5 mm determined according to the literature will lead to an APA rut depth of 5.06 mm, which is significantly lower than the 8 mm VDOT BMD threshold. From another viewpoint, considering the rutting failure limitation of 24.5 mm allowed in airfield pavement (Garg et al., 2018), the corresponding rut depth was observed to be 7.19 mm from Figure 6-9d. All of that indicates that VDOT may need to revise their BMD specification by reducing the APA rut depth threshold to at least 7.000 mm. However, it is vital to recognize that these are initial findings, constrained by the limited number and types of mixtures evaluated in this study. Furthermore, these observations are pending further substantiation from real-world field performance data of sections constructed in Virginia.

6.8. Summary of Findings, Conclusions, and Future Work

This research provided a comprehensive evaluation of the permanent deformation behavior of a control mixture and five BMD optimized SMs through full-scale testing at APT facility under incremental loading conditions. The incorporation of various combinations of RAP contents, binder PGs, RA, and WMA additive, including four HRAP mixtures, allowed for a thorough investigation of their rutting performance. This paper specifically documented the design of the 2020 APT BMD program, the traffic test plan, and the measurement of rutting performance.

The empirical rutting model proposed in this study effectively accounted for the observed environment aging effect and facilitated the normalization of measured rut depth. The results demonstrated that all BMD HRAP mixtures exhibited higher rut depths compared to the control mixture 30_C. This can be attributed to the higher binder content in BMD HRAP mixtures, which aimed to enhance cracking resistance but compromised rutting resistance. Notably, mixture 45_HR showed a high likelihood of reaching the rutting failure even under a loading of 9,000 lbf. Nevertheless, when considering the aging effect, BMD optimized mixture 30_O displayed better rutting performance compared to the control mixture.

Correlation analysis revealed significant and strong positive correlations between APT measured rut depth and multi-level laboratory rutting performance tests. Remarkably, the highest positive correlation ($r = 0.98$) was observed between the APA average value of reheated specimens and the HVS test.

Finally, this research provided a steppingstone towards the validation of the VDOT BMD APA rut depth threshold of 8mm threshold based on actual full-scale accelerated testing and based recommended RSI value for standard traffic by means of fundamental and engineering performance test (i.e., SSR).

Subsequent research will explore the impact of incremental loading protocols on the mixtures using the time-hardening approach and will assess the potential applicability of the conventional load equivalency coefficient of 4.2 to BMD mixtures in general and HRAP mixtures in particular. Concurrently, an in-depth study on the cracking section is in progress, with an attempt to validate the VDOT BMD threshold of 7.5% and 70 for mass loss and CT_{index} by means of Cantabro tests and IDT-CT, respectively.

6.9. Reference

- American Association of State Highway and Transportation Officials. (2017). Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards. Washington, DC.
- ASTM International, (2023). *ASTM E1703*, Standard Test Method for Measuring Rut-Depth of Pavement Surfaces Using a Straightedge, West Conshohocken, PA.
- Choubane, B., Page, G. C., & Musselman, J. A. (2000). Suitability of Asphalt Pavement Analyzer for Predicting Pavement Rutting. *Transportation Research Record*, 1723(1), 107-115.

- Cohen, J. (2013). *Statistical Power Analysis for The Behavioral Sciences*. routledge.
- Cong, P., Liu, N., Tian, Y., & Zhang, Y. (2016). Effects of Long-Term Aging on the Properties of Asphalt Binder Containing Diatoms. *Construction and Building Materials*, 123, 534-540.
- Faccin, C., Specht, L. P., Schuster, S. L., Boeira, F. D., Bueno, L. D., Brondani, C., ... & Nascimento, L. A. H. D. (2022). Flow Number Parameter as a Performance Criteria for Asphalt Mixtures Rutting: Evaluation to Mixes Applied in Brazil Southern Region. *International Journal of Pavement Engineering*, 23(9), 3055-3067.
- Flintsch, G., Xue, W., Diefenderfer, B., & Meroni, F. (2020). Evaluation of Cold Central-Plant Recycling (CCPR) Technique Using Full-Scale Accelerated Pavement Testing. In *Accelerated Pavement Testing to Transport Infrastructure Innovation: Proceedings of 6th APT Conference* (pp. 270-279). Springer International Publishing.
- Garg, N., Kazmee, H., Ricalde, L., & Parsons, T. (2018). Rutting Evaluation of Hot and Warm Mix Asphalt Concrete under High Aircraft Tire Pressure and Temperature at National Airport Pavement and Materials Research Center. *Transportation Research Record*, 2672(23), 117-127.
- Ghanbari, A., Underwood, B. S., & Kim, Y. R. (2022). Development of a Rutting Index Parameter Based on the Stress Sweep Rutting Test and Permanent Deformation Shift Model. *International Journal of Pavement Engineering*, 23(2), 387-399.
- Greene, J., Chun, S., Nash, T., & Choubane, B. (2015). Evaluation and Implementation of PG 76-22 Asphalt Rubber Binder in Florida. *Transportation Research Record*, 2524(1), 3-10.
- Jones, D., Wu, R., Tsai, B. W., & Harvey, J. T. (2011). Warm-Mix Asphalt Study: Test Track Construction and First Level Analysis of Phase 3b HVS and Laboratory Testing (Rubberized Asphalt, Mix Design# 2).
- Kawa, I., Zhang, Z., & Hudson, W. R. (1998). Evaluation of the AASHTO 18-kip Load Equivalency Concept (Vol. 7). Austin, TX, USA: Center for Transportation Research, Bureau of Engineering Research, University of Texas at Austin.
- Kim, D., & Kim, Y. R. (2017). Development of Stress Sweep Rutting (SSR) Test for Permanent Deformation Characterization of Asphalt Mixture. *Construction and Building Materials*, 154, 373-383.
- Meroni, F., Flintsch, G. W., Habbouche, J., Diefenderfer, B. K., & Giustozzi, F. (2021). Three-Level Performance Evaluation o High RAP Asphalt Surface Mixes. *Construction and Building Materials*, 309, 125164.
- Meroni, F., Xue, W., Flintsch, G. W., & Diefenderfer, B. K. (2020). Impact of Mix Design Optimization on HMA Rutting Performance under Accelerated Pavement Testing. In *Accelerated Pavement Testing to Transport Infrastructure Innovation: Proceedings of 6th APT Conference* (pp. 108-117). Springer International Publishing.
- Mota, R. V., Kuchiishi, A. K., Takahashi, M. M., de Souza, G., Camargo, F. F., Bessa, I. S., ... & Bernucci, L. L. (2021). Effect of Binder Rheology and Aggregate Gradation on the Permanent Deformation of Asphalt Mixtures. *International Journal of Civil Engineering*, 19(7), 777-787.

- Tian, Y., Lee, J., Nantung, T., & Haddock, J. E. (2018). Calibrating the Mechanistic–Empirical Pavement Design Guide Rutting Models Using Accelerated Pavement Testing. *Transportation Research Record*, 2672(40), 304-314.
- Tong, B., Habbouche, J., Diefenderfer, S. D., & Flintsch, G. W. (2024). Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia. *International Journal of Pavement Engineering*, 25(1), 2325609.
- Xue, W., Flintsch, G. W., & Diefenderfer, B. K. (2020). Measuring Pavement Permanent Deformation in Accelerated Pavement Testing. *Transportation Research Record*, 2674(5), 340-348.
- Xue, W., Flintsch, G. W., Diefenderfer, B. K., & Bowers, B. F. (2017). Characterization and Modeling of Recycled Pavement Sections (No. CAIT-UTC-NC29). Rutgers University. Center for Advanced Infrastructure and Transportation.
- Yin, F., Chen, C., West, R., Martin, A. E., & Arambula-Mercado, E. (2020). Determining the Relationship among Hamburg Wheel-Tracking Test Parameters and Correlation to Field Performance of Asphalt Pavements. *Transportation Research Record*, 2674(4), 281-291.
- Zou, J., Roque, R., & Byron, T. (2012). Effect of HMA Ageing and Potential Healing on Top-Down Cracking Using HVS. *Road materials and pavement design*, 13(3), 518-533.

CHAPTER 7 - Cracking Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: Insights from Full-Scale Accelerated Testing Program⁷

7.1. Abstract

The Balanced Mix Design (BMD) approach has gained prominence in recent years, showing potential to effectively address challenges associated with incorporating high reclaimed asphalt pavement (HRAP) in asphalt mixtures, such as increased brittleness and susceptibility to premature cracking. This study assessed the cracking performance of a control asphalt mixture and five BMD-optimized asphalt surface mixtures through the use of an accelerated pavement testing (APT) program. The mixtures featured various reclaimed asphalt pavement (RAP) contents, two binder performance grades, a recycling agent, and a warm mix asphalt additive. The analysis integrated continuous longitudinal strain monitoring, APT surveys, and laboratory assessments to rank the mixtures. The results revealed three primary response phases in longitudinal strain and utilized deformation uniformity as a data quality indicator, indicating that localized microcracking may significantly affect the strain response. APT surveys detected only low-severity cracks in most sections, and laboratory evaluations, including permeability and dynamic modulus tests, confirmed no structural damage for the evaluated mixtures, except for a 60% RAP section. All other BMD mixtures demonstrated superior cracking resistance compared to the control mixture, with HRAP BMD mixtures outperforming those with conventional RAP contents, underscoring the efficacy of BMD concept incorporating softer binders and recycling agents. Correlation analysis between APT and VDOT-adopted BMD tests, including indirect tensile cracking test (IDT-CT) and Cantabro tests, examined the current Cracking Tolerance (CT) index and Cantabro mass loss (CML) thresholds. The study also indicated that while current BMD thresholds provided sufficient margins for satisfactory APT cracking performance, rutting resistance may become a concern for over-asphalted BMD mixtures.

Keywords: accelerated pavement testing, RAP, BMD, cracking performance, IDT-CT, Cantabro test

⁷ This paper, with Bilin Tong as the first author, has been submitted for publication in January 2025. It is co-authored by Jhony Habbouche, Ph.D., P.E., Ernesto Urbaez Perez., Gerardo W. Flintsch, Ph.D., P.E., Stacey D. Diefenderfer, Ph.D., P.E., Brian K. Diefenderfer, Ph.D., P.E., Eugene Amarh, Ph.D., Samer Katicha, Ph.D.. Contribution: Conceptualization: Lead; Methodology: Lead; Resources: Equal; Data Processing and Analysis: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead.

7.2. Objective and Scope

The objective of this study was to investigate the cracking performance of asphalt SMs with conventional and high RAP contents using APT. The study specifically addresses the design of six SMs with varying RAP contents, the establishment of APT pavement test cells, and the traffic loading plan. A comprehensive analysis was conducted, incorporating longitudinal strain monitoring, APT field surveys, and multiple laboratory performance evaluations to rank the evaluated mixtures. Additionally, a correlation analysis was performed between APT field performance and laboratory BMD performance tests currently adopted by VDOT, including the IDT-CT and Cantabro tests, to examine and validate the effectiveness of existing VDOT BMD thresholds. Finally, the cracking performance explored in this study was analyzed alongside APT-derived rutting metrics, facilitating a discussion on the equilibrium or “balance” status of primary performance traits within the evaluated mixtures.

7.3. Experimental Design and Research Approach

7.3.1. Evaluated Materials

A total of one control and five BMD 9.5mm dense-graded SMs were investigated in this study. The mixtures incorporated various combinations of RAP contents, two binders (PG 64S-22 and PG 58-28), one RA, and one WMA additive. Laboratory-produced laboratory-compacted specimens were 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 APT project, the materials were well-controlled, and the RAP material used 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. Even though a 100% blending does not exist since the RAP binder may not be fully available, for the sake of this experiment a 100% blending was assumed, and the research team relied on performance tests to confirm enough virgin binder was added for compensation. The WMA additive used is a bio-based product that is also used as an antistrip agent. The mixtures are defined in Table 3-1.

7.3.2. Virginia APT Facility

The primary aim of laboratory testing protocols is to simulate the stress and strain conditions that pavements experience under typical traffic scenarios. However, it is crucial to recognize that the stress or strain states imposed on laboratory specimens may not fully align with those encountered in the field. To address this discrepancy between laboratory simulations, or ME design models derived from laboratory material characterizations, and real-world pavement stress conditions, APT offers a more accurate means of overcoming the inherent limitations of traditional laboratory tests by providing a realistic representation of field pavement stress conditions.

The key element of the APT for BMD project was the Dynatest HVS, specifically the Dynatest Mark VI model shown in Figure 7-1a. This system includes an environmental chamber designed to maintain a stable temperature in the loaded area. The HVS system automatically regulates the surface temperature to ensure that the temperature at a depth of 2 inches beneath the surface remains at the target level. Inside the environmental chamber, a dual-tire carriage, capable of reaching speeds up to 12.4 mph with a tolerance of ± 2 mph, was mounted. The system applied loads using a dual-tire assembly equipped with 11.00R22.5 tires inflated to 105 psi. Moreover, the system could complete up to 12,000 unidirectional or 24,000 bidirectional passes within a 24-hour period. The dual-tire carriage accommodated loads ranging from 6,750 lbs to 22,500 lbs. A laser profiler, mounted on the HVS carriage and shown in Figure 7-1b, scanned the pavement surface to measure the vertical permanent deformation. Data acquisition was facilitated by a system that collected signals from various instruments to monitor pavement conditions. National Instruments input modules were installed in a chassis with a controller to support data collection, and LabVIEW software was used to create an interactive interface for data management and analysis. The entire DAS system, along with its components, was housed in a weather-resistant chamber located near the HVS apparatus.

