Ruggedness and Interlaboratory Studies for Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test: Phase II Report

PUBLICATION NO. FHWA-HRT-22-113

JUNE 2023

U.S. Department of Transportation
Federal Highway Administration

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

The standards for asphalt mixture performance tester (AMPT) cyclic fatigue testing (American Association of State Highway and Transportation Officials (AASHTO) TP 107-22 and AASHTO TP 133-21) enable pavement engineers to predict asphalt mixture performance over a wide range of loading and climate conditions.[1,2] These standards are an important component of ongoing Federal Highway Administration efforts to increase pavement life through fundamental testing and mechanistically based predictive models. For the widespread implementation of AMPT cyclic fatigue testing procedures to take place, precision statements that define the repeatability and reproducibility of the test standards are needed. This report details an interlaboratory study of AMPT cyclic fatigue testing standards that was undertaken to meet this need. Researchers analyzed the results of the interlaboratory study experiments to define the allowable within- and between-laboratory variations in the test results. They developed corresponding precision statements for both AASHTO TP 107-22 and AASHTO TP 133-21 and assessed the implications of the precision statements on the uncertainty in practical applications of the test. Draft revised versions of both standards have also been developed as part of this effort.

Mark E. Swanlund
Acting Director, Office of Infrastructure
Research and Development

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-22-113	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Ruggedness and Interlaboratory Studies for Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test: Phase II Report		5. Report Date June 2023
		6. Performing Organization Code
7. Author(s) Cassie Castorena (ORCID: 0000-0002-5915-0084), B. Shane Underwood (ORCID: 0000-0002-7223-3968), Y. Richard Kim (ORCID: 0000-0003-3295-977X), Jing Ding Kangjin Lee (ORCID: 0000-0003-3458-0747), Nam Tran (ORCID: 0000-0001-8183-4741), and Adam Taylor		8. Performing Organization Report No.
9. Performing Organization Name and Address North Carolina State University Department of Civil, Construction, and Environmental Engineering Campus Box 7908 Raleigh, NC 27695		10. Work Unit No.
		11. Contract or Grant No. DTFH6117C00037
12. Sponsoring Agency Name and Address Office of Infrastructure Research and Development Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101		13. Type of Report and Period Covered Phase II Report; September 2017–October 2021
		14. Sponsoring Agency Code HRDI-10

15. Supplementary Notes
The Contracting Officer's Representative was David J. Mensching, Ph.D., P.E. (HRDI-10; ORCID: 0000-0003-2460-2586).

16. Abstract
This report highlights findings from Phase II of a project designed to develop precision statements for cyclic fatigue testing using an asphalt mixture performance tester (AMPT). These standards include American Association of State Highway and Transportation Officials (AASHTO) TP 107-22 and AASHTO TP 133-21, which apply to 100-mm-diameter and 38-mm-diameter test specimens, respectively.[1,2] Seven laboratories participated in an interlaboratory study of the AMPT cyclic fatigue tests designed according to ASTM E691-20 and ASTM C670-15[4,5]. Researchers in these laboratories conducted dynamic modulus and cyclic fatigue testing of three mixtures for each specimen geometry. The results were used to establish repeatability and reproducibility precision limits for damage characteristic curve and failure criterion results of the AMPT cyclic fatigue tests; these limits reflect the variation in test results that will be exceeded with a probability of 5 percent if the test is executed properly. The researchers introduced a refined functional data metric to capture the variation in damage characteristic curve results. All precision limits were defined as a function of the mixture nominal maximum aggregate size (NMAS) except the failure criterion reproducibility, which did not follow a consistent trend with respect to NMAS. The established repeatability limits quantify the acceptable variation among three test determinations (specimens) obtained within a laboratory on a single material. The reproducibility limits quantify the acceptable variation between average test results of two laboratories conducted on the same material.

17. Key Words Cyclic fatigue testing, asphalt mixture performance tester, AMPT, asphalt mixture, interlaboratory study, repeatability, reproducibility, precision asphalt mixture performance testing standards	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. https://www.ntis.gov

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 98	22. Price N/A

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1,000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2,000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2,000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND VARIABLES

Abbreviations

2S2P1D	two springs, two parabolic elements, and one dashpot
AASHTO	American Association of State Highway and Transportation Officials
AMPT	asphalt mixture performance tester
COV	coefficient of variation
FHWA	Federal Highway Administration
ILS	interlaboratory study
LVDT	linear variable differential transformer
MCMC	Markov Chain Monte Carlo
NMAS	nominal maximum aggregate size
PG	performance grade
PID	proportional, integral, derivative
RAP	reclaimed asphalt pavement
SMA	stone matrix asphalt
SPT	simple performance tester
S-VECD	simplified viscoelastic continuum damage

Variables

α	damage growth rate
a_n	multiplier affected by the number of test determinations, n, listed in table 3
α_T	time–temperature shift factor at a given temperature
a_{Total}	sum of the contributions from uncertainty in the $\|E^*\|$ model, damage characteristic model, and failure criterion
β	2S2P1D fitting coefficient
C	pseudostiffness
C_{11}	fitting coefficient for power law damage-characteristic curve model
C_{12}	fitting coefficient for power law damage-characteristic curve model
C_{11_new}	fitted parameters for damage-characteristic curve model after normalization
C_{12_new}	fitted parameters for damage-characteristic curve model after normalization
$C_{11\text{-}i\_new}$	fitted parameters for specimen i after normalization
$C_{12\text{-}i\_new}$	fitted parameters for specimen i after normalization
$\bar{C}_{11\text{-}i\_new}$	fitted parameters for laboratory i after normalization
$\bar{C}_{12\text{-}i\_new}$	fitted parameters for laboratory i after normalization
$\bar{\bar{C}}_{11\_new}$	fitted parameters for both laboratory results after normalization
$\bar{\bar{C}}_{12\_new}$	fitted parameters for both laboratory results after normalization
$\bar{C}_{11\_new}$	fitted parameters for all three specimens after normalization
$\bar{C}_{12\_new}$	fitted parameters for all three specimens after normalization
C_i	pseudosecant modulus at the current time step
C_{i-1}	pseudosecant modulus at the previous time step
C^*	cyclic pseudosecant modulus
C^*_n	cyclic pseudosecant modulus at the current analysis cycle
C^*_{n-1}	cyclic pseudosecant modulus at the previous analysis cycle
$\bar{C}(S)_i$	the fitted function for ith laboratory after normalization
$\bar{\bar{C}}(S)$	the fitted function of two laboratories being compared after normalization

$C(S_{new})_i$	the fitted function for *i*th replicate after normalization for v_{norm} calculation
$\bar{C}(S_{new})$	the fitted function of all the data after normalization for v_{norm} calculation
$COV_{rpt}\%$	repeatability COV for S_{app}
$COV_{rpd}\%$	reproducibility COV for S_{app}
COV(*x*,*y*)	COV between x and y
$d2s_{rpt}(\%)$	allowable difference between two test results from one operator expressed as a COV
$d2s_{rpd}(\%)$	allowable difference between test results reported by two laboratories expressed as a COV
D^R	pseudostrain energy-based fatigue failure criterion
D_i^R	D^R value for *i*th specimen
D_{fit}^R	average D^R value for all the specimens
DMR	dynamic modulus ratio
$\%\bar{D}$	average *%Damage*
$\%D_P$	predicted damage at the reliability level *P*
%Damage	percent of damage predicted using FlexPAVE™
$E'_{fit\_i,j}$	fitted storage modulus from the fitted curve for the *i*th specimen and *j*th temperature and frequency
$E'_{fit\_avg,j}$	storage modulus from the fitted curve using all of the specimens at the *j*th temperature and frequency
δ	2S2P1D fitting coefficient
E'_{2S2P1D}	storage modulus from 2S2P1D model
E_0	maximum storage modulus value (kPa or psi)
E_{00}	minimum storage modulus value (kPa or psi)
$\|E^*\|$	dynamic modulus
$\|E^*\|_{fingerprint}$	dynamic modulus determined from fingerprint test
$\|E^*\|_{LVE}$	dynamic modulus from 2S2P1D model
ε^R	pseudostrain
ε_{pp}^R	peak-to-peak pseudostrain
ε_{ta}^R	tension amplitude of the pseudostrain
f_V	density distribution function for v_{norm}
γ	2S2P1D fitting coefficient
Γ	gamma function
h	2S2P1D fitting coefficient
η	between-laboratory consistency statistic
θ	within-laboratory consistency statistic
I_{LVE_test}	linear viscoelastic (dynamic modulus model) variation index from test result
$I_{D^R\_test}$	D^R variation index from test results
k	1×10^5, multiplier to change the magnitude for v_{norm} calculation
K	the shape parameter in gamma distribution
K_1	form adjustment factor
κ	2S2P1D fitting coefficient
λ	the scale parameter in gamma distribution
MAR	maximum allowable range
max E'	maximum storage modulus
n	total number of cyclic fatigue test specimens

n_{LVE}	total number of dynamic modulus test specimens
n_S	number of simulations in the 95-percent confidence interval, 4,750
N	number of cycles
N_f	number of cycles to failure
ω	angular frequency (rad/s)
ω_R	reduced angular frequency (rad/s)
p	number of laboratories in ILS
S	damage parameter
S_{all}	all the S values from n damage characteristic curves
S_{app}	apparent damage capacity
$S_{app\text{-}i}$	S_{app} value for the ith simulation
$\overline{S}_{app}$	average S_{app} value across all simulations
$S_{app\text{-}1}{}^{2.5}$	the 2.5 percentile of S_{app} for laboratory 1
$S_{app\text{-}1}{}^{97.5}$	the 97.5 percentile of S_{app} for laboratory 1
$S_{app\text{-}2}{}^{2.5}$	the 2.5 percentile of S_{app} for laboratory 2
$S_{app\text{-}2}{}^{97.5}$	the 97.5 percentile of S_{app} for laboratory 2
S_{factor}	factor used to normalize damage value in v_{norm} calculation
S_{min}	the minimum of the corresponding failure value of each dataset
S_{mean}	the average of the corresponding failure value of each dataset
S_{max}	the maximum of the corresponding failure value of each dataset
S_{new}	damage value after normalization
SE	standard error of actuator strain
s_i	within-laboratory standard deviation
$s_r\%$	repeatability standard deviation (COV)
$s_{\bar{x}}$	standard deviation of laboratory averages
$s_{\bar{x}}\%$	reproducibility standard deviation (COV)
std_{Total}	standard deviation of *%Damage*.
ΔS	increment of damage
ΔS_i	damage growth between the current and previous time step
ΔS_n	damage growth between the current and previous analysis cycle
Δt_R	reduced time step
σ	stress (kPa or psi)
σ_{pp}	peak-to-peak stress (kPa or psi)
t	time measured from the experiment (s)
t_R	reduced time (s)
T	temperature (°C)
τ_E	2S2P1D fitting coefficient
$\overline{x}_i$	laboratory average for laboratory i
$\overline{\overline{x}}$	average of all laboratory averages
v_{norm}	variation index for damage characteristic curve results
Z_P	standard normal deviate corresponding to reliability level P

CHAPTER 1. INTRODUCTION

The American Association of State Highway and Transportation Officials (AASHTO) testing protocols AASHTO TP 107-22 and AASHTO TP 133-19—collectively referred to as procedures for asphalt mixture performance tester (AMPT) cyclic fatigue testing—enable the practical, mechanical performance characterization of asphalt mixtures using cyclic fatigue testing in the AMPT.[1,2] They are the only testing protocols developed specifically for the AMPT to assess fatigue cracking.

AASHTO TP 107-22 is used for testing cylindrical asphalt mixture specimens that are 100-mm in diameter by 130-mm tall (hereafter referred to as large specimens).[1] In contrast, AASHTO TP 133-21 applies to cylindrical asphalt mixture specimens that are 38-mm in diameter by 110-mm tall (hereafter referred to as small specimens).[2] AASHTO TP 107-22 preceded AASHTO TP 133-21. The small-specimen geometry was established to improve testing efficiency and enable field core testing for mixtures with nominal maximum aggregate size (NMAS) values up to 19 mm.

The results of AMPT cyclic fatigue testing, coupled with the results of dynamic modulus ($|E^*|$) testing, are used to obtain two key material functions of the simplified viscoelastic continuum damage (S-VECD) model: the damage-characteristic curve (i.e., pseudostiffness (C) versus the internal state parameter representing damage (S)) and the pseudostrain energy-based fatigue failure criterion (D^R). The damage-characteristic curve and failure criteria are independent of the loading mode and loading history. Consequently, prediction of the damage response to any given loading history of interest is possible using limited test results, thus making the fatigue-cracking characterization of asphalt mixtures efficient compared to empirical methods.

The standards for AMPT cyclic fatigue testing recently underwent a ruggedness evaluation to define tolerance ranges for the experimental factors that contribute to variability in the test results.[3] The standards for AMPT cyclic fatigue testing were revised based on the ruggedness evaluation findings. The AMPT cyclic fatigue standards lack precision statements that define the repeatability (i.e., with laboratory variation) and reproducibility (i.e., between laboratory variation) of the test results when executed properly. Therefore, this study aims to improve the AMPT cyclic fatigue standards by developing precision statements. To achieve this objective, an interlaboratory study (ILS) was executed wherein seven laboratories conducted AMPT cyclic fatigue testing according to the AASHTO TP 107-22 and AASHTO TP 133-21 protocols established based on the ruggedness evaluation results.[1,2] Statistical analysis of the ILS yielded recommended precision statements for AASHTO TP 107-22 and AASHTO TP 133-21. The ILS was designed and analyzed according to ASTM E691-20 with integration of ASTM C670-15 recommendations specifically for construction materials.[4,5]

This report contains four chapters and an appendix. Chapter 1 introduces the study's motivation, objective, and scope. Chapter 2 provides the overall methodology of the study, including the materials, test and analysis methods, and experimental plan. Chapter 3 presents the results of the experiments and statistical analyses of the results, along with the proposed precision statements for AASHTO TP 107-22 and AASHTO TP 133-21. Chapter 4 presents the conclusions and recommendations drawn from the results of this study.[1,2]

This report uses certain naming conventions throughout as follows:

- "Samples" or "loose mixture samples" are plant-mixed asphalt mixture samples in a loose state.
- "Gyratory-compacted samples" are plant-mixed, laboratory-compacted asphalt mixture samples that have been compacted in a Superpave gyratory compactor.
- "Test specimens" or "specimens" refer to plant-mixed materials that have been compacted, cut, and cored in the laboratory.
- "Prefabricated specimens" are specimens prepared by a single laboratory and distributed to ILS participants.

CHAPTER 2. METHODOLOGY

ASTM E691-20 and ASTM C670-15 describe the basic methodology for designing, carrying out, and analyzing the results of an ILS to define the repeatability and reproducibility of a test procedure.[4,5] Repeatability defines the variability in test results obtained by a single operator in a single laboratory on the same material. Reproducibility defines the variability between test results obtained from two different laboratories on the same material. The cyclic fatigue test method's repeatability is represented as the acceptable variation among the three test replicates. The reproducibility of the cyclic fatigue test methods characterizes the acceptable variation between the average test results of two laboratories. This chapter describes the methodology followed in conducting the AMPT cyclic fatigue ILS.

PARTICIPATING LABORATORIES

The seven laboratories that participated in the ILS, the corresponding AMPT testing equipment, and the gyratory compactor type that they used are listed in table 1. The laboratories are codified within the results to preserve anonymity.

Table 1. ILS participants.

Participant	AMPT Equipment*	Gyratory Compactor*
Federal Highway Administration (FHWA)	Controls AMPT	Pine G2
Maine Department of Transportation	Controls AMPT	Pine G1
National Center for Asphalt Technology	Controls SPT (upgraded)	Pine G2
North Carolina State University	Controls AMPT Pro®	Pine G2
University of Massachusetts-Dartmouth	Controls AMPT	Pine G2
University of Nevada-Reno	Controls AMPT	Pine G2
University of New Hampshire	Controls SPT	Controls Servopac®

*The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.
SPT = simple performance tester.

ILS PROTOCOL

An ILS protocol was developed and distributed to the participating laboratories that provided detailed specimen fabrication and testing instructions. The researchers provided the laboratories with links to data-logging forms; AMPT cyclic fatigue test standards; and videos demonstrating sample fabrication, test specimen preparation, and testing procedures. The data-logging forms were used to collect information on specimen fabrication procedures, specimen bulk-specific gravity measurements, and testing details (e.g., thermal equilibration time and ambient laboratory conditions). In addition, participants were provided detailed instruction sheets that specified the specimen and sample identifications provided by the research team. They were given each specimen's and sample's purpose, mixture compaction temperatures, theoretical maximum specific gravity values, target air void contents, and test temperature and frequency requirements. The researchers provided participants with shared folders for uploading their data-logging forms and AMPT test results.

Before beginning the ILS, the research team distributed a prebriefing survey to ask participants about specific aspects of the testing equipment, calibration, and additional information to help address data inconsistencies and/or ensure compliance with the current requirements of the AMPT cyclic fatigue test standards. The research team also held a webinar before beginning the ILS to present the cyclic fatigue test standards that resulted from the ruggedness evaluation and ILS testing requirements. The participants were allowed to ask questions during the webinar.

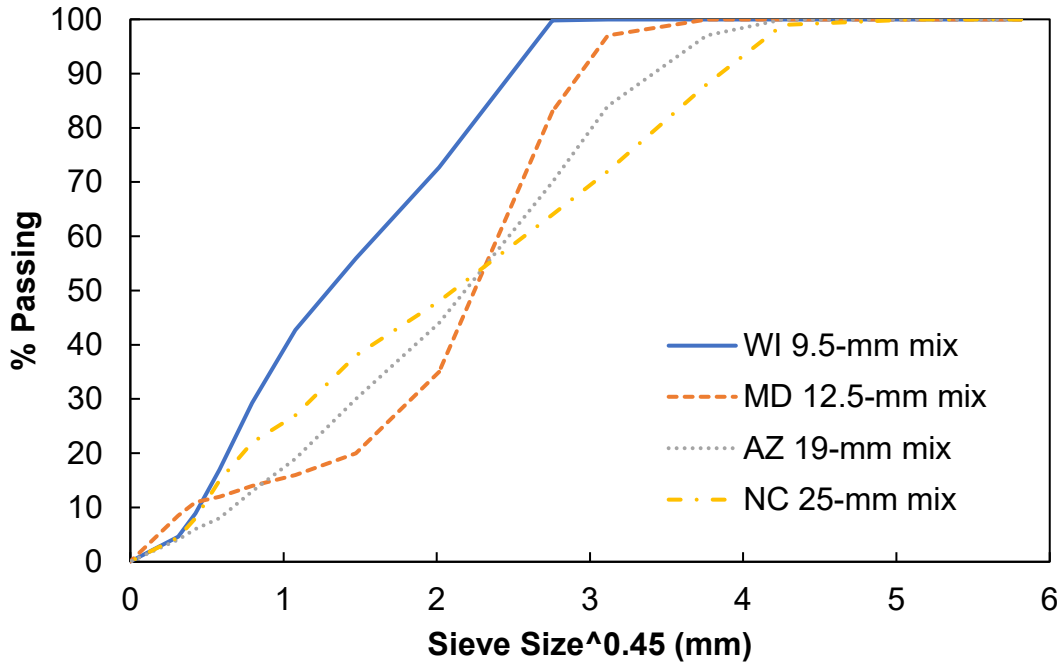
MATERIALS

Four plant-mixed, laboratory-compacted mixtures were used for the ILS: dense-graded 9.5-mm NMAS surface mixture from Wisconsin, stone matrix asphalt (SMA) surface mixture with 12.5-mm NMAS from Maryland, dense-graded 19-mm NMAS surface mixture from Arizona, and dense-graded 25-mm NMAS base mixture from North Carolina. Table 2 provides details about the selected mixtures. Figure 1 shows the gradations for the selected mixtures.

Table 2. Selected mixtures for the ILS.

Source	NMAS (mm)	Virgin Binder Grade	Binder Content (%)	RAP Content (%)	VMA (%)	VFA (%)	Number of Design Gyrations
Wisconsin	9.5	PG 58-28	6.1	25	16.2	75.3	75
Maryland	12.5	PG 64E-22	6.5	15	18.5	81.0	100
Arizona	19.0	PG 70-10	4.8	0	14.2	71.5	75
North Carolina	25.0	PG 64-22	4.2	30	13.6	71.5	65

PG = performance grade; RAP = reclaimed asphalt pavement; VMA = voids in mineral aggregate; VFA = voids filled with asphalt.



Source: FHWA.

Figure 1. Graph. Sieve size 0.45-mm power chart for gradations.

Plant-mixed, laboratory-compacted samples were used based on the results of a survey distributed to potential ILS participants that suggests that most laboratories use plant-mixed samples for test specimen fabrication in practice. The mixtures were sourced from different climatic regions that are expected to yield different test results. Also, the SMA mixture contained polymer-modified virgin asphalt, whereas the other mixtures did not include polymer modification. The research team included differences in NMAS and gradation type to cover a representative range of test variability. For both small- and large-specimen geometries, the researchers used both the 9.5-mm dense-graded surface mixture with an unmodified binder and the 12.5-mm SMA mixture with the polymer-modified binder. The 19-mm mixture was only used in the small-specimen geometry ILS, whereas the 25-mm mixture was only used in the large-specimen geometry ILS. The maximum NMAS permitted in AASHTO TP 133-21 is 19 mm.[(2)]

To prepare the loose mixture samples for distribution, the participants separated and randomized the plant-mixed material into individual samples following the researchers' loose mixture separation procedure.[(6)] Plant-mixed specimen testing using this procedure has yielded AMPT $|E^*|$ and cyclic fatigue test results with comparable variability to lab-mixed specimen testing in the research team's laboratory when this separation procedure is followed. Each separated sample was placed in a quick-release, heavy-duty, wax-lined box, consistent with those used by AASHTO re:source. Each box contained the material required to produce approximately one-half a gyratory-compacted sample. All mixture test specimens were prepared to achieve an air void content of 5 ± 0.5 percent.

PILOT TESTING

ASTM E691-20 specifies pilot testing before beginning an ILS to familiarize participants with the test procedures and ILS protocol.[4] ILS participants were supplied with prefabricated small-test specimens for protocol familiarization. The test specimens were fabricated according to AASHTO PP 99-19 by a single operator, with the research team using the 9.5-mm ILS mixture.[6] The fabricated specimens were randomized and distributed to the participating laboratories. Each laboratory received eight pilot test specimens: three for $|E^*|$ testing conducted according to AASHTO TP 132-19, three for cyclic fatigue testing following AASHTO TP 133-21, and two extra specimens.[2,8] Participants were instructed to conduct a single cyclic fatigue test and $|E^*|$ test for initial feedback. The researchers reviewed the initial test results to identify any procedural issues. If they identified issues, they met with the participating laboratory to resolve the problem prior to completing the additional tests.

To assess pilot $|E^*|$ test results, the differences between a participating laboratory and the research team's results were compared against the reproducibility limits included in AASHTO T 378-17.[9] Note that these limits were established based on a large specimen fabricated by individual laboratories, so the conditions under which they were developed do not coincide with the pilot testing conducted in this study. However, they provided a reference to assess the general agreement of the results obtained by different laboratories. Cyclic fatigue test results were assessed using the proportional, integral, derivative (PID) data quality indicator established in Phase I of this project (i.e., maximum limit of 10 percent on the actuator strain standard error in cycles two through five of the cyclic fatigue test) and the research team's experience evaluating the reproducibility of cyclic fatigue test results from Phase I of this project.[3] Procedure discrepancies identified among the participating laboratories that contributed to biases in initial pilot results included the following:

- Thermal equilibration time differences.
- Temperature calibration errors.
- $|E^*|$ tests using compensating springs.
- Deviations from strain selection guidance in AASHTO TP 133-21.[2]
- PID tuning issues.

The research team resolved these discrepancies before the full-scale ILS testing. The appendix presents a detailed discussion of the steps taken to identify and resolve procedural discrepancies, as well as the final pilot test results obtained after addressing procedural issues. Pilot test results were not used to develop the precision statements.

FULL-SCALE TESTING

The participating laboratories prepared full-scale ILS samples and test specimens using separated loose mixture samples. Full-scale ILS testing used specimens fabricated by the participant laboratories because precision statements for reproducibility are contingent upon variability in both the specimen fabrication and the test procedures. In addition, the repeatability of specimens fabricated by a single laboratory and within individual laboratories is not expected to differ significantly based on the ILS results of AASHTO T 378-17.[10]

For each ILS mixture, the research team shipped six boxes of separated loose mixture samples (i.e., sufficient material to produce three gyratory-compacted samples) to the participating laboratories to perform a compaction study. The compaction study would determine the sample mass required to achieve the specified air void content in extracted test specimens. The research team provided 16 additional boxes of loose mixture for each mixture included in the large-specimen ILS and 8 additional boxes of loose mixture for each mixture included in the small-specimen ILS. These allocations allowed for the fabrication of two extra gyratory samples for both large- and small-specimen fabrication purposes, as needed. Laboratories were required to measure and report the test specimen air void content to ensure they were within the specified range.

Laboratories were required to carry out three $|E^*|$ and three cyclic fatigue tests on each mixture and specimen geometry combination included in the ILS. Large-specimen $|E^*|$ testing adhered to AASHTO R 83-17 and AASHTO T 378-17.[9,11] Dynamic modulus tests of the small specimens followed AASHTO PP 99-19 and AASHTO TP 132-19.[6,8] In both large- and small-specimen $|E^*|$ tests, asphalt specimens are subjected to frequency-sweep testing at 10, 1, and 0.1 Hz at three temperatures. All specimens were tested at 4 and 20 °C. The third test temperature for the 9.5-mm mixture was 35 °C. For the other mixtures, the third test temperature was 40 °C. The test temperatures were selected on the basis of the binder performance grade in accordance with AASHTO R 84-17.[11]

Large-specimen cyclic fatigue testing followed AASHTO TP 107-22, whereas small-specimen cyclic fatigue testing adhered to AASHTO PP 99-19 and AASHTO TP 133-21.[1,2,7] The researchers selected the test temperature based on the guidance given in AASHTO TP 133-21, which specifies test temperatures based on the expected blended binder grade for the mixtures containing reclaimed or recycled materials. When a test resulted in a specimen failure outside the instrumented gauge points (i.e., end failure), as depicted in figure 2 of AASHTO TP 133-21, the test was considered invalid, and an additional specimen was tested.

TEST ANALYSIS PROCEDURES

The results of both AASHTO TP 133-21 and TP 107-22 are the S-VECD damage characteristic curve and failure criterion.[1,2] The linear viscoelastic properties of the asphalt mixture are required to implement the S-VECD model, which is obtained from the $|E^*|$ test. FlexMAT™ 2.0 was used to analyze all AMPT $|E^*|$ and cyclic fatigue test data.[13] The following sections provide information about the analysis of the ILS test results conducted within FlexMAT 2.0 to obtain the damage characteristic curve and failure criterion results.

Dynamic Modulus Testing

The calculation of $|E^*|$, phase angle, and data-quality indicators using AMPT dynamic modulus test results followed the method outlined in each respective $|E^*|$ standard. Consistent with AASHTO TP 133-21 and TP 107-22, the researchers used the two springs, two parabolic elements, and one dashpot (2S2P1D) model to characterize the $|E^*|$ as a function of temperature and frequency.[1,2,14] Equations 1 through 3 show the 2S2P1D model for storage modulus.

$$E'_{2S2P1D} = E_{00} + \frac{E'_1}{\left(\left(\frac{E'_1}{E_0 - E_{00}}\right)^2 + \left(\frac{E'_2}{E_0 - E_{00}}\right)^2\right)} \quad (1)$$

$$E'_1 = (E_0 - E_{00}) \times \left(1 + \delta \times (\omega_R \times \tau_E)^{-\kappa} \times \cos\left(\frac{\kappa\pi}{2}\right) + (\omega_R \times \tau_E)^{-h} \times \cos\left(\frac{h\pi}{2}\right)\right) \quad (2)$$

$$E'_2 = (E_0 - E_{00}) \times \left(\delta \times (\omega_R \times \tau_E)^{-\kappa} \times \sin\left(\frac{\kappa\pi}{2}\right) + (\omega_R \times \tau_E)^{-h} \times \sin\left(\frac{h\pi}{2}\right) + (\omega_R \times \tau_E \times \beta)^{-1}\right) \quad (3)$$

Where:

E_0 = maximum storage modulus value (kPa or psi).
E_{00} = minimum storage modulus value (kPa or psi).
κ, δ, γ, h, β, τ_E = fitting coefficients.
ω_R = reduced angular frequency (rad/s).
E'_{2S2P1D} = storage modulus from 2S2P1D model.

