CAPRI

CONSORTIUM FOR ASPHALT
PAVEMENT RESEARCH AND IMPLEMENTATION

Guidelines and Recommendations for Field Validation of Test Criteria for Balanced Mixture Design (BMD) Implementation

capriasphalt.us

Guidelines and Recommendations for Field Validation of Test Criteria for Balanced Mixture Design (BMD) Implementation

Report No. CAPRI-23.001-R

By

Randy West, Ph.D., PE
Director and Research Professor
National Center for Asphalt Technology
Auburn University, Auburn, Alabama

Nam Tran, Ph.D., PE
Assistant Director and Research Professor
National Center for Asphalt Technology

Fan Yin, Ph.D., PE
Assistant Director and Associate Research Professor
National Center for Asphalt Technology

Carolina Rodezno, Ph.D.
Associate Research Professor and Lead Researcher
National Center for Asphalt Technology

Thomas Harman
Senior Research Engineer
National Center for Asphalt Technology

October 2023

DISCLAIMER

ACKNOWLEDGMENTS

This work was sponsored by The Consortium for Asphalt Pavement Research and Implementation (CAPRI). https://capriasphalt.us/

The authors gratefully acknowledge the following members of CAPRI for their review of this technical report: Jason Blomberg, Missouri DOT; Andrew Brooks, C.W. Matthews; Jeff Kern, Champaign Asphalt; Zane Hartzog, Alabama DOT; and Tyler Wollmuth, North Dakota DOT.

The authors gratefully acknowledge Jim Musselman (Florida DOT) for his efforts while at NCAT in advancing this research.

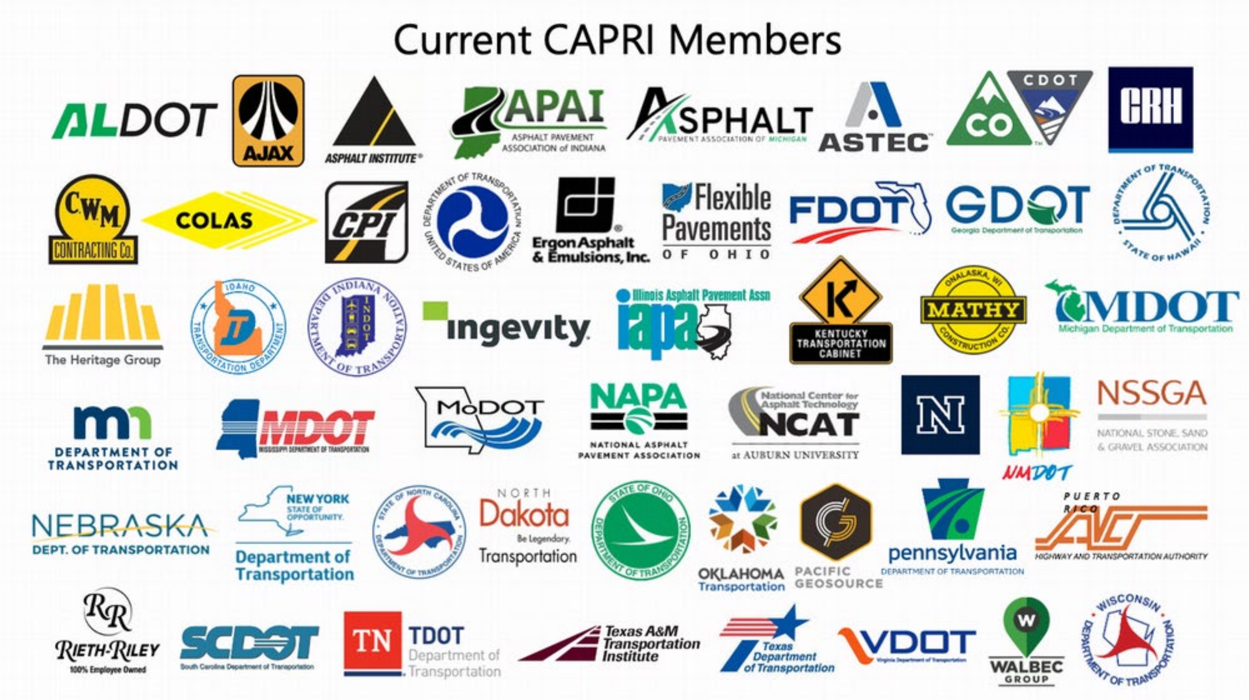


Table of Contents

List of Figures

List of Tables

Foreword

Balanced Mix Design (BMD) tests are important for assessing a mixture's resistance to specific distresses, such as rutting, cracking, and moisture damage. The efficacy of a BMD test as a specification requirement for mix design approval depends on establishing a robust relationship between test results and field performance. Such a relationship is also essential for developing appropriate specification criteria for quality assurance. However, there are only a limited number of studies that have conducted well-controlled field experiments to establish these relationships.

Developing laboratory-to-field relationships requires well-controlled field experiments that consider factors specific not only to an agency's traffic, climate, materials, and existing pavement structures but also to the types of distresses commonly encountered in the state. Highway agencies can conduct these experiments by constructing test sections suitable for their location or by collaborating with other Departments of Transportation (DOTs) to create field experiments for their region.

This document provides guidelines that highway agencies can follow when constructing test sections for field experiments. By adhering to these guidelines, the agencies can obtain results that are suitable for establishing valid correlations between BMD test results and field performance. This process will subsequently lead to the development of appropriate specification criteria. The organization of these guidelines is illustrated in Figure 1.

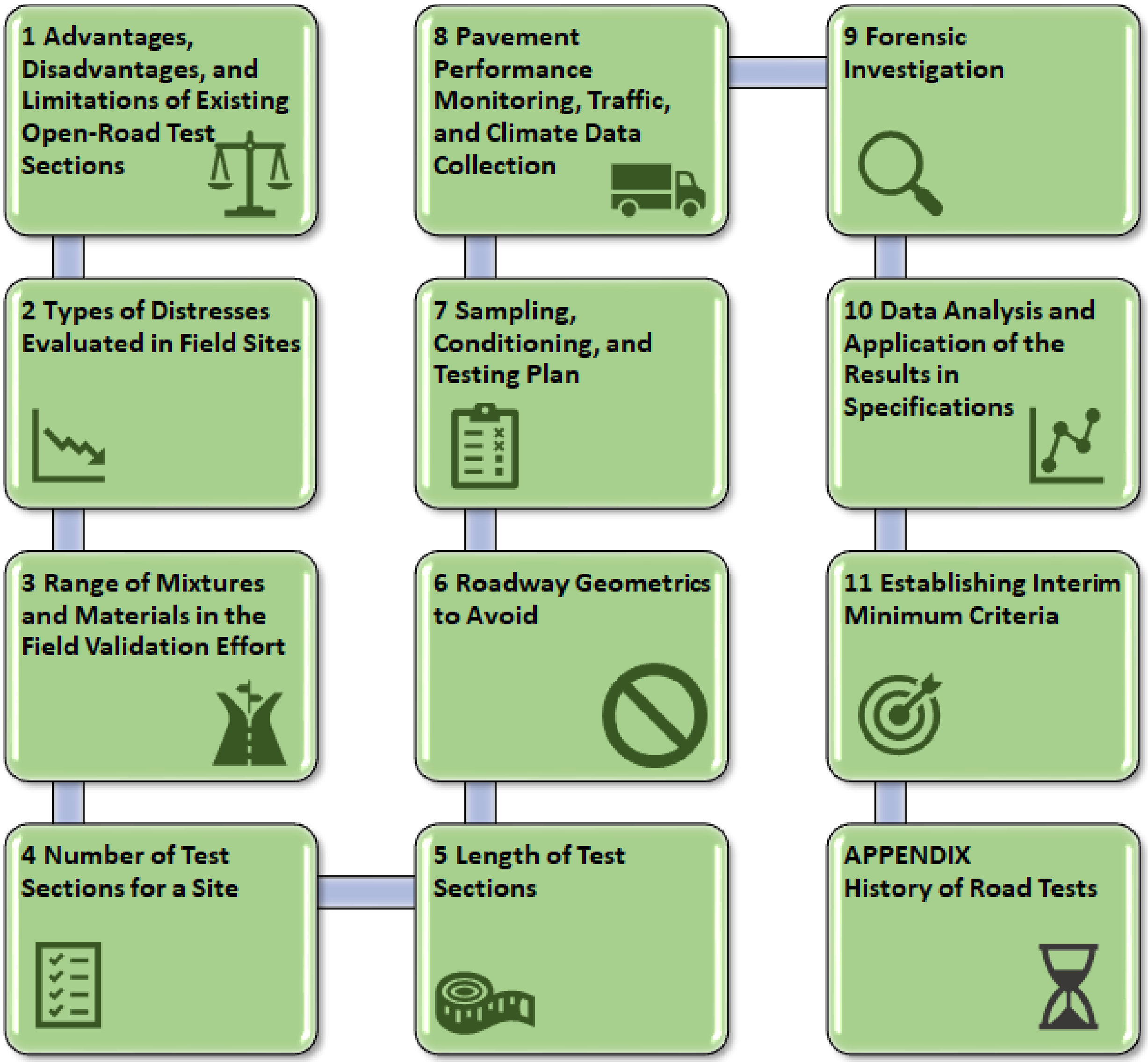


Figure 1. Organization of the Guidelines.

1 Advantages, Disadvantages, and Limitations of Test Section Approaches

Over the years, the asphalt pavement industry and researchers have used different approaches to evaluate the performance of pavement materials and designs including open-road test sections, and accelerated pavement testing facilities. Accelerated pavement facilities can consist of test tracks and heavy vehicle simulators (HVS) or accelerated load frames (ALF), which are capable of accelerated pavement testing. The ideal set of test sections should be easy to design, construct, traffic, monitor, and provide timely information to fulfill the objectives of the study.

Open-Road Test Sections

Open-road test sections involve the construction of pavement test sections with standard construction equipment and methods and trafficking of the test sections with actual vehicles. Test sections that carry actual traffic and are subjected to real environmental conditions and represent the most realistic approach for field validation experiments. Additionally, these sections allow for long-term data collection, providing valuable insights into pavement aging and deterioration over time. Moreover, open-road test sections typically require less initial investment compared to closed test tracks or loading simulators, since the only loading costs are related to traffic monitoring and weigh-in-motion devices, making them cost-effective.

However, open-road test sections also have some disadvantages. Data accumulation in these sections may be slower due to the gradual accumulation of performance data compared to accelerated testing approaches. Moreover, analyzing the effects of factors like weather and traffic volume can be challenging, potentially leading to increased variability in the results. Spatial variability in pavement performance may also occur in open-road test sections due to differences and changes in the underlying layers. In addition, state agencies are reluctant to leave roads in-service with significant distresses that may lead to greater risks for the road users.

The Long-Term Pavement Performance (LTPP) Program represents the most comprehensive research program that utilizes in-service test sections located in the United States and Canada. The program monitors the long-term performance of different pavement structures under different traffic conditions, climatic factors, subgrade soils, and maintenance, and rehabilitation programs. Other in-service roads include the Minnesota Road Research Project (MnROAD) which is a long-term full-scale pavement research facility.

Accelerated Pavement Testing (APT)

Accelerated pavement testing (APT) is the application of controlled moving wheel loads to a pavement or test sections at an accelerated rate compared with normal loading to determine its response in a compressed time. This allows us to have performance results in months rather than years.

APT includes closed test tracks and loading units. Test tracks are specially designed and carry actual trucks. Loading units include mobile or fixed units such as an HVS or an ALF that can apply known loads on an actual pavement section or a section specially built for evaluation.

Closed test tracks, like the NCAT Test Track, enable accelerated testing of numerous test sections simultaneously, allowing for quicker data collection and evaluation of pavement performance under extreme conditions. The sections under evaluation can be built with a consistent underlying support, the mixes to be evaluated can be designed to meet a particular performance test target, and traffic and performance of the sections is closely monitored. They also enable a better evaluation of environmental factors, reducing potential result variability. However, closed test tracks do not replicate real-world distribution of traffic loads so the rate of pavement damage may not precisely match those observed on open roads. Furthermore, setting up and maintaining closed test tracks require focused investments.

HVS and ALF facilities can provide precise control over loading and environmental conditions, but test sections are loaded one at a time, which can create challenges when comparing results of multiple sections tested over a span of months or years. They can be used for limited field validation experiments since the loading wheel that simulates traffic uses a slower speed than typical traffic that can substantially impact the response and performance of the sections. For these experiments, the field results can provide a good ranking of mixture performance, but they are not adequate to set performance criteria. In addition, these simulators may not fully replicate complex traffic loading patterns or environmental effects. The complexity and cost of setting up and operating loading simulators are additional limitations to consider.

Agency Network-level Pavement Management System (PMS) Data

Agency network-level PMS data to establish test criteria is the least expensive approach available to agencies since it relies on field performance data already gathered for other purposes that include supporting planning activities, and allocation of funds. PMS data can offer comprehensive information on pavement conditions and performance across a broad network. These data are advantageous for gaining an overall understanding of pavement conditions and identifying general trends. However, network-level data may lack the granularity required for specific performance evaluation of individual test sections. The accuracy of network-level data may vary depending on the data collection methods and instruments used. Additionally, this approach does not represent a well-controlled experiment, where mix designs are carefully selected, and an adequate characterization of the underlying conditions is conducted. Traffic and construction activities are generally not well documented or may exist in separate databases that are not suitably linked with the PMS database. The primary disadvantage of using PMS data for field validation are the confounding influences of underlying conditions, traffic loading, time of construction, and construction quality. This approach is not recommended since too many confounded factors will make the evaluation of field performance correlations unreliable.

Table 1 summarizes the advantages, disadvantages, and limitations of the different approaches for field validation of tests for mix design approval and quality assurance.

Table 1. Advantages, Disadvantages, and Limitations of Test Section Approaches.

Advantages	**Open-Road Test Section**	**Closed Test Track**	**Accelerated Loading Simulator**	**Agency Pavement Management Data**
Real-world Traffic	✓			✓
Real-world Environmental Conditions	✓			✓
Long-Term Data Collection	✓			✓
Cost Effectiveness	✓	✓		✓
Accelerated Testing		✓	✓	
Controlled Environment		✓	✓	
Controlled and Repeatable Testing			✓	
Comprehensive Data	✓	✓	✓	
Disadvantages				
Slow Data Accumulation	✗			✗
Limited Control	✗	✗	✗	✗
Spatial Variability	✗			✗
Limited Representation of Real-World Conditions		✗	✗	
Limited Flexibility		✗	✗	✗
Complexity and Cost			✗	
Granularity of Data				✗
Data Accuracy				✗

This guide focuses on open-road test sections for Field Validation of Test Criteria for Balanced Mixture Design Implementation.

2 Types of Distresses Evaluated in Field Sites

The primary forms of distress that typically control asphalt pavement service lives differ from state to state. State DOTs should design their BMD validation experiments to feature their most common form of distress(es), which may be obvious or may require an analysis of the state's PMS and/or discussions with the state's district or region maintenance engineers.

Each state's PMS should contain historical data on the condition of its pavement network. Many states now use vehicles equipped with high-speed, automated data collection sensors and high-resolution video to collect information used to rate pavement conditions across their entire highway network. Typically, pavements are rated by 0.01-mile segments for rutting, roughness (quantified as International Roughness Index, IRI), cracking, patching, and surface macrotexture. Some states classify cracking by orientation (i.e., longitudinal or transverse) and location (i.e., in wheelpaths or outside of wheelpaths), which may help in interpreting the cause of the cracking. Most DOTs combine these distress data into a single index or composite score for the segment. Although the composite scores are necessary for network-level analysis and budgetary planning, the analysis needed to help identify the cause of the distresses must be more granular. More specifically, cracking data should be categorized by cracking type and mode presented in Table 2.

Table 2. Categories of Cracking Distress.

Types of Cracking	Mode of Cracking
Load-related	○ Top-down cracking ○ Bottom-up fatigue cracking
Environmental	○ Thermal cracking (Transverse) ○ Block cracking
Reflective	○ Asphalt over concrete ○ Asphalt over asphalt

The following sections briefly describe experiment designs for the most common modes of asphalt pavement distress. Other distresses, such as slippage, longitudinal joint failures, and raveling are more commonly caused by construction issues than mixture quality issues. Therefore, they are not considered in this guide for field validation experiments for BMD tests.

For any field experiment, the pavement cross-section (layer thicknesses) should be the same throughout the test sections, and the base and subgrade moduli should be as consistent as possible in each section. Extra care must also be taken to ensure the layer thicknesses are closely controlled by surveying the grades and elevations of base layers prior to paving experimental layers of new or reconstructed pavements.

Some states may consider designing and building multiple field validation experiments to have field test sections that can be used to evaluate different distresses.

Rutting

Rutting is a distress that is typically manifest in the first few years of service. Although rutting can occur in the upper asphalt layers or in unbound base or subgrade layers, rutting in the upper asphalt layers is the distress of interest in a BMD field validation experiment. An example of rutting on an interstate pavement is presented in Figure 2.

Figure 2. Rutting on an Interstate Pavement.

The most straightforward approach to developing a field validation experiment for rutting is an experiment to evaluate the rutting resistance of the surface layer. Therefore, the experiment can be included in a typical overlay or mill and fill project or combined with several of the cracking test experiments on a deeper pavement rehabilitation, full-depth reconstruction, or new pavement construction project. However, when multiple asphalt layers are used in the project, consideration must be given to the rutting resistance of layers other than the experimental layer so that rutting measured on the surface of the test sections can be attributed only to the experimental layer without the need for a complicated and expensive forensic analysis to try to separate deformations in multiple layers.

Since acceleration and deceleration of heavy vehicles impart high shear stresses in pavements, which can lead to permanent deformation and other forms of distress, it is important to locate the test sections excluding intersections. This recommendation applies to experiments designed to evaluate asphalt mixture resistance to any of the distress types.

In all BMD field validation experiments, it is desirable for the test sections to include a wide range of field performance. For a rutting validation experiment, it is desirable to have a couple of test sections with very little rutting (excellent performance) with less than three to four millimeters of rutting after two years of service. Similarly, it is desirable to have a couple of test sections with marginal to poor performance, for example, 10 to 15 millimeters of rutting after a few years. To achieve a wide range of field performance in the test sections, the DOT will have to develop a special provision (or a supplemental specification) for the mix designs to have different ranges of rutting resistance based on the agency's selected rutting test.

Planning to have test sections with poor performance (in this case, rutting) means that the DOT should plan for early rehabilitation of some of the test sections. Furthermore, it is a good idea to install signs on the shoulders at the beginning of each test section to notify the driving public that the roadway features experimental pavements.

Load-related, Top-down Fatigue Cracking

A field experiment focusing on top-down cracking, as illustrated in Figure 3, is more challenging than other distresses. An experiment intended to produce top-down cracking should feature experimental surface layers with a range of cracking resistance. Generally, surface cracking does not appear for a few years, even for relatively stiff and brittle mixtures. Therefore, in-situ aging is expected to affect the performance of the test sections.

Figure 3. Top-Down Cracking of a Test Section on the NCAT Test Track.

The successful top-down cracking experiment on the NCAT Test Track from 2015 to 2020 was designed to have the experimental 1.5-inch surface layers on top of a common cross-section consisting of two 2.25-inch fatigue-resistant layers consisting of a Superpave mix using a highly modified asphalt binder, on top of a marginal-strength six-inch crushed aggregate base, over the strong native subgrade at the Test Track. Given the heavy loading for the track, this cross-section would have been expected to eventually have bottom-up fatigue cracking. The highly modified Superpave mixture used in the lower layers provided sufficient resistance to bottom-up fatigue cracking. The pavement structure under loading generated high deflections, which resulted in the necessary strains in the surface layers to induce top-down cracking. More information on the NCAT Cracking Group Experiment can be found in NCAT Report 21-03.

Based on the NCAT Test Track experience with the top-down cracking experiment, the recommended pavement design for a top-down cracking experiment is new construction or reconstruction that is intentionally designed to have a shorter fatigue life based on the design asphalt thickness but with very fatigue-resistant lower layers. DOTs can use their preferred pavement design method to determine a cross-section for a short life of five years for the expected traffic, then reduce the asphalt layer thickness by a small amount (e.g., 10%) but use a highly strain-tolerant mix in all layers below the surface experimental layer to avoid bottom-up fatigue cracking.

Load-related, Bottom-up Fatigue Cracking

A validation experiment for bottom-up fatigue cracking (Figure 4) must have the experimental layer as the lowest asphalt layer in either a new or reconstructed pavement. One option would be to design a multi-layer pavement cross-section with test sections using different experimental base layers using common (shared) mixtures in the upper layers.

Figure 4. Extensive Fatigue Cracking of an Asphalt Pavement.

A challenge with this experimental approach is that the cracking must propagate through the entire cross-section to assess the performance of the lower experimental layers. Alternatively, the entire cross-section in each test section could be built with the experimental mix design as either multiple layers or as a single layer. Single-layer construction of asphalt pavements is uncommon and presents unique challenges to achieve satisfactory smoothness but avoids potential issues with debonding or weak layer interfaces that can ruin the experiment. This approach was used in the NCAT Additive Group experiment built in 2021.