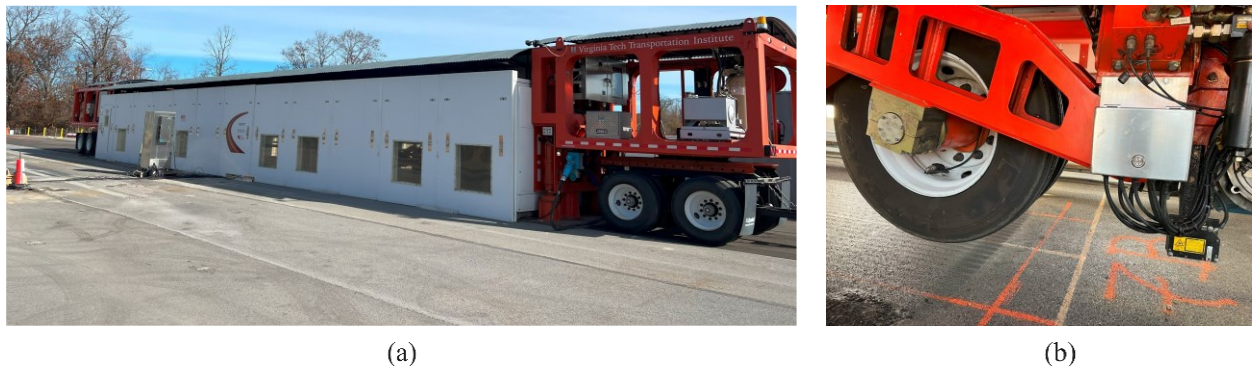


Figure 7-1. Photographs taken of: (a) Virginia APT facility; (b) Laser profiler mounted on HVS carriage. APT = accelerated pavement testing; HVS = heavy vehicle simulator.

7.3.3. Test Section and Instrumentation

Six APT lanes were constructed, each dedicated to evaluating one of six asphalt mixtures under full-scale accelerated loading. Each lane measured 300 feet in length and 10 feet in width. Within each lane, five test sections, labeled A through E, were established (shown in Figure 7-2a). Two sections, marked with blue, were designated for rutting evaluation, and the findings from this analysis have been previously published (Tong et al., 2023). The section marked in red was dedicated to the study of cracking, which is the main focus of this paper. Test sections A, C, and E were instrumented with various sensors, including three strain gauges (KM-100 HAS) placed at the bottom of the asphalt SM layer, spaced 2 feet apart, as well as pressure cells and moisture sensors at the bottom of the granular base layer, and thermocouples at a depth of 2 inches within the asphalt layer. The pavement structure, as depicted in Figure 7-2b, was uniform across all six lanes. Each lane was constructed using a research mixture consisting of two 1.5-inch asphalt lifts,

resulting in a total thickness of 3 inches. The asphalt layer was placed over a 12-inch VDOT 21B base layer, followed by an additional 26-inch layer of VDOT 21B used as the subgrade, all supported by a reinforced foundation. This pavement structure enables the isolation of SMs for performance evaluation. In addition, Figure 7-2b illustrates the typical instrumentation layout for each section. The installation of strain gauges is demonstrated in Figure 7-3a. To protect the strain gauges from damage during the constructions, loose mixture was placed around them before paving, as shown in Figure 7-3b. Sections B and D were designated as backup cells to address any unforeseen issues or damage encountered during the testing process. The test sections were strategically located within the central 200 feet of each lane, with effective tested area within each cell measuring 24 feet in length. The spacing between cells allowed room for the HVS support structure.

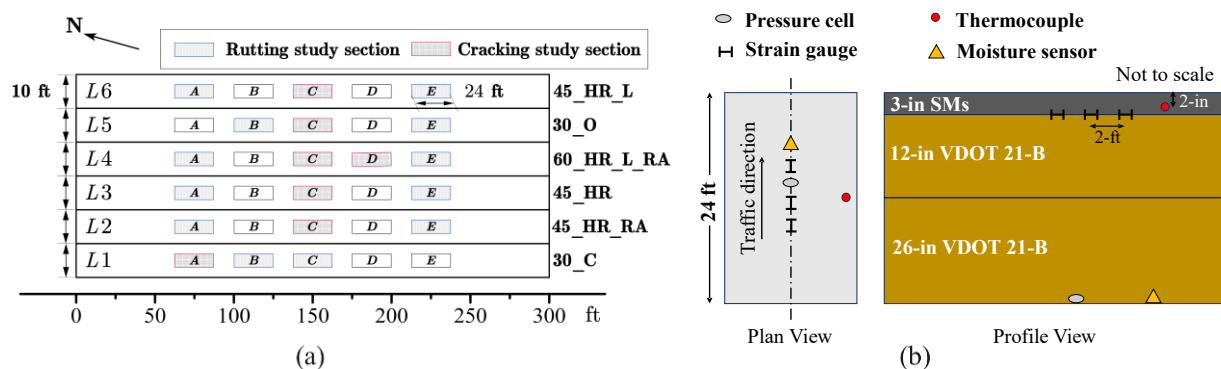


Figure 7-2. (a) Site layout for the 2020 APT project; (b) Test section instrumentation typical layout. APT = accelerated pavement testing.

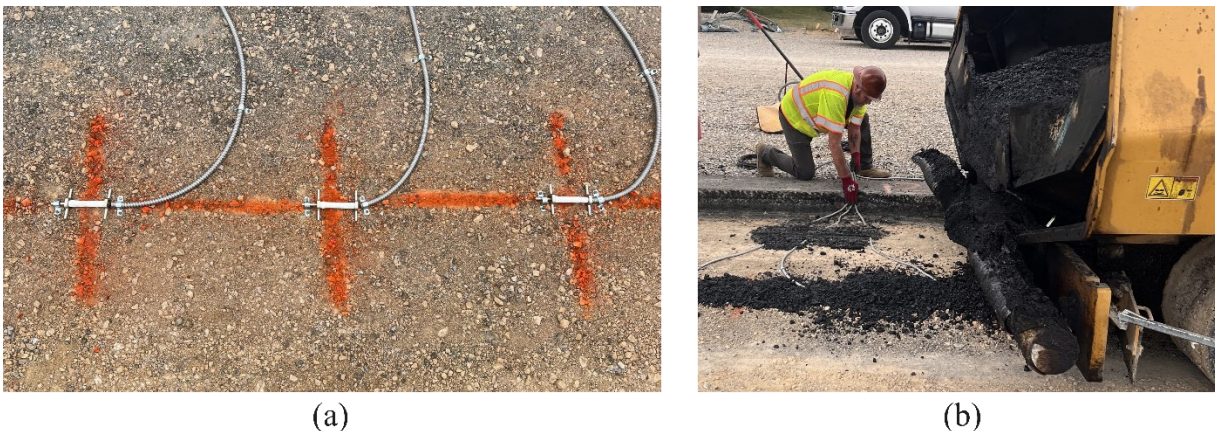


Figure 7-3. (a) Installing longitudinal strain gauge KM-100 HAS; (b) Strain gauge in construction

7.3.4. Loading Protocols

For standardized analysis in pavement design and evaluation, the applied loads and repetitions were commonly converted into equivalent single axle loads (ESALs), with a target of approximately 3,000,000 ESALs for each cracking test section. The calculation of ESALs was

conducted using Equation 1. The concept of 18-kip single-axle equivalency, introduced following the AASHO Road Test completed in 1961 was utilized in this study (Kawa et al., 1998). A single pass at load levels of 40 kN (9,000 lb) and 67 kN (15,000 lb) corresponded to 1.00 ESAL and 8.55 ESALs, respectively.

The loading protocol employed in each cracking test section involved the application of a bi-directional load of 67 kN (15,000 lb) to achieve a target of 3,000,000 ESALs, typically completed over a period of 6 to 8 weeks. This protocol was designed to accelerate the development of pavement distresses. An exception was made for the control section (30_C), where an initial load of 40 kN (9,000 lb) was applied for one week, after which the load was increased to 67 kN (15,000 lb) in alignment with the other sections. To simulate traffic wandering, the APT facility allows the entire loading frame to move laterally. This lateral movement occurred in 2-inch (50.8 mm) increments, with a maximum displacement of ± 17 inches (431.8 mm) from the centerline. The wandering pattern followed a standard normal distribution, with a mean value of 0 inches and a standard deviation of 127 mm (5 inches). Pavement temperature was maintained at 20°C (68°F), measured by a thermocouple embedded 51 mm (2 inches) below the surface. Throughout the testing period, the HVS operated continuously, with brief pauses for daily maintenance and occasional repairs. The temporary staff shortages during the COVID-19 pandemic impacted the testing schedule.

The APT BMD experiment spanned approximately three years. Except for mixture 60_HR_L_RA, only one cracking section was tested for each mixture. For mixture 60_HR_L_RA, initial testing in Lane 4, Section C (the first cracking section for this mixture) suggested that significant precipitation caused severe damage due to moisture, leading to pronounced fatigue distress. To verify these findings, Section D was used for further testing, resulting in two cracking sections for mixture 60_HR_L_RA, a unique situation not observed with the other mixtures. Initially, the dual tire assembly operated bidirectionally at a constant speed of 4 mph. However, due to operational challenges with the APT facility, the speed was increased to an average of 6 mph on May 6, 2021. Consequently, most test cells were loaded at 6 mph, except for mixture 45_HR and part of the cracking section of mixture 45_HR_RA, which continued to be tested at 4 mph.

7.3.5. Data Collection

7.3.5.1. Longitudinal strain gauge monitoring

Unlike rutting, which can be directly evaluated by surface deformation, cracking typically initiates at the bottom of the asphalt layer and becomes visible only after it propagates to the surface. Therefore, instrumentation is essential for monitoring cracking progression before it reaches the surface. Cracking results from the cyclic application of traffic loads and is primarily associated with tensile strain at the bottom of the asphalt layer, which can be measured using strain gauges (*Error! Reference source not found.*). In this study, data was collected continuously for four-minute intervals every three hours using a data acquisition system. This collection frequency

ensured that each data window maximized the capture of pavement responses across the entire scheduled wander pattern. Additionally, this approach allowed continuous monitoring of the pavement's response and enabled evaluation of whether the data effectively captured the different stages of cracking progression throughout the testing period.

An example of three strain gauges response under a pass of load are shown in Figure 7-4a. Considering the inherent heterogeneity of asphalt concrete, as well as factors related to production, construction, and environmental conditions, strain gauges within each section may exhibit varied responses. Inspired by the use of deformation uniformity as a data quality indicator in dynamic modulus testing (DMT), where three linear-variable displacement transducers measure deformation, this study employs deformation uniformity to evaluate the quality of data collected from APT. Additionally, deformation uniformity is explored as a potential means of identifying the initiation points of cracking. Deformation is defined in NCHRP Report 513 (Bonaquist, 2003) as *“the average difference between (or among) the amplitude of the deformation signals expressed as a percentage of the mean deformation”*. It can be calculated using Equation 7-1. If all deformation signals exhibit identical amplitudes, the deformation uniformity value is zero, indicating perfect consistency among the signals. According to AASHTO T378-17, a deformation uniformity threshold of 30% is established as the acceptable limit for DMT.

$$\text{Deformation Uniformity}(\%) = \frac{\sum_{i=1}^n |D_i - \bar{D}|}{n \cdot \bar{D}} \times 100\% \quad \text{Equation 7-1}$$

Where, n = the number of signals, D_i = deformation amplitude of the i -th signal, $\bar{D}$ = mean deformation amplitude.

Figure 7-4b demonstrates the typical longitudinal strain response, which aligned with the characteristic of the measurements from other investigators (Elseifi et al., 2012). The longitudinal strain response exhibits an initial compression (ϵ_{c1}), followed by tension (ϵ_t), and then a second compression (ϵ_{c2}), with the second compression peak consistently lower than the first. Before and after the passage of the moving load, the strain gauge response returns to zero. These zero responses formed the reference table for each data collection. The total strain (ϵ_{total}) is calculated as the difference between the maximum and minimum measured strain.

