

2 CRACKING GROUP EXPERIMENT: VALIDATION OF TOP-DOWN CRACKING TESTS FOR BALANCED MIX DESIGN

2.1 Background

As interest in balanced mix design (BMD) began to grow six years ago, the FHWA and state departments of transportation in Alabama, Florida, Illinois, Maryland, Michigan, Minnesota, Mississippi, New York, North Carolina, Oklahoma, and Wisconsin worked with NCAT and MnROAD to develop an experiment to evaluate numerous cracking tests. The primary objective of the experiment was to determine which laboratory cracking tests had the best correlation with field performance. Two complimentary experiments were planned and built; the experiment at the NCAT Test Track focused on validating tests for top-down cracking, and the experiment at MnROAD focused on validating tests for thermal cracking. This chapter describes the findings of the NCAT Test Track top-down cracking experiment. The findings of the MnROAD thermal cracking test validation will be provided later in a separate report.

The NCAT Test Track Cracking Group experiment included seven test sections, each with a different surface mix. The seven mixtures were intentionally designed to yield a range of field top-down cracking performance. Six of the seven mixtures (i.e. N1, N2, N5, N8, S5, and S6) were designed in accordance with the conventional Superpave requirements of AASHTO M 323 using an N_{design} of 80 gyrations. The mix design for Section S13 was unique in that it was a gap-graded, asphalt rubber mixture designed using the Marshall method using 75 blows per side. Table 1 summarizes the general mix descriptions, virgin binders, recycled materials contents, and continuous grades of the extracted and recovered binders from plant mix samples. For Section S5, the requested binder grade was a PG 58-28 to use with the higher RAP content mixture. However, the “softer” binder supplied for the section actually graded as a PG 64-28, which was verified to contain polymer modification.

Table 1. Summary of Surface Mixtures Used in the NCAT Top-Down Cracking Experiment

Test Track Section	Mixture Description	NMAS ^a (mm)	Virgin Binder Grade	RAP Content	RAS Content	Recovered Binder Cont. Grade
N1	Control (20% RAP)	9.5	PG 67 -22	20%	0%	88.6 -16.6
N2	Control, Higher Density	9.5	PG 67 -22	20%	0%	89.9 -15.9
N5	Control, Low Density, Low AC ^b	9.5	PG 67 -22	20%	0%	88.0 -18.5
N8	Control + 5% RAS	9.5	PG 67 -22	20%	5%	107.3 -5.4
S5	35% RAP, PG 58-28	9.5	PG 64 -28	35%	0%	82.8 -23.0
S6	Control, HiMA ^c Binder	9.5	PG 94 -28	20%	0%	101.4 -21.5
S13	Gap-graded, asphalt-rubber	12.5	Not tested	15%	0%	Not tested

^a Nominal Maximum Aggregate Size; ^b asphalt content; ^c Highly Modified Asphalt

The test sections were built relatively thin for the heavy loading on the Test Track so that the surface layers would experience significant deflections. To avoid bottom-up fatigue cracking, the intermediate and base layers contained the same highly modified binder used in S6. The same mix design was used for the lower two layers; it was a 19.0 mm NMAS Superpave mix containing 17% RAP (19% RAP binder ratio) with an N_{design} of 60 gyrations. The surface mixtures were constructed as a 1.5-inch lift over the highly polymer-modified intermediate and base layers, which were 2.25 inches each.

The as-constructed cross-sections of the test sections are illustrated in Figure 1. Some variations in thicknesses of the layers were identified from construction surveys and later verified with cores.

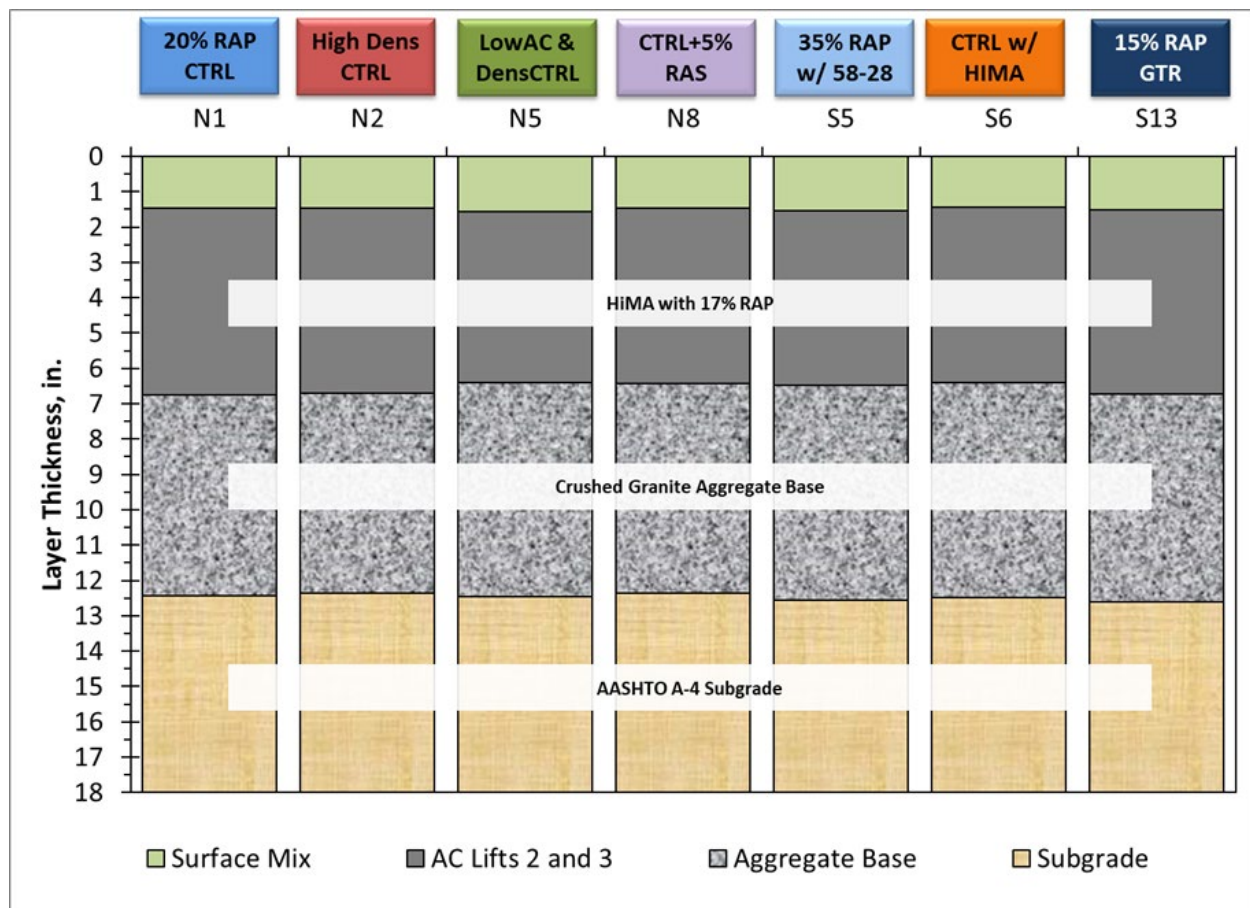


Figure 1. Cross-section of Cracking Group Test Sections on the NCAT Test Track

2.2 As-Constructed Mixture Properties

Table 2 summarizes traditional quality control results for the surface layer mixtures. As noted previously, all of the mixtures except for S13 were dense-graded Superpave mixtures that would be classified as fine-graded by AASHTO M 323. The gap-graded, asphalt-rubber surface mixture for S13 was unique, with a binder content at least 1.6% higher than each of the Superpave mixtures. The as-constructed density levels for two of the sections (N2 and N5) were also intentionally variable for the experiment. Section N2 was compacted to a higher relative density target of 96%, and N5 was constructed to a lower relative density target of 90%. The other six sections had a target density of 92 to 93% of their respective maximum theoretical densities.

Table 2. Traditional Quality Control Properties of the NCAT Top-Down Cracking Group Experiment Test Sections

	N1 Control	N2 Control w/ High Density	N5 Control w/ Low Dens. & AC	N8 Control w/ 5% RAS	S5 35% RAP PG 58-28	S6 Control w/ HiMA	S13 Gap-Graded Asphalt- Rubber*
Sieve Size							
12.5 mm (1/2")	99	100	100	99	99	100	96
9.5 mm (3/8")	97	98	99	98	96	98	85
4.75 mm (#4)	67	70	73	66	73	67	35
2.36 mm (#8)	52	54	54	51	56	52	22
1.18 mm (#16)	41	43	42	41	44	42	19
0.60 mm (#30)	28	28	28	30	29	28	14
0.30 mm (#50)	15	15	15	17	16	15	8
0.15 mm (#100)	9	9	9	11	10	9	5
0.075 mm (#200)	5.4	5.6	5.7	7.1	6.3	5.4	3.6
Total Binder Content (P_b)	5.4	5.4	5.1	5.3	5.8	5.8	7.4
Eff. Binder Content (P_{be})	4.7	4.7	4.4	4.8	5.1	5.0	6.6
RAP Binder Ratio	0.18	0.18	0.19	0.18	0.29	0.17	0.08
RAS Binder Ratio	--	--	--	0.19	--	--	--
Dust/Binder Ratio	1.1	1.2	1.3	1.5	1.2	1.1	0.5
Rice Sp. Gravity (G_{mm})	2.469	2.468	2.478	2.492	2.472	2.459	2.402
Avg. Bulk Sp. Gravity (G_{mb})	2.375	2.372	2.348	2.415	2.393	2.384	2.319
Lab Compaction Temp.	290°F	290°F	290°F	290°F	285°F	325°F	350°F
Air Voids (V_a)	3.8	3.9	5.3	3.1	3.2	3.1	3.4
Agg. Bulk Gravity (G_{sb})	2.634	2.631	2.633	2.672	2.656	2.634	2.631
Avg. VMA	14.7	14.7	15.4	14.4	15.1	14.7	18.4
Avg. VFA	74	73	66	79	79	79	81
Mat Density (% G_{mm})	93.6	96.1	90.3	91.5	92.2	91.8	92.7

*75-blow Marshall hammer compaction used for mix design and QC.

2.3 Laboratory Testing Plan

The five laboratory cracking tests initially selected by the sponsors of the NCAT Test Track Cracking Group Experiment were the energy ratio (ER), the Texas overlay (OT-TX) test, the NCAT modified overlay test (OT-NCAT), the Louisiana semi-circular bend test (SCB-LA), and the Illinois flexibility index test (I-FIT). The Indirect Tensile Cracking Test (i.e., IDEAL-CT) was added to the experimental plan after the experiment was under way based on discussions with the sponsors. The Asphalt Mixture Performance Tester (AMPT) cyclic fatigue test was also added to the experiment later in the project. Table 3 summarizes the cracking tests conducted in the experiment. Further details of the tests were provided in the end-of-cycle report from the 2015 Test Track, (1) and Chen (2).

Table 3. Top-Down Cracking Tests Analyzed in this Experiment

Test	Test Method	Primary Output	Key Reference
Energy Ratio	No standard	ER	Roque et al. (3)
Texas Overlay Test	Tex-248-F	β	Garcia et al. (4,5)
NCAT Overlay Test	Tex-248-F*	β	Ma (6), Garcia et al. (4)
Louisiana SCB	ASTM D8044-16	J_c	Cooper III et al. (7)
Illinois Flexibility Index Test	AASHTO TP 124	FI	Ozer et al. (8)
IDEAL Cracking Test	ASTM D8225-19	CT Index	Zhou et al. (9)
AMPT Cyclic Fatigue	AASHTO TP 133-19	S_{app}	Wang et al. (10)

*With change of frequency, gap, and definition of failure

Each of the cracking tests, except cyclic fatigue, were conducted on samples prepared and conditioned in four ways:

1. Lab-mixed, lab-compacted (LMLC) specimens after short-term oven aging (STOA) according to AASHTO R 30, abbreviated as LMLC-STOA.
2. LMLC-STOA plus an additional eight hours of aging at 135°C in a loose mix condition prior to compaction. The additional eight hours of aging at 135°C is referred to by NCAT as the “critical aging” procedure as described in Chen et al. (10, 11), abbreviated as LMLC-CA.
3. Plant-mixed, lab-compacted (PMLC) specimens reheated (RH) to the compaction temperature, abbreviated as PMLC-RH.
4. PMLC specimens reheated then “critically-aged” prior to compaction, abbreviated as PMLC-CA.

Lab-mixed, lab-compacted specimens were prepared to match the gradations and asphalt contents for the respective mixes obtained from quality control testing as shown in Table 2. The plant-produced mixtures were sampled at the time of construction and placed in five-gallon buckets. The mixtures were later reheated to 150°C for two hours, quartered into appropriate masses for specimens, heated to the respective compaction temperature for one hour, and then compacted using a Superpave gyratory compactor (SGC). For plant-produced mixture samples subject to the critical aging procedure, the buckets of mix were reheated to 150°C for two hours to allow for the mixtures to be quartered and placed in pans in a thin layer less than approximately 20 mm thick, then put in an oven equipped with a timer to condition the loose mixtures for eight hours at 135°C. The eight-hour critical aging protocol was conducted

overnight so that the next morning, the mixtures could be immediately brought to the compaction temperature and then compacted with an SGC. The lab compacted specimens were compacted to $7.0 \pm 0.5\%$ air voids, except for mixture N2 which used a target air void content of $4.0 \pm 0.5\%$ and N5 which used a target air void content of $10.0 \pm 0.5\%$. After the laboratory experimental plan had been completed, it was found that the flexibility index and CT_{Index} were sensitive to the effect of specimen air void content in a counterintuitive way. Therefore, additional specimens were prepared at $7.0 \pm 0.5\%$ air voids for the N2 and N5 mixtures and tested in the I-FIT and IDEAL-CT.

The cyclic fatigue test was added to the experimental plan after the preparation of laboratory mixture samples had been completed. Therefore, the AMPT tests were only conducted on PMLC specimens (reheated and critically-aged).

2.4 Field Performance Results

From October 2015 to February 2021, the Cracking Group Experiment test sections accumulated 20 million ESALs. Their field performance at the end of the second three-year research cycle is summarized in Table 4. It should be noted that when cracking began to appear in any test section, a few cores were taken on cracks to determine if the cracking was confined only to the surface layer or if the cracks extended deeper. Those cores always confirmed that the cracks originated at the surface and there was no evidence of debonding or segregation. Figure 3 shows examples of cores taken to evaluate the cracking; the yellow line highlights the crack and the yellow arrow indicated the direction of traffic. The cracks were typically noticed first as hairline cracks in the outer parts of the wheelpaths and would grow in the direction of traffic and transverse to the direction of traffic in the wheelpaths.

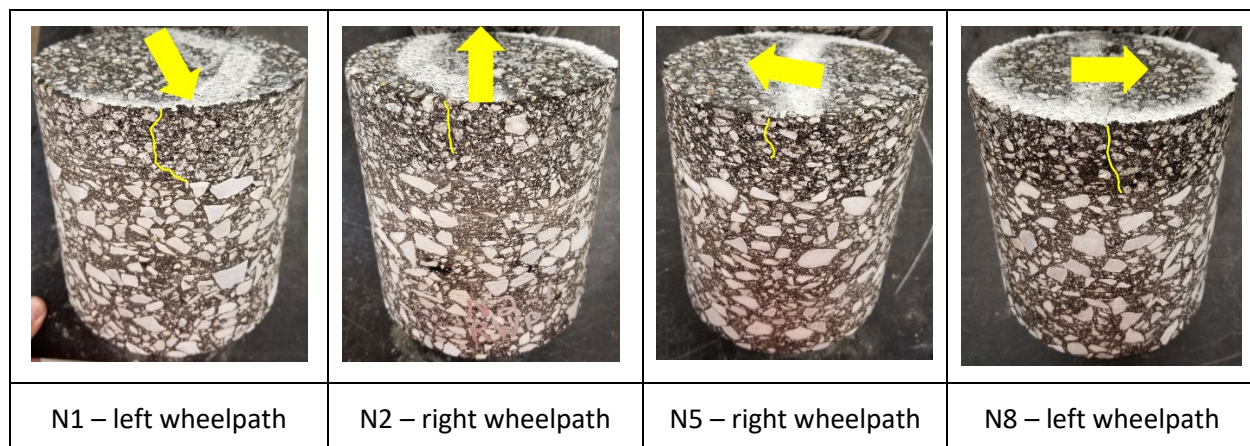


Figure 3. Photographs of Cores Showing Top-Down Cracking

Sections N5 and N8 reached terminal serviceability in February 2020, at which time the surface layers were milled and an overlay was constructed on the test sections to enable the trucks to continue to operate without diverting around those sections. After milling, it was noted only a small area of about 12 feet in length of section N8 had any deterioration in the lower layers, again confirming that the observed cracking was almost entirely top-down.

Table 4. Performance of NCAT Cracking Group Test Sections after 20 Million ESALs

Test Track Section	Mixture Description	Rutting (mm)	Change in IRI (in./mi)	Change in Mean Texture Depth (mm)	Cracking (% of lane area)	
					Feb. 2020	Feb 2021
N1	Control	4.9	5	0.60	11.2	44.5
N2	Control, Higher Density	4.4	11	0.60	7.7	12.5
N5*	Control, Low Density, Low AC	1.4	30	0.61	21.1	47.4*
N8*	Control + 5% RAS	1.8	50	0.76	70.8	99.3*
S5	35% RAP, PG 58-28	3.5	5	0.66	0.2	1.1
S6	Control, HiMA binder	3.3	8	0.78	0	0.9
S13	Gap-graded, asphalt-rubber	5.6	9	0.20	0	0

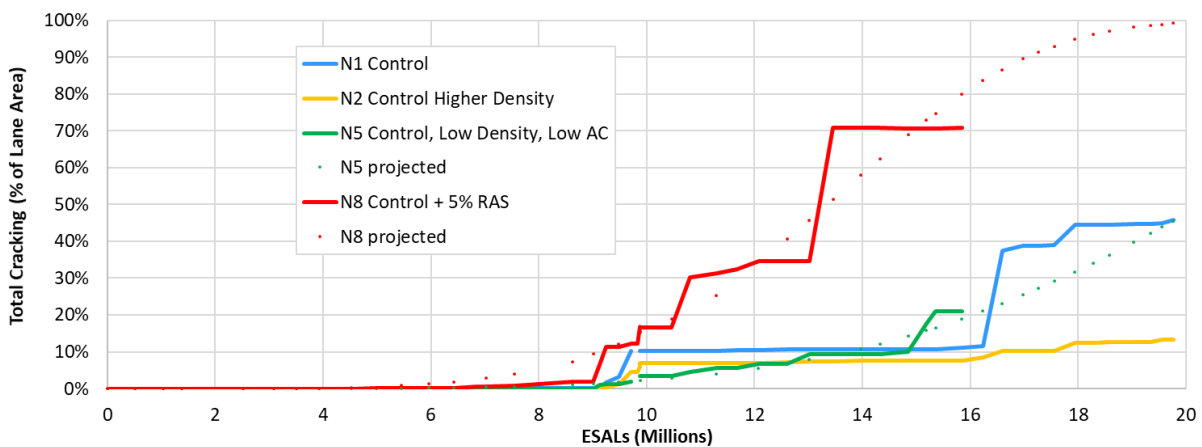
*Rutting, IRI, and MTD data are from Feb. 2020 at 16 Million ESALs prior to mill and overlay of the section. Cracking results reported for Feb 2021 for these sections are projected.

The cracking data in Table 4 are shown for two dates. February 2020 was the last month that all of the sections were still in service and represents the conditions after approximately four and a half years and 16 million ESALs had been applied to the test sections. February 2021 represents the completion of the experiment. After N5 and N8 were milled and overlaid in February 2020, the cracking data for these two sections had to be projected forward based on data prior to their removal from the experiment. A sigmoidal function was fit to the cracking versus ESALs data for these sections using the form of equation 1.

$$Cracking = 1 - e^{(-\theta_1 \times MESALs^{\theta_2})} \quad (1)$$

Where θ_1 and θ_2 are curve fitting coefficients and MESALs are millions of ESALs.

This approach to estimating the amount of cracking beyond 16 million ESALs was found to be reasonable based on comparisons of measured versus predicted data for N1 and N2. This projection of cracking for these two sections was necessary to develop correlations between lab results and field performance using all of the test sections. Figure 4 shows the progression of cracking for the test sections over the two cycles.

