Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-03/4047-2		

4. Title and Subtitle	5. Report Date
ASSESSMENT OF THE TxDOT BINDER QUALITY ASSURANCE PROGRAM	December 2003
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Amy Epps Martin, Eun Sug Park, Edith Arambula, and Clifford Spiegelman	Report 4047-2
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	11. Contract or Grant No. Project No. 0-4047
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Texas Department of Transportation Research and Technology Implementation Office P. O. Box 5080 Austin Texas 78763-5080	Research: August 2000-August 2002
	14. Sponsoring Agency Code

15. Supplementary Notes

Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.
Research Project Title: Analysis and Development of Asphalt Quality Assurance Procedures

16. Abstract

Most state departments of transportation (DOTs) maintain asphalt binder quality assurance (QA) programs to ensure that asphalt binders used in the construction of their road system meet specifications stipulated for each project. The Texas Department of Transportation (TxDOT) samples and approves asphalt materials at the source and utilizes these materials in highway projects without consideration of possible changes in properties that may occur between production and use during construction. Historic concern and limited recent data indicate that binder properties do change, contributing to difficulties during construction operations and poor performance. The primary goal of this project was to evaluate the current TxDOT binder QA program and recommend revisions as necessary toward improving quality. This final report documents an assessment of the current program based on (1) an extensive information search and review that included two detailed surveys of TxDOT districts and nine other state DOTs and (2) results from a comprehensive laboratory testing program that simulated factors that may affect changes in binder properties prior to use. Recommendations for improvement are provided, and these recommendations are evaluated in terms of resource requirements. In a recommended total binder quality system, the binder QA program would be one tool to improve material quality used during construction. Other recommended tools include mandated quality control (QC) plans for both suppliers and contractors that require special handling for binders, training programs for sampling and testing, general education seminars presenting the goals of the system to all involved, and a round-robin program to establish testing variability. Development of a user-friendly database for storing binder QA data that includes storage times, storage temperatures, and pavement performance data is also recommended with a suggestion of frequent analysis of this data, possibly to set frequency rates for field sampling by binder and supplier.

17. Key Words	18. Distribution Statement
Asphalt, Binder, Quality Assurance, Evaluation	No restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	88	

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

ASSESSMENT OF THE TxDOT BINDER QUALITY ASSURANCE PROGRAM

by

Amy Epps Martin
Assistant Research Scientist
Texas Transportation Institute

Eun Sug Park
Assistant Research Scientist
Texas Transportation Institute

Edith Arambula
Graduate Assistant Researcher
Texas Transportation Institute

and

Clifford Spiegelman
Research Scientist
Texas Transportation Institute

Report 4047-2
Project Number 0-4047
Research Project Title: Analysis and Development of Asphalt Quality Assurance Procedures

Sponsored by the
Texas Department of Transportation
In Cooperation with the
U.S. Department of Transportation
Federal Highway Administration

December 2003

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration (FHWA) and TxDOT. This report does not constitute a standard, specification, or regulation, nor it is intended for construction, bidding, or permit purposes. Trade names were used solely for information and not for product endorsement. The engineer in charge was Amy Epps Martin, P.E. (Texas No. 91053).

ACKNOWLEDGMENTS

The authors thank the Texas Department of Transportation (TxDOT) and the Federal Highway Administration (FHWA) for their support in funding this research project. Special thanks goes to Darren Hazlett and Jerry Peterson of TxDOT for their efforts in providing technical guidance, support, and direction to this project. Special thanks are also due to Jacob Bell and Brian Elliott of the Texas Transportation Institute (TTI) for their help in the laboratory and in the field. Thanks also to Amanda Hughes and Sara Khan from TTI for revising TxDOT data into a usable format. Thanks to the many departments of transportation (DOTs) who provided the authors with an understanding of their binder quality assurance (QA) programs and, in some cases, a data set for analysis. The suppliers who donated materials for this project and provided interest and support are also thanked. Finally, special gratitude is due to Charles Glover and Rajesh Ramachandran of the Center for Asphalt and Materials Chemistry in the Department of Chemical Engineering at Texas A&M University for their guidance and help during the laboratory testing program.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

CHAPTER 1. INTRODUCTION

Most state departments of transportation (DOTs) maintain asphalt binder quality assurance (QA) programs to ensure that asphalt binders used in the construction of their road system meet specifications stipulated for each project. These specifications include the binder grade, which is selected based on environmental and traffic conditions expected over the design life of the project. Production of a quality asphalt pavement requires that the binder used during construction meet the requirements for the selected grade.

Binder QA programs may require sampling at the production source (supplier sample), during construction (field sample), or both. Possible sampling points are shown in Figure 1 as boxes and include the following:

- a storage tank at the production site or refinery,
- a storage tank at a supplier terminal,
- a transfer line to load transports at the production site or refinery,
- a blending line to load transports without intermediate tank storage,
- a transfer line from a transport to a storage tank at the contractor site,
- a storage tank at the contractor site,
- a transfer line from the contractor storage tank to the hot-mix asphalt (HMA) plant for asphalt cements, and
- a spray bar on a distributor truck for liquid asphalts (asphalt emulsions or cutback asphalts).

Programs that do not require sampling during construction, either at the HMA plant or at the project site for spray applications of liquid asphalts, do not consider possible changes in material properties that may have occurred between production (at the supplier location) and use during construction (in the field). Some of these changes may be detrimental in terms of performance or create difficulties during construction operations. Performance problems may surface if changes in material properties render a binder used during construction to have insufficient capacity to resist the primary forms of distress under the environmental and traffic conditions for a specific project. Construction difficulties may arise if, for example, contractors select compaction temperatures based on the specified grade and these temperatures are inadequate in terms of consistency for the actual material used.

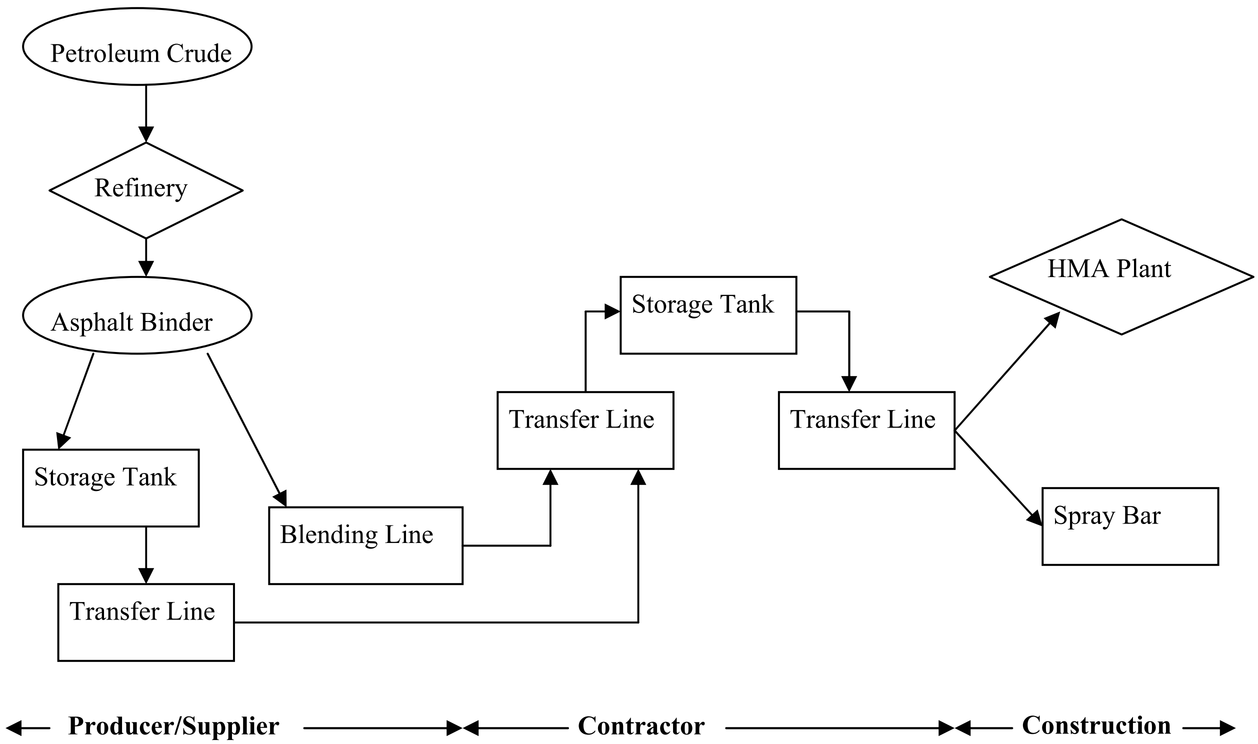


Figure 1. Asphalt Binder from Production to Construction.

The Texas Department of Transportation (TxDOT) samples and approves asphalt materials at the source, and these materials are then utilized in highway projects without consideration of possible changes in properties that may occur between production (at the supplier location) and use during construction (in the field). Historic concern and limited recent data indicate that binder properties do change, contributing to construction and operation difficulties as well as poor performance. The primary goal of this project was to evaluate the current TxDOT QA program for binders and recommend revisions as necessary toward improving quality. The work plan executed to provide the results contained in this report was not as initially proposed due to significant difficulties obtaining corresponding data from supplier and field samples. Modifications to the proposed work plan were required throughout the project, and these modifications were undertaken at the direction of TxDOT personnel.

To evaluate the TxDOT binder QA program, an understanding of factors that may cause changes in binder properties between production (at the supplier location) and use during construction (in the field), performance models that quantify the effect of these changes on performance, current TxDOT QA practices, and other state DOT QA programs is needed. An interim report documents the results of an extensive information search and review and the design of comprehensive field and laboratory testing programs (*1*). This report summarizes documentation included in the interim report, presents results from the comprehensive testing programs, provides recommended changes to the TxDOT binder QA program and corresponding required resources, and concludes with suggested future research.

CHAPTER 2. INFORMATION SEARCH AND REVIEW

A literature search and review was conducted with the assistance of the Texas Transportation Institute (TTI) library staff and completed an extensive survey to accomplish the following goals:

- obtain general definitions of and recommendations for QA programs with an emphasis on binder QA programs,
- identify prospective binder properties directly related to performance that can be measured in a timely manner for use in a QA system,
- identify any performance models that relate off-target values of binder properties to loss of field performance and associated costs,
- identify factors that may cause changes in properties of binders sampled from the source to those sampled just prior to use,
- define the current binder QA program in Texas and its impact on TxDOT districts, and
- define the state-of-the-practice in binder QA programs in Texas and other selected states.

This chapter summarizes the results of each part of the information search and review, including summaries of the relatively small body of literature found and general comparative descriptions of the binder QA programs in Texas and selected states. An interim report provides more detailed descriptions (*1*).

QUALITY CONTROL AND QUALITY ASSURANCE

General references by A. Mitra, D.C.S. Summers, R. Aguayo, and A. Gabor define quality control (QC) and QA and describe the use of statistics to enhance quality and aid in decision making (*2, 3, 4, 5*). QC is generally defined as a system used to maintain a desired level of quality in a product or service. This goal may be achieved through different measures such as planning, design, use of proper equipment and procedures, inspection, and corrective action when a deviation is observed between the product, service, or process output and a specified standard. QA is generally defined as all planned or systematic actions necessary to provide confidence that a product or service will satisfy given needs.

Several people have made significant contributions in the field of QC/QA. W.E. Edwards Deming may be the most recognized (*6*). He conducted a thriving worldwide consulting practice for more than 40 years with clients that included manufacturing companies, telephone companies, railways, carriers of motor freight, consumer researchers, census methodologists, hospitals, legal firms, government agencies, and research organizations in universities and in industry. He suggested the following 14 points for management that are fundamental to the implementation of any quality program:

- Create and publish to all employees a statement of the aim and the purposes of the company or other organization. The management must consistently demonstrate their commitment to this statement.
- Everyone, including top management, must learn the new philosophy.
- Understand the purpose of inspection, for process improvement and cost reduction.
- End the practice of awarding business on the basis of the price tag alone.
- Constantly and continuously improve the system of production and service, to improve quality and productivity and, thus, constantly decrease costs.
- Institute training on the job.
- Institute leadership.
- Drive out fear, so that everyone may work effectively for the company.
- Break down barriers between departments.
- Eliminate slogans, exhortations, and targets for the work force asking for zero defects and new levels of productivity.
- Eliminate work standards (quotas) on the factory floor. Substitute leadership.
- Create pride in the job being done.
- Institute a vigorous program of education and self-improvement.
- Put everybody in the company to work to accomplish the transformation.

These 14 points, integral to a successful QC/QA program, were integrated into the recommendations presented in this report.

Statistics can be utilized in both QC and QA environments to aid in decision making. QC uses process control charts to compare material properties during production with required test values and to determine when a change in the process is required to consistently produce material that meets specifications. Statistics can also be used in this setting to determine if a particular process can produce material that meets specific requirements. Confidence intervals are used in QA to account for material, sampling, and testing variability and to determine when a material fails a single property or multiple properties required in a specification. In this report, results obtained through statistical analysis techniques demonstrate the potential for establishing a binder QA program with continuous improvement and availability of information relevant to decision making toward improving quality.

With regard to binder quality, suppliers and contractors are responsible for maintaining their own QC system. The owner, generally a DOT, defines and maintains the QA system to ensure a binder has all properties required by the specification and related to adequate performance to guard against premature failure. Many states utilize the American Association of State Highway and Transportation Officials (AASHTO) PP26 Standard Practice for Certifying Suppliers of Performance-Graded (PG) Asphalt Binders as a guideline for establishing their QC/QA systems (*7*). This standard defines PG suppliers and their responsibilities in terms of assuring specification compliance. The supplier must submit a QC plan to the agency that details the testing procedures and frequency to assure compliance.

AASHTO PP26 provides guidance for minimum QC plan components and a standard form for reporting data. QC plan requirements include transport inspection guidelines and initial, reduced, and minimum testing frequencies. This standard also provides sampling and laboratory accreditation requirements. If historical compliance is demonstrated, the standard defines an approved supplier certification program that agencies may use to minimize disruption in the construction process. Agency responsibilities outlined in AASHTO PP26 include acceptance of the QC plan, administration of the certification program, and inspection of supplier facilities. The standard also describes provisions for split sample and QA sampling and testing, but it does not specify guidelines for sampling and testing frequencies or specific acceptable tolerances for specification parameters. For reduced testing frequencies in supplier QC plans, the variability of each test is suggested for the tolerance level.

The Northeast Center of Excellence for Pavement Technology (NECEPT) is currently addressing deficiencies in AASHTO PP26 through a pooled-funds study (*8, 9, 10, 11*). These deficiencies include failure to specify sampling and testing frequencies for QA samples, sampling locations for QA samples to account for changes in binder properties that may occur subsequent to production, acceptable tolerances for specification compliance that consider all possible sources of variability, and corrective action for noncompliance. Their goals include development of a QC/QA system that includes multiple components to address these inadequacies. They have developed a binder technician and laboratory certification program, a split sampling program to establish expected testing variability, a QC program for suppliers, a QA program that includes conflict resolution guidelines and payment schedules incorporated in a simulation program that ensures a balance between agency and supplier risk, and a regional database with common specification certificates of analysis to support these programs.

BINDER PROPERTIES RELATED TO PERFORMANCE

The recently implemented specification system for binders used in HMA was developed during the Strategic Highway Research Program (SHRP) and utilizes laboratory tests that measure fundamental physical properties that can be directly tied to field performance of asphalt-aggregate mixtures. This system specifies binder properties for unmodified or modified asphalt cements used in HMA to ensure safety, provide for ease in pumping and handling, guard against excessive aging, and mitigate the three major forms of distress in asphalt concrete pavements: permanent deformation, fatigue cracking, and thermal cracking (AASHTO MP1) (*12*). The PG binder specification system was developed based on unmodified asphalt cements, but the equipment and form of the specification is expected to be applicable to modified binders. National Cooperative Highway Research Program (NCHRP) Project 9-10 explored the applicability of the PG specification to modified binders and assessed what changes are needed to support evaluation of these materials (*13*).