To induce fatigue cracking, the pavement thickness design should be prepared to generate sufficient tensile strains in the bottom layer to lead to fatigue cracking within a few years under traffic. A simple approach to design a bottom-up fatigue cracking experiment is to use a pavement cross-section that is considerably thinner than needed for the given subgrade modulus and the typical 20-year design traffic. DOTs can use their preferred pavement design method to determine a cross-section for a short life of five years for the expected traffic. This goal of having some sections reach “failure” poses a challenge for test sections on an open highway. The agency would need to plan for rapid repair of failing sections to minimize disruptions to the highway’s users.

Although the fatigue cracking resistance of lower layers is not significantly affected by aging, it may take several years of trafficking before fatigue cracking propagates to the pavement surface. As noted previously, it is important to build the test sections to have subgrade and base moduli as consistent as possible and for the thicknesses of asphalt layers to be as consistent as possible so that the primary variable affecting the occurrence of cracking is the fatigue resistance of the asphalt mixtures in the test sections.

Environmental, Thermal Cracking

Thermal cracking is a common distress for asphalt pavements in many central and northern states that commonly experience sub-zero (Fahrenheit) temperatures (Figure 5). Thermal cracks are initiated at the pavement surface when rapid drops in temperature cause contraction of the pavement.

Figure 5. Thermal Crack in an Asphalt Pavement.

Contraction causes tensile stresses in the asphalt that can exceed its tensile strength of these mixtures at low temperatures. The surface layer is more susceptible to cracking due to embrittlement of the binder from exposure to ultra-violet (UV) radiation and oxidation. However, thermal cracks quickly propagate through the entire depth of asphalt layers and often into the base and subgrade layers. Overlaying a pavement with thermal cracks, even if the upper layers are milled off, will quickly lead to reflection cracking over the thermal cracks.

Therefore, building a field validation experiment for thermal cracking is best achieved with a new pavement or a completely reconstructed pavement.

The MnROAD-NCAT Cracking Group Experiment from 2016 to 2022 featured thermal cracking as the targeted distress for eight test sections on I-94 in Minnesota. Details of this experiment are described in NCAT Report 23-03 (Vrtis, et al., 2023). Originally, the test cells for the experiment were in the driving and passing lanes of the mainline roadway. The pavement was reconstructed from the subgrade up including two asphalt layers. Each cell used the same experimental mix in the upper and lower lifts, with the lower layer placed three inches thick, and the upper layer placed two inches thick. A 10-foot outside shoulder and a 4-foot inside shoulder were constructed at the same cross-section as the travel lanes.

After two winters, the test cells began to show a variety of cracks. Limited coring identified that much of the cracking was caused by a weak or no bond between the upper and lower asphalt layers. The hypothesis for the origin of the distresses is that it began with water entering the pavement structure at the longitudinal joint between the travel lanes and the longitudinal joints between the shoulders and the travel lane. The water then migrated into the interface between the lifts and likely weakened the bond. Traffic on the travel lanes and freeze-thaw cycles accelerated the damage initiated by the weakened bond leading to widespread fatigue damage of the upper layer in many of the cells. This cracking compromised the experiment.

Fortunately, thermal cracking was evident in the inside and outside shoulders, but the other types of cracking were not present due to the absence of traffic. Therefore, the thermal cracking in the shoulders was monitored for two more years to establish the correlations with the lab tests. Despite the original test cells being compromised, the experiment was still useful as a thermal cracking validation experiment because the shoulders were paved with the same experimental mixtures. This example illustrates the possible usefulness of paved shoulders for environmental-related cracking (i.e., thermal and block cracking) validation experimental studies. The only caution for this approach is the possibility of “sympathy” cracks that can translate laterally from one lane to another, from a travel lane to a shoulder, or vice versa.

Reflection Cracking

Reflection cracking is a very common type of distress of asphalt overlays on both concrete and asphalt pavements. Reflective cracking, as presented in Figure 6, is caused by the concentration of tensile strains in the overlay due to load-related and environmental mechanisms. For asphalt overlays on concrete pavements, reflection cracking is considered inevitable despite a variety of attempted solutions such as interlayer paving geosynthetics, bond breakers, and numerous incarnations of crack relief layers; because the strain magnitudes are too high even for highly strain-tolerant asphalt mixtures (Allen, 1985). For this reason, the preferred approach for the rehabilitation of concrete pavements is slab fracturing prior to an asphalt overlay (Gu et al., 2020). However, the design of asphalt mixtures can affect the rate of reflection cracking propagation and deterioration once reflection cracks reach the pavement surface.

Figure 6. Reflection Cracks in an Asphalt Overlay on a Concrete Pavement.

For a reflection cracking experiment involving an asphalt overlay on an existing asphalt pavement, it can be very challenging to find an existing project that has a consistent extent and severity of cracking over a mile or so that can be subdivided into an adequate number of test sections. For that reason, MnROAD and NCAT have taken an approach to building experiments for reflection cracking that uses "*artificial cracks*" in the underlying asphalt pavement. In this approach, a new or existing asphalt pavement with no cracks is saw-cut full-depth in a grid pattern.

At the NCAT Test Track, saw-cuts in the longitudinal direction were made at 3-ft. intervals across the lane and every 15-ft. in the transverse direction. The 1/8th-inch wide artificial cracks/saw-cuts (width of the saw blade) were filled with a dry, fine-graded sand to avoid the cracks healing in the hot months. The reflection cracking treatments and overlay test sections were constructed over the artificial cracks following conventional paving practices.

The 2022 MnROAD Reflective Cracking Challenge used a similar approach. First, two new asphalt 2.5-inch layers were paved, then the saw-cuts were made completely through both layers, but only in the transverse direction at 25-ft. or 50-ft. intervals. No sand was added to the artificial cracks. After saw-cutting, the top 1.0-inch was milled off then the experimental mixtures were constructed as a 1.5-inch overlay.

For an asphalt overlay on concrete pavement, the formation and propagation of reflection cracks are controlled by the properties of the asphalt and the condition of the concrete pavement. A study by Dave et al. (2021) developed a decision tree-based tool for selecting asphalt mixtures and overlay designs for concrete pavements to prolong overlay lives based on the performance of 12 MnROAD test sections. However, this study was not specifically designed to validate cracking tests and had only four sections with common cross-sections but different surface mix designs. Other experimental variables were overlay thicknesses, the in-place density of the asphalt overlays, single and two-layer overlays, tack-coat application rates and methods, and the use of various interlayers. The existing 9.5 inch jointed reinforced (skewed and dowelled joints) PCC pavement was constructed in 1973, and over time, had some maintenance including some joint and slab replacements and diamond griding. At the time of the construction of the asphalt overlays, the PCC pavement was reported to be in fair condition with some joint faulting, spalling, and mid-panel cracks. Load-transfer efficiency (LTE) across the PCC joints, based on FWD testing prior to the overlays, was an uncontrolled experimental variable and ranged from 17% to 58% in the driving lane. The following parameters were found to have a significant influence on reflective cracking performance and were used as inputs in the reflection cracking model: traffic loading, climate, asphalt overlay thickness, LTE, asphalt overlay mixture type, and existing base layer modulus. Therefore, for an asphalt-on-concrete reflection cracking study to assess the impact of mix type and the relationship between lab test results and field performance, it is important to choose a test site with consistent concrete joint LTE and base moduli.

Moisture Damage Susceptibility

The two tests used by most highway agencies to assess the moisture susceptibility of asphalt mixtures (Figure 7) are AASHTO T 283 (Tensile Strength Ratio, TSR) and AASHTO T 324 (Hamburg Wheel Tracking Test, HWTT). Aschenbrener (1995) reported that the HWTT provided an excellent correlation with 20 Colorado pavements of known field stripping performance. Schram and Williams (2012) concluded that the HWTT stripping inflection point (SIP) and the ratio between the stripping slope and the creep slope were effective in discriminating 13 Iowa mixtures with a range of good to bad stripping in the field. In contrast, they found AASHTO T 283 to be ineffective in identifying stripping susceptibility. However, NCHRP Report 779 found that no stripping was evident in 34 HMA and WMA test sections in 12 states, despite 11 mixes having HWTT SIPs less than 5000 cycles but only two mixes having TSRs below 0.75 (West et al., 2014).

Figure 7. Cores from an Asphalt Pavement Showing Moisture Damage in a Near-Surface Layer.

Consequently, many highway agencies are uncertain about which test is better to protect against the risk of pavement failure due to moisture damage or what criteria to use for their selected method. Therefore, a series of robust validation experiments are recommended to assess the reliability of these or other test methods as indicators of the moisture susceptibility of asphalt mixtures.

For this validation effort, the utilization of multiple APT facilities to test pavement sections is recommended. ALFs, HVSs, and similar facilities can be used to apply loads to test sections with a range of moisture susceptibilities in controlled environments, including wet/semi-saturated conditions, in a relatively short period of time. Such facilities allow for “testing to failure” and enable forensic investigations without risking or inconveniencing the public. The use of multiple APT facilities would enable simultaneous testing of mixtures common to different areas of the U.S. and other countries. Several APT facilities have expressed interest in participating in this research. The following six tasks are proposed:

1. Design a coordinated series of experiments using multiple APT facilities to encompass a range of materials and environments. The use of APTs would enable controlled moisture conditions, avoid risks to the public due to expected failures of test sections, and facilitate detailed forensic analyses.
2. Select mixtures for each site with a range of moisture susceptibilities.
3. Construct test sections at each site with the regionally selected asphalt mixtures.
4. Sample the plant-produced mixtures and conduct mixture performance tests (TSR, HWTT, and possibly others)
5. Monitor and collect traffic/load data, environmental data, and performance data using destructive and nondestructive testing.
6. Analyze and summarize results, establish correlations, and make recommendations for appropriate criteria for use in specifications.

Table 3 summarizes the recommended approaches in this section.

Table 3. Summary of Recommended Approaches.

Type of Distress	Targeted Layer	Construction	Design Considerations	Additional Items
Rutting	Surface Layer	Overlay, or Mill & Fill	Lower Layers have High Rut Resistance	Avoid intersections
Top-down Cracking	Surface Layer (e.g., 1.5-inches)	New or Reconstruction with a fatigue-resistance intermediate layer	Consider designing for a short design life	Resource: NCAT 2015-2020 Test Track
Bottom-up Cracking	Sufficient tensile strains in the bottom layer	New or Reconstruction	Considerably thinner than needed	Resource: NCAT Additive Group 2021
Thermal Cracking	Surface Layer	Overlay, or Mill & Fill		Resource: MnROAD-NCAT Cracking Group 2016-2022
Reflective Cracking	Surface Layer	Artificial Cracks (sand / no sand options)		Resource: MnROAD-NCAT Reflective Cracking Challenge
Moisture Susceptibility	Surface layer	APT Facility	AASHTO T283 or HWTT	Resource: List of six proposed research tasks

Resources:

Link to NCAT Test Track Reports:
https://eng.auburn.edu/research/centers/ncat/testtrack/reports.html

Link to MnROAD Resource Website:
https://www.dot.state.mn.us/mnroad/

3 Range of Mixtures and Materials in the Field Validation Effort

To establish robust laboratory-to-field relationships, it is crucial to construct a validation experiment consisting of field test sections where the only variable is the asphalt mixtures for a single layer that is most relevant to the distress(es) of interest. For example, if the focus is on *top-down cracking*, the validation experiment should primarily target surface mixtures, whereas, for *bottom-up cracking*, the emphasis should be on base mixtures. It is essential to maintain consistency among the test sections in terms of underlying support conditions, traffic, layer thickness, and pavement age to avoid confounding their field performance differences.

When selecting mix designs for the validation experiment, the primary objective is to encompass a wide range of results for the BMD tests being considered by the agency rather than specifically covering mix design variables like binder grade, aggregate type, or RAP content. An experimental design that includes mix designs with different BMD test results is expected to yield a range of field performance, enabling the discrimination between mixtures with good and poor performance. In this regard, it is recommended to include one or two test sections using currently accepted mix designs as a reference point. Additionally, the experiment should explore mix designs beyond current practices to achieve a wider range of BMD test results. Table 4 presents a list of common mix design strategies for improving performance test results. For a more comprehensive discussion of mix design strategies aimed at adjusting BMD test results, refer to the National Asphalt Pavement Association (NAPA) online *Balanced Mix Design Resource Guide*. It should be noted that some mix design adjustment strategies may impact the mixture's volumetric properties, requiring the relaxation or elimination of the agency's volumetric requirements.

Table 4. Common Mix Design Strategies to Adjust BMD Test Results.

Rutting Resistance	Cracking Resistance	Moisture Resistance
• Adjust aggregate gradation • Use a stiffer asphalt binder • Polymer modification • Lower asphalt binder content • Increase recycled materials content • Add fiber additives	• Increase asphalt binder content • Lower recycled materials content (*) • Use a softer (better quality) asphalt binder • Polymer modification (in most cases) • Add a rejuvenator	• Add an anti-strip agent • Change binder source • Change aggregate type

(*) - Crack-resistant mixes can be developed with high recycled material content.

Additional Resource: FHWA-UNR TechBrief, "*Adjustment of Asphalt Mix Design/Job Mix Formula to Satisfy Mechanical Test Properties*."

For agencies that have already established preliminary BMD criteria, it is highly recommended to include asphalt mixtures in the validation experiment that both pass and fail the proposed criteria for each form of distress being evaluated. This inclusion allows for the validation or adjustment of the criteria based on field performance data. For instance, if the aim is to validate the Hamburg Wheel Tracking Test (HWTT) and Indirect Tensile Asphalt Cracking Test (IDEAL-CT) criteria for rutting and top-down cracking evaluation, respectively, the experiment should include asphalt mixtures in different zones of the HWTT-versus-IDEAL-CT performance diagram, as depicted in Figure 8.

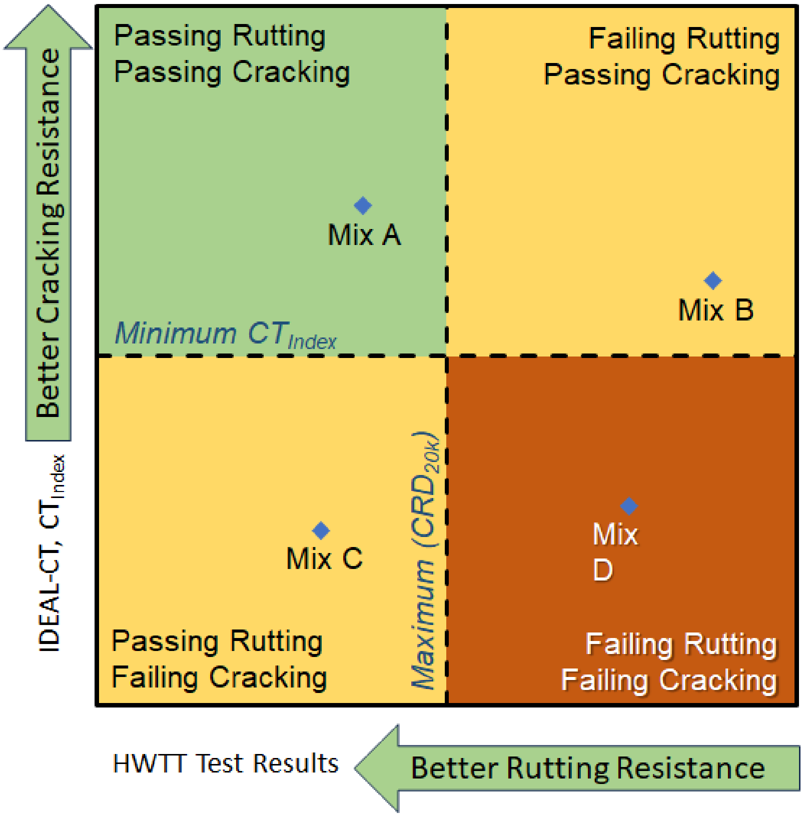


Figure 8. Example of using HWTT-versus-IDEAL-CT Performance Diagram to Select Asphalt Mixtures for Field Validation Experiment.

Since the validation experiment intends to generate a wide range of field performance among the test sections, it is essential for the agency to assume the risk of some test sections performing poorly or even failing prematurely. Without such instances, it becomes challenging to determine the appropriate criteria that effectively distinguish mixtures with good and poor performance. Therefore, when planning the validation experiment, it is imperative to include a contingency plan for repairing, removing, or replacing test sections that reach a predefined distress condition.

By incorporating these guidelines and recommendations, the field validation effort can comprehensively evaluate and establish the necessary criteria for BMD implementation.

4 Number of Test Sections for a Site

A critical aspect of the field validation experimental plan is determining the appropriate number of test sections. The decision regarding the number of test sections has implications for establishing correlations between laboratory test results and field performance, as well as the overall confidence in the test criteria.

It is essential to consider the trade-off between data richness and the associated costs when making this decision. While a larger number of test sections can provide more data points and enhance confidence in the test criteria, it also increases the upfront mix design work, testing efforts, instrumentation requirements *(if applicable)*, and the time needed for monitoring pavement performance. In addition, it impacts production logistics, stockpiles, and operations.

To strike a balance between cost and robustness, a field experiment with six test sections is generally recommended. This configuration allows for the inclusion of test sections exhibiting both good and poor field performance. Table 5 presents a generalized matrix of six test sections, covering mixtures with a wide range of rutting and cracking resistance.

For instance, Section ① employs an experimental mixture with low-rutting and medium-cracking resistance to simulate the low-performing scenario for field performance, while Section ⑥ utilizes an experimental mixture with high-rutting and medium-cracking resistance to represent the high-performing scenario. The remaining sections encompass mixtures spanning various levels of rutting and cracking resistance. The partial-factor matrix provides two sections for the three resistance levels for both modes of failure. By incorporating six test sections, it becomes feasible to establish robust laboratory-to-field correlations for rutting and cracking performance, as depicted in Figure 9 (a) (b).

Table 5. Example Field Validation Experimental Matrix with Six Test Sections.

Rutting Resistance	**Cracking Resistance**		
	Low	Medium	High
Low		①	②
Medium	③		④
High	⑤	⑥	

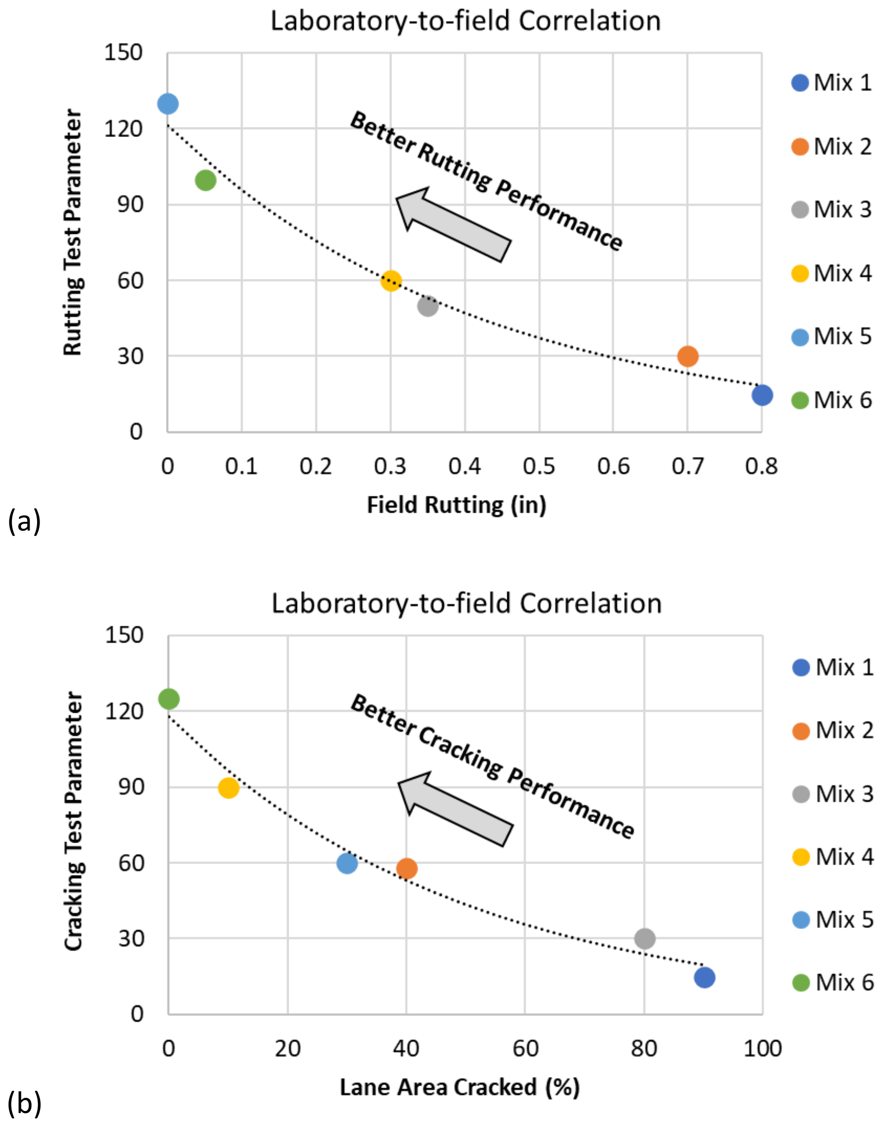


Figure 9. Hypothetical Laboratory-to-field Correlation Results from a Validation Experiment; (a) Rutting Correlation Results, (b) Cracking Correlation Results.