**Figure 4. Cracking versus ESALs for the Four Test Sections with Significant Cracking**

The field performance of the Cracking Group Experiment test sections satisfied the primary objective of the study, which was to generate a wide range in the extent and severity of top-down cracking. As can be seen in Table 4, each of the test sections performed very well with regard to rutting and change in texture, which is an indicator of raveling. After 20 million ESALs of trafficking, three of the test sections (S5, S6, and S13) had little to no cracking, one section (N2) had a low amount of low severity cracking, two sections (N1 and N5) had a moderate extent of cracking, about 45% – although their severity levels were quite different, and one section (N8) had extensive, high severity cracking.

The difference in severity level of cracking for N1 (Ctrl) and N5 (Ctrl Low Density & AC) is evident in two ways. Photos from the sections (Figure 5) show that N1 had low severity cracking in January 2021, whereas the cracking in N5 observed on December 19, 2019 had begun to lead to shallow potholes through the surface layer. This is also evident in the change in IRI for these sections. Section N1 had a change of only five inches per mile in roughness over five and a half years, whereas N5 increased in roughness by 30 inches per mile over four and a half years.

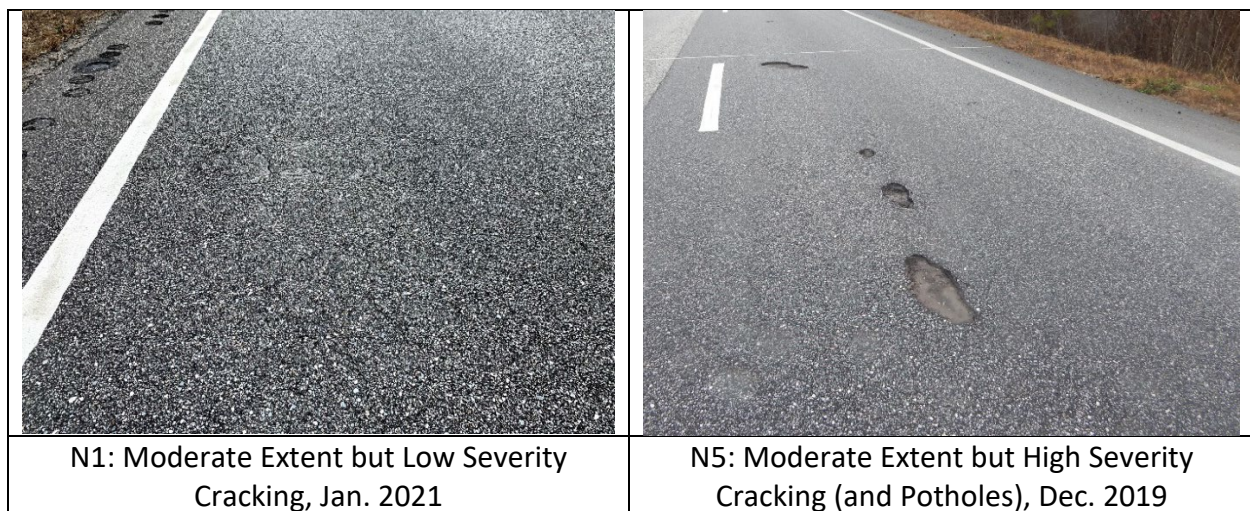


Figure 5. Cracking in N1 (Control) and N5 (Control, Low Density, Low AC)

Figure 6 shows a photo of part of N8 (Ctrl +5% RAS) from December 2019. At that point in time, the cracking was about 70% of the lane area with potholes starting to develop.

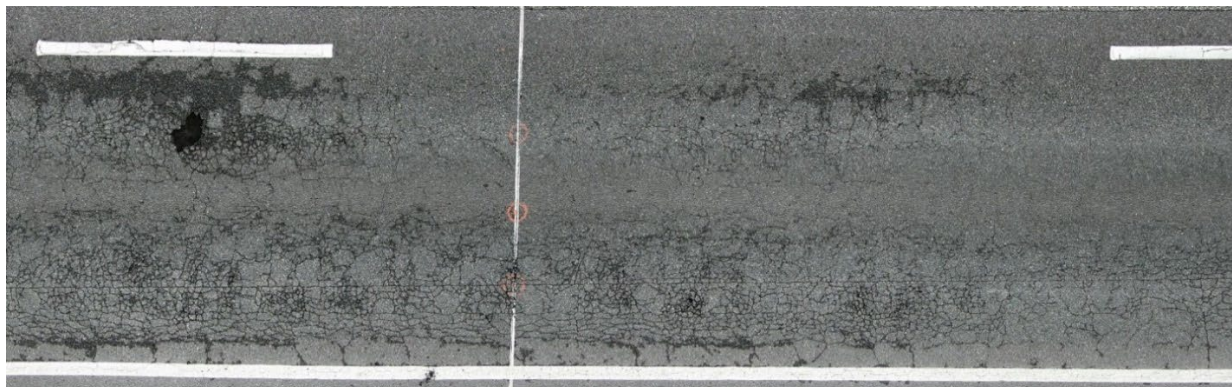


Figure 6. Photograph of Distress in Part of N8 in December 2019

2.5 Pavement Response Analysis

Part of the field investigation of the Cracking Group sections included measurements made from embedded instrumentation under truck loading along with backcalculation of in-situ material properties from falling weight deflectometer (FWD) testing. The goals with each of these monitoring programs included the following:

- Characterize seasonal temperature effects on pavement responses;
- Evaluate differences in pavement responses driven primarily by surface lift mixture differences; and
- Quantify effects of pavement cracking on measured pavement responses through non-destructive testing.

To that end, each section in the experiment was instrumented with asphalt strain gauges (ASGs), earth pressure cells (EPCs), and temperature probes during the construction process. The ASGs were placed to measure bending-induced tensile strain at the bottom of the asphalt base layer in the direction of traffic while the EPCs were placed at the asphalt concrete/granular base interface and granular base/subgrade interface to measure vertical pressures at those depths, respectively. While 100% of the EPCs survived installation and both research cycles, the ASG survivability was extremely low and did not produce sufficient data to include in the following analyses. Data were collected from the embedded gauges twice per week during the first test cycle followed by weekly in the second test cycle. Data collection alternated mornings and afternoons to capture both the short term (i.e., daily) and long term (i.e., seasonal) temperature fluctuations.

The FWD testing program consisted of testing approximately three times per month along the inside, outside, and between wheelpaths at four random locations in each section. Each set of deflection data at a given location consisting of three replicate drops at three drop heights representing 6, 9, and 12-kip loadings. The measured deflection data were used in EVERCALC 5.0 to back-calculate the layer properties to establish a time history of in situ properties during the two-year test cycle. Only data from the 9-kip load level resulting in backcalculated properties with a root-mean-square-error (RMSE) of less than 3% are presented below.

2.5.1 *Measured Pavement Responses*

Data collected over both test cycles were normalized to a reference temperature of 68°F and plotted against time in Figures 7 and 8 for base and subgrade pressures, respectively. The temperature normalization process followed a previously-established procedure (1) and used data collected from both test cycles. The gap in both data sets corresponds to the reconstruction time between test cycles where traffic was not applied to the sections.

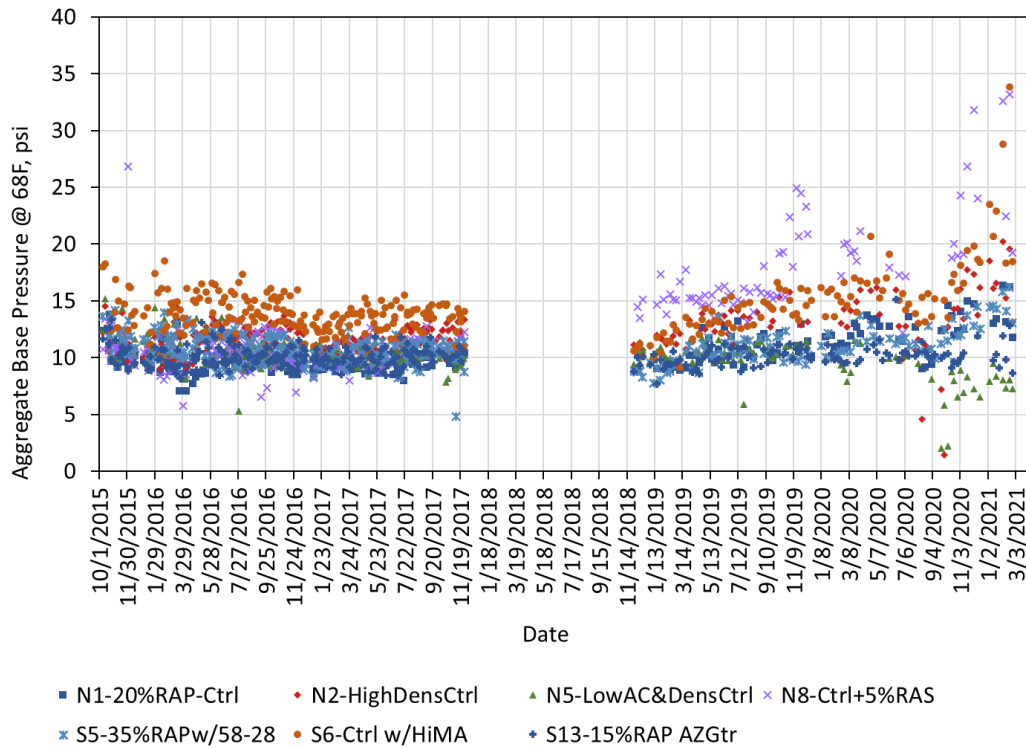


Figure 7. Base Pressure at 68°F Versus Time

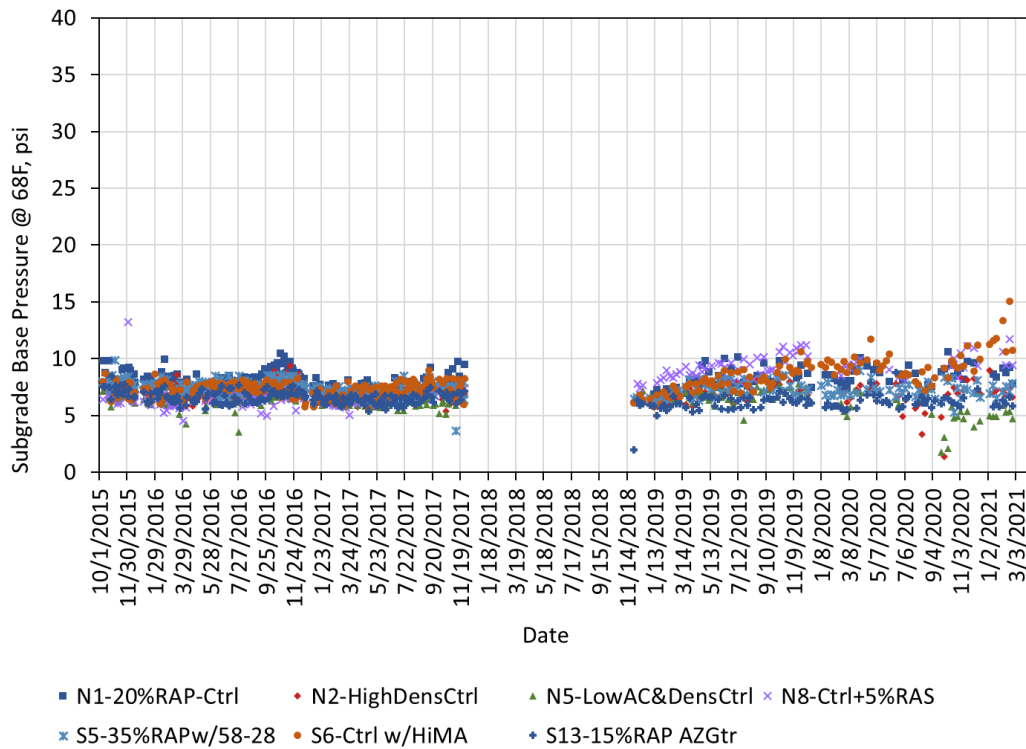


Figure 8. Subgrade Pressure at 68°F Versus Time

As shown in Figures 7 and 8, the data from the first test cycle (2015 through 2017) show no apparent effects of pavement damage as each data set is relatively constant over time, or even decreasing as with S6 (Ctrl w/ HiMA). The damage that had appeared during the first test cycle was not severe enough to increase the base and subgrade pressure readings. However, the effects of pavement damage become much more evident with certain sections in the second test cycle. The most severely damaged section was N8 (Ctrl + 5%RAS) which reached 20% of lane area cracked in mid-February 2019. A sharp increase of 10 psi in the aggregate base pressure (Figure 7) and a more general upward trend of about 4 psi in the subgrade pressure (Figure 8) correspond to this increasing cracking damage. Clearly, though the cracking was top-down, it was severe enough to affect the load carrying capacity of the section resulting in elevated stress levels in the pavement foundation. Section N5 also reached 20% of lane area cracked 11 months later in mid-January 2020. The steadiness of both pressure data sets, and even a slight decline at the end of the second test cycle in Figures 7 and 8, suggest that though there was extensive cracking at the surface, it was not severe enough to affect stress levels deeper in the pavement structure. Both of these sections were milled and inlaid in mid-February 2020.

The remaining sections survived to the end of the second test cycle with no rehabilitation but with varying degrees of cracking across the lane area as noted in Table 4. It is important to emphasize that the measurements shown in Figures 7 and 8 correspond to particular locations in each section and do not represent the overall condition of the pavement, but rather only in close proximity to the particular pressure cells. That may help explain the increasing pressure measurements toward the end of the second test cycle for some of the sections. Notable are N1 (20%RAP-Control), N2 (High Density Control), S5 (35%RAP w/58-28) and S6 (Ctrl w/ HiMA), which all show similar trends with slightly increasing pressure measurements during the second test cycle with a sharper increase in the last few months. This suggests that the cracking in the vicinity of the pressure cells had become severe enough to begin compromising the sections' structural integrity. The exception was S13 (15%RAP GTR), which did not have any observable cracking and no appreciable changes in pressure measurements.

2.5.2 Backcalculated Asphalt Concrete Moduli

Backcalculated AC moduli, normalized to 68°F following procedures similar to normalizing pressure responses, are plotted in Figure 9 for both test cycles. The vertical spread in the respective data series represents spatial variability within each test section and it appears that the modulus was decreasing toward the end of the second cycle as cracking became more widespread and severe. To better visualize the data, the data was subdivided into grouped sections and examined by offset (between wheelpaths, inside wheelpath and outside wheelpath) as shown in Figures 10 through 15.

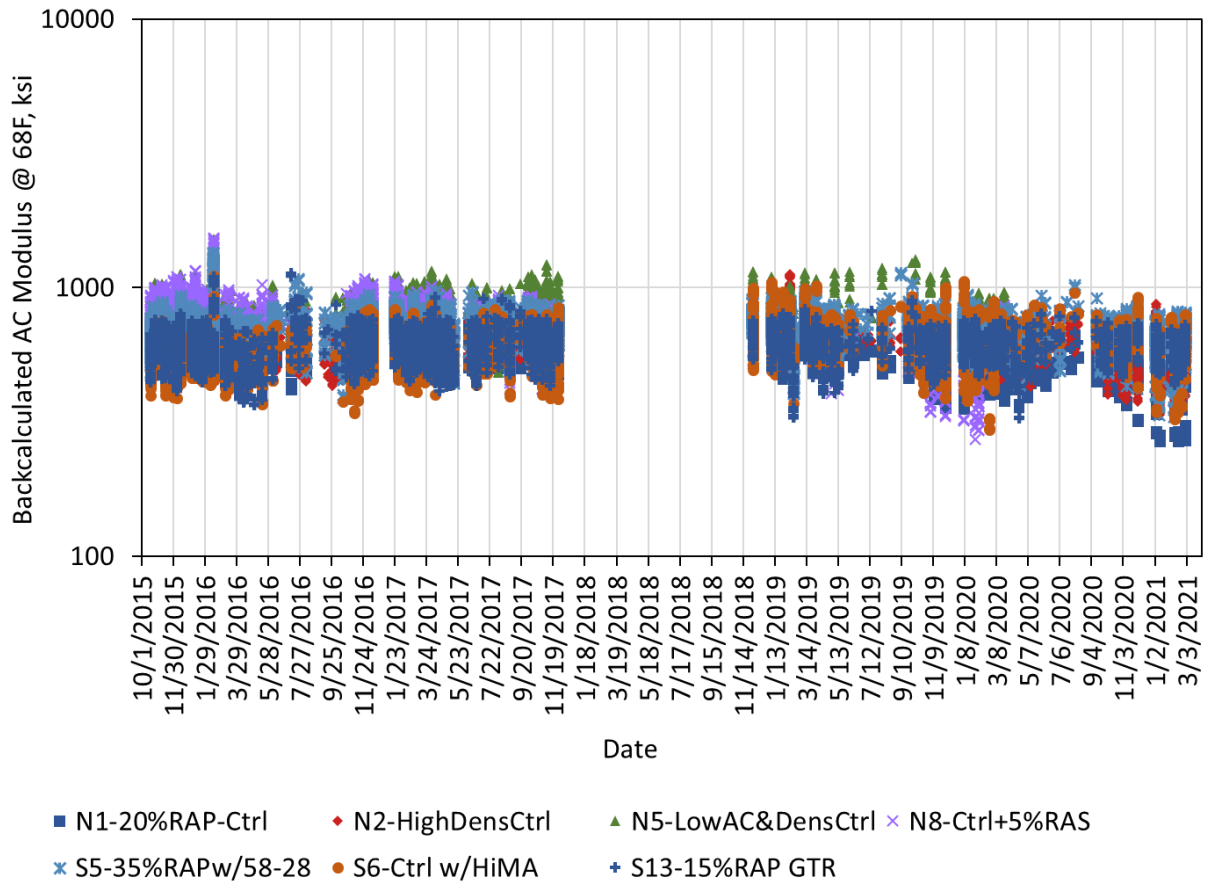


Figure 9. Backcalculated AC Moduli at 68°F for All Test Locations

Sections N5 (LowAC&DensCtrl) and N8 (Ctrl+5%RAS) experienced the most widespread and severe damage and are shown in Figure 10, which contains data from both test cycles and all test locations. It is clear that N8 had a substantial decline in AC modulus during the second test cycle as damage became more severe up until it was milled and overlaid. This trend is not as readily apparent in Section N5, and the data appear more scattered. The trendlines in Figure 10 quantify the decline with an approximate 33% decrease in AC modulus in Section N8 and only an 8% decrease in Section N5.