The properties specified in the PG system are consistent for all binders; only the temperatures at which these properties must be met vary. Each property specified is measured using a characterization test described in this section. For a specific project, predicted pavement temperatures and traffic conditions determine the binder grade needed for satisfactory performance.

The characterization tests required to specify a binder measure physical properties related to pavement performance directly through engineering principles. A historical database of past performance is not needed to use test results as a prediction tool, although validation is required and has been completed in terms of laboratory mixture performance tests (*14*). A characterization test related to rutting performance is conducted on binder that has been short-term aged in the Rolling Thin Film Oven (RTFO) (American Society for Testing and Materials [ASTM] D2872), to simulate the critical state for this type of distress after mixture production and construction (*7, 15*). Tests related to cracking performance are conducted on binder that has been short-term aged in the RTFO and long-term aged in the Pressure Aging Vessel (PAV) (AASHTO PP1) to simulate the critical state for both fatigue and thermal cracking (*7, 15*).

A dynamic shear test (AASHTO TP5) characterizes binder resistance to rutting and fatigue cracking (*7*). This test is used to evaluate the time- and temperature-dependent behavior of binders at intermediate and high temperatures. A controlled stress dynamic shear rheometer (DSR) measures the viscoelastic behavior of the material in terms of complex shear modulus (G^*) and phase angle (δ). The DSR applies a sinusoidal variation in shear stress (τ) to a thin film of binder at a frequency of 10 rad/s, and the resulting sinusoidal variation in shear strain (γ) is measured (*12*).

The complex shear modulus (G^*) provides a measurement of the total material resistance to repeated shear stress, including the elastic or recoverable deformation and the viscous or non recoverable deformation. The phase angle (δ) provides an indication of the relative amount of elastic response as compared to viscous response, with $G^*\cos\delta$ or the component in phase with the stress measuring the elastic response and $G^*\sin\delta$ or the loss modulus relating the viscous response. Phase angles vary from 0 to 90°, with a zero angle representing a purely elastic material and a right angle representing a purely viscous material. At low temperatures, binders behave more like elastic solids, with δ approaching zero. To completely characterize a binder, both properties are needed as functions of temperature and time of loading, as two binders may have equivalent G^* values but behave differently due to the relative amount of elastic versus viscous response to applied shear stress, indicated by the phase angle (δ).

The specification combines both rheological properties by specifying a minimum value of $G^*/\sin\delta$ for short-term aged binders. This parameter controls permanent deformation by limiting the dissipated energy in a controlled stress repetitive shear loading test. The minimum $G^*/\sin\delta$ is set at 2.20 kPa in the specification for a loading frequency of 10 rad/s. In the development of the specification, this limit was determined based on measured $G^*/\sin\delta$ for unaged and commonly used AC-10 binders and an average measured value of aging index (ratio of absolute viscosity after RTFO to viscosity before RTFO) for these materials that historically has shown adequate performance in terms of resistance to permanent deformation in moderate climates (represented by the conventional 140 °F [60 °C] viscosity measurements) (*16*).

The specification for long-term aged binders requires a maximum $G^*/\sin\delta$ value of 5000 kPa for a loading rate of 10 rad/s, as measured in the DSR. This parameter is assumed to control fatigue cracking in thin pavement structures by limiting the dissipated energy in a controlled strain repetitive loading test. The maximum value for $G^*/\sin\delta$ was selected based on a large study of 42 binders, with 15 percent failing to meet the specified maximum value (*16*). The effects of pavement structure and mixture stiffness in terms of HMA resistance to fatigue cracking are not currently included in the PG binder specification.

The bending beam rheometer (BBR) and the direct tension tester (DTT) are used to determine the low-temperature behavior of binders. The BBR characterizes binder stiffness at temperatures too low for accurate measurement with the DSR. With both pieces of equipment, binder stiffness is evaluated over a wide range of temperatures critical to performance in the field. The BBR subjects a small beam of binder to a constant creep load and measures the resulting deflection at a temperature related to the lowest service temperature encountered by a pavement (AASHTO TP1) (*7, 12*). Using beam theory, the binder stiffness (S) is calculated.

This stiffness provides a measure of the binder resistance to creep loading at low temperatures, simulating thermal stresses incurred in pavements as temperatures decrease. The creep rate (m) is also determined from test results as the change in stiffness with time as measured on a log-log plot. The BBR testing temperature is 50°F (10°C) higher than the low pavement temperature expected in the field to reduce testing time to 240 s using the principle of time-temperature superposition (*12*). This provides results equivalent to the creep stiffness and creep rate after a two-hour loading time at the minimum pavement temperature. The binder specification sets limits on the stiffness and m-value at a 60-s loading time. These parameters represent critical properties of the binder that directly relate to HMA resistance to thermal cracking. For adequate resistance to this form of distress, the binder plays a predominant role. For a given change in temperature, binders with more resistance to thermal cracking will exhibit smaller induced tensile stresses (controlled by stiffness) and relax these induced stresses at a faster rate (controlled by the m-value). The specification requires a creep stiffness at 60 s to be less than 300 MPa and an m-value at this same time of loading to be at least 0.30. If the stiffness is between 300 and 600 MPa, the requirement for direct tension failure strain may be used to pass the specification.

The DTT provides an indication of the strain that can be sustained by a binder prior to failure. Although relationships exist to relate the creep stiffness measured with the BBR to the strain at break for unmodified binders, these relationships do not apply to all binders, especially modified ones. The DTT pulls a dog-bone-shaped sample of binder at a slow constant rate until failure (*12*). This test is performed at low temperatures on PAV residue of binders with creep stiffnesses between 300 and 600 MPa. The failure strain (ε_f) is calculated and defines the load where the failure stress reaches a maximum. Failure stress is defined as the ratio of failure load and original cross-sectional area (36 mm^2). The SHRP specification requires that the failure strain be at least 1 percent.

The recently completed NCHRP Project 9-10 recommended significant changes to the Superpave binder specification for modified binders (*13*). These changes addressed deficiencies in the original specification that included a lack of consideration for the following:

- storage stability,
- additives used in modification,
- the effect of non-Newtonian behavior on mixing and compaction temperatures,
- damage accumulation from repeated traffic loading,
- pavement structure effects,
- traffic speed (other than grade shifting), and
- the effect of cooling rate and variable glass-transition temperatures on low-temperature behavior.

As part of NCHRP Project 9-10, researchers developed screening tests to evaluate storage stability and additives. Based on an extensive laboratory study involving binder and mixture testing, they also recommended new binder parameters to improve characterization of the binder contribution to the three primary forms of asphalt concrete distress. These new parameters are the viscous component of creep stiffness (G_v) measured in a repeated shear creep test at high temperatures, the number of cycles to crack propagation (N_p) measured in a repeated shear controlled stress test at intermediate temperatures, and the critical thermal cracking temperature based on both failure stress and failure strain criteria at representative cooling rates. Researchers also developed new procedures for determining glass-transition temperature and mixing and compaction temperatures for modified mixtures.

NCHRP 9-10 researchers recommended a three-level grading system to accommodate different levels of reliability and available data. Level 1 is based only on environmental conditions, with Level 2 also incorporating traffic conditions. Environmental conditions, traffic speed and volume, and pavement structure are all considered in Level 3. Other recommendations included changes to mixture testing procedures. For binder QA purposes, measurement of the new binder parameters after short-term aging in the RTFO was suggested. As implementation of these results following a field validation experiment proceeds, further changes to the TxDOT binder QA program may become necessary.

MODELS RELATING BINDER PROPERTIES TO PERFORMANCE

The literature on models relating binder properties to performance is extremely limited (*9*). Most researchers recognize the need for these types of models for a number of different applications, but robust models are not available at this time. One limited study conducted at the University of Nevada, Reno, produced a report by Stephane Charmot titled, "Pay Adjustment Factors for Superpave Performance Graded Asphalt Binders," that provides the following (*17*):

- recent models that relate Superpave binder properties to mixture performance, and
- pay factors associated with inadequate performance for each type of distress (rutting, fatigue cracking, and low temperature cracking) due to off-target Superpave binder properties.

The Nevada Department of Transportation (NDOT) developed a pay factor system based on Charmot's results (*17*). Key economic factors in developing such a system include inflation, discount rate, and analysis period. In Charmot's life-cycle cost analysis, a discount rate of 4 percent with no inflation was used over an analysis period of 30 years for rutting and fatigue or 22 years for low-temperature cracking.

Charmot analyzed mixture performance test results and binder test results gathered during the SHRP validation studies. He then developed pay factors due to an inadequate binder based on a methodology that incorporates the following two alternatives, one when an adequate binder is used and one when an inadequate binder is used:

- calculation of total present worth,
- transformation to an equivalent uniform annual cost, and
- conversion to a total cost over the expected performance life.

The difference in total costs as a percent of binder cost is then subtracted from 100 to determine the pay factor. Maintenance costs, user costs, and nonuser costs were not considered in the life-cycle cost analysis because they were considered equivalent for both the adequate and inadequate binder scenarios. Only rehabilitation costs were considered affected by a reduction in performance life. A brief discussion of the data used for each primary form of distress follows:

- Rutting: Mixture resistance to rutting was defined as the number of Repeated Simple Shear Test at Constant Height (RSST-CH) cycles to 2 percent permanent shear strain after short-term oven aging. Binder rutting performance was assessed by $G^*/\sin\delta$ values after RTFO. The RSST-CH cycles were converted to Equivalent Single Axial Loads (ESALs) using the SHRP relationship. The sensitivity analysis showed the rutting pay factor model to be stable, with the most significant effect from HMA thickness.
- Fatigue Cracking: Mixture resistance to fatigue cracking was defined as the number of cycles in the flexural beam fatigue test (68 °F [20 °C], 10 Hz) to reduce the flexural stiffness by 50 percent after short-term oven aging. Binder fatigue performance was assessed by $G^*\sin\delta$ values after RTFO and after RTFO and PAV. The sensitivity analysis showed the fatigue cracking pay factor model is also stable, with the most significant effect from HMA thickness as expected for this type of distress.
- Low-Temperature Cracking: Mixture resistance to thermal cracking was measured in terms of a transverse cracking index after seven years for six test pavements in Pennsylvania. Binder low-temperature cracking performance was assessed by S values and m-values at -29.2 °F (-34 °C) after RTFO and PAV. Maintenance costs had to be considered for this type of distress. Two different sets of pay factors were developed based on the two different binder properties. The sensitivity analysis showed the low-temperature cracking pay factor model is very stable, with the most significant effect from HMA specific gravity.

In the absence of identifying other viable models, the resulting models from this study were considered when the benefits of recommended changes to the TxDOT binder QA program were evaluated.

FACTORS AFFECTING BINDER PROPERTIES PRIOR TO CONSTRUCTION

A possible limitation of the current TxDOT binder QA program is the inability to account for binder properties that may change between production (at the supplier location) and use during construction (in the field). A number of factors may affect or cause these changes. Based on the literature review, Table 1 provides a list of these factors that can be separated into three categories based on the location of the binder (Figure 1) during its journey from production to use during construction. The highlighted factors in Table 1 were selected for inclusion in a laboratory testing program to identify factors that have the most impact on measured binder properties that may change between production (supplier sample) and use during construction (field sample).

Table 1. Factors That May Affect Binder Properties Prior to Construction.

Category	Factors *
Supplier Location	Storage Time
	Storage Temperature (Overheating)
	Blending
	Changing Crude Source
	Refinery Process (Temperature and/or Pressure)
	Contamination in Tanks
Transportation	Contamination in Tanks
	Overheating
Contractor Location	Storage Time
	Storage Temperature (Overheating)
	Contamination/Mixing Different Binders
	Separation
	Dilution
	Presence of Modifier

** Factors that appear shaded were selected for inclusion in the experiment design for this project.*

Aging is one critical effect caused by extended storage time at elevated temperatures. This effect is generally the result of one or more of the following six processes, rendering an increase in the binder stiffness and resulting in a brittle material with reduced resistance to cracking (*18*):

- oxidation,
- volatization,
- thixotropy,
- polymerization,
- separation, and
- syneresis.

The most important processes in terms of the factors suggested in Table 1 are steric hardening (thixotropy), volatization, and oxidation. Researchers anticipate that the effect of aging resulting from these processes is one of the primary mechanisms causing changes in binder properties from production to use during construction. This expectation was confirmed in the laboratory experiment conducted as part of this project. Researchers expect other primary effects to be related to contamination or mixing of different materials either in the blending or modification process.

Physical and/or chemical changes in properties are a particular problem with polymer-modified asphalts. Most researchers believe excessive heating will cause certain polymers to depolymerize (partially) into monomers that have very low viscosities. The result may be that an expensive modified asphalt required because of its superior properties may be placed in construction with properties commensurate with a lower grade that will result in poor performance. Increased storage temperature is one of the factors explored in the laboratory experiment described in a subsequent chapter.

BINDER QA PROGRAM IN TEXAS

Currently, TxDOT samples and approves asphalt materials at the source based on procedures set forth in October 1998 (*19*). The source is defined as either the production site (refinery) or the supplier terminal, and the TxDOT procedures use the terms supplier and producer interchangeably. Prior to the approval process by TxDOT, the supplier must provide test results that indicate specification compliance. In addition, TxDOT samples materials for QA testing according to Test Method Tex-500-C, with the supplier present (*20*). TxDOT obtains samples from tanks if batched or as transports are being loaded if blended. TxDOT may also sample transports on a random basis prior to departure from the production site or the supplier terminal. The TxDOT Asphalt Branch of the Materials Section, Construction Division, subsequently referred to as the TxDOT laboratory in Austin, conducts as many tests on these supplier samples as deemed necessary to verify specification compliance. These verification tests constitute the current TxDOT binder QA program. TxDOT covers costs for all materials that meet the specification and by the supplier if a material fails to meet the requirements. If transport samples fail, TxDOT cancels shipment rights for the originating tank. TxDOT approves asphalt cements for up to 60 days and liquid asphalts (asphalt emulsions and cutback asphalts) for a maximum of 30 days. Advance acceptance prior to verification or QA testing is also possible if the supplier has established a QC plan and a good record of compliance, defined as test results for three consecutive samples verified by TxDOT through QA testing and provision of acceptable test results by the supplier. TxDOT can withdraw this privilege if a sample does not meet the specification.

In addition to the established QA program that relies on monitoring the quality of binders at the supplier source, a program of random sampling in the field by TxDOT districts has also been suggested to increase overall binder quality (*21*). Guidelines for taking samples as close to the point of use as possible, making the contracting community aware of the program in advance, detecting any problems early in the project, and giving priority for completing the QA testing were presented in a May 1999 memo from Mr. Michael Behrens to all district engineers (*21*). Testing may take place at either the TxDOT laboratory in Austin or in a district laboratory that has the capability to conduct the required tests. In addition, the May 1999 memo states that all remedial actions for noncompliance with specifications are available, including pay-factor adjustments.

TxDOT does not require the field sampling QA program at this time, but suggestions made to the district engineers stem from recent attempts to revise the asphalt binder specification for PG asphalts to include QC/QA testing of samples taken as close to the point of use during construction as possible. Provisions for bonus/penalty pay-factor adjustments were also explored. Three draft versions that include these types of revisions were proposed over a two-year period from 1996 to 1998 (*22, 23, 24*). Figure 2 highlights the similarities and differences of the three draft versions.