When interpreting the correlation results, caution must be exercised to ensure that the test sections exhibit notably different field performance and that a strong correlation exists between the field performance and laboratory test results, with more details described in Section 10 of this report.

5 Length of Test Sections

The length of test sections and associated considerations, such as location, labeling, and monitoring, are pivotal to ensuring both the validity and representativeness of the performance evaluation. This section provides a comprehensive look at these crucial aspects.

Location of Test Sections: For in-service test sections (open-road sections), their location should be selected so that they will be subjected to a consistent speed, exclude intersections, and have grades below 2%. In addition, the location of the sections should not create a safety risk to road users or people monitoring the sections. No maintenance, preservation treatments, or rehabilitation activities should be conducted during the monitoring period unless these activities are part of the evaluation program. If possible, sections should be located close to a traffic weigh station or weigh-in-motion (WIM) sensors should be installed to capture load data along with climatic data throughout the evaluation period.

- **Length of Test Sections**: The requisite length hinges on the type of test section, be it open-road or closed test track sections. A common length for in-service pavement test sections is 500-foot. This length provides ample tonnage for material sampling from each section during production, and it's sizable enough to facilitate both manual and automated pavement condition assessments. If manual evaluations or measurements are involved, traffic control becomes essential to ensure the safety of the survey crew.

  The LTPP program open-road test sections included general pavement studies (GPS) and specific pavement studies (SPS), where test sections consisted of 500-foot monitoring portion and a 50-foot materials sampling section at each end. In addition, on GPS test sections, maintenance control zones extend 500-foot in front and 250-foot beyond the limits of the sections as indicated in Figure 10. For SPS that consist of multiple test sections as part of a single project, the maintenance control zone is extended as indicated in Figure 11 (Elkins and Ostrom, 2021).

  The MnROAD mainline test sections are also 500-foot long, but with a 50-foot transition zone between cells.

  At the NCAT Test Track, each test section is 200-foot with 25-foot transition zones between the sections.

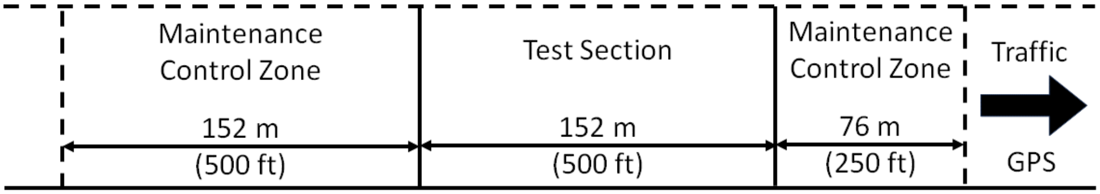


Figure 10. LTPP GPS Test Section General Layout (Elkins and Ostrom, 2021).

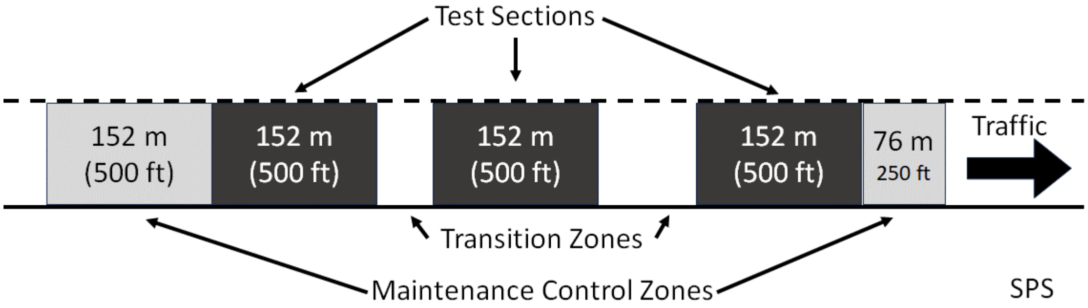


Figure 11. LTPP SPS Test Section General Layout (Elkins and Ostrom, 2021).

- **Safety Considerations:** The safety of road users and monitoring crews is paramount. When establishing test sections, especially those requiring manual assessment, appropriate traffic management plans should be in place. This includes signs, barriers, and trained personnel to guide traffic and protect the survey teams.
- **Subgrade Uniformity and Drainage:** Ensure that the subgrade is uniformly compacted and well-drained. Inconsistent subgrade conditions can significantly affect pavement performance. Adequate drainage measures, such as proper slope and drainage channels, should be in place to prevent water accumulation and its detrimental effects on the pavement structure. Regular inspection and maintenance of subgrade conditions throughout the testing process are essential for accurate performance evaluation.
- **Labeling of Test Sections:** During the original Strategic Highway Research Program (SHRP) and later under the LTPP program, test sections were marked with road signage, as shown in Figure 12. Labeling test sections offer several advantages, both for research personnel and for the public. Here are some of the primary benefits:
  Safety: For both research personnel and the public, labeled sections can signify areas where extra caution is needed. For instance, research personnel might be working close to the roadway, or there might be equipment set up nearby.
  Identification and Monitoring: For research personnel, labeled test sections make it easier to identify which sections are under observation, facilitating easier monitoring, data collection, and evaluation.
  Consistency in Data Collection: Proper labeling ensures that data is consistently collected from the correct section, reducing the risk of errors in measurement or monitoring.

Public Awareness: For the public, labels can indicate that a specific portion of the roadway is part of a study or experiment. This awareness can help in understanding any unusual roadway behavior or characteristics they might observe.

Coordination with Maintenance Crews: Road maintenance crews can easily identify which sections should not be interfered with (e.g., no repaving or patching) during the test period.

Ease of Communication: When discussing results, challenges, or observations, having labeled sections provides a common reference point for all stakeholders, ensuring everyone is on the same page.

Efficiency in Reporting: When compiling data and findings, labeled sections allow for organized and clear reporting. It's easier to correlate observed effects or anomalies with specific sections when they are properly labeled.

Transparency and Public Trust: Demonstrating to the public that certain road sections are under study can foster transparency. It shows that the responsible agencies are actively working to improve infrastructure and understand road behavior better.

Facilitation of Collaboration: When multiple teams or agencies collaborate on a project, having labeled test sections ensures that all parties involved can easily refer to and understand the specifics of the study area.

Training and Onboarding: For new personnel or those unfamiliar with the study, labeled sections serve as a direct and clear indication of the test areas, making it easier to get acquainted with the project.

Figure 12. Examples of SHRP/LTPP Test Section Road Signs.

- **GPS Coordinates for Test Sections**: GPS coordinates should also be incorporated into the delineation of the test sections; to provide highly accurate and consistent locations for performance monitoring.
- **Meaningful Pavement Condition Monitoring**: The length of the test sections should be sufficient to conduct meaningful pavement condition monitoring and performance evaluation. It is essential to capture enough pavement area to collect representative

data on distresses, rutting, cracking, roughness (e.g., IRI), and other performance indicators.

- **Transition/Buffer Zones**: The inclusion of transition or buffer zones is a prudent approach. These zones provide a buffer between the test sections; allowing the asphalt plant operation to switch to the new job mix formula (JMF).
    - Typically, a minimum of 40 to 60 tons of mix should be produced and placed prior to the start of the test section and sampling; to allow the production to achieve a steady state. Asphalt transport trucks are capable of payloads from 13 to 26 tons, typically around 18 to 20 tons.
    - E.g., for a 1.5-inch surface mix placed 12-feet wide, the buffer zone will be around 500-feet.
- **Production Logistics/Minimum Tonnage/Test Section Length**: Agencies need to engage the Contractor community early in the process to discuss and assess production logistics to ensure the desired job mix formulas (JMF) can be produced in the most efficient and effective way. Additional aggregate stockpiles, asphalt binders, quarry/pit operations, and other production process factors may facilitate the need for additional tonnage of each mixture. For example, in one state the contractor community recommended a minimum of 1500 to 2000 tons of each mix (approximately ½ day of production); to meet the proposed mixtures for a field validation project.
- **Sampling of Materials**: Sufficient length is necessary to allow for the sampling of materials from the test sections. Mix constitutive properties, BMD testing and cores are essential for conducting detailed laboratory testing to understand the material properties and performance characteristics of the pavement.
    - State agencies typically pull plant mix samples either from the transport truck at the production facility or pull samples at the roadway. Whatever method is commonly specified should be carried out for the test section construction.
    - Similar to production quality assurance (QA), sampling should be random and stratified to ensure a representative sample is taken.
    - The sampling rate for the BMD tests is a function of the combined variability of the material, sampling, conditioning, and testing.
    - Shadow projects support by NCAT and conducted by state DOTs has shown typical variability, as defined by the coefficient of variation (COV = Sample Standard Deviation / Sample Average), ranges between 15 to 20% for BMD performance indicators such as the IDEAL-CT (CT_{Index}) and the HWTT.
    - **Number of Samples per Test Section**: Table 6 shows the likelihood of a three replicate average representing the true population mean for different sampling rates, based on a range of COV. For example, if the IDEAL-CT test has a COV of

15% and the sampling rate is three samples per test section, then there is a likelihood that 20% of the test results will not be representative or 1-in-5. For a COV of 15% and nine samples, the likelihood falls to 5% or 1-in-20.

Table 6. Relationship between sample size (n) based on COV with 3-replicate testing per field test section and the standard error of the mean (SEM).

	COV (3 Replicates)		
Sample Size, n	10%	15%	20%
3	16%	20%	24%
4	12%	14%	15%
5	9%	10%	11%
6	7%	8%	9%
7	6%	7%	7%
8	5%	6%	6%
9	5%	5%	5%
10	4%	4%	4%
12	3%	3%	3%

Where: The SEM yields the likelihood of accepting a result statistically outside the true mean of the field test section.

Recommendation: For test sections where 3-replicate testing is performed, a likelihood of 5% is recommended, where seven to nine samples per test section should be taken.

- **Number of Replicates**: Similar to the COV, the number of replicates per sample impacts the likelihood of achieving a testing average that is representative of the true population mean. Table 7 shows the relationship between the number of replicates for a combined COV of 15%.

Table 7. Relationship between the number of replicates for a combined COV or 15% per field test section and the standard error of the mean (SEM).

	No. Replicates (Pop. COV 15%)		
Sample Size, n	3	4	5
3	20%	7%	6%
4	14%	6%	4%
5	10%	5%	3%
6	8%	4%	2%
7	7%	3%	2%
8	6%	3%	1%
9	5%	3%	1%
10	4%	2%	1%
12	3%	2%	1%

Where: The SEM yields the likelihood of accepting a result statistically outside the true mean of the field test section.

Recommendation: For test sections where 5-replicate tests are performed, four samples per test section should be taken, or where 4-replicate tests are performed, five samples per test section should be taken to obtain a SEM of 5%.

Note: This recommendation would apply to BMD tests like the IDEAL-CT index. BMD tests with higher variation (COV), would require a greater sampling and testing rate.

- **Variability Reduction**: Longer test sections can help reduce spatial variability in pavement conditions, especially in open-road sections where environmental and traffic variations may impact performance. By having longer sections, the influence of localized variations is better averaged out.
- **Traffic and Load Considerations**: The length should also consider the traffic volume and load characteristics of the road. Test sections should be long enough to capture the effects of repeated loading over time, which can be crucial for assessing performance under real-world traffic conditions.
- **Budget and Resource Constraints**: While longer test sections may provide more comprehensive data, it's essential to consider practical limitations such as budget and available resources. Balancing the need for representative data with the available resources is crucial.
- **Statistical Significance**: For the statistical analysis of the data, the length should be sufficient to ensure statistical significance in the results. Adequate sample size and length are crucial for drawing meaningful conclusions from the collected data.

Recommendation: Based on the above considerations, it is recommended that the length of the test sections be carefully chosen, taking into account the specific objectives of the validation efforts, the type of test section, traffic characteristics, and available resources. A minimum length of five hundred feet or more for open-road test sections is generally desirable to account for variability and achieve statistically significant results. For closed test track sections, the length can be shorter due to controlled testing conditions. Ultimately, the selected length should strike a balance between practicality, representativeness, and statistical significance to ensure the success and reliability of the validation efforts.

Example BMD Test Sections

Background: The state DOT has identified *top-down cracking* and *rutting* as key performance challenges based upon a review of their asset/pavement management system. They have conducted laboratory assessment of several of the BMD tests and have selected the IDEAL-CT test for cracking resistance and the HWTT for rutting resistance/moisture sensitivity. Shadow testing of last year's Superpave design mixes has provided them with a range of typical test results. Based on the *Guidelines and Recommendations for Field Validation of Test Criteria for*

Balanced Mixture Design (BMD) Implementation, they have adopted Table 5 *Field Validation Experimental Matrix with Six Test Sections* to design their open-road experiment.

The state DOT has established an Agency-Industry taskforce to identify challenges and address concerns in constructing the sections. NCAT provided a 1-day BMD workshop to kick off the taskforce. The state DOT has taken on the responsibility of designing the six mixtures.

Project Logistics:

- Asphalt Production Facility, 250 TPH (tons per hour)
- Mill and Fill, 1.5-inch surface mix (6 job mix formulas)
- 18-ton haul trucks
- Transition/Buffer Zone = 3 trucks / 54 tons
- BMD Test Section = 600 tons / 1.0 miles
- Two to three sections per day
- Four-replicates for each BMD test
- Five-samples per test section
- Sublot of 126 tons (600 tons / 5 samples) or 7 trucks

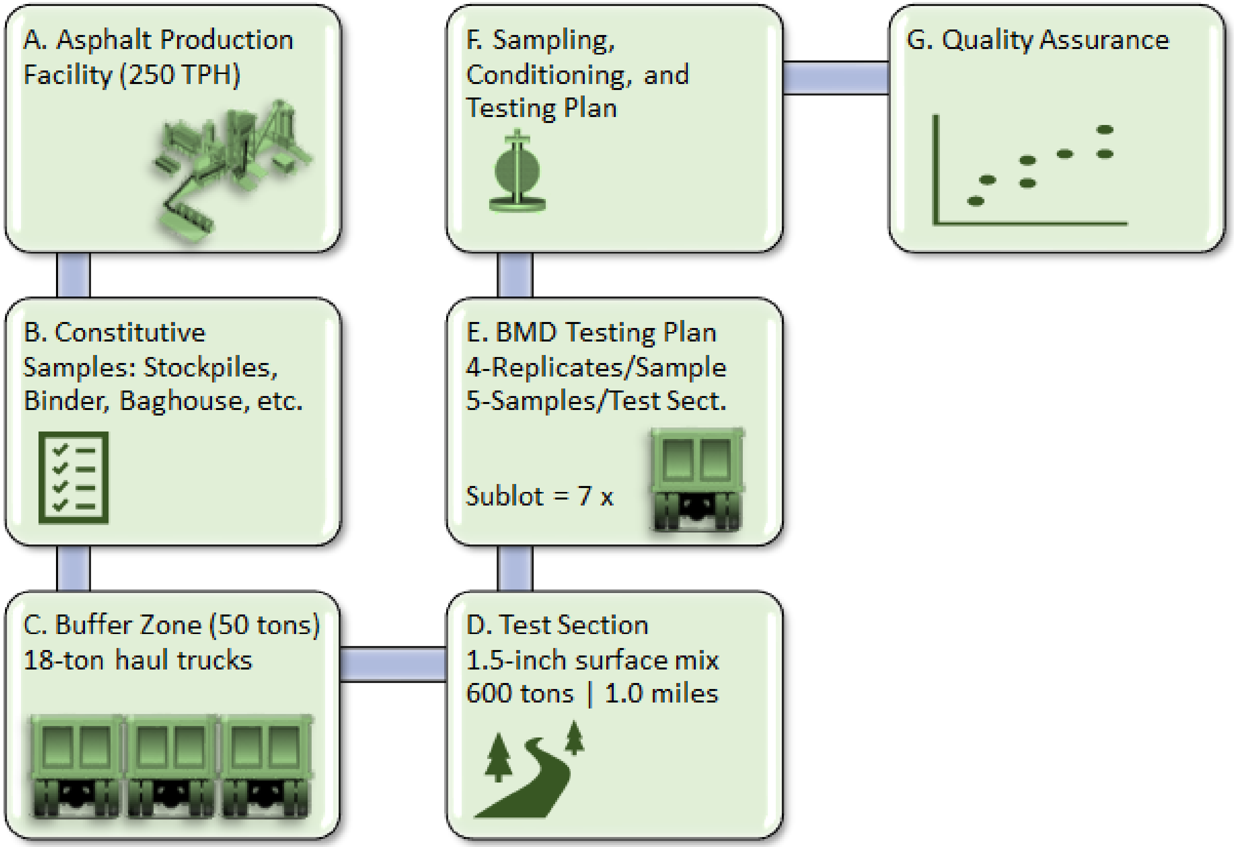


Figure 13. Flow of Test Section Production, Testing, and Placement.

6 Roadway Geometrics to Avoid

In the context of asphalt balance mix design and field validation, careful consideration of roadway geometrics is essential, as certain features may pose challenges or affect the performance evaluation of the asphalt mix. The following test outlines specific roadway geometrics that should be avoided during the field validation process:

> **Intersections**: It is advisable to not include intersections as part of test sections for field validation. Intersections typically experience higher traffic volumes, turning movements, and braking forces, contributing to increased stress and wear on the pavement. For accurate validation, it is more appropriate to focus on straight sections of roadways away from intersections.
>
> **Horizontal Grades**: Roadway sections with significant horizontal grades should be avoided for field validation whenever possible. Grades over 2% can impact the behavior of the asphalt mix due to the varying loads and forces acting on the pavement. It is preferable to choose relatively flat sections of the road to obtain more consistent and representative results during the validation process.
>
> **Curves**: Similarly, road sections with sharp curves should be avoided during field validation. These curves impose additional stresses on the pavement, leading to uneven loading and the potential for rutting or cracking. Selecting straighter sections of the roadway will help ensure that the performance evaluation is not significantly influenced by the presence of curves.
>
> **Variable Traffic Speeds**: The traffic speed within the test section should not change, and the typical traffic flow should be constant. This allows for a more accurate assessment of the asphalt mixture performance under similar traffic conditions.

By avoiding intersections, steep grades, sharp curves and variable speeds, the field validation process can focus on roadway geometrics that provides a more representative evaluation of the asphalt mix’s performance under typical operating conditions. This approach will help in obtaining accurate data and making informed recommendations for the mix design.

7 Sampling, Conditioning, and Testing Plan

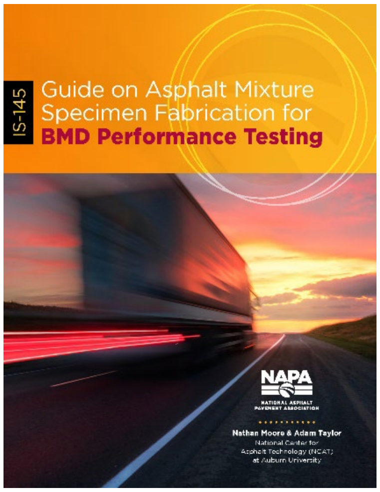
The sampling, conditioning, and testing plan is a critical aspect of the BMD validation effort. It ensures that the collected samples are representative of the actual pavement and that the testing provides accurate and reliable data for analysis. Considerations for this aspect of the BMD validation plan are as follows:

1. **Sampling Methods:** Provide detailed information on the types and numbers of samples required for the BMD tests of interest. The primary type of sample of interest is the asphalt mixture produced for each test section. The sampling locations should be strategically chosen to ensure representativeness of the entire pavement section length and generally should be consistent with the agency's mixture sampling location for acceptance testing. Sampling of the raw materials such as asphalt binder, aggregates, recycled materials, and additives are recommended for possible secondary analyses. Cores from the ends of the transition zones of the test sections are recommended for density tests and possible secondary analyses.
2. **Representativeness:** The sampling plan should emphasize the importance of representative sampling to obtain reliable data. Random sampling, stratified sampling, or systematic sampling methods may be employed based on the specific needs and characteristics of the project. Ensuring that the samples are representative of the pavement's overall composition is crucial for accurate analysis.
3. **Sample Storage and Reheating:** Provide guidelines for sample storage to avoid or minimize alterations to the material properties. Properly sealed containers and storage conditions are recommended to maintain sample integrity. Long-term storage of samples in a temperature-controlled environment is recommended. If reheating is necessary for preparing test specimens, an appropriate temperature and duration should be specified to prevent degradation.
    a. **Reheating:** Establish and document how samples are to be reheated.
    b. **Lag Time**: Is defined as the time from obtaining the mixture sample at the plant or paving site to laboratory compaction, within ± 15 minutes. Establish a consistent lag time and document the lag time for each set of compacted specimens for BMD testing in the test result notes.
    c. **Dwell Time**: Is defined as the time between laboratory compaction to the start of preconditioning for the selected performance test, with ± 15 minutes. Establish a consistent dwell time for the testing and document it with the test results.

4. **Sample Fabrication Resource**: Guide on Asphalt Mixture Specimen Fabrication for BMD Performance Testing, by Nathan Moore & Adam Taylor, NCAT.

   Go to:

   https://www.asphaltpavement.org/
   Select: NEWS & RESOURCES
   Selection: ONLINE STORE
   Under *Search our Product Catalog* select *Asphalt Production > Testing*, and then click on Search to find the Guide.