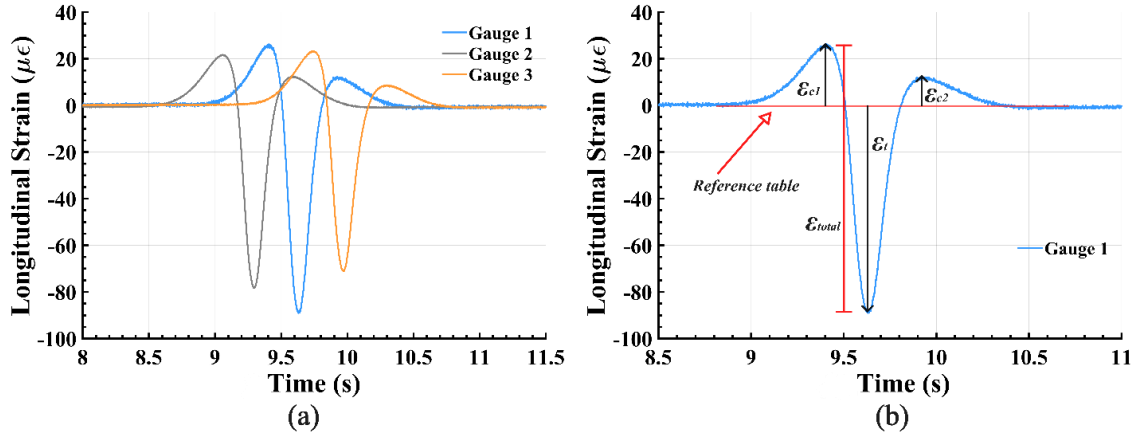


Figure 7-4. Typical response of mixture 45_HR_RA during initial testing (a) Strain response under one loading pass for three strain gauges; (b) Strain gauge 1 example. HR = high reclaimed asphalt pavement (RAP); RA = recycling agent.

7.3.5.2. Insights from data processing

A typical 4-minute data collection window captures approximately 31 passes of load, as illustrated by the blue lines in Figure 7-5. Within each 4-minute window, both tensile strain ϵ_t and compressive strain ϵ_{c1} were extracted for each pass, along with their peak values (highlighted by the red star maker). The peak values correspond to the strain responses when the moving load was applied above the strain gauges, centered on the wander pattern, where higher responses are expected compared to those at the edges. The peak tensile strain ϵ_t was used to monitor the pavement performance in this study. Meanwhile, the horizontal lines between loading passes represent the strain gauge response during the non-loading phase, corresponding to the rest period between passes of load. The midpoint of the non-loading response segment, marked by the green solid square, was used to indicate the residual strain $\Delta\epsilon$. From the data processing, three primary phases can be summarized. At phase I (initial testing stage), the responses to consecutive loading passes were relatively uniform and repeatable, with the ϵ_{c1} and ϵ_t magnitude primarily determined by the position of the load within the wander pattern. During non-loading intervals, the residual strain recovered to a stable value near to 0, indicating that the reference table stabilized during the 4-minute data collection period shown in Figure 7-5.

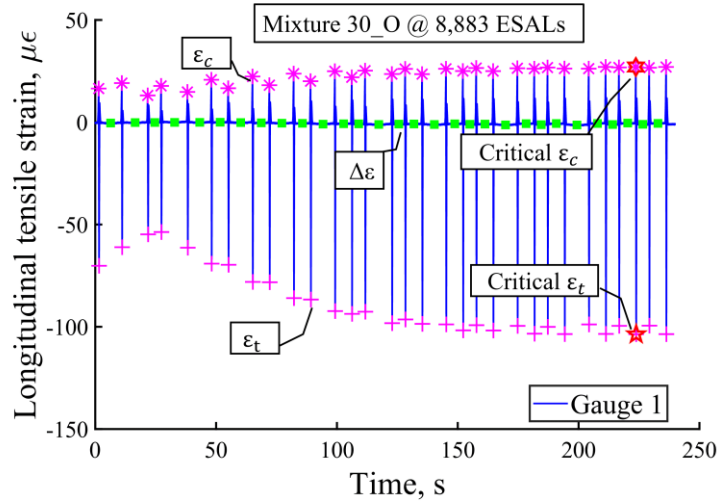


Figure 7-5. Typical 4-min data collection window for mixture 30_O at phase I: Stable reference table at initial stage of testing. O = optimized.

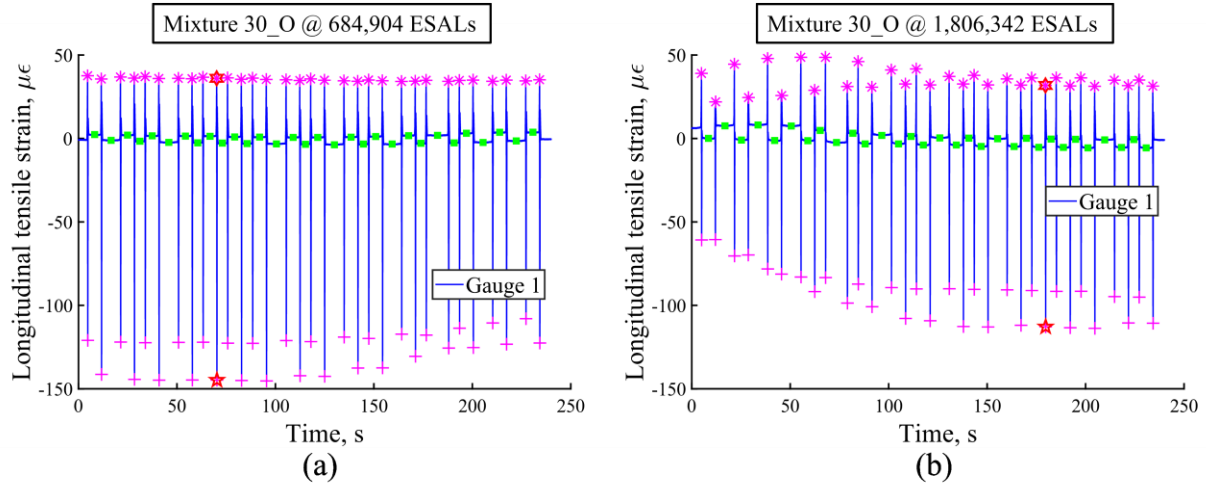


Figure 7-6. Typical 4-min data collection window for mixture 30_O at (a) phase II: Fluctuated reference table occurs; (b) phase III: Increased reference table occurs. O = optimized.

Subsequently, with the accumulation of loading passes, both ϵ_{c1} and ϵ_t magnitudes increased, which is attributed to the reduction in stiffness. The system then transitions to Phase II, as shown in Figure 7-6a, where the higher ϵ_{c1} and ϵ_t were measured compared to that of Figure 5, while the reference table began to fluctuate, and the $\Delta\epsilon$ no longer returned to zero. It was observed that this phase varied in duration depending on the evaluated materials. At this phase, ϵ_{c1} and ϵ_t may reach their peak values for the testing period, while the residual strain continues to increase. This stage is also consistent with previous studies, which noted the presence of some noise in the mid-phase of data collection, although ϵ_{c1} and ϵ_t could still be extracted (Elseifi et al., 2012). Finally, Phase III is characterized by a further increase in residual strain, introducing more uncertainty in the response collection in terms of ϵ_{c1} and ϵ_t magnitudes. This could be attributed to the initiation of microcracking near the strain gauges and varying degrees of crack development.

Other potential contributing factors include material heterogeneity, construction, and installation issues. During data processing, Phase III exhibited ε_{c1} and ε_t magnitudes in three distinct ways:

(1) a decrease in ε_{c1} and ε_t magnitudes, as illustrated in Figure 7-6b. In some extreme cases, residual strain exceeding ε_{c1} and ε_t in the final testing stage, rendering the extraction of these responses infeasible. This phenomenon was captured in the Figure 7-7a for the control mixture;

(2) a stabilization of ε_{c1} and ε_t magnitudes over an extended period, where crack development has reached a steady state, with no further significant increase. From an energy perspective, it may result from the dissipation of applied load energy due to microcracking; and

(3) a rapid increase in ε_{c1} and ε_t magnitudes from loading, likely linked to the extent of crack development.

A particular scenario was observed in the control mixture 30_C. An increase in ε_t for gauge 2 was observed exclusively in a single loading direction, unlike the strain responses in other mixtures. The discrepancy in the tensile strain response between consecutive loadings at the same position exceeded $100 \mu\epsilon$, as depicted in Figure 7-7b. Regardless of the scenario, the emergence and growth of reference fluctuations become a key issue in data processing. Therefore, in addition to ε_{c1} and ε_t , the reference table or residual strain $|\Delta\varepsilon|$ was also monitored in this study.

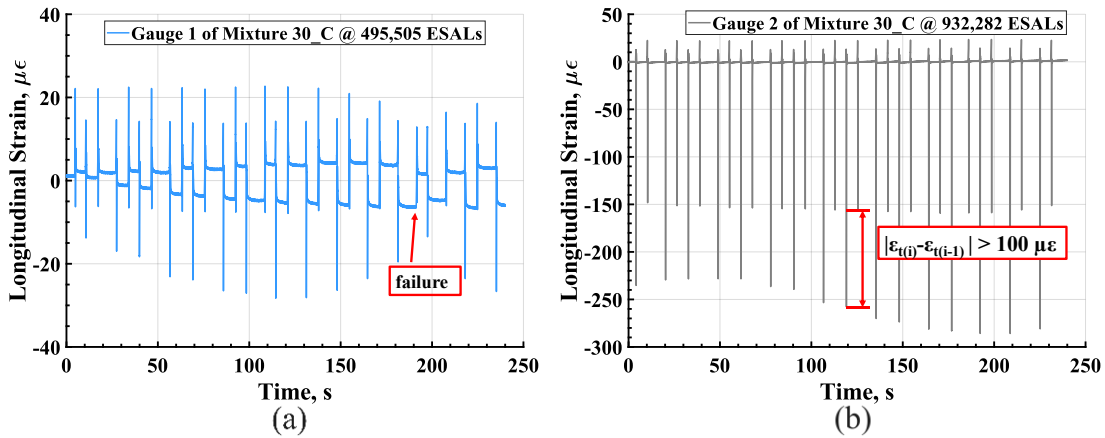


Figure 7-7. Longitudinal strain response of mixture 30_C: (a) failed to extract tensile and compressive strain gauge 1 @ 495,505 ESALs; (b) crack initiation point at gauge 2 @ 932,282 ESALs. C = control.

Based on the longitudinal tensile strain monitoring, this paper proposes using peak residual strain to quantify the potential number of ESALs to cracking initiation point, N_i . The use of N_i instead of the number of cycles to failure, N_f , is due to the sensitivity of the strain gauges employed to localized spots. These gauges can capture localized responses, such as the accumulation of factors like microcracking. However, pavement failure is typically defined as the point at which 20% of the entire surface area is cracked. Consequently, strain gauges, which affected by localized spots, are likely to show variations well before the actual failure point.

7.3.5.3. Crack survey of evaluation

Cracking development can be assessed by visually observing surface cracks on the pavement surface in APT (Long et al., 1996; Tsai et al., 2003; Ozer et al., 2018). As part of the daily maintenance of the testing, technicians conduct a straightforward visual inspection and record the time when hairline cracking is initially observed. However, visual inspections during APT operations can be challenging, as various environmental factors may hinder accurate detection of initial cracking. For example, dust from underlying materials can obscure microcracking on the surface. A field survey was conducted at the end of the APT testing program for each section, after washing the pavement surface with water. The loading area was mapped, and all visible cracks were marked and some examples are shown in Figure 7-8. The crack lengths detailed in this study are derived entirely from the cumulative length of observable micro-cracks identified through APT field surveys within each section.

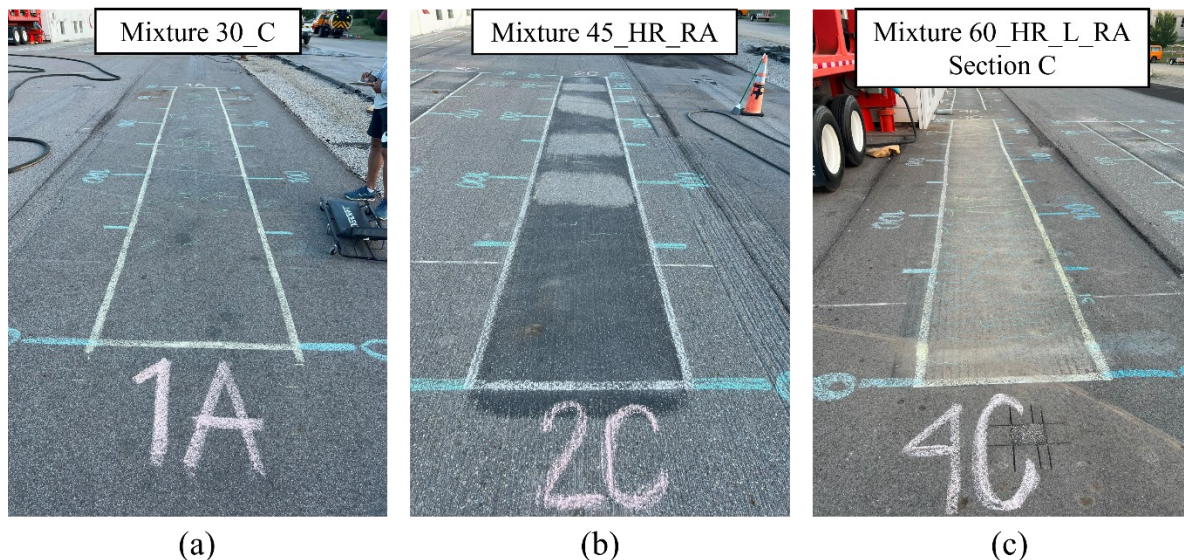


Figure 7-8. Examples of cracking sections after testing with cracks marked for mixtures: (a) 30_C; (b) 45_HR_RA; (c) 60_HR_L_RA. C = control; HR = high reclaimed asphalt pavement (RAP); RA = recycling agent; L = softer virgin binder.