The first version requires obtaining four samples per day during construction and includes both a bonus and penalty pay structure for compliance over the entire project and noncompliance within specific limits for part of the project, respectively. For preconstruction, the contractor is required to provide a complete set of test results indicating specification compliance. The TxDOT laboratory in Austin then conducts verification testing and bears the cost of this process. If the specification compliance is not confirmed, the contractor supplies a second sample and complete set of test results to TxDOT. For the second round of confirmation testing, the contractor bears the costs.

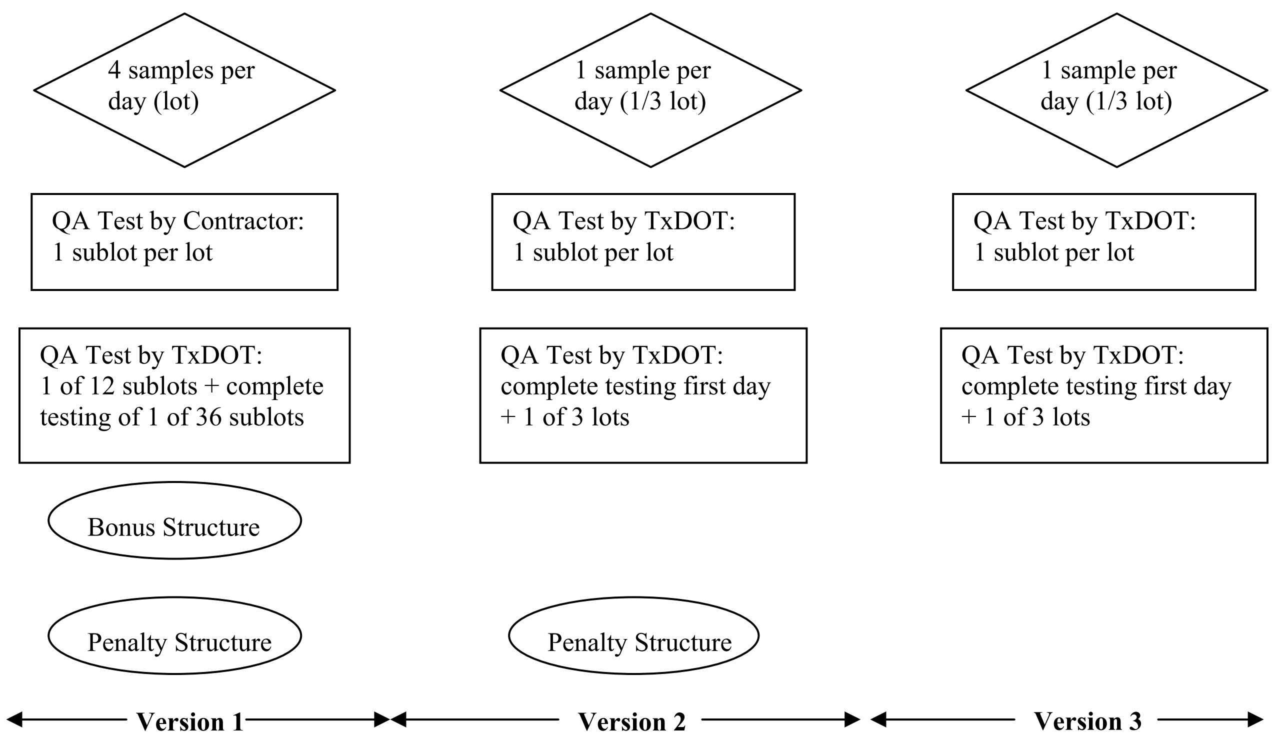


Figure 2. Previously Proposed Binder QC/QA Programs for TxDOT.

During construction, the specification requires that samples be taken and labeled as specific lots and sublots. A lot in the sampling plan is defined as the amount of binder used during one day's production of HMA for a specific project. Each lot contains four sublots. The contractor samples materials with TxDOT personnel present. In this version, QC testing by the contractor is optional, and QA testing by the contractor is required. The QA testing requires the contractor to determine the rutting parameter ($G^*/\sin\delta$) from DSR results after short-term aging in the RTFO for one sublot per lot selected at random (*12*). TxDOT district laboratories conduct verification testing for this high-temperature rutting parameter on a minimum of one out of every twelve sublots. For one out of every 36 sublots, the TxDOT laboratory in Austin conducts complete specification verification. Pay-factor adjustments are then determined based on the high-temperature properties as measured in QA testing if the contractor QA results and the verification results are consistent according to a specified maximum difference. If the results differ by more than this maximum, the remaining sublots in the lot in question are tested, and either an agreement is made to use all of the QA tests or all of the verification tests to characterize the lot or referee testing is undertaken by the TxDOT laboratory in Austin. A schedule is also provided to allow accumulation of penalty pay factors based on the RTFO-DSR rutting parameter.

The second version of the proposed QC/QA specifications reduced the number of samples per day to one and eliminated the bonus pay-factor adjustment. The only change made in the third version was to eliminate pay-factor adjustments altogether. Other changes from the first version in both subsequent versions (two and three) included a definition of a lot in the sampling plan as three consecutive sublots with one sublot sampled each day and required QA testing to be conducted by TxDOT instead of the contractor. Required QA testing includes determination of the RTFO-DSR rutting parameter for one sublot per lot selected at random. TxDOT also conducts confirmation testing on the first day of production and for a minimum of one for every three lots thereafter. This testing includes all tests to ensure complete specification compliance. Penalty pay factors in the second version are adjusted based on QA testing by lot, unless QC testing conducted by the contractor can isolate a particular sublot in the lot classified as noncompliant. The maximum allowable difference in the QC and QA test results is 0.5 kPa in this version. Other than the changes noted, the second and third versions replicate the first version.

After evaluation of each of these versions of possible QC/QA specifications, TxDOT decided that this type of specification required excessive administration and that district personnel were not available at the time (*21*). As a result, the decision to implement field sampling in a QA program was left to the individual districts and was not required. As part of this project, changes to the current QA program, such as those presented in the three draft versions described were evaluated (*22, 23, 24*).

Survey of TxDOT Districts and Suppliers

Two evaluation surveys for TxDOT district personnel and binder suppliers were developed (*1*). The survey questions addressed satisfaction with the current TxDOT binder QA program, suppliers and contractors for each district, and sampling and testing of binders including resources and commonly failed tests. These surveys were faxed to all TxDOT districts and suppliers that serve Texas after contact was made by phone. No surveys were received from suppliers, but 14 out of 25 TxDOT districts responded.

An appendix of an interim report contains a summary of the TxDOT district survey responses in tabular form in a common format for ease of comparison with survey results from state DOT personnel responsible for the overall binder QA program (*1*). This interim report also contains many tables that further highlight the similarities and differences between the perceptions of the 14 TxDOT districts. This section presents a summary of the results from this TxDOT district survey.

Approximately half of the districts are satisfied with the current TxDOT binder QA program, and half are not. Districts that at least take field samples from some suppliers believe the program is fair and achieves a stated goal of obtaining the material as specified on the road in order to produce asphalt concrete that lasts its intended design life. The districts were not asked specifically to identify the goal of the current TxDOT binder QA program, so an assessment of district understanding of the primary motivation behind the program could not made. Five districts, including two that do not currently take any field samples (Beaumont [BMT] and Bryan [BRY]), think that the program is not fair, and a total of seven districts feel that the current program does not achieve its goal. Wichita (WFS) is an anomaly in assessing the program as fair but unable to achieve its goal. BRY cites infrequent testing as a reason for its assessment, and this district has a field sampling program in development. El Paso (ELP) suggested that guidelines needed to be developed for materials that fail the specification. Four districts identified the current program as ineffective, and three other districts cited the lack of contractor QC as a shortcoming of the existing program.

The survey indicated confusion among the districts in term of responsibility for a quality product following construction. Half of the responses indicate the contractor is responsible, while four districts accept the responsibility as the DOT. Two districts spread responsibility between the contractors and the suppliers, and BRY splits responsibility between the contractor and the DOT. According to the survey of TxDOT personnel who oversee the binder QA program, responsibility transfers from the contractor following construction and acceptance by the DOT. To improve the program, the primary goals and responsibilities should be clear to all involved.

Some districts that take field samples do not have a laboratory, and they send their samples to Austin for testing. Amarillo (AMA) and Lufkin (LFK) have the largest laboratory testing capabilities, testing field samples from some or all suppliers, respectively. Most of the other districts with a laboratory utilize one or two technicians for binder testing.

Eight of the 14 districts surveyed collect field samples from all suppliers, and three districts collect these samples from some suppliers. Eleven of the districts indicate that DOT personnel or a contract employee hired by the DOT take the sample, and three districts specify that the contractor is also present. Eight districts respond that these personnel undergo some training. Most samples are taken from either the contractor storage tank or closest to the point of use, in-line at the HMA plant. Sampling frequencies vary by district from daily to monthly and from once per truckload to once per project or as requested.

All TxDOT districts with laboratories utilize AASHTO equipment and test standards when testing binder field samples. Three districts indicate that a QA officer is in charge of calibrating the equipment at least on a yearly basis. Six districts cite an annual calibration frequency, and two other districts calibrate more frequently. Eleven district laboratories have DSR equipment, with 10 also having a RTFO. One district also has the Brookfield viscometer and penetration equipment. Three districts also have Brookfield viscometers, and three different district laboratories contain penetration equipment. Nine districts use the DSR and RTFO equipment for obtaining high-temperature properties before and after short-term aging, and a few other districts utilize penetration equipment and Brookfield viscometers. Testing usually includes an abbreviated program based on available equipment, and frequencies vary by district from daily to weekly or once for every five samples. Only Houston (HOU) conducts multiple replicate tests, while the other districts utilize single replicates to check for specification compliance.

Half of the districts surveyed have relatively low failure rates (0-3 percent) for field samples, with 90 to 100 percent of district test results in agreement with the supplier results contained in the current binder QA program. Three districts indicate that retesting of the same sample is the prescribed procedure if a test result does not satisfy the specification. Only two districts (Dallas [DAL] and ELP) specified a test (DSR) for PG asphalt cements where the material fails to meet the specification most often. No tests were cited for asphalt emulsions. These results are offered, taking into account the fact that these districts only conduct limited testing of field samples.

Contractor Visit and Interview

A visit with Bill Thomas of Young Brothers in Bryan focused on the concerns and responsibilities of HMA plant owners in relation to the binder QA program in Texas. As the binder QA program in Texas is now formulated, HMA plant owners are not involved in binder acceptance testing. They assume that the binder purchased from the supplier meets the required specifications. Young Brothers has three binder tanks, and they generally use the material in a single tank over a 24-hour period. Generally, they only use one grade of binder in HMA production at a rate of 220 tons per hour. They only utilize one supplier, and they do not conduct any binder tests. They report tracking numbers for the binder printed on the work orders obtained from the suppliers to the TxDOT district.

BINDER QA PROGRAMS IN TEXAS AND OTHER STATES

In addition to the evaluation surveys of TxDOT districts and Texas (TX) binder suppliers, additional information was gathered through a two-part phone survey of state DOTs, including TxDOT. The goal of this additional information search was to collect general and then detailed information from binder QA programs in both Texas and nine other selected states. States were selected based on contacts or others suggested by these contacts. The more general survey involved collecting general information, any documentation including specifications, and a sample data set (over a one-year period). Information gathered in the more detailed survey through multiple phone conversations and e-mail included the following:

- contact information;
- general satisfaction, goals, and shortcomings;
- responsibility for premature failures;
- size of the program (number of major suppliers, major contractors, laboratories, technicians);
- impact on suppliers and contractors;
- general sampling, testing, and handling requirements and output;
- DOT sampling and testing of both supplier and field samples;
- equipment;
- specification compliance requirements;
- pay factor / penalty systems;
- cost estimates; and
- analysis of results.

A detailed review and analysis of the statistical validity of each state binder QA program was not completed as initially proposed because of time and resource limitations. TxDOT may pursue this type of analysis through an ongoing statistical support contract or a multi-year project focused specifically toward achieving this goal.

An appendix of an interim report contains a summary of the state DOT survey responses in tabular form in a common format for ease of comparison with survey results from TxDOT districts (*1*). This interim report also contains many tables that further highlight the similarities and differences between the 10 state binder QA programs. This section presents a summary of the results from this state survey.

All of the states except TX are satisfied with their current binder QA program. Following completion of this project, TxDOT's satisfaction with their program is expected to improve. The goal of all of the states' programs is to obtain the material that was specified on the road. Minnesota (MN) also cited a secondary goal of saving time and effort through a coordinated program where multiple states share certification and inspection of suppliers. California (CA) indicated that there must also be no delay in construction caused by the binder QA program. A specific goal of the TX program is to promote fairness to all parties through a program that requires minimum resources. In most of the states, responsibility for a quality product transfers from the contractor following construction and acceptance by the DOT. The DOT is then responsible for premature failures, usually after the first year in service. This system works well in many of the states where penalties are assessed to the contractors based on an estimate of the difference in performance of the as-constructed and as-designed or as-specified pavement. Most of the states felt that their program was fair to contractors, but many questioned the issue of fairness with respect to suppliers because any penalties assessed the contractors are usually passed on to the suppliers even if the there is a lack of QC during transportation or at the contractor location, a problem cited by half of the states. Three states plan to introduce or expand a required contractor QC plan in the near future, and one state recognizes that resources are not available to maintain this type of system.

TX and CA have the largest number of major binder suppliers, but many of the other states have larger laboratory testing programs in terms of the number of laboratories and the number of technicians assigned to the binder QA program. The workload in terms of number of tests per year varies from state to state and is difficult to compare because of differences in sampling and testing frequencies and abbreviated testing requirements. Most states with large testing programs require testing of field samples for acceptance by the DOT. Testing of supplier samples is left to the suppliers themselves and is required in almost all of the states, although each state differs in terms of the frequency of complete and abbreviated specification compliance testing. In some of the states, the DOT tests supplier samples at the beginning of the season, for new binders, or in special situations. Currently, the TX system is opposite, requiring DOT testing of supplier samples and no regular system of testing field samples.

In terms of sampling either supplier or field samples for testing by the DOT, most states allow the contractor to take the sample according to AASHTO guidelines with a DOT witness present. In Nevada (NV) and Oregon (OR), this witness is not required but is present some of the time in OR. In CA, DOT employees take samples, and these personnel are trained, as they are in half of the states surveyed. In TX, neither the contractor nor the DOT is present; TxDOT hires a contract employee with no formal training to take supplier samples. Most states take field samples from either the contractor storage tank or closest to the point of use, in-line at the HMA plant, or from the emulsion distributor truck. Sampling frequencies are also adjustable in some states to account for a continued record of compliance or noncompliance or to adjust the laboratory workload.

Most of the states, including TX, require supplier QC plans, but currently only Utah (UT) requires some form of a contractor QC plan. Thus, the impact on contractors is minimal in most states unless construction is shut down for a serious problem that may be related to binder quality. Different states have different supplier requirements that may include an annual inspection, certification, or an advance acceptance program to reduce delays. Certification in two states is good for a combined group of states, reducing the number of resources required for each individual state.

All states surveyed utilize AASHTO equipment and test standards when testing binders in their QA programs. Laboratories are AASHTO accredited through the efforts of one or two people in the majority of states. A complete set of testing equipment is found in the central laboratory in each state, while regional laboratories may only have a limited set of equipment. Less than half of the states have a formal technician training program. Three states participate in round-robin testing programs, and four states allow for adjustment of testing frequencies. Testing frequencies vary by state, with some samples remaining untested, some undergoing an abbreviated specification compliance testing program, and others subjected to a complete testing sequence.