   Link and QR code to YouTube on Asphalt Mixture Sample Preparation Guide (NAPA IS-145), which includes:
   - Aggregate processing,
   - Aggregate batching,
   - Asphalt mixture heating and mixing,
   - Asphalt mixture field sampling, reheating, and splitting, and
   - Mixture aging and compaction.

   https://www.youtube.com/watch?v=diTrsPrHuSQ

5. **Sample Conditioning:** Provide detailed descriptions of the sample conditioning procedures to ensure consistency for the testing of the mixtures in the validation experiment. Different conditioning procedures may be used for samples or specimens for various BMD tests.

6. **Testing Procedures:** Document the testing procedures for each type of sample, whether it's BMD performance indicators, asphalt binder, aggregates, or cores. Specify the testing equipment, temperatures, specimen handling protocols, and standards to be followed to achieve consistent and reliable results.

7. **In-Place Density:** In-place density is an important property that impacts pavement performance. Specify the testing method and criteria to be used for in-place density.

8. **Additional Information:** Identify other characteristics that should be gathered during construction, such as: silo time, haul time, mix temperatures at delivery and at the start of compaction, tack coat materials and application rate, pavement layer thicknesses, etc. This information can complement the BMD tests and provide a comprehensive understanding of the pavement's performance.

9. **Conventional Testing:** Besides BMD tests, it is recommended that traditional QA tests be conducted and recorded for mixtures from each test section such as asphalt content, gradation, and volumetric properties.

10. **Quality Control and Assurance:** Include guidelines for quality control (QC – contractor process control) and assurance during the sampling and testing process. Require periodic checks on testing equipment, calibration, and repeatability of test results to ensure data accuracy and reliability.
11. **Split Samples:** Include split samples for BMD testing to be conducted by the agency, the contractor, and possibly an independent third-party lab. Document all logistical differences between specimen conditioning (reheating, lag, and dwell times). This will provide information to help assess the role of Independent Assurance (AI) in the future use of BMD testing in QA.
12. **Reporting and Documentation:** Stress the importance of thorough documentation throughout the sampling, conditioning, and testing process. Detailed records of sampling locations, procedures, and test results should be maintained for future reference and verification. Take and archive photographs of the test sections, sampling methods and locations, sample containers, storage facilities, sample handling procedures, and test methods to complement the documentation. Consider developing something similar to the SHRP LTPP field protocol data worksheet.

Recommendation: The sampling, conditioning, and testing plan should be comprehensive and well-documented, tailored to the specific needs and objectives of the BMD validation effort. Emphasize the importance of representativeness and quality control to ensure the accuracy and reliability of the data collected. Collaboration with experienced materials testing laboratories and adherence to established standards and protocols are essential to achieve meaningful results for the field validation test sections.

Setting aside additional samples for future testing is highly recommended (e.g., NCAT Test Track sampling approach): As test and conditioning procedures evolve over time, the ability to access samples becomes instrumental in the validation of emerging concepts.

EXAMPLE: SAMPLING AT THE NCAT TEST TRACK

Generally, 20 to 40 five-gallon buckets of mix are taken for each section at the NCAT Test Track.

In addition,

- Five gallons of the asphalt binder,
- Ten five-gallon buckets of each aggregate stockpile, including recycled materials, and
- One five-gallon bucket each of mineral filler, additives, and baghouse fines.

However, amounts are typically much more than this and can exceed 100 buckets of the mix, depending on the study.

8 Pavement Performance Monitoring, Traffic, and Climate Data Collection

Two key components of BMD specifications are acceptance criteria for rutting and cracking resistance of asphalt mixtures. These acceptance criteria can only be set based on field performance collected from test sections to ensure the asphalt mixtures produced based on BMD specifications in the future meet the DOTs performance expectations. Hence, it is essential to monitor the field performance of the test sections for both cracking and rutting.

This section provides state DOTs with a plan for monitoring their test sections' short- and long-term field performance. The following sections will provide information on the frequency of pavement performance, traffic, and climate data collection and protocols for collecting this information. This information helps state DOTs prioritize their resources and ensure that they collect quality data to develop their BMD specifications.

Pavement Performance, Traffic, and Climate Data Collection

The performance of asphalt pavements can be evaluated based on several key indicators, including cracking, rutting, roughness, and frictional properties, with rutting and cracking data being the most important for validating BMD specifications. Some state DOTs may also be interested in collecting friction data to ensure changes from volumetric mix design to BMD do not negatively impact the frictional properties of asphalt mixtures, potentially affecting the pavement's ability to provide safe driving conditions, especially in wet or icy conditions.

Since rutting often occurs early in pavement life, rutting data will be collected semi-annually in the first two to three years, or until rut depths are unchanged for the two consecutive measurements. From that point, rutting data can be collected annually. In contrast, cracking occurs later in pavement life, so cracking data can be collected annually until cracks are first detected in the test section. From that point, cracking data will be collected semi-annually. However, DOTs can collect both cracking and rutting performance data simultaneously semi-annually or annually using automated performance data collection vehicles.

Furthermore, friction testing data can be collected at the frequency DOTs often collect their frictional properties, which can be annually or biennially.

The frequencies recommended for performance data collection allow for detecting changes in the pavement surface caused by traffic loads and environmental effects. The amount of cracking and rutting may increase quickly if DOTs test the mixtures with the cracking and/or rutting resistance that does not meet the anticipated acceptance criteria.

Finally, traffic and climate data will be collected for the test sections throughout the performance evaluation period. This information is essential for evaluating the performance of the test sections, especially when they are in different traffic and/or weather conditions.

Protocols for Collecting Pavement Performance, Traffic, and Climate Data

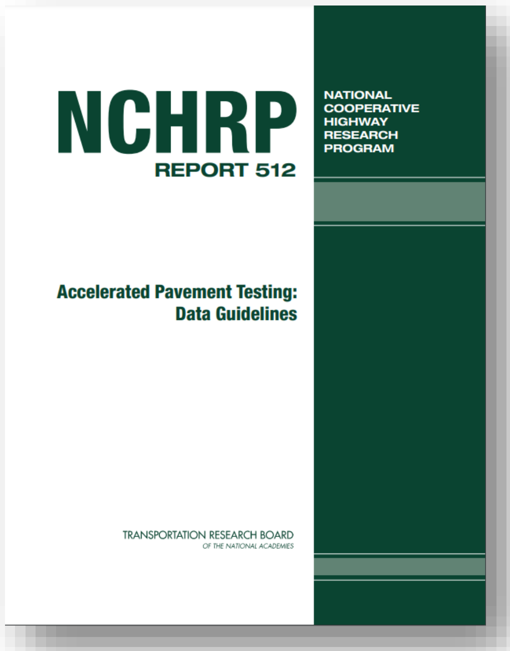
The protocols for collecting pavement performance, traffic, and climate data for test sections are critical to ensure the accuracy and reliability of the data collected. The following are some of the key considerations for the data collection:

- **Training and Certification.** The personnel responsible for collecting performance data should be trained and certified to ensure they are competent and knowledgeable in the data collection procedures and protocols. The training should include both classroom and field training, and the personnel should be regularly evaluated to ensure their continued competence.
- **Equipment and Tools.** The equipment and tools for collecting pavement performance, traffic, and climate data should be calibrated according to the manufacturer's instructions and verified using reference standards to ensure accuracy and reliability. The equipment and tools for collecting this data include laser scanners and/or cameras for rutting and cracking and a locked-wheel friction tester for frictional properties. In addition, automatic traffic recorders and WIM systems can collect traffic volume and weight data, and weather stations and pavement temperature sensors can gather climate and mid-depth pavement temperature data. Also, the same equipment should be used for data collection throughout the field validation period if possible.
- **Data Collection Procedures.** The data collection procedures should follow the same standardized methods throughout the collection period, such as the state data collection procedures for their pavement management systems (for automated data collection) or the LTPP data collection procedures (for manual data collection) (Simpson et al. 2006), to ensure that the data collected is consistent and comparable over time. The procedures should include instructions for selecting the sections where data are collected, measuring the performance indicators, and recording the data.
- **Data Management and Storage.** The collected data should be managed and stored in a centralized database to ensure accessibility and reliability. The database should include information on the test sections, the performance indicators, and the data collection dates. The data should be regularly backed up and secured to prevent data loss or corruption.
- **Data Quality Control.** Quality control measures should be implemented to ensure the accuracy and reliability of the data collected. This may include regular site visits to verify the data collected, cross-checking the data with other sources, and conducting statistical analyses to identify outliers and errors. In addition, all states are required to have Federal Highway Administration (FHWA) approved Data Quality Management Plans for their network-level condition data.

By following these protocols for collecting performance data, state DOTs can help ensure that the data collected is accurate, reliable, and comparable over time. This will enable them to evaluate the performance of their asphalt pavements and prioritize maintenance and repair efforts to ensure their safety, durability, and sustainability.

DATA COLLECTION

The collection and the analysis of data is an important element needed for developing meaningful conclusions and recommendations from a test project. Saeed & Hall (2003) developed a summary for APT facilities of some of the various types of data collected during their studies. While different than open-road test sections, many of the recommendations are relevant to pavement and material characterization data that are typically obtained during the design, construction, and monitoring of test pavements after construction and loading.



NCHRP
REPORT 512

NATIONAL
COOPERATIVE
HIGHWAY
RESEARCH
PROGRAM

Accelerated Pavement Testing:
Data Guidelines

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Link: https://www.k-state.edu/pavements/trb/A2B09/index_files/nchrp_rpt_512.pdf

9 Forensic Investigation

Despite the meticulous planning, design, and construction efforts of DOTs, unexpected performance issues or pavement failures can still occur during field performance evaluations. In such cases, it becomes imperative for DOTs to conduct forensic investigations to identify the causes behind these unpredictable performance results.

This section describes a procedure for conducting forensic investigations to determine the underlying factors contributing to unexpected performance or failures observed in test sections. The process is primarily based on the guidelines outlined in NCHRP Report 747 (Rada et al., 2013) and consists of several key steps. These steps include defining the unexpected performance result or problem, performing visual inspections, formulating hypotheses of what may have caused the expected performance, conducting tests to evaluate the hypotheses, analyzing the results, and developing conclusive findings.

The procedure outlined in this section is tailored for DOT personnel with expertise in pavement engineering and materials testing. Moreover, it can be customized to meet the unique needs and requirements of individual DOTs. By following this systematic and methodical approach to forensic investigation, DOTs can effectively identify root causes of unexpected performance, thereby gaining valuable insights to develop their BMD specifications and improve their practices.

Defining the Problem

Defining the problem is the first step in conducting a forensic investigation of test sections. It involves identifying specific instances of unexpected pavement performance that require investigation and determining the relevant information about the pavement necessary for further evaluation. Additionally, it entails defining the scope of the investigation, establishing clear objectives, and developing a plan for the investigation.

The plan will outline the specific inspections and tests to be employed to form and evaluate hypotheses, delineate the roles and responsibilities of the investigation team, and establish a timeline for the investigation. The investigation can be executed efficiently and effectively by constructing a well-defined plan.

The insights gained from the investigation results are instrumental in understanding unexpected pavement performance, thereby facilitating the development of improved BMD specifications to prevent future pavement failures.

Conducting Visual Inspections

The next step in the forensic investigation process is conducting thorough visual inspections. These inspections serve as a vital means of gathering essential information to form hypotheses regarding the causes of unexpected performance observed in test sections. Inspectors can gain valuable insights into the underlying factors contributing to the observed performance issues by carefully inspecting the pavement and identifying distresses.

Following a standardized classification and rating system for distresses is essential. Such a system allows for consistently identifying and assessing distress types and severity levels. ASTM D6433, *Standard Practice for Classification of Asphalt Pavement Distresses*, provides a comprehensive framework for characterizing and rating various types of distresses, including cracking, rutting, and other surface defects.

During visual inspections, it is important to identify the specific location and extent of distresses or failures within the pavement, which may involve noting areas of cracking, rutting, or other surface deformations. Mapping out the distribution and severity of distresses allows investigators to understand spatial patterns and potential correlations with underlying causes.

For example, a significant amount of cracking is observed in certain pavement sections. This observation suggests localized factors contributing to the distress, such as inadequate compaction during construction or insufficient bonding between layers. By documenting these observations and linking them to potential causes, inspectors can refine their hypotheses and guide subsequent testing and analysis.

In addition, maintaining detailed records of observed distresses and associated information is essential during visual inspections. These records should include relevant data such as distress type, location, and severity. Additionally, taking photographs or videos can visually document the pavement's condition.

These records ensure that all essential details are documented for further analysis. For instance, a visual inspection reveals a section of the pavement with extensive alligator cracking and associated moisture damage. The inspector documents the distress type, extent, location, and detailed notes about the surrounding environmental conditions and traffic. This comprehensive record becomes a valuable resource for later stages of the investigation, enabling a thorough understanding of the distress mechanisms and aiding in developing findings.

Formulating Possible Hypotheses

Formulating possible hypotheses is essential to conducting a forensic investigation of asphalt pavements. These hypotheses serve as potential explanations for the observed distresses or failures and guide further testing and analysis. They can be developed based on a combination of information about the pavement gathered in the prior step and the experience of the investigation team.

When formulating hypotheses, it is important to consider several possibilities causing the observed distresses. For example, when visual inspections reveal widespread fatigue cracking on the pavement surface, possible hypotheses can include factors such as inadequate thickness design, high traffic loading, poor asphalt mixture quality, inadequate asphalt binder content, or other construction-related issues. By exploring different potential causes, investigators can ensure a comprehensive investigation and increase the likelihood of identifying the underlying factors.

However, as testing and analysis progress, the hypotheses should be refined based on the obtained results. The findings from laboratory and field testing can provide valuable insights that help eliminate less likely hypotheses and strengthen the ones that remain viable. This iterative refinement process allows the investigation team to focus on the most promising hypotheses.

For instance, laboratory testing of asphalt samples reveals inadequate asphalt binder content in the distressed sections. This information strengthens the hypothesis that poor asphalt mixture quality contributes to the observed distress. Consequently, the investigation team may adjust their focus on conducting further testing to gather additional evidence related to asphalt mixture properties and reviewing acceptance data collected during construction.

Conducting Tests for Evaluating the Hypotheses

Conducting tests to evaluate the hypotheses is essential to the forensic investigation of asphalt pavements. These tests, which can include non-destructive testing and destructive testing in the laboratory or field settings, are important for assessing the properties of the pavement materials and structure to help identify the underlying causes of the distresses or failures.

Non-destructive testing techniques, such as ground-penetrating radar (GPR), falling weight deflectometer (FWD), or pavement surface profiling, provide valuable insights into the condition and characteristics of the pavement without causing any damage. For example, GPR can help assess subsurface layer thicknesses and identify potential voids or moisture-related issues, while FWD can determine the structural capacity and stiffness of the pavement layers.

Destructive testing involves extracting pavement samples for laboratory testing, allowing for a more detailed examination of the materials. Cores can be obtained and subjected to various laboratory tests, such as asphalt binder extraction and characterization and testing of field cores. These tests provide quantitative data on material properties, composition, and performance characteristics. For example, laboratory tests on extracted asphalt samples may reveal inadequate binder content, low ductility values, or high susceptibility to aging, suggesting a potential contribution of binder-related issues to the observed distresses. The data collected from these tests will be carefully organized for further analysis in the next step.

Analyzing the Results

The data collected from testing should be carefully analyzed. Statistical techniques, data visualization, and expert interpretation should be employed to identify patterns, correlations, and causal relationships. The analysis methods, assumptions, and limitations should be carefully assessed. In addition, peer review and consultation with experts in the field can add valuable insights and contribute to the credibility of the investigation's results. By thoroughly examining the data, the investigation team can uncover critical insights and determine the root causes of the observed distresses or failures.

Developing Conclusions

Identifying the root cause of the distresses or failures is the primary objective in developing conclusions. This process involves carefully analyzing the results obtained from testing and analysis to determine the underlying factors contributing to the observed pavement performance issues. By considering the evidence, investigators can pinpoint the specific root cause or causes. For example, after extensive testing and analysis, it may be determined that the distresses observed on the pavement are primarily caused by poor compaction during construction, leading to inadequate density and weakened structural integrity. This conclusion would be supported by laboratory tests on extracted cores, field observations, and other relevant data.

Finally, a clear and concise report should be prepared to document the investigation’s findings. The report should include detailed descriptions of the distresses or failures, the testing methods employed, the results of testing and analysis, and the conclusions. Where appropriate, the report should include visual aids accompanied by clear explanations to ensure the information is easily understandable and accessible to all relevant stakeholders.

This section discusses a procedure for forensic investigations to identify the root causes of unexpected asphalt pavement performance. DOTs can customize the procedure to meet their unique needs and requirements. The results can enable a thorough understanding of the distress mechanisms and provide valuable insights to develop their BMD specifications and improve their practices.

10 Data Analysis and Application of the Results in Specifications

The primary goal of a field validation experiment is to establish a correlation between field performance and the laboratory test results, and if that relationship is strong and covers a reasonably wide range of field performance, then that relationship can be used to set criteria for the laboratory test(s). Without a field validation experiment to help set appropriate criteria for use in specifications, there are risks associated with setting unreasonable limits that are too high or too low.

If criteria are set too high, it may be very challenging for contractors to meet the specification, possibly causing mix designers to use unusual materials or limiting the number of contractors to bid on the work. These risks could increase the costs of paving projects, limit the use of sustainable materials, and limit competitiveness among contractors. On the other hand, setting criteria too low could result in poor-performing pavements, increasing the agency’s pavement management costs, and increasing user costs due to higher vehicle operating costs.

As noted in section 8, the field performance (e.g., rut depths, feet of thermal cracking, percent of lane area cracking, mean texture depth, friction, etc.) of the test sections should be monitored on a regular frequency so that time series plots can be constructed for the distress as they progress over time. In some cases, a few test sections, upon reaching certain distress limits, may have to be rehabilitated to avoid further damage to the pavement structural and reduce risks to the traveling public. The time series plots will allow fitting a distress progression model that may be useful for extrapolating the trend to a common point in the life of the experiment to compare different test sections. This approach was used in the NCAT Cracking Group Experiment, as illustrated in Figure 14, and explained in NCAT Report 21-03.

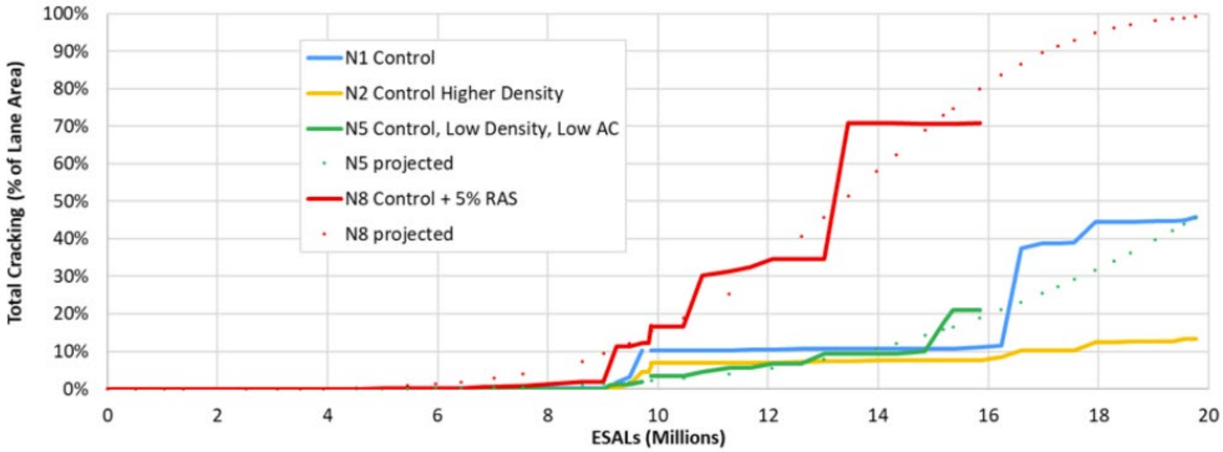


Figure 14. Time/Traffic Series Plot of Cracking from the 2015-2021 NCAT Cracking Group Experiment.

Figure 15 shows a good example of a strong lab-to-field validation correlation from the Pooled-Fund Low-Temperature Cracking study (Marasteanu et al., 2012). This is referred to as a scatterplot, which graphically represents the relationship between two quantitative variables measured for the same items. This example presents DCT fracture energy versus measured thermal cracking (m/500 m) from eight MnROAD test sections and two sections from Wisconsin. The age of the pavements ranged from 5 to 15 years. This correlation plot was the basis for setting the recommended DCT criterion of 400 J/m^2 for long-term aged specimens for projects of moderate risk. Later it was recommended to adjust the criterion to 460 J/m^2 for short-term aged specimens.