7.3.5.4. Collection and lab evaluation of field cores

Laboratory tests were conducted on field cores extracted from each APT test section to assess the potential degradation of the evaluated mixtures. As illustrated in Figure 7-9, three cores were collected from random locations within the loaded area (labeled LA-1 to LA-3) for each tested section, and three corresponding cores were taken from the unloaded area (labeled UA-1 to UA-3). The permeability was measured on field cores (shown in Figure 7-10a) using a flexible wall falling head permeameter in accordance with Virginia Test Method-120 (VTM 120) (Maupin, 2010). The permeability test results can reflect the degree of internal air void connectivity within the evaluated specimens. Following the permeability tests, two small-scale specimens, each with a diameter of 38 mm (1.5 in.) and a height of 110 mm (4.3 in.), were extracted from each field core

as shown in Figure 7-10b. These specimens were prepared for small-scale DMT as represented in Figure 7-10c in accordance with AASHTO TP 132-19. The test frequencies were set at 0.1, 1, and 10 Hz, with the temperature maintained at 20°C. This temperature setting was chosen to prevent damage to the test specimens. The DMT tests, along with the permeability tests used in this study, were conducted to determine if structural damage could be detected in any of the tested pavements.

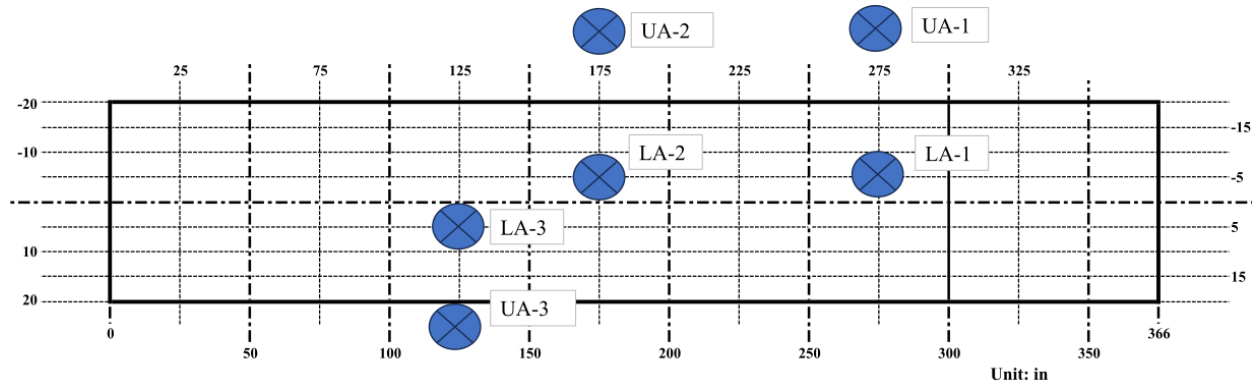


Figure 7-9. An example of the location of field cores within the test section in Lane 1. LA = loaded area; UA = corresponding unloaded area.

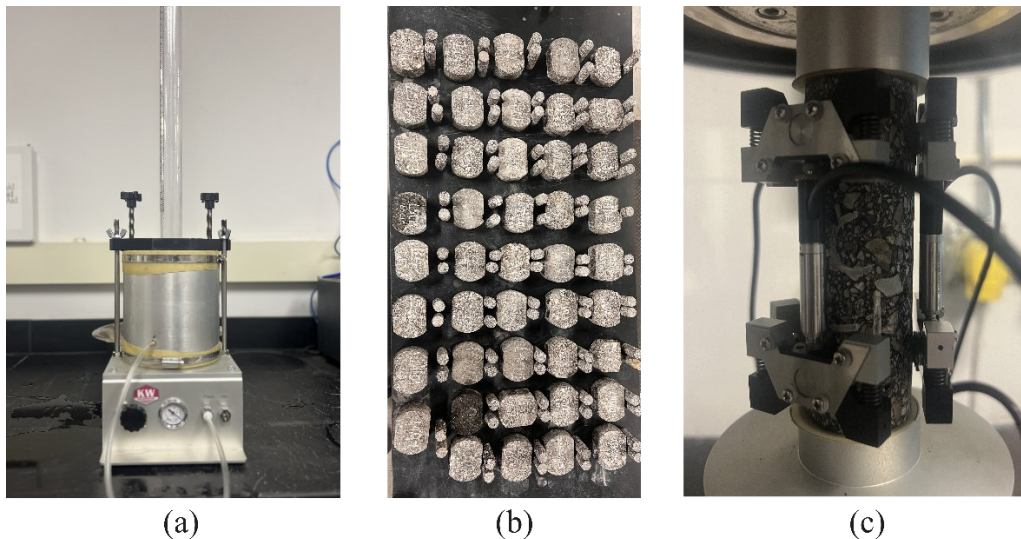


Figure 7-10. (a) Permeability test; (b) A portion of field cores for DMT; (c) Small-scale specimen DMT. DMT = dynamic modulus test.

7.4. Results and Discussion

7.4.1. Longitudinal Strain Monitoring

7.4.1.1. Tensile Strain Monitoring

The peak longitudinal tensile strain (ϵ_t), representing the pavement response from the loading passes at the center of the wandering pattern, was used to monitor the development of longitudinal strain versus ESALs shown in Figure 7-11. The pattern was similar to that observed in total strain monitoring. Also shown in Figure 7-11 is a vertical red dashed line representing the number of ESALs to initiation of cracking, N_i . More details are explained in the following section.

Every strain gauge remained functional after construction, except for one strain gauge in mixture 45_HR and 45_HR_L, which became non-responsive shortly after construction. In Figure 7-11, it can be observed that the strain gauges within each tested section collected similar responses during the initial phase, specifically within the range of less than 0.1×10^6 ESALs (equivalent to approximately 10,000 passes of load). Additionally, in this early stage, the typical response patterns for each strain gauge were clear, consistent, and free from noise. These observations suggest that the strain gauges demonstrated stable behavior during the initial testing period. Furthermore, the results imply that proper installation and compaction efforts ensured that the strain gauges remained securely bonded within the asphalt layer at this stage. However, as the cumulative number of load passes increased, the responses recorded by the strain gauges began to diverge. This variability likely reflects the transition of the strain gauges into the previously discussed three phases of response at different times. The onset of this divergence may be attributed to the development of localized microcracking in the vicinity of the strain gauges, influencing their recorded responses.

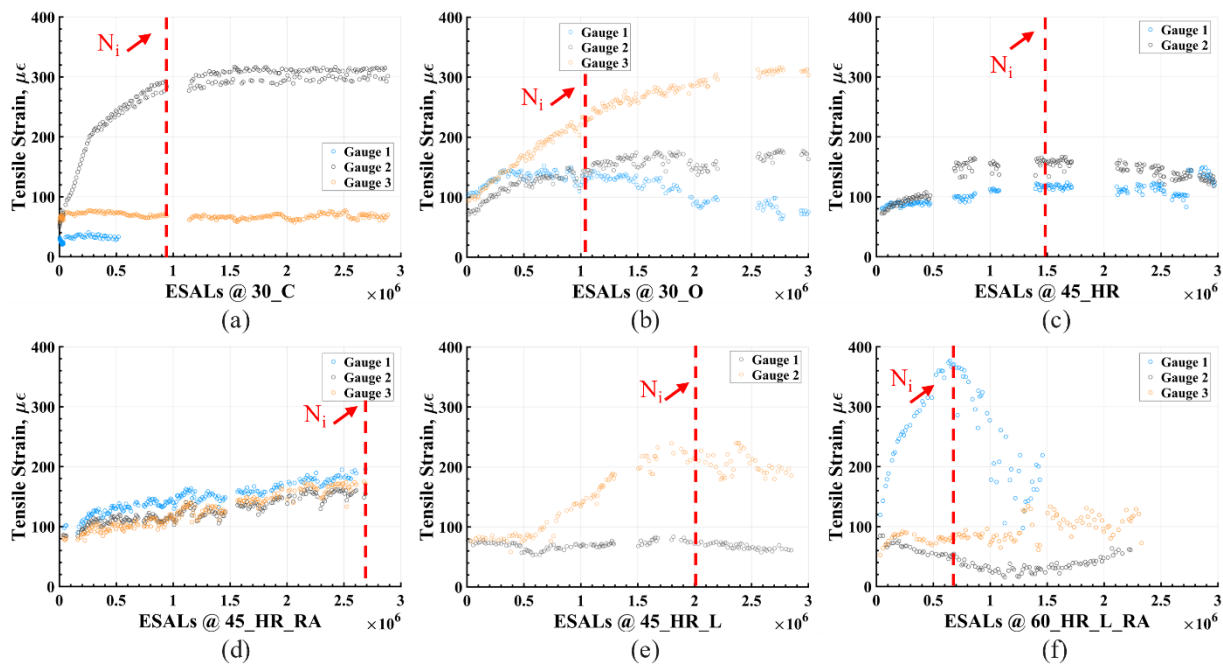


Figure 7-11. Longitudinal strain in the center line monitored along with testing for different mixtures. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; N_i = the number of ESALs to cracking initiation.

Specifically, for the 30% mixtures, both mixture 30_C (Figure 7-11a) and 30_O (Figure 7-11b) showed a drop in tensile strain at gauge 1. In the case of mixture 30_C, gauge 1 was considered to be in the post-Phase III stage, making it impossible to extract subsequent tensile and compressive strain data captured in Figure 7-7a. Such behavior was not observed in mixture 30_O. Additionally, gauge 2 of mixture 30_C and gauge 3 of mixture 30_O accumulated tensile strain exceeding $300 \mu\epsilon$, with gauge 2 of mixture 30_C showing particularly rapid accumulation.

However, it was observed that the increase in ε_t of mixture 30_C confined to a single loading direction, contrasting sharply with the strain responses seen in other mixtures. Notably, the peak tensile strain response showed a discrepancy exceeding 100 $\mu\varepsilon$ between successive loadings, as illustrated in Figure 7-6b. This marked variance suggested a potential complete fracture on one side—either left or right—of the strain gauge. This behavior, in comparison to the BMD mixtures, displayed a relatively less ductile characteristic. The cracking resistance of mixture 30_O was considered to be better than that of mixture 30_C.

Next is mixture 45_HR, in Figure 7-11c where tensile strain showed a declining trend after 1.5 million ESALs. The tensile strain response remained below 180 $\mu\varepsilon$ throughout the testing period, which could be attributed to the higher binder content in this mixture compared to the others. Mixture 45_HR_RA, illustrated in Figure 7-11d, consistently demonstrated uniform and gradual longitudinal strain development across the three strain gauges. The tensile strain increased almost linearly from 80 $\mu\varepsilon$ to 200 $\mu\varepsilon$, highlighting the stability and resistance of this mixture. Mixture 45_HR_RA is considered to exhibit the best cracking resistance among the mixtures evaluated. In Figure 7-11e, for mixture 45_HR_L, strain gauge 1 stabilized at approximately 200 $\mu\varepsilon$ without a significant drop in response, while strain gauge 2 maintained high stability throughout the loading process. Mixture 45_HR_L is ranked as the second-best in cracking resistance, following Mixture 45_HR_RA, and outperforms Mixture 45_HR.

In contrast, Mixture 60_HR_L_RA, as shown in Figure 7-11f, exhibited the highest variability among the responses recorded by the three installed strain gauges. Strain gauge 1 displayed rapid strain accumulation followed by a subsequent drop, while strain gauge 2 showed a declining response early in the loading process. These observations suggest that Mixture 60_HR_L_RA had the poorest cracking resistance among the tested mixtures. This behavior is likely attributed to continuous heavy rainfall during the early loading phase, which may have caused moisture-induced damage in Lane 4, Section C, where Mixture 60_HR_L_RA was tested.