Single replicate test results are compared to specification limits that include tolerance intervals in half of the states evaluated. The basis for these tolerances is different for each state, ranging from proficiency or round-robin test results to ASTM or AASHTO precision and repeatability statements. Each state defines compliance and rejection limits in a schedule. The other half of the states, including TX, do not allow for any variability in the result from the specification limit. In these states, the supplier is expected to account for any variability and ensure that the specified value can be met.

Most states are satisfied with their binder QA program, as illustrated by their relatively low failure rates (less than 5 percent), especially for PG asphalt cements. Each state prescribes a different procedure following failure of a material to meet a specified test, but the majority require retesting the same sample and testing of samples immediately surrounding the failed sample. These results are used to estimate the quantity of material out of specification for calculation of penalties through pay factors. A few states compare failed test results with other results from the supplier, round-robin testing programs, or AASHTO repeatability limits. Complete resampling and retesting or testing by a third party is another less-common option in a few states, with the supplier or the contractor paying for testing of noncompliant material in TX and UT, respectively. Tests for PG asphalt cements, where the material fails to meet the specification most often according to the survey results, include the DSR on unaged or short-term aged (in the RTFO) and a toughness and tenacity test the intermountain west states (Colorado [CO], NV, and UT) include in their PG+ specification. For asphalt emulsions, a number of states cited Saybolt viscosity and penetration of the residue as tests where the material most often fails the specification.

When materials fail the specification, pay factors are calculated in seven of the 10 states. Pay factors are also determined in Maryland (MD), although there is no formal system. The two states without pay factors (CA and TX) are also the largest states that probably use the largest volume of binders in asphalt construction per year. Issues associated with these large states may partially explain the lack of a formal pay-factor system. Penalties are assessed based on only one binder property in four states and on an accumulation of failing binder properties in three states. Often dependent on the materials involved, properties measured, environmental conditions, and facility type, each state uses different schedules and equations to determine the penalty assessed of the contractor.

The final category analyzed through the detailed survey of state binder QA programs was the analysis of benefits and costs. In all states, the main use of the data is to allow the DOT to accept the material and responsibility for use in asphalt pavement construction. The majority of states use an electronic database for a variety of purposes, including forensic analyses and historical analysis of the quality of materials from each supplier and the performance of different binders. Other benefits cited include improved communication with suppliers, laboratory assessment, research, and the ability to track binder use and costs to the state. The larger states of CA and TX currently have inadequate databases that do not allow for some of these benefits. Again, increased resources are required, but creation of electronic databases is forthcoming in all states surveyed. No detailed cost information and therefore benefit to cost (B/C) ratios were available from any of the 10 states. Only a qualitative sense of confidence is obtained in all states, except TX. Unfortunately, none of the 10 states has any quantitative confidence estimate of the quality of material utilized in asphalt pavement construction due to limited resources and the lack of a need to quantitatively justify the program. Four of the states highlighted the fact that their binder QA programs attempt to balance resources while at the same time assessing the quality of materials used in asphalt pavement construction and qualitatively obtaining a sense of confidence in these materials.

CHAPTER 3. ANALYSIS OF EXISTING DATA

In addition to the qualitative comparison of binder QA programs documented in the previous chapter, existing binder data from three states were quantitatively evaluated. Data received through the information search from Colorado and Oregon was statistically analyzed using cluster analysis to compare test results required by specification to their corresponding specified values. A different classification tool called classification and regression trees (CART) was used to statistically analyze existing data from Texas toward the same goal. This second type of analysis was pursued with the Texas data because the cluster analysis did not produce meaningful results for TxDOT's use in decision making. This chapter provides descriptions of these two analysis methods, followed by a summary of the resulting classifications and their implications. More detailed results are presented in an interim report (*1*).

CLUSTER ANALYSIS

Two approaches were used to examine the Colorado and Oregon data through cluster analysis. One approach compared each individual test result with its required value in the specification. The second approach compared all test results to their specified values simultaneously. The analysis also focused on statistically describing results from each test and the collection of tests for each type of binder material. For each test, central tendency, variation, and shape and type of the distribution of results were examined through graphical and mathematical techniques. The focus of this analysis was to show, using data from the other states, what information can be obtained if the Texas data included results from field samples stored in an organized, easily accessible manner. One goal was to understand the variability to facilitate establishment of a rational basis for pay factors and determination of the confidence level that the material used meets the specification.

As a first step with the Colorado PG data set, correlation of different binder test parameters was explored to aid in selecting those most relevant for use in a QA program. Then statistical distributions of the selected parameters were examined using kernel estimation, a nonparametric smoothing method. This initial analysis showed bimodal distributions, with one group of measurements that generally exceeded the specification in one mode and a second group of measurements clustered around the specified value. As a result of the multimodality of the data, cluster analysis was chosen as a more appropriate tool.

Cluster analysis is an exploratory data analysis tool for solving classification problems. Its objective is to sort cases into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. Each cluster thus describes, in terms of the data collected, the class to which its members belong. As a result, cluster analysis can reveal similarities in data that may have been otherwise impossible to find.

The results from cluster analysis can be used in several ways. Cluster analysis aids in the identification of outliers (observations lying very far from the main body of the data) by assigning them to one cluster. These outliers may be the result, for instance, of measurement errors or typing errors made while entering the observations into a database. Outliers can be discarded so as not to affect the result of the analysis. When future QA tests are conducted, they can be assigned to clusters, enabling prediction of tests that might cause problems and whose results should therefore be examined more closely. This assignment can be done using different statistical procedures to find a cluster where observations have relationships between variables similar to the one under investigation. For experimental design purposes, clusters can be used as blocks. Thus, it would be important to pick an equal number of samples from each cluster to make the analysis less biased and to reduce supplier-to-supplier variability. Other anticipated advantages of this type of analysis include identification of materials (and corresponding suppliers) that consistently fail specific property requirements.

Cluster analysis groups observations so that the observations in each group are similar with respect to the clustering variables. The various clustering techniques fall into two categories, hierarchical and nonhierarchical. Hierarchical cluster analysis is an iterative procedure. Initially, each data point is a cluster. In each succeeding step, the two "closest" clusters are merged, reducing the total number of clusters by one. This continues until there is only one cluster, or the desired predetermined number of clusters is reached.

Determining which clusters are "closest" requires a measure of the distance between clusters. The various hierarchical clustering algorithms differ mainly in the way they compute distance. Sharma (*25*) gives a summary of the various clustering algorithms together with the empirical studies comparing the performance of different clustering algorithms. From the survey, it appears that single-linkage, average-linkage, and Ward's method perform best.

For this analysis, Ward's method was chosen. Ward's method does not compute distances between clusters, but rather forms clusters by maximizing within-cluster homogeneity.

The main problem with all hierarchical clustering methods is that only observations with complete data can be used with the program utilized. In this project, 62 percent of the observations have missing data, so these methods are of limited use.

In nonhierarchical clustering, the data are partitioned into *g* predetermined clusters. This requires that the researcher have some *a priori* knowledge of how the data will cluster. This is usually obtained by clustering the data using one or more hierarchical techniques. Observations with missing data can also be handled. Once the cluster centroids or seeds are identified, observations are assigned to the seed closest in value based on available information.

CART ANALYSIS

Existing data from Texas were statistically analyzed using CART, with the majority of records labeled Pass, Fail, and For Information Only. This type of analysis was used to develop simple rules that produce classification trees and corresponding classes with these three labels. For each type of material, several critical properties ($x_1, \ldots, x_p$) were identified and used in the CART analysis to decompose the data using binary (two way) splitting rules. In each of the resulting subsets of data, a majority-voting rule determined the class label (Pass, Fail, or For Information Only). For example, a splitting rule of ($x_1 \leq 150$) meant that all data with x_1 values less than or equal to 150 were assigned into one class, and the remaining data were assigned to another class. The most common label in the subset determined the overall class label. For example, the 38 Pass labels in 50 cases with ($x_1 \leq 150$) identified this class as Pass. CART recursively split and resplit the properties until a simple tree was produced that accurately reflected the classifications in the existing database, if possible. An example output tree for the Cationic Rapid Setting (CRS) CRS-2 emulsion material from this analysis is shown in Figure 3, with Saybolt2 indicating the Saybolt viscosity measured at 122 °F (50 °C).

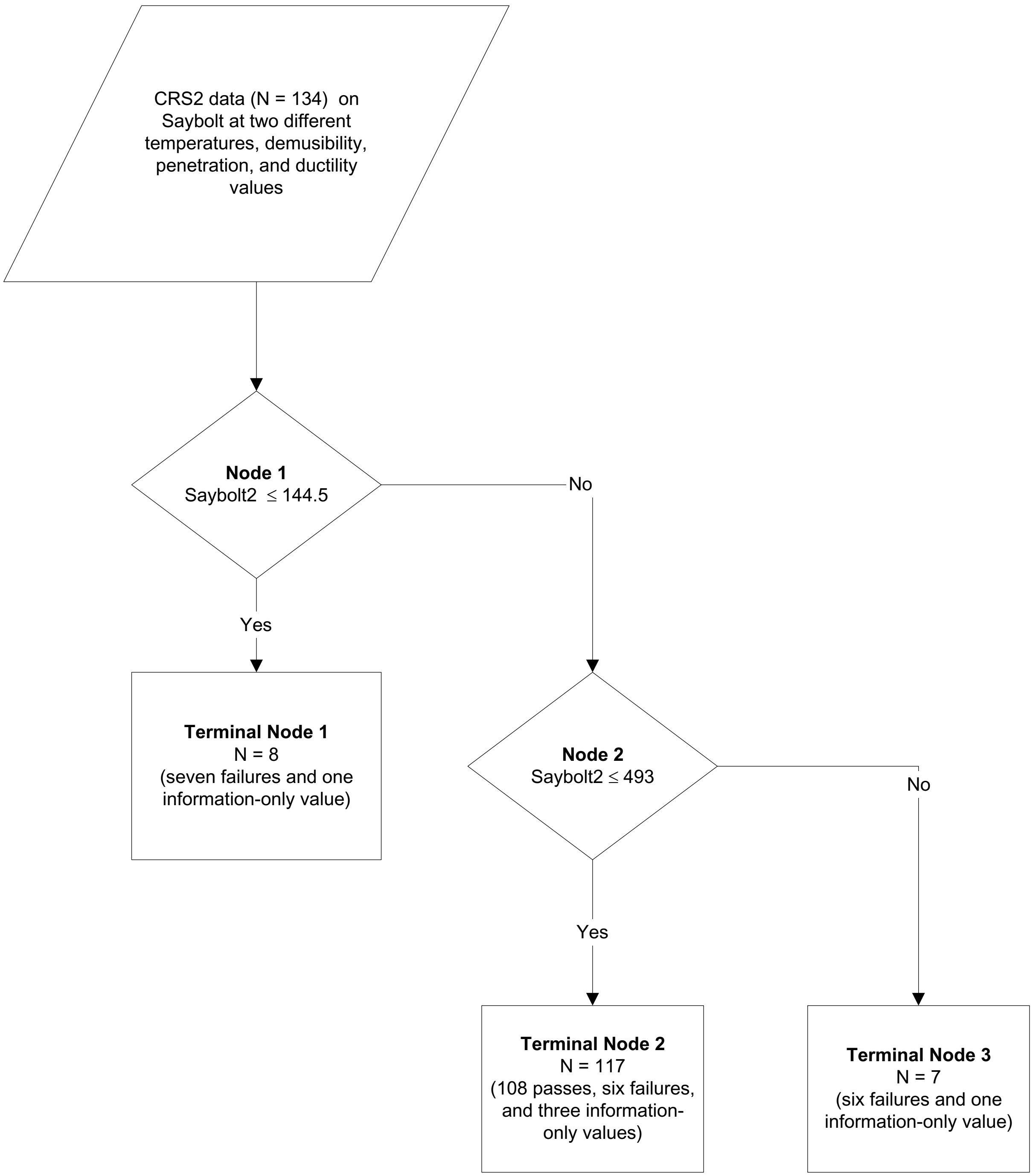


Figure 3. CRS-2 CART Tree.

COLORADO

Colorado PG data covered a one-year (2000) time period and consisted of the results from three QA tests (DSR, RTFO-DSR, PAV-DSR) for eight different binders (PG binder grades labeled Binders 1-8) produced by twelve suppliers (Supplier A-M, without Supplier I, to avoid confusion with J). DSR represents the $G^*/\sin\delta$ value measured on an unaged binder. RTFO-DSR is used for the $G^*/\sin\delta$ value for a short-term aged binder, and PAV-DSR indicates the $G^*\sin\delta$ value measured on a binder that has been both short-term and long-term aged. Of the 577 observations, only 217 had complete data. The DSR data were missing from some observations, but this test was performed more frequently than either the RTFO-DSR or PAV-DSR tests.

To standardize the data, each property was transformed in the following manner:

$$std.value = \frac{value - spec}{spec}$$

where:

spec	=	the specified value for the test,
value	=	a test result, and
std.value	=	the standardized test result for further analysis

The standardized QA test results for DSR, RTFO-DSR, and PAV-DSR were then relabeled as STDSR, STRTFO, and STPAV, respectively. These standardized results must all be greater than zero to meet the specification.

The goal of this analysis was to separate suppliers based on the quality of their binder. Hierarchical clustering with Ward's method was used to identify the number of clusters and cluster seeds. Then, a nonhierarchical cluster solution was obtained for the data. One obvious outlier for RTFO-DSR was identified and deleted. This outlier might be due to an error when results were typed into the database. The number of clusters was chosen to be four based on several statistics that measure cluster homogeneity. By comparing descriptive statistics for each cluster to those for the entire data set, the following observations are offered:

- Cluster 1 contains below-average STDSR values, below-average STRTFO values, and above-average STPAV values.

- Cluster 2 contains STDSR and STRFO values far below average and below-average STPAV values.
- Cluster 3 contains STDSR and STRFO values far above average and STPAV values far below average.
- Cluster 4 contains above-average STDSR values and STRTFO and STPAV values far above average.

In summary, the best cluster is Cluster 4, and the worst cluster is Cluster 2 based on the number of failures or results not passing the specification. Cluster analysis could not locate all failures into one cluster. The first three clusters have observations with failures. Cluster 2 has all 10 STDSR failures; one-third (seven out of 21) of all STRTFO failures are in Cluster 1, and the remaining two-thirds (14 out of 21) are in Cluster 2. Two-thirds of all STPAV failures (two out of three) are in Cluster 2, with one-third (one out of three) in Cluster 3.

Approximately 35.5 percent of all tests were grouped into Cluster 1, 40 percent into Cluster 2, 9.5 percent into Cluster 3, and 15 percent into Cluster 4. Most of the suppliers have observations in each cluster. Almost all suppliers, except G, K, and M, have the majority of observations in the first two clusters, ones that reflect bad (compared to other clusters) performance for STDSR and STRTFO. For Suppliers A and C, more than 50 percent of the observations are in Cluster 2, the worst cluster. Some suppliers, like Supplier F, have a significant percentage in every cluster, which might indicate unstable performance (test results vary significantly). This can be explained by the fact that for some suppliers, performance changes by binder. Therefore, cluster analysis for each binder was also conducted separately to compare the results. To be consistent in this secondary analysis, the number of the clusters was chosen to be three, based on several statistics that measure cluster homogeneity for the data for Binder 1. Detailed clustering analysis results for each binder are presented in an interim report (*1*).

In summary, cluster analysis resulted in a good separation of suppliers (i.e., observations from one supplier belong mainly to one cluster) if there was high correlation among variables in the data set. When the correlation among variables was low, cluster analysis did not seem to be useful in that there was not a good separation of suppliers (i.e., observations from one supplier were evenly split among two or more clusters).