To best reflect the effect of the differences in cyclic fatigue tests alone on the damage characteristic curve and failure criterion, the research team analyzed the $|E^*|$ test results of all laboratories for a given mixture and specimen geometry together and coupled the averaged result with all of the cyclic fatigue testing obtained from each lab. One way in which the 2S2P1D model is used within the S-VECD model is the calculation of the dynamic modulus ratio (DMR), defined in equation 4. The $|E^*|$ from the 2S2P1D model yields $|E^*|_{LVE}$ in this equation, calculated at the test temperature and frequency corresponding to cyclic fatigue testing. In equation 4, $|E^*|_{fingerprint}$ is defined as the $|E^*|$ determined from the fingerprint portion of the cyclic fatigue test. *DMR* is used in AASHTO TP 107-22 and AASHTO TP 133-21 to characterize the potential difference between the specimens tested in the $|E^*|$ test and the specimens used in the cyclic fatigue test.[1,2] This value is also used to calculate parameters that define the damage characteristic curve.

The AMPT cyclic fatigue test standards stipulate that the *DMR* must fall between 0.85 and 1.15 to ensure sufficient consistency in the test results based on the ruggedness evaluation.[3] In some cases, the use of the average $|E^*|$ test results from all laboratories to calibrate the 2S2P1D model yielded *DMR* values that fell outside of the specified limits; in these cases, a subset of the $|E^*|$ test results of the collective laboratories was identified for calibrating the 2S2P1D model that yielded *DMR* values within the specified range for all laboratories. The appendix includes the full-scale ILS dynamic modulus results and a summary of individual laboratories' $|E^*|$ results that were used collectively to generate $|E^*|$ inputs for cyclic fatigue analysis of each mixture and specimen geometry combination.

$$DMR = \frac{|E^*|_{fingerprint}}{|E^*|_{LVE}} \tag{4}$$

Cyclic Fatigue Testing

The damage characteristic curve constitutes the relationship between the normalized pseudosecant modulus *C*, which is a material integrity indicator, and damage *S*. The research team followed the standard equations for calculating these quantities, as presented in the AASHTO TP 107-22 and TP 133-21 test protocols.[1,2] The pseudosecant modulus and damage are defined in equations 5 through 8; equations 5 and 6 are used to calculate the pseudosecant modulus and damage, respectively, for the tensile portion of the initial loading cycle, whereas equations 7 and 8 are used in later cycles.[15]

$$C = \frac{\sigma}{\varepsilon^R \times DMR} \tag{5}$$

Where:

σ = stress (kPa or psi).
ε^R = pseudostrain.

$$\Delta S_i = \begin{cases} \left(-\frac{DMR}{2} \left(\varepsilon^R\right)^2 \left(C_i - C_{i-1}\right) \right)^{\alpha/\alpha+1} \left(\Delta t_R\right)^{1/\alpha+1} & C_i \le C_{i-1} \\ 0 & C_i > C_{i-1} \end{cases} \tag{6}$$

Where:

ΔS_i = damage growth between the current and previous time step.
C_i = pseudosecant modulus at the current time step.
C_{i-1} = pseudosecant modulus at the previous time step.
Δt_R = reduced time step.
α = continuum damage power term, equal to 1+1/*m* where *m* is the maximum slope of the relaxation modulus master curve in log space.

$$C^* = \frac{\sigma_{pp}}{\varepsilon_{pp}^R \times DMR} \tag{7}$$

Where:

C^* = cyclic pseudosecant modulus.
σ_{pp} = peak-to-peak stress (kPa or psi).
ε_{pp}^R = peak-to-peak pseudostrain.

$$\Delta S_n = \begin{cases} \left(-\frac{DMR}{2} \left(\varepsilon_{ta}^R \right)^2 \left(C^*_n - C^*_{n-1} \right) \right)^{\alpha/\alpha+1} \left(\Delta t_R \right)^{1/\alpha+1} \left(K_1 \right)^{1/\alpha+1} & C^*_n \le C^*_{n-1} \\ 0 & C^*_n > C^*_{n-1} \end{cases} \tag{8}$$

Where:

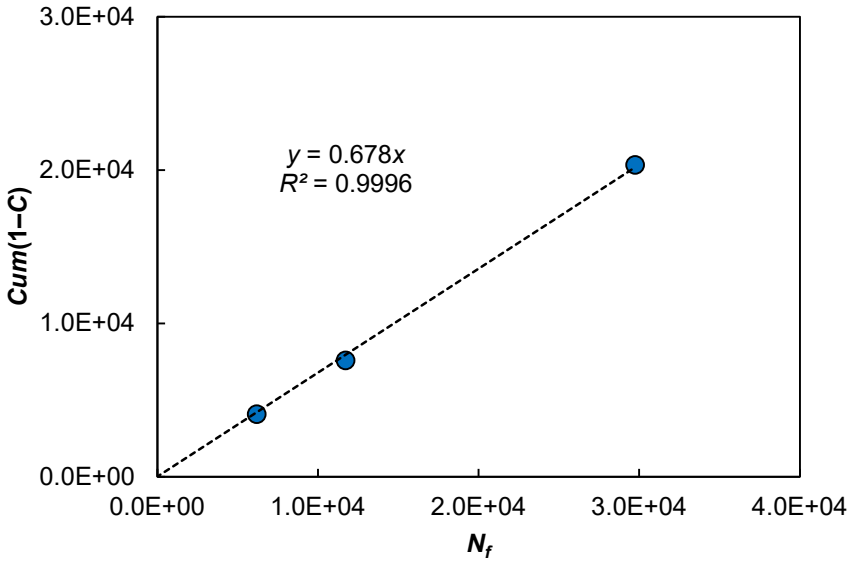
ΔS_n = damage growth between the current and previous analysis cycle.
ε_{ta}^R = tension amplitude of the pseudostrain.
K_1 = form adjustment factor.
C^*_n = cyclic pseudosecant modulus at the current analysis cycle.
C^*_{n-1} = cyclic pseudosecant modulus at the previous analysis cycle.

Equation 9 is used to model the relationship between C and S, constituting the damage characteristic curve using the collective results of three test specimens.[16] The fitting coefficients C_{11} and C_{12} are reported as one of the primary test results from the AMPT cyclic fatigue test. S represents the microstructural damage state variable.

$$C = 1 - C_{11} S^{C_{12}} \tag{9}$$

The failure criterion (D^R) is another cyclic fatigue test result, defined in equation 10 and shown in figure 2. The D^R is the average reduction in pseudostiffness up to failure.[17] The D^R of a given mixture is determined as the average D^R value of the collective test specimens. Thus, the deviation of the individual data points from the best fit line is a measure of within-laboratory specimen-to-specimen variability in the failure criterion results. The N_f is defined as the cycle where the product of peak-to-peak stress and cycle number reaches a peak value in AASHTO TP 133-21 and TP 107-22.[1,2]

$$D^R = \frac{\int_0^{N_f} (1 - C) dN}{N_f} \tag{10}$$



Source: FHWA.
Cum = cumulative.

Figure 2. Graph. D^R failure criterion.

STATISTICAL ANALYSIS

The researchers conducted repeatability and reproducibility analysis according to ASTM E691-20 and ASTM C670-15.[4,5] However, an inherent limitation of these standards is that they are "concerned exclusively with test methods which yield a single numerical figure as the test result."[4] The damage characteristic curve is a cyclic fatigue test result; this functional relationship does not lend directly to the analysis procedure prescribed in ASTM E691-20 or C670-15. Therefore, the researchers also investigated an alternative means to define the repeatability and reproducibility of the damage characteristic curve using a recently developed functional data metric.[18]

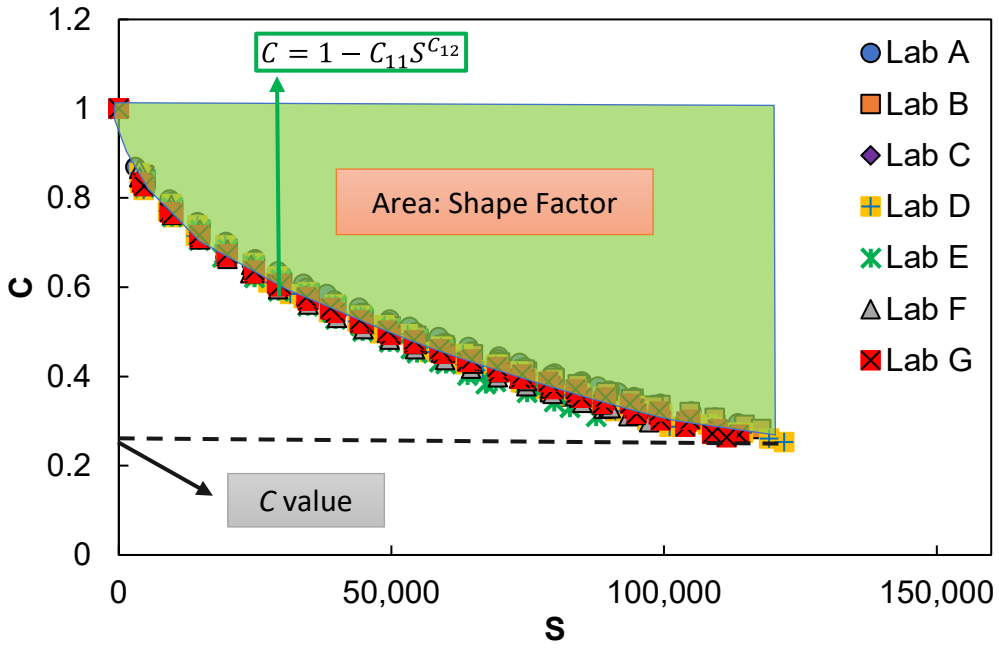
Single-Point Measure Analysis Conducted Using ASTM E691 and ASTM C670

There are two underlying assumptions regarding the interpretation of ILS test results according to ASTM E691-20.[4] First, all laboratories are assumed to have the same variability level under the repeatability (within-laboratory) conditions for a given mixture. Second, mean laboratory test results are assumed to vary according to a normal distribution in each mixture. Following the terminology in the standard, a test result represents the value of a characteristic obtained by carrying out a specified test method. A test determination represents the value of a characteristic or dimension of a single test specimen derived from one or more observed values. An example of a test determination is a single test specimen's damage characteristic curve.

ASTM E691-20 provides guidance for preparing a precision statement for test results.[4] ASTM C670-15 provides additional guidance for preparing a precision statement using the results of test determinations.[5] Based on the definition of test result and test determination, this study includes three test determinations for each laboratory per material. The test results are the best-fit damage characteristic curve across three specimens collectively and the average D^R value from three

specimens. A test determination is the damage characteristic curve and D^R value of one test specimen. Given that a laboratory would typically execute three test determinations to obtain a test result and not obtain two separate test results, repeatability of the cyclic fatigue test standards should reflect variability among test determinations within a lab. Therefore, the repeatability and reproducibility precision state development herein incorporate guidance from ASTM C670-15.[5] The ASTM E691-20 and C670-15 standards strive to define limits for the repeatability and reproducibility of a test that will only be exceeded with a probability of 5 percent if the test is executed properly.[4,5]

As noted, ASTM E691-20 is written for tests that yield a single numerical figure as the test result, which is directly applicable to the failure criterion result, D^R.[4] However, the damage characteristic curve results are functional data rather than a single numerical figure. To comply with the standard, the research team selected the following six single-point values to represent the damage characteristic curve for initial analyses: C value at S_{min}, S_{mean}, and S_{max} and shape factor at S_{min}, S_{mean}, and S_{max}, S_{min}, S_{mean}, and S_{max} represent the minimum, average, and maximum of all the S at failure for each material/specimen geometry. The shape factor represents the area between $C = 1$ and the C versus S curve at predefined levels of S. In this analysis, the shape factor was computed when S equals S_{min}, S_{mean}, or S_{max}. Figure 3 shows an example of the C value and shape factor at S_{max}.



Source: FHWA.
C = C value; S = shape factor.

Figure 3. Graph. An example of *C* value and shape factor.

The analysis framework prescribed in ASTM E691-20 can be divided into three steps:[4]

- Step 1: Evaluate test consistency. Test statistics are used to determine whether the collected data are adequately consistent to form the basis for a test method precision statement.

- Step 2: Investigate inconsistent data. The data that do not meet the consistency requirement are investigated to identify any possible errors, such as clerical or procedural errors. Then, the analysts should decide whether to include or delete the inconsistent data.

- Step 3: Obtain the precision statistics. The repeatability and reproducibility standard deviations are calculated to establish the repeatability and reproducibility statements.

The researchers executed these three analysis steps as described in the following sections.

Test Consistency

The statistical analysis for the estimates of precision statistics is simply a one-way analysis of variance (within and between laboratories) for each material. Therefore, severe outliers can significantly affect the analysis and invalidate the results. The consistency of the data should be examined before developing precision statistics. Equations 11 through 14 present the statistics used in this analysis according to ASTM E691-20.[4] Statistics *h* and *k* were used to evaluate the between-laboratory consistency and within-laboratory consistency, respectively.

$$s_{\bar{x}} = \sqrt{\sum_{i=1}^{p} (\bar{x}_i - \bar{\bar{x}})^2 / (p-1)} \tag{11}$$

Where:

$s_{\bar{x}}$ = standard deviation of laboratory averages.
$\bar{x}_i$ = laboratory average for laboratory *i*.
$\bar{\bar{x}}$ = average of all laboratory averages.
p = number of laboratories in ILS.

$$s_r = \sqrt{\sum_{i=1}^{p} s_i^2 / p} \tag{12}$$

Where:

s_r = repeatability standard deviation.
s_i = within-laboratory standard deviation.

$$\eta = (\bar{x}_i - \bar{\bar{x}}) / s_{\bar{x}} \tag{13}$$

$$\theta = s_i / s_r \quad (14)$$

Where:

η = between-laboratory consistency statistic.
θ = within-laboratory consistency statistic.

The consistency statistics η and θ were calculated for each laboratory, geometry, and material combination to evaluate the consistency of the data collected in the ILS. The critical value of each statistic is based on the number of laboratories and number of tests each laboratory conducts. This study was designed to include seven laboratories and three cyclic fatigue tests for each laboratory. The corresponding critical values for η and θ for these conditions are ±2.05 and 2.03, respectively (ASTM E691-20).[4] Consistency statistics were calculated for both the D^R and damage characteristic curve results described in the previous section.

Investigate Inconsistent Data

The inconsistent data flagged in the consistency test were investigated to make sure that there were no procedural or reporting errors. It should be noted that the critical values of the h and k consistency statistics are calculated based on the 0.5 percent significance level rather than the 5 percent often used in statistical analyses of materials (ASTM E691-20).[4] There are 42 h (or k) values for one single-point measure representing seven laboratories and six mixtures at a particular testing geometry. Theoretically, 0.21 out of 42 values are expected to stand out as inconsistent according to the 0.5-percent significance level. Therefore, any data flagged as inconsistent require careful investigation for rejection from precision statement development. Data ultimately deemed inconsistent were removed from the dataset before the repeatability and reproducibility statistics were computed.

Obtain the Precision Statistics

Each laboratory conducted three test determinations in the AMPT cyclic fatigue ILS rather than obtaining three test results. Therefore, the repeatability and reproducibility statistics developed follow the guidance in ASTM C670-15, which applies to the development of repeatability statements based on individual test determinations and results. ASTM E691-20 only provides guidance for developing precision statements based on test results.[4,5]

The repeatability standard deviation, s_r, represents the standard deviation of test determinations obtained under repeatability conditions. In this study, the reproducibility standard deviation was taken to be $s_{\bar{x}}$, which represents the standard deviation of test results obtained under reproducibility conditions (i.e., quantifies the difference between the test results of two laboratories). For the C at a specific S value, single-point measures of damage characteristic curve variability, 1−the average C was used as the basis for the calculation of the average value across laboratories because the standard deviation in the C value increases as the C value decreases (and correspondingly the S value increases).

The corresponding repeatability and reproducibility coefficients of variation (COVs) were calculated using equation 15 and equation 16, respectively.

$$s_r\% = \frac{s_r}{\bar{\bar{x}}} \times 100 \tag{15}$$

$$s_{\bar{x}}\% = \frac{s_{\bar{x}}}{\bar{\bar{x}}} \times 100 \tag{16}$$

Where:

$s_{\bar{x}}\%$ = reproducibility COV.

$\bar{\bar{x}}$ = average of all laboratory averages for shape factor, 1−average of all laboratory averages for *C* value.

Due to the nature of the damage characteristic curve, the *C* or *S* value selected to carry out a single-point measure is highly likely to affect the repeatability and reproducibility results. Plots of repeatability and reproducibility COVs at different 1−*C* values were evaluated to identify the sensitivity of the repeatability and reproducibility to the chosen point within the damage characteristic curve.

When the test determinations are the basis for repeatability, the single-operator precision statement is defined as the maximum allowable difference between multiple test determinations or the maximum acceptable difference between test results obtained as the average of multiple determinations (ASTM C670-15).[5]

For the repeatability of one test result with *n* test determinations, the maximum allowable range (MAR) (i.e., the difference between the highest and lowest test determinations) expressed as a COV is calculated using equation 17.

$$MAR = a_n \times s_r\% \tag{17}$$

Where:

a_n = multiplier affected by the number of test determinations, *n*, listed in table 3.

$s_r\%$ = repeatability COV.

Table 3. Multiplier for repeatability statistic.

Number of Test Determinations	Multiplier
2	2.8
3	3.3
4	3.6
5	3.9
6	4.0
7	4.2
8	4.3
9	4.4
10	4.5

ASTM C670-15 does not give very prescriptive instructions for defining the reproducibility of test results. It only specifies that "If the test method calls for reporting the average of more than one test determination, multilaboratory precision is expressed as a maximum allowable difference between averages of such groups obtained by two laboratories."[5] Based on this guidance, the reproducibility precision of the AMPT cyclic fatigue tests was defined using $s_{\bar{x}}$, which constitutes the standard deviation of the test results between two laboratories. Correspondingly, equation 18 was used to define the reproducibility limit in terms of the allowable difference in test results reported by two laboratories, expressed as a COV.

$$d2s_{rpd}(\%) = 2.8 \times s_{\bar{x}}(\%) \tag{18}$$

Where:

$d2s_{rpd}(\%)$ = allowable difference between test results reported by two laboratories expressed as a COV.

Advanced Analysis of Damage Characteristic Curve Repeatability and Reproducibility Using a Functional Data Metric

To capture the variation of the entire damage characteristic curve among test replicates better than the single-point measures, Ding et al. developed a variance index (v) [18] Herein, the definition of the variance index has been modified by first normalizing the damage characteristic curve and then performing the same calculation as shown in Ding et al. Because the curve is first normalized, the name of the index used for this project is "normalized variance index," v_{norm}, and distinct definitions of v_{norm} for repeatability and reproducibility are given. However, using v_{norm} to define the repeatability and reproducibility of damage characteristic curve results requires some deviation from the ASTM standards because the v_{norm} follows a gamma, rather than a normal, distribution.

The v_{norm} is defined in equations 19 through 21. Equations 20 and 21 show that when the v_{norm} is calculated, the S values are first normalized to avoid bias based on the damage characteristic curve length (i.e., S at failure). As shown by the integral in equation 19, v_{norm} encompasses the total area between individual C versus S curves and the curve that resulted from optimization of

all individual curves with normalization based on the number of curves considered and a scaling factor, k. The multiplier k was applied to change the magnitude of v_{norm} from approximately 10^{-5} to 1 for convenience. Further details pertaining to the development of v, and thus v_{norm}, are provided in Ding et al.[18]

$$v_{norm} = \frac{k \cdot \sum_{i=1}^{n} \int_0^1 (C(S_{new})_i - \bar{C}(S_{new}))^2 dS_{new}}{n-1} \tag{19}$$

$$S_{new} = S_{all} \times S_{factor} \tag{20}$$

$$S_{factor} = \frac{1}{S_{mean}} \tag{21}$$

Where:

$k = 1 \times 10^5$, multiplier to change the magnitude.
S_{all} = all the S values from n replicates.
$C(S_{new})_i$ = the fitted function for i^{th} replicate after normalization.
$\bar{C}(S_{new})$ = the fitted function of all the data after normalization.
n = the total number of cyclic fatigue test specimens.

Equations 22 through 24 are used to obtain the fitted functions after normalization (i.e., $C(S_{new})_i$ and $\bar{C}(S_{new})$) in equation 19, using the originally defined damage characteristic curve model coefficients in equation 9 and S_{factor} (defined in equation 21).

$$C(S_{new}) = 1 - C_{11\_new} S_{new}^{C_{12\_new}} \tag{22}$$

$$C_{11\_new} = C_{11} / (S_{factor}{}^{C_{12}}) \tag{23}$$

$$C_{12\_new} = C_{12} \tag{24}$$

Where:

C_{11_new}, C_{12_new} = the fitted parameters for damage-characteristic curve model after normalization.
C_{11}, C_{12} = the fitted parameters before normalization.
Using v_{norm} to define repeatability.

Figure 4 illustrates the calculation of v_{norm} to evaluate damage characteristic curve repeatability. Damage characteristic curve results for three specimens are shown in figure 4-A. The S_{mean} value

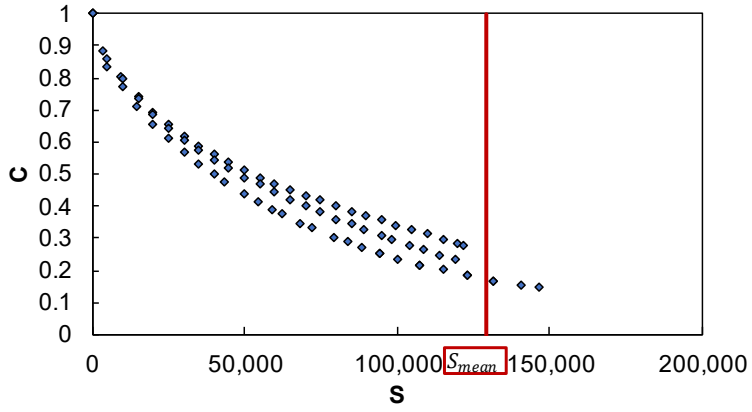
is calculated by averaging all S at failure for three specimens, and the corresponding S_{factor} is calculated using equation 21. All S values are normalized using equation 20, and the corresponding rescaled damage characteristic curve results are shown in figure 4-B. Fitted functions for individual specimens (i.e., fit_individual) and the overall fit from the aggregated results of all specimens (i.e., fit_all) are shown in figure 4-B as well. The fitting coefficients in these cases are determined using equations 23 and 24. The individual specimen curves are used to calculate $C(S_{new})_i$, and the overall fit is used to calculate $\bar{C}(S_{new})$ and determine v_{norm} using equation 25 (obtained by substituting equation 22 and the equivalent for the overall fit into equation 19 and solving the integral).

$$v_{norm} = \frac{k \cdot \sum_{i=1}^{n} \left(\frac{\bar{C}_{11\_new}^2}{2\bar{C}_{12\_new} + 1} + \frac{C_{11-i\_new}^2}{2C_{12-i\_new} + 1} - 2\frac{\bar{C}_{11\_new} C_{11-i\_new}}{\bar{C}_{12\_new} + C_{12-i\_new} + 1} \right)}{n - 1} \tag{25}$$

Where:

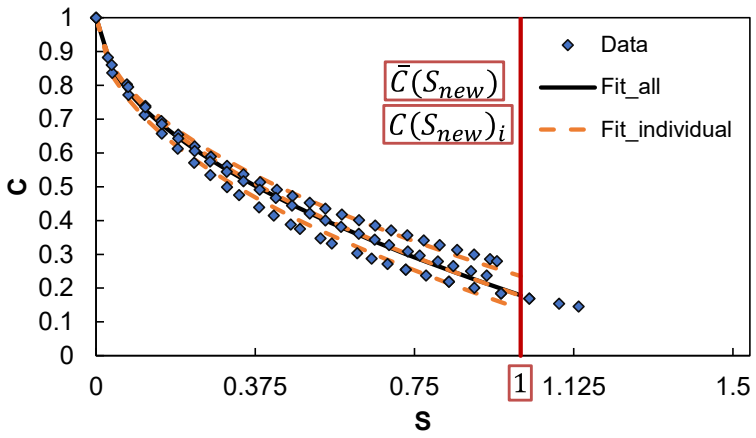
$C_{11\text{-}i\_new}$, $C_{12\text{-}i\_new}$ = the fitted parameters for specimen i after normalization.

$\bar{C}_{11\_new}$, $\bar{C}_{12\_new}$ = the fitted parameters for all three specimens after normalization.



Source: FHWA.

A. Finding S_{mean} for three specimens.



Source: FHWA.

B. Normalizing *S* values and fitting individual and overall damage characteristic curves.

Figure 4. Graphs. An example for calculating repeatability v_{norm}.

Using v_{norm} to Define Reproducibility

Equation 26 provides the v_{norm} to define damage characteristic curve reproducibility. The denominator is 1 in equation 19 when v_{norm} is calculated to evaluate the reproducibility of the results of two laboratories because $n = 2$ in this case.

$$v_{norm} = k \cdot \sum_{i=1}^{2} \int_0^1 (\bar{C}(S_{new})_i - \bar{\bar{C}}(S_{new}))^2 dS_{new} \tag{26}$$

Where:

$\bar{C}(S)_i$ = the fitted function for *i*th laboratory after normalization.
$\bar{\bar{C}}(S)$ = the fitted function of two laboratories being compared after normalization.

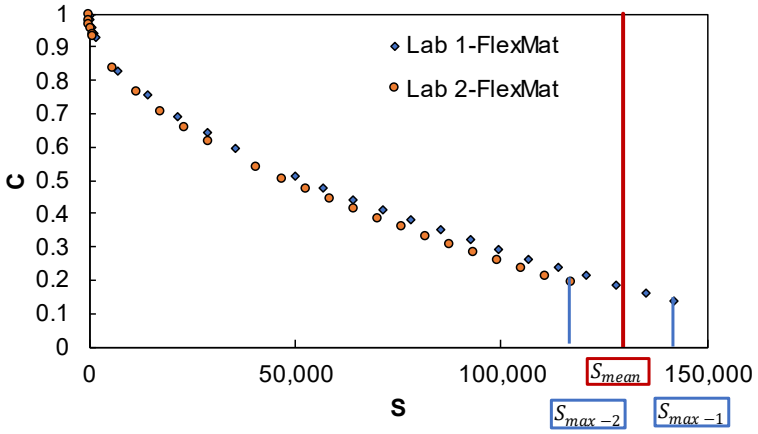
Figure 5 illustrates the v_{norm} approach for calculating the reproducibility of damage characteristic curves using equation 26. Figure 5-A shows the selected data points of fitted curves for the individual laboratories obtained from the collective results of three test specimens. The damage at failure for each fitted curve is the S_{max} across all the specimens of each laboratory. The selected data points for each fitted curve are based on predefined damage values of 0, 100, 300, 500, and 1,000, and damage values with an interval of $0.05 \times S_{max}$ from $0.05 \times S_{max}$ to S_{max}. The S_{mean} value is calculated by averaging S_{max} of two laboratories, and the S_{factor} for reproducibility is calculated using equation 21. Figure 5-B presents the fitted curve across two laboratory results with the selected data points. The fitted curves for each laboratory and the fitted curve across two laboratory results are then rescaled, and $\bar{\bar{C}}(S_{new})$ is determined for each of the three fitted curves. Corresponding rescaled model fitting coefficients are determined using equations 23 and 24. The reproducibility v_{norm} captures the reproducibility of two test results and is calculated using equation 27. Recall that a test result is the damage characteristic curve that best fits the collective $C(S)$ data from three test specimens

$$v_{norm} = k \cdot \sum_{i=1}^{2} \frac{\overline{\overline{C}}_{11\_new}^{2}}{2\overline{\overline{C}}_{12\_new} + 1} + \frac{\overline{C}_{11-i\_new}^{2}}{2\overline{C}_{12-i\_new} + 1} - 2\frac{\overline{\overline{C}}_{11\_new}\overline{C}_{11-i\_new}}{\overline{\overline{C}}_{12\_new} + \overline{C}_{12-i\_new} + 1} \quad (27)$$

Where:

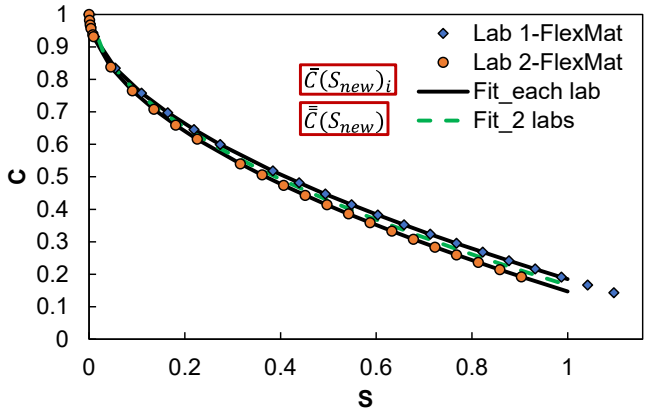
$\overline{C}_{11\text{-}i\_new}$, $\overline{C}_{12\text{-}i\_new}$ = the fitted parameters for laboratory i after normalization.

$\overline{\overline{C}}_{11\_new}$, $\overline{\overline{C}}_{12\_new}$ = the fitted parameters for both laboratory results after normalization.



Source: FHWA.

A. Finding S_{mean} for two laboratory results.



Source: FHWA.

B. Normalizing S values and fitting overall damage characteristic curves.

Figure 5. Graphs. An example for calculating reproducibility v_{norm}.