7.4.1.2. Peak Residual Strain Monitoring

The development of peak $|\Delta\varepsilon|$ with each data collection 4-min window versus ESALs (presented in Figure 7-12) is used to complement peak ε_t monitoring, providing a comprehensive evaluation of pavement cracking performance in this study. The value of N_i in this study was defined as the point where two or more strain gauges consistently exhibit a peak residual strain $|\Delta\varepsilon|$ exceeding 5 $\mu\varepsilon$ for each mixture or single strain gauge exceeds residual strain $|\Delta\varepsilon|$ more than 15 $\mu\varepsilon$, which may present challenges in extracting ε_{c1} and ε_t . For mixture 30_C, as previously discussed, N_i is defined as the point at which the discrepancy in peak tensile strain between bi-directional passes at the same loading position exceeds 100 $\mu\varepsilon$. For mixture 30_O, shown in Figure 7-12b, and mixture 45_HR, shown in Figure 7-12c, N_i s were identified when gauges 1 & 2 continuously obtained $|\Delta\varepsilon|$ exceeding 5 $\mu\varepsilon$. For the mixture 45_HR_RA in Figure 7-12d, residual strain mostly remained at a relatively low level ($|\Delta\varepsilon| < 5 \mu\varepsilon$) throughout the entire loading period except gauge 3, which exhibited $|\Delta\varepsilon| > 5 \mu\varepsilon$ continuously at the end of the testing period. As a

result, N_i for this mixture was determined as the final accumulative ESALs. For mixture 45_HR_L, the N_i was identified when the $|\Delta\varepsilon|$ recorded by gauge 2 exceeded $15\ \mu\varepsilon$, shown in Figure 7-12e. Finally, for the mixture 60_HR_L_RA in Figure 7-12f, the residual strain for the different strain gauges rapidly accumulated to values greater than $5\ \mu\varepsilon$ at the beginning of the loading phase, and continued to increase throughout the test.

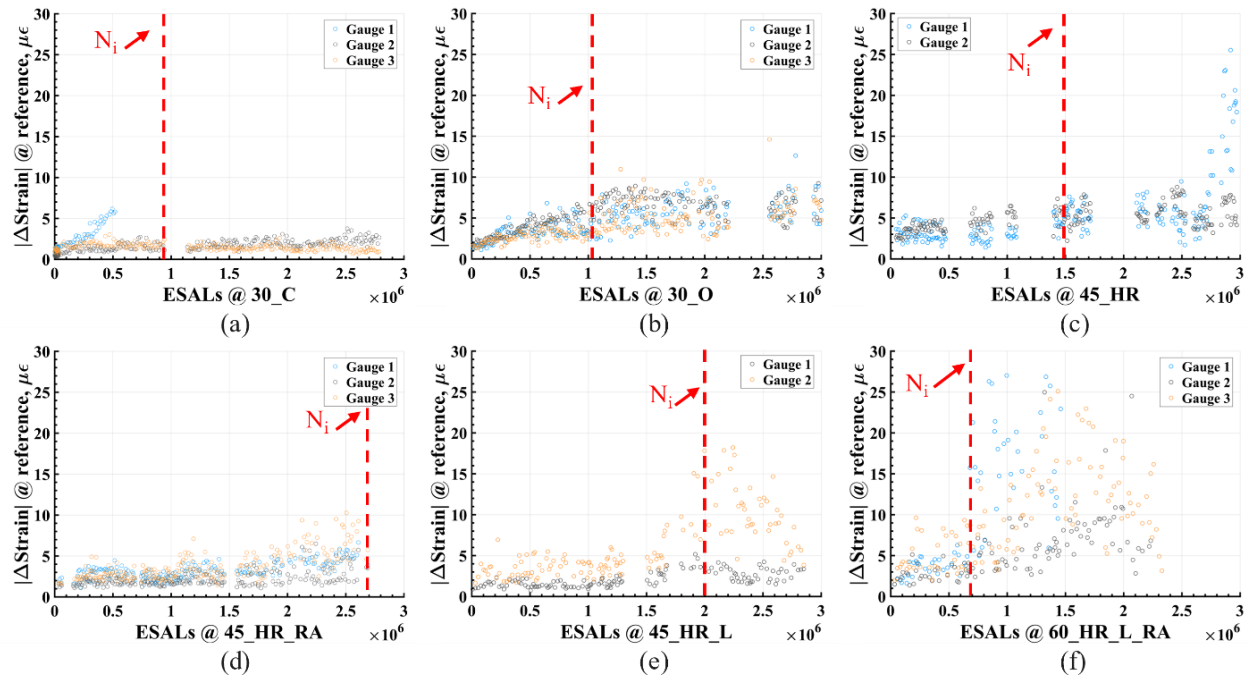


Figure 7-12. Peak residual strain monitored along with testing for different mixtures. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; N_i = the number of ESALs to cracking initiation.

The values of N_i are summarized in Table 7-1. Overall, the ranking of cracking resistance among the mixtures is as follows (ranked from highest cracking resistance to least): 45_HR_RA, 45_HR_L, 45_HR, 30_O, 30_C, and 60_HR_L_RA. However, further experimental work is required to better understand the influence of microcracking at different locations relative to strain gauges and the degree of crack development on strain gauge responses, which will be the focus of future studies.

Table 7-1. Summary of cracking performance measures with rankings for the evaluated mixtures.

Mixtures	30_C	30_O	45_HR	45_HR_RA	45_HR_L	60_HR_L_RA
N_i , ESALs	932,282	1,010,944	1,484,575	2,689,201	2,058,116	680,525
Ranking	E	D	C	A	B	F

Note: C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; N_i = the number of ESALs to cracking initiation; A to E = the order ranking of the mixtures from the best (A) to the least (F) performing.

7.4.1.3. Deformation Uniformity

Deformation uniformity for two-gauge mixtures is presented in Figure 7-13a, while three-gauge mixtures are illustrated in Figure 7-13b. Mixture 30_C is categorized as a two-signal mixture due to the inability to extract ε_t and ε_c from gauge 1 at approximately 0.5 million ESALs. A higher value indicates a higher non-uniformity of deformation. Using the 30% threshold recommended by AASHTO T378-17 for DMT as a data quality indicator, the strain gauge responses for all mixtures during the APT testing period exceeded this threshold except for 45_HR and 45_HR_RA. For the lower-performing mixtures, such as Mixture 30_C and Mixture 60_HR_L_RA, reliable data collection was observed within the first 279,671 and 87,627 ESALs, respectively.

Deformation uniformity indicates the similarity of local deformation values across the material, thereby providing insight into the uniformity of the material's local properties. When deformation uniformity exceeds a certain threshold, it may suggest the occurrence of localized stress concentrations, potentially leading to early cracking or excessive deformation in specific areas, which can eventually result in pavement failure. For localized points, the transition of deformation uniformity from below the threshold to exceeding the 30% threshold during repeated loading is likely associated with the development of microcracking. Using deformation uniformity as a metric to quantify cracking performance reveals that, for three-gauge mixtures, the ranking of cracking resistance is Mixture 45_HR_RA, Mixture 30_O, and Mixture 60_HR_L_RA. This ranking aligns with the conclusions drawn in the previous section using N_i as a quantification method. However, for two-gauge mixtures, reliable quantification is more challenging. The calculation of deformation uniformity for two-signal mixtures tends to exhibit significant instability, making it less reliable as a performance indicator (Kim et al., 2008).

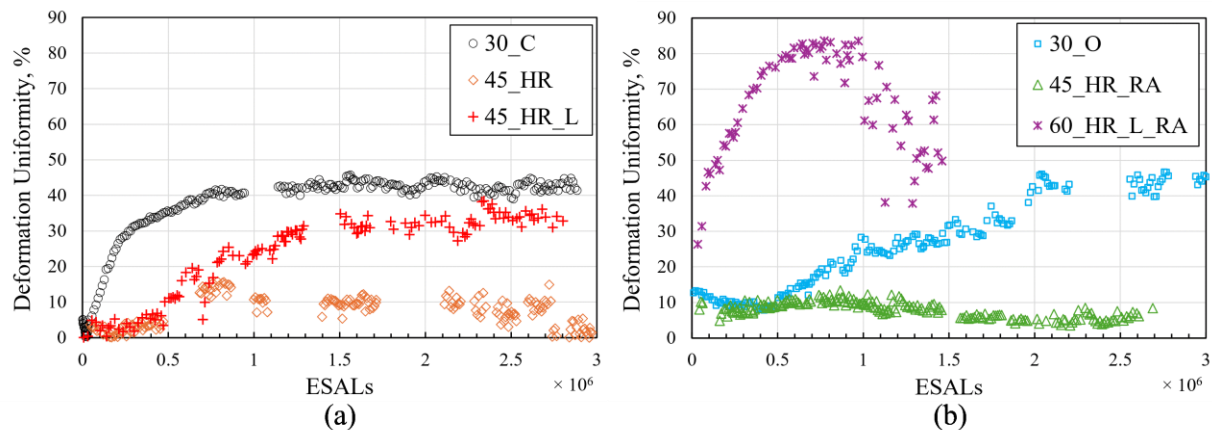


Figure 7-13. Deformation uniformity for (a) two-signal sections (b) three-signal sections. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder.

7.4.2. Cracking Evaluation

7.4.2.1. APT Field Survey

To assess cracking during the APT field survey, the research team referred to the standard practice outlined in ASTM D6433-18 for measuring the Pavement Condition Index (PCI). No interconnected cracks were observed on the pavement surface, except in Lane 4, Section C, for Mixture 60_HR_L_RA. Consequently, this study utilized crack length as the primary metric to evaluate pavement performance. The identified cracks are summarized in Table 7-2. Summary of identified surface crack lengths at different severity per mixture. It was observed that the majority of cracks in the testing sections were hairline cracks with low severity, exhibiting a crack width of approximately 0.1 mm. Some cracks were classified as medium severity, with crack widths ranging from 0.1 to 0.4 mm, while others were considered high severity, with crack widths exceeding 0.4 mm (primarily greater than 0.75 mm). These severity levels were defined relative to the cracking distress observed in Lane 4, Section C, where crack widths were classified as high severity. Currently, no standardized definition for crack severity in APT sections exists. For reference, according to ASTM D6433-18, most low-severity cracks are defined as having widths less than 10 mm. These measurements are intended solely as a relative comparison tool among the six evaluated mixtures, serving the specific purpose of establishing a performance ranking. Notably, this assessment does not delve into the extent of cracking failure or structural damage.

Table 7-2. Summary of identified surface crack lengths at different severity per mixture

Cracks	30_C	30_O	45_HR	45_HR_RA	45_HR_L	60_HR_RA_L
Low	20,152	14,831	10,105	3,790	6,974	2,765
Medium	0	360	1,800	0	1,800	8,295
High	0	180	300	0	0	44,243
Total	20,152	15,371	12,205	3,790	8,774	55,303

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

The total crack lengths for each tested section are shown in Figure 7-14. It was evident that all BMD mixtures, with one exception—mixture 60_HR_L_RA—outperformed the control mixture 30_C in terms of observed surface cracking. Notably, the 45% RAP mixtures had less observed surface cracking than the BMD-optimized 30% RAP mixture. Among the mixtures tested, the 45% RAP mixture containing a RA had the least amount of observed surface cracks. However, severe cracking distress was observed in Lane 4 Section C for Mixture 60_HR_L_RA. This distress could be related to one or more of the following potential causes: mixture production variability, lack of crack resistance of the mixture, construction uniformity issues, heavy rainfall during testing period, localized stress concentrations, and insufficient bonding between pavement layers. As a result, Lane 4, section D (shown in Figure 7-2a) was used as a backup section for testing. No cracking was observed before 228,753 ESAL, which performed much better than that of Section C. However, due to the lack of instrumentation data and shorter testing period, the

performance of Lane 4 section D was excluded from further discussion here. Overall, in terms of crack lengths, the mixtures were ranked (from least cracking to greatest) as follows: 45_HR_RA, 45_HR_L, 45_HR, 30_O, 30_C, and 60_HR_L_RA. This ranking aligns with the observations from the longitudinal tensile strain analysis shown previously, which demonstrated the effectiveness of strain gauges in monitoring pavement performance. The use of strain gauges enables the early detection of microcracking development, providing a means to predict cracking on the surface before it becomes visibly apparent.

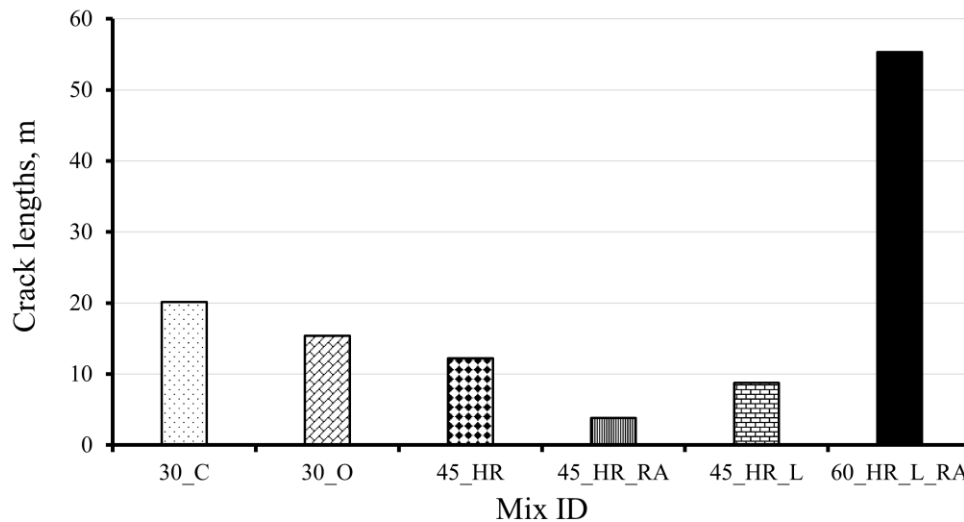


Figure 7-14. Field survey on cracking length. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; CT = cracking tolerance.