Suppliers were separated into three well-defined groups using statistical clustering methods for each binder. In each group, measured DSR values for all three aging states (original, after RTFO, and after RTFO and PAV) were similar. Thus, groups of suppliers were found more likely than others to be out of specification for a particular binder. With this result, the Colorado archived data provided useful information about the Colorado PG binders and suppliers.

Clustering by binder is recommended because, for some suppliers, performance in terms of specification compliance changes by binder. In addition, this type of analysis may contribute to the definition of a formal classification scheme, indicating rules for assigning new binders to clusters for identification and diagnostic purposes.

OREGON

Oregon emulsion data was also evaluated to determine if cluster analysis could be used to identify materials and corresponding suppliers that consistently fail specific property requirements. Unfortunately, for the data set evaluated, all emulsion test results met specifications, so cluster analysis was not pursued.

TEXAS

In contrast, Texas data cannot be easily used in a binder QA program. After extensive effort to archive data in a usable form, the statistical information that could be extracted was summarized and analyzed using CART. PG64-22 and PG70-22 binder data were analyzed including critical selected properties measured in the DSR (DSR on unaged binder, RTFO-DSR, and PAV-DSR) and the BBR (BBR stiffness *S* and *m*-value). For CRS-2 and CRS-2P emulsions, Saybolt viscosity measured at two temperatures, demulsibility, penetration of the residue, and ductility of the residue were selected as critical properties.

There were 322 data records from 20 suppliers for the PG64-22 data, with some missing values for each variable and all but three records labeled Pass, Fail, or For Information Only. CART analysis produced a classification tree with six classes. Class 6, with a PAV-DSR value greater than 3.5 MPa, contained all three of the Fail values from two of the suppliers, one of 27 For Information Only values, and five of 289 Pass values.

There were 543 data records from 21 suppliers for the PG70-22, data with some missing values for each variable and all but 17 records labeled Pass, Fail, or For Information Only. CART analysis produced a classification tree with three classes. Class 1, with a STRTFO value of less than 0.002 (or a STDSR value less than 0.009 for missing STRTFO values), contained the bulk of the Failures and For Information Only values (eight of 11 Fail and 54 of 79 For Information Only) and only one of the 436 Pass values. Class 2 contained two more of the 11 Fail values, seven additional For Information Only values, and no Pass values. Class 2 required STRTFO values greater than 0.002 (or STDSR values greater than 0.009 for missing STRTFO values) and standardized *m*-values (STM) values greater than -0.002. Class 3 contained the remaining values, including all but one of the Pass values, one Fail value, and 18 For Information Only values. Conclusions from this analysis point to Fail classification based on low RTFO and DSR values. Most For Information Only values grouped with the Fail values, and some suppliers produced an unusually large percentage of Fail and For Information Only values.

There were 273 data records from 15 suppliers for the PG76-22 data, with a typical record labeled Pass (216 values), Fail (1 value), or For Information Only (55 values). The PG76-22 data were not analyzed using CART due to the small number of failures.

There were 134 data records from nine suppliers for the CRS-2 data with a typical record labeled Pass (108 values), Fail (19 values), or For Information Only (five values). Two records labeled Meets Specifications Only were not analyzed. CART analysis produced a classification tree with three classes. Classes 1 and 3 combined contained 13 of 19 Fail values and two of five For Information Only values. Class 1 required Saybolt viscosity values at 122 °F (50 °C) less than 144.5 s if data were available. Class 3 required Saybolt viscosity values at 122 °F (50 °C) greater than 493 s. Conclusions from this analysis point to classification of a Failure based on low or high Saybolt viscosity values. A single supplier with both the largest number (13 of 19) and largest percentage (68 percent) of Fail values was identified. The other Fail values were distributed over five other suppliers (one of eight samples, two of 22 samples, two of 40 samples, and one of three samples).

There were 297 data records from 13 suppliers for the CRS-2P data, but the records were labeled Pass (248 values), Fail (25 values), For Information Only (nine values), Meets Specifications Only (14 values), and Variation from Specifications is Immaterial (one value). Analysis of the CRS-2P data did not produce meaningful classification rules, possibly due to a significant number of data records that were categorized with labels other than Pass, Fail, or For Information Only.

CHAPTER 4. FIELD EXPERIMENT

The initial strategy for evaluating the TxDOT binder QA program was to validate and further examine differences in properties between corresponding supplier and field samples and identify factors responsible for these changes. This chapter describes the design of the field experiment utilizing this initial strategy and the results obtained with its implementation.

Due to anticipated difficulties in obtaining samples from the field and their corresponding supplier test results, an extensive laboratory experiment that utilized supplier samples and simulation of storage conditions and contamination was also designed. The next chapter describes this laboratory simulation experiment.

Based on discussions with TxDOT personnel, specific factors highlighted in Table 1 were selected for inclusion in the field and laboratory simulation experiments to evaluate any changes in properties of asphalt cements and emulsions. Both experiments were designed to identify the factors that have the most impact on RTFO-DSR. This test was selected due to its direct relationship with performance in terms of resistance to rutting in the early life of an asphalt concrete pavement, frequent use as a QA parameter by other state DOTs, and equipment availability in the TxDOT districts. Penetration tests on emulsion samples were also included in both experiment designs because RTFO-DSR properties are not currently measured for emulsion residues, and the initial strategy was to compare supplier and field sample results.

DESIGN

For the field experiment, all of the factors highlighted in Table 1 could not be included in the design because some of them (Contamination and Storage Temperature) were uncontrollable in the field. In addition, the number of test runs needed to be restricted due to difficulties anticipated in obtaining samples from the field and their corresponding test results from the supplier tank. As a result of these limitations, a screening design shown in Tables 2 and 3 was proposed. The factors in the field experiments were Modifier (with two levels: modified PG76-22 [L1] and unmodified PG64-22 [L2] for asphalt cements or modified CRS-2P [L1] and unmodified CRS-2 [L2] for emulsions) and Storage Time (with two levels: more than one week [1] and less than one week [-1]). Storage Time was taken as the sum of the storage times at both the supplier and contractor locations. Storage Temperature was used as a covariate (an uncontrollable variable that influences the response but is itself unaffected by any other experimental factors) in contrast to the laboratory experimental design where Storage Temperature was an additional controllable factor. Supplier (with two levels determined as field samples were identified) was used as a block to increase precision in the estimation of factor effects. The response variable was the percent change in RTFO-DSR before and after each treatment (storage) was applied.

Table 2. Field Experimental Design for Asphalt Cements.

Row	Modifier	Storage Time	Supplier	Storage Temperature
1	Modified (L1)	More than one week (1)	1	
2	Unmodified (L2)	Less than one week (-1)	1	.
3	Modified (L1)	More than one week (1)	1	.
4	Modified (L1)	Less than one week (-1)	1	.
5	Modified (L1)	Less than one week (-1)	1	.
6	Unmodified (L2)	More than one week (1)	1	.
7	Unmodified (L2)	More than one week (1)	2	.
8	Unmodified (L2)	More than one week (1)	2	.
9	Modified (L1)	Less than one week (-1)	2	.
10	Unmodified (L2)	Less than one week (-1)	2	.
11	Modified (L1)	More than one week (1)	2	.
12	Unmodified (L2)	Less than one week (-1)	2	.

Table 3. Field Experimental Design for Emulsions.

Row	Modifier	Storage Time	Supplier	Storage Temperature
1	Modified (L1)	More than one week (1)	1	.
2	Unmodified (L2)	Less than one week (-1)	1	.
3	Modified (L1)	More than one week (1)	1	.
4	Modified (L1)	Less than one week (-1)	1	.
5	Modified (L1)	Less than one week (-1)	1	.
6	Unmodified (L2)	More than one week (1)	1	.
7	Unmodified (L2)	More than one week (1)	2	.
8	Unmodified (L2)	More than one week (1)	2	.
9	Modified (L1)	Less than one week (-1)	2	.
10	Unmodified (L2)	Less than one week (-1)	2	.
11	Modified (L1)	More than one week (1)	2	.
12	Unmodified (L2)	Less than one week (-1)	2	.

RESULTS

An extensive effort was undertaken to locate field samples where all data and storage information was available. The course of action to achieve this goal was to contact the supplier and gather information about shipping of PG64-22 or PG76-22 binders and/or CRS-2 or CRS-2P emulsions to TxDOT projects. The information required for each shipment included Storage Time and Storage Temperature, TxDOT district, project identification (ID), Control-Section-Job (CSJ) number, and the contractor assigned to the job. The Storage Time and Storage Temperature data were most often expressed by experience and common practice rather than by an accurate measurement on each shipment.

The contractor was called next to retrieve information on Storage Time and Storage Temperature while the binder was at their site. As with suppliers, this information was commonly an approximation as no specific measurements were available.

The next step was to contact the TxDOT district office handling the project and ask for the supplier material laboratory number. This number consists of a C, the two last digits of the current year, the number 37, and a four digit serial number. For samples with a specific project or CSJ number, the TxDOT contract administrator or the TxDOT construction office was able to obtain this information from the material test history report. If the supplier lab number was for some reason missing, as was often the case, and the project was still ongoing, it was possible to call the TxDOT area office and ask them to retrieve the number from the shipment delivery ticket where the corresponding laboratory number always appears. Sometimes a project had several supplier laboratory numbers or alternatively more than one project was using material from the same supplier tank. If the district had obtained field samples, the assigned ID on the 202 TxDOT form was requested. This number varies because no specific format is defined for the supplier material. If no samples were obtained and the project was still ongoing, TxDOT personnel were requested to take field samples and send them to the TxDOT laboratory in Austin with a proper identifying note on the 202 form. Finally, a petition was sent to the TxDOT laboratory in Austin to retrieve the supplier and field test results using the supplier laboratory number and the identification on the 202 form.

This search produced only eight results for asphalt cement samples and none for emulsion samples. The main obstacle in finding field sample results and their corresponding supplier test data was the poor system for sample identification. Sometimes field samples results were available, but it was difficult to track them back to their corresponding supplier material approval test results, as their identification provided no link. In other cases with existing supplier test data, no field samples were taken at the time of the project, samples taken were not properly identified on the 202 TxDOT form, or samples taken were lost during transportation to the TxDOT laboratory in Austin.

Table 4 presents a limited set of results for asphalt cements. Row numbers correspond to the factor-level combinations presented in Table 2. RTFO-DSR test results for the supplier samples are identified as X0 and as X1 for field or project samples. The percent difference between these two results is indicated as (X1 - X0) / X0*100.

Table 4. Field Experimental Results for Asphalt Cements (nearest 0.1 kPa).

Row	Supplier Sample X0	Field Sample X1	Percent Change (X1 - X0)/ X0*100
1	2.4	2.3	-3.8
2	3.1	2.8	-9.8
3	2.4	2.0	-16.3
4	2.3	1.3	-42.7
5	2.3	1.1	-106.7
6	4.1	4.2	1.4
9	3.5	3.4	-2.3
10	2.9	3.9	37.9

The majority of these results indicate a negative percent change in the RTFO-DSR result or $G^*/\sin\delta$ parameter. Only two showed an increase. In some cases, the value drops so low that it fails to meet the minimum Superpave grade specification threshold for RTFO samples of 2.20kPa. When comparing rows 1 and 3 as well as 4 and 5, samples from the same supplier used on different projects showed significant variability.

The failure to meet specifications and test result variability among field and supplier samples is a major concern, since this lack of QC can cause early pavement failures and reduced service life. A good example of this problem occurred on a resurfacing project in El Paso County on Loop 375 completed in April 2000. Air blown PG76-16 was used for this project, and within three months of placement, the surface looked aged and gray in color, with extensive fatigue cracking and aggregate loss. A forensic study conducted on that project established that although there were many factors contributing to poor performance, including construction practices and structural design, one factor with significant impact was related to binder quality (*26*). Penetration tests performed on recovered binder from core samples were very low, suggesting that the binder was aged and brittle and therefore unable to perform as designed and bond properly to the existing surface layer.

CHAPTER 5. LABORATORY SIMULATION EXPERIMENT

With insignificant data for the field experiment, the revised strategy for evaluating the TxDOT binder QA program was to analyze differences in properties between supplier samples before and after simulated storage and/or contamination to identify factors responsible for these changes. This chapter describes the design of the laboratory simulation experiment utilizing this revised strategy and the results obtained with its implementation.

As with the field experiment, both asphalt cements and emulsions were evaluated in a laboratory testing program to identify the factors that have the most impact on RTFO-DSR. Penetration tests were also performed on emulsion samples because RTFO-DSR properties are not currently measured for emulsion residues. The Saybolt viscosity test was considered, but this test was not feasible due to the use of unsealed containers that allowed for extensive water loss in the sample after storage at elevated temperatures. Fourier Transform Infrared Spectroscopy (FTIR) testing was also included in the laboratory simulation experiment because of its ability to track chemical compound formations related to binder aging, which helped to better understand and explain RTFO-DSR result fluctuations. In addition, Gel Permeation Chromatography (GPC) tests were performed on emulsions as a quality control check to assure that water was effectively removed by the stirred-can method developed during TxDOT Research Project 0-1710 and that only the binder residue was left behind for further testing (*27*).

DESIGN

The factors for asphalt cements in the laboratory experiment were Modifier (with two levels: modified PG76-22 [L1] and unmodified PG64-22 [L2]), Contamination (with three levels: no contamination [L1], contamination of 100 gallons of a 6000 gallon transport truck [L2], and contamination of 500 gallons of a 20,000 gallon contractor tank [L3]), Storage Time (with three levels: one week [L1], one month [L2], and two months [L3]), and Storage Temperature (with two levels: 335 °F [168 °C] [-1] and 375 °F [191 °C] [1]). In addition to these factors, Supplier (with two levels: Supplier 1 [1] and Supplier 2 [2]) was introduced as a block to remove excess variation due to differences in manufacturing process among suppliers. Each factor-level combination corresponds to a different treatment, and two replicate samples (with two measurements on each sample) were tested for each combination. Prior to treatment (storage at elevated temperature), each asphalt cement sample was fabricated by pouring about 0.13lb (60 grams) of the material into an ointment tin, flushing the tin with nitrogen to simulate storage in a closed tank by precluding aging at the surface, and sealing the lid with a stiff asphalt cement. After treatment, the response variable was measured as the relative difference (multiplied by 100) in RTFO-DSR after storage (treatment) with respect to the initial value obtained from control samples that were not stored at elevated temperatures. Test runs corresponding to treatments were randomized within each block (Supplier) to average out the effects of extraneous factors that cannot be directly controlled. This resulted in an augmented D-optimal randomized block design shown in Table 5 that allows for estimation of all main effects and two-way interactions (*28*).

The factors for emulsions in the laboratory experiment were Modifier (with two levels: modified CRS-2P [L1] and unmodified CRS-2 [L2]), Contamination (with two levels: no contamination [L1], contamination of 100 gallons of a 6000 gallon transport truck [L2]), Storage Time (with three levels: two days [L1], one week [L2], and one month [L3]), and Storage Temperature (with two levels: 150 °F [66 °C] [-1] and 180 °F [82 °C] [1]). Supplier (with two levels: Supplier 1 [1] and Supplier 2 [2]) was used as a block in the design shown in Table 6. Again, two replicate samples (with two measurements on each sample) were tested for each factor level combination (row), and all main effects and two-way interactions were estimated. For the emulsion samples, water was removed to produce a residue by the stirred-can method developed during TxDOT Research Project 0-1710 (*27*). These samples were not sealed because the water vapor released during storage at elevated temperature was hypothesized to preclude aging and simulate storage in a closed tank.

Table 5. Laboratory Experimental Design for Asphalt Cements.