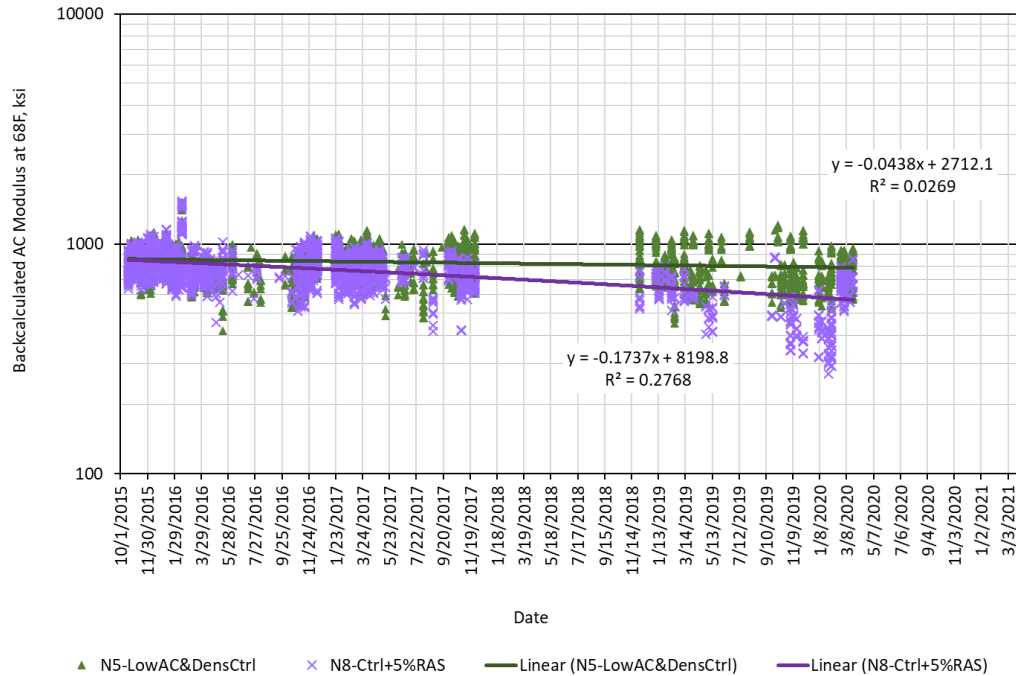
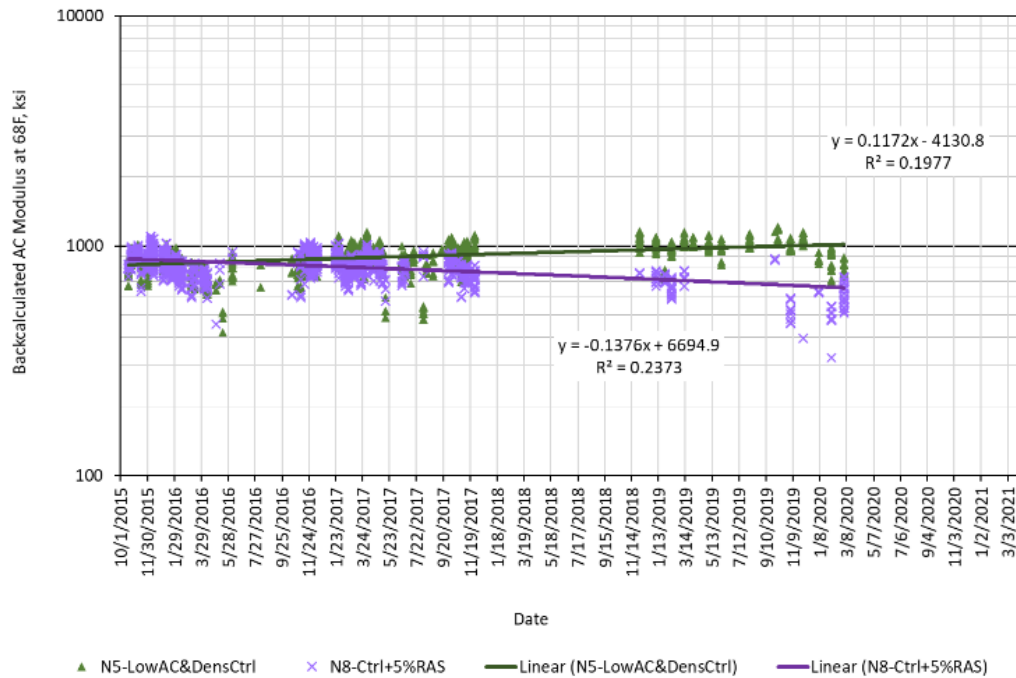


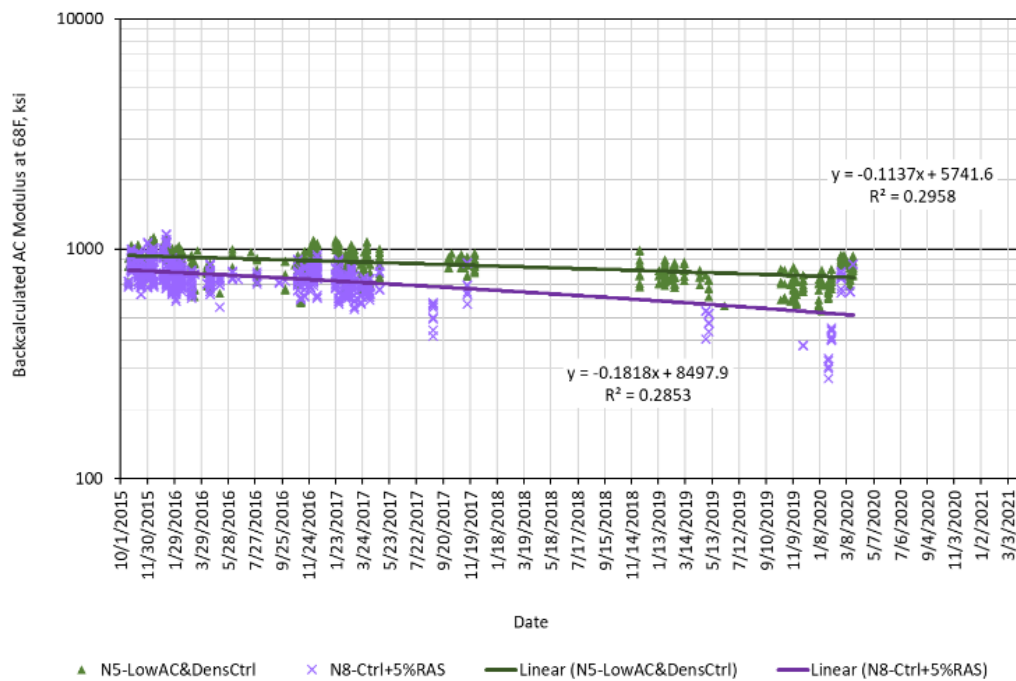
Figure 10. N5 and N8 Backcalculated AC Moduli at 68°F for All Test Locations

To more closely examine the effect of transverse offset in the lane, Figure 11 subdivides the data from Figure 10 into between wheelpaths (Figure 11a), inside wheelpath (Figure 11b) and outside wheelpath (Figure 11c) with linear trendlines fit to the respective data sets. The slopes of the trendlines indicate if cracking at the surface is affecting structural integrity. Larger negative slopes (and corresponding higher R^2) indicate more significant damage while smaller slopes (and corresponding smaller R^2) indicate less damage.

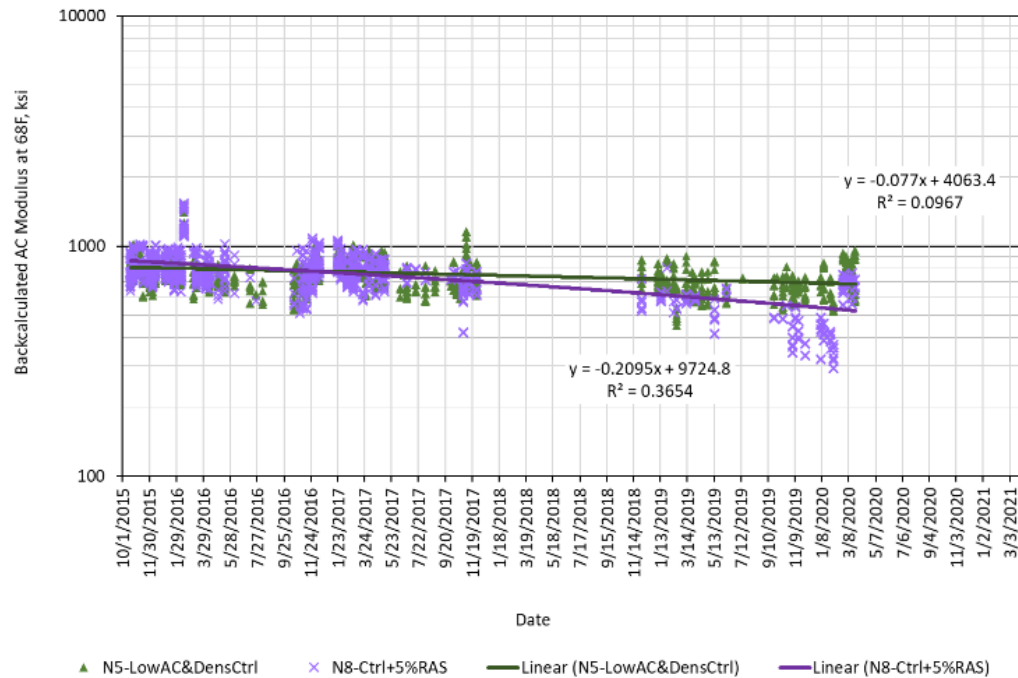
In all three offsets (between, inside, and outside) Section N8 (Ctrl+5%RAS) experienced more pavement damage with a similar order of magnitude between all three. The most severe offset was the outside wheelpath where the AC modulus at 68°F decreased by 39% during the entire experiment. In contrast, N5 (LowAC&DensCtrl) did not appear to suffer structural degradation between the wheelpaths and had a relatively slighter damaging affect in the inside and outside wheelpaths. In the inside wheelpath, N5 experienced a 20% modulus reduction from start to finish. It should be noted that these sections were resurfaced before the others so their degradation should not be directly compared to the others that experienced the full 20 million ESALs.



a) Between Wheelpaths



b) Inside Wheelpath



c) Outside Wheelpath

Figure 11. N5 and N8 Backcalculated AC Moduli at 68°F By Wheelpath

Since the other sections did not appear to have changing modulus versus time during the first test cycle, the remaining plots (i.e. Figures 12 through 15) focus only on the second test cycle. Figure 12 contains data for N1 (20%RAP-Control) and N2 (High Density Control) at all locations while Figure 13 subdivides the data by wheelpath. The trendlines in Figure 12 indicate similar behavior from both sections with a decline in modulus of 32% (N1) and 24% (N2) averaged across each section. Similar trends were observed by wheelpath (Figure 13), indicating that the structural degradation was equally experienced throughout the section.

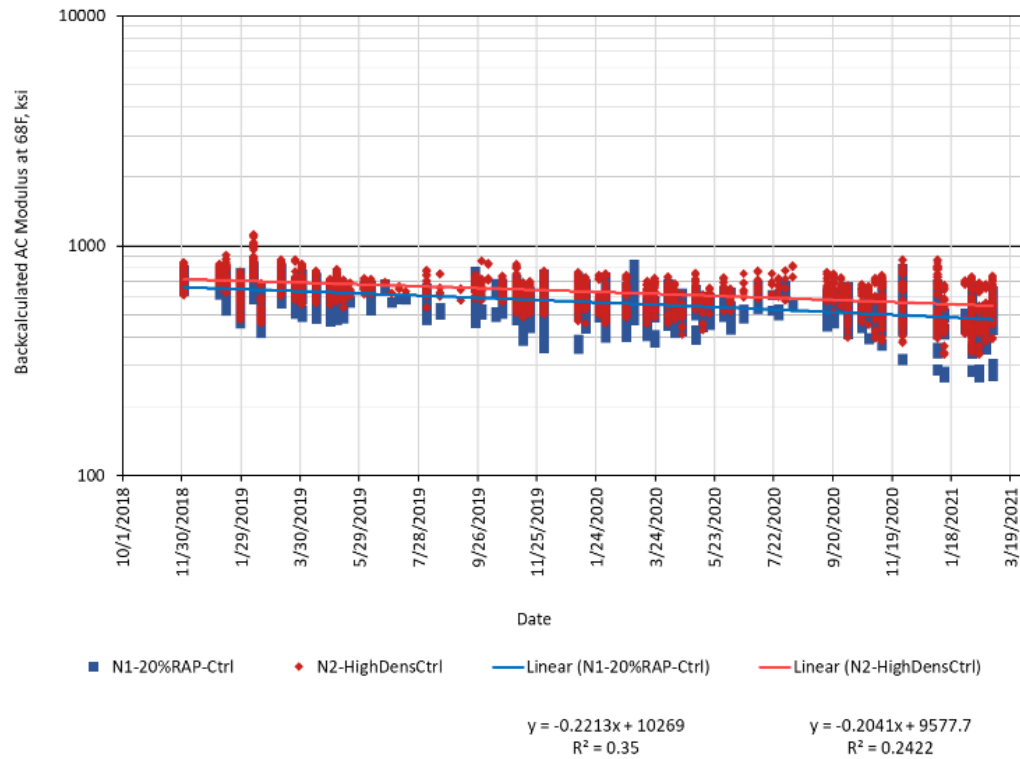
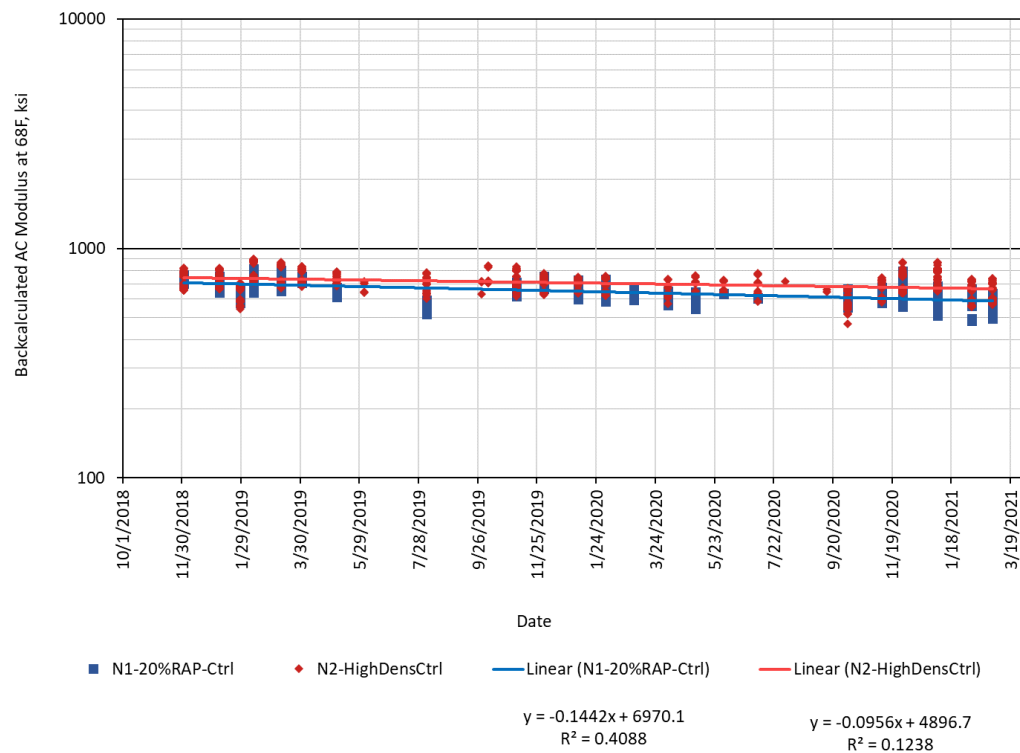
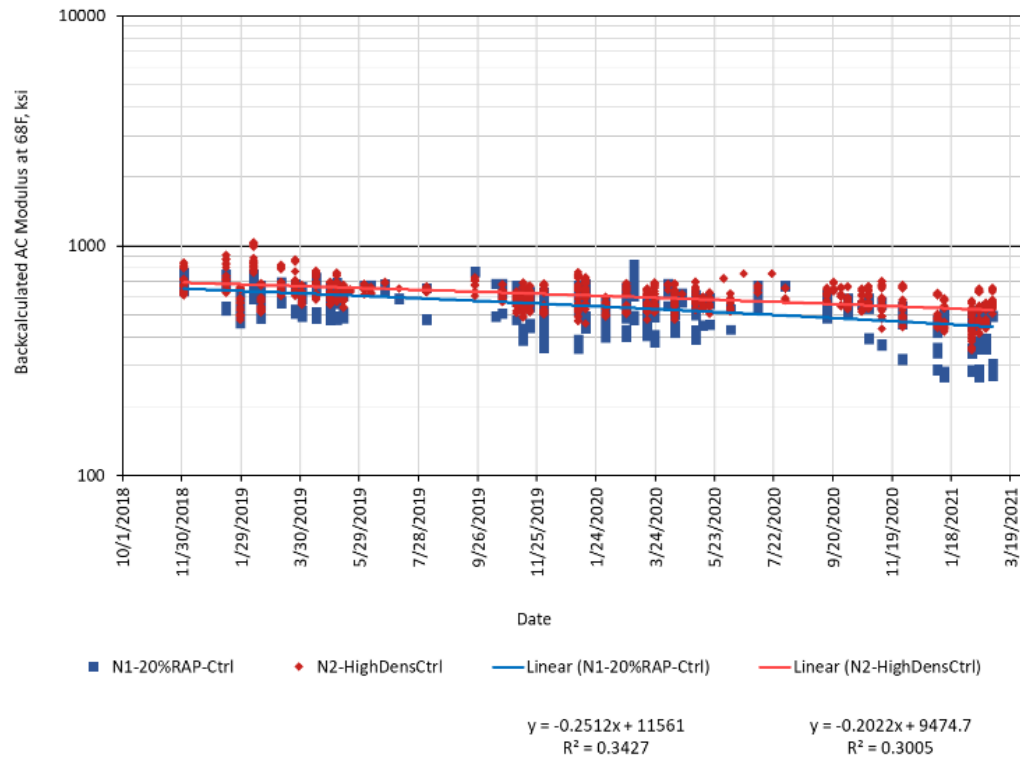


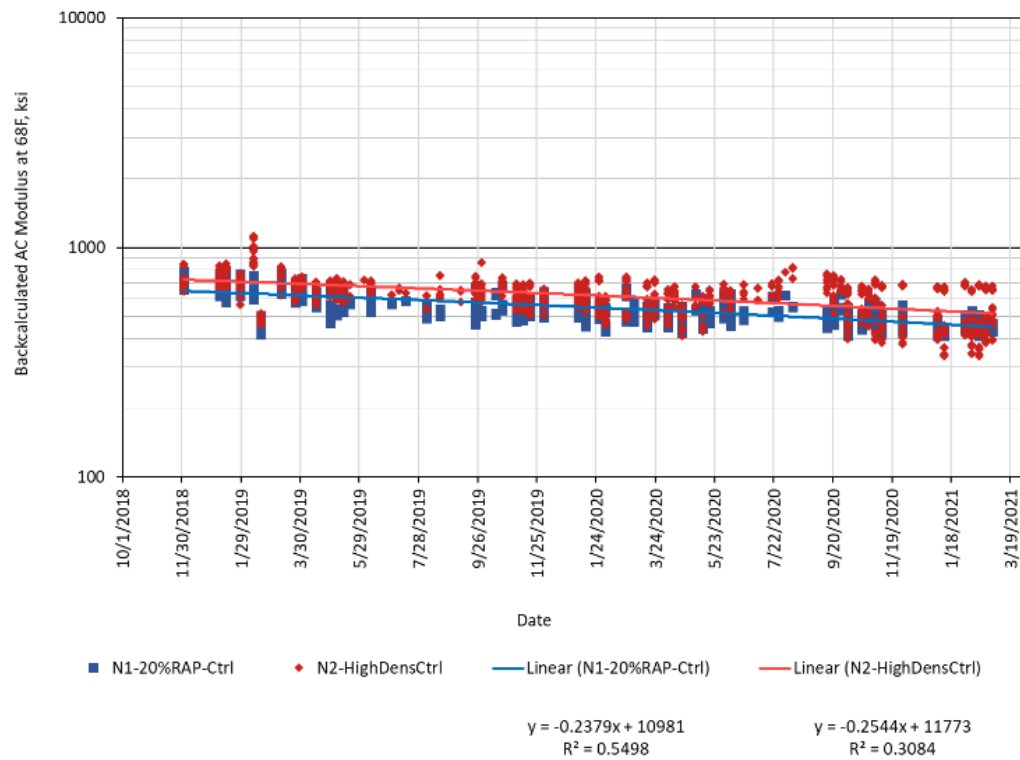
Figure 12. N1 and N2 Backcalculated AC Moduli at 68°F for All Test Locations (Second Test Cycle)



a) Between Wheelpaths



b) Inside Wheelpath



c) Outside Wheelpath

Figure 13. N1 and N2 Backcalculated AC Moduli at 68°F by Wheelpath (Second Test Cycle)

It is evident from Figure 14 that the south tangent sections (S5-35%RAP w/58-28, S6-Ctrl w/ HiMA & S13-15%RAP GTR) did not experience the same level of section-wide structural degradation as the north tangent sections during the second test cycle with nearly flat trendline slopes and low R^2 values. Of these sections, S13 experienced only a 2% decline in modulus while S5 (12% decrease) and S6 (20% decrease) were greater. Recall from the discussion of pressure responses that S13 had essentially no change in pressure at 68°F, which is consistent with no changes in AC modulus. The other sections experienced some increases in pressure, corresponding to their decreasing moduli. Subdividing the data by wheelpath as shown in Figure 15 found similar trends.