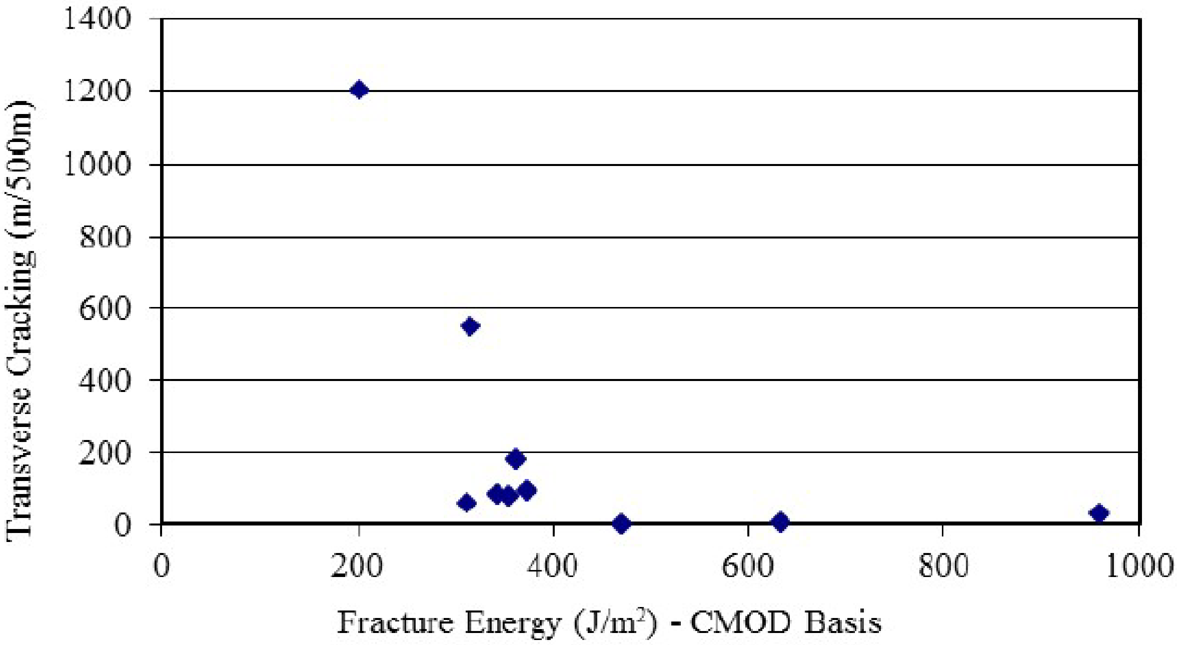


Figure 15. Relationship Between DCT Fracture Energy (at PG_{low} + 10°C) versus Field Measured Transverse Cracking.

Buttlar et al. (2019) later pointed out that the age of the sections was not considered in Figure 16. Given that an asphalt pavement becomes more susceptible to thermal cracking as it ages due to binder oxidation which results in a loss of its ability to relax thermal stresses. To address this issue, Buttlar et al. added another feature of the correlation by using spheres of different sizes to represent the age of the test section, as shown in Figure 16. The updated correlation plot also includes many more sites and corresponding DCT results compared to Figure 15.

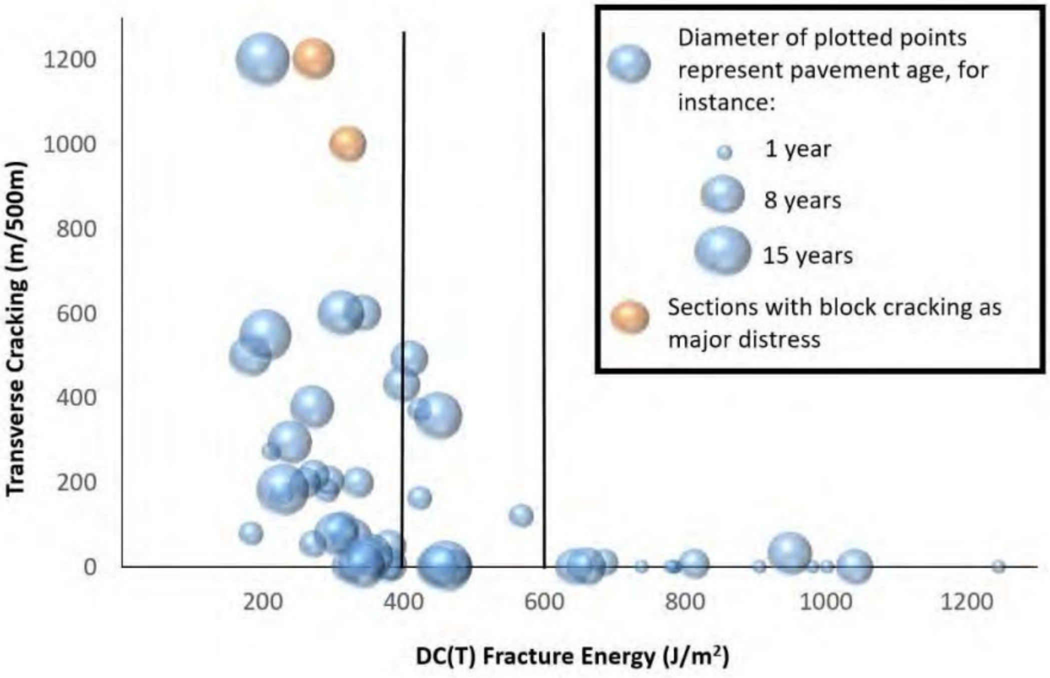


Figure 16. Expanded and Enhanced Correlation Between DCT Fracture Energy (at PG_{low} + 10°C) Versus Field Measured Transverse Cracking.

Constructing a scatterplot is a simple process in Microsoft Excel. The lab test results are presented in one column and the corresponding field performance results at a particular time are in another column. Here is a link to a 4.5-minute video on YouTube to explain the process:

https://www.youtube.com/watch?v=Kk5GG6zi46Q

Figure 17 shows an example correlation scatterplot between the N_{flex} Factor cracking test and passes of the FHWA Accelerated Loading Facility (ALF) to reach 240 inches of fatigue cracking. The plot shows a best-fit regression equation from Excel. This example shows a linear equation that provided the least-squares best fit to the data and the R^2 value, also known as the coefficient of determination. Linear relationships with a higher R^2 indicate that the regression equation has a better fit to the data. More information on linear regressions and R^2 can be found in this video:

https://www.youtube.com/watch?v=zPG4NjIkCjc

The R^2 can be interpreted as the percentage of the regression model explained or caused by the dependent variable. So, for Figure 17, we can say that 59% of the variation in the adjusted ALF passes to 240 inches of cracking can be explained by differences in N_{flex} Factor for the test sections. That means that 41% of the variation in the field cracking data is due to other factors or errors.

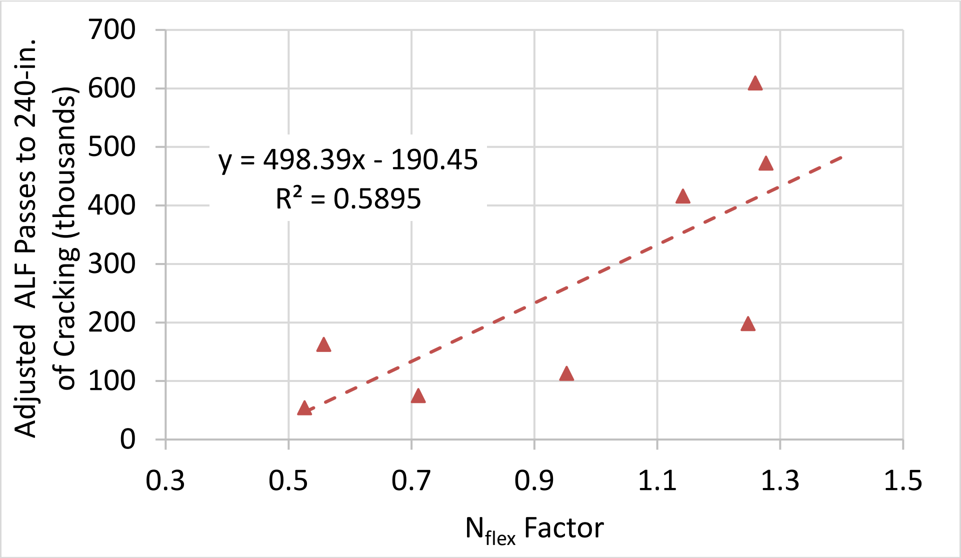


Figure 17. Scatterplot of N_{flex} Factor and ALF Passes to 240 inches of Fatigue Cracking with a Best-Fit Linear Regression and its Coefficient of Determination, R^2.

There is often confusion about other forms of regression equations, such as polynomials, exponential, power functions, etc., and the appropriateness of the R^2 values that Excel calculates for these forms of regression equations. Many statistical guides warn that R^2 is not an appropriate "goodness-of-fit" indicator for non-linear regressions. However, the confusion lies in the definitions of linear and non-linear. Linear models do not necessarily mean the equation is a straight line. In statistics, the term 'linear' in linear regression refers to the regression model being linear *"in the parameters." (Frost, 2019)*

> *"In the parameter"* refers to the relationship of variables in a mathematical model where the unknowns or coefficients appear as multipliers (not exponents or inside functions). In simpler terms, even if the equation might look curved or complicated when graphed, as long as the unknown variables or coefficients are not raised to powers or found inside functions (like sin or cos), it's considered linear "in the parameters."

Therefore, the other forms of regression equations provided by Excel are linear in the parameters, as shown in Figures 18 and 19, and provide slightly higher R^2 values indicating that they are slightly better regression models for relating the independent variable (N_{flex} Factor) and dependent variable (ALF passes to 240-inches of fatigue cracking).

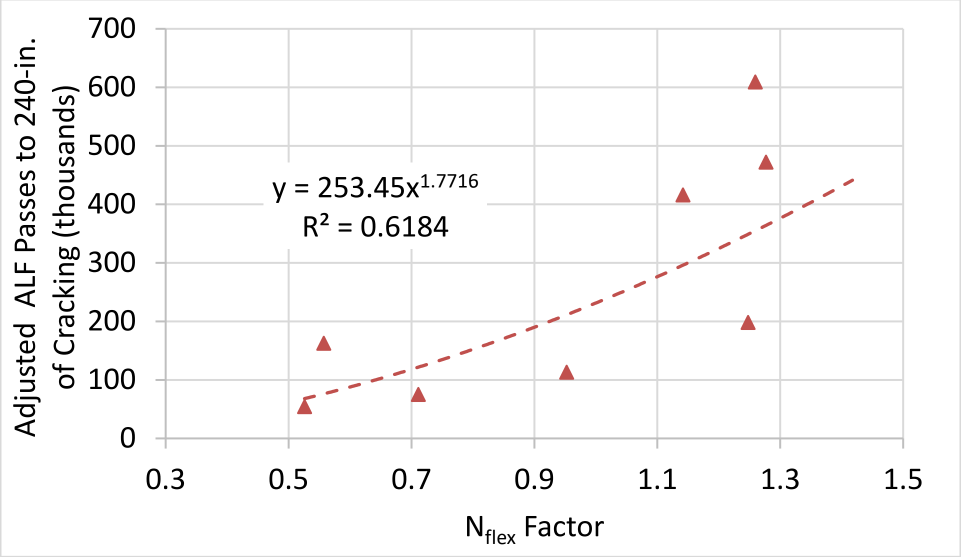


Figure 18. Scatterplot of N_{flex} Factor and ALF Passes to 240 inches of Fatigue Cracking with a Best-Fit Power Function Regression and its Coefficient of Determination, R^2.

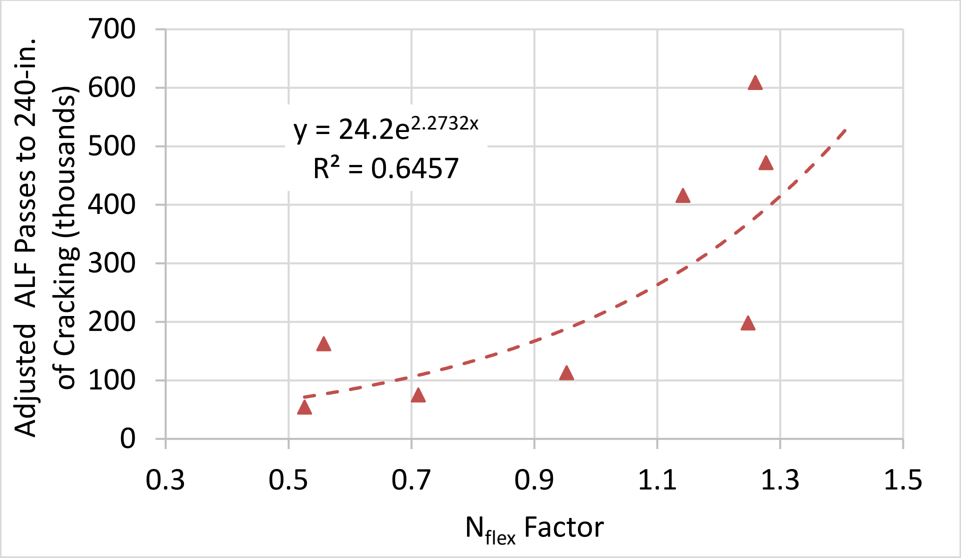


Figure 19. Scatterplot of N_{flex} Factor and ALF Passes to 240 inches of Fatigue Cracking with a Best-Fit Exponential Regression and its Coefficient of Determination, R^2.

A good explanation of R^2 and its limitations is provided in this article, entitle “*How To Interpret R-squared in Regression Analysis*.”

https://statisticsbyjim.com/regression/interpret-r-squared-regression/

One limitation of using R^2 as the only goodness of fit indicator is that it can be influenced by one or two data points that are separated from the rest of the data. Figure 20 shows an example where the correlation includes one point in the upper left part of the graph that is separated from the rest of the data. Without that point, the relationship between the dependent and independent variables would have been much weaker. This is why the field validation experiment should be designed to cover an evenly distributed range of laboratory test results and field performance data, as previously discussed in Chapter 3.

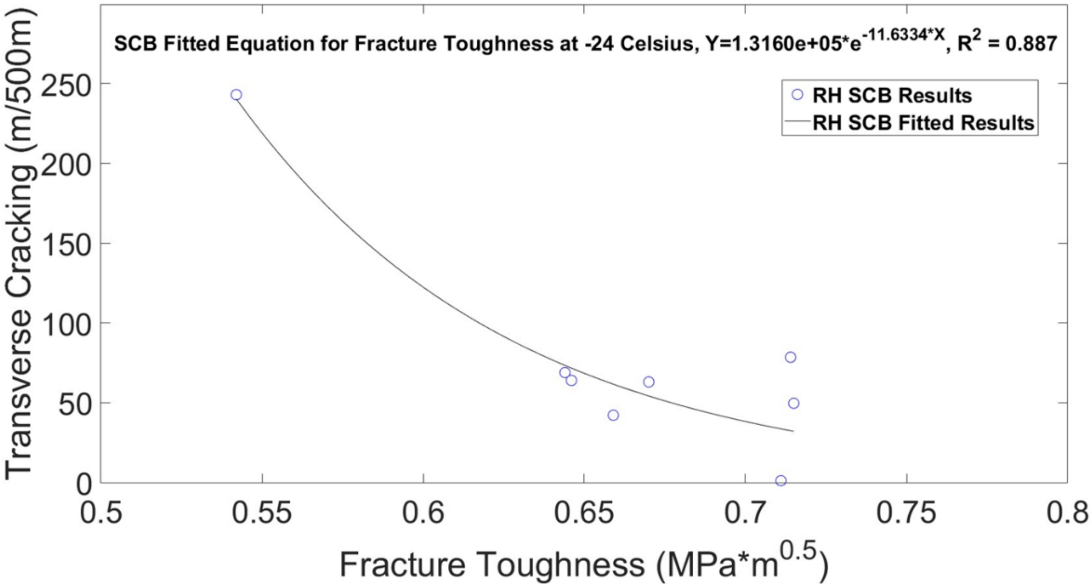


Figure 20. Correlation of Transverse (Thermal) Cracking versus Fracture Toughness from the Low-Temperature Semi-Circular Bend Test.

For regression analyses, most statistical references recommend analyzing the residuals. A residual is the difference between a variable's observed value and its predicted value from the regression equation. In other words, a residual quantifies how far away a point is from the regression line. The analysis of residuals is performed by plotting the independent values on the X-axis and the corresponding residual values on the Y-axis. If the residual plot reveals a pattern where the residual values are above or below the line for part of the plot, it indicates the presence of a bias in that region of the regression model.

Besides having a strong lab-to-field relationship, it is also important for the test to be able to statistically discern between good, fair, and poor field performance. In other words, it is desirable for mixes with different field performance to have statistically different test results. Figure 21 shows an example chart with statistical groupings of the Texas Overlay Test results from the NCAT Cracking Group Experiment (West et al., 2021). In this chart, the letters A through E represent statistical groupings. Results with the same letter are not statistically different from one another. This analysis considers the testing variability of the results. Test methods with relatively large variabilities will not be able to provide suitably discernable results between mixes and are not recommended for use in specifications. As shown in Figure 21, the test results are able to statistically discern the mixtures with different top-down cracking performance in the field, as indicated by five grouping letters (A through E) across the mixtures.

The grouping analysis is conducted in two steps. The first step is a one-way analysis of variance (ANOVA) which is a statistical method for testing differences among means of three or more groups when there is a single independent variable. In the case of a field validation experiment, the single independent variable is the mixtures used in the test sections, so the ANOVA will determine if any of the mixtures have test results that are statistically different from any other mixtures.

The second step is to determine which mixture or groups of mixtures have similar results and which mixtures are statistically different. There are several different post hoc tests that are applicable to different circumstances. Commonly used post-hoc tests in engineering materials are Tukey's honest significant difference (HSD) test, the Tukey-Kramer test, and the Games-Howell test. The Tukey-Kramer test is an adaptation of the Tukey HSD test when the sample sizes of the groups are not equal. The Tukey-Kramer test and the Tukey test will yield the same results when the sample sizes of the groups are equal. The Games-Howell test is a pairwise comparison test that is appropriate when the variances of the groups are unequal.

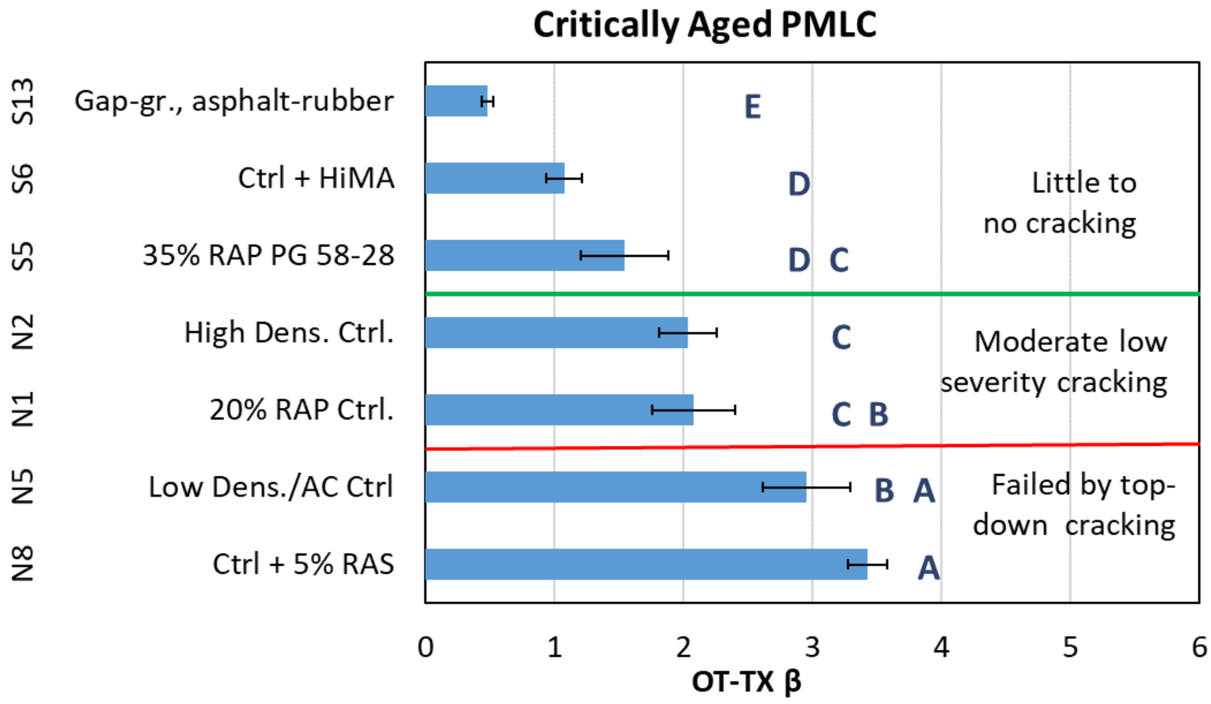


Figure 21. Chart Showing Statistical Groupings of Overlay Test Results from the NCAT Cracking Group Experiment.

Example of Field Validation Experimental Matrix with Eight (8) Test Sections

In 2013, the Federal Highway Administration (FHWA) pavement test facility constructed a Sustainability Study and loaded it under their Accelerated Load Frames (ALF). Figure 22 summarizes the cracking performance versus one of the proposed BMD cracking tests, the IDEAL-CT (ASTM D8225).