Row	Modifier	Contamination	Storage Time	Storage Temperature	Supplier
1	No (L2)	100 of 6000 gal (L2)	One week (L1)	335 °F (-1)	1
2	No (L2)	500 of 20,000 gal (L3)	Two months (L3)	375 °F (1)	1
3	No (L2)	100 of 6000 gal (L2)	One month (L2)	375 °F (1)	1
4	Yes (L1)	100 of 6000 gal (L2)	One week (L1)	335 °F (-1)	1
5	No (L2)	None (L1)	One month (L2)	375 °F (1)	1
6	Yes (L1)	100 of 6000 gal (L2)	One month (L2)	375 °F (1)	1
7	Yes (L1)	500 of 20,000 gal (L3)	Two months (L3)	335 °F (-1)	1
8	No (L2)	500 of 20,000 gal (L3)	One month (L2)	335 °F (-1)	1
9	No (L2)	None (L1)	One week (L1)	335 °F (-1)	1
10	Yes (L1)	500 of 20,000 gal (L3)	One week (L1)	375 °F (1)	1
11	Yes (L1)	None (L1)	One month (L2)	335 °F (-1)	1
12	No (L2)	None (L1)	Two months (L3)	375 °F (1)	1
13	Yes (L1)	100 of 6000 gal (L2)	Two months (L3)	375 °F (1)	2
14	Yes (L1)	100 of 6000 gal (L2)	One month (L2)	335 °F (-1)	2
15	Yes (L1)	500 of 20,000 gal (L3)	One month (L2)	375 °F (1)	2
16	Yes (L1)	500 of 20,000 gal (L3)	One week (L1)	335 °F (-1)	2
17	Yes (L1)	None (L1)	Two months (L3)	335 °F (-1)	2
18	No (L2)	None (L1)	One month (L2)	335 °F (-1)	2
19	No (L2)	100 of 6000 gal (L2)	One week (L1)	375 °F (1)	2
20	No (L2)	100 of 6000 gal (L2)	Two months (L3)	335 °F (-1)	2
21	No (L2)	500 of 20,000 gal (L3)	One week (L1)	375 °F (1)	2
22	No (L2)	500 of 20,000 gal (L3)	Two months (L3)	335 °F (-1)	2
23	Yes (L1)	None (L1)	Two months (L3)	375 °F (1)	2
24	Yes (L1)	None (L1)	One week (L1)	375 °F (1)	2

Table 6. Laboratory Experimental Design for Emulsions.

Row	Modifier	Contamination	Storage Time	Storage Temperature	Supplier
1	Yes (L1)	None (L1)	One week (L2)	150 °F (-1)	1
2	Yes (L1)	100 of 6000 gal (L2)	Two days (L1)	150 °F (-1)	1
3	Yes (L1)	100 of 6000 gal (L2)	One month (L3)	180 °F (1)	1
4	No (L2)	None (L1)	One month (L3)	150 °F (-1)	1
5	Yes (L1)	None (L1)	One month (L3)	150 °F (-1)	1
6	No (L2)	100 of 6000 gal (L2)	One week (L2)	150 °F (-1)	1
7	Yes (L1)	None (L1)	One week (L2)	180 °F (1)	1
8	No (L2)	100 of 6000 gal (L2)	Two days (L1)	150 °F (-1)	1
9	No (L2)	None (L1)	Two days (L1)	180 °F (1)	1
10	No (L2)	100 of 6000 gal (L2)	One month (L3)	180 °F (1)	1
11	Yes (L1)	100 of 6000 gal (L2)	One week (L2)	180 °F (1)	1
12	Yes (L1)	None (L1)	Two days (L1)	180 °F (1)	1
13	No (L2)	None (L1)	One month (L3)	180 °F (1)	2
14	Yes (L1)	100 of 6000 gal (L2)	One month (L3)	150 °F (-1)	2
15	Yes (L1)	100 of 6000 gal (L2)	One week (L2)	150 °F (-1)	2
16	No (L2)	None (L1)	One week (L2)	150 °F (-1)	2
17	No (L2)	100 of 6000 gal (L2)	One week (L2)	180 °F (1)	2
18	No (L2)	None (L1)	Two days (L1)	150 °F (-1)	2
19	No (L2)	100 of 6000 gal (L2)	Two days (L1)	180 °F (1)	2
20	No (L2)	None (L1)	One week (L2)	180 °F (1)	2
21	Yes (L1)	None (L1)	One month (L3)	180 °F (1)	2
22	No (L2)	100 of 6000 gal (L2)	One month (L3)	150 °F (-1)	2
23	Yes (L1)	None (L1)	Two days (L1)	150 °F (-1)	2
24	Yes (L1)	100 of 6000 gal (L2)	Two days (L1)	180 °F (1)	2

RESULTS

The laboratory experimental data collected for asphalt cements and emulsions under various factor-level combinations were analyzed to identify the important factors that affect or cause a statistically significant change in RTFO-DSR. Table 7 contains the row numbers corresponding to factor-level combinations in Table 5 and RTFO-DSR test results reported to the nearest 0.1 kPa. Tables 8 and 9 contain RTFO-DSR test results corresponding to the row numbers in Table 6. The difference between these tables is the type of container used for sample storage. First, unlined cans were used and induced rust formation inside the can and on top of the sample after storage at elevated temperatures. To avoid rust formation, a new set of samples was prepared in epoxy lined cans. Unlined sample results, presented in Table 8, were analyzed to examine any effect of rust development. After measuring RTFO-DSR on 10 lined emulsion samples, the equipment was recalibrated. After calibration, the whole set of lined samples was tested again. With the 10 observations obtained prior to and after the calibration procedure, a regression analysis was conducted to obtain an equation and adjust the post-calibration results to the pre-calibration ones to allow for comparison with unlined emulsion results measured before calibration. Adjusted results are presented in Table 9.

Table 7. Laboratory Experiment Results for Asphalt Cements (nearest 0.1 kPa).

Row	Percent Change Y = (Y1 - Y0)/Y0*100	
1	187.3	192.1
2	2836.1	2751.6
3	5.8	3.8
4	1.8	4.4
5	1173.1	1240.2
6	244.0	250.1
7	2516.6 2608.4	4027.9 4229.6
8	1351.3	1357.8
9	783.5	823.6
10	130.7	130.5
11	771.1	759.1
12	2029.1	2116.7
13	1294.3	1376.9
14	522.5	459.8
15	580.2	574.9
16	212.8	212.4
17	1033.8	1009.1
18	2950.5	3294.7
19	1905.6	1928.8
20	13,860.8	15,594.3
21	387.2	363.1
22	12,568.0	12,563.5
23	375.8	365.8
24	94.9	75.8

Table 8. Laboratory Experiment Results for Unlined Emulsions (nearest 0.1 kPa).

Row	Percent Change Y = (Y1 - Y0) / Y0 * 100	
1	-22.4	-15.1
2	7.1	1.6
3	-11.7	-4.1
4	-8.9	-6.9
5	-21.6	-19.7
6	-7.8	-9.5
7	-8.7	-8.7
8	7.7	7.3
9	-6.0	-0.8
10	13.0	8.6
11	31.8	19.6
12	22.1	8.4
13	8.4	14.9
14	5.4	15.7
15	1.9	6.7
16	15.1	20.4
17	25.8	23.5
18	4.4	4.5
19	2.1	8.0
20	7.4	10.7
21	3.4	6.3
22	1.0	3.9
23	13.2	11.3
24	14.2	13.0

Table 9. Laboratory Experiment Adjusted Results for Lined Emulsions (nearest 0.1 kPa).

Row	Percent Change =(Y1 - Y0)/Y0*100	
1	-21.1	-18.2
2	-10.1	-18.1
3	-27.5	-28.8
4	-0.3	1.8
5	-23.1	-20.5
6	-10.9	-12.4
7	-25.2	-25.0
8	-23.3	-16.8
9	-0.5	-0.4
10	3.8	7.9
11	-24.8	-19.7
12	-24.2	-24.0
13	1.5	1.6
14	34.6	33.5
15	18.2	17.8
16	-11.4	-9.3
17	4.9	6.3
18	4.2	0.9
19	0.5	3.8
20	-3.4	-8.1
21	41.2	49.2
22	9.3	9.2
23	32.0	26.4
24	24.4	28.2

Test results were obtained by measuring $G^*/sin\delta$ at each factor-level combination for two samples, with the exception of row 7 in Table 7 for asphalt cements for which two additional samples were taken due to excessive variability in the measured RTFO-DSR test results. Two repeated readings were taken on each sample, and the average test result over those two readings was reported as the RTFO-DSR test value (Y1) for each factor-level combination. These average test results were compared against the average (over four readings on two samples) RTFO-DSR test result (Y0) from the control samples for the corresponding Modifier and Supplier factor combination. Control samples were assumed contamination-free at the supplier site; the only contamination locations simulated in the experiment were the transport truck and the contractor tank. Note that Modifier and Supplier are the only relevant factors for control samples, thus there were only four different Y0 values. The change in RTFO-DSR was estimated based on the differences between average test results for control samples (Y0) and average test results for the stored samples obtained by simulation of storage conditions and contamination (Y1). The relative difference (Y1 - Y0)/Y0, representing the percent change (after multiplying by 100) for the stored sample compared to that of the control sample, was considered as the response variable (Y) used to simulate the change in the RTFO-DSR property between supplier and field samples.

One specific testing anomaly was encountered when measuring the results presented in Table 7. A majority of both unmodified and modified samples formed an extremely stiff crust approximately 0.12in (3mm) in thickness at the surface after storage at elevated temperatures. This effect was noted after one-week, one-month, and two-month storage times. No logical pattern was found to explain the presence of this crust in some samples and its absence in others. When the material was prepared for the RTFO procedure, homogeneity was difficult to achieve, and small stiff flakes remained in the material. Care was taken to avoid these flakes when preparing the RTFO-DSR sample for testing, but some results may not be representative due to the presence of this stiff material. A possible explanation of crust formation stems from leakage of the nitrogen seal, allowing binder oxidation that may be critical at elevated storage temperatures. Chemical aging analysis of the crust was attempted without success due to excessive material stiffness. The testing difficulty for these flaky crusted samples may be addressed by a new method for short-term aging of binders using the Stirred Air-Flow Test (SAFT). This test, developed under TxDOT project 0-1742, improves the standard RTFO method by providing more reliable results for asphalts that develop a skin at the surface and polymer-modified materials (*29*). Stiff crust formation is not considered a major issue in the supplier or contractor storage tanks, as the ratio of surface area to volume is considerably smaller when compared to the ratio of the tin used in the laboratory experiment.

Another important issue concerning RTFO-DSR results was related to a data quality review process undertaken to reduce variability that might bias conclusions and recommendations. One source of variability due to multiple technicians was identified. Technician 1 prepared approximately half of the asphalt cement samples (12 factor-level combinations), and the other half were prepared by Technician 2. Since the samples prepared by

Technician 1 showed much greater sample-to-sample variability in terms of the value measured after storage than those made by Technician 2, all of the samples done by Technician 1 were replaced with new samples prepared and retested by Technician 2.

Asphalt Cement Data Analysis

To gain a better understanding of the data, some exploratory data analyses were conducted before proceeding with Analysis of Variance (ANOVA). A box-plot of the data was first constructed. The box-plot provides information about center, spread, and symmetry/skewness of the data. The lower edge of the box corresponds to the 25th percentile (25 percent of the measurements lay below this value), and the upper edge of the box corresponds to the 75th percentile (75 percent of the measurements lay below this value). Thus the box contains 50 percent of the data. The difference between the 75th percentile and the 25th percentile (i.e., the height of the box) is often called the interquartile range and is considered a robust measure of spread since it is not affected by outliers. The centerline of the box represents the sample median (50 percent of the measurements lay below this value). The whiskers of the box extend to the largest/smallest observations that are not outliers, and the circles beyond the whiskers of the box represent outliers. Observations corresponding to rows 20 and 22 in Table 7 were identified as extreme outliers, based on the box-plot of the response variable Y (Figure 4). The identified outliers can mask the effect of important factors and/or lead to unreasonable conclusions. Therefore, the observations corresponding to rows 20 and 22 were replaced by missing values for further analysis.

Figure 5 shows that for asphalt cements, the modified binder (PG76-22) generally leads to a smaller change compared to the unmodified binder (PG64-22) after storage. The pairs of rows (3,5), (1,9), and (19,21) in Tables 5 and 7 show a higher level of contamination leads to a smaller change (with all the other factor levels fixed) for the unmodified binder (PG64-22). In other words, adding more modified binder (PG76-22) (more contamination for PG64-22) appears to improve the unmodified binder (PG64-22) in terms of susceptibility to aging.

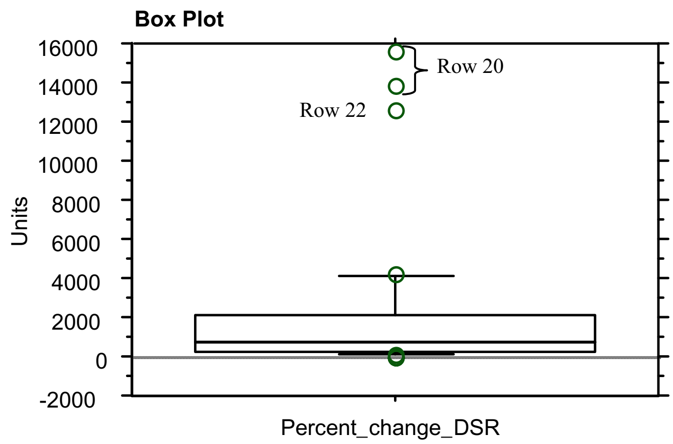


Figure 4. Box-plot for Percent Change in RTFO-DSR for Asphalt Cements.

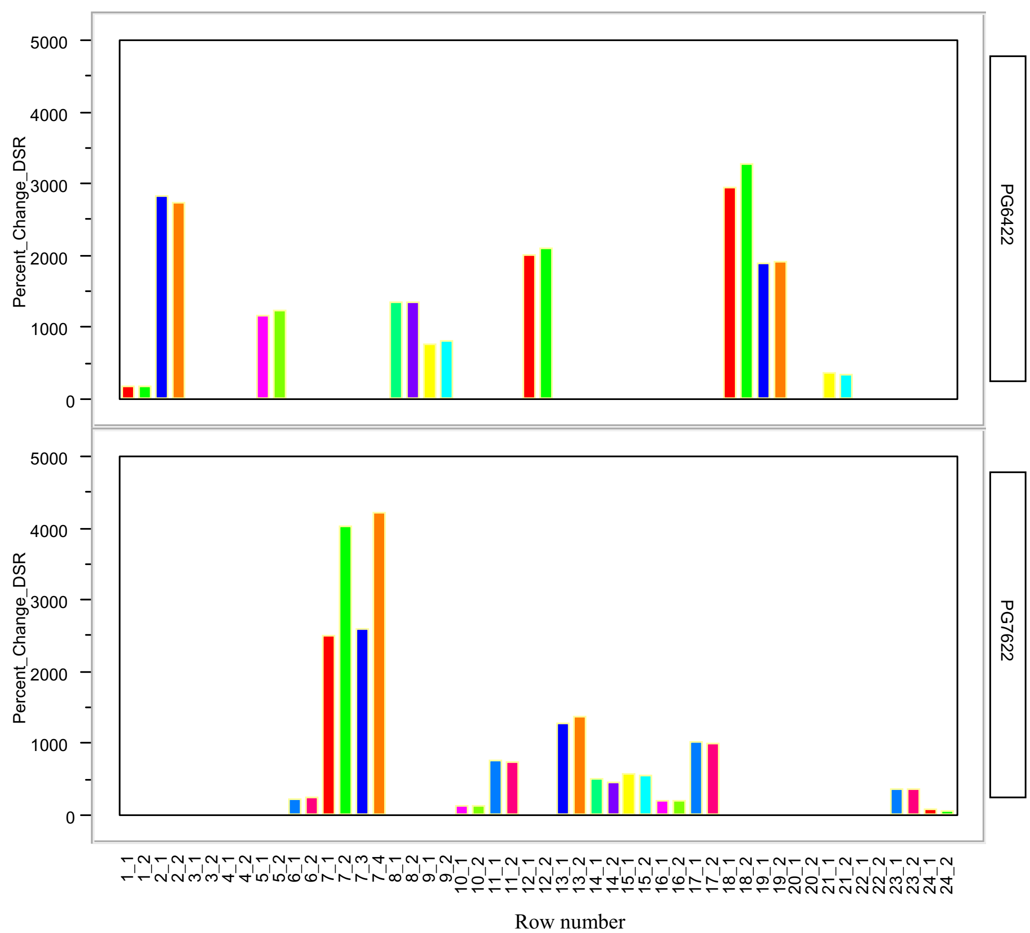


Figure 5. Percent Change in RTFO-DSR for Asphalt Cements by Modifier.