Precision Statements

Ding et al. showed that v_{norm} values obtained from different operators for a given material follow a gamma distribution, defined in equation 28.[18] The analysis procedures in ASTM E691–20 and C670-15 are only directly applicable to test results that follow a normal distribution. Therefore,

v_{norm} is not directly amenable to the guidance for precision statement development given in the ASTM standards.[4,5] However, the research team used the underlying principle behind the precision statement developed in the ASTM standards as the basis for developing a precision statement using v_{norm} herein. A gamma distribution was fit using the ILS results and was subsequently used to define the maximum limit for v_{norm} that will be exceeded with a probability of 5 percent if the test is executed properly.

$$v_{norm} \sim Gamma(K, \lambda), f_V(v) = \frac{v^{K-1} e^{-v/\lambda}}{\lambda^K \Gamma(K)} \quad (28)$$

Where:

f_V = probability density function for v_{norm}.
K = the shape parameter in gamma distribution.
Γ = the gamma function
λ = the scale parameter in gamma distribution.

ANALYSIS OF THE PRECISION STATEMENT IMPLICATIONS ON UNCERTAINTY IN THE TEST RESULTS' PRACTICAL APPLICATIONS

Uncertainty in FlexPAVE™ %Damage Predictions

One of the primary uses of the AMPT cyclic fatigue test results is the mechanistic prediction of fatigue damage evolution using FlexPAVE.[19] FlexPAVE version 1.1, developed under the sponsorship of FHWA, is a pavement-performance analysis software package that uses $|E^*|$ and S-VECD models to integrate the effects of loading rate and temperature into a structural model that then computes the pavement's responses and damage evolution. The FlexPAVE program calculates the long-term fatigue damage and rut depth of asphalt pavements under changing environmental and loading conditions. The output from FlexPAVE fatigue cracking simulations is the percentage of damage (*%Damage*), which is computed as the ratio of the damaged cross-sectional area to the total effective cross-sectional area; additional details are provided elsewhere.[20] Ghanbari recently established a method to estimate how the uncertainties from material model inputs propagate to uncertainty in *%Damage* predictions obtained from FlexPAVE.[21] This study adopted this framework to evaluate the implications of the combined uncertainty in the damage characteristic curve and failure criterion results on the uncertainty in *%Damage* predictions.

Ghanbari used Bayesian inference-based Markov Chain Monte Carlo (MCMC) methods to estimate how the uncertainties from the material model's input parameters propagate to uncertainty in the FlexPAVE Version 1.1 predicted pavement performance.[21] First, predictive envelopes at predefined statistical levels were generated independently for each material model (i.e., $|E^*|$, damage characteristic curve, and failure criterion) using the MCMC method. The predictive envelope defines the range that a new observation will fall into with a defined level of probability. Based on predictive envelope, 1,000 possible results for each of the three models were generated. Next, 1,000 possible sets of FlexPAVE inputs were generated by randomly sampling from the distributions of material model inputs without replacement.

FlexPAVE simulations were then conducted using these inputs to generate 1,000 sets of *%Damage* results that reflect the uncertainty in pavement performance simulation as a consequence of the material model uncertainty. This analysis framework was carried out using different materials, pavement structures, traffic conditions, climate conditions, and base and subgrade moduli. Ghanbari's analysis demonstrated that the uncertainty in *%Damage*, defined by the standard deviation in the predicted *%Damage*, depends on the average *%Damage* and material model uncertainty, irrespective of the pavement simulation conditions (i.e., structure, traffic, climate), as shown in equation 29. Correspondingly, equation 30 was defined to relate uncertainty in *%Damage* predictions to uncertainty in the $|E^*|$ and cyclic fatigue model uncertainties. The a_{Total} in equation 30 is defined by summing the contributions from uncertainty in the $|E^*|$ model, damage characteristic model, and failure criterion, as shown in equations 31-36.

$$\%D_P = \%\bar{D} + Z_p \times std_{Total} \tag{29}$$

$$std_{Total} = a_{Total}(\%\bar{D})^{0.544} \tag{30}$$

$$a_{Total} = 0.28a_1 + 0.55a_2 + 0.83a_3 \tag{31}$$

$$a_1 = 0.02561 \times \left(I_{LVE\_test}\right) \tag{32}$$

$$a_2 = 0.02519 v_{norm}^{0.5017} \tag{33}$$

$$a_3 = 5.2657 \times (I_{D^R\_test}) \tag{34}$$

$$I_{D^R\_test} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}\left(D_i^R - D_{fit}^R\right)^2} \tag{35}$$

$$I_{LVE\_test} = \frac{1}{n_{LVE}}\sum_{i=1}^{N}\sum_{j=1}^{9}\left|\frac{E'_{fit\_i,j} - E'_{fit\_avg,j}}{E'_{fit\_avg,j}}\right| \tag{36}$$

Where:

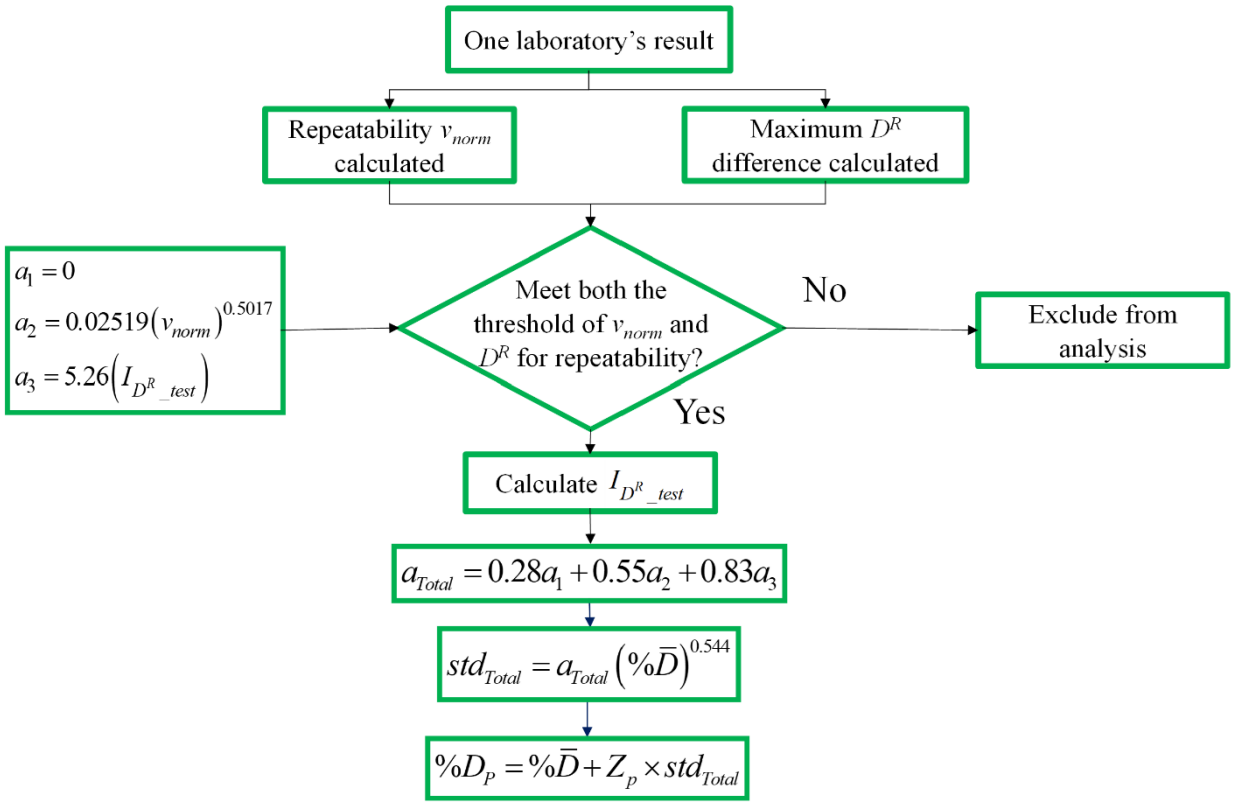
$\%D_P$ = predicted damage at the reliability level P.

Z_P = standard normal deviation corresponding to reliability level P.

std_{Total} = standard deviation of *%Damage*.
$\overline{\%D}$ = average *%Damage*.
I_{LVE_test} = linear viscoelastic ($|E^*|$ model) variation index from test results.
$I_{D^R\_test}$ = D^R variation index from test results.
n = number of cyclic fatigue test specimens.
n_{LVE} = number of dynamic modulus test specimens.
D_i^R = D^R value for ith specimen.
D_{fit}^R = average D^R value for all the n specimens.
$E'_{fit\_i,j}$ = fitted storage modulus from the fitted curve for the ith specimen and jth temperature and frequency.
$E'_{fit\_avg,j}$ = storage modulus from the fitted curve using all of n_{LVE} specimens at the jth temperature and frequency.

According to Ghanbari's framework, all the ILS results that meet the proposed repeatability and reproducibility precision limits were used to identify the maximum uncertainty in the *%Damage* predictions from FlexPAVE.[21] Figure 6 presents the procedure used to assess the repeatability of *%Damage* predictions, which defines the uncertainty in *%Damage* in one test result (i.e., three test specimens) obtained in a single laboratory on a single material. As shown in this figure, ultimately, v_{norm} is selected for defining damage characteristic curve repeatability and, thus, used herein.

The research team calculated the repeatability v_{norm} and maximum D^R difference across three specimens for each laboratory and mixture. If both v_{norm} and maximum D^R results met the repeatability precision limits—indicating the repeatability was acceptable—then the repeatability v_{norm} and $I_{D^R\_test}$ were used to calculate a_{Total} in equation 31. Since this analysis focused on identifying uncertainty induced from the cyclic fatigue test alone, a_1—which reflects the contribution of variation in the $|E^*|$ curve results to *%Damage* uncertainty—was set to zero. The researchers then used the calculated a_{Total} to calculate the standard deviation in *%Damage* using equation 30. Finally, the standard deviation was used to calculate the 95 percent confidence interval of *%Damage* using equation 29. A *%Damage* of 15 percent was used in all analyses conducted because it reflects a critical point in the transfer function that relates *%Damage* to field observations of cracking.[22] At less than 15 *%Damage*, little fatigue cracking is observed on pavement surfaces, but after this limit, small increments in *%Damage* correspond to notable increases in field cracking.



Source: FHWA.

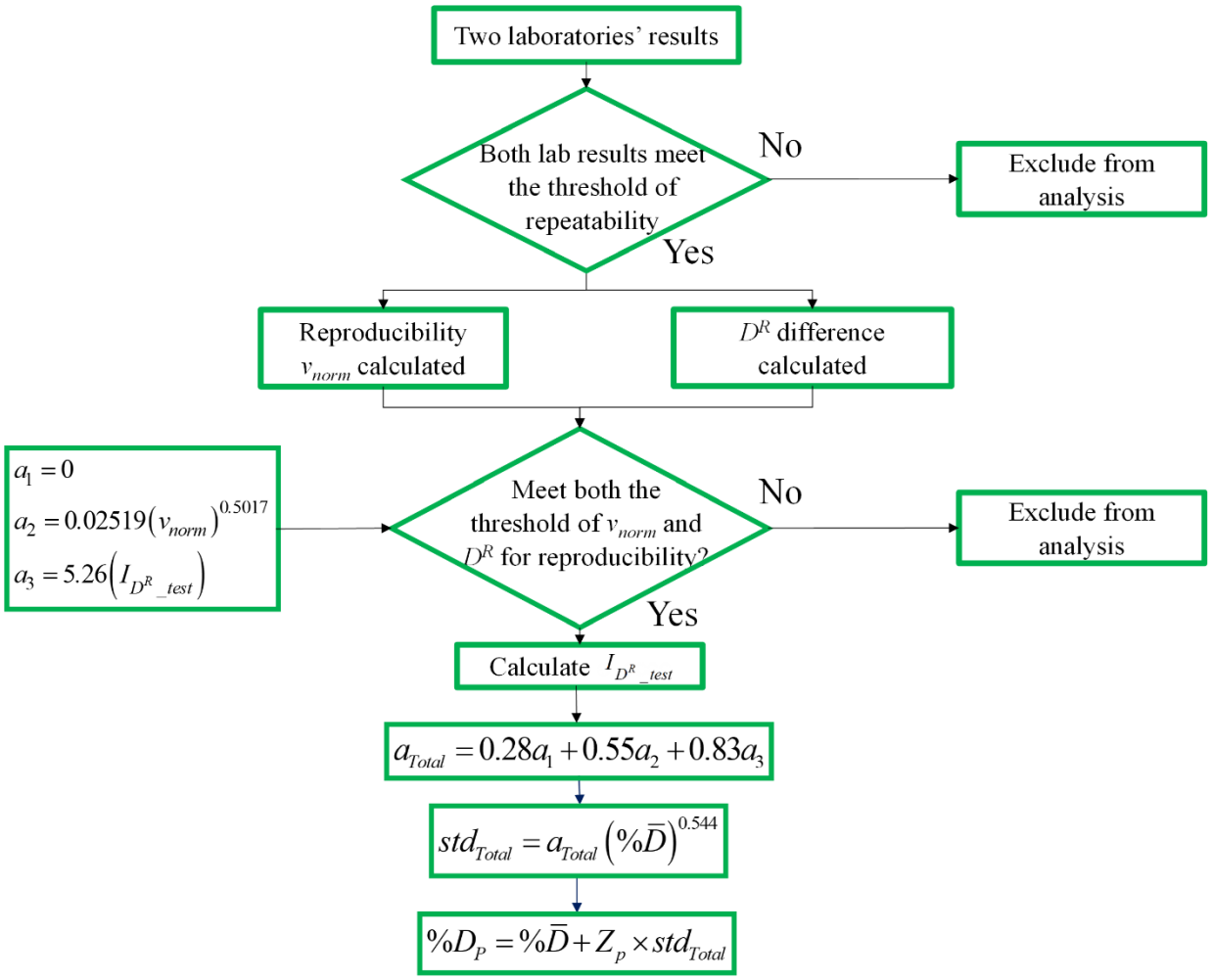
Figure 6. Flowchart. The procedure to assess uncertainty in *%Damage* predictions under repeatability conditions.

Figure 7 presents the procedure used to assess uncertainty in *%Damage* predictions for any two laboratory results obtained using the same material under reproducibility conditions. First, the research team evaluated the individual results of the two laboratories under consideration. If the results exceeded the repeatability limits for v_{norm} or D^R, the data were not used for reproducibility analysis. If the differences in the test results between the two laboratories exceeded reproducibility limits for these parameters, the data were also rejected from the analysis. If the pair of laboratories' results met both repeatability and reproducibility limits, then the reproducibility v_{norm} and $I_{D^R\_test}$ were used to calculate a_{Total} in equation 31. Consistent with repeatability analyses, a_1—which reflects the contribution of variation in the $|E^*|$ curve results to *%Damage* uncertainty—was set to zero. It should be noted that reproducibility $I_{D^R\_test}$ uses the average D^R for each laboratory as D_i^R. The calculated a_{Total} was then used to calculate the standard deviation in *%Damage* using equation 30. Finally, the standard deviation was used to calculate the 95 percent confidence interval of *%Damage* using equation 29. An average *%Damage* of 15 percent was used in all analyses, consistent with the repeatability analyses.

The repeatability uncertainty was computed for each of the seven laboratory's results for each mixture and specimen geometry combination. The reproducibility uncertainty was calculated for each pair of the seven laboratory's results for each mixture and geometry combination. In total,

there were 42 results for the repeatability analysis and 126 (i.e., 21×6) results for the reproducibility analysis before results that did not meet the proposed precision limits were rejected.

Ultimately, the results of the FlexPAVE analysis defined the maximum and median repeatability and reproducibility standard deviations of *%Damage* for each mixture and specimen geometry combination. These maximum and median standard deviations were used to perform assessments of the maximum and common uncertainty expected for test results that meet the precision statements established in this study.



Source: FHWA.

Figure 7. Flowchart. The procedure to assess uncertainty in *%Damage* predictions under reproducibility conditions.

Uncertainty in Apparent Damage Capacity

AMPT cyclic fatigue test results can also be used to calculate an index parameter to indicate asphalt mixture fatigue resistance, termed the apparent damage capacity (S_{app}) and defined in equation 37.[23] The S_{app} value incorporates the effects of the material's modulus and toughness

on its fatigue resistance and measures the amount of fatigue damage the material can tolerate under loading. The temperature for the S_{app} calculation is equal to the average of high- and low-temperature grades at the standard 98-percent reliability level minus 3 °C for the climate where the mixture is to be placed. Higher S_{app} values indicate increased fatigue resistance. The S_{app} value is sensitive to mixture factors (e.g., aggregate gradation, asphalt binder content, reclaimed asphalt pavement content, and asphalt binder grade), compaction, and aging and meets general expectations regarding the effects of these parameters on fatigue-cracking performance.[23] Similar to the *%Damage* analysis, the implications of the combined variation in the damage characteristic curve and failure criterion results permitted under the proposed precision limits on the uncertainty in S_{app} were evaluated.

$$S_{app} = 1000^{\frac{\alpha}{2}-1} \frac{a_T^{\frac{1}{\alpha+1}} \left(\frac{D^R}{C_{11}} \right)^{1/C_{12}}}{\left| E^* \right|^{\alpha/4}} \tag{37}$$

Where:

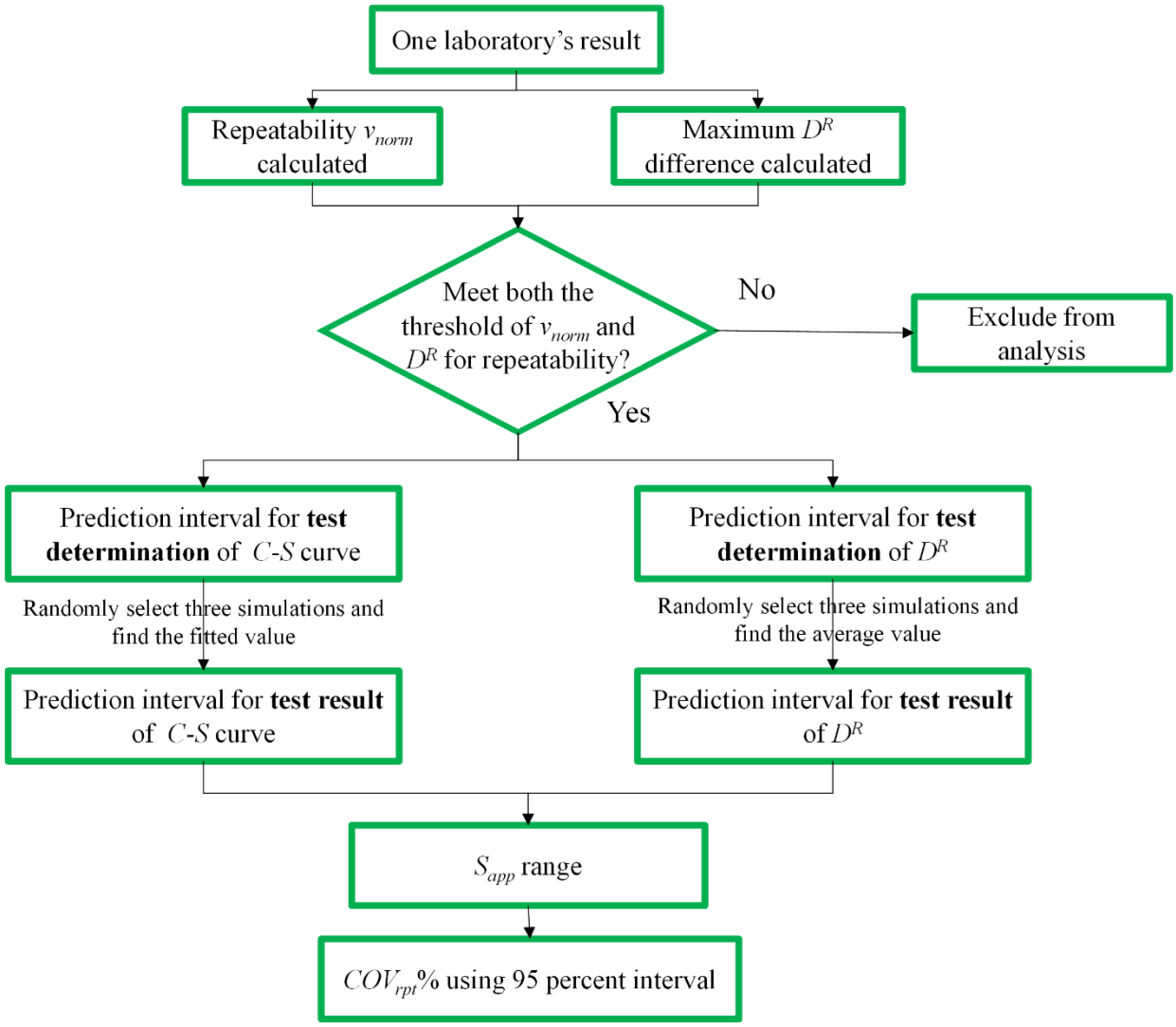
α = damage growth rate.

a_T = time–temperature shift factor at a given temperature.

To evaluate the uncertainty in S_{app}, the S_{app} values were calculated using the results of individual test determinations. The researchers calculated the S_{app} value for each laboratory, termed the "measured" S_{app} values herein. The measured S_{app} values were used to assess the variation of S_{app} values among test determinations within a given laboratory and the reproducibility of S_{app} results among laboratories. In addition, the researchers developed a framework to further evaluate S_{app} uncertainty due to the combined uncertainty that is analogous to the framework that Ghanbari used to quantify *%Damage* uncertainty.[21] The repeatability of S_{app} values among test determinations could be directly assessed from the measured S_{app} values.

However, the repeatability of S_{app} results could not be directly measured from the ILS results since each laboratory only generated one test result on a given mixture and specimen geometry combination. The researchers used supplementary analysis to assess the repeatability of S_{app} results. Figure 8 presents the flowchart for assessing the uncertainty in S_{app} in one test result (i.e., three test specimens) obtained in a single laboratory on a single material. Consistent with the *%Damage* analysis, data were only incorporated into the analysis if the repeatability limits for v_{norm} or D^R were met when the uncertainty in S_{app} was analyzed. The researchers used results meeting the repeatability limits to generate independent predictive intervals for damage characteristic curves and D^R using the MCMC method. The predictive intervals represent the

range within which a new test determination (either damage characteristic curve or D^R value) will fall within with 95 probability.



Source: FHWA.

Figure 8. Flowchart. The procedure to assess uncertainty in S_{app} values under repeatability conditions.

Then, 5,000 possible damage characteristic curves and D^R determinations were simulated based on the prediction intervals. Next, three sets of damage characteristic curves and D^R determinations were randomly selected from the 5,000 sets of simulated parameters to reflect three test determinations that comprise a possible test result. The collective results of the three simulated test determinations were used to calculate an average D^R value and damage characteristic curve fitting coefficients in equation 9, which constitute an AMPT cyclic fatigue test result.

The process of randomly sampling three test determinations to constitute a possible test result was conducted 5,000 times with replacement and used to calculate 5,000 corresponding simulated S_{app} values from a given lab. Consistent with the other analysis within this study, consistent $|E^*|$ were used in all S_{app} calculations for a given specimen geometry and mixture combination, irrespective of the lab.

The 95 percent confidence interval was defined using the 5,000 simulated S_{app} values. The values falling within the bounds, corresponding to 2.5 and 97.5 percentiles, were used to calculate the S_{app} COV according to equation 38. The analysis was conducted for each mixture, laboratory, and specimen geometry combination.

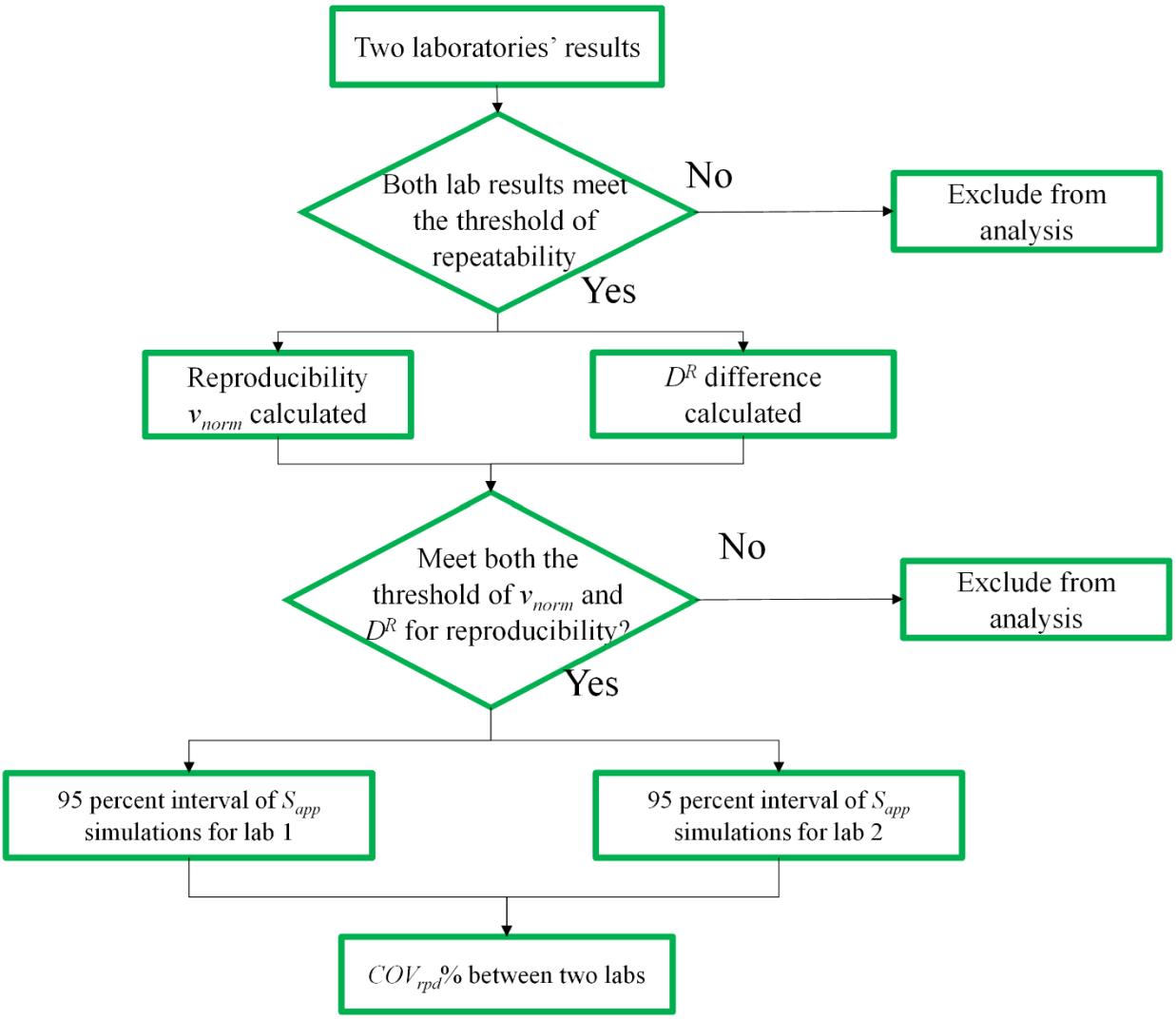
$$COV_{rpt}\% = \frac{\sum_{i=1}^{N}\left(S_{app-i} - \bar{S}_{app}\right)/(n_S - 1)}{\bar{S}_{app}} \times 100 \quad (38)$$

Where:

$COV_{rpt}\%$ = repeatability COV for S_{app}.
n_S = number of simulations in the 95-percent confidence interval, 4,750.
$S_{app\text{-}i}$ = S_{app} value for the ith simulation.
$\bar{S}_{app}$ = average S_{app} value across n_S simulations.

Supplementary analysis was also conducted to provide a measure of the maximum expected variation in S_{app} among the laboratories, given the proposed reproducibility precision limits. Figure 9 presents the framework used to assess the uncertainty in S_{app} obtained using the same material and specimen geometry under reproducibility conditions. Consistent with the *%Damage* analysis, only data that met both repeatability and reproducibility precision limits were included in the analysis of the uncertainty in S_{app} under reproducibility conditions. If a pair of laboratories' results met both repeatability and reproducibility limits, then the 2.5 percentile and 97.5 percentile S_{app} values from each of the two laboratories—obtained from the repeatability analysis from the process depicted in figure 8—were used to calculate the reproducibility COV% for S_{app}, according to equations 39 and 40. That is, the reported reproducibility COV% reflects

the maximum possible between-laboratory variations in S_{app} results based on the 95 percent confidence intervals for S_{app} results within a given lab.



Source: FHWA.