7.4.2.2. Laboratory Evaluation on Field Cores

The average permeability and standard deviation of cored replicates from the loaded and unloaded areas is presented in Figure 7-15. Compared to the VDOT maximum allowable permeability of 150×10^{-5} cm/s (Maupin, 2010), all field cores exhibited values significantly below this threshold. Nearly half of the cores demonstrated non-permeable characteristics, likely due to the lower designed air void content in the BMD mixtures compared to the control mixtures. Notably, two field cores from the loaded area of mixture 60_HR_L_RA showed complete cracking damage, making it impossible to conduct permeability test. This scenario was unique to mixture 60_HR_L_RA and was not observed in any of the other mixtures.

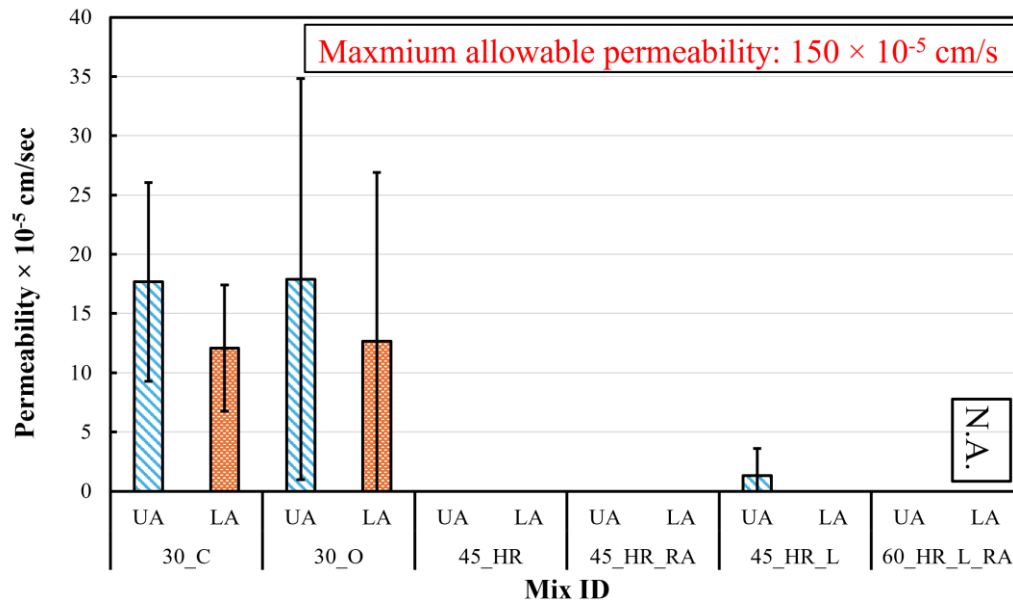


Figure 7-15. Permeability test results. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder. R = reheats; UA = unloading area; LA = loading area; N.A. = not available; I-bars indicate $\pm$ one standard deviation.

The dynamic modulus results at 20°C and 10 Hz are presented in Figure 7-16. In the figure, 'R' denotes reheated specimens, which were sampled as loose mixtures during construction, reheated, fabricated, and tested. A 20% or 50% reduction in stiffness is commonly used as a criterion to quantify material deterioration. However, upon comparison, no such stiffness reduction was observed in the tested specimens, either between the cores extracted from the loaded and unloaded areas or between the cores and reheated specimens. The only exception was mixture 60_HR_L_RA, where two field cores were already completely damaged at the time of extraction, rendering them incapable of sustaining load during testing; therefore, their modulus was assumed to be zero. The comparison between the modulus of field cores and reheated specimens varied among different mixtures, which can be attributed to differences in air void content and compaction effort due to traffic. Overall, these small-scale DMT test results together with the permeability test are consistent with previous findings, indicating that five of the evaluated mixtures did not experience structural damage or crack distress.

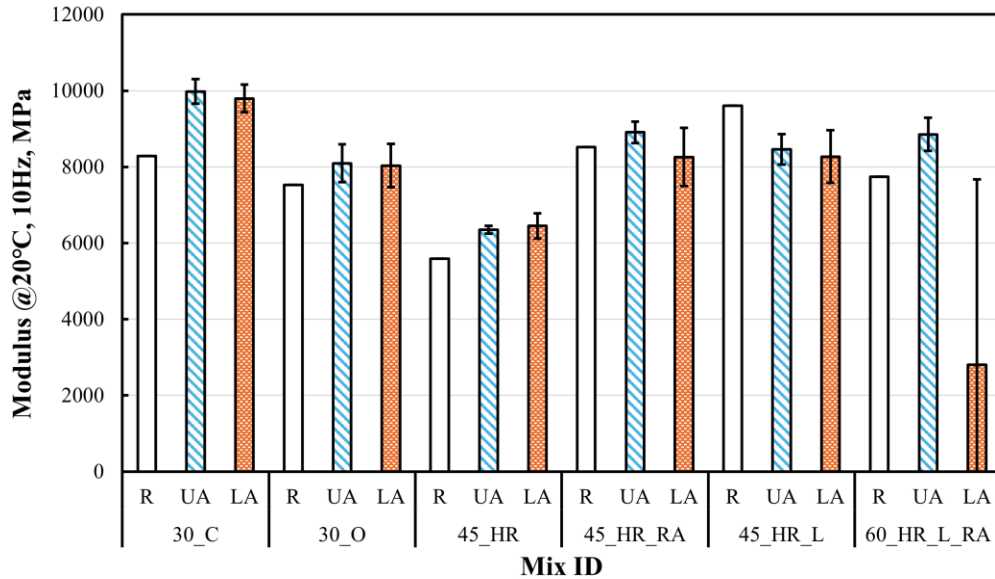


Figure 7-16. Dynamic modulus test results at 20°C and 10Hz. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder. R = reheats; UA = unloading area; LA = loading area; I-bars indicate $\pm$ one standard deviation.

7.5. Examining VDOT BMD Thresholds

This section aimed to evaluate and validate the current VDOT BMD thresholds for the CT Index and Cantabro Mass Loss (CML) by establishing correlations between APT results and laboratory test indices. The relevant laboratory tests have been extensively studied in prior research (0), and their detailed procedures are not reiterated in this chapter.

7.5.1. Indirect Tensile Cracking Test (IDT-CT)

IDT-CT, also known as IDEAL-CT, is a fundamental test commonly used by state agencies and was incorporated into the BMD approach for VDOT to evaluate mixture resistance to cracking. The IDT-CT was conducted in accordance with ASTM D8225-19. A higher CT index indicates a higher cracking resistance. The minimum limit of 70 was specified by the VDOT BMD special provision for specimens fabricated from reheated materials. IDT-CT was performed on two types of specimens in this study:

- Non-Reheated Specimens (non-reheats): laboratory-compacted non-reheated specimens that were compacted on-site by the producer staff using loose mixture sampled at the plant without reheating.
- Reheated Specimens (reheats): Laboratory-compacted reheated specimens that were compacted after reheating the loose mixture sampled at the plant. Mixtures were not subjected to any additional oven aging.

The testing was performed on non-reheated and reheated specimens that were sampled over four intervals during production. Tong et al. (2024b) reported no statistically significant

differences among the four samples for either the non-reheated or reheated specimens in terms of CT index. This finding supports the use of average values for analysis in this section.

7.5.2. Cantabro Test

Another BMD test adopted by VDOT is the Cantabro test. According to AASHTO T 401, the Cantabro test method is used to indirectly assess the cohesion, bonding, and effects of traffic abrasion on asphalt mixtures. Several researchers have recommended the Cantabro test as an effective method for durability assessment (Cox et al., 2017). CML was determined on the three gyratory specimens used to determine volumetric properties. Similar to the previous IDT-CT procedure, tests were conducted on both non-reheated and reheated specimens. A lower CML value indicates better durability performance. Tong et al. (2024a) reported significant high correlation between CML and CT index for the evaluated mixtures. Consequently, CML was incorporated in this section, and its relationship with the APT results was analyzed.

7.5.3. Laboratory Test Comparisons and Correlations

Table 7-3 summarizes the results of all testing and presents the ranking for each mixture (designated as A for best performing mixture through F for worst performing mixture) with respect to each test.

Table 7-3. Summary and ranking of different tests for the evaluated mixtures.

Measures	30 C	30 O	45 HR	45 HR RA	45 HR L	60 HR L RA
CT index non-reheats	138.8 (F)	242.2 (C)	943.7 (A)	342.2 (B)	220 (D)	219.7 (E)
CT index reheats	54.1 (F)	75.9 (E)	405.2 (A)	158 (B)	94.3 (D)	101.2 (C)
CML non-reheats	7.4 (F)	5.8 (E)	2.4 (A)	4.3 (B)	5.0 (D)	4.6 (C)
CML reheats	8.7 (F)	8.2 (E)	3.7 (A)	6.2 (C)	6.3 (D)	6.1 (B)
APT Rut Depth, mm	8.5 (B)	6.5 (A)	20.8 (F)	13.3 (D)	10.2 (C)	13.9 (E)

Note: C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; Ni = the number of ESALs to cracking initiation; CT = cracking tolerance; CML = Cantabro mass loss; reheats = reheated specimens; non-reheats = non-reheated specimens; A to E = the order ranking of the mixtures from the best (A) to the least (F) performing.

Figure 7-17a illustrates the relationship between the CT index of non-reheated and reheated specimens and APT crack lengths. No clear trend is evident initially; however, when excluding the APT results for mixture 45_HR and 60_HR_L_RA, a strong correlation between the CT index and crack lengths emerges for the remaining mixtures (shown in Figure 7-17b). The correlation coefficients between the CT index of non-reheated and reheated specimens and APT crack lengths reach -0.88 and -0.94, respectively. Mixture 45_HR was excluded because the binder content of the produced mixture was higher than the mix design binder content plus acceptable tolerance and resulted in an over-asphalted mixture having significantly lower air void contents (Tong et al.,

2024a). Mixture 60_HR_L_RA was excluded due to environmental factors affecting the APT results as discussed previously. Overall, the correlation exhibited in Figure 7-17b suggests that the IDT-CT, as adopted by the current VDOT BMD, has predictive capability for APT field pavement cracking performance. The current VDOT BMD thresholds for the CT index are 95 for non-reheated specimens and 70 for reheated specimens. Based on the linear regression relationship shown in Figure 7-17b, these thresholds correspond to APT field crack lengths of 22.7 m and 15.9 m, respectively, which fall within the range of acceptable cracking performance in the APT. Consequently, no structural damage or crack failures are anticipated, further supporting the validity of the current VDOT BMD thresholds for the CT Index.

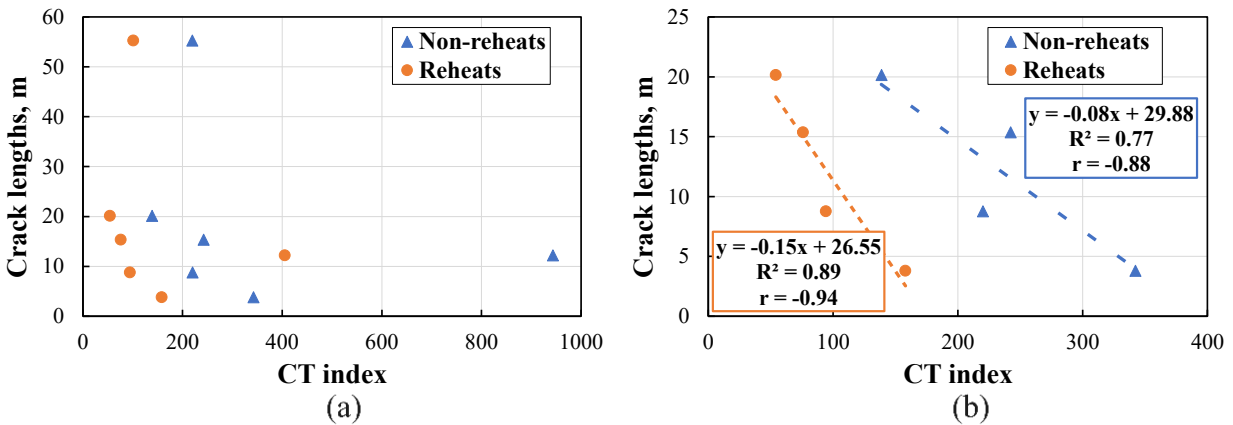


Figure 7-17. Correlation between CT index and APT crack lengths (a) for all evaluated mixture; (b) for mixtures excluding mixture 45_HR and mixture 60_HR_L_RA. R^2 = coefficient of determination; r = correlation coefficient.