ANOVA was then employed to analyze the asphalt cement data with outliers in rows 20 and 22 removed from Tables 5 and 7. Since there is more than one observation for each factor-level combination in Table 7, it was possible to carry out a lack-of-fit *F* test. This test is used to determine whether anything left out of the model is statistically significant. A significant lack-of-fit test means that there is some significant effect (interaction) left out of the model, and therefore, it is usually desirable for the lack-of-fit test to be insignificant. A model with Modifier, Contamination, Storage Time, and Storage Temperature as main effects, all possible two-way interaction effects among them, and Supplier as a block effect resulted in an insignificant lack-of-fit test, which indicated that the model was adequate with respect to the terms included. Table 10 contains the ANOVA results for asphalt cements.

Table 10. Analysis of Variance for Percent Change in RTFO-DSR for Asphalt Cements.

Source	Degrees of Freedom	Sum of Squares	F-Value	P-Value
Modifier	1	2175621.9	21.0265	0.0001
Contamination	2	1315363.3	6.3562	0.0059
Storage Time	2	4958016.7	23.9586	<.0001
Storage Temperature	1	272275.6	2.6314	0.1173
Supplier	1	1377516.4	13.3132	0.0012
Modifier*Contamination	2	4323079.6	20.8904	<.0001
Modifier*Storage Time	2	314662.6	1.5205	0.2381
Contamination*Storage Time	4	9732153.8	23.5144	<.0001
Modifier*Storage Temperature	1	230629.4	2.2289	0.1480
Contamination*Storage Temperature	2	2120289.6	10.2459	0.0006
Storage Time*Storage Temperature	2	3646563.8	17.6213	<.0001

Based on Table 10, there are statistically significant interaction effects (joint effects of two factors) including Modifier*Contamination, Contamination*Storage Time, Contamination*Storage Temperature, and Storage Time *Storage Temperature at the level α=0.05. Due to these statistically significant interaction effects, the individual factor effects (main effects) were only assessed conditional on each level of the other factor, with the exception of the block factor Supplier. Interaction plots for the significant interaction effects are presented in Figure 6.

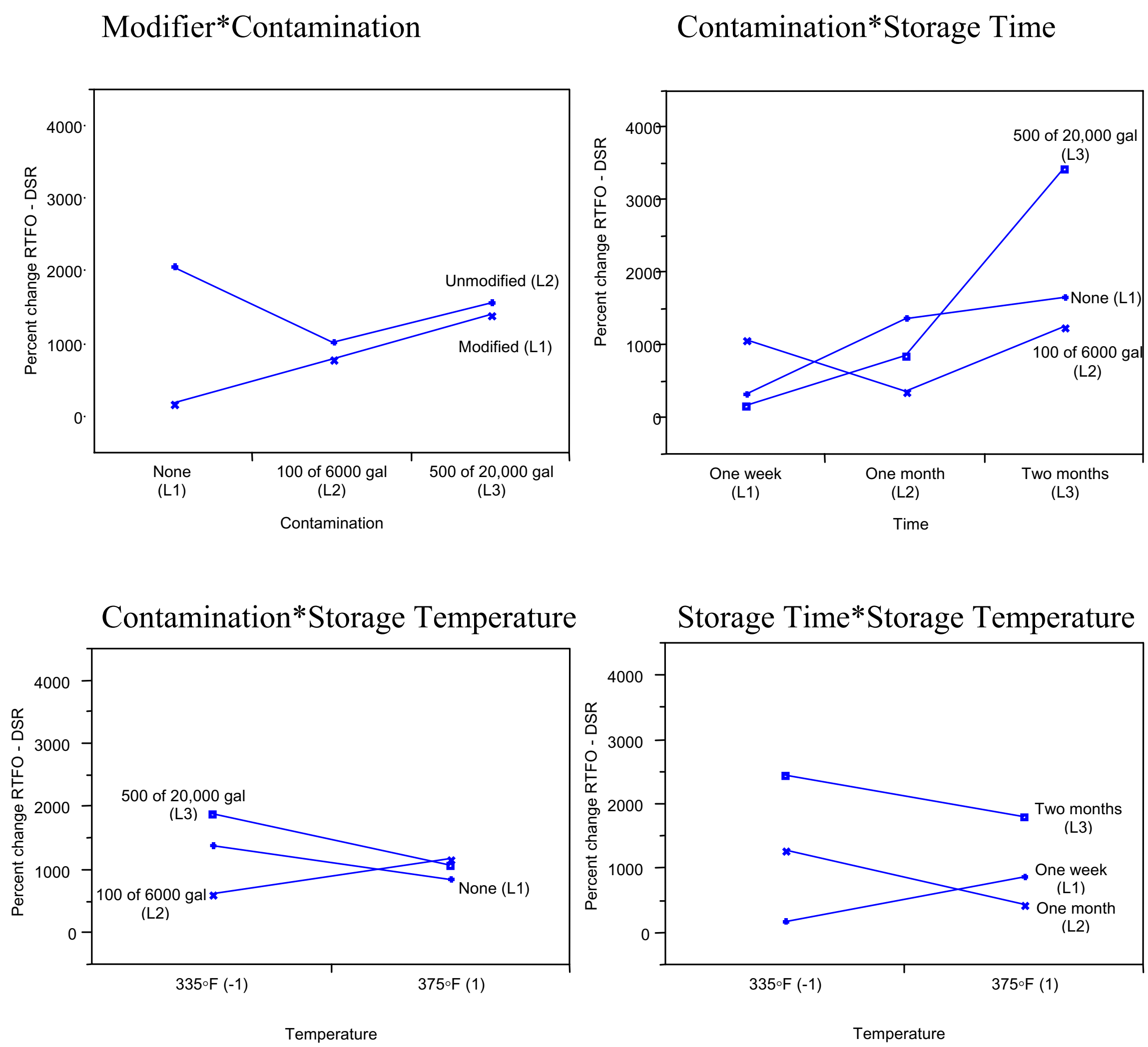


Figure 6. Statistically Significant Interaction Effects for Asphalt Cements.

The conclusions drawn based on Table 10 and the interaction plots in Figure 6 are as follows:

- The effect of Contamination was different for each level of Modifier. Contamination of the modified material (PG 76-22) led to a greater change in material response. Contamination of the unmodified material (PG 64-22) with a stiffer material that appears to be more resistant to aging led to a smaller change in material response as expected.
- The effect of Contamination was different for each level of Storage Time. The largest change occurred for a Storage Time of two months at the largest level of Contamination.

The corresponding mean response for this factor-level combination was significantly different from that for all other Storage Time and Contamination combinations.

- No Contamination effect was shown for a Storage Temperature of 375 °F (191 °C).
- The material response to Storage Temperature was statistically significant for Storage Times of one week and one month, but for the longest Storage Time of two months, there was no Temperature effect.
- One supplier exhibited a greater relative change in the response variable compared to the other supplier.

Besides statistically significant effects, a practical significant change in RTFO-DSR test results occurred when the material showed a percent change greater than 100. This percent change difference in material response was considered significant because it is equivalent to the change due to RTFO short-term aging that causes an approximate shift by one binder grade according to ASTM D3381 or Superpave specifications (*12, 15*). With the exception of rows 3, 4, and 24; all rows showed a percent change in RTFO-DSR test results of more than 100. Thus, the factor-level combinations included in the experiment are critical in terms of binder aging and performance.

In addition to RTFO-DSR, FTIR testing was also conducted on short-term aged asphalt cement samples. FTIR is an analytic technique used to identify organic compounds by measuring the absorption of various infrared light wavelengths by the irradiated sample. The wave number, shape, and intensity of the infrared absorption spectra are used to identify specific molecular compounds and structures. FTIR results were analyzed to discover peak regions of the spectra that could predict the response variable or relative change in RTFO-DSR. The FTIR test results were arranged in order of decreasing correlation with the response variable (relative change in RTFO-DSR), and then a prediction model was applied to the ranked values adding one term at a time in order until the error of prediction increased (*30*). The resulting peak region for the asphalt cement data consisted of wavelength numbers from 1695 up to 1714, which was consistent with the knowledge that the best indicator of oxidation or carbonyl area (C=O) is recorded in between wavelengths numbers of 1650 and 1820. This confirmed that the relative changes in RTFO-DSR were indeed caused by an oxidation process that formed chemical radicals within the binder structure available for reaction and combination with oxygen when the sample was exposed during the RTFO test.

Emulsion Data Analysis

Next, emulsion residue data were analyzed to determine important factors that affect RTFO-DSR values. As mentioned, emulsion residue was obtained by a stirred-can method using GPC as QC to check that the water was effectively removed from the sample. This quality check was considered important because even small amounts of water remaining in the sample can make it soft and bias the RTFO-DSR results. In the GPC procedure, polymer components are separated on the basis of their hydrodynamic size by columns filled with porous gel particles. Molecules smaller than the pore size can enter the particles and therefore have a longer retention time in the column than larger molecules that cannot enter the particles. Retention time determines the types of molecules or compounds in the sample.

The common percentage of removed water in an emulsion sample is between 25 to 30 percent. The percentage of removed water varied from 17 to 31 percent and from 12 to 38 percent for the unlined and lined emulsion data, respectively. These variations were possibly a consequence of elapsed times between the end of treatment (storage) and the dewatering process that varied between 32 and 160 days. It is possible that unsealed containers used to store the samples allowed water to evaporate during this elapsed time.

Exploratory data analyses on emulsion data revealed that changes in RTFO-DSR values of emulsion residues are considerably smaller in magnitude than those of asphalt cements. One possible explanation for the magnitude difference stems from the hypothesis that water vapor released from the emulsion surface provided a better barrier to oxidation than the nitrogen blanket utilized for asphalt cements. In addition, emulsion storage time and temperature were smaller compared to those for asphalt cements.

Several of the percent change results presented in Tables 8 and 9 are negative. For the unlined emulsions, seven factor-level combinations (rows) yielded a negative percent change, and all of them were from Supplier 1 (Figures 7 and 8). A greater number of rows showed negative results for the lined emulsion data: 13 in total for the adjusted values and 19 for the post-calibration results. Among the 13 cases, 11 were from Supplier 1, and only two were from Supplier 2. This was consistent with the unlined observations. The relatively small negative or positive percent change in RTFO-DSR (less than 20 percent in absolute value) may represent a practically insignificant change in the material in terms of performance for the conditions considered. These results suggest that storage and handling guidelines for emulsions should focus on water loss and separation of components rather than on aging.

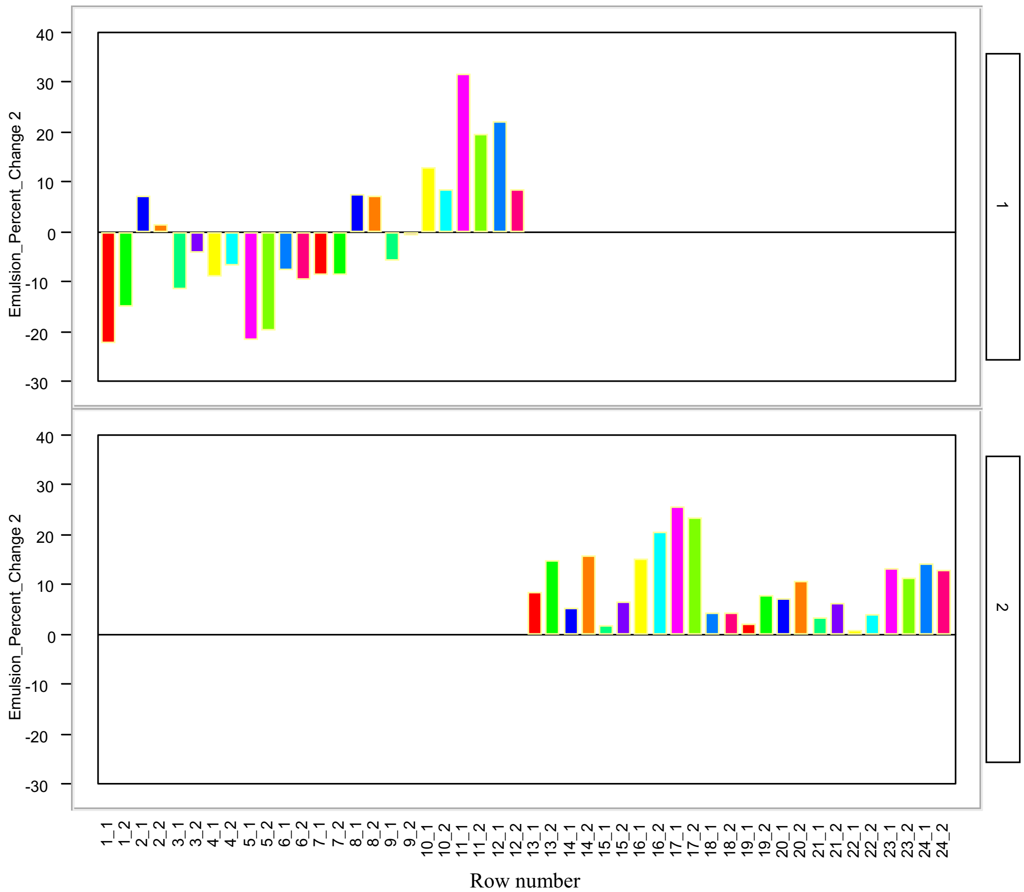


Figure 7. Percent Change in RTFO-DSR for Unlined Emulsions by Supplier.

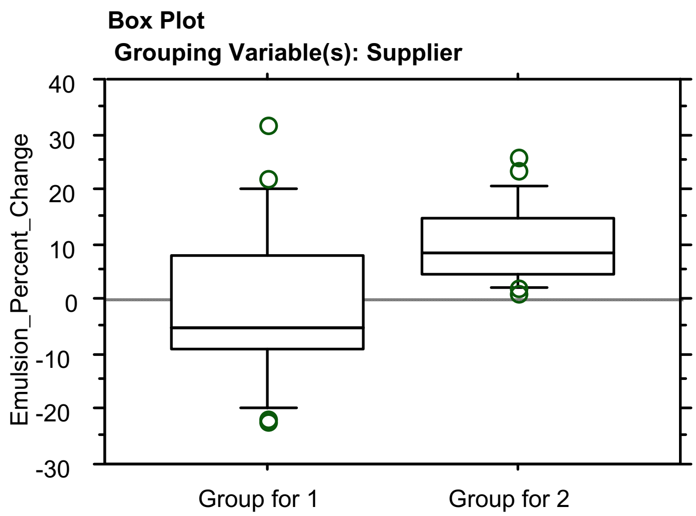


Figure 8. Box-plot for Percent Change in RTFO-DSR for Unlined Emulsions.