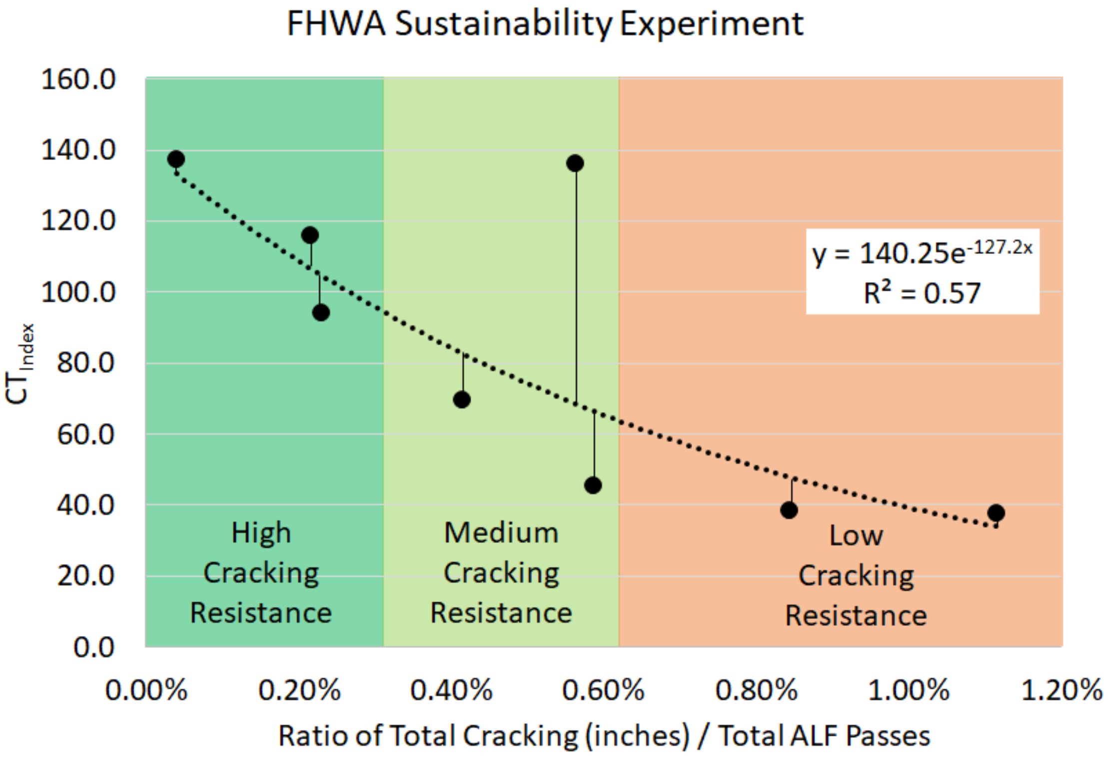


Figure 22. Example of Laboratory-to-field Correlation based on the FHWA ALF Sustainability Study

Note: The high-medium-low cracking resistance ranges are for illustrative purposes only.

In general, an R^2 value of 0.60 or higher is considered satisfactory for utilizing the correlation to set the test criteria. In addition, the standard error of the regression, often called the Residual Standard Error (RSE), should be included in the analysis. The RSE gives an estimate of the standard deviation of the residuals, which are the differences between the observed and predicted values. A smaller RSE indicates a better fit of the regression model to the data, as the observed values are closer to the predicted values.

Equation 1. Residual Sum of Squares, RSS.

$$\text{RSS} = \sum_{i=1}^{n}\left(\text{Y}_i - \hat{\text{Y}}_i\right)^2$$

Where:

- n is the number of observations
- Y_i is the actual value of the dependent variable for the i^{th} observation, e.g., Average CT_{Index} measured in test section 2 ($i = 2$)
- $\hat{\text{Y}}_i$ is the predicted value of the dependent variable for the i^{th} observation, e.g., $\text{CT}_{\text{Index}} = \alpha\, e^{\beta x}$, where α and β are regression coefficients from your field test sections and x is measured cracking per 10,000 ESALs.

Equation 2. Residual Stand Error, RSE.

$$\text{RSE} = \sqrt{\frac{RSS}{(n-p)}}$$

Where p is the number of predictors (independent variables) in your model, not counting the intercept. E.g., for the CT_{Index} model above there is only one independent variable (x), such that $p = 1$.

The specific threshold for correlation may vary depending on factors such as the number of data points, the range of field performance, and the range of laboratory test results analyzed.

By carefully considering the number of test sections, balancing cost, and ensuring diverse field performance representation, the field validation effort can provide meaningful correlations between laboratory test results and field performance for establishing effective test criteria.

Assessment of model relating CT_{Index} to cracking:

- The R^2 of the model is close to 0.60, as recommended, however:
- The RSE = 28.0. This indicates that 95% of the data is within ±2 RSE or ±56.0, this appears to be too wide of a band to set criteria.

You may note that one of data points appears to be significantly away from the regression line (x=0.56, y=136.0). There are several potential or combination of reasons for this point have a high residual (distance from the regression model), which include: a) variable subgrade support under the ALF sections, b) age of section at time of loading, c) sampling bias, and d) relationship between CT_{Index} and measured performance. For illustrative purposes, let's assume we determine this data point to be suspect and remove it from the analysis, as such:

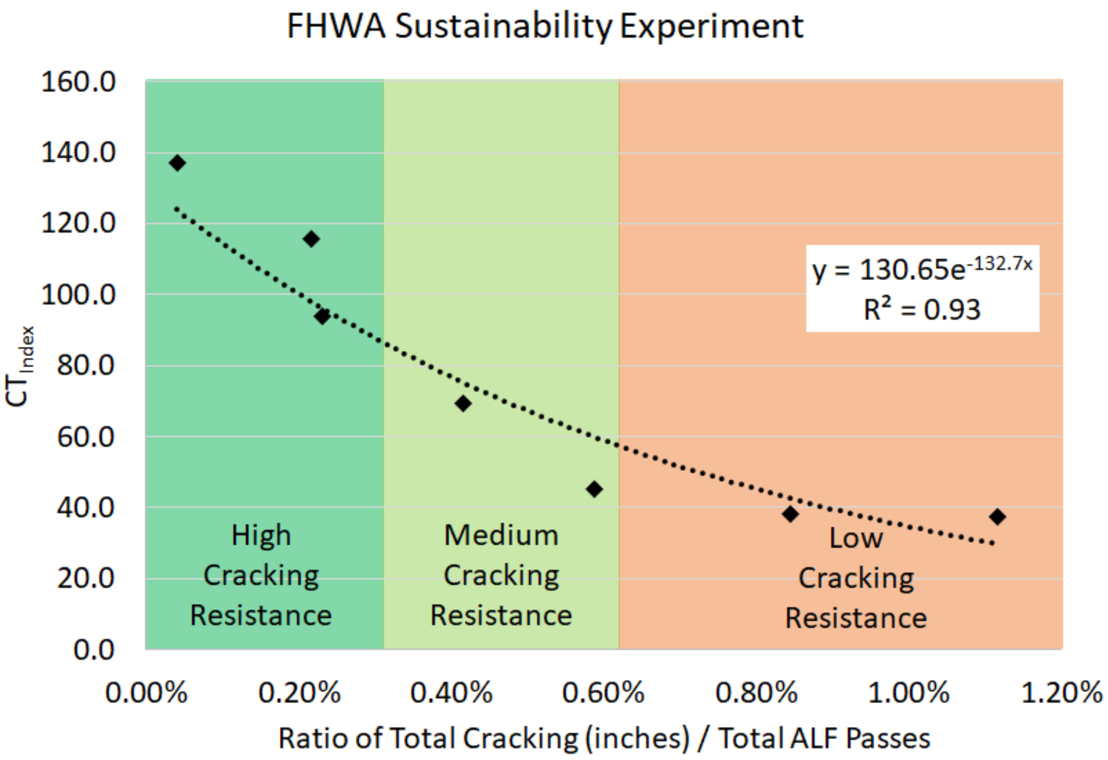


Figure 23. Example of Laboratory-to-field Correlation based on the FHWA ALF Sustainability Study with potential outlier removed.

- Then the predicted model becomes:

$$y = 130.65\ e^{-132.7x}$$

Where: $n = 7$, $R^2 = 0.93$, and RSE = 11.7. This appears to be a good model.

There is no defined value for RSE that determines if a model is good or poor. However, RSE provides a value directly related to the BMD test result that can be used to assess risk.

Agencies will have to set their criteria based on risk tolerance and roadway application. In this study a wide range of IDEA-CT results were obtained with a similar wide range of field performance.

Overview of the ALF Sustainability study:
https://highways.dot.gov/research/laboratories/pavement-testing-laboratory/pavement-testing-facility-overview

11 Establishing Interim Criteria

Establishing interim minimum test scores for specifications is crucial for ensuring that agencies can continue their construction projects while awaiting the completion of field trials. These interim criteria should be carefully developed to maintain a certain level of quality in pavement construction. Here are some considerations for this section:

1. **Benchmark Testing:** Recommend agencies to conduct benchmark testing on submitted mix designs or plant-produced mixes to establish interim specification criteria. Benchmark testing involves testing representative samples of the proposed mixes to evaluate their performance against predetermined performance thresholds. This can be based on historical data or performance data from similar projects.
2. **Shadow Projects:** Discuss the potential use of shadow projects to generate interim specification criteria. Shadow projects involve constructing small test sections alongside ongoing construction projects using different mixtures or construction techniques. Performance data from these shadow projects can help establish provisional criteria based on observed performance trends.
3. **Data Analysis:** Emphasize the importance of proper data analysis in establishing interim criteria. Statistical analysis, such as comparing mean values or percentile values of performance indicators, can provide a more objective and comprehensive understanding of the mixture performance. See section 10 for R^2 *and its limitations are provided in this article link*.
4. **Consistency:** Discuss the need for consistency in data collection and testing procedures to ensure the reliability of the interim specification criteria. Consistent and standardized testing protocols should be followed to eliminate potential bias or variations in test results.
5. **Risk Assessment:** Consider conducting a risk assessment to evaluate the potential consequences of using interim criteria. This assessment can help agencies identify potential risks associated with the proposed criteria and make informed decisions.
6. **Adaptability:** Suggest that the interim specification criteria be adaptable and subject to revision based on the results of the ongoing field trials. As more data becomes available from the field trials, the interim criteria can be refined and adjusted accordingly.
7. **Communication with Contractors:** Encourage communication and collaboration with contractors during the establishment of interim criteria. Contractors can provide valuable input and feedback based on their experiences, which can enhance the effectiveness of the interim criteria.

8. **Documentation:** Stress the importance of documenting the process of developing interim criteria, including the rationale behind the chosen performance thresholds and the methodologies used for data analysis. This documentation will be useful for transparency and future reference.
9. **Sharing:** Sharing your approaches regionally and nationally, as well as lessons learned is an effective catalyst in accelerating the adoption and ultimately the benefit of balanced mix design.

Recommendation: Developing interim minimum test scores for specifications requires a careful and data-driven approach. Agencies should aim for a performance-based approach and consider benchmark testing, shadow projects, and proper data analysis. Consistency, adaptability, and communication with stakeholders are key to successfully implementing interim criteria until field trials can be completed, and more comprehensive specifications can be established.

A Journey to Performance – this is an idealized scenario of BMD implementation at a "Typical State DOT."

Sandy, the State DOT Bituminous Engineer, has taken on the challenge to implement BMD to address performance issues and provide a sustainable pathway forward. Based on last year's State PMS, Sandy's state is responsible for 7,350 centerline miles of pavement, which totals just over 18,000 lane miles. Ninety-four percent of the state's roadways are paved with asphalt, with an additional 3% concrete and 3% composite (asphalt over concrete). The PMS uses the composite Pavement Condition Index (PCI) to rate the network. Table 8 shows last year's rating:

Table 8. State DOT PMS Pavement Condition Index Ratings.

PCI Score	Condition	Interstate	State Route	Region/District	Low-Volume
96 – 100	Very Good	13%	13%	5%	2%
76 – 95	Good	53%	44%	50%	59%
46 – 75	Fair	32%	31%	28%	27%
21 – 45	Poor	2%	12%	16%	9%
0 – 20	Very Poor	0%	0%	1%	3%

Overall, 58% of all pavements rate as Good or better, with 30% rating as Fair, and the remaining 12% rating as Poor to Very Poor. Over the last decade, pavement performance has been trending downward. Ten years ago, 66% of all pavements rated as Good or better, with 21% rating as Fair, and 13% rating as Poor to Very Poor.

Pavement design is based on a life-cycle cost analysis period of 40 years, which includes remaining service life (RSL), salvage value, and any removal costs. Asphalt surface layers have a design life of 20 years, with a service life typically ranging from 15 to 22 years. On the extremes, service life varies from 10 to more than 30 years.

PCI calculations includes indexes for: rutting (RUT), fatigue cracking (FAT), transverse cracking, raveling (RAV), patching, and loading. Table 11.2 shows the individual scores for three of the PCI indexes.

Table 9. State DOT PCI Summary of Key Composite Indexes.

PCI Indexes	Statewide Average	Minimum Value	
RUT	91.1	52	Rutting Resistance
FAT	73.7	40	Fatigue Cracking Resistance
RAV	92.7	72	Related to Moisture Susceptibility

The $150 million dollar state annual paving program includes 10% reconstruction, 41% asphalt overlays, and 49% pavement preservation (typically chip seals). Three Superpave binder grades are specified in the state: PG 52-28, PG 64-22, and PG 76-22. Table 11.3 provides a summary of the Superpave mix design parameters used for surface mixes.

Table 10. Typical Overlay/Surface Mixes Superpave Mix Design Requirements.

Traffic	NMAS	Gradation	N_{design}	VMA	VFA	$P_{0.075}/P_{be}$	Allowable RAP
Low	9.5mm	Fine	50	15.0	70 to 80	0.6 to 1.2	25 to 40%
Medium	12.5mm	Fine	75	14.0	65 to 78	0.6 to 1.2	20 to 30%
High	12.5mm	Coarse	100	14.0	65 to 75	0.8 to 1.6	15 to 25%

The breakdown of last year's surface mixes by traffic level were:

- 10% Low
- 60% Medium
- 30% High

Sandy's review of the information, along with conversations with the contractor community, provides the following insights:

- Lower PCI's are being driven by fatigue cracking.
- The state does not have a rutting issue.
- The majority of the paving program uses 12.5mm fine-graded mixes.
- Contractors typically design mixes on the lower allowable RAP range, citing challenges meeting all the Superpave volumetric criteria.
- The State DOT would like to increase the RAP content for a more sustainable product.
- The Contractors are also interested in higher-RAP as they explore developing environmental product declarations (EPD).
- Sandy is developing a BMD field validation experiment to establish criteria.

Laboratory Trials – During the past two years, Sandy has sampled 35 production surface mixes and performed BMD testing. In reviewing the NCHRP 10-107 report entitled, "*Guide for Implementing Balanced Mix Design Specification*," by NCAT, Sandy selected one rutting and two cracking tests:

- Hamburg Wheel Tracking Test.
- Indirect Tensile Asphalt Cracking Test, and
- Disk-Shaped Compact Tension Test.

Sandy likes that the HWTT provides both an assessment of rutting resistance (rut depth at 10,000 passes) and moisture susceptibility (stripping inflection point, SIP). Sandy's laboratory technicians provided feedback on the specimen prep time for the IDEAL-CT and DCT, preferring the IDEAL-CT. Table 11.4 provides a summary of the testing results:

Table 11. Summary of Benchmark Testing for High- and Medium-Traffic Surface Mixes.

Traffic	Parameter	HWTT-SIP	HWTT Rut Depth 10k passes (mm)	IDEAL-CT (CT_{Index})	DCT Fracture Energy (J/M^2)
Medium	Mixes, *n*	22			
	Average, $\bar{Y}$	13,700	5	66.5	481.3
	COV	23%	19%	18%	22%
High	Mixes, *n*	13			
	Average, $\bar{Y}$	16,200	4	59.5	422.7
	COV	15%	17%	19%	21%

COV is based on the average COV for each set of replicates tested.

Sandy uses Table 5, "*Example Field Validation Experimental Matrix with Six Test Sections*," to layout the design of the field validation sections. Sandy establishes a joint Agency-Industry Taskforce to meet and vet the proposed layout. The Taskforce encourages the use of BMD Approach D, which replaces the Superpave volumetric requirement with BMD testing; to explore higher RAP mix designs.

REFERENCES

1. Simpson, A. L., Ostrom, B. K., & Schmalzer, P. N. (2006). Guidelines for the collection of long-term pavement performance data (No. FHWA-HRT-06-067). United States. Federal Highway Administration.

2. Rada, G. R., Jones, D., Harvey, J., Senn, K., & Thomas, M. (2013). Guide for Conducting Forensic Investigations of Highway Pavements (with supplemental material on CD-ROM) (Vol. 747). Transportation Research Board.

3. Brown, E. R., and R. B. Powell. A General Overview of Research Efforts at the NCAT Pavement Test Track. Presented at 2nd International Symposium on Maintenance and Rehabilitation of Pavements and Technological Control, Auburn, AL, 2001.

4. Choubane, B., and Greene, J., Accelerated Pavement Testing: Celebrating over 100 Years of Innovation and Economic Benefits, TRB Centennial Paper, 2020.

5. Coetzee, N. F., W. Nokes, C. Monismith, J. Metcalf, and J. Mahoney. Full-Scale/Accelerated Pavement Testing: Current Status and Future Directions. In Transportation in the New Millennium, Transportation Research Board, Washington, D.C., 2000.

6. FHWA. The Long-Term Pavement Performance Program. FHWA Publication No. FHWA-HRT-15-049. Federal Highway Administration, Washington, DC, September 2015.

7. Federal Highway Administration, Turner-Fairbank Highway Research Center. Pavement Testing Facility Overview. https://highways.dot.gov/research/laboratories/pavement-testing-laboratory/pavement-testing-facility-overview

8. Harvey, J., Roesler, J., Coetzee, N., and Monismith, C. L. Caltrans Accelerated Pavement Test (CAL/APT) Program Summary Report, Six-Year Period: 1994 – 2000. Report No. FHWA/CA/RM-2000/15, Pavement Research Center, Institute of Transportation Studies, University of California, Berkeley. June 2000.

9. Hudson, W. R., C. L. Monismith, J. F. Shook, F. N. Finn, and E. L. Skok, Jr. AASHO Road Test Effect on Pavement Design and Evaluation after 50 Years. Transportation Research Circular E-C118: Pavement Lessons Learned from the AASHO Road Test and Performance Section on Pavements (AFD00) Standing Committee on Full-Scale Accelerated Pavement Testing (AFD40) of the Interstate Highway System, Transportation Research Board, Washington, D.C., 2017.

10. Kelsey, C. "The Benefits of Accelerated Pavement Testing on Infrastructure", *GeoSynthetica*, December 9, 2019, https://www.geosynthetica.com/accelerated-pavement-testing-benefits/. Accessed July 17, 2023.

11. Long-Term Pavement Performance. U.S. Department of Transportation, Federal Highway Administration, Turner-Fairbank Highway Research Center. https://highways.dot.gov/research/long-term-infrastructure-performance/ltpp/long-term-pavement-performance

12. McKendrick, W. B., and W. N. Carey. Status of Large-Scale Road Tests. AASHO Road Test, 1957. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?referer=http://scholar.google.com/&httpsredi

13. Metcalf, J. B. NCHRP Synthesis 235: Application of Full-Scale Accelerated Pavement Testing. Transportation Research Board, Washington, D.C., 1996.

14. Metcalf, J., Romanoschi, S., Yongqi, L., and Rasoulian, M., Construction and Comparison of Louisiana's Conventional and Alternative Base Courses under Accelerated Loading. Report No. FHWA/LA.88/312 Institute of Recyclable Materials, Louisiana State University, Baton Rouge, LA August 1998.

15. Mitchell, T. "WesTrack: The Road to Solutions." Public Roads 60.2 1996.

16. National Research Council, Highway Research Board. The WASHO Road Test Part 1: Design, Construction, and Testing Procedures. Special Report 18, Publication 310, 1955. http://onlinepubs.trb.org/Onlinepubs/sr/sr18.pdf

17. National Research Council, Highway Research Board. The WASHO Road Test Part 2: Test Data, Analyses, Findings. Special Report 22, Publication 360, 1955. http://onlinepubs.trb.org/Onlinepubs/sr/sr18.pdf

18. National Research Council, Highway Research Board. The AASHO Road Test: History and Description of Project. National Academy of Sciences-National Research Council, Washington, D.C. 1961.

19. National Research Council, Highway Research Board. The AASHO Road Test: History and Description of Project. National Academy of Sciences-National Research Council, Washington, D.C. 1962.

20. Older, C. Bates Experimental Road or Highway Research in Illinois. Issue 21 of Bulletin. State of Illinois, Department of Public Works and Buildings, Division of Highways, 1924.

21. Puccinelli, J., Schmalzer, P., Senn, K., and McDonald, L. Long-Term Pavement Performance Warm-Mix Asphalt Study Final Report, Volume Ⅱ: SPS-10 Experimental Matrix and Research Plan. FHWA Publication No. FHWA-HRT-22-019. Office of Infrastructure Research and Development, Federal Highway Administration, McLean, VA, 2023.

22. Saeed, A., Hall Jr, J. W., *NCHRP Report 512, Accelerated Pavement Testing: Data Guidelines.* Transportation Research Board, Washington, D.C., 2003.