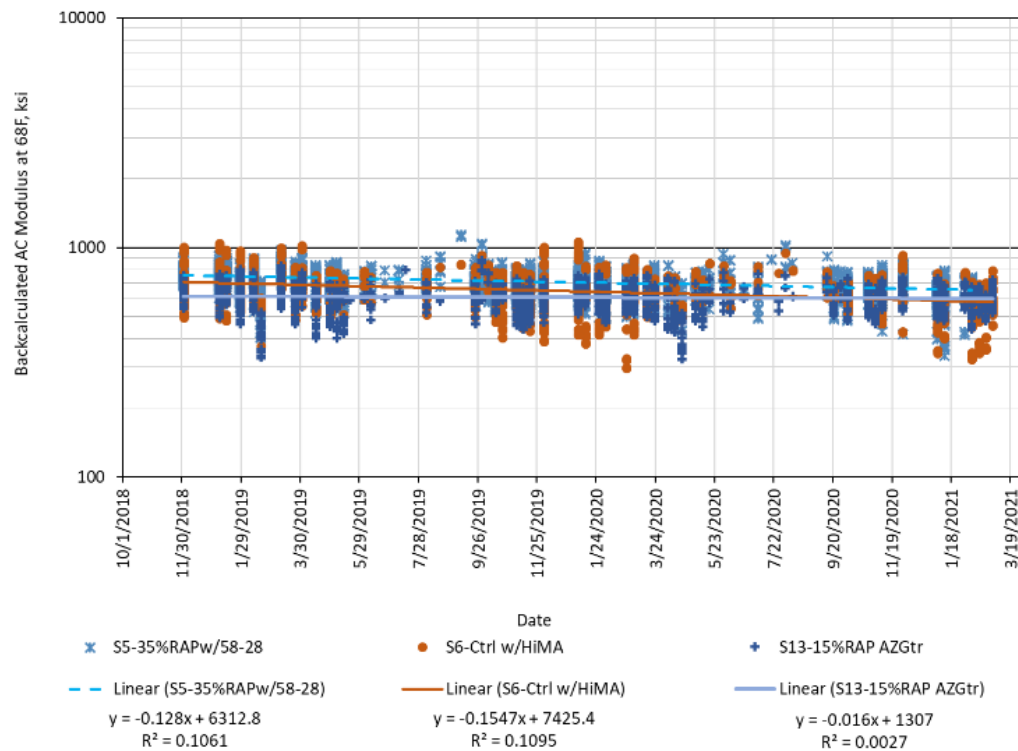
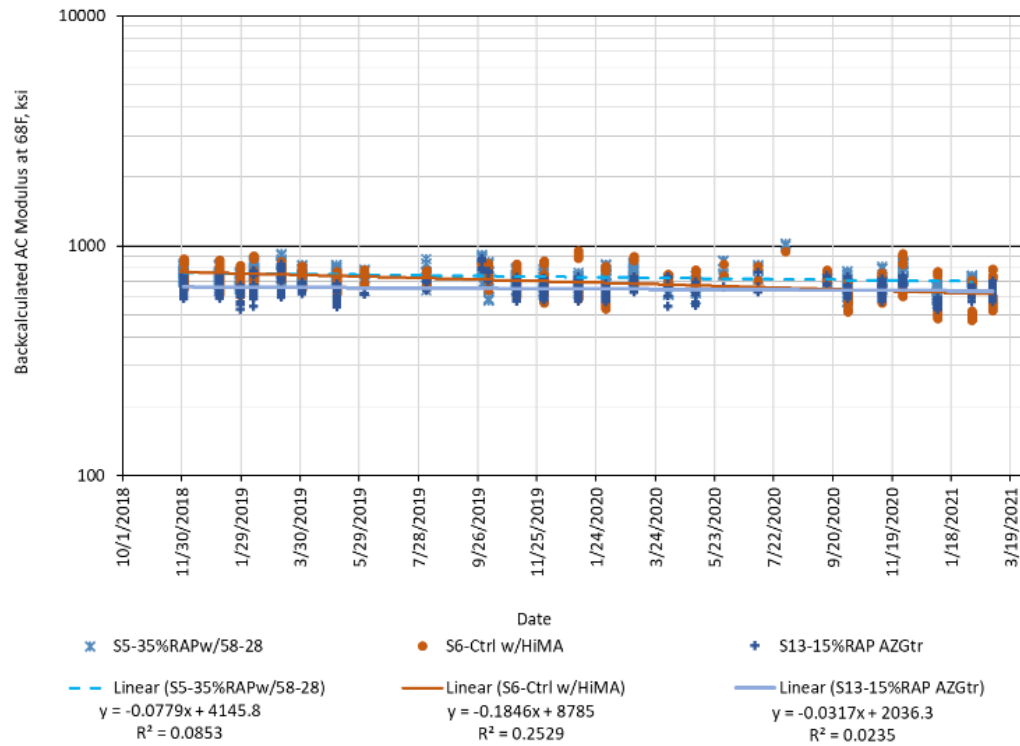
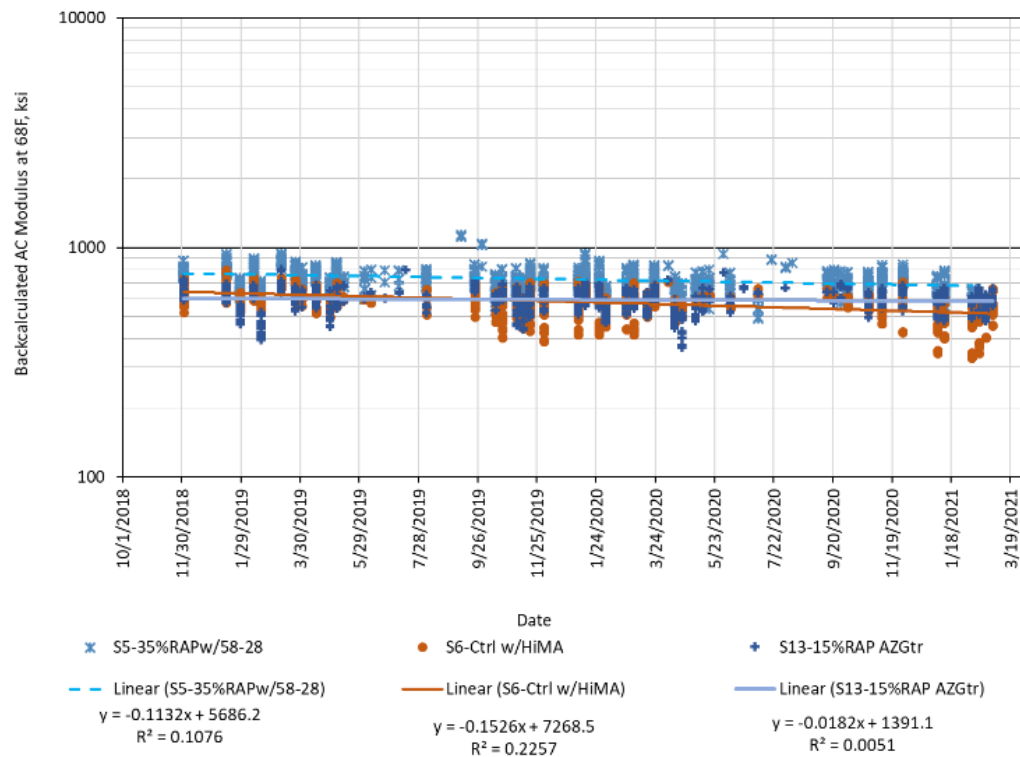


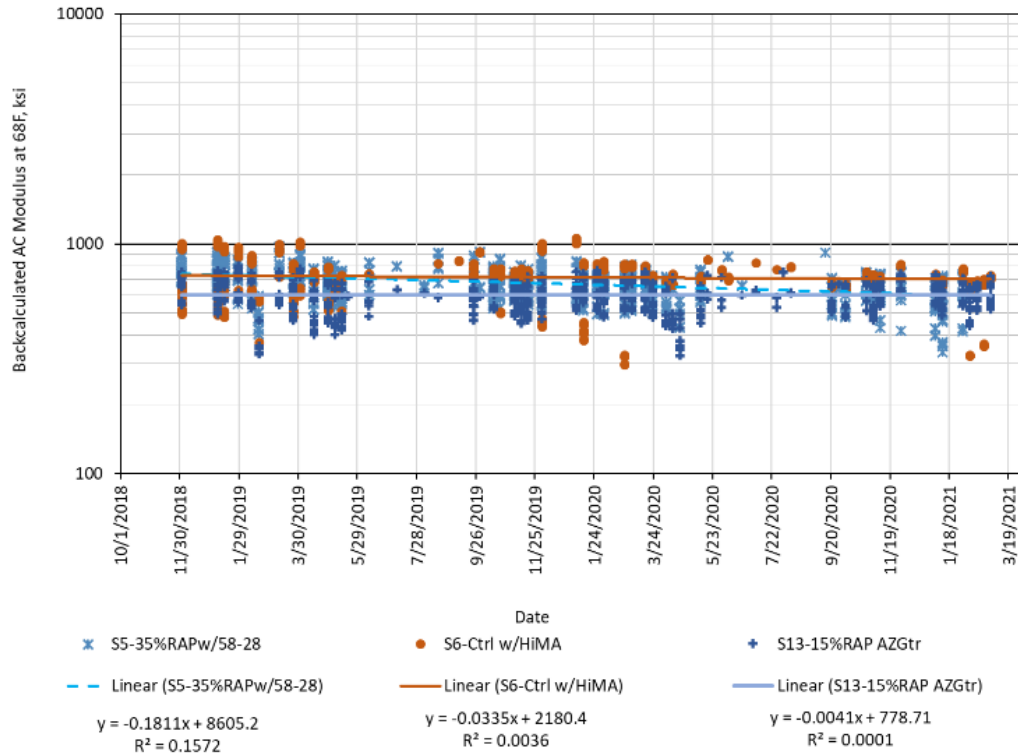
Figure 14. S5, S6, and S13 Backcalculated AC Moduli at 68°F at All Test Locations (Second Test Cycle)



a) Between Wheelpaths



b) Inside Wheelpath



c) Outside Wheelpath

Figure 15. S5, S6, and S13 Backcalculated AC Moduli at 68°F by Wheelpath (Second Test Cycle)

Table 5 summarizes the average decrease in modulus over two research cycles as described above. It is important to emphasize that Sections N5 (LowAC&DensCtrl) and N8 (Ctrl+5%RAS) were milled and inlaid in February 2020 so their respective decreases correspond to 16 million ESALs compared to 20 million ESALs for the other test sections. Of the remaining sections, the best structural performer was S13 (15%RAP GTR) which was also evident from the pressure measurements. The other sections landed between these sets with somewhat similar structural performance.

Table 5 AC Modulus Decrease by Test Section

Section	Average Section-Wide AC Modulus Decrease
N1 (20%RAP-Control)	32%
N2 (High Density Control)	24%
N5 (LowAC&DensCtrl)*	8% (20% in Outside Wheelpath)
N8 (Ctrl+5%RAS)*	33% (39% in Outside Wheelpath)
S5 (35%RAP w/58-28)	12%
S6 (Ctrl w/HiMA)	20%
S13 (15%RAP GTR)	2%

*Taken out of service in February 2020.

2.6 Results of Laboratory Tests and Field Performance Correlations

This section presents the results of the laboratory cracking tests for each of the methods selected by the sponsors of the experiment. For a complete description of the tests, associated methods of analysis, and aging condition, please see the work by Chen (2). The number of

replicates per test are noted in the discussion for each method. Outlier analyses were conducted on results for each set of replicates following ASTM E 178 using a 90% confidence level.

As previously noted, all specimens were initially prepared at $7.0 \pm 0.5\%$ air voids except for N2 (Control High Density) which used a target air void content of $4.0 \pm 0.5\%$ air voids, and N5 (Control Low Density & AC) which used target air void content of $10.0 \pm 0.5\%$ air voids, representing the nominal target in-place relative density for each section. Later in the experimental work, additional I-FIT and IDEAL-CT tests were conducted on N2 and N5 mixtures also compacted to $7.0 \pm 0.5\%$ air voids to avoid the effect of specimen air voids on the results of these cracking tests. For these two tests, results are reported for the specimens prepared to $7.0 \pm 0.5\%$ air voids and at the nominal target in-place air void contents for N2 and N5. However, the correlations of lab results to field performance are presented only for test results of specimens prepared to $7.0 \pm 0.5\%$ air voids. The field cracking data used in correlation analyses corresponds to the percent of lane area cracked after 20 million ESALs in Table 2. As previously noted, sections N5 (Control Low Density & AC) and N8 (Control +5% RAS) were milled and inlaid in February 2020; thus, the field cracking data of these two sections was projected to 20 million ESALs based on non-linear extrapolation using Equation 1.

2.6.1 Energy Ratio Results

The Energy Ratio results are summarized in Table 6. ER results are determined from analysis of three tests with the trimmed means of those tests used to calculate the single ER value. Although there were at least three replicates for each component test in the ER procedure, the final ER value does not have replicates, which limits many types of statistical analyses such as Analysis of Variance (ANOVA).

According to Roque et al. (3), $DCSE_{HMA}$ is the amount of energy required to initiate cracking. From their testing of cores from pavements at least 10 years old, they proposed a minimum $DCSE_{HMA}$ criteria of 0.75 kJ/m^3 to screen out extremely stiff mixtures and a minimum ER criterion of 1.95 for pavements subject to one million ESALs or more per year. However, the Test Track sections are exposed to about five million ESALs per year, a much higher level of traffic than the ER validated range.

From the results in Table 6, it can be seen that only the critically-aged lab prepared sample (LMLC-CA) for N8 had a mean $DCSE_{HMA}$ result below 0.75 kJ/m^3 , indicating that this mixture is very stiff and susceptible to top-down cracking. That is a correct assessment as shown in the field performance results. However, the ER parameter does not correctly identify this mix as being more susceptible to top-down cracking than the other mixes in each set of sample type and mix aging condition. Furthermore, the mix with the best field performance, S13 (Gap-Gr, asphalt-rubber), has the lowest ER in each mix set, which is also an incorrect assessment in terms of top-down cracking resistance.

Table 6. Results of Energy Ratio Tests

Test Section and Mixture Description	Resilient Modulus (GPa)	Creep Compliance Rate	IDT Fracture Energy (kJ/m³)	DCSE _{HMA} (kJ/m³)	Energy Ratio
	Trimmed Means				
LMLC-STOA					
N1: Control	10.41	2.94E-09	3.6	3.30	5.11
N2: Control, Higher Density	15.32	1.85E-09	3.4	3.08	7.38
N5: Control, Low Dens. & AC	10.13	3.50E-09	1.5	1.31	1.85
N8: Control + 5% RAS	12.34	6.77E-10	1.6	1.35	8.06
S5: 35% RAP, PG 58-28	9.70	3.98E-09	2.7	2.48	3.81
S6: Control, HiMA Binder	6.96	3.01E-09	4.9	4.67	6.85
S13: Gap-gr., asphalt-rubber	9.33	4.61E-09	2.1	1.97	2.08
LMLC-CA					
N1: Control	14.12	4.16E-10	2.2	1.90	18.03
N2: Control, Higher Density	15.97	2.62E-10	2.3	1.95	28.69
N5: Control, Low Dens. & AC	10.71	8.68E-10	1.4	1.17	5.55
N8: Control + 5% RAS	15.03	2.07E-10	0.8	0.58	12.46
S5: 35% RAP, PG 58-28	12.87	4.59E-10	2.0	1.74	15.20
S6: Control, HiMA Binder	8.55	9.44E-10	4.0	3.74	16.69
S13: Gap-gr., asphalt-rubber	9.87	2.22E-09	2.2	1.97	3.79
PMLC-RH					
N1: Control	9.94	3.79E-09	4.8	4.52	5.52
N2: Control, Higher Density	12.41	1.98E-09	3.9	3.58	7.43
N5: Control, Low Dens. & AC	7.93	4.31E-09	3.4	3.19	3.57
N8: Control + 5% RAS	12.75	4.98E-10	1.8	1.57	12.82
S5: 35% RAP, PG 58-28	7.38	3.46E-09	6.0	5.76	7.39
S6: Control, HiMA Binder	7.28	2.44E-09	5.4	5.17	9.18
S13: Gap-gr., asphalt-rubber	7.40	5.17E-09	2.7	2.52	2.24
PMLC-CA					
N1: Control	13.08	8.31E-10	2.7	2.43	11.36
N2: Control, Higher Density	16.99	2.16E-10	1.8	1.49	26.42
N5: Control, Low Dens. & AC	11.04	9.13E-10	1.9	1.68	7.60
N8: Control + 5% RAS	17.28	1.93E-10	1.0	0.79	17.16
S5: 35% RAP, PG 58-28	9.89	1.23E-09	3.9	3.67	12.35
S6: Control, HiMA Binder	8.73	1.04E-09	6.8	6.50	25.55
S13: Gap-gr., asphalt-rubber	10.54	1.48E-09	3.2	2.98	8.36

Figure 16 shows a bar chart for the ER results of the critically-aged plant mix samples. The chart is divided into three sections based on the observed field performance. It is clear from this chart that ER does not distinguish the mixtures with different cracking susceptibilities. Since there are no replicates of ER results for each mixture, standard deviations for ER cannot be determined and statistical comparisons are not possible.

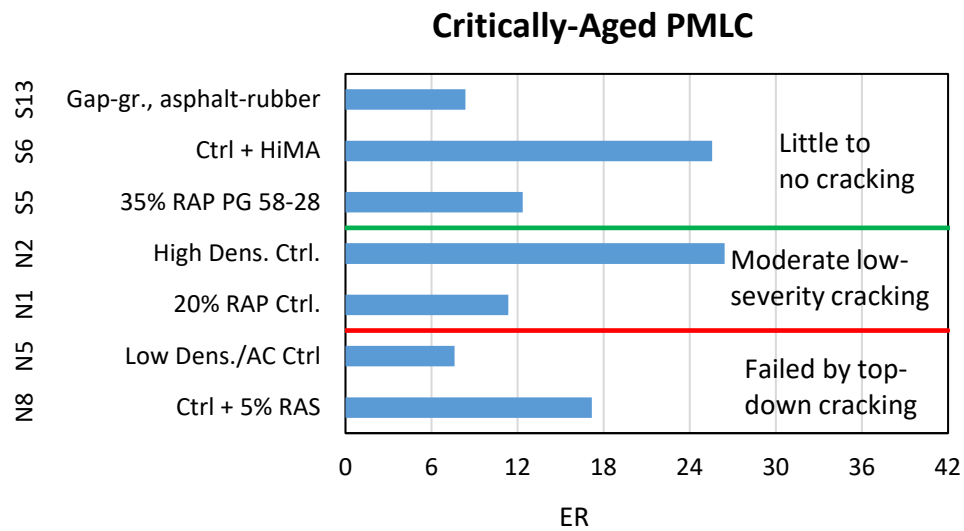


Figure 16. Chart of ER Results and Field Performance Groupings

Figure 17 shows correlation plots of Energy Ratio versus the observed cracking data on the Test Track for each of the four sample preparation and mix aging sets. The R^2 values shown are for the best-fit least-squares regressions among linear, exponential, logarithmic, or power functions as determined by Excel. It is evident from each of the four plots that ER is a not a good indicator of top-down cracking resistance for these asphalt mixtures.