Figure 9. Flowchart. The procedure to assess uncertainty in S_{app} values under reproducibility conditions.

$$COV_{rpd}\% = \max\left(COV(S_{app-1}^{2.5}, S_{app-2}^{97.5}), COV(S_{app-2}^{2.5}, S_{app-1}^{97.5})\right) \quad (39)$$

$$COV(x,y)=\sqrt{\left(x-\frac{x+y}{2}\right)^2+\left(y-\frac{x+y}{2}\right)^2}\Big/\frac{x+y}{2}\times 100 \quad (40)$$

Where:

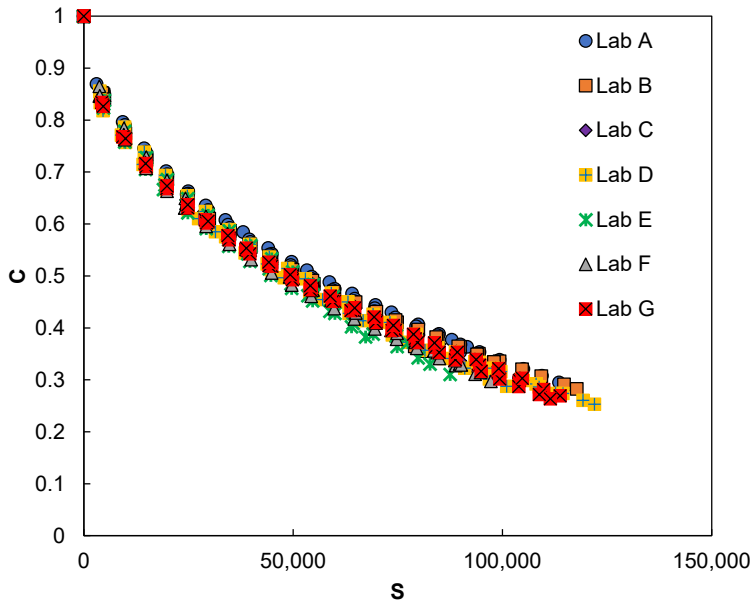
$COV_{rpd}\%$ = reproducibility COV for S_{app}.
$S_{app\text{-}1}^{2.5}$ = the 2.5 percentile of S_{app} for laboratory 1.
$S_{app\text{-}1}^{97.5}$ = the 97.5 percentile of S_{app} for laboratory 1.
$S_{app\text{-}2}^{2.5}$ = the 2.5 percentile of S_{app} for laboratory 2.
$S_{app\text{-}2}^{97.5}$ = the 97.5 percentile of S_{app} for laboratory 2.
COV(x,y) = COV between x and y.

Consistent with the *%Damage* analysis, the researchers computed the S_{app} repeatability uncertainty for the results for each mixture and specimen geometry combination from the seven laboratories. The researchers computed the reproducibility uncertainty for each pair of the seven laboratory's results for each mixture and geometry combination. In total, there were 42 results for the repeatability analysis and 126 (i.e., 21×6) results for the reproducibility analysis before the results that did not meet the proposed precision limits were rejected. The maximum and median repeatability and reproducibility S_{app} COV% of S_{app} for each mixture and specimen geometry combination were extracted to provide measures of the typical and maximum possible uncertainty that can be expected for laboratory results meeting the proposed precision limits.

CHAPTER 3. RESULTS

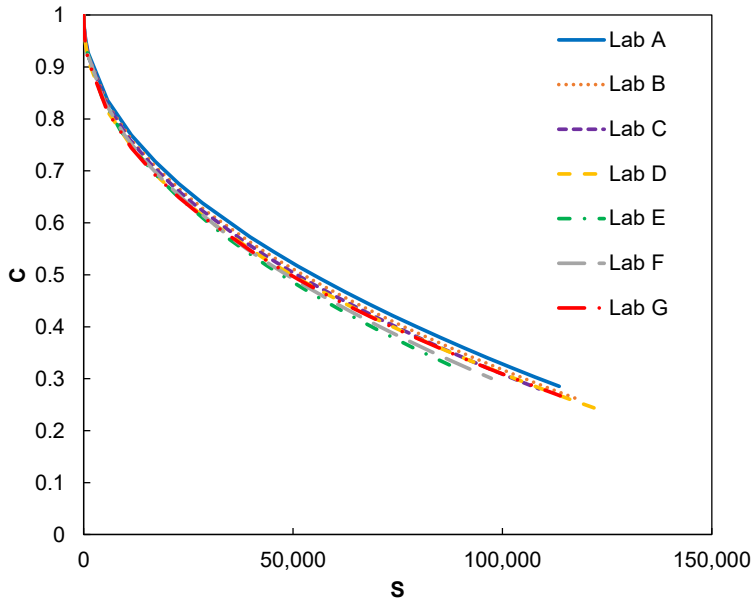
ILS RESULTS OVERVIEW

Figure 10 through figure 33 show the ILS test results for each mixture and test geometry. Graphs of individual specimen damage characteristic curves convey the results of individual test determinations. In contrast, the fitted damage characteristic curve graphs convey the test result (the curve best fit to the results of the three test determinations). The bar graphs of the D^R results show the average D^R values from each laboratory, with error bars denoting the range of individual specimen values from each laboratory. The cumulative $(1-C)$ values versus fatigue life graphs provide another means to visualize failure criterion repeatability and reproducibility. Based on visual observation of the results, the 9.5-mm mixture (small- and large-specimen geometries) has the least variation among and within laboratories, and the 19-mm small and 25-mm large materials have the most variation among and within the laboratories. Trends among the repeatability and reproducibility of the two specimen geometries for the 9.5-mm and 12.5-mm mixtures are difficult to ascertain visually.



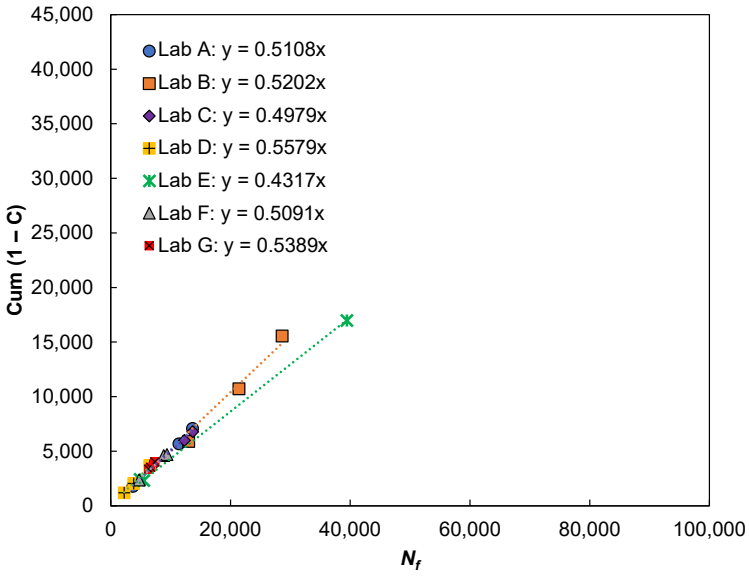
Source: FHWA.

Figure 10. Graph. Individual specimen damage characteristic curves for 9.5-mm small specimens.



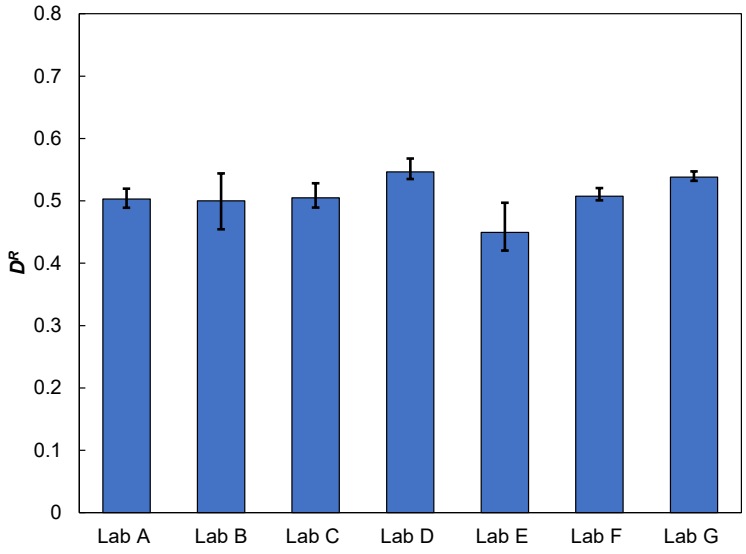
Source: FHWA.

Figure 11. Graph. Fitted damage characteristic curves for 9.5-mm small specimens.



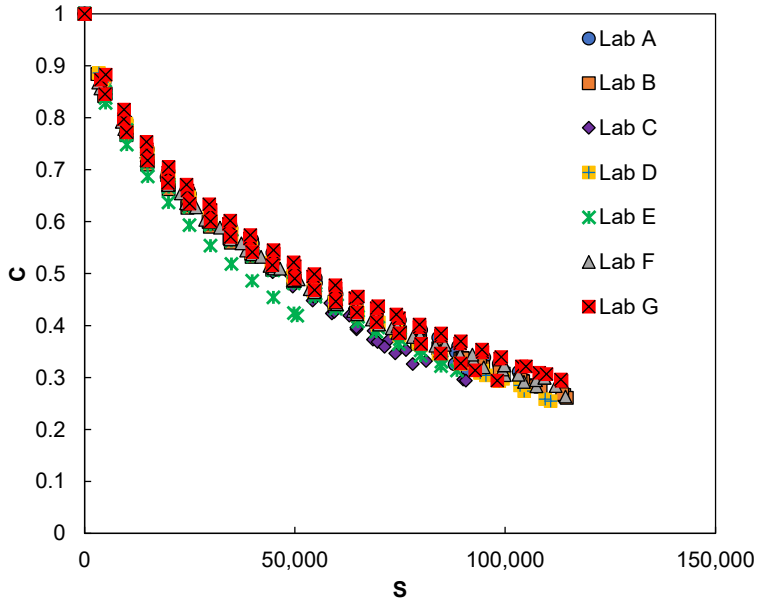
Source: FHWA.

Figure 12. Graph. Cumulative (1−*C*) results from individual tests for 9.5-mm small specimens.



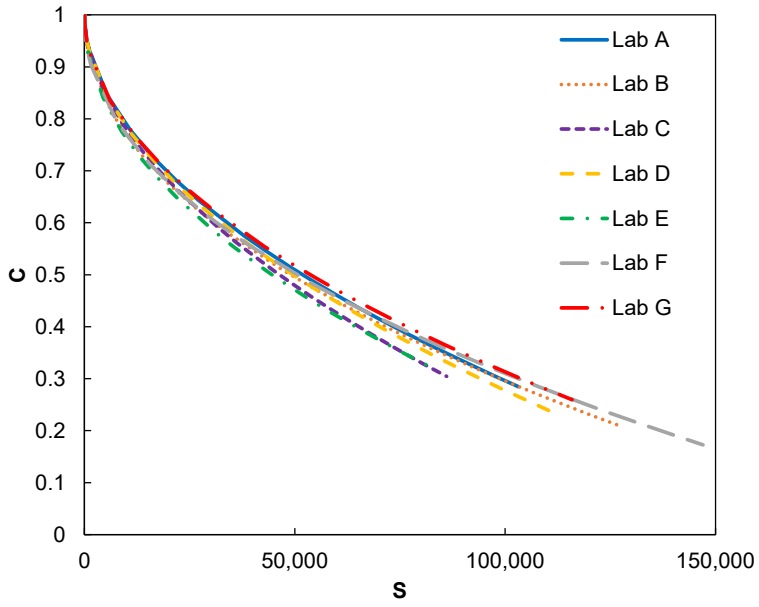
Source: FHWA.

Figure 13. Graph. D^R results for 9.5-mm small specimens.



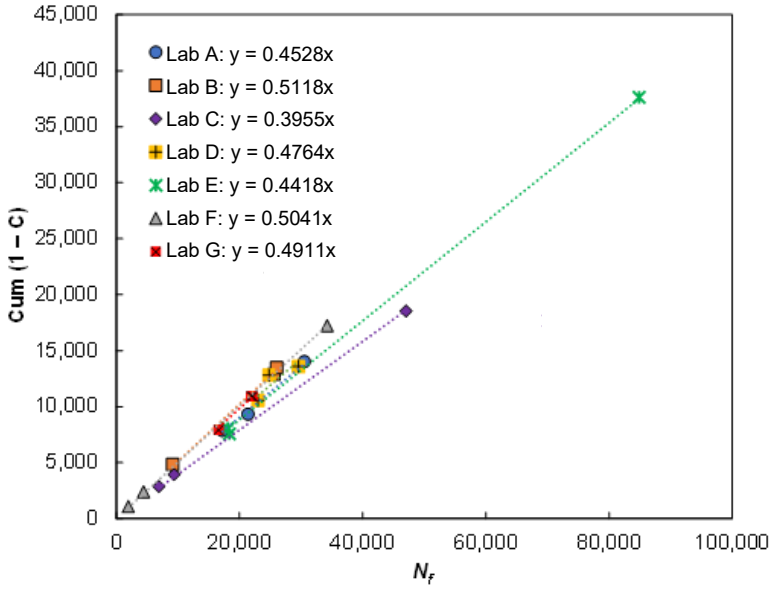
Source: FHWA.

Figure 14. Graph. Individual specimen damage characteristic curves for 9.5-mm large specimens.



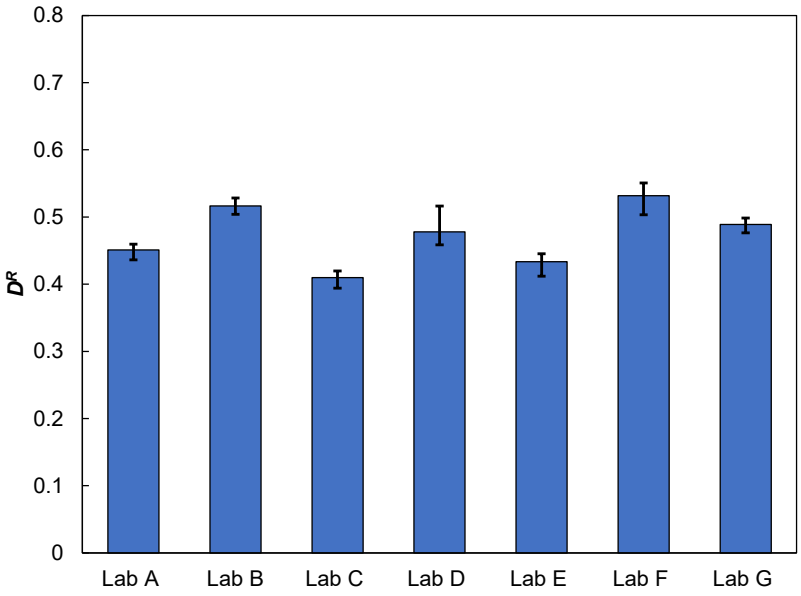
Source: FHWA.

Figure 15. Graph. Fitted damage characteristic curves for 9.5-mm large specimens.



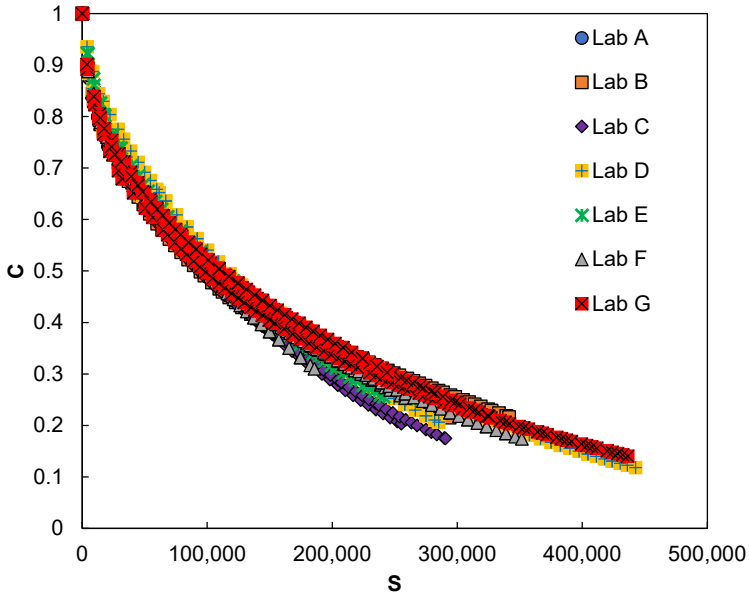
Source: FHWA.

Figure 16. Graph. Cumulative (1−*C*) results from individual tests for 9.5-mm large specimens.



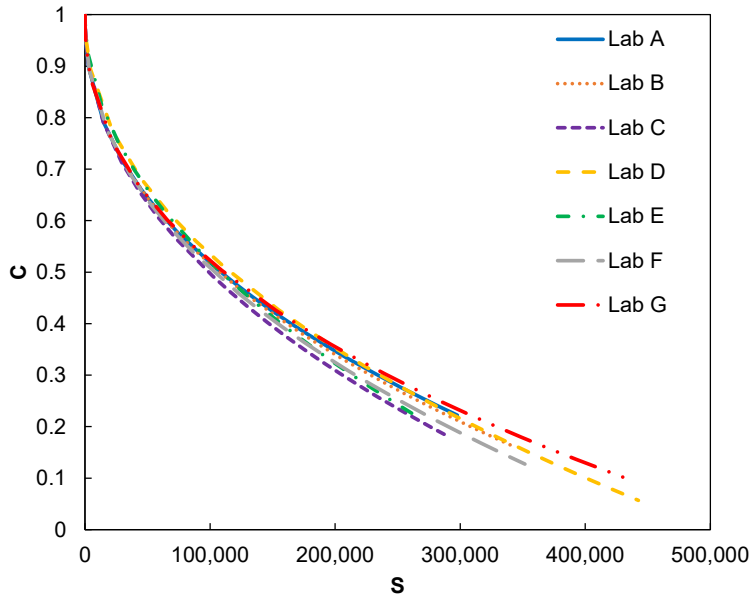
Source: FHWA.

Figure 17. Graph. D^R results for 9.5-mm large specimens.



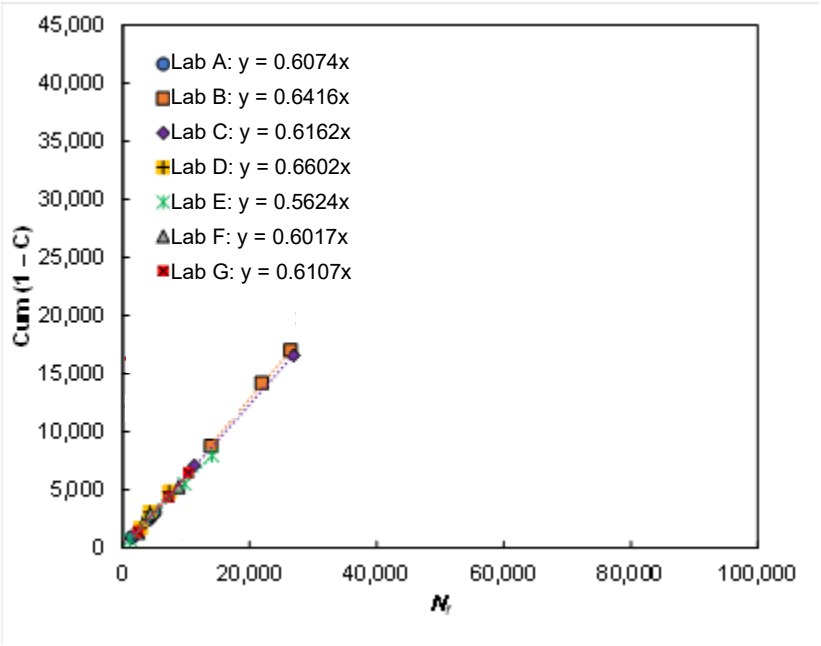
Source: FHWA.

Figure 18. Graph. Individual specimen damage characteristic curves for 12.5-mm small specimens.



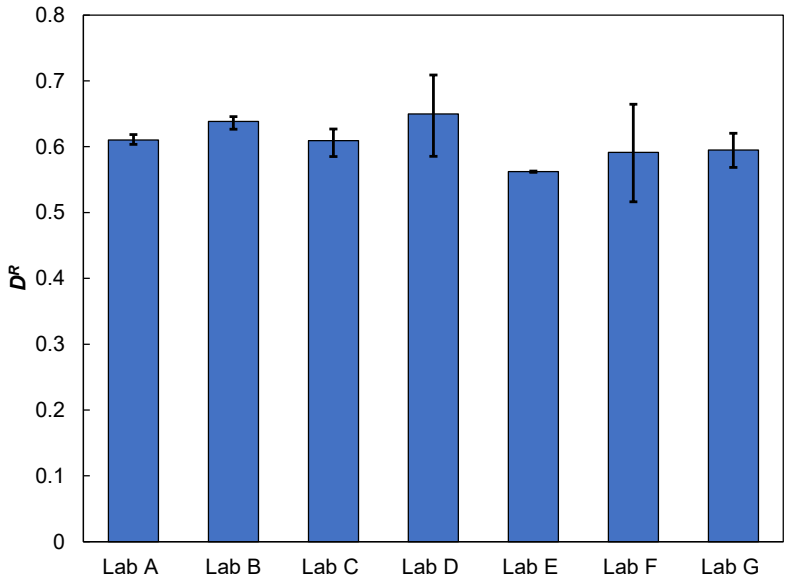
Source: FHWA.

Figure 19. Graph. Fitted damage characteristic curves for 12.5-mm small specimens.



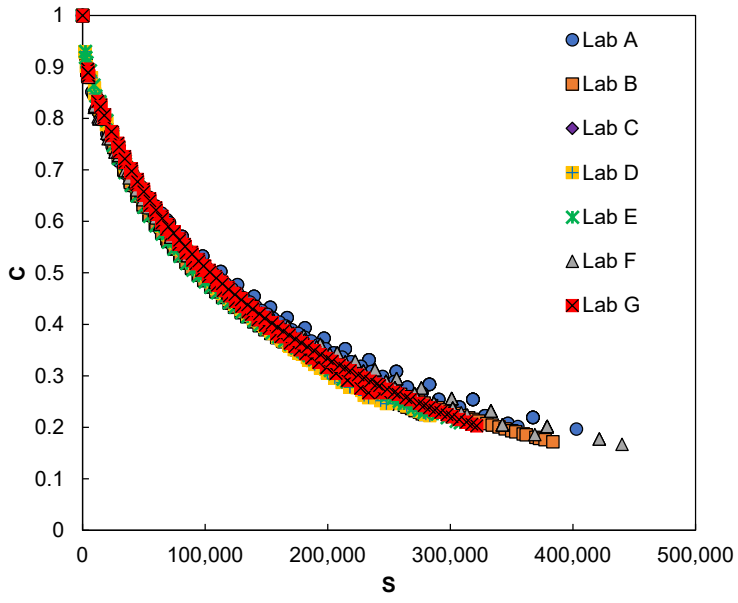
Source: FHWA.

Figure 20. Graph. Cumulative (1−*C*) results from individual tests for 12.5-mm small specimens.



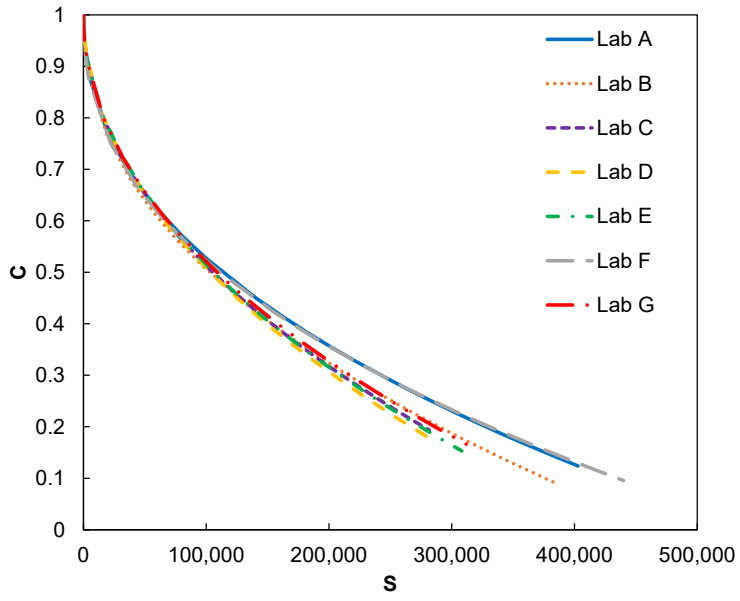
Source: FHWA.

Figure 21. Graph. DR results for 12.5-mm small specimens.



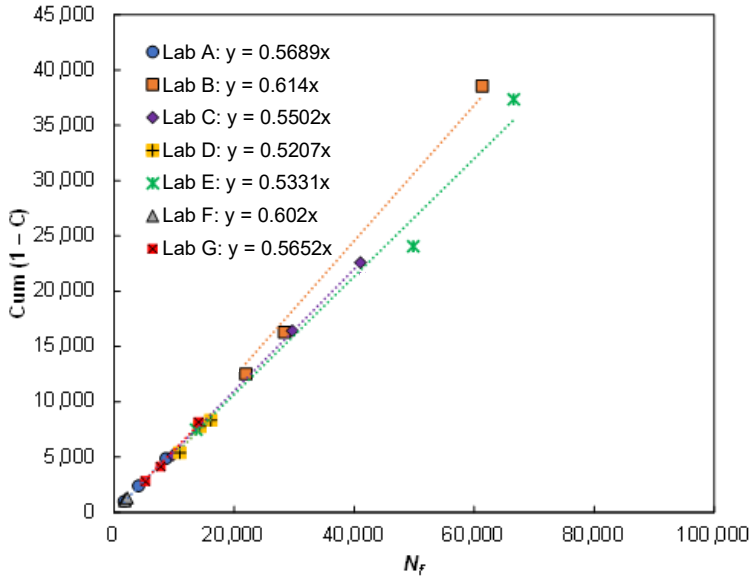
Source: FHWA.

Figure 22. Graph. Individual specimen damage characteristic curves for 12.5-mm large specimens.



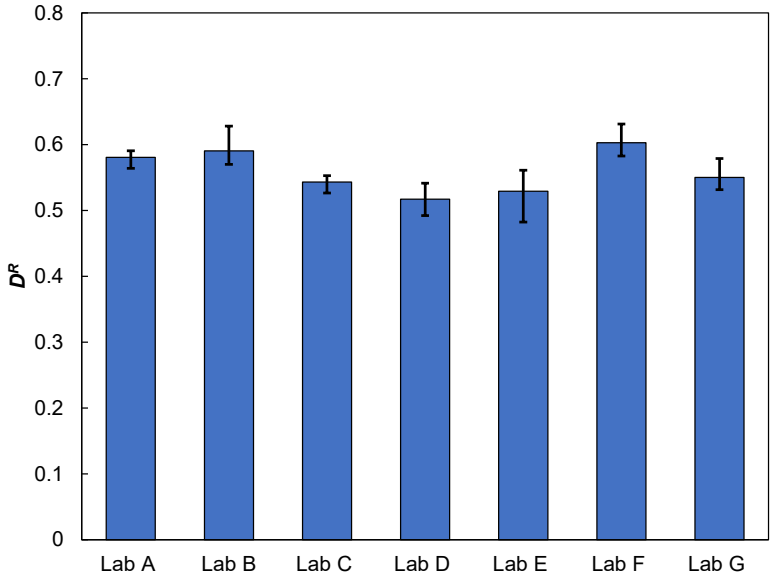
Source: FHWA.

Figure 23. Graph. Fitted damage characteristic curves for 12.5-mm large specimens.



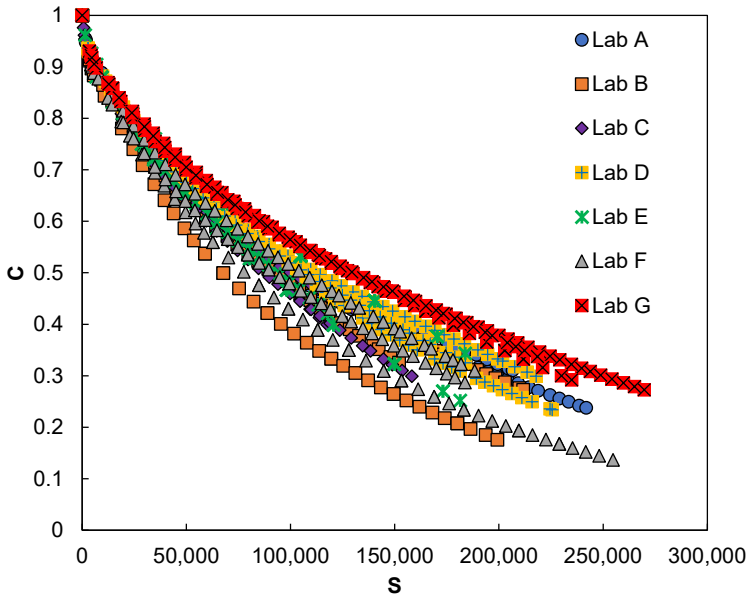
Source: FHWA.

Figure 24. Graph. Cumulative (1−*C*) results from individual tests for 12.5-mm large specimens.