Figure 7-18 illustrates the relationship between APT crack lengths and CML for all mixtures. The observed trend closely mirrors the IDT-CT results. After excluding the influence of mixture 45_HR and mixture 60_HR_L_RA, the CML of non-reheated and reheated specimens demonstrated strong correlations with APT crack lengths, with correlation coefficients of 0.97 and 0.96, respectively. The current VDOT CML threshold is 7.5% for reheated and non-reheated specimens, although a previous study suggested a threshold of 5.8% for non-reheated specimens (Tong et al., 2024b). Based on the regression established in Figure 7-18b, these thresholds correspond to crack lengths of 12.9 m and 12.8 m, respectively. Similar to the findings related to the IDT-CT analysis, the APT field pavement cracking performance for these mixtures—excluding mixture 45_HR and mixture 60_HR_L_RA—was deemed acceptable, as it did not result in structural damage. This further validates the effectiveness of the current VDOT BMD thresholds for CML.

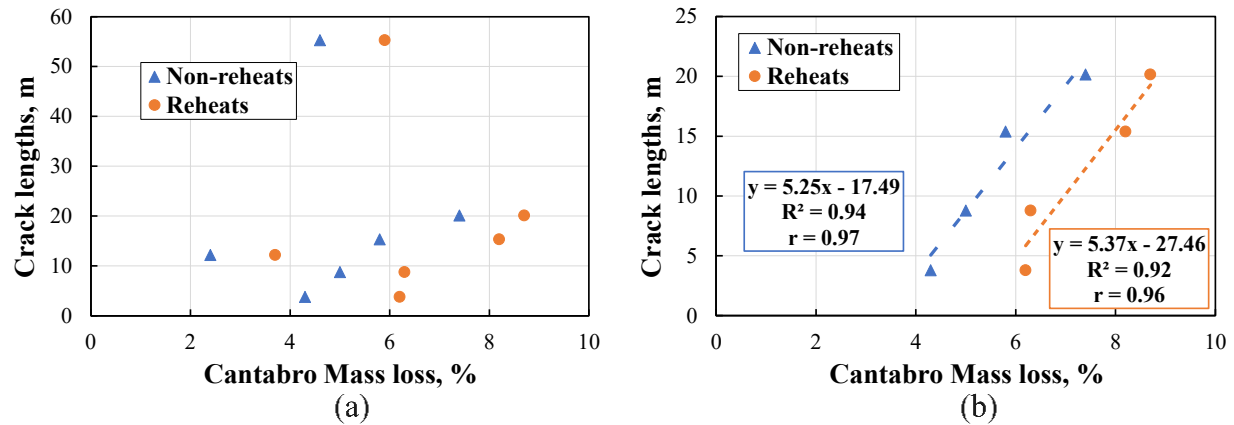


Figure 7-18. Correlation between Cantabro mass loss and APT crack lengths (a) for all evaluated mixture; (b) for mixtures excluding mixture 45_HR and mixture 60_HR_L_RA. R^2 = coefficient of determination; r = correlation coefficient.

7.6. Correlation with APT Rutting Studies

Another study of the Virginia 2020 BMD APT program focused on rutting performance evaluation, with detailed findings presented by Tong et al. (2023). The rutting performance of each mixture was assessed using the normalized APT rut depth, as summarized in **Error! Reference source not found.** Analyzing the correlation between cracking lengths from the APT cracking study sections and the rut depth from the APT rutting study sections (as shown in Figure 7-19) provides insights into the balance of primary performance characteristics, including cracking and rutting, for the designed mixtures.

Figure 7-19a reveals that not all mixtures exhibit a balanced performance. Specifically, mixture 45_HR and mixture 60_HR_L_RA are notably positioned away from the other mixtures. For mixture 45_HR, the higher asphalt content compared to other mixtures resulted in a compromised rutting performance, reaching rut depth of 20.8 mm, without achieving the highest cracking performance. In the case of mixture 60_HR_L_RA, continuous heavy rainfall led to fatigue failure, with the crack length significantly exceeding that of other mixtures. In Figure 7-19b, after excluding mixtures 45_HR and 60_HR_L_RA, a clearer trend emerges among the remaining mixtures: an improved cracking performance often coincides with a compromise in rutting performance. The correlation between rut depth and crack length is significant, with a Pearson correlation coefficient of -0.83. Among the BMD mixtures, including 30_O, 45_HR_L, and 45_HR_RA, a perfect linear correlation is observed, with a Pearson correlation coefficient equal to -1. As discussed, the evaluated mixtures did not exhibit significant signs of fatigue failure, except for mixture 60_HR_L_RA. In these tested sections, surface cracks did not connect or form noticeable crack areas, indicating that the current BMD thresholds provide sufficient margins to ensure satisfactory field pavement cracking performance. However, for rutting performance, the 12.5 mm rut depth (Greene et al., 2015)—a typical threshold for rutting failure in pavements—may pose a concern for BMD mixtures. This suggests that further refinement of BMD thresholds, particularly concerning rutting performance, may be considered. Alternatively, greater attention

should be directed toward mixture design to prevent excessive asphalt content. The incorporation of HRAP necessitates careful consideration of factors such as the addition of RA, the use of softer binders, and increased binder content when applying BMD concept.

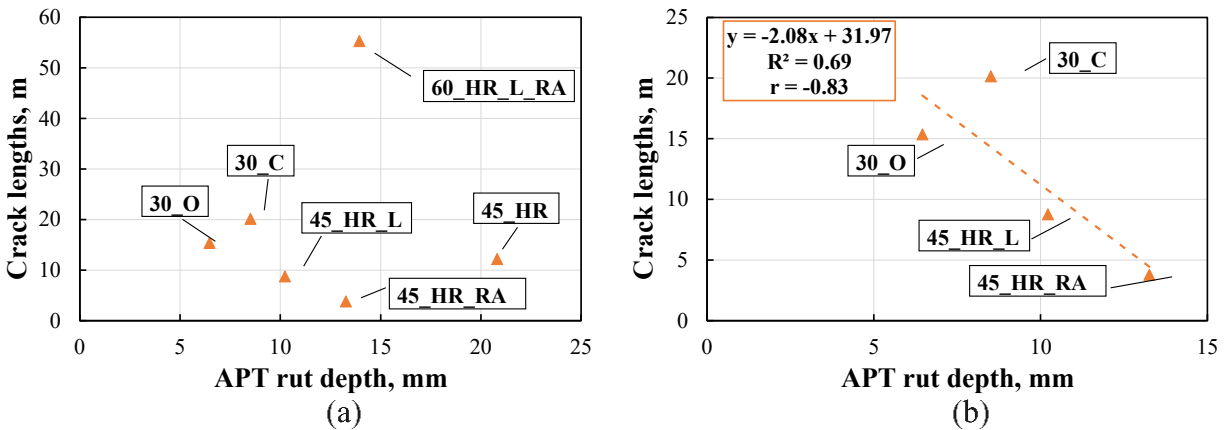


Figure 7-19. Correlation between APT crack lengths and APT rut depth. C = control; O = optimized; HR = high reclaimed asphalt pavement; RA = recycling agent; L = softer virgin binder; CT = cracking tolerance.

7.7. Summary of Findings, Conclusions, and Future Work

This research presents an evaluation of the cracking performance of a control mixture and five BMD-optimized surface mixtures through laboratory and full-scale testing at the Virginia APT facility. These mixtures included various combinations of RAP content, two binder performance grades, a single RA, and a WMA additive. A comprehensive analysis was carried out, integrating longitudinal tensile strain monitoring, APT field surveys, and a series of laboratory performance evaluations to rank the mixtures. The primary outcomes are summarized below:

(1) This paper offered a detailed summary of the primary response phases observed in longitudinal tensile strain development under APT. It also illustrated that strain gauges embedded at the bottom of the asphalt surface layer may demonstrate varying behaviors, potentially due to the development of microcracking in localized regions. Additionally, this study proposed a method for monitoring residual strain to quantify the potential number of ESALs to the cracking initiation point. The application of this method aligns well with field survey results in ranking the mixtures consistently.

(2) Deformation uniformity can serve as a data quality indicator for the collected instrument data and may provide a quantification method for evaluating cracking performance. However, its application was challenging for those sections where only two strain responses were operational throughout the testing period.

(3) The APT survey revealed that most cracks were low-severity (crack width = 0.1 mm), with no interconnected cracks forming cracking areas, except in the mixture 60_HR_L_RA section. Laboratory assessment of field cores, including permeability tests and small scale DMT, further confirmed that no structural damage or cracking distress was observed in any of the evaluated mixtures, except for mixture 60_HR_L_RA.

(4) Crack lengths for each APT section were measured and used as a metric to evaluate cracking performance. The ranking from best to worst in terms of crack length is as follows: 45_HR_RA, 45_HR_L, 45_HR, 30_O, 30_C, and 60_HR_L_RA. All BMD mixtures, except 60_HR_L_RA, demonstrated superior crack length performance compared to the control mixture. This supported the effectiveness of using the BMD concept, particularly with the incorporation of softer binders and/or recycling agents.

(5) Through correlation analysis, the gap between APT performance and the BMD tests utilized by VDOT, including the IDT-CT and Cantabro tests, can be bridged. This analysis examined the current VDOT BMD thresholds for CT index and CML.

(6) The combination of APT rutting and cracking performance data offered valuable insights into the balance of primary performance characteristics for the evaluated mixtures. Except for mixtures 45_HR and 60_HR_L_RA, the other mixtures demonstrated a linear relationship, where improved cracking resistance often coincided with a trade-off in rutting performance. The investigation also suggested that the current BMD thresholds provide sufficient margins to ensure satisfactory pavement cracking performance in the APT, though rutting resistance may remain a concern for over-asphalted BMD mixtures.

(7) Further experimental work is necessary to deepen the understanding of how microcracking at different locations relative to strain gauges affects strain gauge responses. This topic will be the focus of future investigations.

The conclusions are constrained by the limited scope of the evaluated mixtures. Further validation necessitates additional assessments encompassing a broader range of mixtures and more field performance data.

7.8. Reference

- American Association of State Highway and Transportation Officials. (2019). *AASHTO TP 132-19: Determining the Dynamic Modulus for Asphalt Mixtures Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT)*. Washington, D.C.
- American Association of State Highway and Transportation Officials. (2022). *AASHTO T 401-22: Cantabro Abrasion Loss of Asphalt Mixture Specimens*. Washington, D.C, 2022.
- ASTM International. (2019). *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.
- Cox, B. C., Smith, B. T., Howard, I. L., & James, R. S. (2017). State of Knowledge for Cantabro Testing of Dense Graded Asphalt. *Journal of Materials in Civil Engineering*, 29(10), 04017174.
- Elseifi, M. A., Mohammad, L. N., & Zhang, Z. (2012). Assessment of Stress and Strain Instrumentation in Accelerated-Pavement Testing. *International Journal of Pavement Research and Technology*, 5(2), 121.
- Greene, J., Chun, S., Nash, T., & Choubane, B. (2015). Evaluation and Implementation of PG

- 76-22 Asphalt Rubber Binder in Florida. *Transportation Research Record*, 2524(1), 3-10.
- Harvey, J., & Popescu, L. (2000). Accelerated Pavement Testing of Rutting Performance of Two Caltrans Overlay Strategies. *Transportation Research Record*, 1716(1), 116-125.
- Kawa, I., Zhang, Z., & Hudson, W. R. (1998). *Evaluation of the AASHTO 18-Kip Load Equivalency Concept* (Vol. 7). Austin, TX, USA: Center for Transportation Research, Bureau of Engineering Research, University of Texas at Austin.
- Kim, W., Labuz, J., Chadbourn, B., & Loken, M. (2008). Uniformity of Axial Displacement in Element Testing. *Geotechnical Testing Journal*, 31(3), 274-278.
- Long, F., Harvey, J., Scheffy, C., & Monismith, C. L. (1996). Prediction of Pavement Fatigue for California Department of Transportation Accelerated Pavement Testing Program Drained and Undrained Test Sections. *Transportation research record*, 1540(1), 105-114.
- Maupin Jr., G. W. (2010). Examination of an Implemented Asphalt Permeability Specification (No. VTRC 11-R7), Virginia Transportation Research Council, Charlottesville.
- Ozer, H., Al-Qadi, I. L., Singhvi, P., Bausano, J., Carvalho, R., Li, X., & Gibson, N. (2018). Prediction of Pavement Fatigue Cracking at an Accelerated Testing Section Using Asphalt Mixture Performance Tests. *International Journal of Pavement Engineering*, 19(3), 264-278.
- Steyn, W. J. (2012). *Significant Findings from Full-Scale Accelerated Pavement Testing* (Vol. 433). Transportation Research Board.
- Tong, B., Habbouche, J., Diefenderfer, S. D., & Flintsch, G. W. (2024a). Multi-Level Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents: A Case Study in Virginia. *International Journal of Pavement Engineering*, 25(1), 2325609.
- Tong, B., Habbouche, J., Diefenderfer, S. D., Flintsch, G. W., & Boz, I. (2024b). Assessment of the Production Variability and Composite Performance Index for Conventional and High Reclaimed Asphalt Pavement Balanced Mix Design Mixtures. *Transportation Research Record*, 03611981241287785.
- Tong, B., Habbouche, J., Flintsch, G. W., & Diefenderfer, B. K. (2023). Rutting Performance Evaluation of BMD Surface Mixtures with Conventional and High RAP Contents under Full-Scale Accelerated Testing. *Materials*, 16(24), 7611.
- Tsai, B. W., Harvey, J. T., & Monismith, C. L. (2003). The Application of Weibull Theory in Asphalt Concrete Fatigue Performance Prediction. In *Transportation Research Board Annual Meeting Presentation*, Washington, DC.