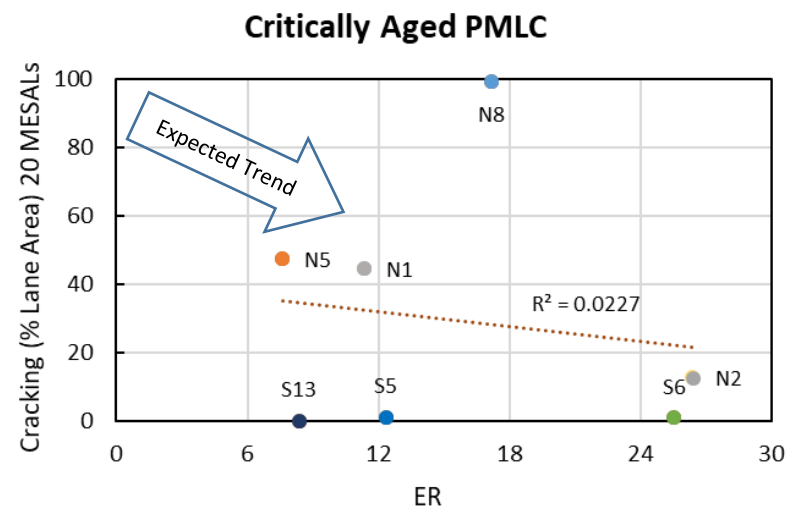
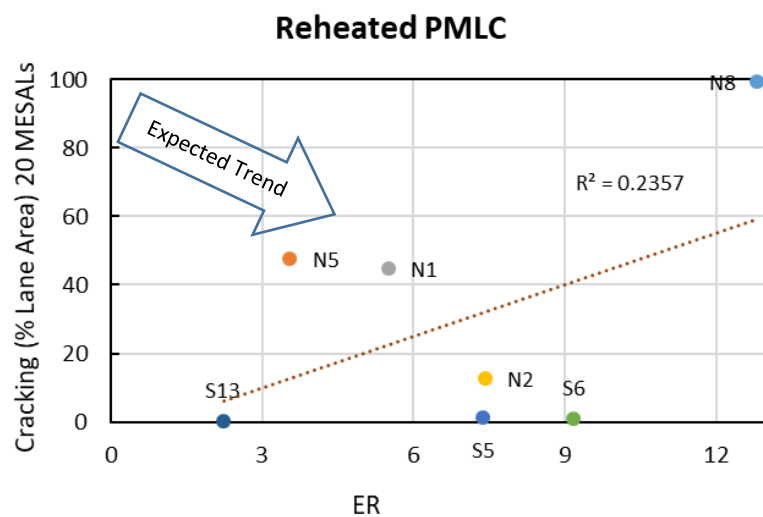
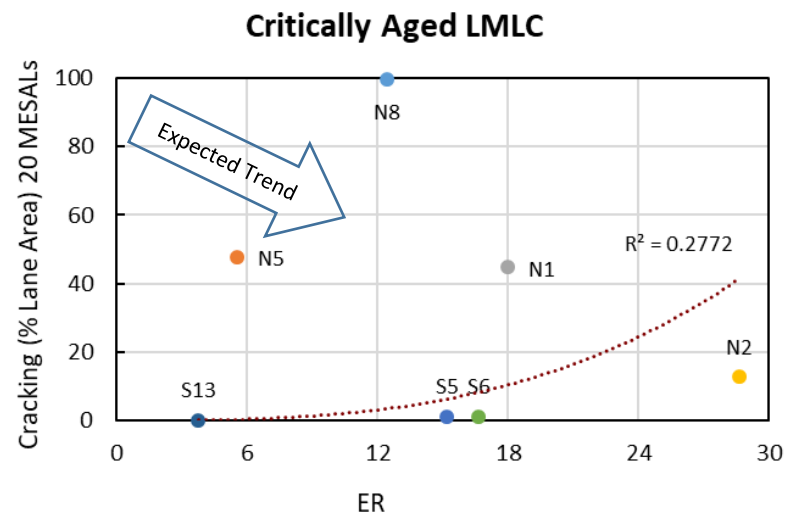
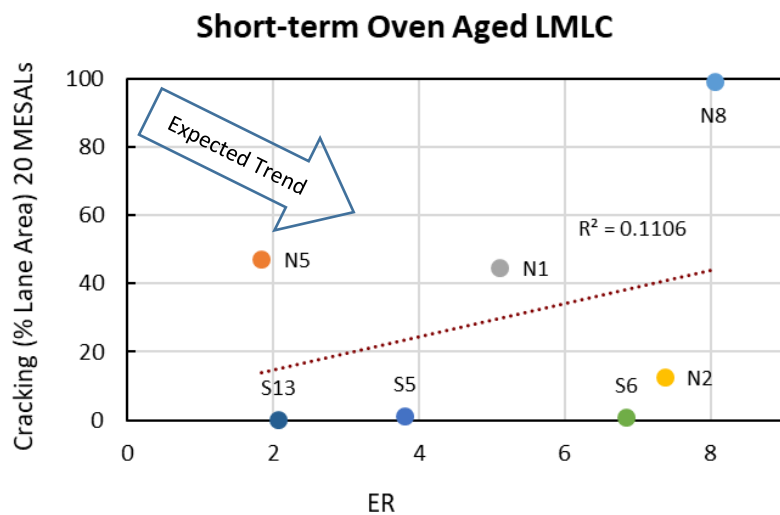


Figure 17. Correlations of ER with Field Performance for the Lab and Plant Samples Subject to Different Aging Conditions

2.6.2 Texas Overlay Results

During the first half of this experiment, the OT-TX results were analyzed and reported as cycles to failure, N_f , according to Texas method in 2018 (1). Since that time, the Texas DOT has begun using two other parameters, critical fracture energy (G_c) and the crack progression rate (β) developed by Garcia et al. to assess cracking resistance of asphalt mixtures (4). Critical fracture energy has been shown to be a measure of mixture toughness rather than cracking resistance, whereas the crack progression rate parameter has been found to be an excellent indicator of cracking resistance and is less variable than the number of cycles for failure parameter, N_f (4, 5). Since the primary purpose of this study is to determine the relationships of the cracking test results to the field performance on the Test Track, this report will focus only on the new crack resistance index parameter, β .

Table 7 summarizes those results for each of the seven mixes prepared and conditioned as outlined previously. The results shown are based on a minimum of four replicates after the outlier analysis. In most cases, five replicates were tested. For the β parameter, a lower number indicates better cracking resistance. Texas DOT recently established a preliminary BMD criterion for maximum cracking progression rate (i.e. β) of 0.45 applicable to most mix types including SMA, Superpave, and certain other dense-graded mixtures. A lower maximum criterion of 0.40 has been proposed for crack attenuating mixtures (CAM) and thin overlay mixtures (TOM). These preliminary criteria are applicable to short-term conditioned lab-produced mix and reheated plant-mix samples.

An examination of the results in Table 7 indicates that the general trends are logical. Within each set of mix sample type and aging condition, the β for mix N8 is the highest and the β for mix S13 is the lowest. For each mixture, the β increased after critical aging, except for the plant mix samples (N8).

Table 7. Results of Texas Overlay Tests Using the β Parameter

Test Section and Mixture Description	LMLC-STOA		LMLC-CA		PMLC-RH		PMLC-CA	
	Avg.	COV	Avg.	COV	Avg.	COV	Avg.	COV
N1: Control	1.00	9%	2.34	29%	0.88	32%	2.08	16%
N2: Control, Higher Density	0.84	28%	2.04	35%	0.60	16%	2.03	11%
N5: Control, Low Dens. & AC	0.90	25%	2.38	20%	0.85	7%	2.96	11%
N8: Control + 5% RAS	2.31	22%	3.25	5%	3.54	10%	3.43	4%
S5: 35% RAP, PG 58-28	0.55	12%	2.04	19%	0.60	14%	1.54	22%
S6: Control, HiMA Binder	0.42	3%	0.89	22%	0.95	38%	1.07	13%
S13: Gap-gr., asphalt-rubber	0.40	20%	0.57	13%	0.32	2%	0.48	9%

Figure 18 shows a bar chart for the OT-TX β results for the critically-aged plant mix samples. The whiskers represent plus and minus one standard deviation of the β parameter from the replicate tests. The chart is divided by the cracking performance of the test sections. It is apparent that the OT-TX β parameter does a very good job of distinguishing the cracking resistance of the mixtures. For this study, a β value of about 1.75 appears to be a good criterion to separate excellent cracking resistance (i.e. little to no cracking on the Test Track) from moderate cracking resistance of the critically-aged PMLC results.

The ANOVA indicated that mixtures had a statistically significant effect on the β parameter. The Games-Howell post-hoc pairwise comparison was then conducted to determine which mixtures were statistically different from one another. The letters (A, B, C, D, E) down the middle of the chart indicate the results of the Games-Howell comparisons. Mixtures that do not share a letter are significantly different at a 95% confidence level. From this analysis, the gap-graded, asphalt rubber mix in S13 was superior to all other mixtures. Statistically, the β parameter separated the mixtures into five groups, although there was overlap among most of the groups. This indicates that the OT-TX β is a powerful test in discerning cracking resistance of asphalt mixtures.

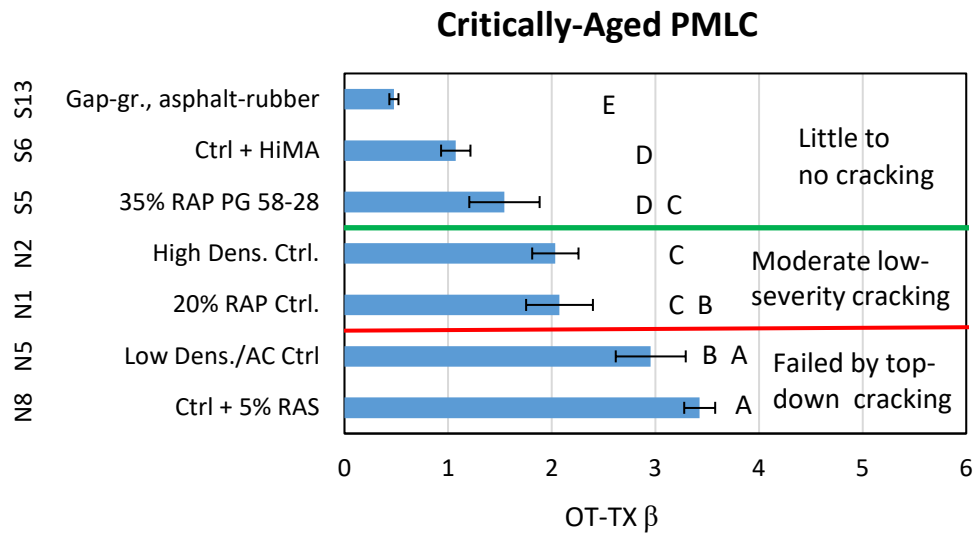


Figure 18. Chart of Statistical Comparisons of OT-TX β among Mixtures with Performance Groupings

Figure 19 shows correlation plots of OT-TX β results versus the observed cracking on the Test Track for the lab-prepared and plant-produced samples after the different aging conditions. These relatively high R^2 values for each of the four plots indicate that the OT-TX β is a very good indicator of top-down cracking resistance for asphalt mixtures. It is logical to expect the critically-aged results to better correlate with field performance, since the binders in those samples are aged to represent the condition of the binders after about five years of field aging for surface layers on the Test Track. For this test, the coefficients of determination are similar for the critically-aged samples and the corresponding reheated or STOA samples, which indicates that even the results after reheated plant mix or STOA lab mix correlates as well with the field performances as the critically-aged samples.

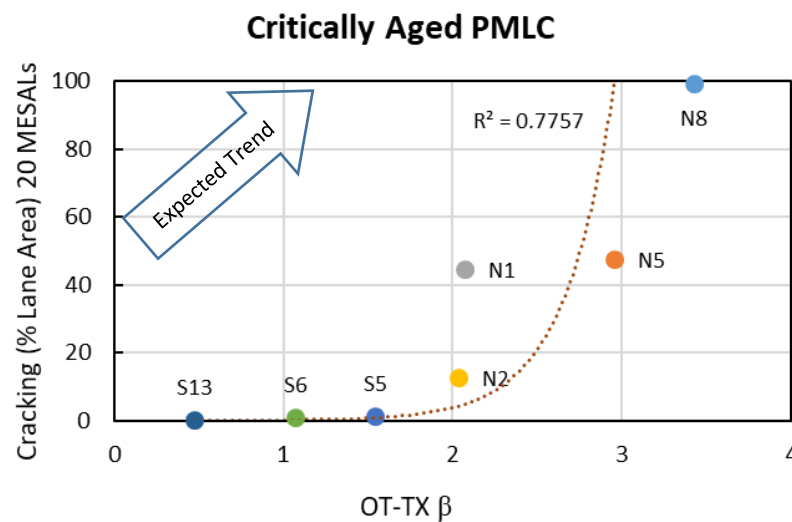
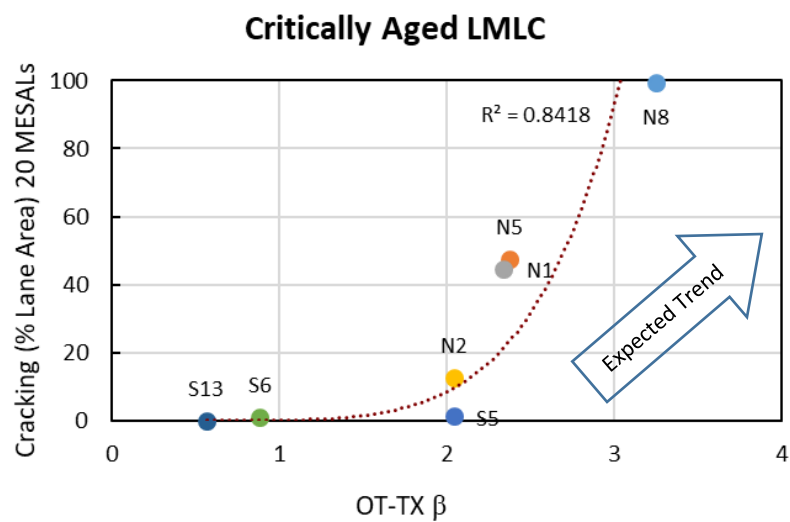
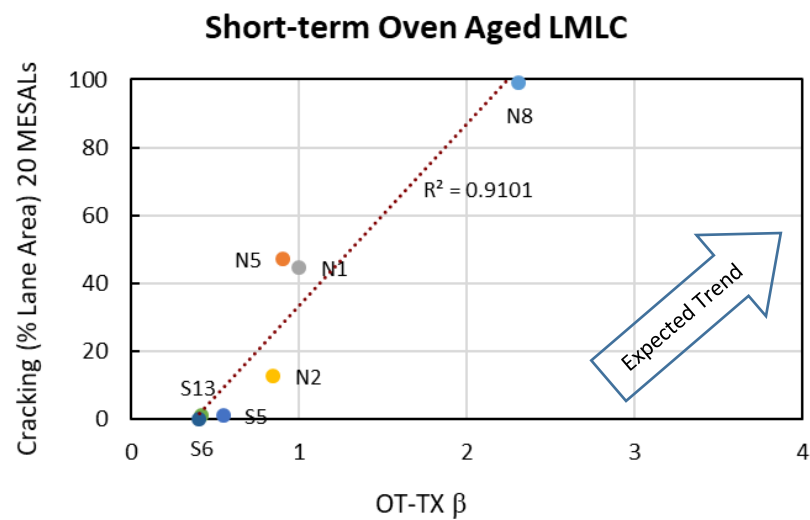
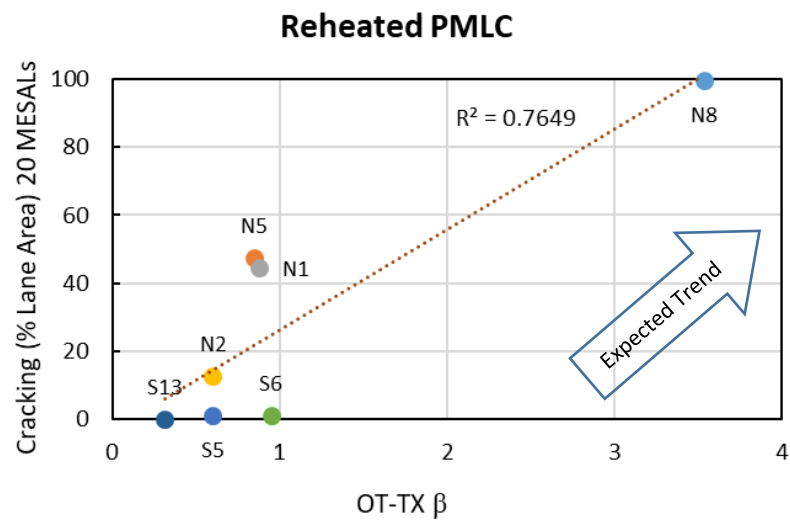


Figure 19. Correlations of OT-TX β with Field Performance for the Lab and Plant Samples Subject to Different Aging Conditions

2.6.3 NCAT Overlay Test Results

Table 8 summarizes the results of the OT-NCAT procedure also using the β parameter. For this test, a minimum of three replicates were tested. The variability of the β values from the OT-NCAT method compared to the OT-TX method is generally lower. The average COVs of the OT-TX results for all four sets ranged from 12 to 20% compared to 6 to 14% for the OT-NCAT method.

For the LMLC-STOA set, very similar OT-NCAT β values were obtained for four mixtures (N2, S5, S6, and S13). The same is true for the PMLC-RH set although the group was a little different (N1, N2, S6, and S13). However, after critical aging, the β values increased and the differences between mixes increased. Thus, for the OT-NCAT β parameter, the critical aging of loose mix is helpful in making distinctions among mixtures with different field performance.

Table 8. Results of the OT-NCAT Tests Using the Beta Parameter

Test Section and Mixture Description	LMLC-STOA		LMLC-CA		PMLC-RH		PMLC-CA	
	Avg.	COV	Avg.	COV	Avg.	COV	Avg.	COV
N1: Control	0.27	2%	0.48	18%	0.27	9%	0.50	14%
N2: Control, Higher Density	0.24	9%	0.42	12%	0.24	8%	0.41	9%
N5: Control, Low Dens. & AC	0.33	10%	0.66	9%	0.32	4%	0.60	16%
N8: Control + 5% RAS	0.41	4%	1.15	35%	0.60	8%	0.95	29%
S5: 35% RAP, PG 58-28	0.25	2%	0.52	10%	0.29	4%	0.33	6%
S6: Control, HiMA Binder	0.25	5%	0.29	11%	0.26	12%	0.27	9%
S13: Gap-gr., asphalt-rubber	0.24	8%	0.25	8%	0.24	7%	0.26	6%

Figure 20 shows a bar chart of OT-NCAT β results for the critically-aged plant mix samples. The chart is divided by the cracking performance of the test sections. From this chart, it can be seen that the OT-NCAT β parameter correctly ranked and provided some differentiation among the mixtures that corresponded to their field cracking performance.

As with the OT-TX method, the ANOVA indicated that mixtures had a statistically significant effect on the β parameter. The Games-Howell post-hoc pairwise comparison determined which mixtures were statistically different from one another. The letters in the middle of the chart indicate the results of the Games-Howell comparisons; mixtures that do not share a letter are significantly different at a 95% confidence level. The Games-Howell analysis separated the mixtures into four statistical groups, although there was some overlap among the groups. Although no single mixture was distinct from the others, this analysis indicates that the OT-NCAT β is able to distinguish among the top-down cracking susceptibilities of the mixtures in this study. A β value between 0.41 and 0.33 appears to separate the mixtures with good cracking resistance from those with moderate cracking resistance for the critically-aged plant mix samples. The midpoint of this range is 0.37 and is suggested as a preliminary maximum criterion for β based on critically-aged PMLC results of this experiment. Further research, including benchmarking studies and additional field validation studies, should be considered before implementation of any criteria.

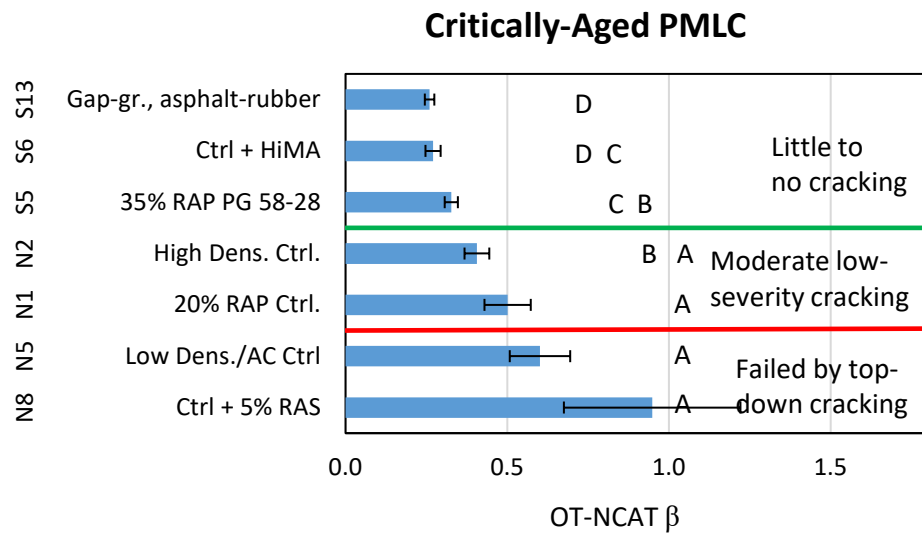


Figure 20. Statistical Comparisons of OT-NCAT β among Mixtures with Performance Groupings

Figure 21 shows correlation plots of OT-NCAT β results versus the observed top-down cracking on the Test Track for the four sample preparation and aging condition sets. As with the OT-TX method, the high R^2 values indicate that the OT-NCAT β is a very good indicator of propensity for top-down cracking. The PMLC-CA correlation to field performance has the highest R^2 value of all tests in this study.