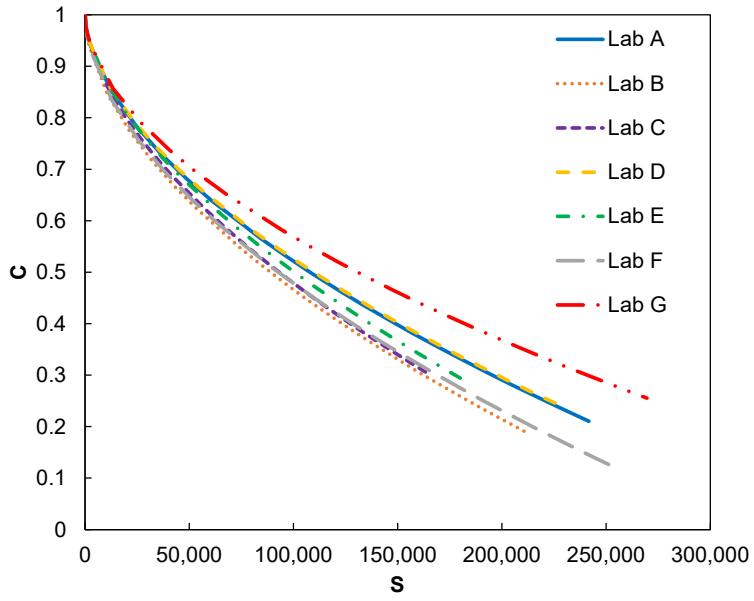
Source: FHWA.

Figure 25. Graph. D^R results for 12.5-mm large specimens.



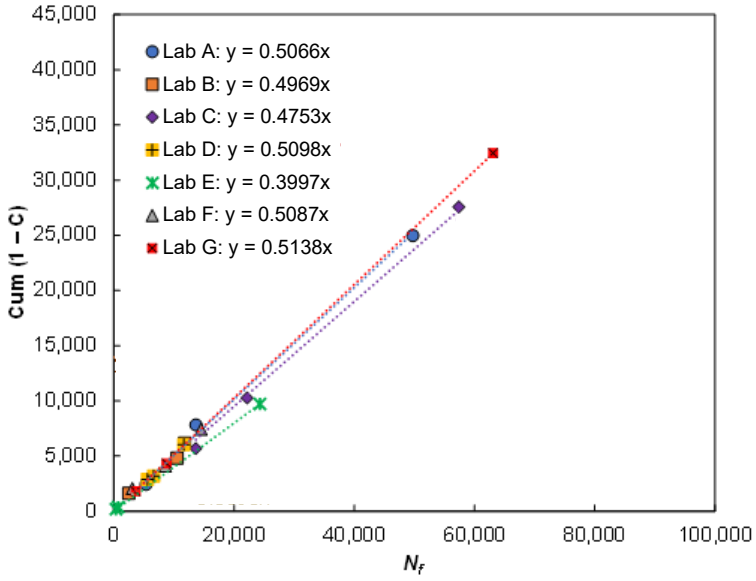
Source: FHWA.

Figure 26. Graph. Individual specimen damage characteristic curves for 19-mm small specimens.



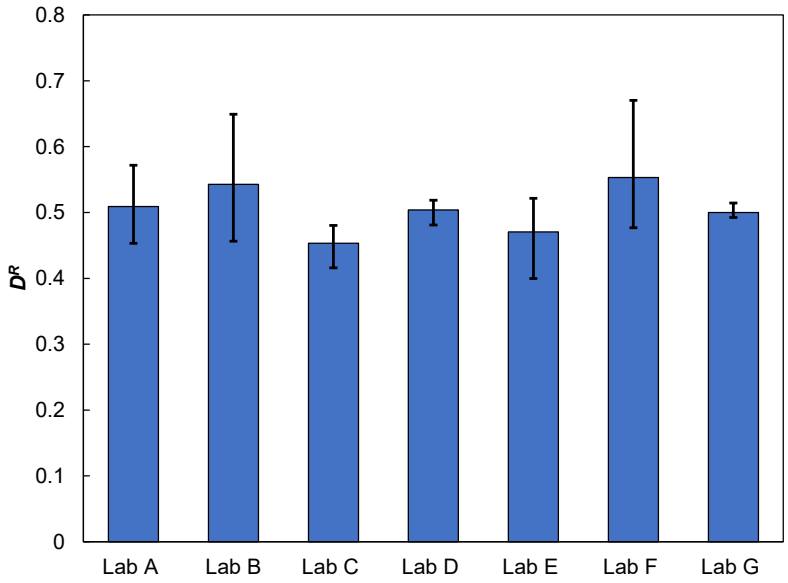
Source: FHWA.

Figure 27. Graph. Fitted damage characteristic curves for 19-mm small specimens.



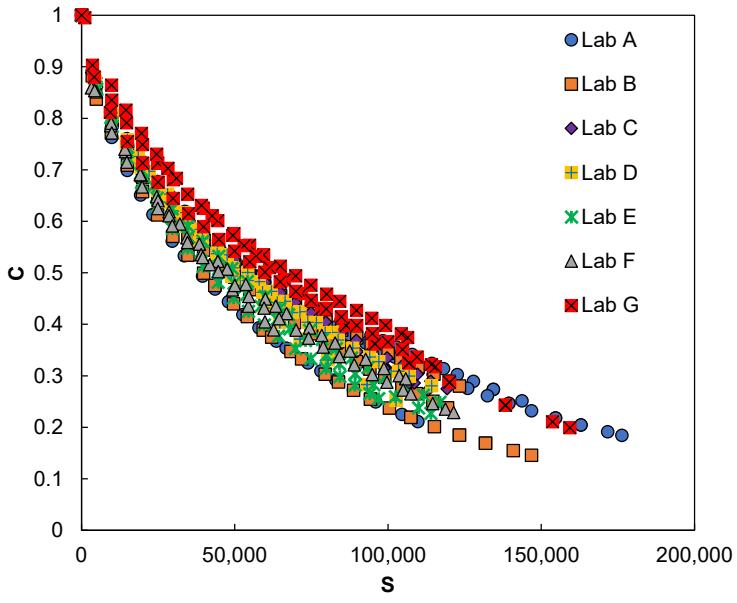
Source: FHWA.

Figure 28. Graph. Cumulative (1−*C*) results from individual tests for 19-mm small specimens.



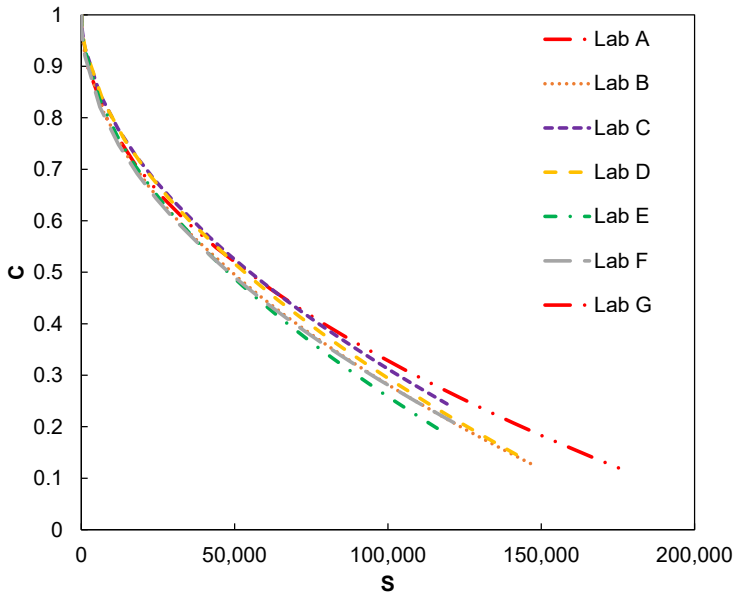
Source: FHWA.

Figure 29. Graph. D^R results for 19-mm small specimens.



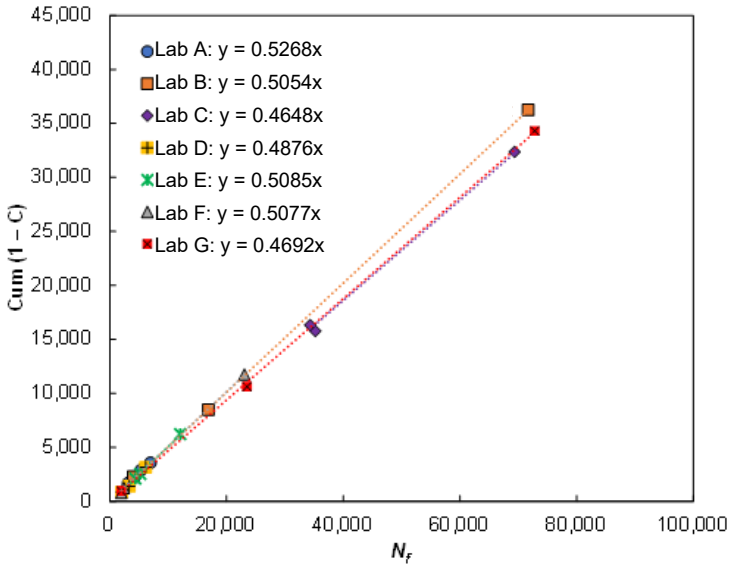
Source: FHWA.

Figure 30. Graph. Individual specimen damage characteristic curves for 25-mm large specimens.



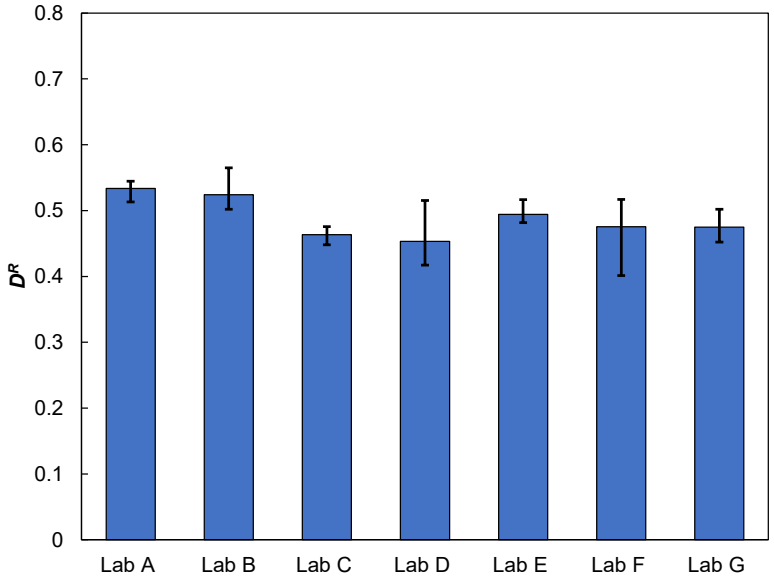
Source: FHWA.

Figure 31. Graph. Fitted damage characteristic curves for 25-mm large specimens.



Source: FHWA.

Figure 32. Graph. Cumulative (1−*C*) results from individual tests for 25-mm large specimens.



Source: FHWA.

Figure 33. Graph. D^R results for 25-mm large specimens.

ILS RESULTS STATISTICAL ANALYSIS

Consistency Statistics

The researchers evaluated the consistency of the test results using the seven selected single-point parameters to assess the damage characteristic curve and failure criterion per ASTM E691.[(4)] The analysis yielded 84 consistency statistics (42 for η and 42 for θ) for each single-point parameter. Table 4 summarizes the percentage of the consistency statistics that exceed the critical value for each parameter. The analysis indicated a high degree of consistency within and between laboratories. All of the flagged data failed the within-laboratory consistency test. All of the inconsistent results belong to laboratory E's 9.5-mm large-mixture results corresponding to the C values at S_{min}, S_{mean}, and S_{max} and shape factors at S_{mean} and S_{max}.

The researchers reviewed laboratory E's 9.5-mm large-specimen results and identified no procedural or data quality issues. They further investigated the impact of laboratory E's inconsistent results by conducting the repeatability and reproducibility analysis for the 9.5-mm mixture large-specimen results, both including laboratory E's results and excluding them. The comparison indicated that excluding the laboratory E 9.5-mm mixture large-specimen data provides more consistent trends with respect to NMAS than when the results are included. Therefore, laboratory E's 9.5-mm mixture large-specimen data were ultimately excluded from the statistical analysis used to develop the precision statement. The following results do not include laboratory E's 9.5-mm large-specimen data.

Table 4. Percentage of data exceeding critical consistency value.

Percent	C at S_{min}	C at S_{mean}	C at S_{max}	Shape at S_{min}	Shape at S_{mean}	Shape at S_{max}	D^R	All
η (between -lab)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
θ (within-lab)	2.38	2.38	2.38	0.00	2.38	2.38	0.00	2.04

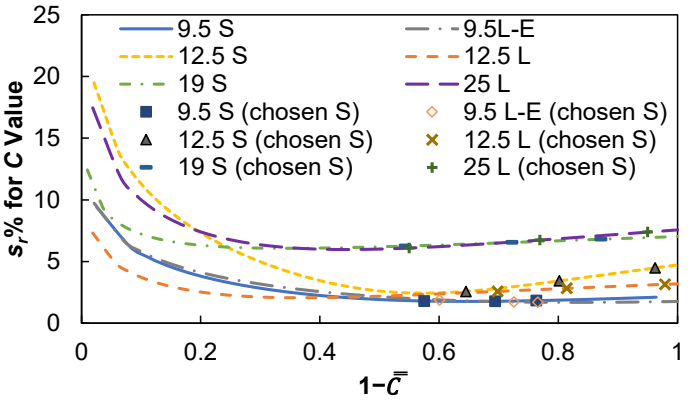
Single-Point Measure Repeatability and Reproducibility Analysis Conducted Following ASTM E691

The repeatability (s_r%) and reproducibility ($s_{\bar{x}}$%) COV results for damage characteristic curve are presented in table 5. The column labeled "9.5-mm large-E" excludes laboratory E data; the remaining tables presented in this report follow this naming structure. The results show that the COVs change with the chosen S value for analysis in terms of both the C value and shape factor. However, there is no clear trend between the chosen S value and the corresponding COV that suits all the mixtures with different geometry. The repeatability and reproducibility COVs for the damage characteristic curve parameters are generally higher for the 19-mm and 25-mm mixtures than the smaller NMAS mixture results for a given specimen geometry, which matches expectations. However, trends for NMAS are less clear in the D^R results. The D^R COV results show that the 19-mm small mixture has the highest repeatability COV, and 9.5-mm large mixture has the highest reproducibility COV, even when laboratory E's data are excluded.

Table 5. Repeatability and reproducibility COV.

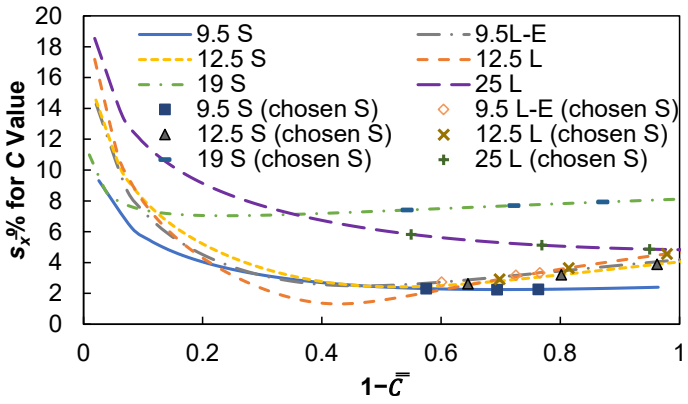
Parameters	9.5-mm Small	9.5-mm Small	9.5-mm Large-E	9.5-mm Large-E	12.5-mm Small	12.5-mm Small	12.5-mm Large	12.5-mm Large	19-mm Small	19-mm Small	25-mm Large	25-mm Large
	s_r%	$s_{\bar{x}}$%	s_r%	$s_{\bar{x}}$%	s_r%	$s_{\bar{x}}$%	s_r%	$s_{\bar{x}}$%	s_r%	$s_{\bar{x}}$%	s_r%	$s_{\bar{x}}$%
C at S_{min}	1.79	2.30	1.88	2.74	2.57	2.62	2.58	2.91	6.24	7.40	6.10	5.82
C at S_{mean}	1.77	2.24	1.71	3.17	3.42	3.21	2.82	3.65	6.54	7.69	6.74	5.12
C at S_{max}	1.82	2.25	1.69	3.32	4.47	3.88	3.15	4.56	6.80	7.92	7.39	4.86
Shape at S_{min}	2.17	2.64	2.41	2.52	2.76	2.46	2.19	1.58	6.07	7.15	6.02	6.80
Shape at S_{mean}	1.91	2.41	2.08	2.54	2.43	2.46	2.35	2.15	6.19	7.34	6.08	5.83
Shape at S_{max}	1.82	2.33	2.00	2.59	2.85	2.80	2.60	2.98	6.35	7.50	6.42	5.35
D^R	5.24	6.19	4.14	9.27	6.41	4.85	4.83	5.82	12.85	7.07	7.68	6.23

Figure 34 and figure 35 present the variation in the repeatability and reproducibility COVs as a function of the average *C* value for *C* values at a specific *S* value and shape factors, respectively. The trend for each mixture in terms of either *C* value or shape factor was captured by using a finely spaced increment with ΔS, where ΔS is the increment of damage (1,000 units until a maximum value of 10^5). Therefore, the trends conveyed by the lines in figure 34 and figure 35 reflect the true trend since interpolation to form the lines was minimal. The average *C* values corresponding to S_{min}, S_{mean}, and S_{max} are identified as points (chosen *S*) to provide context to the results shown in table 5. Both *C* values and shape factor show similar trends in their COVs with respect to the average *C* value at which the parameter is evaluated. These results indicate that shape factor, although expressed as area of damage characteristic curve results, provides similar information to the *C* value alone. The results also suggest that defining the test precision based on *C* (or shape factor) at the average *S* corresponding to $(1-\bar{\bar{C}}) = 0.7$ may help avoid bias to the chosen *S* value when a single point measure of the damage characteristic curve is used. The results clearly indicate higher variation within and across laboratories in the 19-mm and 25-mm mixtures compared to the smaller NMAS mixtures.



Source: FHWA.
L = large; L-E = large with exclusion of laboratory E; S = small.

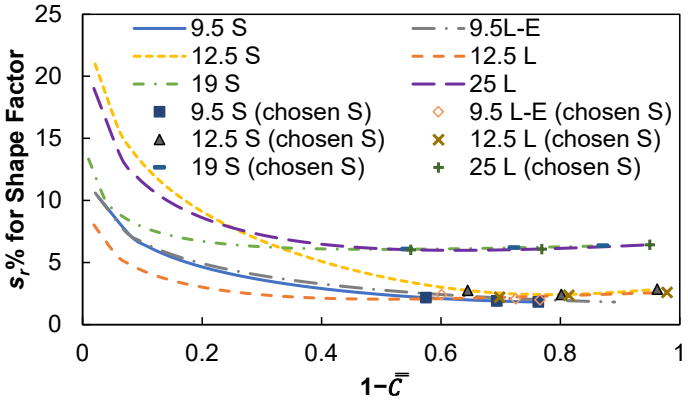
A. Sensitivity analysis of repeatability COV (s_r%) on single-point *C* value selection.



Source: FHWA.

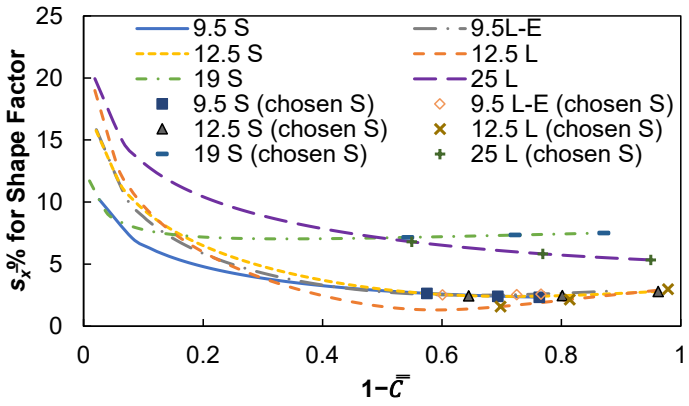
B. Sensitivity analysis of reproducibility COV ($s_{\bar{x}}$%) on single-point *C* value selection.

Figure 34. Graphs. Sensitivity analysis of repeatability and reproducibility COV on single-point *C* value selection.



Source: FHWA.

A. Sensitivity analysis of repeatability COV (s_r%) on single-point shape factor selection.



Source: FHWA.

B. Sensitivity analysis of reproducibility COV ($s_{\bar{x}}$%) on single-point shape factor selection.

Figure 35. Graphs. Sensitivity analysis of repeatability and reproducibility COV on single-point shape factor selection.

The researchers developed repeatability and reproducibility limits for the single-point measures (*C* value and shape factor) using the repeatability and reproducibility COVs with $(1-\bar{\bar{C}})=0.7$ (i.e., an average *C* value of 0.3) for damage characteristic curve results based on the results of the sensitivity analysis. Limits for D^R were defined using the repeatability and reproducibility standard deviation because D^R is a single-point measure that is independent of the selection of *S*. Following the ASTM E1169-20 and ASTM C670-15 standards, the repeatability and reproducibility limits define the variation in the result that will not be exceeded with 95 percent confidence.[4,5] The results are shown in table 6 through table 8. The limits for the repeatability of three specimens are conveyed as the maximum allowable difference among three specimens (COV%), according to equation 17. The reproducibility limits convey the maximum allowable difference between two test results obtained in different laboratories (COV%), defined according to equation 18. In general, the repeatability and reproducibility limits rise with increasing NMAS for a given specimen geometry. The only exception is the D^R reproducibility limits, which do not reflect a clear trend with respect to NMAS. Note that the repeatability limits for $|E^*|$ and phase angle in AASHTO T 378-17 depend on modulus and NMAS but the reproducibility limits are independent of NMAS.[9] Thus, the trends in repeatability and reproducibility regarding NMAS observed in this study align with the findings of AASHTO T 378 ILS.

Table 6. Repeatability and reproducibility limits (COV) for *C* value at $(1-\bar{\bar{C}})=0.7$.

Mixtures	Repeatability (*MAR%*)	Reproducibility ($d2s_{rpd}$%)
9.5 S	5.85	6.28
9.5 L	5.75	8.49
12.5 S	9.38	7.90
12.5 L	8.52	8.18
19 S	21.45	21.44
25 L	21.48	14.78

Table 7. Repeatability and reproducibility limits (COV) for shape factor at $(1-\bar{\bar{C}})=0.7$.

Mixtures	Repeatability (*MAR%*)	Reproducibility ($d2s_{rpd}$%)
9.5 S	6.26	6.72
9.5 L	7.03	7.04
12.5 S	8.33	6.72
12.5 L	7.25	4.46
19 S	20.37	20.48
25 L	19.81	17.03

Table 8. Repeatability and reproducibility precision statements (standard deviation) of D^R.

Mixture	Average D^R	Repeatability (*MAR*)	Reproducibility ($d2s_{rpt}$)
9.5 S	0.507	0.088	0.088
9.5 L-E	0.479	0.065	0.124
12.5 S	0.608	0.129	0.083
12.5 L	0.559	0.089	0.091
19 S	0.505	0.214	0.100
25 L	0.488	0.124	0.085

Advanced Analysis of Damage Characteristic Curve Repeatability and Reproducibility Using a Functional Data Metric

Table 9 and table 10 show the v_{norm} values calculated using the ILS test results in terms of repeatability and reproducibility, respectively. The corresponding precision limit results are shown in table 11. These precision limits were defined by separately fitting the gamma function given in equation 28 to the distribution of v_{norm} repeatability and reproducibility results and using the resultant gamma functions to define the v_{norm} limits that would not be exceeded with 95 percent confidence. Analysis for 9.5-mm large mixture was conducted without laboratory E's data. Precision limit analysis for repeatability in table 11 identified the repeatability of one test result (i.e., repeatability among three test determinations).

Generally, increasing the NMAS yields higher repeatability and reproducibility limits for v_{norm} (and hence, indicates higher variance) for a given specimen geometry. Trends with respect to specimen geometry vary, indicating separate precision limits may be warranted for the two specimen geometries. In two cases, 12.5-mm small-specimen geometry and 25-mm mixture large-specimen geometry, the reproducibility v_{norm} limits are smaller than the corresponding repeatability limits. Because v_{norm} for assessing repeatability evaluates variation among test determinations (i.e., among three individual test specimens), and v_{norm} for assessing reproducibility evaluates variation among test results (i.e., results of three test specimens each in two laboratories), they are not directly comparable. The variation in three individual specimens is expected to exceed the variation in the average results of two sets of three specimens, which may explain the observed trends for the 25-mm mixture and 12.5-mm small-specimen results.

Table 9. Repeatability v_{norm} for different laboratories and different materials.

Laboratory	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A	5.5	14.6	0.7	98.3	6.4	527.9
B	12.1	1.5	38.3	21.7	274.8	142.7
C	2.3	6.3	38.7	0.3	38.2	11.2
D	14.7	0.4	88.4	3.7	19.9	66.0
E	20.6	107.0	21.4	16.6	163.0	68.8
F	7.3	4.2	49.9	25.9	117.7	68.6
G	2.9	39.8	7.5	14.9	1.9	38.4

Table 10. Reproducibility v_{norm} for different laboratories and different materials.

Laboratory Comparison	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A-B	6	9	2	58	170	89
A-C	15	46	49	82	105	19
A-D	23	8	8	131	1	62
A-E	64	65	26	99	32	174
A-F	43	11	20	1	113	81
A-G	27	6	4	55	174	76
B-C	2	40	37	8	7	42
B-D	6	9	10	23	188	15
B-E	33	38	21	13	54	15
B-F	18	6	11	65	8	1
B-G	8	20	13	6	680	280
C-D	2	18	84	4	117	9
C-E	18	7	19	0	22	83
C-F	7	88	9	97	3	51
C-G	3	87	104	4	516	102
D-E	14	32	47	3	39	46
D-F	6	29	35	154	130	24
D-G	0	28	16	16	144	184
E-F	2	77	10	119	22	11
E-G	12	104	80	6	332	400
F-G	5	9	54	69	592	290

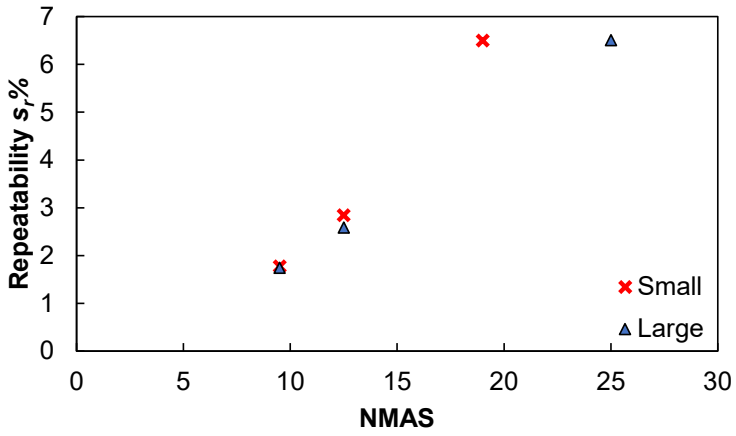
Table 11. Repeatability and reproducibility precision statements with v_{norm}.

Mix/Geometry	Repeatability Limit 95 Percent Confidence	Reproducibility Limit 95 Percent Confidence
9.5 S	22	46
9.5 L-E	38	82
12.5 S	108	86
12.5 L	88	170
19 S	314	580
25 L	401	313

Evaluation of Repeatability and Reproducibility Trends for NMAS

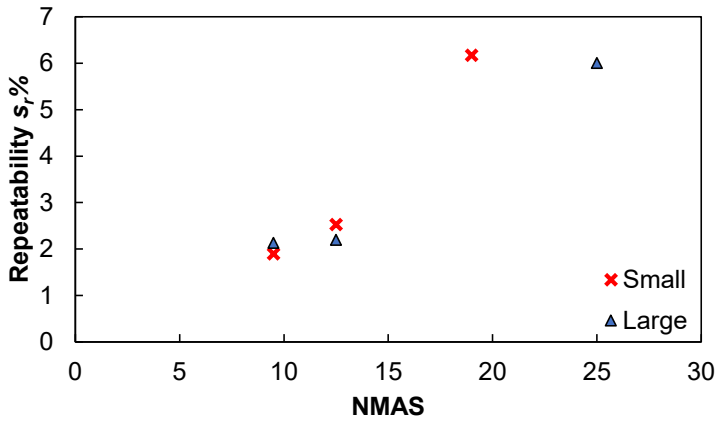
Figure 36 and figure 37 present the evaluation of repeatability and reproducibility trends for NMAS and specimen geometry for damage characteristic curve (*C* value, shape factor, v_{norm}) and D^R, respectively. For *C* value and shape factor, the repeatability statistic, s_r%, and reproducibility statistic, $s_{\bar{x}}$%, were extracted for an average $(1-\bar{\bar{C}}) = 0.7$. The analysis for 9.5-mm large mixture does not include laboratory E's results.