23. Vargas, A., *Pavement Preservation Group Study - Summary of Findings,* 2019. https://eng.auburn.edu/research/centers/ncat/files/pgsummary2019.pdf

24. Vargas-Nordcbeck, A., Vrtis, M. C., & Worel, B. (2019). Performance of Chip Seal Treatments in Two Different Climatic Regions. Transportation Research Record, 2673(12), 728–736. https://doi.org/10.1177/0361198119853546

25. White, T. D., J. M. Albers, and J. E. Haddock. Limiting Design Parameters for Accelerated Pavement Testing System. Journal of Transportation Engineering, 1992.

26. Wynand, S. Synthesis of Highway Practice 433: Significant Findings from Full-Scale Accelerated Pavement Testing. Transportation Research Board of the National Academies, Washington D.C., 2012.

27. Allen, H., 1985, Methods and Materials for Reducing Crack Reflectance, Office of Research and Development, Minnesota Department of Transportation, FHWA/MN/RD-84/09.

28. Aschenbrener, T. 1995. "Evaluation of Hamburg Wheel-Tracking Device to Predict Moisture Damage in Hot Mix Asphalt." Transportation Research Record: Journal of the Transportation Research Board, No. 1492, Washington, DC, pp. 193–201.

29. Dave, E.H, J.E. Sias, and K.E. Hazlett, 2021, Developing Best Practices for Rehabilitation of Concrete with Hot Mix Asphalt (HMA) Overlays Related to Density and Reflective Cracking, Minnesota Department of Transportation, Report No. NRRA202109.

30. Gu, F., R. West, B. Bowers, and R. Willis, 2020, Influences of Slab Fracturing Techniques on Long-Term Performance of Rehabilitated PCC Pavements. International Journal of Pavement Engineering.

31. Schram, S. and R.C. Williams, 2012, Ranking of HMA Moisture Sensitivity Tests in Iowa, Iowa Department of Transportation, Report No. RB00-012.

32. Vrtis, M., C. Rodezno, R. West, J. Podolsky, J. Calvert, and D. Van Deusen, 2023, MnROAD Cracking Group Experiment: Validation of Low-Temperature Cracking Tests for Balanced Mix Design, NCAT Report 23-03.

33. West, R., C. Rodezno, G. Julian, B. Prowell, R. Frank, L. Osborne, and T. Kriech (2014) Field Performance of Warm Mix Asphalt Technologies, NCHRP Report 779, Transportation Research Board, National Cooperative Highway Research Program.

34. Buttlar, W., P. Rath, H. Majidifard, E. Dave, and H. Wang, 2019, Relating DC(T) Fracture Energy to Field Cracking Observations and Recommended Specification Thresholds for Performance-Engineered Mix Design, Relationship Between Laboratory Cracking Tests and Field Performance of Asphalt Mixtures. Transportation Research Circular Number E-C251.

35. Marasteanu, M., W. Buttlar, H. Bahia, R. Williams, et al., 2012, Investigation of Low-Temperature Cracking in Asphalt Pavements, National Pooled Fund Study -Phase II, Minnesota Department of Transportation, Report No. MN/RC 2012-23.

36. West, R., D. Timm, B. Powell, N. Tran, F. Yin, B. Bowers, C. Rodezno, F. Leiva, A. Vargas, F. Gu, R. Moraes, and M. Nakhaei, 2021, Phase VII (2018-2021) NCAT Test Track Findings, NCAT Report 21-03.

37. Frost, J., (2019) Regression Analysis an Intuitive Guide for Using and Interpreting Linear Models, Statistics by Jim Publishing, State College, PA.

APPENDIX

A1 Literature Review: Full-scale Road Test Sections

FULL-SCALE ROAD TEST SECTIONS

Bates Road (Illinois) – 1920 - 1923

The Bates Road project, constructed between 1920 and 1923 by the State of Illinois Division of Highways, marked one of the early endeavors in full-scale road test sections for asphalt pavements. The primary aim of this project was to gain a deeper understanding of the unresolved challenges in rural pavement design during that era. Various aspects related to pavement performance were investigated, including subgrade soil drainage, the impact of repeated bearing pressures on soils, the effects of temperature changes on pavement surfaces, the distribution of wheel loads on pavement stresses, and the influence of moving wheel loads.

The test road, located near Bates, Illinois (southwest of Springfield) and stretching east to Farmingdale Road on what is now the old US-54 (Older, 1924), comprised six groups of test sections, each incorporating different pavement types and configurations. The pavement types tested included:

1) Vitrified brick surfacing with bituminous joint filler on a macadam base.
2) Asphalt surfacing on a macadam base.
3) Asphalt surfacing on a concrete base.
4) Vitrified brick with bituminous joint filler on a concrete base.
5) "Monolithic" brick, which refers to brick with cement grout filler laid on a concrete base.
6) One-course concrete, both plain and with various types of embedded steel.

Within each group, multiple test sections were constructed, each featuring variations in thicknesses and strengths. The length of each test section was 200 feet. For instance, the group with asphalt surface on a macadam base had six test sections, each incorporating different types and thicknesses of wearing and base courses. To assess the performance of these various pavement configurations, the Bates Road was subjected to traffic using Liberty Trucks, a type of U.S. Army vehicle produced during World War I. These trucks were equipped with dual 5-inch solid rubber tires on the rear axle, and their operation on the test road provided valuable data for evaluation.

The traffic testing and data collection on the Bates Road were concluded in 1923. Subsequently, the test road underwent reconstruction and became part of the state highway system. The Bates Road project stands as a significant historic effort in the field of pavement engineering, providing valuable insights into rural pavement design challenges and paving the way for further advancements in asphalt pavement technologies (Choubane & Greene, 2020).

WASHO Road Test (Idaho) – 1952 - 1954

The Western Association of State Highway Officials (WASHO) Road Test, established in 1952, was a collaborative effort involving 12 western states and the Bureau of Public Roads. The test track was located on US-191, situated 11 miles south of Malad, Idaho, and 2 miles north of the Idaho-Utah border. The primary objective of this road test was to assess the effects of different axle loadings on asphalt pavement sections.

Two test track ovals were constructed, each comprising two parallel lanes, each 12 feet wide with a 6-foot shoulder. Each loop consisted of two 1,900-foot tangents, comprising a total of five 300-foot test sections with a 100-foot transition between each section. In total, there were 20 test sections. These test sections featured two different thicknesses (2 and 4 inches) of asphalt pavements over various thicknesses of a granular base.

Construction on the test sections was completed in September 1952, and regular test traffic commenced shortly afterward. The test traffic included tractor-semi-trailer combinations, with the tractor-drive axles and trailer axles loaded to specified weights. In one loop, 22,400-lb. single-axle vehicles operated in the outer 12-foot lane, while 18,000-lb. single-axle vehicles operated in the inner lane. In the other loop, 40,000-lb. tandem-axle vehicles ran in the outer lane, and 32,000-lb. tandem-axle vehicles ran in the inner lane. Traffic testing on both loops continued intermittently until May 1954.

Some of the notable findings from the WASHO Road Test included demonstrating the benefits of paved shoulders and evaluating the relative advantages of four-inch versus two-inch thick asphalt pavements (McKendrick et al., 1957).

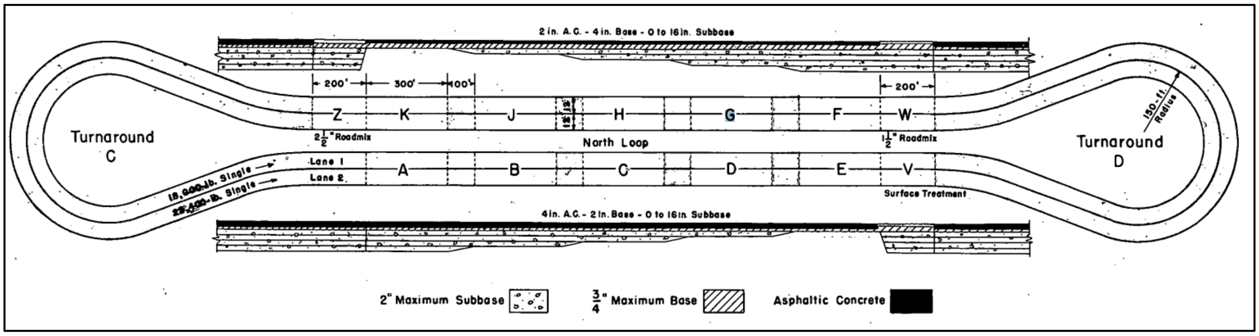


Figure 24. Schematic Layout of Test Loops at the WASHO Road Test (NRC, 1955).

AASHO Road Test (Illinois) – 1958 - 1960

The American Association of State Highway Officials (AASHO) Road Test is considered to be the first modern test road (Choubane & Greene, 2020). It was located near Ottawa, Illinois, about 80 miles SW of Chicago, along the alignment of I-80. Construction occurred from 1956 to 1958, and the facility was under test traffic from October 1958 until November 1960 (NRC, 1961). The main objective was to investigate the performance of different pavement structures and conditions under moving loads of known magnitude and frequency. The AASHO Road Test was composed of several separate major experiments: one related to flexible pavements with asphalt concrete surfacing, one related to rigid pavements with concrete surfacing, and one related to bridges.

The test facilities were made up of four large loops (Loops 3 – 6) and two smaller loops (Loops 1 & 2). Each loop was a segment of a four-lane divided highway with parallel roadways (tangents) that were connected by a turnaround at each end. Tangent lengths were 6,800 ft. for the larger loops and 4,400 ft. and 2,000 ft. for the smaller loops. An example of one of the larger test loops is shown in Figure 25. Each tangent was constructed with various pavement structural sections, where the pavement design (i.e., thickness) typically varied from section to section. The minimum length of a section was 100 ft. in the larger loops and 15 ft. in the smallest loop. Short transition pavements also separated the test sections (NRC, 1961). The AASHO Road Test included 836 test sections, with a total of 10 traffic lanes, each tested under a specific axle load. These loads ranged from 2,000 pounds on a single axle to 48,000 pounds on a tandem axle (McKendrick et., al, 1957).

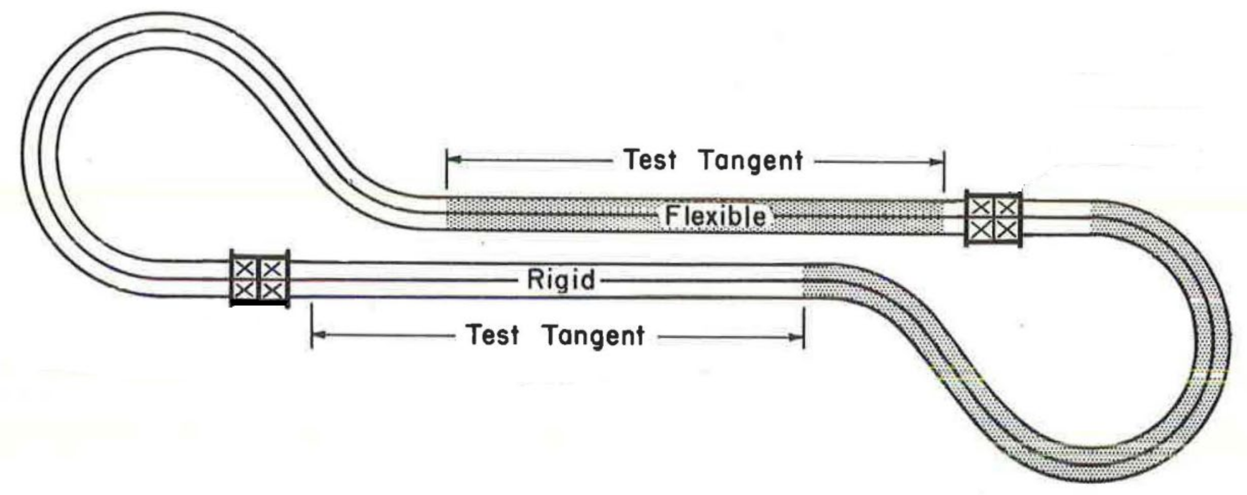


Figure 25. AASHO Road Test - Test Loop (NRC 1961).

Test traffic on the pavement sections began in November 1958, and Loops 2 – 6 were subjected to traffic for slightly more than two years. Every vehicle in any of the ten traffic lanes had the same axle load and axle configuration. The assignment of axle loads and vehicle types to the various loops and lanes is shown in Figure 26. During the loading operations, vehicles were operated at 35 mph. A total accumulation of 1,114,000 axle-load applications was applied during the 25-month traffic testing period. To accomplish this, soldiers of the U. S. Army Transportation Corps Road Test Support Activity drove more than 17 million miles (NRC, 1961).

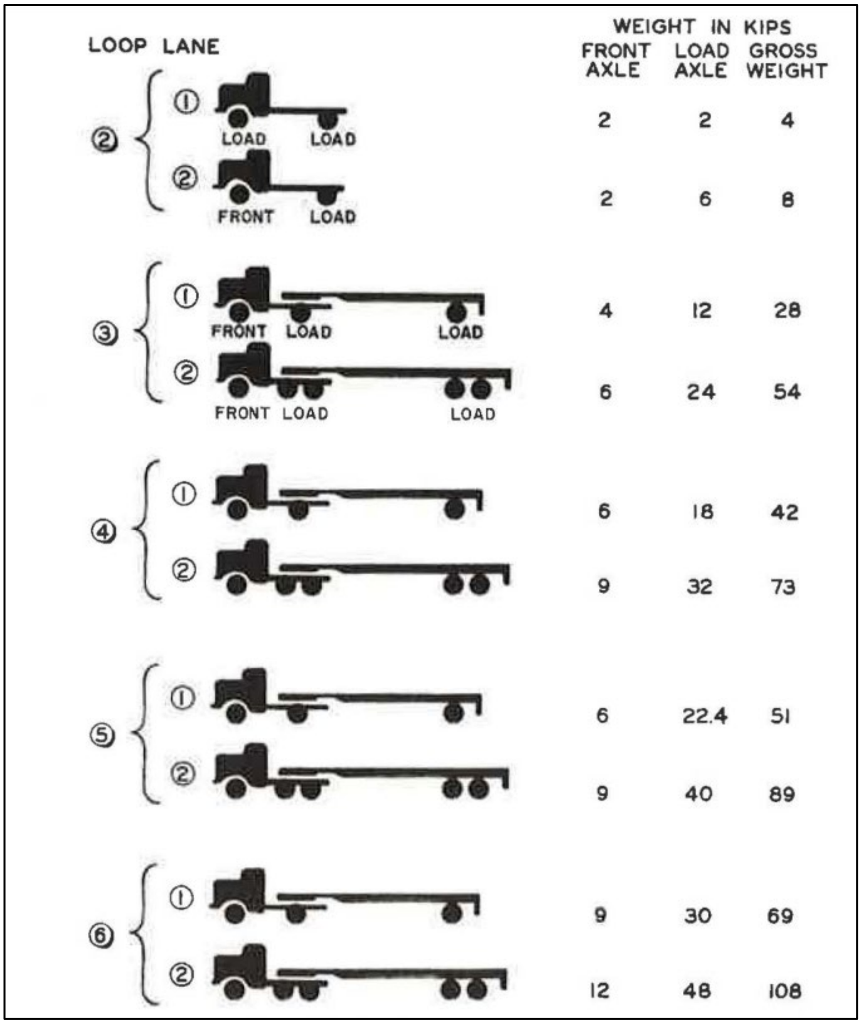


Figure 26. AASHO Road Test - Typical Test Vehicle Axle Loadings (NRC, 1962).

During the loading operations, performance measurements were routinely determined. Each measurement program was broken into one of the following three categories (NRC, 1962):

> (1) Data on the roughness and visible deterioration of the surface of each section. This included measurements of changes in the pavement profile in both the transverse and longitudinal directions, as well as the extent of cracking and patching in the surface.
> (2) A determination early in the life of each section of any transient load effects that might be directly correlated with the ultimate performance of the section. This involved measuring strains and deflections, which would serve as the basis for determining pavement capacity.
> (3) Collection of data that could potentially lead to a better understanding of pavement mechanics. This included determining the severity of pumping of rigid pavements, changes in layer thickness in flexible pavements, pavement temperatures, subsurface conditions, and numerous other measurements.

With respect to pavement maintenance, since the objectives of the study were focused on the overall performance of the test sections, maintenance operations were held to a minimum in any section that was still considered under study. When the "present serviceability" of any section dropped to a specified level, the section was eliminated from the study, and maintenance or reconstruction was performed as needed.

There were several significant findings from the AASHO Road Test that continue to impact flexible pavement designs to this day, such as (a) the development of the use of equivalent single-axle loads (ESALs), (b) the serviceability–performance concept, and (c) effects of layer thickness and strength. It also changed the way that people conduct pavement research by illustrating the power of factorial experiments, high-quality data, and statistical analysis (Hudson, et., al 2007).

MnROAD (Minnesota) – 1993

The Minnesota Road Research Project (MnROAD) was constructed by the Minnesota Department of Transportation (MnDOT) between 1989 and 1993. This extensive pavement research facility is located approximately 40 miles northwest of Minneapolis-St. Paul, near the town of Albertville, along Interstate 94. MnROAD comprises two test roadways situated along I-94, and it serves as a testing ground for over 50 distinct test sections, each representing various combinations of paving materials and designs. To gather comprehensive data on pavement response and environmental conditions, more than 4500 sensors were installed within these test sections. The test sections are distributed across three separate roadway segments within MnROAD:

1) A 3.5-mile section of I-94, which is the original westbound roadway of I-94.
2) Another 3.5-mile section of the I-94 mainline roadway.
3) A 2.5-mile Low Volume Road, specifically designated for testing purposes.

For the I-94 segments, MnROAD utilizes in-service highway traffic for loading by diverting the westbound traffic onto the test sections. A bypass is provided to shift traffic off the mainline testing segment when more detailed evaluation is necessary. This allows for continuous real-world loading and monitoring of the test sections. Additionally, a 2.5-mile section looping through the MnROAD facility serves as a low-volume road test. In this case, a semi-trailer truck is used for loading, providing controlled conditions for studying the pavement's performance under specific loading scenarios. (Choubane & Greene, 2020). The MnROAD layout is shown in Figure 27.

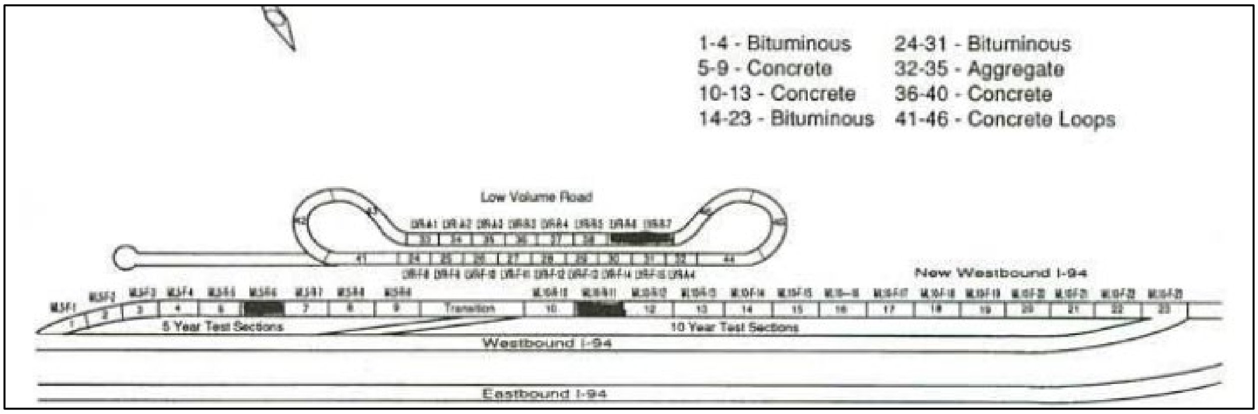


Figure 27. Test Cell Locations at MnROAD (Metcalf, 1996).

WesTrack (Nevada) – 1995

WesTrack was a 2.9-kilometer oval loop located approximately 35 miles southeast of Reno, Nevada. in Western Nevada. The Federal Highway Administration (FHWA) sponsored its construction. The primary objectives of WesTrack were to 1) continue the development of performance-related specifications (PRS) and determine the impact of material variations (particularly binder content, aggregate gradations, and density) on pavement performance and 2) provide early field validation of the Superpave performance prediction models and mixture analysis procedures (formerly called the Level III mixture design). Construction of the first pavement test sections on WesTrack was completed in October 1995, and truck loading was initiated in March 1996 (Mitchell, 1996). The Layout of the WesTrack test sections is shown in Figure 28.

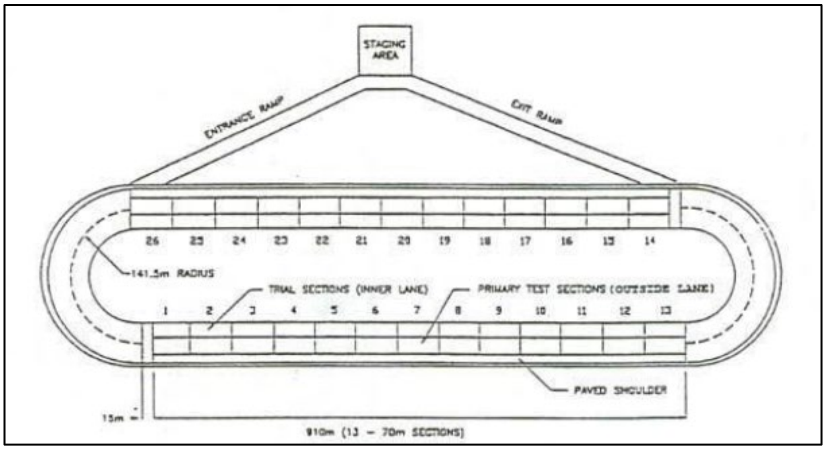


Figure 28. Layout of WesTrack Test Sections (Metcalf, 1996).