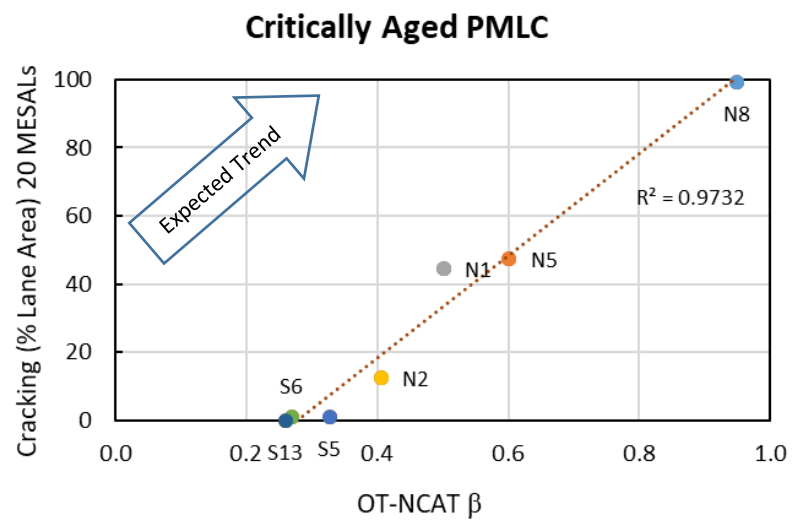
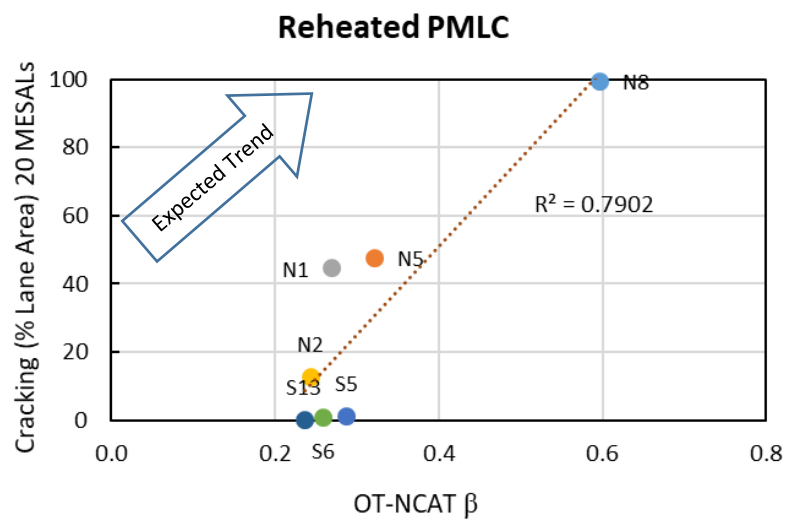
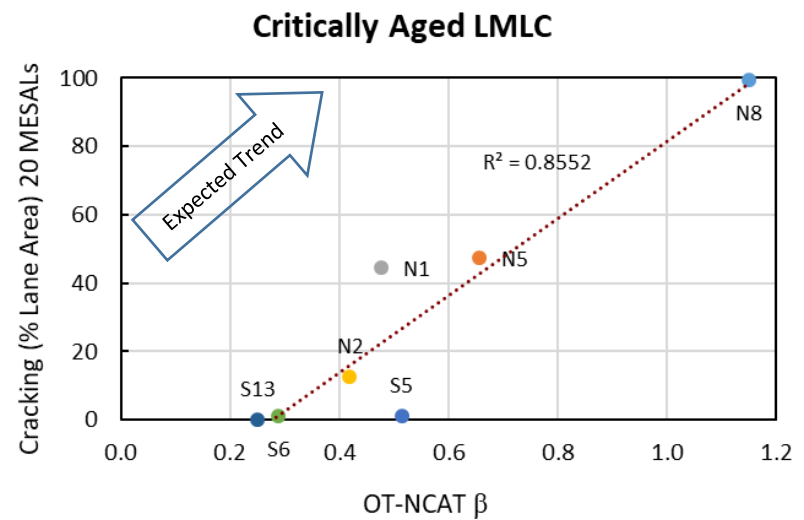
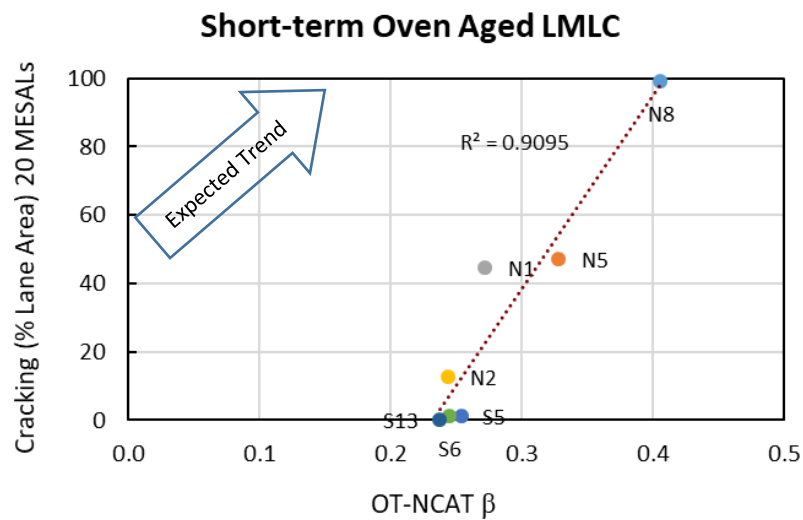


Figure 21. Correlations of OT-NCAT β with Field Performance for the Lab and Plant Samples Subject to Different Aging Conditions

2.6.4 Semi-circular Bend Test (Louisiana Method) Results

Table 9 shows results of the critical strain energy release rate, J_c , the key test output for the Louisiana SCB test. For this test, at least four replicates were tested at each of the three notch depths. The critical strain energy release rate, also referred to as J-integral, is calculated as shown in Equation 2.

$$J_c = -\left(\frac{1}{b}\right) \frac{dU}{da} \quad (2)$$

Where:

J_c is the critical strain energy release rate (kJ/m²);

b is the average sample thickness (mm);

a is the notch depth (mm);

U is the strain energy to failure (N-mm); and

dU/da is the change of strain energy with notch depth, or more simply, the slope of the best fit linear regression between notch depth and the strain energy to failure.

Since the regression is fit through all of the data, the standard error of the slope was used to estimate the standard deviation of the slope by dividing the estimate of the total regression error (Se) by the sum of squared differences between the x values (Sxx); in this case, notch depths. Since the average thickness of specimens is a constant, the slope is the only variable in the J_c equation. Thus, the estimated standard deviation of the slope multiplied by the average thickness was used to estimate the variability of J_c results. A more detailed explanation of this analysis is provided by Moore (13). The COVs reported in Table 11 are the estimated standard deviation divided by the average J_c .

Table 9. SCB Critical Strain Energy Release Rate, J_c , and the Estimated COV

Test Section and Mixture Description	LMLC-STOA		LMLC-CA		PMLC-RH		PMLC-CA	
	Avg.	COV	Avg.	COV	Avg.	COV	Avg.	COV
N1: Control	0.29	27%	0.38	6%	0.36	46%	0.34	9%
N2: Control, Higher Density	0.54	10%	0.45	9%	0.61	19%	0.27	12%
N5: Control, Low Dens. & AC	0.32	15%	0.26	13%	0.34	53%	0.22	8%
N8: Control + 5% RAS	0.23	15%	0.23	12%	0.39	33%	0.25	9%
S5: 35% RAP, PG 58-28	0.30	9%	0.27	16%	0.34	50%	0.28	10%
S6: Control, HiMA Binder	0.23	30%	0.34	16%	0.37	21%	0.32	9%
S13: Gap-gr., asphalt-rubber	0.61	17%	0.56	15%	0.51	57%	0.57	25%

As can be seen, critical aging had an inconsistent effect on J_c for lab-prepared mixtures, but generally decreased the J_c results, as expected, for the plant-produced mixtures.

Since 2016, the Louisiana Department of Transportation and Development has required a minimum J_c value of 0.6 kJ/m² for high traffic mixtures and 0.5 kJ/m² for medium to low traffic mixtures. These criteria are based on compacted specimens that have been long-term oven aged for five days at 85°C in accordance with AASHTO R30. Although the long-term aging in R30 and the critical aging procedure used in this study are likely to provide different aging effects on the mixtures, only the S13 mixture met the lower J_c criteria for the LMLC-CA and PMLC-CA sets.

This mixture, as well as those used in S5 and S6, performed very well under the extremely high traffic loading on the Test Track.

Figure 22 shows a bar chart for the J_c results of the critically-aged plant mix samples. The whiskers in this chart are the estimated standard deviations of J_c as described previously. Since these standard deviations were not determined in the standard way, ANOVA and pairwise comparison analyses were not conducted for these results. From this chart, it appears that J_c is able to distinguish the S13 mixture as superior and the two mixes that failed first from the other mixtures, but the overall rankings are not consistent with the field cracking data from the Test Track.

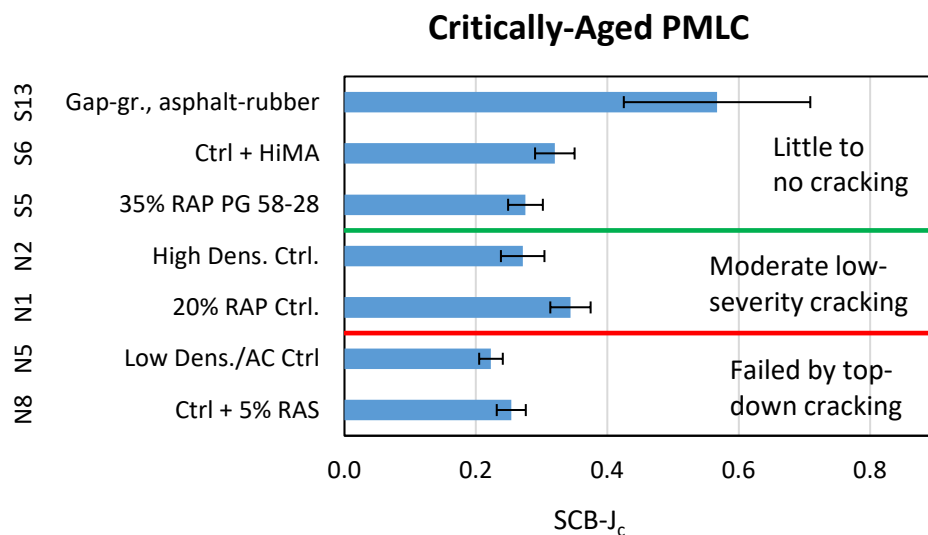


Figure 22. Chart of SCB- J_c Results and Field Performance Groupings

Figure 23 shows correlation plots of J_c results versus the observed top-down cracking for the four sample preparation and aging condition sets. Only the critically-aged PMLC set has a reasonably strong correlation between J_c and field performance. However, even for this set of data, some results are hard to rationalize. For example, N2 and N8 have similar J_c values (0.27 and 0.28, respectively), but these test sections had dramatically different top-down cracking performance on the Test Track.

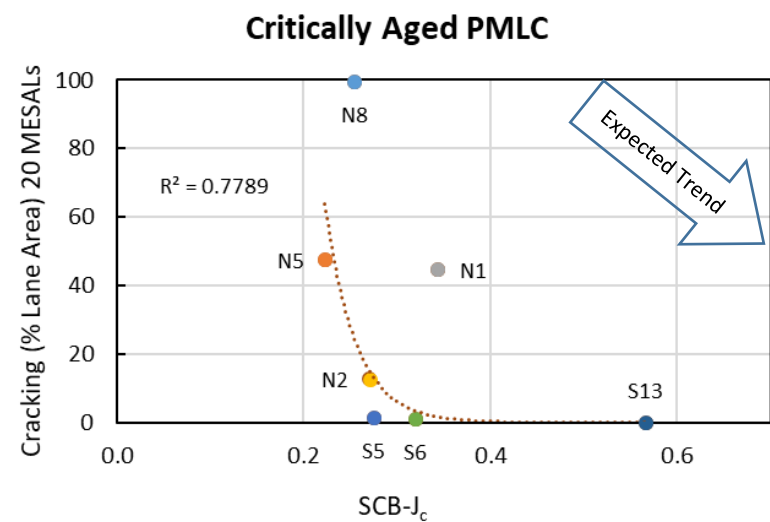
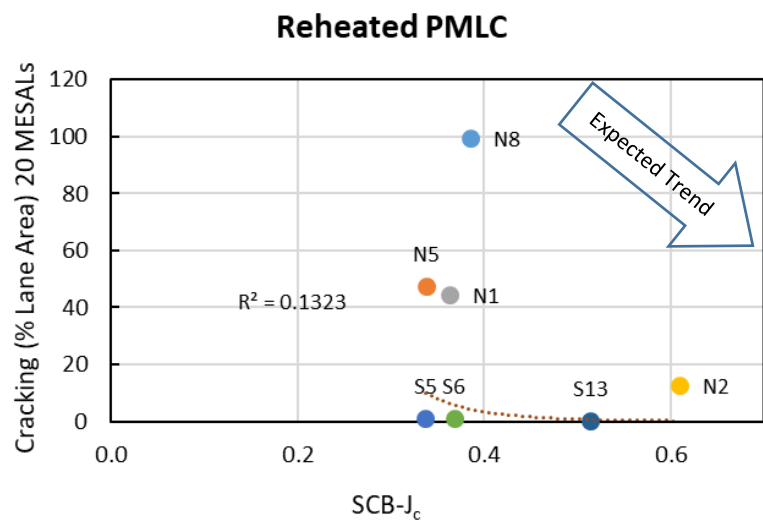
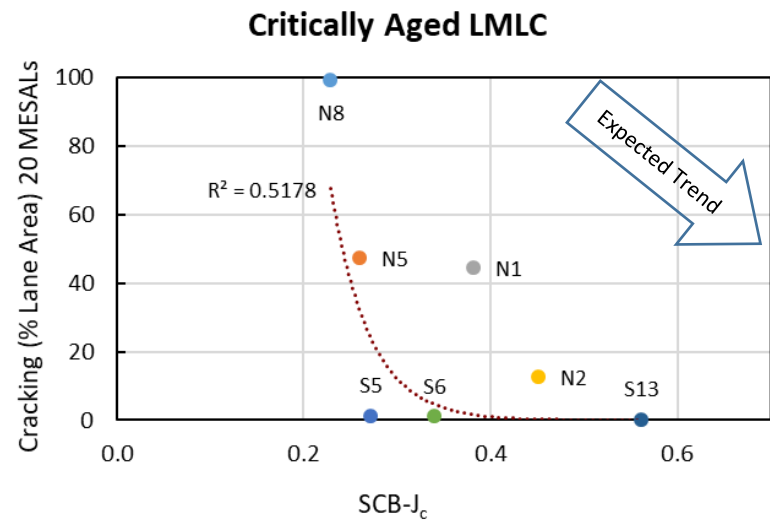
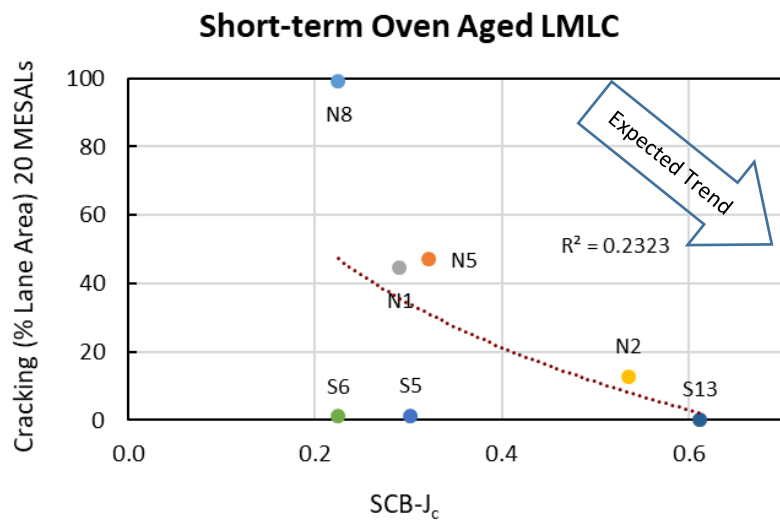


Figure 23. Correlations of SCB- J_c with Field Performance for the Lab and Plant Samples Subject to Different Aging Conditions

2.6.5 Illinois Flexibility Index Test Results

Table 10 summarizes the results of the I-FIT test. These results are based on a minimum of five replicates. According to the test, a higher flexibility index indicates better crack resistance. Overall, these results follow the expected trends within each sample type and conditioning method set. There appears to be a good spread in the average results, but the COVs are high (i.e. above 40%) for some mixtures. As expected, critical aging causes a substantial decrease in FI results. The impact of specimen air void contents in FI results can be seen by comparing the two sets of results for N2 and N5. For N2, FI results were slightly higher for specimens at 4% air voids compared to 7% air voids for STOA and reheated mix samples, but were lower after critical aging. For N5, FI results were substantially higher for specimens at 10% air voids compared to 7% air voids for each sample type and aging condition. Overall, the effect of specimen air void content on FI results is counterintuitive, that is the results seem to indicate that a lower density improves cracking resistance. That is clearly not the case as evident in the field cracking performance of N1 compared to N2.

Table 10. Results of IFIT Tests

Test Section and Mixture Description	LMLC-STOA		LMLC-CA		PMLC-RH		PMLC-CA	
	Avg.	COV	Avg.	COV	Avg.	COV	Avg.	COV
N1: Control	4.16	23%	0.63	50%	3.58	8%	0.59	51%
N2: Ctrl, Higher Dens. 7% Va	2.24	21%	0.25	76%	1.46	25%	1.38	74%
N2: Ctrl, Higher Dens. 4% Va	2.65	31%	0.10	76%	1.86	13%	0.10	67%
N5: Ctrl, Low Dens. & AC 7% Va	1.37	13%	0.21	53%	1.34	16%	0.67	93%
N5: Ctrl, Low Dens. & AC 10% Va	4.02	18%	0.74	34%	2.69	29%	0.80	35%
N8: Control + 5% RAS	0.43	44%	0.03	71%	0.39	18%	0.07	68%
S5: 35% RAP, PG 58-28	5.21	22%	0.70	30%	6.27	10%	1.79	16%
S6: Control, HiMA Binder	14.68	24%	3.43	20%	4.53	6%	3.77	16%
S13: Gap-gr., asphalt-rubber	15.12	34%	5.15	21%	10.40	42%	4.34	18%

The Illinois Department of Transportation (IDOT) currently requires one hour of short-term aging for mixtures containing low absorption aggregates and two hours of short-term aging for mixtures containing high absorption aggregates. Additionally, for surface mixtures only, IDOT uses a long-term aging protocol of three days at 95°C after I-FIT specimens are cut and notched. IDOT's current criteria for FI for dense-graded mixtures is a minimum of 8.0 after short-term aging, and a minimum of 5.0 after long-term aging. For SMA, the minimum FI is 16.0 after short-term aging, and 10.0 after long term aging. IDOT also has a separate criterion for a 4.75 mm mixture used for crack relief layers. Of the LMLC-STOA set of dense-graded Superpave mixtures, only the mixture from S6 exceeds the IDOT minimum criterion of 8.0 after short-term aging, but it does not meet the minimum FI of 5.0 after long-term aging. The mixture from S13 is similar to an SMA but it does not meet IDOT's minimum SMA criterion of 16.0 after short-term aging or the minimum FI of 10.0 after long-term aging. However, comparisons of these results with the Illinois criteria must be taken with caution since the aging procedures are different. A few other states have different preliminary criteria for flexibility index.