Figure 36 shows that within-laboratory variation increases as the NMAS increases for a given specimen geometry, irrespective of the measure in terms of *C* value, shape factor, v_{norm}, and D^R. The damage characteristic curve measures (*C* value, shape factor, v_{norm}) generally indicate similar repeatability between the large- and small-specimen geometry for the 9.5-mm and 12.5-mm mixtures, regardless of the parameter of evaluation. The 25-mm mixture large-specimen results show similar repeatability to 19-mm mixture small-specimen results in terms of *C* value and shape factor but a higher repeatability limit based on the v_{norm} approach. Although it shows the same increasing trend with NMAS for a given specimen geometry, the trend with D^R is somewhat different than the damage characteristic curve trends. The small-specimen geometry D^R results of the 9.5-mm and 12.5-mm mixtures have higher within-laboratory variation than large-specimen geometry. Furthermore, the 25-mm mixture large-specimen results indicated less within-laboratory variation in D^R compared to the 19-mm mixture small-specimen geometry results.



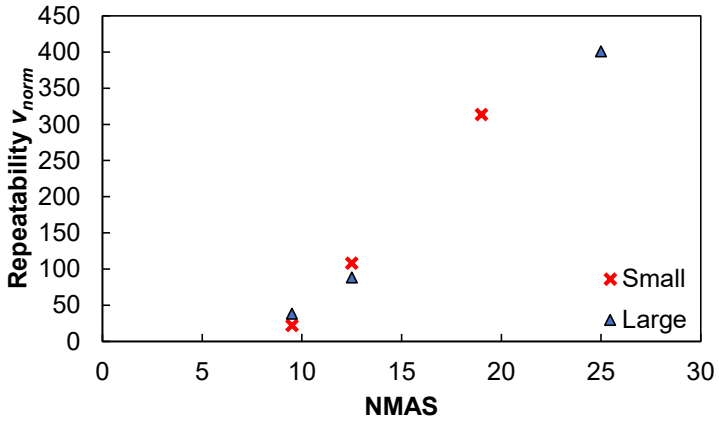
Source: FHWA.

A. Repeatability analysis using single-point *C* value.



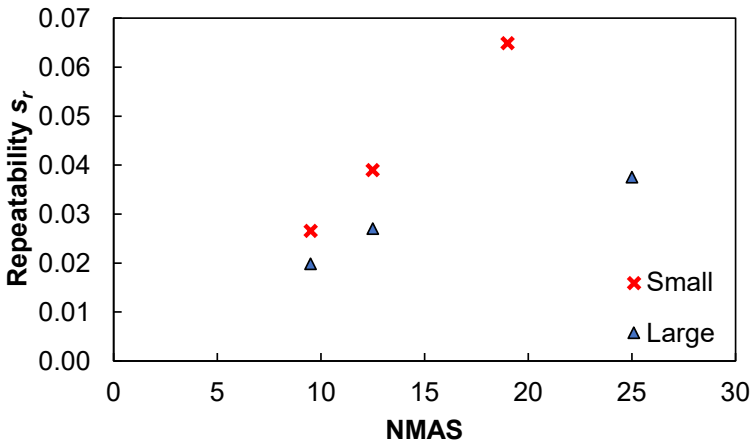
Source: FHWA.

B. Repeatability analysis using single-point shape factor.



Source: FHWA.

C. Repeatability analysis using v_{norm}.

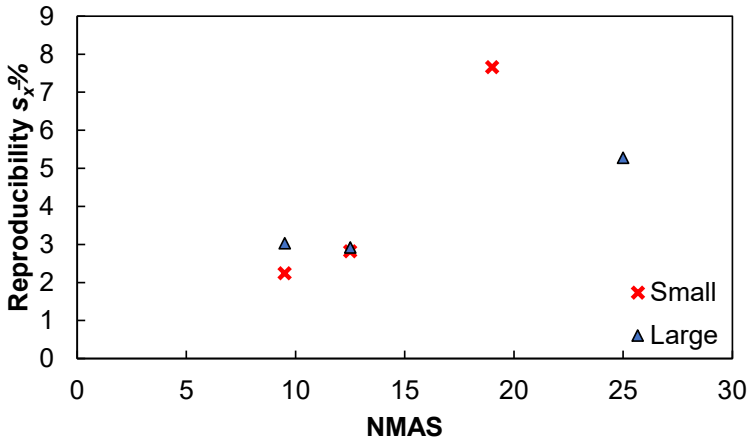


Source: FHWA.

D. Repeatability analysis using D^R.

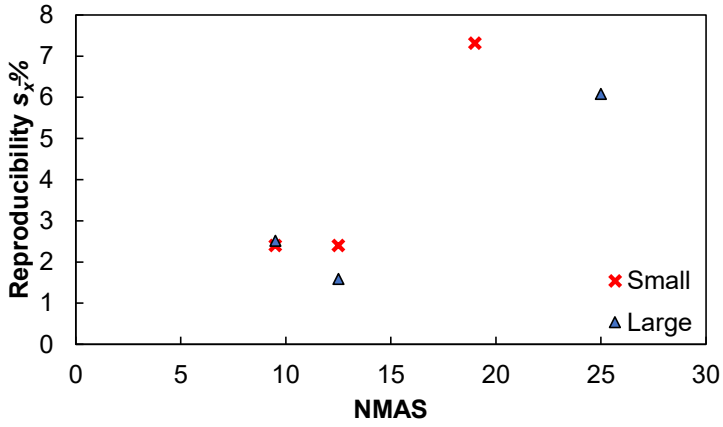
Figure 36. Graphs. Repeatability analysis comparison between single-point measure and v_{norm} approach.

Figure 37 shows similar trends in v_{norm}, with increasing NMAS for a given specimen geometry. The *C* value and shape factor measures indicate higher between-laboratory variation in the 19-mm and 25-mm mixtures compared to the 9.5-mm and 12.5-mm mixtures; however, trends among 9.5-mm and 12.5-mm mixtures are somewhat inconsistent. The 25-mm mixture has lower between-laboratory variation than the 19-mm mixture in all damage characteristic curve measures, regardless of the parameters. All damage characteristic curve measures indicate the highest between-laboratory variation in the 19-mm mixture small-specimen results. Since the 25-mm mixture was used with the large-specimen geometry and the 19-mm mixture was used with the small-specimen geometry, these results match expectations. However, D^R measures again present different trends with NMAS compared to damage characteristic curve measures. All mixtures have similar between-laboratory variances for D^R, irrespective of NMAS and geometries, although the 9.5-mm large mixture has somewhat higher reproducibility variation than any other mixtures.



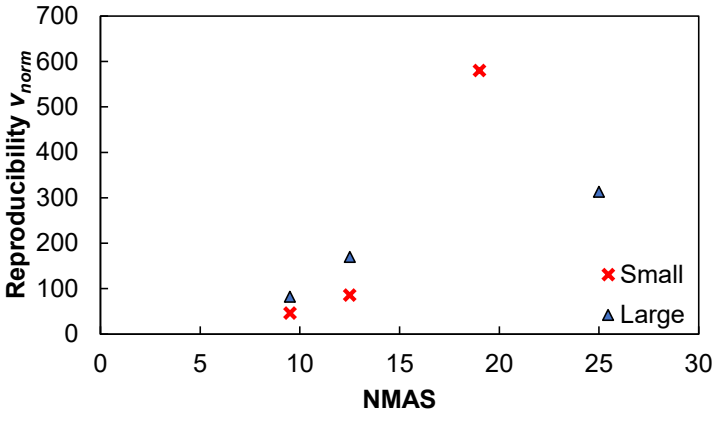
Source: FHWA.

A. Reproducibility analysis using single-point *C* value.



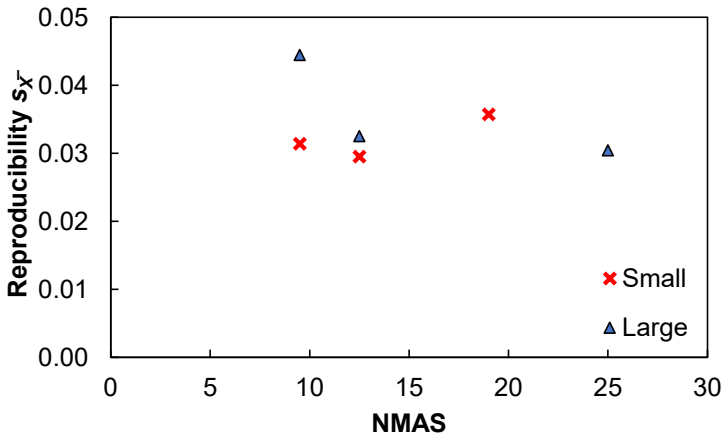
Source: FHWA.

B. Reproducibility analysis using single-point shape factor.



Source: FHWA.

C. Reproducibility analysis using v_{norm}.



Source: FHWA.

D. Reproducibility analysis using D^R.

Figure 37. Graphs. Reproducibility analysis comparison between single-point measure and v_{norm} approach.

Establishment of the Recommended Precision Statements

Damage Characteristic Curve

Given the clear NMAS dependence observed in the v_{norm} results and its advantage of considering the entire damage characteristic curve rather than a single point within the curve, v_{norm} is recommended for quantifying the repeatability and reproducibility of damage characteristic curve results. The researchers propose the v_{norm} limits given in table 11 for defining the repeatability and reproducibility of damage characteristic curve results of the AMPT cyclic fatigue tests; these limits are a function of NMAS and specimen geometry.

Failure Criterion

Given the clear NMAS dependence of D^R repeatability shown in figure 36, is the researchers recommend that the NMAS and specimen geometry-specific D^R repeatability limits provided in table 8 be incorporated into the AMPT cyclic fatigue test standard precision statement. However, the D^R reproducibility results in figure 37 do not indicate NMAS dependence. In fact, the reproducibility precision for different mixtures within a given specimen geometry are close to each other except for the 9.5-mm large-mixture specimen results, which exhibit the highest between-laboratory variation (for unknown reasons). Given the lack of a clear trend in the reproducibility of D^R with respect to NMAS, the collective ILS results of each specimen geometry were aggregated together and used to define specimen geometry-specific limits for the difference in D^R test results between two laboratories that are not expected to be exceeded with 95 percent confidence; this approach follows the underlying premise behind the precision statement development guidance given in ASTM E691-20 and C670-15.[4,5] Correspondingly, the researchers took the following steps to define the precision statement for D^R reproducibility:

1. Calculate the average D^R value of three specimens for each lab, mixture, and geometry combination.

2. Compute the difference between the average D^R of each pair of laboratories for each mixture with different NMAS and geometry. Each pair of laboratories yielded two results (i.e., laboratory A minus laboratory B and laboratory B minus laboratory A for 9.5-mm small mix).

3. Divide the data computed in step 2 into two groups according to the mixture geometry (small versus large).

4. Test whether the data are normally distributed using the Shapiro-Wilk test for each group.[24]

5. Fit a normal distribution to the data in step 3 in each group with mean and standard deviation using the maximum-likelihood fitting method.[25]

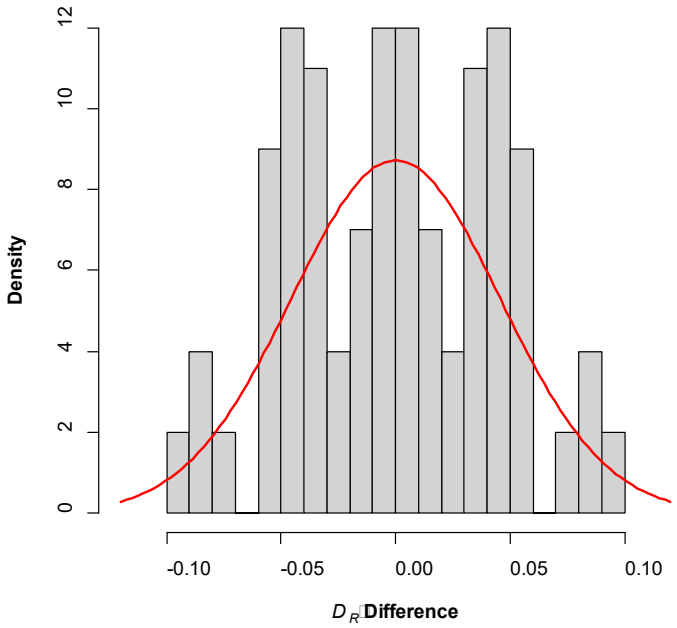
6. Calculate the 95-percent confidence interval to find the maximum allowable difference in D^R test results between two laboratories.

A summary of the D^R reproducibility analysis results, aggregating all mixture results for a given geometry, is presented in table 12. The number of data points for the large-specimen geometry mixture is smaller than for small-specimen geometry because laboratory E data were excluded from the 9.5-mm large-specimen data, as previously discussed. Both small- and large-specimen geometry results yielded *p*-values higher than 0.05 in the Shapiro-Wilk test, which indicates the hypothesis that the results are normally distributed was not rejected.[24]

Table 12. D^R reproducibility analysis results aggregating NMAS for different geometry.

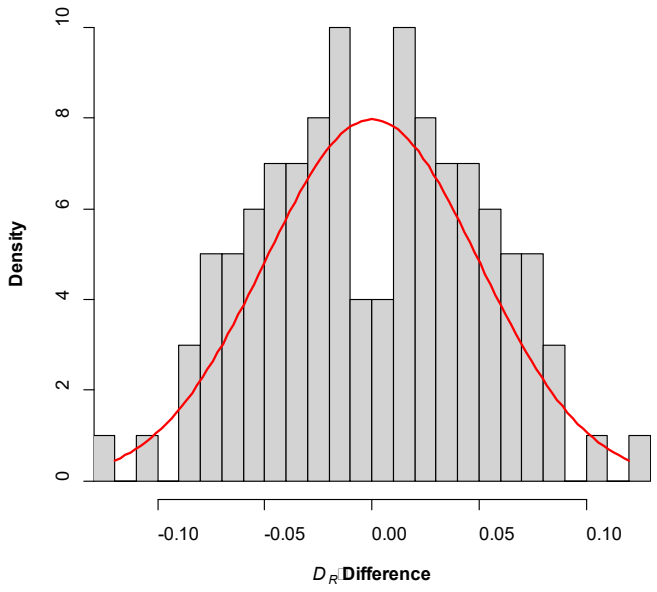
Geometry	Number of Data Points	Shapiro-Wilk Test *p*-value[24]	Normal Distribution	Maximum Allowable Difference
Small	126	0.075	Yes (0, 0.046^2)	0.090
Large	114	0.38	Yes (0, 0.050^2)	0.098

Figure 38 presents the corresponding histogram of D^R differences between laboratory pairs, with the corresponding best fit normal distribution. The results show that the maximum allowable difference between two laboratory results for D^R is 0.090 for mixtures with the small geometry and 0.098 for mixtures with the large geometry using a confidence level of 95 percent; it is recommended that these limits be incorporated into the reproducibility precision statement in the AMPT cyclic fatigue test standards. Table 13 presents, the recommended precision statements for the repeatability and reproducibility of D^R.



Source: FHWA.

A. Histograms of D^R difference with the fitted normal distribution for mixtures with small geometry.



Source: FHWA.

B. Histograms of D^R difference with the fitted normal distribution for mixtures with large geometry.

Figure 38. Graphs. Histograms of D^R difference with the fitted normal distribution.

Table 13. Recommended repeatability and reproducibility precision statements (maximum allowable difference) of D^R.

Mixture	Repeatability (*MAR*)	Reproducibility ($d2s_{rpd}$)
9.5 S	0.088	0.090
9.5 L-E	0.065	0.098
12.5 S	0.129	0.090
12.5 L	0.089	0.098
19 S	0.214	0.090
25 L	0.124	0.098

PROPOSED PRECISION STATEMENTS

The following provides a summary of the proposed precision statements for the AMPT cyclic fatigue test standards:

- Precision—Criteria for judging the acceptability of cyclic fatigue test results, including damage characteristic curve and D^R obtained by this method, are given in table 14 and table 15. The v_{norm} is calculated by an analytical integration of the sum of the squared difference between the fitted curve for each specimen and the fitted curve across all the specimens divided by the number of specimens. The v_{norm} precision limits given in table 14 and table 15 are based on the test results from seven laboratories.

- Single operator precision (repeatability)—The figures in columns 3 and 4 of table 14 are the acceptable limits that are appropriate for the mixtures described in column 2. Results obtained in the same laboratory by the same operator using the same equipment in the shortest practical period should be accepted unless the v_{norm} or D^R range exceeds that given in table 14.

- Multilaboratory precision (reproducibility)—The figures in columns 3 and 4 of table 15 are the acceptable limits that are appropriate for the mixtures described in column 2. Results obtained by two different operators testing the same material in two different laboratories should be accepted unless the v_{norm} or D^R difference exceeds that given in table 15.

The precision limits given in table 14 and table 15 are based on six mixtures with different NMAS and geometries tested in seven laboratories.

Table 14. Single-operator precision for damage characteristic curve and D^R results.

Applicable Standard	NMAS (mm)	v_{norm} Limit	Acceptable Limit for D^R (*MAR*)
AAHTO TP 107[1]	9.5	38	0.065
AAHTO TP 107	12.5	88	0.089
AAHTO TP 107	19.0	401	0.124
AAHTO TP 107	25.0	401	0.124
AAHTO TP 133[2]	9.5	22	0.088
AAHTO TP 133	12.5	108	0.129
AAHTO TP 133	19.0	314	0.214

Table 15. Multilaboratory precision for damage characteristic curve and D^R results.

Applicable Standard	NMAS (mm)	v_{norm} Limit	Acceptable Limit for D^R ($d2s_{rpd}$)
AAHTO TP 107[1]	9.5	82	0.098
AAHTO TP 107	12.5	170	0.098
AAHTO TP 107	19.0	313	0.098
AAHTO TP 107	25.0	313	0.098
AAHTO TP 133[2]	9.5	46	0.090
AAHTO TP 133	12.5	86	0.090
AAHTO TP 133	19.0	580	0.090

ANALYSIS OF THE PRECISION STATEMENT IMPLICATIONS ON UNCERTAINTY IN PRACTICAL APPLICATIONS OF THE TEST RESULTS

Data Excluded From the Analysis

Table 16 and table 17 present the individual laboratory results and laboratory pair results that did not meet the proposed precision limits for repeatability and reproducibility, respectively. Given the proposed precision limits, these results were excluded from the analysis of the uncertainty in FlexPAVE *%Damage* and S_{app}. Of the 42 total results, 6 laboratories (14.3 percent) did not meet the repeatability limits, and 44 out of 126 laboratory pairs (34.9 percent) did not meet the reproducibility limits. Recall that the v_{norm} and D^R precision limits were both established based on a 95-percent confidence interval. Because each was established using a 95-percent confidence level, it is expected that roughly 5 percent of laboratories would exceed each of the limits. A test will not meet the repeatability limit if either of the v_{norm} or D^R exceeds the repeatability precision limits. Therefore, the probability of a test not meeting repeatability limits is approximately 10 percent (1–95 percent×95 percent), higher than 5 percent. Within the six laboratories that did not meet the repeatability limits, two (4.8 percent) were excluded due to failure to meet the D^R precision limit, and four (9.5 percent) were excluded due to failure to meet v_{norm} limit. The number of laboratory pairs that do not meet reproducibility results reflects a much higher percentage than 10 percent because data were discarded for consideration in the reproducibility analysis if repeatability (within a laboratory) or reproducibility (between-laboratory) criteria were exceeded.

Table 16. Data failing repeatability limits.

Laboratory	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A	—	—	—	X	—	X
B	X	—	—	—	—	—
C	—	—	—	—	—	—
D	—	—	—	—	—	—
E	—	X	—	—	—	—
F	—	—	X	—	—	—
G	—	X	—	—	—	—

—No data.
X = excluded due to not meeting the precision limits.

Table 17. Data excluded from reproducibility analysis.

Laboratory Pair	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A-B	X	—	—	X	—	X
A-C	—	—	—	X	—	X
A-D	—	—	—	X	—	X
A-E	X	X	—	X	—	X
A-F	—	—	X	X	—	X
A-G	—	X	—	X	—	X
B-C	X	X	—	—	—	—
B-D	X	—	—	—	—	—
B-E	X	X	—	—	—	—
B-F	X	—	X	—	—	—
B-G	X	X	—	—	X	—
C-D	—	—	—	—	—	—
C-E	—	X	—	—	—	—
C-F	—	X	X	—	X	—
C-G	—	X	X	—	—	—
D-E	X	X	—	—	—	—
D-F	—	—	X	—	—	—
D-G	—	X	—	—	—	—
E-F	—	X	X	—	—	—
E-G	—	X	—	—	—	X
F-G	—	X	X	—	X	—

—No data.

Uncertainty in FlexPAVE *%Damage* Predictions

The uncertainty in *%Damage*, conveyed by the standard deviation calculated using Ghanbari's framework for each mixture and specimen geometry combination in terms of repeatability and reproducibility, is shown in table 18 and table 19, respectively.[19] The X's correspond to those excluded due to not meeting the precision limits. Table 20 and table 21 summarize the maximum *%Damage* standard deviation result for each mixture and specimen geometry combination of all of the ILS results that met the proposed precision limits. Table 22 and table 23 summarize the median *%Damage* standard deviation for each mixture and specimen geometry combination. These maximum and median standard deviations were used to calculate the 95-percent confidence intervals for *%Damage*. The maximum repeatability uncertainty of *%Damage* increases with NMAS for a given geometry, which matches expectations given that the repeatability precision limits for v_{norm} and D^R both increase as a function of NMAS. The 19-mm small-specimen results show that this NMAS and geometry combination yielded the highest uncertainty in repeatability *%Damage* of all conditions evaluated. The small-specimen 19-mm mixtures results have the second highest repeatability v_{norm} limit of all conditions evaluated, including both small- and large-specimen geometries, while the large-specimen 25-mm mixture results have the highest. However, the 19-mm small-specimen results have a higher repeatability D^R limit than the 25-mm large-specimen results; this result suggests that the influence of D^R uncertainty had a more pronounced effect on the uncertainty of *%Damage* than the damage characteristic curve uncertainty.

Table 18. Repeatability standard deviation of *%Damage* for all the laboratories and mixtures.

Laboratory	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A	0.37	0.43	0.16	X	1.05	X
B	X	0.26	0.53	0.77	2.48	1.26
C	0.40	0.36	0.70	0.25	0.88	0.42
D	0.51	0.53	1.50	0.49	0.57	1.30
E	0.90	X	0.30	0.87	1.73	0.80
F	0.33	0.50	X	0.69	2.21	1.47
G	0.22	X	0.56	0.61	0.27	0.76
Median	0.39	0.43	0.54	0.65	1.05	1.03
Maximum	0.90	0.53	1.50	0.87	2.48	1.47

Table 19. Reproducibility standard deviation of *%Damage* for all the laboratories and mixtures.

Laboratory Pair	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A-B	X	0.78	0.34	X	1.10	X
A-C	0.25	0.79	0.44	X	1.14	X
A-D	0.69	0.42	0.54	X	0.11	X
A-E	X	X	0.75	X	0.70	X
A-F	0.44	0.95	X	X	1.05	X
A-G	0.63	X	0.26	X	0.89	X
B-C	X	X	0.64	0.60	0.99	0.95
B-D	X	0.54	0.29	0.96	1.19	0.89
B-E	X	X	0.98	0.78	1.11	0.51
B-F	X	0.29	X	0.60	0.27	0.49
B-G	X	X	0.62	0.52	X	1.47
C-D	0.47	0.88	0.93	0.36	1.12	0.27
C-E	0.76	X	0.70	0.16	0.44	0.84
C-F	0.19	X	X	1.15	X	0.55
C-G	0.41	X	X	0.18	1.82	0.72
D-E	X	X	1.22	0.22	0.69	0.79
D-F	0.51	0.82	X	1.54	1.15	0.50
D-G	0.10	X	0.74	0.55	0.77	1.03
E-F	0.62	X	X	1.34	1.04	0.37
E-G	1.02	X	0.84	0.33	1.38	X
F-G	0.41	X	X	0.99	X	1.05
Median	0.47	0.79	0.67	0.6	1.05	0.75
Maximum	1.02	0.95	1.22	1.54	1.82	1.47

Table 20. Maximum repeatability variation for each mixture.

Mixture	std_{Total}	Average *%Damage*±1.96 std
9.5 S	0.90	15±1.76
9.5 L	0.53	15±1.05
12.5 S	1.50	15±2.94
12.5 L	0.87	15±1.70
19 S	2.48	15±4.86
25 L	1.47	15±2.88

std = standard deviation.

Table 21. Maximum reproducibility variation for each mixture.

Mixture	std_{Total}	Average *%Damage*±1.96 std
9.5 S	1.02	15±2.00
9.5 L	0.95	15±1.85
12.5 S	1.22	15±2.39
12.5 L	1.54	15±3.03
19 S	1.82	15±3.56
25 L	1.47	15±2.89

Table 22. Median repeatability variation for each mixture.

Mixture	std_{Total}	Average *%Damage*±1.96 std
9.5 S	0.39	15±0.76
9.5 L	0.43	15±0.83
12.5 S	0.54	15±1.07
12.5 L	0.65	15±1.28
19 S	1.05	15±2.05
25 L	1.03	15±2.01

Table 23. Median reproducibility variation for each mixture.

Mixture	std_{Total}	Average *%Damage*±1.96 std
9.5 S	0.47	15±0.91
9.5 L	0.79	15±1.54
12.5 S	0.67	15±1.31
12.5 L	0.60	15±1.18
19 S	1.05	15±2.05
25 L	0.75	15±1.48

An increase in the maximum *%Damage* uncertainty with increasing NMAS was expected under reproducibility conditions, given that the reproducibility limits for v_{norm} increase with NMAS and the reproducibility limits for D^R are independent of the NMAS. The maximum uncertainty in *%Damage* under reproducibility conditions increases with NMAS for the small-specimen geometry but not for the large-specimen geometry; the reason for this trend is unknown. The 19-mm small-specimen geometry results encompass the maximum uncertainty in *%Damage* under reproducibility conditions for all mixtures and geometries evaluated. Because repeatability assesses variation among three individual test specimens and reproducibility assesses variation among test results (i.e., results of three test specimens each in two laboratories), they are not directly comparable. Thus, trends among repeatability and reproducibility *%Damage* uncertainty for a given mixture are varied.

Comparison between median and maximum confidence intervals in percent damage predictions provides context to the typical versus extreme uncertainty ranges expected in practice. For

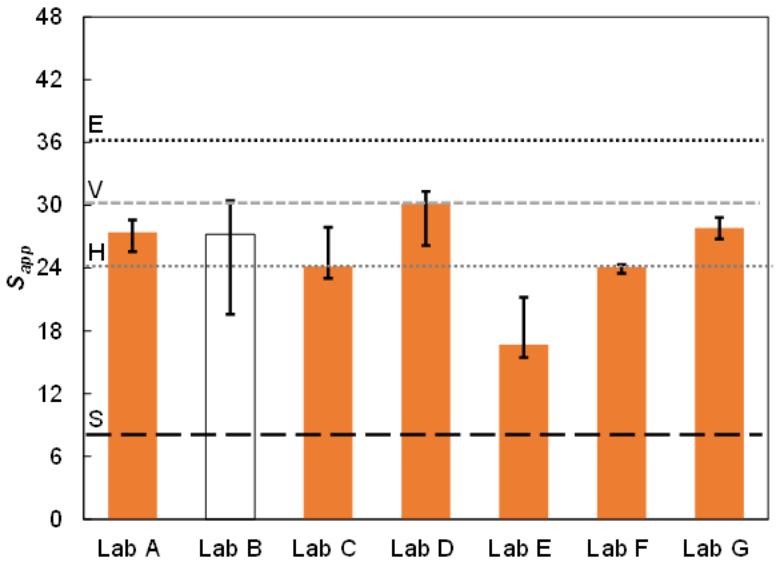
example, even though the 95-percent confidence interval of predicted *%Damage* at an average *%Damage* of 15 percent under repeatability conditions for the 19-mm mixture using the small-specimen geometry is from 10.14 percent to 19.86 percent (approximately ±30 percent error), the median 95-percent confidence interval is 12.95 percent to 17.05 percent (corresponding to roughly ±14 percent error). For the 9.5-mm mixture with the small-specimen geometry, the median condition yields a 95-percent confidence interval for *%Damage* of 14.24–15.76 percent (corresponding to roughly ±5 percent error).