CHAPTER 8 - Summary, Conclusions, and Recommendations

8.1. Overview

This dissertation assessed the application of the BMD concept in designing HRAP SMs (RAP contents exceeding 30%) in Virginia, offering optimized methodologies and insights. It included a comprehensive evaluation of BMD mixtures, specifically examining one control mixture and five dense-graded BMD SMs designed according to the current VDOT BMD special provision. The study incorporated a multi-level laboratory performance assessment, followed by the VDOT-adopted BMD tests with their associated variability analysis. To address the challenges in analyzing and comparing multiple BMD performance metrics, this dissertation introduced the CPIs to evaluate the composite performance for asphalt mixture. To tackle the difficulties encountered during both the design and production phases of BMD, this research proposed the integration of ML techniques to develop an effective ML-BMD framework. Furthermore, an in-depth analysis of pavement response and performance data from an APT experiment was conducted. Correlation analyses were performed to bridge various evaluation scales of performance tests. The study examined the performance-based specifications related to VDOT BMD tests and their associated thresholds.

8.2. Summary of Primary Findings

The primary findings of this dissertation are categorized into four key areas: The impact of BMD on volumetric properties and aggregate gradations, laboratory BMD performance evaluation, APT experiment and their correlation with laboratory performance tests, and the application of ML in BMD. Details are provided as follows.

8.2.1. BMD Impact on Volumetric Properties and Gradations

- The optimum binder content in all BMD mixtures exceeded that of the control mixture, indicating that the BMD process may require additional binder for certain mixtures.
- Mixtures containing a RA component and/or a softer binder (i.e., 45_HR_RA, 45_HR_L, and 60_HR_RA_L) exhibited lower optimum binder contents than mixtures without these components (e.g., 45_HR).
- The binder content of the 45_HR mixture was 6.8%, which was much higher than expected/typical, which may result in an over-asphalted characteristic during production and performance.
- The voids filled with asphalt (VFA) for mixtures 30_O, 45_HR_RA, and 60_HR_L_RA were 87.0, 84.0, and 90.0%, respectively, much higher when compared to the control mixture (30_C).
- Precision in plant control over mixture gradation was demonstrated, leading to minimal variability among field samples.

- The combined effect of variations in different volumetric properties including VTM and VFA during production may influence the primary performance of the mixtures.

8.2.2. Laboratory Evaluation

- Multi-level laboratory tests revealed that the ranking order differed across different tests. For instance, the CT index (basic level) and S_{app} (advanced level) tests did not yield the same ranking order in cracking evaluation.
- That different levels of rutting performance tests produced relatively consistent results, as reflected by the strong positive linear correlation found between RSI and APA.
- Rutting resistance showed a negative correlation with durability. Based on S_{app} , fatigue cracking resistance showed a negative correlation with durability, while based on CT index and G-Rm, cracking resistance showed a positive correlation with durability.
- Compared with mass-controlled volumetric pills used for the Cantabro test, compacting IDT-CT and APA specimens at VTM of $7 \pm 0.5\%$ exhibited less statistically significant difference between designs and produced specimens and among the different samples during production.
- The produced HRAP BMD mixtures complied with the prescribed VDOT BMD special provisions in terms of CML and CT index. Items that failed to meet the corresponding specifications were limited to the 30% RAP mixtures, with the control mixture exhibiting the most pronounced deviation.
- All mixtures except mixture 45_HR, complied with current APA rut depth threshold of 8 mm.
- The aging effect introduced during the reheating process may compromise the durability and cracking resistance of the mixtures as expected.
- 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.
- The designed BMD mixture could become unbalanced due to production variability.

8.2.3. Machine Learning Application in BMD

- The robustness of the ML models was evaluated and validated through 30 independent training/testing dataset splits in a random manner.
- Volumetric properties and gradations could serve as factors for developing ML models to address challenges encountered in the design and production phases of BMD.
- Among the ML models, RF, XGB, and SVR exhibited superior performance compared to LR, effectively capturing the complex interactions within the mixture properties.
- ML models reliably delivered accurate predictions across all BMD tests, this was corroborated by the identification of significant influencing factors such as Type of specimens, AC, $\log|G^*|$, 75 μm sieve size passing percentage, and VTM. The impact of the

reheating process on mixtures' durability and cracking performance was also identified, supported by previous studies.

- The ML models faced challenges in achieving stable R^2 values in the APA RD test, likely due to limited data availability and the use of average values from APA specimens.
- ML models can be integrated in the current BMD framework and have the potential to improve multiple BMD performance indices.

8.2.4. APT Experiment

- All BMD HRAP mixtures exhibited higher rut depths compared to the control mixture 30_C. This can be attributed to the higher binder content in BMD HRAP mixtures, which was used to enhance cracking resistance but compromised rutting resistance. Notably, mixture 45_HR showed a high likelihood of reaching the rutting failure even under a loading of 9,000 lbf. Nevertheless, when considering the aging effect, BMD optimized mixture 30_O displayed better rutting performance compared to the control mixture.
- The empirical rutting model proposed in this study effectively accounted for the observed environment aging effect and facilitated the normalization of measured rut depth.
- Significant and strong positive correlations between APT measured rut depth and multi-level laboratory rutting performance tests. Remarkably, the highest positive correlation ($r = 0.98$) was observed between the APA average value of reheated specimens and the APT experiment.
- Only low-severity cracks (crack width < 1 mm) were detected during the APT experiment, with no interconnected cracks forming larger cracking areas, except in the section of mixture 60_HR_L_RA.
- Laboratory assessments of field cores, including permeability tests and small specimen DMT, confirmed that fatigue cracking failure was not observed in any of the evaluated mixtures, except for mixture 60_HR_L_RA.
- In terms of cracking performance, based on measured crack lengths, the mixture ranking (A-F) is as follows: 45_HR_RA, 45_HR_L, 45_HR, 30_O, 30_C, and 60_HR_L_RA.
- All BMD mixtures, except 60_HR_L_RA, demonstrated superior cracking resistance compared to the control mixture, with HRAP BMD mixtures outperforming those with conventional RAP contents.
- Except for mixtures 45_HR and 60_HR_L_RA, the other mixtures demonstrated a linear balance relationship, where improved cracking resistance often coincided with a trade-off in rutting performance.
- The current BMD thresholds provide sufficient margins to ensure satisfactory pavement cracking performance in the field, though rutting resistance may remain a concern for BMD mixtures.

8.3. Conclusions

This dissertation successfully achieved its primary objectives by evaluating the application of the Balanced Mix Design (BMD) concept for designing durable surface mixtures in Virginia, with a particular emphasis on mixtures containing relatively higher RAP contents (exceeding 30%). The specific conclusions corresponding to each main task are summarized as follows:

- The multi-level performance tests revealed discrepancies between basic/intermediate performance test results and advanced performance test results. The comparisons conducted highlighted the significant benefits of incorporating softer binders and/or recycling agents in high RAP (HRAP) mixtures. Additionally, the findings demonstrated that the BMD approach provides an effective framework for designing asphalt mixtures capable of simultaneously improving fatigue cracking resistance and rutting performance. Furthermore, the study showed that HRAP BMD surface mixtures can be optimized to achieve superior overall performance compared to conventionally designed control mixtures.
- The BMD variability study revealed relatively precise gradation control; however, exceedances of volumetric property tolerances were observed in BMD-optimized mixtures during production. Performance metrics, including durability, cracking resistance, and rutting susceptibility, were assessed using the Cantabro test, the IDT-CT, and the APA test, respectively. The test results indicated that the produced mixtures could become unbalanced in performance. The study emphasized both the necessity and the effectiveness of adopting the BMD approach for high RAP (HRAP) mixtures. However, the potential aging effects introduced during the reheating process may negatively impact durability and cracking resistance, highlighting the need for careful consideration of production practices.
- Regarding the optimization methods. This dissertation proposed a novel index CPI associated with a 3D-plot that can be used to evaluate the relative composite performance of different mixtures and a ternary plot to assist in evaluating the contribution of individual major performance results in BMD optimized mixtures. The use of 3-D plots provided a more intuitive method for evaluating the balanced status among multiple performance indices of produced mixtures. Additionally, a CPI_R was introduced to assess the composite performance of the produced mixture -durability, cracking, and rutting, facilitating the multiple BMD performance indices evaluation. Furthermore, ML models can address challenges prevalent in the existing BMD framework, thereby enhancing its implementation during both the design and production phases. Moreover, the integrating ML models provided avenues for simultaneously improving multiple BMD performance indices.
- Through APT rutting study, this research provided a steppingstone towards the validation of the VDOT BMD APA rut depth threshold of 8mm threshold based on actual APT rut depth and recommended RSI value. A revised threshold was also proposed. The APT cracking study demonstrated that all BMD mixtures exhibited superior cracking resistance compared to the control mixture, with HRAP BMD mixtures outperforming those containing

conventional RAP contents. These findings reinforced the effectiveness of the BMD approach, particularly when incorporating softer binders and recycling agents. The current CT index and CML thresholds validated through the APT cracking performance. The study further revealed that while current BMD thresholds provide adequate margins to ensure satisfactory field cracking performance, rutting resistance may become a concern for over-asphalted BMD mixtures, highlighting the need for balanced performance considerations.

8.4. Significance

This research demonstrated that the performance of higher RAP content mixtures can exceed that of lower RAP content mixtures through the application of BMD approaches. The research introduced composite evaluation methods that significantly enhance the comparison and analysis of multiple performance metrics, addressing the limitations of traditional comparative approaches. Lessons learned and experiences gained from this research provide valuable insights and references for laboratory testing, APT evaluations, and even field performance data collection and interpretation. Furthermore, the integration of cutting-edge machine learning (ML) techniques into mixture design represented an innovative endeavor, creating a robust platform avenue for simultaneously improving multiple performance attributes of asphalt mixtures. This dissertation not only advanced the implementation of BMD for surface mixtures but also paved the way for sustainable and performance-driven practices in pavement engineering. These contributions offered guidance and practical recommendations for the design, production, construction, and performance monitoring for HRAP asphalt Mixtures. Moreover, the findings are expected to shape future research and industry practices, promoting innovation and sustainability in pavement engineering.

8.5. Recommendations

8.5.1. Implementation Recommendations

1. Transportation agencies should continue the efforts toward the full implementation of BMD for SMs and establish corresponding performance thresholds for selected BMD tests, adapted to reflect local practices and interests.
2. The transportation agencies should consider incorporating additional tools, beyond simply increasing binder content, such as RAs and/or softer binders, to optimize the design and production of BMD SMs.
3. Mixtures containing up to 45% RAP can achieve satisfactory performance when properly controlled and, where desired, designed using current BMD specifications modified to include tools like RAs and/or softer binders. However, the use of higher RAP content mixtures is not suitable for all locations or circumstances. It should be pursued in areas where sufficient RAP quantities are available, producers have demonstrated expertise, approved RAP management practices are in place, and production facilities are equipped to handle such mixtures.

4. Transportation agencies can use the CPI as part of their acceptance program for BMD mixtures to determine a pay factor for possible bonuses or penalties.
5. With the extensive database of BMD test results available, transportation agencies should consider integrating ML models to optimize the design and production processes for BMD.

8.5.2. Recommendations for Future Research

1. Further research is necessary to gain a deeper understanding of how crack development impacts strain gauge responses.
2. Laboratory testing is recommended to evaluate the effectiveness of ML models in simultaneously enhancing multiple performance indices of asphalt mixtures.
3. Laboratory testing is recommended to investigate the impact of various aging conditions on BMD mixture performance.
4. Further sampling and testing, using a broader range of mixtures, can be conducted to address discrepancies observed in multi-level performance tests. These efforts should be complemented by validation through APT and field performance evaluations.
5. Future monitoring of field performance is essential to validate the application of the BMD concept in designing HRAP SMs.

Appendix: Chapter 3 – Statistic Analysis

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. ANOVA with Post-hoc Analysis (Tukey's HSD Test)

Group-1	Group-2	P_value					
		30_C	30_O	45_HR	45_HR_RA	45_HR_L	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- 4. Nonparametric (Kruskal-Wallis) Test Results

Group-1	Group-2	Adj. Sig.												
		30_C		30_O		45_HR		45_HR_RA		45_HR_L		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- 5. ANOVA with Post-hoc Analysis (Tukey's HSD Test) Results for Produced Samples

Group-1	Group-2	P_value											
		30_C		30_O		45_HR		45_HR_RA		45_HR_L		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- 6. Nonparametric (Kruskal-Wallis) Test with Pairwise Comparisons for Produced Samples

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