Figure 24 shows a bar chart for the FI results of the critically-aged plant mix samples. The whiskers represent plus and minus one standard deviation of FI. It is clear that FI does a good job of ranking the mixtures according to their top-down cracking performance.

The ANOVA indicated that mixtures had statistically different FI results and the Games-Howell post-hoc pairwise comparison determined which mixtures were statistically different from one another. It can be seen from the pairwise comparisons in Figure 24 that FI did not statistically distinguish the moderately performing mixtures from the poor performing mixtures due partly to the relatively high variability of some FI results. Also, the mixture from S5, which performed very well on the Test Track, had FI results that were not statistically grouped with the other good performing mixtures (i.e., S6 and S13). Setting a preliminary FI criterion between the results for S5 and N2 seems reasonable; therefore, a minimum FI threshold of 1.5 for critically-aged PMLC mixtures should provide good resistance to top-down cracking. Further research, including benchmarking studies and additional field validation studies, should be considered before implementation of any criteria.

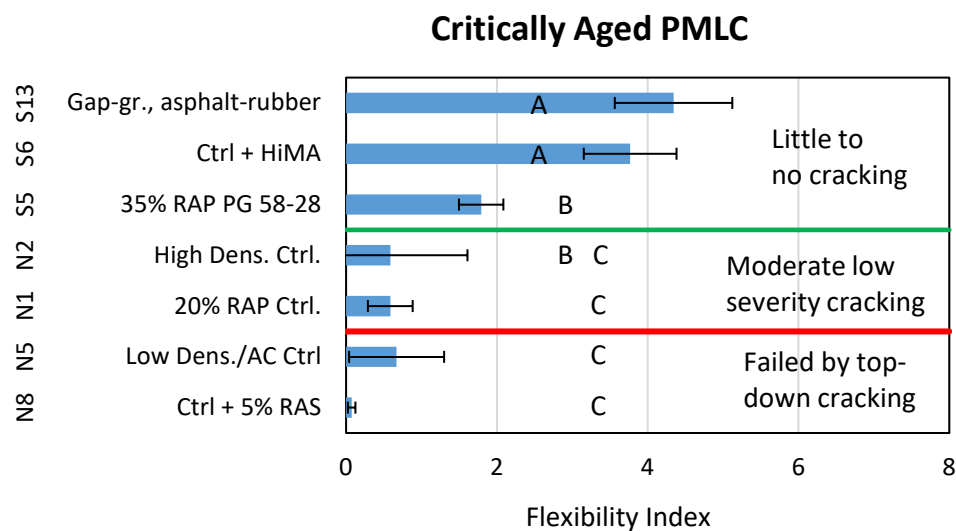


Figure 24. Chart of Statistical Comparisons of FI among Mixtures with Performance Groupings

Figure 25 shows the correlation plots of FI versus the top-down cracking observed on the Test Track for the four sample preparation and aging condition sets. Each chart shows a strong correlation between FI and the amount of top-down cracking at the conclusion of the experiment. These results indicate that FI is a good indicator of top-down cracking resistance.

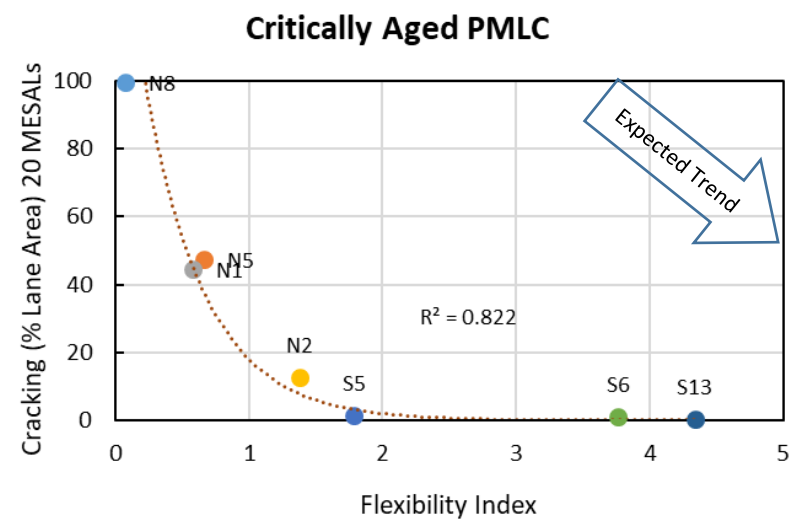
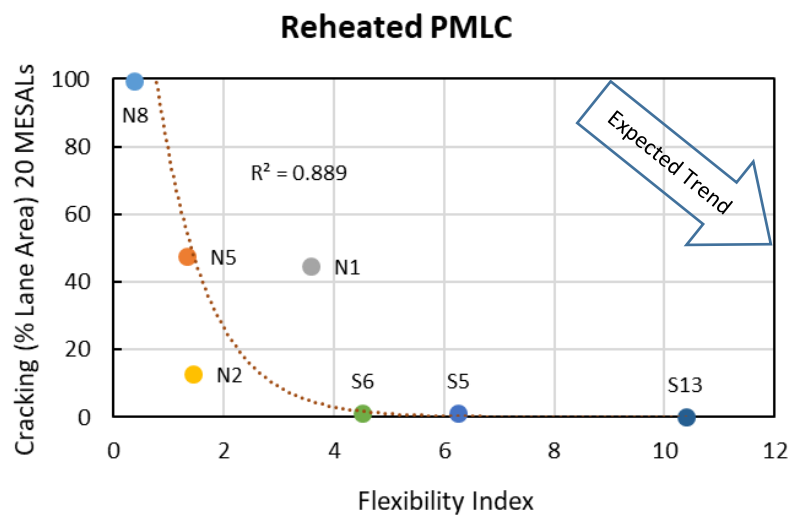
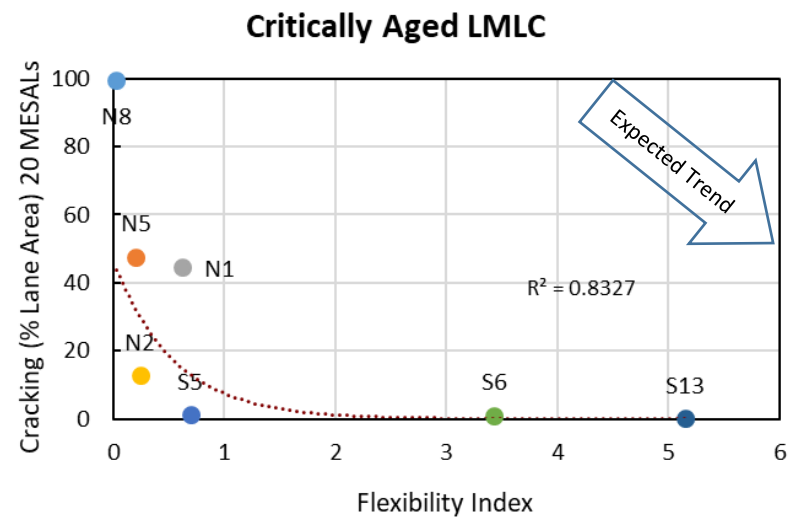
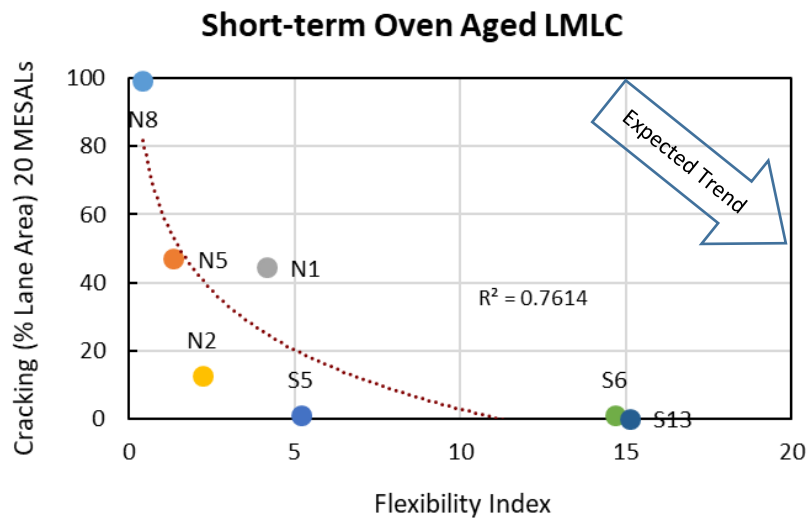


Figure 25. Correlations of FI with Field Performance for the Lab and Plant Samples Subject to Different Aging Conditions

2.6.6 IDEAL Cracking Test Results

Table 11 summarizes the results of the IDEAL-CT. These results are based on a minimum of five replicates. For this test, a high CT_{Index} indicates better crack resistance. As expected, CT_{Index} results for each mixture decreased substantially with the critical aging procedure. It can also be seen that the CT_{Index} results for the LMLC-CA set are similar to the corresponding results of the PMLC-CA, set which indicates that the lab preparation of the mixtures provides similar results as plant-produced mixtures.

As with the FI results, the effect of specimen air void contents on CT_{Index} can be seen by comparing the two sets of results for N2 and N5. For N2, CT_{Index} results were substantially lower for specimens at 4% air voids compared to 7% air voids for each sample type and aging condition. For N5, CT_{Index} were substantially higher for specimens at 10% air voids compared to 7% air voids for. Overall, the effect of specimen air void contents on CT_{Index} results is counterintuitive, that is the results seem to indicate that a lower density improves cracking resistance. However, the field cracking performance of N1 compared to N2 shows that is not true.

Table 11. Summary of CT_{Index} Results

Test Section and Mixture Description	LMLC-STOA		LMLC-CA		PMLC-RH		PMLC-CA	
	Avg.	COV	Avg.	COV	Avg.	COV	Avg.	COV
N1: Control	30.2	10%	7.3	13%	26.2	21%	8.8	9%
N2: Ctrl, Higher Dens. 7% Va	27.2	9%	10.3	17%	20.7	10%	10.8	18%
N2: Ctrl, Higher Dens. 4% Va	13.9	10%	6.1	13%	13.2	14%	5.1	18%
N5: Ctrl, Low Dens. & AC, 7% Va	19.2	7%	6.5	17%	15.9	14%	7.6	12%
N5: Ctrl, Low Dens. & AC, 10% Va	33.2	13%	11.8	11%	23.8	21%	8.6	12%
N8: Control + 5% RAS	10.9	23%	2.8	28%	6.7	30%	2.4	23%
S5: 35% RAP, PG 58-28	41.6	17%	10.7	17%	32.4	15%	16.3	9%
S6: Control, HiMA Binder	80.8	16%	22.2	22%	32.9	11%	18.7	20%
S13: Gap-gr., asphalt-rubber	133.1	27%	63.4	19%	208.1	49%	68.4	19%

Figure 26 shows the bar chart for the CT_{Index} results of the critically-aged plant mix samples. The CT_{Index} ranking of the mixtures is consistent with the field performance of the test sections. For the critically- aged set of PMLC samples, a CT_{Index} criterion of 15 appears to discriminate good performing mixtures from moderately performing mixtures based on their top-down cracking performance on the Test Track.

The ANOVA indicated that some mixtures had statistically different CT_{Index} results. The Games-Howell post-hoc pairwise comparison determined which mixtures were statistically different from one another as indicated by the letters down the middle of the chart. Mixtures that do not share a letter are significantly different at a 95% confidence level. It can be seen that the CT_{Index} result for the mixture from S13 is superior to all other mixtures and that the mixture from N8 is the worst mixture, which is consistent with the top-down cracking performance on the Test Track. The pairwise comparisons of CT_{Index} also statistically distinguished the mixtures with little to no cracking from the mixtures having a moderate amount of low severity cracking. It is important to note that none of the Games-Howell groupings overlapped, which further indicates that CT_{Index} is effective in distinguishing the cracking resistance of mixtures.

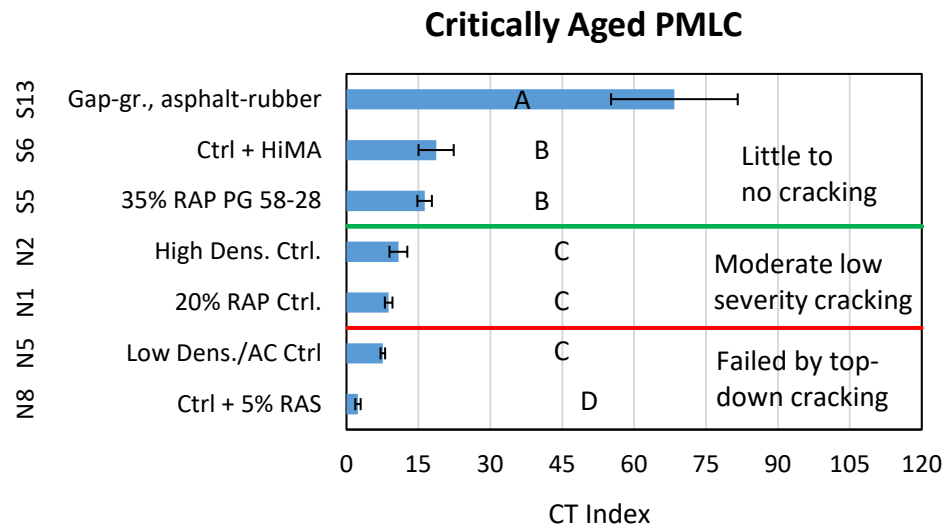


Figure 26. Chart of Statistical Comparisons of CT_{Index} among Mixtures with Performance Groupings

Figure 27 shows the correlations of CT_{Index} with the top-down cracking observed on the Test Track for the four sample preparation and aging condition sets. Each chart shows a strong correlation between the CT_{Index} and top-down cracking at the conclusion of the experiment. These results indicate that CT_{Index} is a good indicator of top-down cracking resistance.

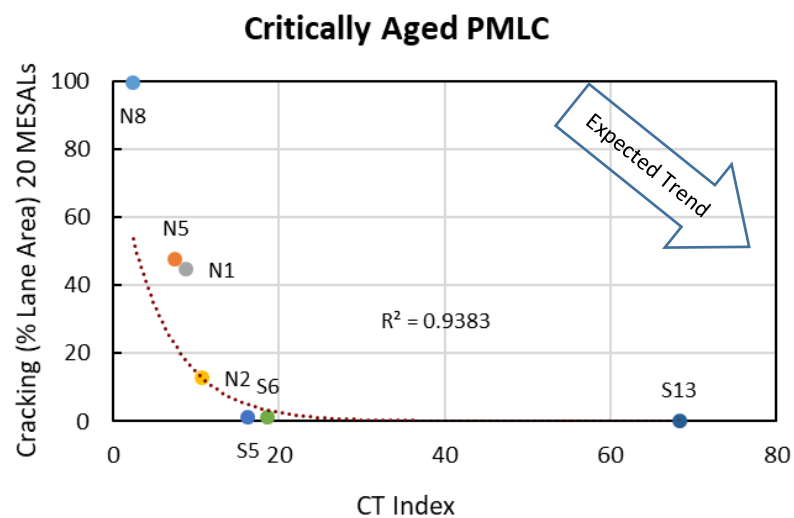
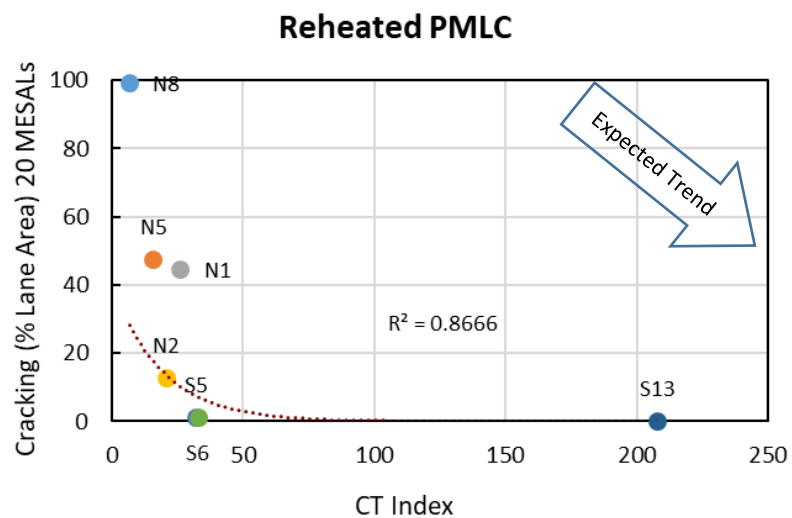
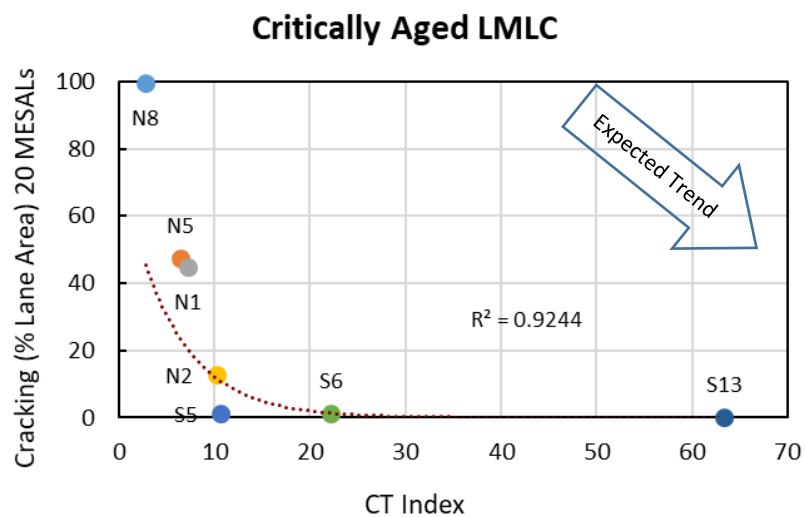
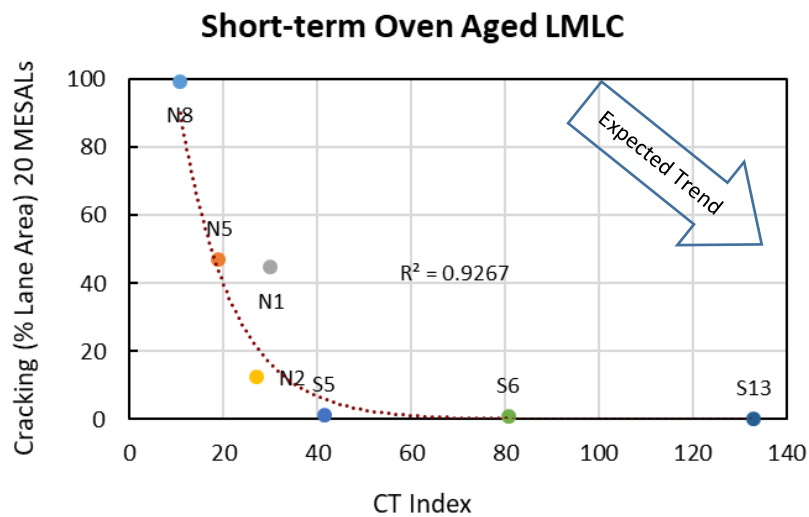


Figure 27. Correlations of CT_{Index} with Field Performance for the Lab and Plant Samples Subject to Different Aging Conditions

2.6.7 AMPT Cyclic Fatigue Test

Table 12 shows the results for the AMPT cyclic fatigue test fatigue cracking index parameter, S_{app} . The cyclic fatigue test was added to the experimental plan after the work on lab-prepared mixtures was completed, so the test was only conducted on plant-produced samples. Three replicates were tested for each of the seven plant-produced mixture at the reheated condition and the critically aged condition.