Similar trends are observed for reproducibility. The maximum 95-percent confidence interval of all conditions under reproducibility conditions coincides with the 19-mm mixture with the small-specimen geometry, yielding a *%Damage* span from 11.44–18.56 percent (roughly ±23 percent error), but the median confidence interval for the same condition is notably lower at 12.95-17.05 percent (roughly ±14 percent error). The practical implications of these errors merit further investigation in future work. Efforts to improve the consistency of specimen fabrication procedures may also enable a reduction in the uncertainty of test outcomes, particularly for the larger NMAS mixtures

Uncertainty in Apparent Damage Capacity

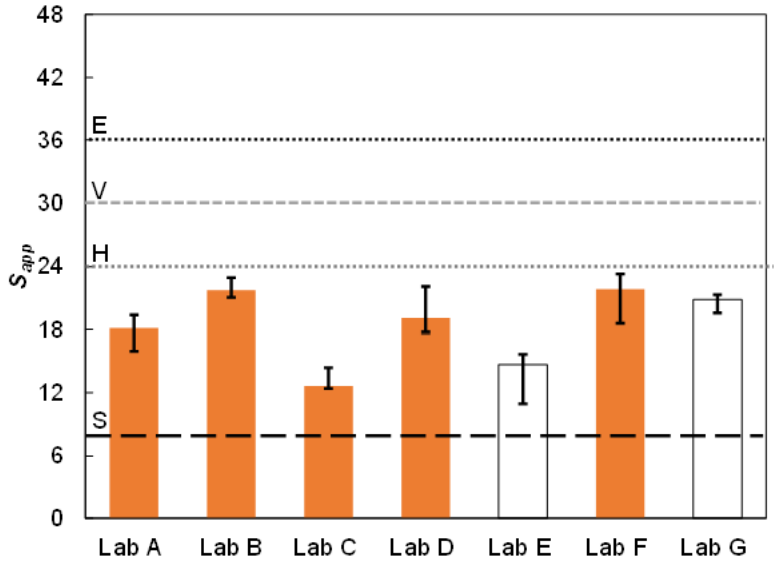
Figure 39 through figure 44 present the measured S_{app} values for the different mixture and specimen geometry combinations. The height of the bars indicates the representative S_{app} values determined from the damage characteristic curve and failure criterion results from a given laboratory. The error bars convey the maximum and minimum values determined from the individual test determinations. Solid bars convey that the representative laboratory's results met the proposed repeatability limits for v_{norm} and D^R, whereas the hollow bars indicate that the laboratory's results failed to meet the proposed repeatability limits for v_{norm} and/or D^R. The horizontal lines convey the proposed S_{app} thresholds for standard (S), heavy (H), very heavy (V), and extreme (E) traffic levels.[23] Note that consistent dynamic modulus results were used for all laboratories when calculating the S_{app} value for a given mixture and specimen geometry combination, consistent with the statistical analysis to develop the precision limits.

Figure 39 through figure 42 demonstrate that the S_{app} values obtained from the large-specimen geometry were generally lower than those for the small-specimen geometry for both the 9.5-mm and 12.5-mm mixtures. Differences in the dynamic moduli and α values among the two specimen geometries could have contributed to these trends. In addition, it is apparent that different test determinations within a given laboratory and the representative S_{app} results from different laboratories can yield different traffic designations according to the proposed S_{app} thresholds. These findings suggest that the limits may merit reconsideration given the measured uncertainty in the test results. While figure 43 and figure 44 indicate all laboratories' representative S_{app} results for the 19-mm and 25-mm mixtures have the same standard traffic level designation according to the proposed S_{app} thresholds, the results do show considerable variation within and across laboratories with respect to the mean values.



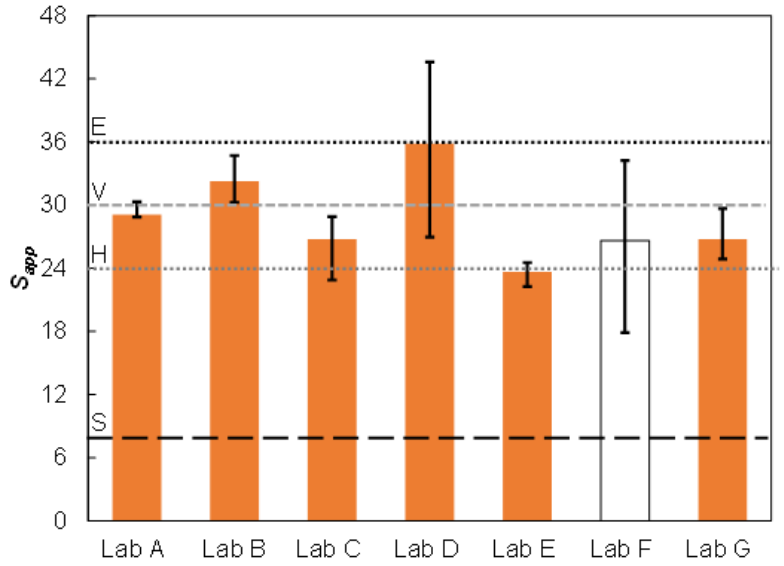
Source: FHWA.
The hollow bar indicates test results that failed to meet the proposed repeatability precision limits.

Figure 39. Graph. Measured S_{app} values for the 9.5-mm mixture small-specimen geometry.



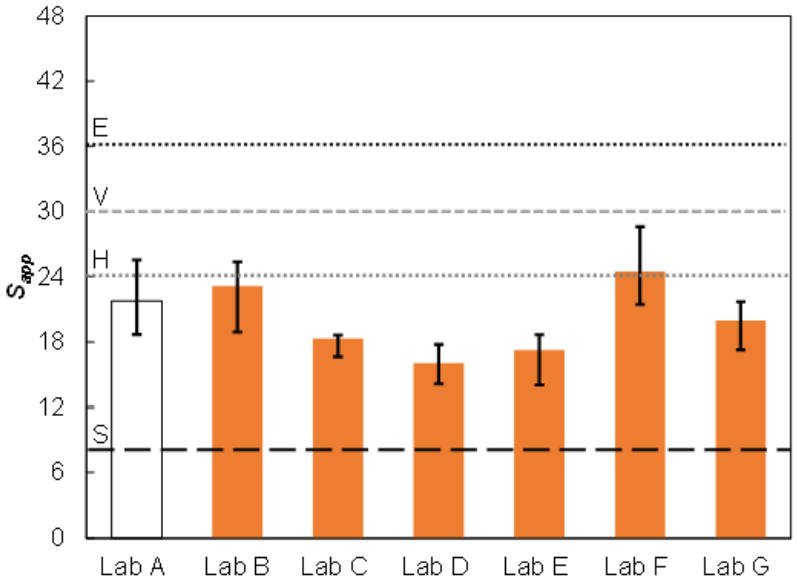
Source: FHWA.
The hollow bars indicate test results that failed to meet the proposed repeatability precision limits.

Figure 40. Graph. Measured S_{app} values for the 9.5-mm mixture large-specimen geometry.



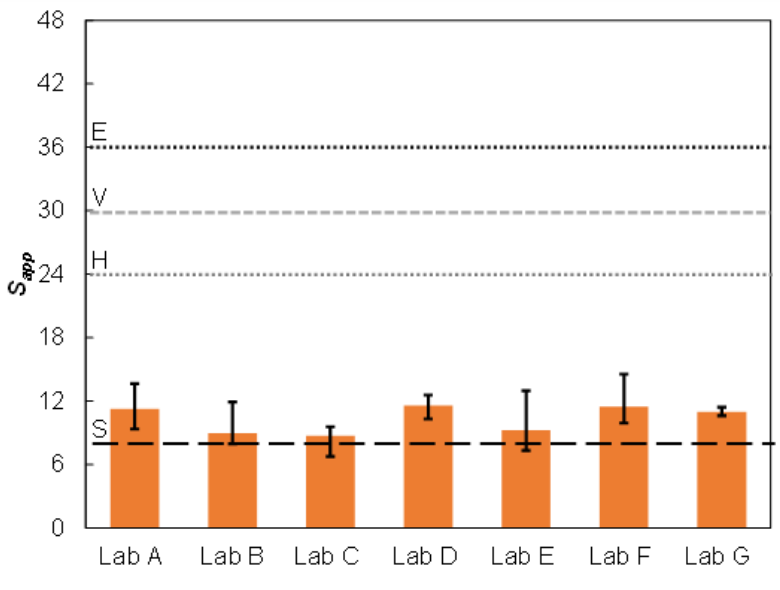
Source: FHWA.
The hollow bar indicates test results that failed to meet the proposed repeatability precision limits.

Figure 41. Graph. Measured S_{app} values for the 12.5-mm mixture small-specimen geometry.



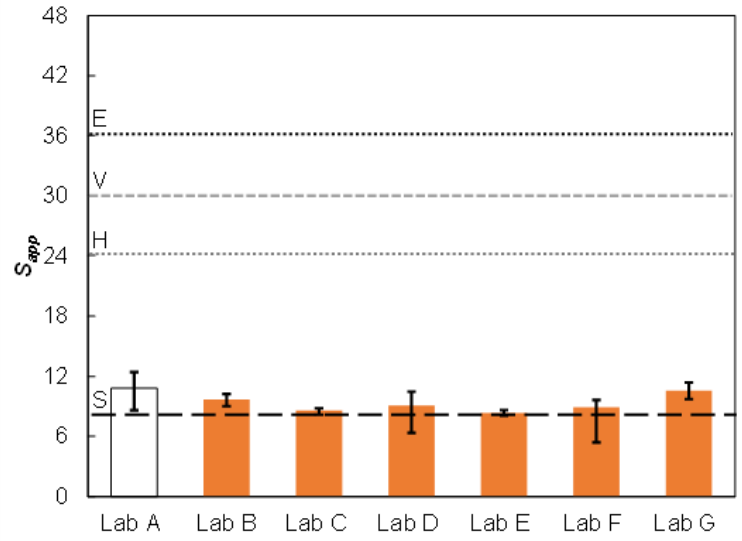
Source: FHWA.
The hollow bar indicates test results that failed to meet the proposed repeatability precision limits.

Figure 42. Graph. Measured S_{app} values for the 12.5-mm mixture large-specimen geometry.



Source: FHWA.

Figure 43. Graph. Measured S_{app} values for the 19-mm mixture small-specimen geometry.



Source: FHWA.
The hollow bar indicates test results that failed to meet the proposed repeatability precision limits.

Figure 44. Graph. Measured S_{app} values for the 25-mm mixture large-specimen geometry.

Table 24 and table 25 show the uncertainty in S_{app} under repeatability conditions measured among test determinations and predicted among test results from the analysis in terms of the COV% for each mixture and specimen geometry combination. The X's represent the laboratory's results that were excluded due to not meeting the repeatability precision limits. As described in chapter 2, the repeatability COV% values for test results were calculated using simulated S_{app} results that fell within a 95-percent confidence interval. The S_{app} repeatability

COV% values among test determinations were calculated directly from the ILS test results. The simulated S_{app} values were generated based on the observed within-laboratory variability in both damage characteristic curve and failure criterion results. Figure 45 shows the maximum S_{app} COV% values for each mixture and specimen geometry combination reported in table 24 and table 25.

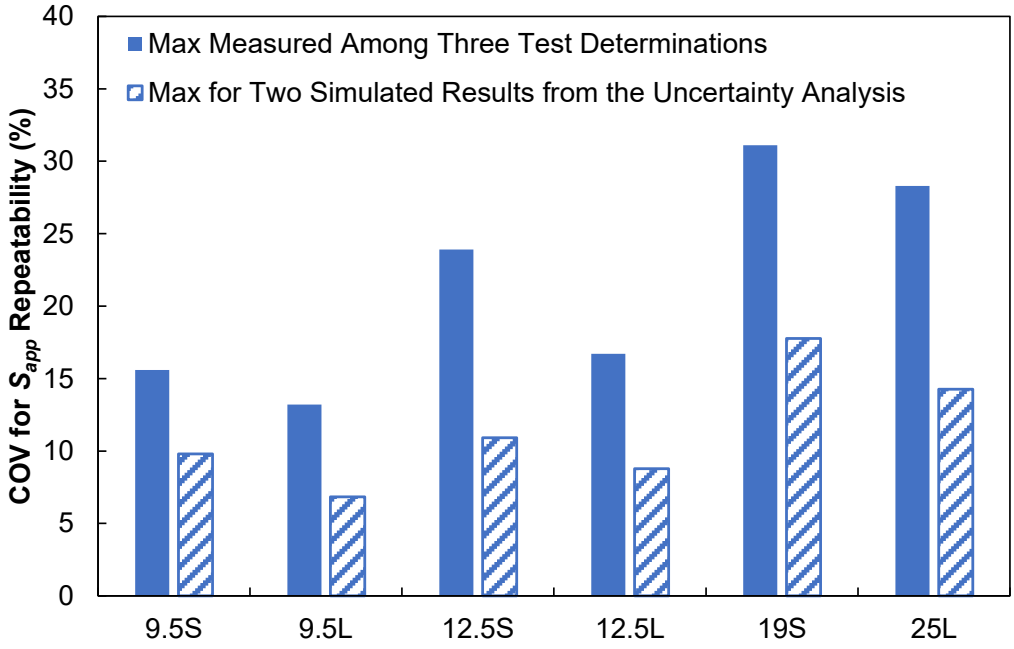
The results demonstrate that the uncertainty in S_{app} under repeatability conditions generally increases as the NMAS increases for a given specimen geometry, which matches expectations since the precision limits increase with respect to NMAS. Furthermore, the results demonstrate that the variation in S_{app} values among three test determinations is considerably higher than that among two test results (i.e., two sets of three determinations). The maximum S_{app} COV% values for test results span from approximately 7 percent to 18 percent. The maximum COV% value for the repeatability of S_{app} results of 9.5-mm and 12.5-mm mixtures falls below 11 percent. The repeatability COV% values are somewhat higher for the 19-mm and 25-mm mixtures, with a maximum median value of 11 percent and a maximum value of 17.8 percent.

Table 24. Measured repeatability COV% of S_{app} among test determinations for all the laboratories and mixtures.

Laboratory	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A	6.38	10.21	2.64	16.75	19.14	X
B	X	4.63	7.95	5.93	20.80	6.24
C	10.38	7.72	11.60	11.35	17.83	3.85
D	8.80	13.12	23.93	15.26	10.07	26.55
E	15.64	X	5.01	14.87	31.14	3.31
F	1.71	11.46	X	12.51	23.71	28.27
G	3.62	X	8.36	16.75	3.06	7.81

Table 25. Repeatability COV% of S_{app} results for all the laboratories and mixtures from the uncertainty analysis.

Laboratory	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A	3.24	3.32	1.50	X	10.90	X
B	X	2.75	2.92	7.14	17.77	8.40
C	4.19	3.24	4.34	3.15	6.83	3.22
D	3.87	6.84	10.91	5.26	3.83	12.04
E	9.79	X	1.70	8.78	12.21	4.78
F	2.51	5.33	X	5.76	16.90	14.26
G	1.79	X	5.54	5.11	2.41	5.31



Source: FHWA.
Max = maximum.

Figure 45. Graph. Comparison of the maximum S_{app} COV% values measured among test determinations and calculated among test results from the uncertainty analysis.

Table 26 and table 27 present the measured S_{app} COV% values and the S_{app} COV% values from the uncertainty analysis, respectively. The X's represent the laboratories' or laboratory pairs' results that were excluded due to not meeting the precision limits. The reproducibility COV% values from the uncertainty analysis reflect the maximum possible value from a pair of laboratories' results based on the 95-percent confidence intervals for the simulated S_{app} values in each laboratory and, thus, constitute an extreme limit possible given the precision limits.

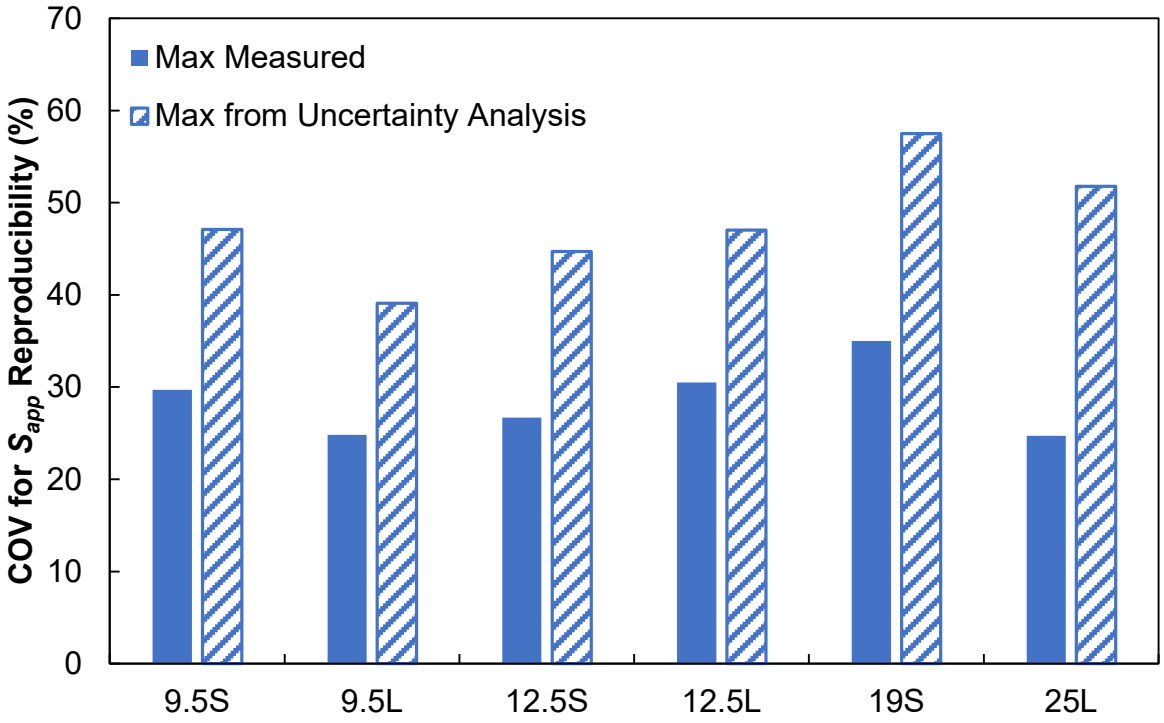
Table 26. Measured reproducibility COV% of S_{app} for all the laboratories and mixtures.

Laboratory Pair	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A-B	X	14.28	5.71	X	5.58	X
A-C	4.66	20.15	8.42	X	24.32	X
A-D	5.82	4.76	11.55	X	0.18	X
A-E	X	X	15.39	X	14.52	X
A-F	7.37	19.84	X	X	0.30	X
A-G	3.02	X	2.26	X	11.20	X
B-C	X	X	14.10	12.04	18.87	8.36
B-D	X	9.56	5.86	20.51	5.40	13.43
B-E	X	X	21.01	15.50	8.98	9.78
B-F	X	5.64	X	10.33	5.28	14.86
B-G	X	X	7.96	8.08	X	7.03
C-D	10.46	24.79	19.88	8.57	24.15	5.09
C-E	22.24	X	7.02	3.49	9.97	1.42
C-F	2.72	X	X	22.24	X	6.53
C-G	7.67	X	X	3.99	35.04	15.35
D-E	X	X	26.71	5.09	14.35	3.67
D-F	13.16	15.15	X	30.52	0.13	1.44
D-G	2.81	X	13.79	12.53	11.37	20.36
E-F	19.58	X	X	25.63	14.23	5.12
E-G	29.66	X	13.16	7.47	25.51	X
F-G	10.37	X	X	18.33	X	21.77

Table 27. Reproducibility COV% of S_{app} for all the laboratories and mixtures from the uncertainty analysis.

Laboratory Pair	9.5 S	9.5 L	12.5 S	12.5 L	19 S	25 L
A-B	X	23.70	13.21	X	50.34	X
A-C	16.47	29.61	17.97	X	49.58	X
A-D	17.09	20.32	30.35	X	24.20	X
A-E	X	X	20.22	X	49.39	X
A-F	16.54	33.16	X	X	44.97	X
A-G	11.22	X	13.97	X	33.08	X
B-C	X	X	26.11	27.45	53.07	24.78
B-D	X	24.64	26.67	38.73	41.12	45.33
B-E	X	X	28.31	39.76	52.88	28.70
B-F	X	18.67	X	30.48	57.48	48.66
B-G	X	X	22.17	27.18	X	29.48
C-D	23.23	39.08	42.63	21.44	40.32	31.67
C-E	43.67	X	16.26	22.52	37.30	14.28
C-F	13.74	X	X	35.07	X	35.19
C-G	17.43	X	X	17.27	48.67	28.21
D-E	X	X	44.70	25.73	40.13	32.02
D-F	23.30	34.07	X	46.02	35.51	41.20
D-G	11.66	X	38.92	28.88	21.29	48.50
E-F	38.48	X	X	47.02	56.57	35.54
E-G	47.09	X	24.06	29.95	48.48	X
F-G	17.50	X	X	34.80	X	51.76

Figure 46 shows the maximum S_{app} COV% values for each mixture and specimen geometry combination reported in table 26 and table 27. Trends in the reproducibility COV% values with respect to NMAS are less clear than for repeatability. The lack of clear trends in D^R variability among laboratories for NMAS may explain this result. The measured maximum reproducibility S_{app} COV% values are considerably lower than those calculated from the uncertainty, highlighting those calculated through the uncertainty analysis constitute an extreme scenario that is possible, but unlikely, given the proposed precision limits. The maximum measured S_{app} reproducibility COV% values spanned from roughly 25 to 35 percent for the different mixture and specimen geometry combinations, whereas the maximum possible values from the uncertainty analysis spanned from roughly 39 to 58 percent.



Source: FHWA.

Figure 46. Graph. Comparison of the maximum S_{app} COV% values measured among test determinations and calculated among test results from the uncertainty analysis.

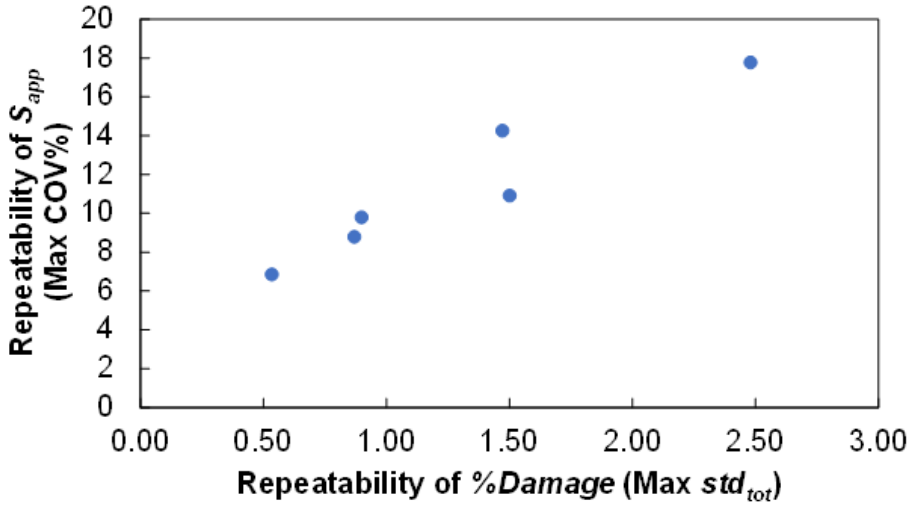
The reproducibility COV% values are considerably larger than the repeatability COV% values. The researchers attributed this result to differences in how the repeatability and reproducibility COV% values were calculated. Increased uncertainty in test results between laboratories compared to within a given laboratory also contributed to this result. The implications of the relatively high S_{app} COV% values observed under reproducibility conditions merit further investigation in future research.

Table 28 summarizes the maximum and median COV% values from the uncertainty analysis regarding repeatability and reproducibility for each mixture and specimen geometry combination.

Table 28. Maximum and median COV% of S_{app} for each mixture from the uncertainty analysis.

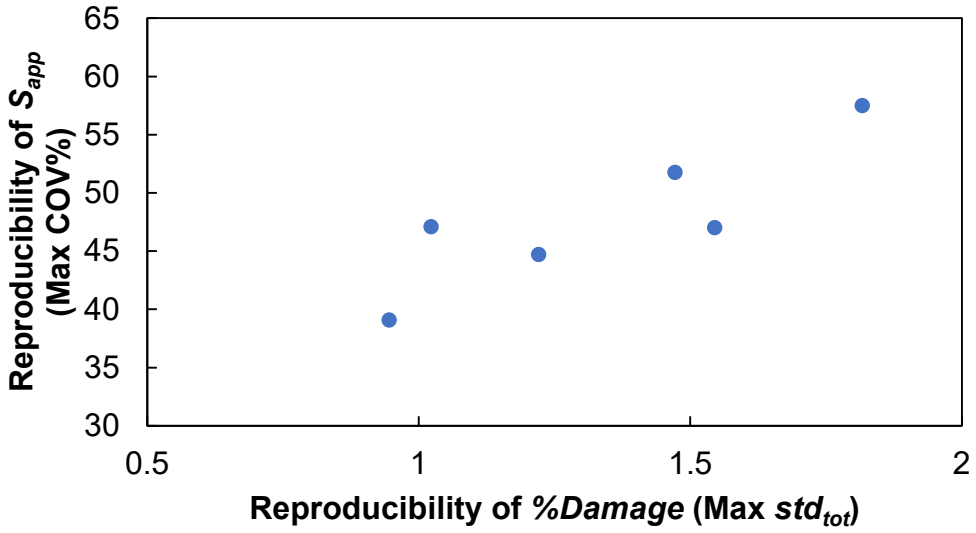
Mixture	Maximum Repeatability COV%	Maximum Reproducibility COV%	Median Repeatability COV%	Median Reproducibility COV%
9.5 S	9.79	47.09	3.55	17.43
9.5 L	6.84	39.08	3.32	27.13
12.5 S	10.91	44.70	3.63	25.08
12.5 L	8.78	47.02	5.51	29.95
19 S	17.77	57.48	10.90	46.72
25 L	14.26	51.76	6.86	33.61

Comparisons between the uncertainty measures for *%Damage* and S_{app} under repeatability conditions and reproducibility conditions are shown in figure 47. The comparisons demonstrate that the measures of uncertainty in *%Damage* and S_{app} are correlated. The uncertainty in *%Damage* and S_{app}, given the uncertainty in AMPT cyclic fatigue test results, merits consideration when practical inferences from these parameters are made.



Source: FHWA.
std_{tot} = standard deviation of *%Damage*.

A. Comparison between uncertainty of *%Damage* and uncertainty of S_{app} in terms of repeatability.



Source: FHWA

B. Comparison between uncertainty of *%Damage* and uncertainty of S_{app} in terms of reproducibility.

Figure 47. Graphs. Comparison between uncertainty of *%Damage* and uncertainty of S_{app}.

CHAPTER 4. CONCLUSIONS AND RECOMMENDED CHANGES TO THE STANDARDS FOR AMPT CYCLIC FATIGUE TESTING

The following summarizes the primary findings and recommended changes to the standards for AMPT cyclic fatigue testing developed as a result of this ILS. Herein, repeatability quantifies the acceptable variation among three test determinations (specimens) obtained within a laboratory on a single material. The reproducibility quantifies the acceptable variation between the average test results of two laboratories conducted on the same material. The research team's findings and recommendations are as follows:

- Statistical tests of the consistency of the ILS results acquired from the seven participating laboratories indicate that the results obtained are highly consistent according to ASTM E691-20.[4] Only a single laboratory's results of a single mixture tested using the large-specimen geometry failed the consistency tests and were omitted from the statistical analysis used to develop the proposed precision limits for the AMPT cyclic fatigue test standards.

- Statistical analysis of the damage characteristic curve repeatability and reproducibility conducted using single-point measures of the damage characteristic curve to directly align with the guidance in ASTM E691-20 and ASTM C670-15 demonstrates that the inferred uncertainty in test results and trends among different mixtures and specimen geometries can vary with the chosen point of reference (e.g., S value at which C is evaluated).[4,5]

- Modified forms of the variance index originally proposed by Ding et al., termed v_{norm} and developed herein to quantify repeatability and reproducibility of damage characteristic curves, better capture variation in the curve than the single-point measures.[18] The v_{norm} results demonstrated clear increases in within- and between-laboratory variability in damage characteristic curve with increasing NMAS values for both small- and large-specimen geometries.

- The v_{norm} results followed a gamma rather than a normal distribution, which precluded the direct use in precision-limit development according to ASTM E691-20 and ASTM C670-15.[4,5] However, fitted gamma distributions of the test results were used to define v_{norm} limits for repeatability and reproducibility conditions that will be exceeded with a probability of 5 percent, which aligns with the guiding principles for precision-limit development, according to the ASTM standards. The established repeatability and reproducibility limits are a function of NMAS and specimen geometry.

- The D^R constitutes a single-point measure and, therefore, is directly amenable to precision-statement development, according to ASTM E691-20 and ASTM C670-15.[4,5] The D^R results demonstrate that within-laboratory variation expands with increasing NMAS for a given specimen geometry. However, the reproducibility of D^R does not depend on the NMAS. Correspondingly, repeatability precision limits for D^R were established as a function of NMAS and specimen geometry, whereas reproducibility limits were established as a function of specimen geometry only.

- The ILS results show that cyclic fatigue test results for a given mixture and specimen geometry, which meet the proposed precision limits, can yield different traffic designations according to the proposed S_{app} thresholds.[23] This finding suggests that the spacing of S_{app} thresholds for different traffic designations may merit reconsideration. Under the proposed precision limits, the S_{app} COV under reproducibility conditions can be as high as 57 percent for 19-mm mixtures with the small-specimen geometry. An analysis of the uncertainty in FlexPAVER version 1.1 predictions yielded results that were correlated to those from S_{app}. The implications of the observed uncertainty in S_{app} and FlexPAVE, given the proposed precision limits, merits further consideration in future work. Possible measures in specimen fabrication procedures to reduce test variability also merit investigation to reduce uncertainty in the test results.