The track consisted of two tangent pieces and the superelevated curves connecting them. Each tangent contained 13 test sections (26 test sections total), each 70 m long. There were no test sections along the curves. The track had two 3.7 m wide lanes; the inside lane was the “trial” lane, and the outside lane was the “test” lane. The trial lane was used to fine tune the mixtures prior to placement in the test lane. The pavement structural cross section was uniform throughout the 26 test sections and consisted of 150 mm of HMA placed over 300 mm of aggregate base and 450 mm of compacted fill subgrade (Mitchell, 1996).

Each of the 70-m-long test sections consisted of three zones. The first 25 m (in the direction in which the trucks were moving) was a transition zone, which allowed for variability during the construction of the section and for reducing the impact of any truck dynamic motion generated by damaged pavement in the previous test section. The next 40 m was the test area in which all nondestructive pavement performance measurements were made, and the last 5 m was the area in which destructive sampling and test measurements were made periodically.

Ten million ESALs were applied to the track over a two-year period. Four identical trucks were used – triple-trailer combinations loaded with tied-down steel plates. A fully loaded truck weighed 676 kN and applied approximately 10.3 ESALs per pass to the test sections. Driverless vehicles were used, guided by wires under the pavement (Mitchell, 1996). data collection at WesTrack included the following:

1) Geotechnical information, including test pits and boring logs prior to construction.

2) Falling weight deflectometer (FWD) testing prior to construction, after the construction of each layer, and monthly during the two-year loading.
3) Laboratory testing of all materials before and during construction for design and quality control. In addition, additional samples were tested from each section to fully characterize the in-place construction, e.g., for asphalt binder content and air voids.
4) Superpave binder and volumetric mix design testing, as well as complete Superpave mixture analysis testing.
5) Strain gauges as well as temperature and moisture content sensors in the pavement sections
6) Performance monitoring (visual distress, transverse, and longitudinal profiling, FWD, and friction) at two- or, in some cases, four-week intervals during the loading period.
7) Strain gauge and accelerometer data was collected from sensors mounted on the axles of one of the trucks to provide dynamic loading information.

NCAT Test Track (Alabama) 2000

In 2000, the National Center for Asphalt Technology (NCAT) completed the construction of an innovative oval test track in Lee County, Alabama, situated approximately 20 miles from Auburn University on a 309-acre site. This oval test track, with a length of 1.7 miles, is a state-of-the-art facility designed to facilitate comprehensive pavement research and experimentation. The test track comprises 46 individual test sections, each measuring 200 feet in length and set up to accommodate different experiments. Funding for the construction of this track was a collaborative effort involving highway agencies and industry sponsors (Brown and Powell, 2001).

Triple-trailer trucks are used to apply traffic loads on the test track, simulating real-world conditions and allowing researchers to study the performance of various pavement designs and materials under heavy truck traffic. The test sections on the track are categorized into three main types:

- Structural Experiments: These sections represent different pavement structures with varying thicknesses, closely resembling real-world pavements. Embedded strain and pressure sensors in the pavement structure allow for the analysis of pavement response to loads, facilitating the validation of mechanistic-empirical pavement design procedures. Monitoring performance on a continuous basis helps evaluate rutting, fatigue cracking, roughness, texture, friction, and noise.

- Surface Mix Experiments: These sections are built with robust cross-sections that limit distress to the experimental surface layers. Researchers can study the performance of various asphalt surface mixtures under different conditions, collecting data on performance indicators like rutting, cracking, and skid resistance.

- Pavement Preservation Studies: This category focuses on evaluating various pavement preservation techniques and materials. Researchers can explore strategies to extend pavement life and delay costly rehabilitation by testing different preservation methods and their effectiveness.

The NCAT oval test track provides an invaluable platform for advancing pavement research and development. It not only supports the study of traditional asphalt materials and pavement design but also offers opportunities for non-pavement materials-related projects, such as evaluating new heavy vehicle suspension systems, alternative fuels, and improved vehicle electronics and safety (Choubane and Greene, 2020).

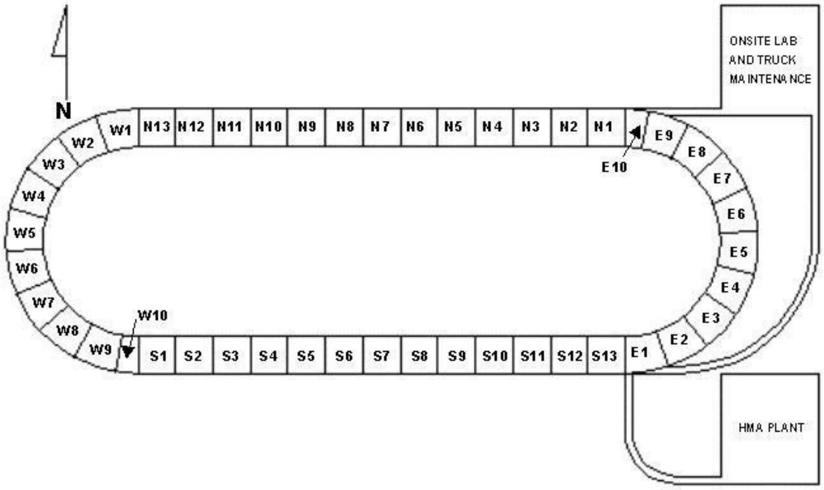


Figure 29. Layout of NCAT Test Track.

NCAT Pavement Preservation Studies (Alabama) 2012

In 2012, the National Center for Asphalt Technology (NCAT) conducted pavement preservation studies on Lee County Road 159 in Auburn, Alabama. This road, characterized by low-volume but high-truck traffic, served as the location for applying 23 pavement preservation treatments or treatment combinations to 100-foot-long test sections. Additionally, two sections with significantly different initial conditions were left untreated as control sections. At the time of treatment, the existing pavement on Lee County Road 159 was 14 years old, consisting of a 5.5-inch hot-mix asphalt layer, and exhibited varying conditions ranging from good to poor (Vargas, 2019).

Expanding the research scope in the 2015 research cycle, NCAT applied 34 pavement preservation treatments or treatment combinations to a section of U.S. Route 280, which represented a higher volume roadway. These new sections on U.S. Route 280 served as a complement to the existing test sections on the Test Track and Lee County Road 159. The sections on U.S. Route 280 were 0.1 miles long, located on the outside lane of the two-lane eastbound highway. Control sections with low and high levels of cracking, rutting, International Roughness Index (IRI), and texture were also included in this study. At the time of treatment, the existing pavement on U.S. Route 280 was nine years old, with an average thickness of 9.9 inches (Vargas-Nordcbeck et al., 2019). Periodic condition assessments are conducted on the preservation sections to evaluate surface distress, ride quality, friction, and structural integrity of the pavements. These assessments include laser profiling, crack mapping, friction testing, and deflection measurements. These condition assessments provide valuable data on the performance of different pavement preservation treatments over time, allowing researchers to make informed decisions about the most effective preservation strategies for maintaining and extending the life of the roadways.

MnROAD Pavement Preservation Studies (Minnesota) 2015

In 2015, the Minnesota Department of Transportation (MnDOT) became a part of the NCAT Pavement Preservation Study and proceeded to construct test sections in the summer of 2016. These test sections were established on County State Aid Highway 8 (CSAH 8) in Pease, Minnesota, to accompany the existing test sections on Lee County Road 159. The purpose of this collaboration was to assess the performance of pavement preservation treatments using locally available materials and following the same design procedure and construction contractor as the original Lee County Road 159 sections.

Before the treatment in August 2016, CSAH 8 had been last constructed in 2005, comprising a pavement structure consisting of a 7.0-inch Hot-Mix Asphalt (HMA) layer over a 6.0-inch granular base, which in turn rested on a sand and gravel subgrade. CSAH 8 experiences an estimated Average Annual Daily Traffic (AADT) of 510 vehicles per day, with approximately 7% of the traffic being heavy vehicles. The heavy vehicle traffic primarily consists of implements of husbandry, such as vehicles used in corn fields and dairy farms that are located along CSAH 8. These vehicles contribute to the loading on the pavement sections, and their movements. As the sections are placed consecutively without major turning points, the vehicle traffic is assumed to be consistent for all chip seal sections on CSAH 8 (Vargas-Nordcbeck et al., 2019).

To collect performance data on the Minnesota sections, a data collection vehicle equipped with an inertial profiler, an INO laser rut measurement system (LRMS), and high-resolution cameras was used periodically. These instruments allowed for the assessment of various pavement performance indicators such as ride quality (IRI), rutting, and the condition of the pavement surface. The collaboration between MnDOT and NCAT in conducting this pavement preservation study on CSAH 8 provides valuable insights into the effectiveness of different preservation treatments, allowing for informed decisions on pavement maintenance and management strategies for Minnesota roadways.

Long-Term Pavement Performance (LTPP) Studies

The LTPP Program was initiated to collect pavement performance data within the scope of the Strategic Highway Research Program (SHRP). This program encompasses two main classes of studies: the General Pavement Study (GPS) and the Specific Pavement Study (SPS) experiments. The goal of these studies was to investigate critical pavement details that influence pavement performance. The combined GPS and SPS experiments include over 2,500 test sections located on in-service highways throughout North America. The LTPP program continually monitors and collects pavement performance data on all active sites, covering various aspects such as inventory, maintenance, monitoring (deflection, distress, and profile), rehabilitation, materials testing, traffic, and climatic conditions.

The GPS experiments were designed to examine general performance based on different pavement types commonly used in North America. In contrast, the SPS experiments were focused on investigating the impact of specific features, such as drainage, layer thickness, and various rehabilitation or maintenance treatments on pavement performance. The test sections for the SPS experiments were specifically constructed for the LTPP study.

When establishing the design criteria for both GPS and SPS experiments, some common elements were considered. For instance, all test sections were situated in the outside single lane, commonly referred to as the "slow lane" of the roadway. Initially, the planned length for test sections was 1,500 feet. However, this was later shortened to 500 feet due to difficulties in finding homogeneous sections of the required length. The decision to use 500 feet was based on the notion that performance consistency over this length would adequately represent general roadway conditions, and it allowed for the measurement of profile wavelengths for assessing smoothness.

In the GPS experiments, individual highway agencies would nominate previously constructed projects for inclusion in the study. A research matrix was used to plan the various experiments. After the highway agencies identified the sections, LTPP contractors conducted field trips to confirm section characteristics and identify and mark out the 500-foot sections for the study. These field visits were essential as they sometimes revealed significant differences between the actual section characteristics and the initial expectations. In some cases, variations in layer types and thicknesses or traffic generators like on/off ramps or intersections complicated the identification of suitable sections for study. Cores were obtained at the ends of each section during the field visits to confirm layer types and thicknesses. The identified section locations were documented, marked out, and confirmed for the study (FHWA, 2015).

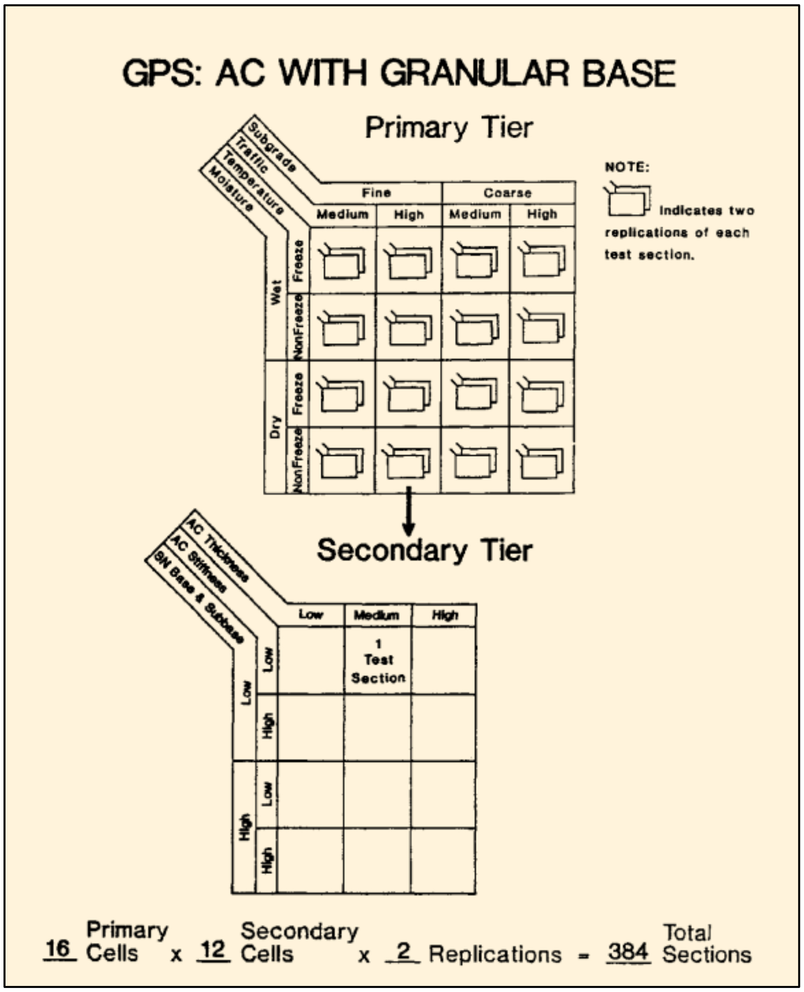


Figure 30. GPS-1: Asphalt Concrete with Granular Base.

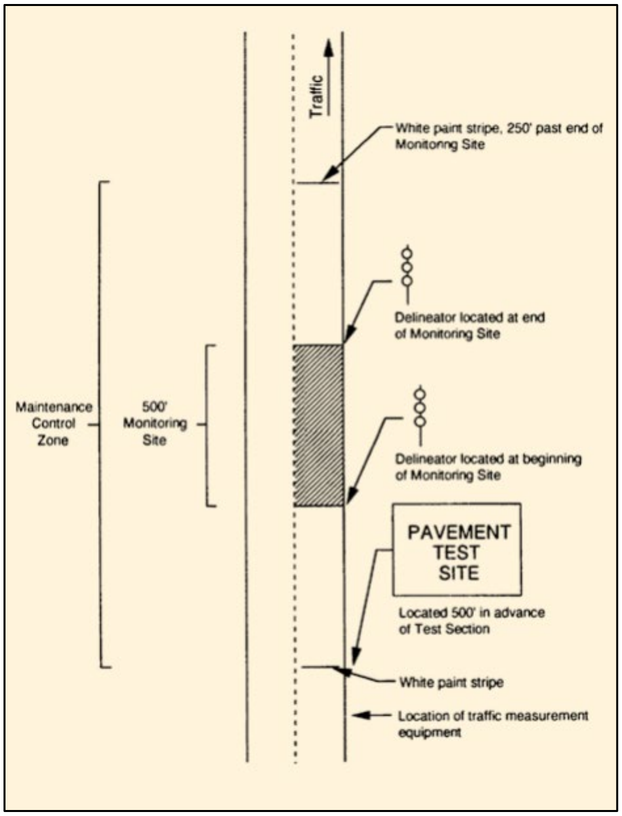


Figure 31. Designated Layout for GPS Sections.

The Specific Pavement Study (SPS) experiments are composed of projects selected from various locations across the United States. Each project is carefully chosen and constructed to represent specific site conditions and experimental factors. These factors are combined to form an experimental matrix, which outlines the projects needed to ensure a statistically sound and robust experiment.

One recent SPS experiment, known as SPS-10, focused on studying the performance of Warm Mix Asphalt (WMA) in comparison to Hot Mix Asphalt (HMA). The experiment, titled "Warm Mix Asphalt Overlay of Asphalt Pavement Study," involved creating an experimental matrix with the following variables: moisture conditions (wet/dry), temperature conditions (freeze/no-freeze), traffic intensity (high/low), and the number of projects (two in each category). Each project location consisted of a minimum of three test sections, each 500 feet in length, constructed contiguously along a section of highway. The goal was to keep all factors, including traffic, climate, subgrade conditions, existing pavement structure, and asphalt production plant, consistent across all test sections within a project location (Puccinelli et al., 2023).

To ensure effective data collection and monitoring of pavement performance, the experimental design required a minimum section length of 800 feet. Within this length, a 500-foot monitoring section was designated for non-destructive performance monitoring (Figure 31). A 50-foot buffer zone was placed on each side of the monitoring area to separate it from the destructive sampling area. The destructive sampling area, also 100 feet on each side, was constructed simultaneously with the same specifications as the monitoring area. This arrangement allowed for material sampling without disturbing the monitored area, which was limited to the outside (truck) lane only. Transition zones were established between the test sections and were a minimum length of 800 feet.

Each project involved the production of at least three different mixtures, including one HMA and two WMA mixtures, all produced at the same asphalt production plant. Additionally, each mixture could only be placed on the test section once the asphalt production plant had achieved steady-state operation (Puccinelli et al., 2023). Through the SPS-10 experiment, researchers aimed to gain valuable insights into the performance of WMA compared to HMA under various conditions and factors, contributing to advancements in pavement design and maintenance practices.

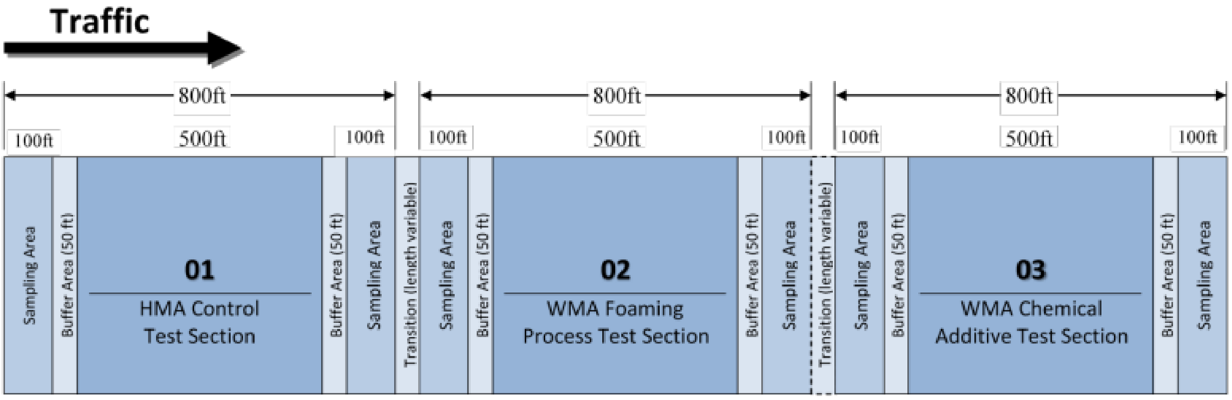


Figure 32. Diagram of a Typical SPS-10 Site Layout (Puccinelli et al., 2023).

Accelerated Pavement Test Facilities

Accelerated Pavement Testing (APT) facilities play a crucial role in pavement research and development by providing a controlled environment for full-scale testing. These facilities utilize accelerated pavement testers to apply controlled loading, simulating traffic on pavement test sections. APT facilities offer several advantages, including faster testing times, increased safety for researchers and the public, and the ability to evaluate different factors simultaneously. Numerous APT programs have been established in the United States since the 1990s, focusing on various aspects of pavement performance. Some notable APT facilities in the US include:

1) **FHWA Turner Fairbanks Highway Research Center**: Utilizes two ALF machines that apply rolling wheel loads on a 45-foot test length of any test pavement. Each machine can apply an average of 35,000-wheel passes per week, simulating the rear dual-wheel of a typical truck (FHWA).

2) **Indiana Department of Transportation** (INDOT): Collaborated with Purdue University to initiate an APT program in 1992, using pavement test tracks with control over water table and pavement temperature. Loads of up to 20,000 lbs. are applied using dual wheels or super-single half-axle assembly at 5 mph, either uni- or bi-directionally (White et al., 1992).

3) **California Department of Transportation** (Caltrans): Implemented their APT program in 1994, utilizing two Heavy Vehicle Simulators (HVS) for full-scale pavements testing. Wheel loads of up to 45,000 lbs. are applied on a half axle, moving in either bi-directional or unidirectional mode at speeds up to 6.2 mph (Harvey et al., 2000).

4) **Louisiana Department of Transportation and Development** (DOTD): Constructed a full-scale pavement testing area using an ALF in 1995, applying rolling wheel loads on a 39 ft section of pavement at a speed of 10.4 mph (Metcalf, et al., 1998).

5) **Florida Department of Transportation** (FDOT): Initiated an APT program in 2000, consisting of multiple linear test tracks with a Heavy Vehicle Simulator (HVS) for accelerated loading (Choubane and Greene, 2020).

These APT facilities contribute significantly to the advancement of pavement design, evaluation, and maintenance practices by providing valuable data and insights into pavement performance under various conditions and loading scenarios.