The S_{app} index parameter was developed by North Carolina State University (NCSU) to include the effects of a mixture's modulus and toughness on the amount of fatigue damage the material can tolerate under loading. The S_{app} value is determined at the average of the high-temperature and low-temperature binder performance grade from LTPPBind Online for the project of interest minus three degrees Celsius. The AMPT test data are processed through the *FlexMAT*[™] program to calculate the representative S_{app} value. Higher S_{app} values indicate better fatigue resistance of the mixture.

Table 12. Summary of Fatigue Index Parameter, S_{app} , from AMPT Cyclic Fatigue Tests

Test Section and Mixture Description	PMLC			
	Reheated		Critically-Aged	
	Representative S_{app}	COV of S_{app}	Representative S_{app}	COV of S_{app}
N1: Control	26.46	20%	18.13	15%
N2: Control, Higher Density	35.38	8%	24.83	4%
N5: Control, Low Dens. & AC	26.49	5%	15.29	24%
N8: Control + 5% RAS	7.58	27%	6.90	9%
S5: 35% RAP, PG 58-28	43.95	32%	47.23	10%
S6: Control, HiMA Binder	53.83	38%	52.29	14%
S13: Gap-gr., asphalt-rubber*	32.54	4%	29.20	11%

*Results reported are from NCSU analysis

NCAT conducted the cyclic fatigue tests in accordance with AASHTO TP 107 and analyzed the data using *FlexMAT*[™]. Results for the S13 mixture were lower than expected, so NCSU analyzed the data and found that the mixture had a second peak in the phase angle vs. C (or modulus) plot. They noted that this was not surprising based on findings in literature that asphalt-rubber mixtures are a two-phase system and exhibit two peaks in the stress-strain plots under tensile loading. None of the other mixtures had the second peak. The results shown in Table 12 for S13 are the S_{app} values determined by NCSU.

Figure 28 shows the bar chart for S_{app} results for the critically-aged set of samples. Note that the order of the mixes in this chart differs from the other bar charts so that the Games-Howell groupings could be shown in a logical way. Results for S13 were moved down in the ranking to third place based on its representative S_{app} value. From this chart, it appears that S_{app} correctly ranks the mixtures in terms of their top-down cracking performance on the Test Track, with the possible exception of the gap-graded, asphalt-rubber mixture in S13 which had no cracking. The five Games-Howell groupings are not consistent with the three field performance groupings.

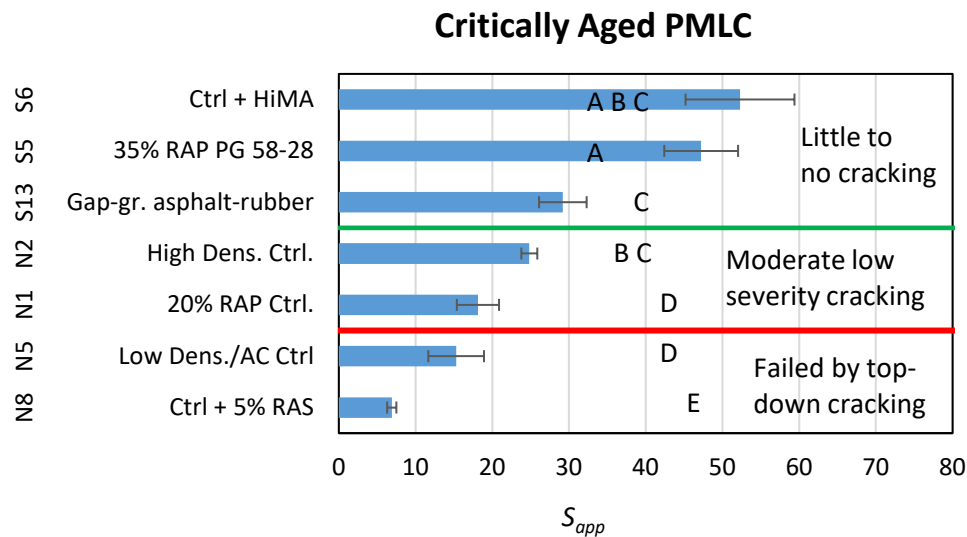


Figure 28. Chart of Comparisons of S_{app} among Mixtures with Performance Groupings

Table 13 shows NCSU's recommended S_{app} criteria for short-term aged mixtures based on the field performance of 105 asphalt mixtures with a range of RAP contents, binder grades, and warm mix asphalt technologies (12). The NCAT Test Track would be considered Very Heavy traffic in this table. The field performance of the Test Track Cracking Group Experiment appears to support the recommended criteria for the Very Heavy traffic level as the S_{app} criterion of 30 separates the test sections with little to no cracking from the test sections with moderate low-severity top down-cracking. The exception is the gap-graded, asphalt rubber mixture from S13 with a S_{app} of 29.2 that marginally failed the minimum criterion of 30. This indicates that the Cyclic Fatigue test or its criteria may need to be adjusted for this mixture type.

Table 13. Recommended Criteria for S_{app} Fatigue Index Parameter

Traffic (million ESALs)	Traffic Tier	S_{app} Limits
Less than 10	Standard	$S_{app} > 8$
Between 10 and 30	Heavy	$S_{app} > 24$
Greater than 30	Very Heavy	$S_{app} > 30$
Greater than 30 and slow traffic	Extremely Heavy	$S_{app} > 36$

Figure 29 shows the correlation of S_{app} values with the observed top-down cracking from the Test Track. For both the reheated and critically-aged sets, the plots show strong correlations between the cyclic fatigue test index parameter and top-down cracking performance. These results indicate that S_{app} is a good indicator of top-down cracking resistance.

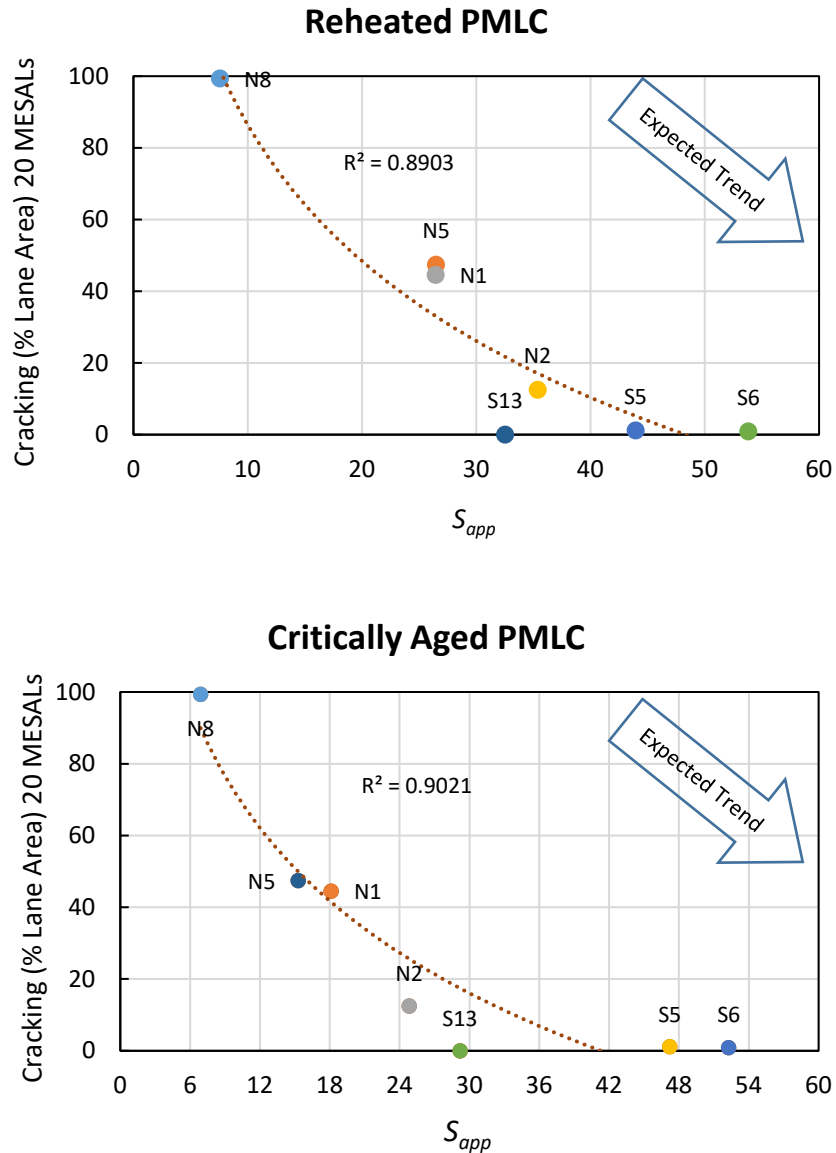


Figure 29. Correlations of S_{app} with Field Performance based on Reheated (Top) and Critically-Aged (Bottom) PMLC Samples

2.7 Summary of Analyses for the Cracking Tests

A summary of important statistical measures from the analyses of the top-down cracking tests evaluated in this study is provided in Table 14. The average COV is the mean within-lab COV from all seven mixtures for each of the four sample preparation and aging condition sets, except for the AMPT cyclic fatigue which only had reheated and critically-aged PMLC results. This is an indication of the within-lab variability, or repeatability, of the tests. All of these COVs are very consistent with those reported in published literature (13). The test with the lowest COV is the NCAT version of the overlay test, β parameter. The test with the highest COV is the I-FIT. The second column indicates the number of statistical groupings from the post-hoc Games-

Howell pairwise comparison analysis. A higher number of groupings is an indication of the test's ability to provide statistical distinctions among different mixtures. The Texas-OT method β parameter and the AMPT cyclic fatigue S_{app} parameter have the highest number of groupings. Although, the CT_{Index} had one less grouping than OT-TX β and Cyclic Fatigue S_{app} comparisons, no mixture had CT_{Index} results that belonged to more than one statistical group. This absence of overlapping groupings is also a strong indicator of the ability of the IDEAL-CT to provide results that are discerning among different mixtures. The last column in Table 14 provides the range of coefficients of determination, R^2 , for the correlations with top-down cracking field performance for the four sample preparation and conditioning sets. The tests that consistently provided R^2 values above 0.85 were the IDEAL-CT test and the AAMPT Cyclic Fatigue test, with only two correlations available for the Cyclic Fatigue test. Other tests that also provided strong lab to field correlations were the OT-TX, OT-NCAT and I-FIT.

Table 14 Summary of Statistical Measures of Top-Down Cracking Tests

Test and Parameter	Average COV	Games Howell Groups	Range of R^2
Energy Ratio, ER	Not available	Not applicable	0.03 to 0.28
Texas Overlay Test, β	17%	5	0.76 to 0.91
NCAT Overlay Test, β	10%	4	0.79 to 0.97
Louisiana SCB, J_c	20%	Not applicable	0.13 to 0.78
Illinois Flexibility Index Test, FI	34%	3	0.76 to 0.89
IDEAL Cracking Test, CT_{Index}	18%	4	0.87 to 0.94
AMPT Cyclic Fatigue, S_{app}	16%	5	0.89 to 0.90

2.8 Conclusions

After four and a half years of in-service aging and trafficking there was a very good spread in the observed top-down cracking of the seven test sections in the Cracking Group experiment, illustrating that surface mixture properties are a key factor in top-down cracking performance.

Based on the analysis of the seven cracking tests evaluated in this experiment, the following conclusions are provided:

- The energy ratio results failed to match the field performance for top-down cracking. This test also lacks practicality for routine use due to the complexity and time to complete the three parts of the test, so it is not recommended for implementation.
- The Texas overlay test β parameter is a very good indicator of a mixture's resistance to top-down cracking. It is a much more discerning indicator than cycles to failure. The effect of air voids on test results is not counterintuitive as with some other cracking tests. For OT-TX tests on critically-aged mixtures, a β value of about 1.75 appears to separate asphalt mixtures with good from moderate cracking resistance. The key disadvantages of this test are the time required to prepare specimens and cost of the equipment. For these reasons, it is not a practical cracking test for day-to-day use in BMD and testing for quality assurance.
- The NCAT-modified version of the overlay test is also a very good indicator of resistance to top-down cracking. A maximum β value of 0.37 is recommended as a preliminary criterion for critically-aged mixtures based on the PMLC results of this experiment. The NCAT version of the OT has a lower coefficient of variation than the Texas procedure

and the testing time is faster. Like the Texas version, the effect of air voids on the results of the NCAT-OT method is not counterintuitive, but it suffers from the same disadvantages of tedious sample preparation and high equipment cost.

- The results of the Louisiana SCB test did not adequately distinguish the top-down cracking resistance of the mixtures in this experiment. Two of the mixtures that performed very well on the NCAT test Track had J_c results very similar to those of mixtures that had moderate low-severity cracking. Other disadvantages of this test are that it does not lend itself to standard methods of variability analysis, as well as the time and cost of preparing the notched semi-circular specimens.
- The flexibility index from the I-FIT procedure does a fair job of correlating with the top-down cracking field performance of this experiment. However, due to the high variability of FI for several mixtures, the critically-aged PMLC results were not statistically discerning among some good and moderately performing mixtures or between good and poor performing mixtures. Like the Louisiana SCB test, a disadvantage of the I-FIT is the time and cost of preparing the specimens. Another concern about this test is that FI results are affected by specimen air void contents in an incorrect manner. Specimens with lower air voids have lower FI results, which is counter to the field cracking performance as evidenced by the Test Track performance of sections N1 versus N2 in this experiment.
- The CT_{Index} from the IDEAL-CT method is a very good indicator for resistance to top-down cracking. It has strong correlations to the field performance of the NCAT test sections and the results are statistically discernable from mix to mix. For critically-aged samples, a minimum CT_{Index} of 15.0 is recommended as a preliminary criterion for good resistance to top-down cracking. As with the I-FIT procedure, a concern about the IDEAL-CT is that CT_{Index} is affected by specimen air void contents in an incorrect manner. Until this issue is corrected, it is recommended that the test only be conducted on specimens compacted to $7.0 \pm 0.5\%$. The IDEAL-CT method is well suited to everyday use in BMD and testing for quality assurance.
- The AMPT cyclic fatigue test index parameter, S_{app} , correlated very well with the observed top-down cracking in the test sections for this experiment. The results also support NCSU's recommended minimum S_{app} criterion of 30 for short-term aged mixture samples for Very Heavy traffic pavement applications. However, the S_{app} results for the gap-graded, asphalt rubber mixture appear to be lower than what they should be based on excellent field performance on the Test Track and the results of other cracking tests in this study. This may indicate that the cyclic fatigue test or its criteria needs to be adjusted for this mixture type. Significant disadvantages of this test are the time and cost to prepare specimens, cost of the equipment, and complexity of data analysis. For these reasons, it is not well suited for routine use in BMD or quality assurance testing. For further information on equipment costs and time to complete many of the BMD tests, see the Balance Mix Design Resource Guide (13).

The NCAT critical-aging procedure was initially developed and further validated as part of this study and has been detailed elsewhere (2, 13, 14). Key findings from those studies are as follows:

- Top-down cracking occurs after several years of in-service aging as the asphalt binder in the surface layer stiffens and loses relaxation properties. The rate of aging likely differs based on climatic factors and characteristics of the binders, but information from several national studies indicate that top-down cracking begins to be evident at a critical field aging condition of approximately 70,000 cumulative degree days (CDD).
- To evaluate the resistance of surface mixtures to top-down cracking, it is appropriate to first condition surface mixtures using a laboratory aging protocol that results in binder and mixture properties that are similar to those from field aging of approximately 70,000 CDD. Given that there is a significant aging gradient with depth from the surface, it is not believed to be necessary to further condition non-surface mixtures beyond what occurs during plant production or during the laboratory short-term aging procedure in AASHTO R30 that is intended to simulate aging during production. Ongoing research is further evaluating this assumption.
- A loose-mix aging protocol of five days at 95°C was found to simulate field aging at approximately 70,000 CDD. An alternate loose-mix aging protocol at 135°C was evaluated for periods of 6, 12, and 24 hours. Analysis of the rheological and chemical binder properties indicate loose-mix aging for eight hours at 135°C would be similar to five days at 95°C. No significant difference in the oxidation-hardening relationship was evident for loose-mix aging at 95°C versus 135°C for the limited set of materials in the aging protocol development experiment. NCAT has referred to the loose-mix aging at 135°C for eight hours to simulate 70,000 CDD as the “critical-aging” protocol.
- Statistical analysis comparing cracking tests results of lab-prepared and plant-produced mixtures indicate that after critical-aging there is no statistical difference between lab and plant mixtures for each of the cracking test parameters reported in this chapter. This finding suggests that laboratory prepared mixtures can be used to correctly assess the cracking resistance of surface mixtures as part of a BMD procedure.

References

1. West, R., D. Timm, B. Powell, M. Heitzman, N. Tran, C. Rodezno, D. Watson, F. Leiva, and A. Vargas. *Phase VI (2015-2018) NCAT Test Track Findings*, NCAT Report 18-04, National Center for Asphalt Technology, Auburn, AL, 2018.
2. Chen, C. *Validation of Laboratory Cracking Tests for Field Top-down Cracking Performance*. Dissertation, Auburn University, Auburn, AL, 2020.
3. Roque, R., B. Birgisson, C. Drakos, and B. Dietrich. Development and Field Evaluation of Energy-based Criteria for Top-down Cracking Performance of Hot Mix Asphalt. *Journal of the Association of Asphalt Paving Technologists*, Vol. 73, 2004, pp. 229-260.
4. Garcia, V.M., Miramontes, A. Garibay, J. Abdallah, I., Carrasco, G. Lee, R., and Nazarian, S., Alternative Methodology for Assessing Cracking Resistance of Hot Mix Asphalt Mixtures with Overlay Tester, *Journal of the Association of Asphalt Paving Technologists*, 2017, pp. 527-548.
5. Garcia, V.M., Miramontes, A. Garibay, J. Abdallah, and Nazarian, S., Improved Overlay Test for Fatigue Resistance of Asphalt Mixtures, Research Report 0-6815-1, Center for Transportation Infrastructure Systems, 2016.
6. Ma, W. *Proposed Improvements to Overlay Test for Determining Cracking Resistance of Asphalt Mixtures*. MS thesis. Auburn University, Auburn, Ala., 2014.
7. Cooper III, S. B., W. King, and M. S. Kabir. Testing and Analysis of LWT and SCB Properties of Asphalt Concrete Mixtures. Presented at Louisiana Transportation Conference, Baton Rouge, LA, 2013. http://www.ltrc.lsu.edu/pdf/2016/FR_536.pdf.
8. Ozer, H., I. L. Al-Qadi, J. Lambros, A. El-Khatib, P. Singhvi, and B. Doll. Development of Fracture-based Flexibility Index for Asphalt Concrete Cracking Potential Using Modified Semi-circle Bending Test Parameters. *Construction and Building Materials*, Vol. 115, 2016, pp. 390-401.
9. Zhou, F., S. Im, L. Sun, and T. Scullion. Development of an IDEAL Cracking Test for Asphalt Mix Design and QC/QA, *Journal of the Association of Asphalt Paving Technologists*, 2017, pp. 549-577.
10. Wang, Y. D., B. S. Underwood, and Y. R. Kim. Development of a Fatigue Index Parameter, S_{app} , for Asphalt Mixes Using Viscoelastic Continuum Damage Theory. *International Journal of Pavement Engineering*, 2020. <https://www.tandfonline.com/action/showCitFormats?doi=10.1080/10298436.2020.1751844>.
11. Moore, N. *Evaluation of Laboratory Cracking Tests Related to Top-Down Cracking in Asphalt Pavements*. MS thesis. Auburn University, Auburn, AL, 2016.
12. FHWA Tech Brief. *Cyclic Fatigue Index Parameter (S_{app}) for Asphalt Mixture Performance Engineered Mixture Design*, Office of Construction and Pavements, FHWA-HIF-19-091, 2019.
13. Yin, F. and R. West, Balanced Mix Design Resource Guide, Information Series-143, National Asphalt Pavement Association, 2020.
14. Chen, C., F. Yin, P. Turner, R., West, and N. Tran. Selecting a Laboratory Loose Mix Aging Protocol for the NCAT Top Down Cracking Experiment, *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, Washington, D.C., 2018.

15. Chen, C., F. Yin, A. Andriescu, R. Moraes, D. Mensching, N. Tran, A. Taylor, and R. West. Preliminary Validation of the Critical Aging Protocol for NCAT Top-down Cracking Experiment. *Journal of the Association of Asphalt Paving Technologists*, 2020.