- The researchers recommend that the proposed precision statements presented in chapter 3 be incorporated into the AMPT cyclic fatigue test procedures.

APPENDIX. INITIAL PILOT TEST RESULTS EVALUATION AND RESOLUTION OF PROCEDURAL DISCREPANCIES

The research team evaluated the initial pilot dynamic modulus test results from participating laboratories according to the data quality statistics specified in AASHTO TP 132-19 and the agreement to the reference results.[8] Agreement to the reference results was assessed using the precision limits for $|E^*|$ and phase angle specified in AASHTO T 378-17 since precision limits were not available for AASHTO TP 132-19.[8,9] In general, laboratories met the data quality statistics requirements. However, the research team identified cases where the differences between the participating laboratory's results and those of the research team exceeded the reproducibility precision limits specified in AASHTO T 378-17.[9] In the majority of these cases, the participating laboratory did not provide adequate thermal equilibration time after placing the test specimen in the AMPT chamber. AASHTO TP 132-19 specifies that a specimen can be first preconditioned to the test temperature in an external chamber and then transferred to the AMPT chamber.[8] If the transfer, linear variable differential transformer (LVDT) bracket, and LVDT are attached, and the AMPT chamber returned to the testing temperature in 5 min, then the standard allows the user to begin the test as soon as the AMPT chamber returns to the testing chamber.

The pilot test results suggest that this procedure did not allow adequate thermal equilibration time and led to discrepancies with the reference results where thermal equilibration was achieved. In cases identified as having insufficient thermal equilibration times, the participants were asked to retest following the temperature conditioning guidance established for AASHTO TP 133-21 in Phase I of this contract using a heat transfer model.[2] The temperature conditioning guidance in AASHTO TP 133-21 specifies the required thermal equilibration time for a small test specimen in the AMPT chamber as a function of the difference between the test temperature and room temperature, whether the specimen is conditioned directly in the AMPT or first preconditioned in an external chamber, and the test setup time for cases where the specimen is preconditioned in an external chamber.[2] Increasing the thermal equilibrations to this guidance was found to resolve the majority of cases that failed to meet the reproducibility limits. This result suggests that the temperature conditioning guidance in AASHTO TP 132-19 should be revised.[8] The research team recommends revisions that are, in part, informed by the experiences gained during these pilot studies.

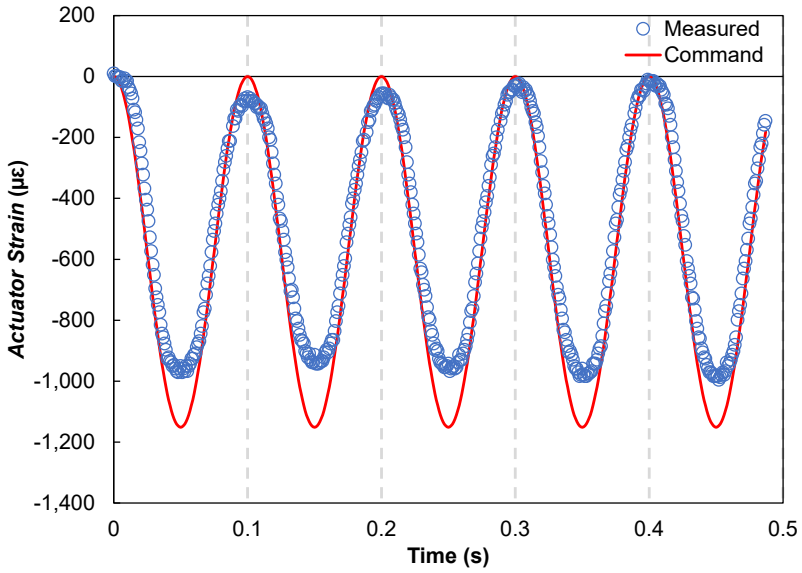
In two cases, a discrepancy between the reference results and the participating laboratories' results still existed after thermal equilibration time differences were resolved. At this stage, these two laboratories were asked to check the temperature of their AMPT chamber using an external temperature probe and compare the reading to the AMPT software's reported value. In one case, a difference between the temperature probe and AMPT software readings was identified, indicating a temperature calibration error. The laboratory recalibrated its AMPT temperature chamber, which was found to resolve the discrepancy with the reference results. In the other case, no temperature calibration error was found. Furthermore, in this case, the discrepancy between the reference results and the participating laboratory's results was only evident at 35 °C. The researchers found that the laboratory was using compensation springs at 35 °C, whereas the other laboratories (including the research team's laboratory) were not. Compensation springs are allowed in AASHTO TP 132-19 to avoid positive strain drift during testing.[8] Repeating the tests

without the compensation springs yielded results that agreed with the reference results, according to the AASHTO T 378-17 precision limits.[9] This finding suggests that the allowance of compensation springs merits further evaluation in future work. Compensation springs were not used in any subsequent ILS testing to avoid potential bias. Again, the research team recommends revisions to the allowance (or disallowance) of compensating springs within AASHTO TP 132 that are, in part, informed by the experiences gained while evaluating troubleshooting with this laboratory.

The researchers evaluated the initial AMPT cyclic fatigue pilot test results based on adherence to the strain selection guide in AASHTO TP 133-21 and the average actuator strain standard error in cycles two through five of the cyclic fatigue test.[2] AASHTO TP 133-21 specifies that the average actuator strain standard error in cycles two through five of the cyclic fatigue test shall not exceed 10 percent. AASHTO TP 133-21 specifies initial strain selection based on the linear viscoelastic $|E^*|$ measured from either AASHTO TP 132-19 or AASHTO T 378 testing or the fingerprint portion of the cyclic fatigue test. Thus, a comparison between the strains specified in AASHTO TP 133-21, according to the reported $|E^*|$ values, was used to assess adherence to the standard. Identified deviations from the standard were discussed with the participants and resolved in subsequent tests.

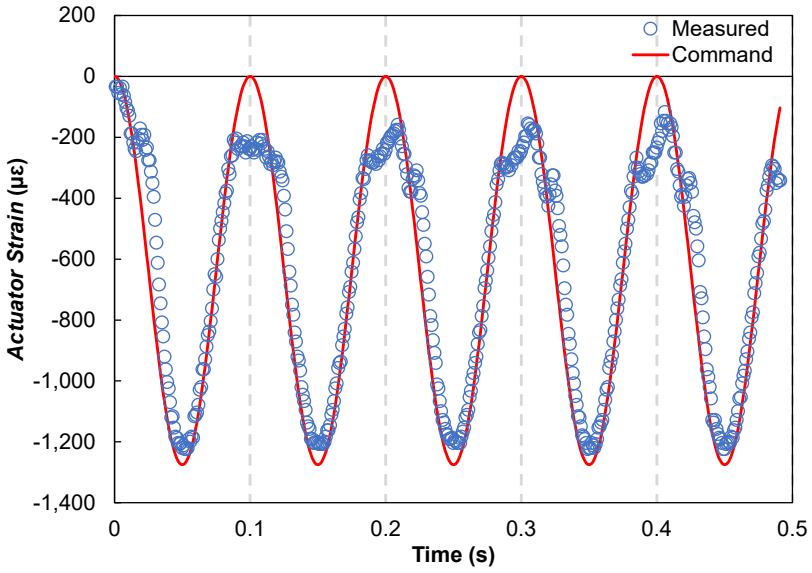
The researchers reviewed the actuator strain waveform in cases where the maximum limit on actuator strain standard error was exceeded. If the actuator strain curve appeared smooth, such as in the case shown in figure 48, the laboratory was instructed to perform PID tuning and then repeat the cyclic fatigue test. In these cases, the repeated test results were in good agreement with those of the research team in terms of the damage characteristic curve and failure criterion. In one case, the actuator strain curve was not smooth, as shown in figure 49. The “hiccup” in the actuator strain signal indicates a loose connection in the cyclic fatigue test setup. In this case, the research team worked with the participating laboratory to identify the source and resolve the loose connection. After resolving this issue, the test was repeated, and the actuator strain signal appeared smooth. Also, after resolving the loose connection, the damage characteristic curve and failure criterion results agreed with the research team’s results.

The final pilot test results obtained after resolving all procedural issues are presented in the subsequent sections.



Source: FHWA.

Figure 48. Graph. Example of actuator strain results without a loose connection.

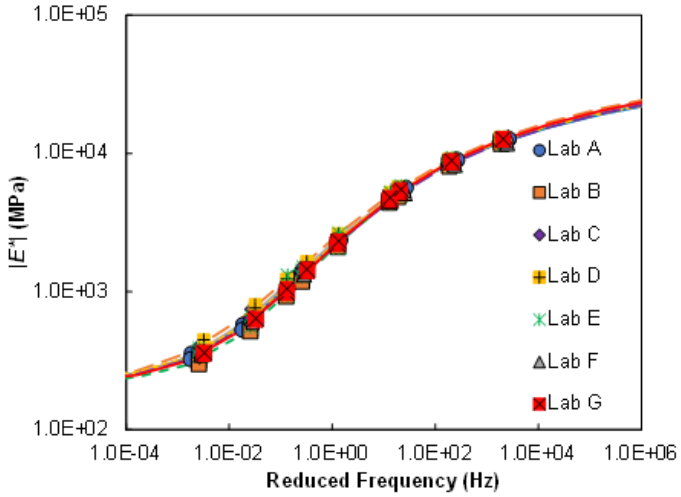


Source: FHWA.

Figure 49. Graph. Example of actuator strain standard error results indicating a loose connection in the cyclic fatigue setup.

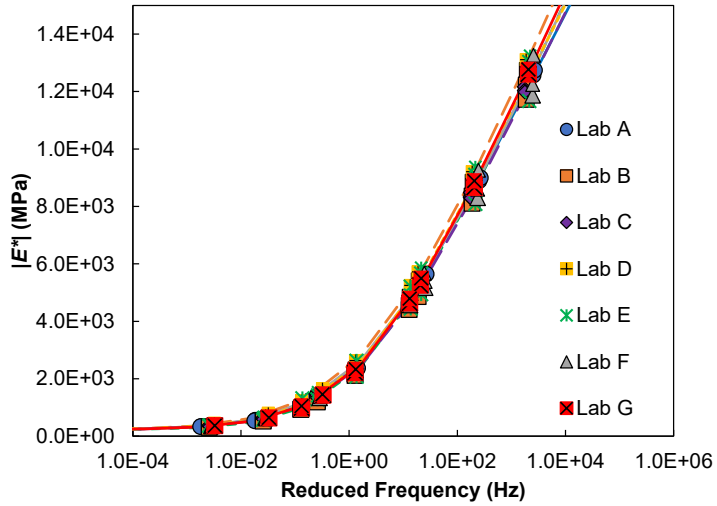
FINAL PILOT TEST RESULTS

Figure 50 and figure 51 show the |*E**| pilot test results using log-log and semilog scales, respectively. Figure 52 and figure 53 show the damage characteristic curve and failure criterion pilot cyclic fatigue test results, respectively.



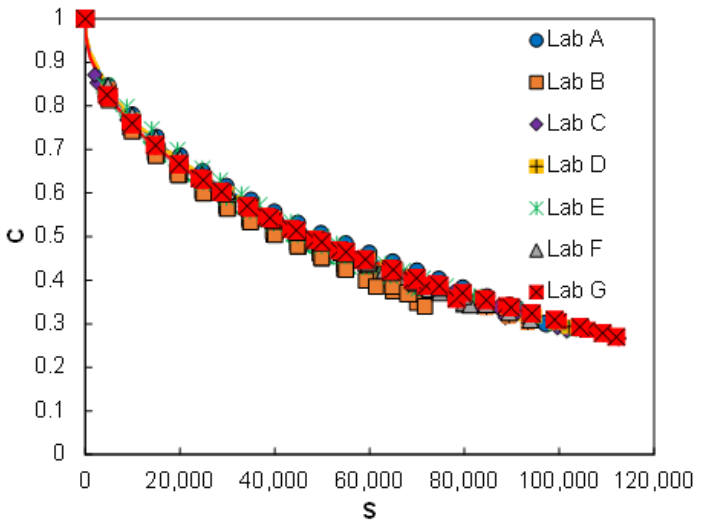
Source: FHWA.

Figure 50. Graph. Dynamic modulus master curve for pilot test results (log-log scale).



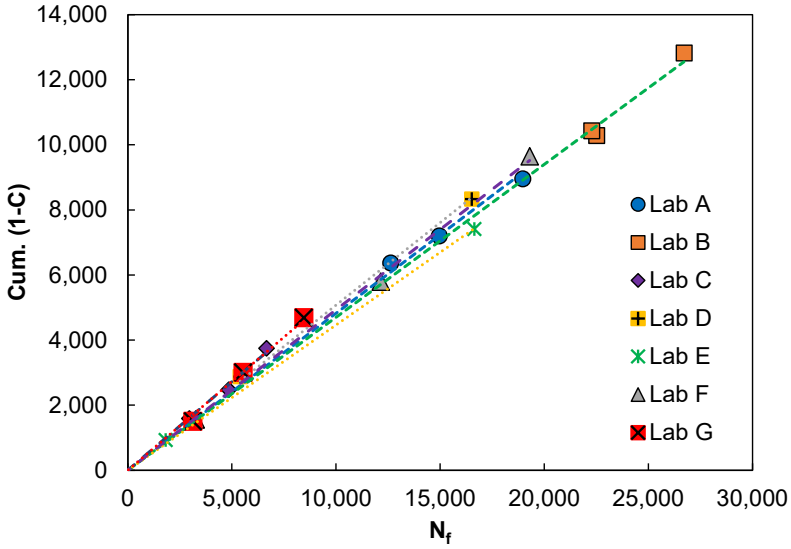
Source: FHWA.

Figure 51. Graph. Dynamic modulus master curve for pilot test results (semilog scale).



Source: FHWA.

Figure 52. Graph. Damage characteristic curves for pilot test results.

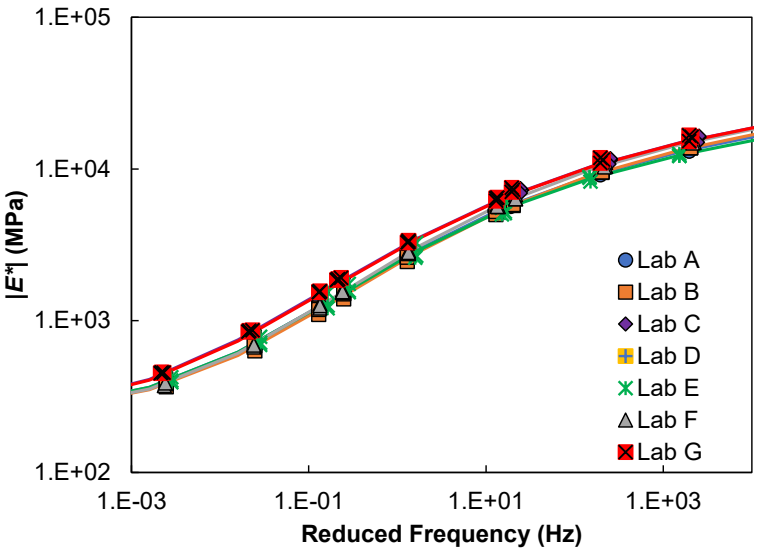


Source: FHWA.

Figure 53. Graph. Failure criterion for pilot test results.

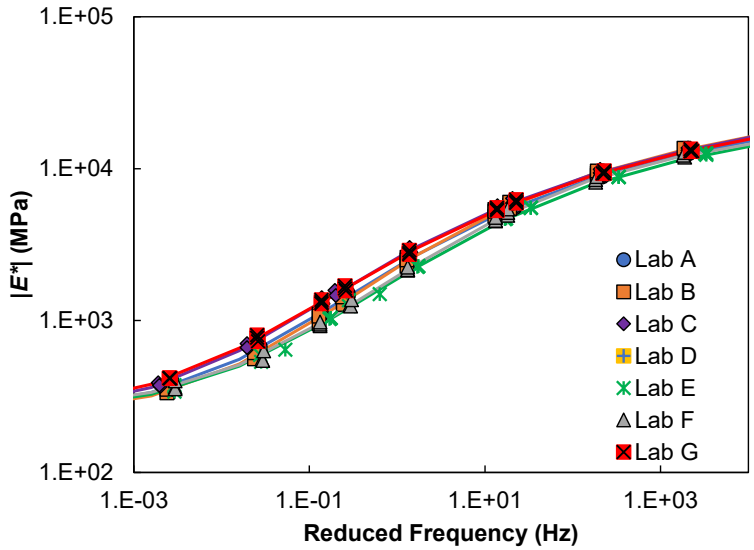
FULL-SCALE ILS DYNAMIC MODULUS TEST RESULTS

Figure 54 through figure 59 show the full-scale |*E**| test results for the 9.5-mm large specimen, 9.5-mm small specimen, 12.5-mm large specimen, 12.5-mm small specimen, 19-mm small specimen, and 25-mm large specimens, respectively. Table 29 summarizes the |*E**| test results used for cyclic fatigue test analysis.



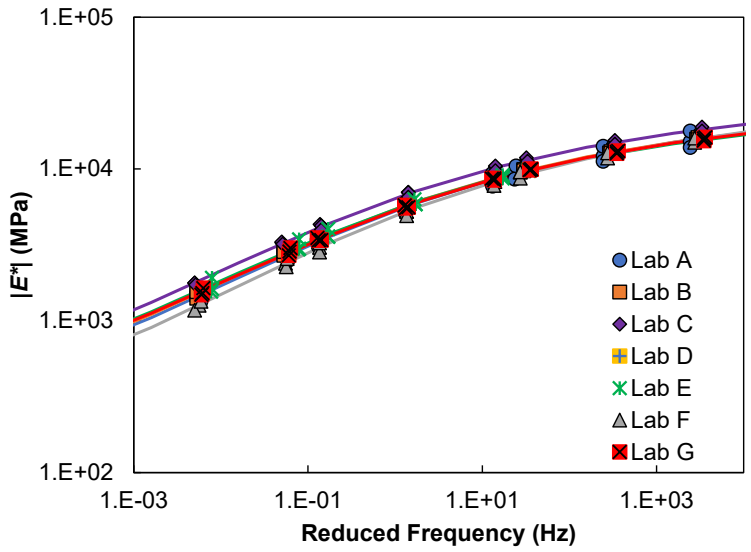
Source: FHWA.

Figure 54. Graph. Dynamic modulus test results for 9.5-mm mixture large specimen.



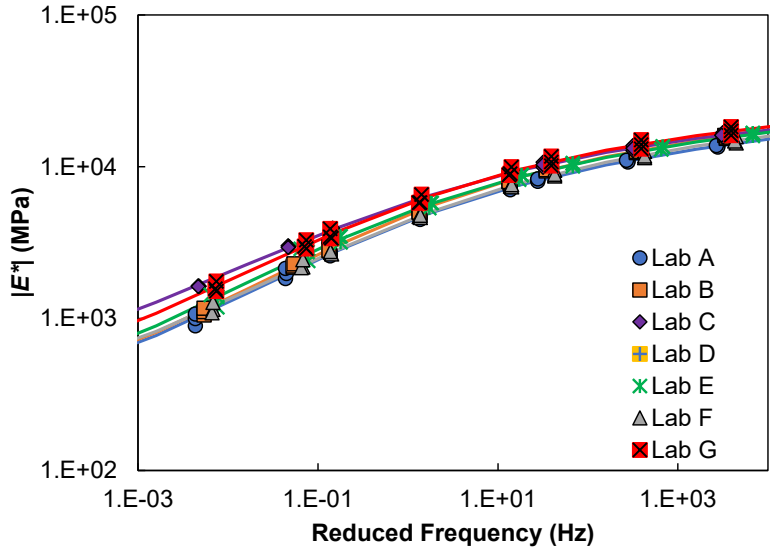
Source: FHWA.

Figure 55. Graph. Dynamic modulus test results for 9.5-mm mixture small specimen.



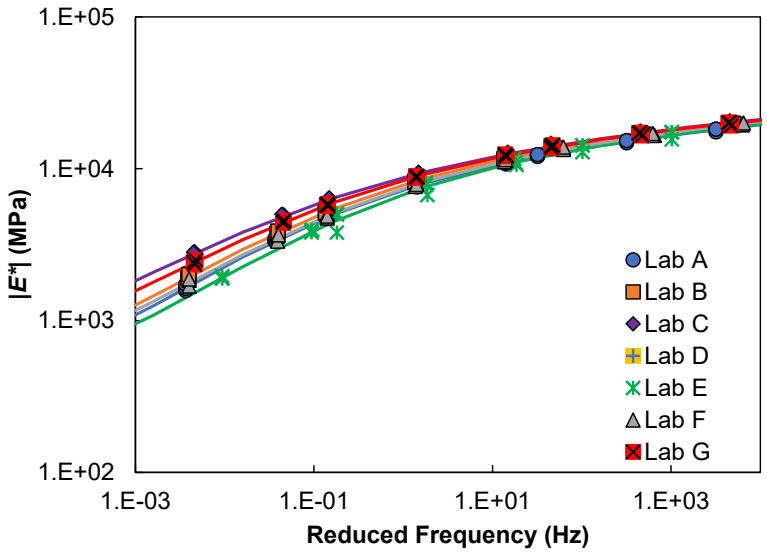
Source: FHWA.

Figure 56. Graph. Dynamic modulus test results for 12.5-mm mixture large specimen.



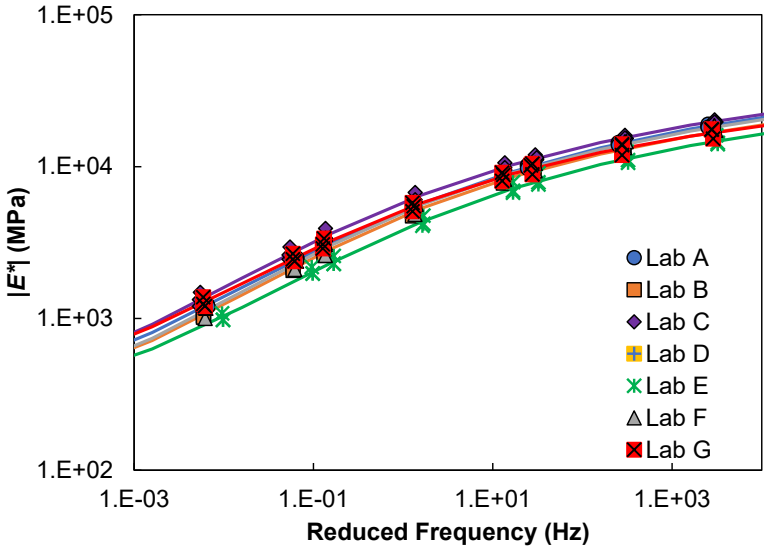
Source: FHWA.

Figure 57. Graph. Dynamic modulus test results for 12.5-mm mixture small specimen.



Source: FHWA.

Figure 58. Graph. Dynamic modulus test results for 19-mm mixture small specimen.



Source: FHWA.

Figure 59. Graph. Dynamic modulus test results for 25-mm mixture large specimen.

Table 29. Dynamic modulus data used for ILS cyclic fatigue analysis.

Specimen	Lab A	Lab B	Lab C	Lab D	Lab E	Lab F	Lab G
9.5-mm Large	Yes	Yes	Yes	No	Yes	Yes	Yes
9.5-mm Small	Yes	Yes	Yes	No	Yes	Yes	Yes
12.5-mm Large	Yes	Yes	Yes	No	Yes	Yes	Yes
12.5-mm Small	Yes	Yes	Yes	No	Yes	Yes	Yes
19-mm Small	Yes	Yes	Yes	No	Yes	Yes	Yes
25-mm Large	Yes	Yes	Yes	No	Yes	Yes	Yes

REFERENCES

1. AASHTO. 2022. *Determining the Damage Characteristic Curve and Failure Criterion Using the Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test*. TP 107-22. Washington, DC: American Association of State Highway and Transportation Officials.

2. AASHTO. 2021. *Determining the Damage Characteristic Curve and Failure Criterion Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test*. TP 133-21. Washington, DC: American Association of State Highway and Transportation Officials.

3. Castorena, C., B. S. Underwood, Y. R. Kim, K. C. Lee, N. Tran, and A. Taylor. 2021. *Ruggedness and Interlaboratory Studies for Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Tests: Phase I Report*. FHWA-HRT-21-057. Washington, DC: Federal Highway Administration.

4. ASTM. 2014. *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method*. E691-20. West Conshohocken, PA: ASTM International.

5. ASTM. 2015. *Standard Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials*. C670-15. West Conshohocken, PA: ASTM International.

6. LaCroix, A. T. 2012. *Performance Prediction of the NCAT Test Track Pavements Using Mechanistic Models*, PhD Dissertation, North Carolina State University, Raleigh, NC.

7. AASHTO. 2019. *Preparation of Small Cylindrical Performance Test Specimens Using the Superpave Gyratory Compactor (SGC) or Field Cores*. PP 99-19. Washington, DC: American Association of State Highway and Transportation Officials.

8. AASHTO. 2019. *Determining the Dynamic Modulus for Asphalt Mixtures Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT).* TP 132-19. Washington, DC: American Association of State Highway and Transportation Officials.

9. AASHTO. 2017. *Determining the Dynamic Modulus and Flow Number for Asphalt Mixtures Using the Asphalt Mixture Performance Tester (AMPT).* T 378-17. Washington, DC: American Association of State Highway and Transportation Officials.

10. Bonaquist, R. F. 2011. *Precision of the Dynamic Modulus and Flow Number Tests Conducted with the Asphalt Mixture Performance Tester*. NCHRP Report 702. Washington, DC: Transportation Research Board of the National Academies of Sciences, Engineering, and Medicine.

11. AASHTO. 2017. *Preparation of Cylindrical Performance Test Specimens Using the Superpave Gyratory Compactor (SGC)*. R 83-17. Washington, DC: American Association of State Highway and Transportation Officials.

12. AASHTO. 2017. *Developing Dynamic Modulus Master Curves for Asphalt Mixtures Using the Asphalt Mixture Performance Tester (AMPT)*. R 84-17. Washington, DC: American Association of State Highway and Transportation Officials.

13. FHWA. 2021. *FlexMAT*™ (software). Version 2.0.

14. Olard, F., and H. D. Benedetto. 2003. "General 2S2P1D Model and Relation Between the Linear Viscoelastic Behaviours of Bituminous Binders and Mixes." *Road Materials and Pavement Design* 4, no. 2: 185–224.

15. Underwood, B. S., Y. R. Kim, and M. N. Guddati. 2010. "Improved Calculation Method of Damage Parameter in Viscoelastic Continuum Damage Model." *International Journal of Pavement Engineering* 11, no. 6: 459–476.

16. Lee, K., C. Castorena, and Y. R. Kim. 2017. "Improving the Reliability of Damage Characteristic Curves in the Simplified Viscoelastic Continuum Damage Model." *Transportation Research Record: Journal of the Transportation Research Board* 2672, no. 28: 493–502.

17. Wang, Y., and Y. R. Kim. 2017. "Development of a Pseudo Strain Energy-Based Fatigue Failure Criterion for Asphalt Mixtures." *International Journal of Pavement Engineering* 20, no. 10: 1182–1192.

18. Ding, J., K. C. Lee, C. Castorena, Y. R. Kim, and B. S. Underwood. 2021. "Use of Resampling Method to Construct Variance Index and Repeatability Limit of Damage Characteristic Curve." *Transportation Research Record: Journal of the Transportation Research Board* 2675, no. 7: 194–207.

19. FHWA. 2020. *FlexPAVE*™ (software). Version 1.1.

20. Kim, Y. R., M. N. Guddati, Y. T. Choi, D. Kim, A. Norouzi, Y. Wang, B. Keshavarzi, M. Ashouri, A. Ghanbari, and A. D. Wargo. 2022. *Hot-Mix Asphalt Performance Related Specification Based on Viscoelastoplastic Continuum Damage (VEPCD) Models*. Report No. FHWA-HRT-21-093. Washington, DC: Federal Highway Administration.

21. Ghanbari, A. 2020. "Performance and Reliability Analysis of Asphalt Mixtures in a Performance-Related Specification." Doctoral dissertation. North Carolina State University.

22. Wang, Y. D., A. Ghanbari, B. S. Underwood, and Y. R. Kim. 2021. "Development of Preliminary Transfer Functions for Performance Predictions in FlexPAVE™." *Construction and Building Materials* 266: 121182.

23. Kim, Y. R., Y. Wang, and S. Underwood. 2019. *Cyclic Fatigue Index Parameter (S_{app}) for Asphalt Performance Engineered Mixture Design*. FHWA-HIF-19-091. Washington, DC: Federal Highway Administration.

24. Shapiro, S. S., and M. B. Wilk. 1965. “An Analysis of Variance Test for Normality (Complete Samples).” *Biometrika* 52, no. 3/4: 591–611.

25. Rice, J. A. 2006. *Mathematical Statistics and Data Analysis*. Boston, MA: Cengage Learning.

Recycled
Recyclable

Recommended citation: Federal Highway Administration,
Ruggedness and Interlaboratory Studies for Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test: Phase II Report
(Washington, DC: 2023) https://doi.org/10.21949/1521943

HRDI-10/06-23(WEB)E