Unlined Emulsions Analysis of Variance

ANOVA was employed to analyze the unlined emulsion data presented in Tables 6 and 8. A model with Modifier, Contamination, Storage Time, and Storage Temperature as main effects, all possible two-way interaction effects among them, and Supplier as a block effect resulted in a significant lack-of-fit test, which indicated that the model lacked some term. Even with the addition of all possible two-way interactions between Supplier and the other factors, the lack-of-fit *F* test was still significant. Thus, data were analyzed separately for each Supplier. Several candidate models were compared in terms of the adjusted R^2 values. A model with Modifier, Contamination, Storage Time, and Storage Temperature as main effects, Modifier*Contamination, Modifier*Storage Time, Contamination*Storage Time, and Modifier*Storage Temperature as interaction effects resulted in the highest adjusted R^2 value and was selected as the final model for both Suppliers. Table 11 and Table 12 contain the ANOVA results for Suppliers 1 and 2, respectively.

Table 11. Analysis of Variance for Percent Change in RTFO-DSR for Unlined Emulsions, Supplier 1.

Source	Degrees of Freedom	Sum of Squares	F - Value	P - value
Modifier	1	210.7030	9.4550	0.0096
Contamination	1	1444.4638	64.8183	<.0001
Storage Time	2	1048.2381	23.5191	<.0001
Storage Temperature	1	29.7617	1.3355	0.2703
Modifier*Contamination	1	190.7530	8.5598	0.0127
Modifier*Storage Time	2	1107.8693	24.8571	<.0001
Contamination*Storage Time	2	813.4087	18.2503	0.0002
Modifier*Storage Temperature	1	62.9421	2.8244	0.1187

Table 12. Analysis of Variance for Percent Change in RTFO-DSR for Unlined Emulsions, Supplier 2.

Source	Degrees of Freedom	Sum of Squares	F - value	P - value
Modifier	1	218.24867	19.2035	0.0009
Contamination	1	156.02883	13.7288	0.0030
Storage Time	2	69.81407	3.0714	0.0837
Storage Temperature	1	120.24825	10.5805	0.0069
Modifier*Contamination	1	61.43955	5.4060	0.0384
Modifier*Storage Time	2	667.03484	29.3459	<.0001
Contamination*Storage Time	2	300.95400	13.2403	0.0009
Modifier*Storage Temperature	1	50.25025	4.4215	0.0573

Based on these tables, Modifier*Contamination, Modifier*Storage Time, and Contamination*Storage Time were all statistically significant at $\alpha = 0.05$, and thus the individual effects of Modifier, Contamination, or Storage Time were only assessed conditional on each level of the other factor. In addition, Storage Temperature was not statistically significant at $\alpha = 0.05$ for Supplier 1 (Table 11), but this factor was statistically significant at $\alpha=0.05$ for Supplier 2 (Table 12). Interaction plots for the significant interaction effects are presented in Figures 9 and 10.

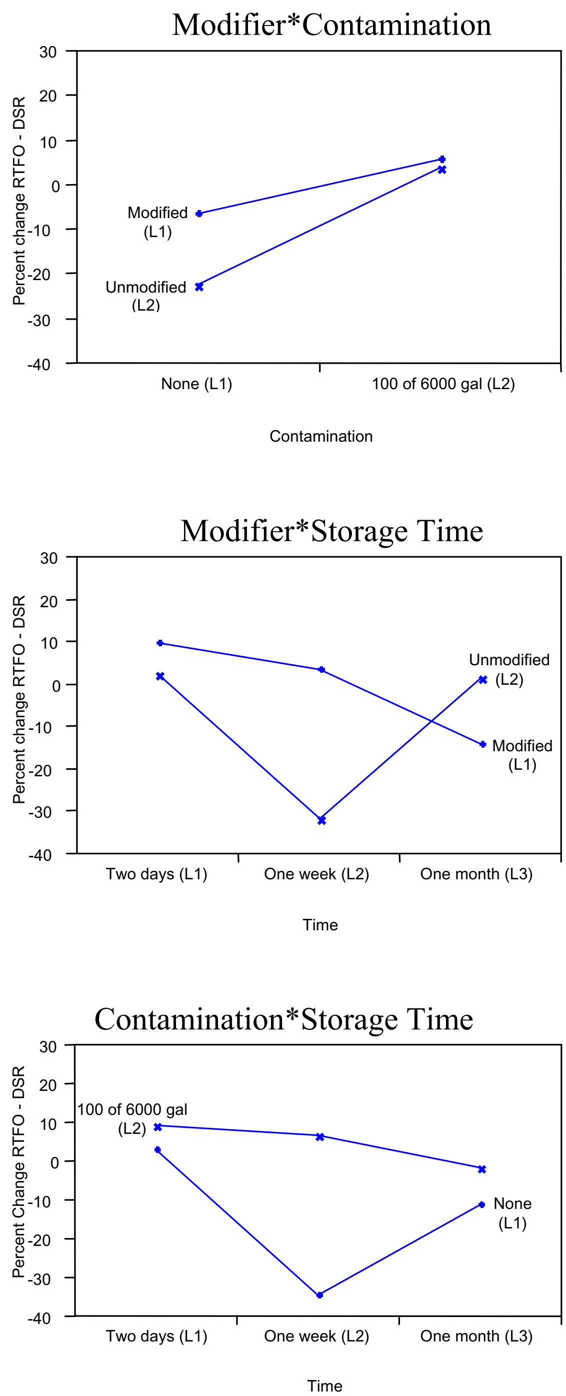


Figure 9. Statistically Significant Interaction Effects for Unlined Emulsions, Supplier 1.

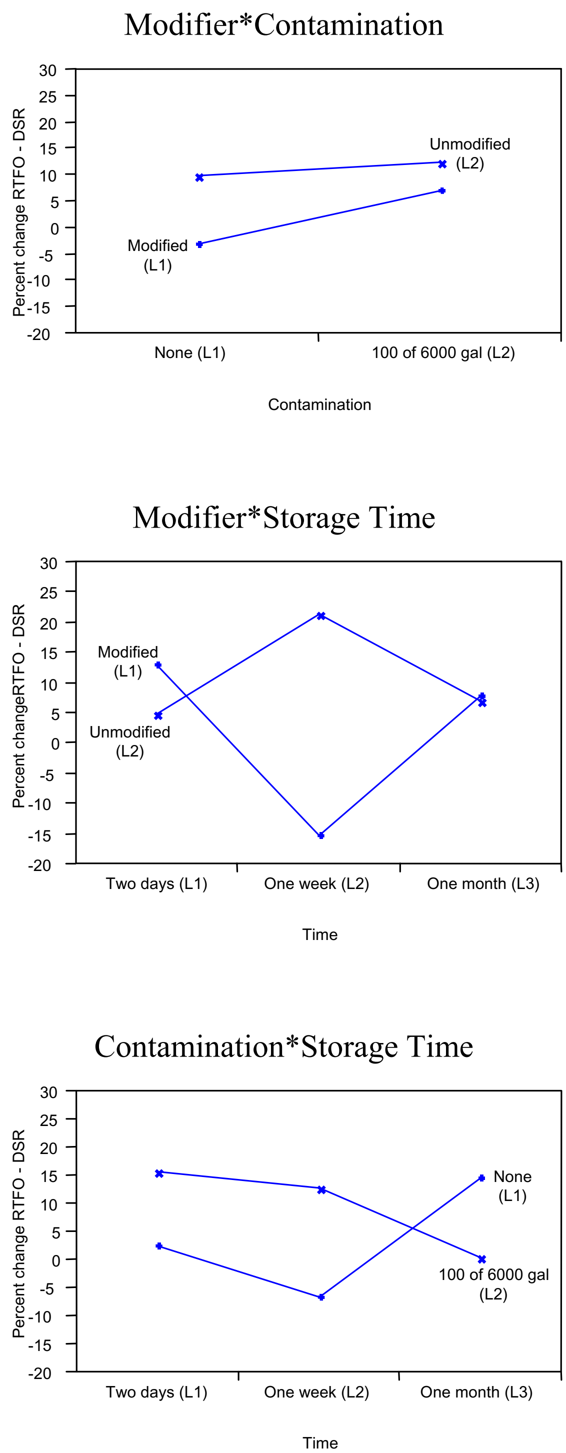


Figure 10. Statistically Significant Interaction Effects for Unlined Emulsions, Supplier 2.

The conclusions drawn based on Tables 11 and 12 and interaction plots in Figures 9 and 10 are as follows.

For Supplier 1,

- The effect of Contamination was different for each level of Modifier, with a stronger negative effect (greater change in material response) for the unmodified material CRS-2 (L2).
- The effect of Modifier was different for longer Storage Times. At Storage Time=one week (L2), the modified material CRS-2P (L1) exhibited a larger change in response compared to the unmodified material CRS-2 (L2), but the reverse was true when Storage Time=one month (L3).
- No contamination effect was shown for the shortest Storage Time of two days (L1) and the longest Storage Time of one month (L3).
- No temperature effect was shown.

For Supplier 2,

- The effect of Contamination was different for each level of Modifier. Contamination of the modified material CRS-2P (L1) led to a greater change in material response, but the unmodified material CRS-2 (L2) was resistant to contamination.
- The effect of Modifier was different for shorter Storage Times. At a Storage Time of two days (L1), the modified material CRS-2P (L1) exhibited a larger change compared to the unmodified material CRS-2 (L2). The reverse was true at a Storage Time of one week (L2).
- No contamination effect was shown for the shortest Storage Time of two days (L1) and the longest Storage Time of one month (L3).
- A larger change in material response was shown for the average storage temperature (150 °F [66 °C])

FTIR tests were also performed on unlined emulsion residues, and the results were analyzed with the same methodology used for asphalt cements. In this case, there were several distinctive peak regions. Among the identified peaks, one of them was larger in terms of wavelength numbers from 995 up to 1030. This spectral region is related to the measurement of sulfoxide (S=O), a compound often added as a cross-linking agent in emulsion products. The fact that sulfoxide compounds are not usually identified as oxidation indicators reaffirms the initial hypothesis that the relatively small negative or positive observed changes in RTFO-DSR do not represent a critical change in material properties for the conditions considered.

Lined Emulsions Analysis of Variance

Adjusted lined emulsion data presented in Table 9 were also analyzed using ANOVA. As with unlined emulsions, a model with Modifier, Contamination, Storage Time, and Storage Temperature as main effects, all possible two-way interaction among them, and Supplier as block effect resulted in a significant lack-of-fit test. Data from each Supplier were then analyzed separately. A model with Modifier, Contamination, Storage Time, and Storage Temperature as main effects and Modifier*Contamination, Modifier*Storage Time, Contamination*Storage Time, and Modifier*Storage Temperature as interaction effects was selected for both suppliers because it showed the highest R^2 value among several other models explored. The ANOVA results for each supplier, presented in Tables 13 and 14, show that Modifier*Contamination, Modifier*Storage Time, and Modifier*Storage Temperature are statistically significant for both suppliers at a level of $\alpha=0.05$. This implies that the individual effects of Modifier, Contamination, Storage Time, and Storage Temperature can only be assessed conditional on each level of the other factor. The Contamination*Storage Time interaction effect was only statistically significant for Supplier 2 but not for Supplier 1.

Based on ANOVA Tables 13 and 14 and interaction plots presented in Figures 11 and 12, the following conclusions are drawn for each Supplier:

For Supplier 1,

- The effect of Contamination was different for each level of Modifier, showing a decreasing percent change for the unmodified material CRS-2 (L2), while the modified material CRS-2P (L1) seemed resistant to Contamination.
- Storage Time affects the change in response for the unmodified material CRS-2 (L2) but has no effect for the modified material CRS-2P (L1). A stronger negative effect was observed for the unmodified material CRS-2 (L2).
- Storage Temperature affects the change for the unmodified material CRS-2 (L2) with a greater change for the higher temperature, but it has no effect for the modified material CRS-2P (L1).

For Supplier 2,

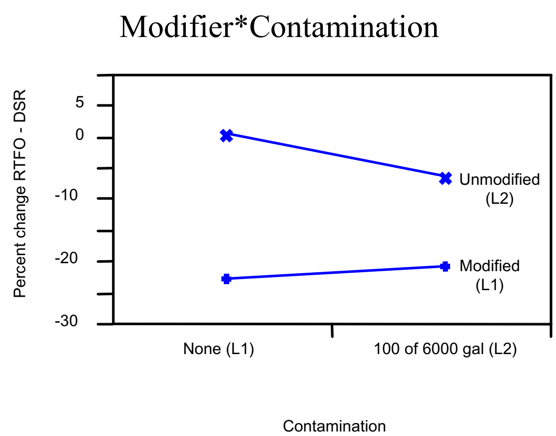
- The Contamination effect was different for each level of Modifier. It led to a slight increasing percent change in response for the unmodified material CRS-2 (L2), while the modified material CRS-2P (L1) seemed resistant to Contamination.
- The modified material CRS-2P (L1) showed a slightly larger percent change in response at the longer Storage Time of one month (L3).
- The effect of Time was different for each contamination level. For the level of contamination at the transport truck (L2) the longer storage time leads to a larger change, but this trend was not observed for the level of no contamination (L1).
- Storage Temperature had a different effect at each Modifier level. The modified material CRS-2P (L1) showed a slightly larger percent change in response when compared to the unmodified material CRS-2 (L2).

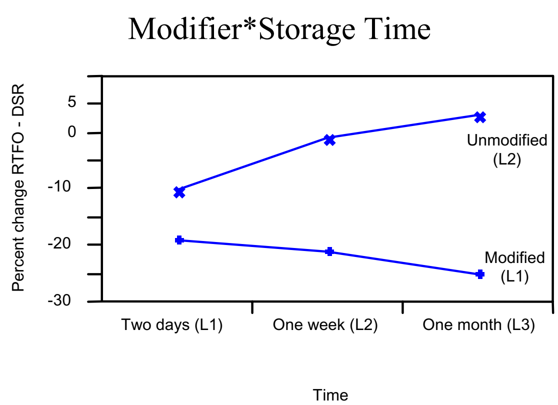
Table 13. Analysis of Variance for Percent Change in RTFO-DSR for Adjusted Lined Emulsions, Supplier 1.

Source	Degrees of Freedom	Sum of Squares	F - Value	P - value
Modifier	1	931.9796	129.2895	<.0001
Contamination	1	22.4369	3.1126	0.1031
Storage Time	2	66.5349	4.6151	0.0326
Storage Temperature	1	18.6124	2.5820	0.1341
Modifier*Contamination	1	82.9610	11.5088	0.0053
Modifier*Storage Time	2	377.4600	26.1817	<.0001
Contamination*Storage Time	2	7.72307	0.5357	0.5986
Modifier*Storage Temperature	1	421.6619	58.4952	<.0001

Table 14. Analysis of Variance for Percent Change in RTFO-DSR for Adjusted Lined Emulsions, Supplier 2.

Source	Degrees of Freedom	Sum of Squares	F - value	P - value
Modifier	1	2,111.6758	314.5147	<.0001
Contamination	1	3.2388	0.4824	0.5006
Storage Time	2	367.0837	27.3369	<.0001
Storage Temperature	1	59.3753	8.8434	0.0116
Modifier*Contamination	1	113.7419	16.9408	0.0014
Modifier*Storage Time	2	82.2527	6.1254	0.0147
Contamination*Storage Time	2	132.3190	9.8539	0.0029
Modifier*Storage Temperature	1	66.1495	9.8524	0.0086





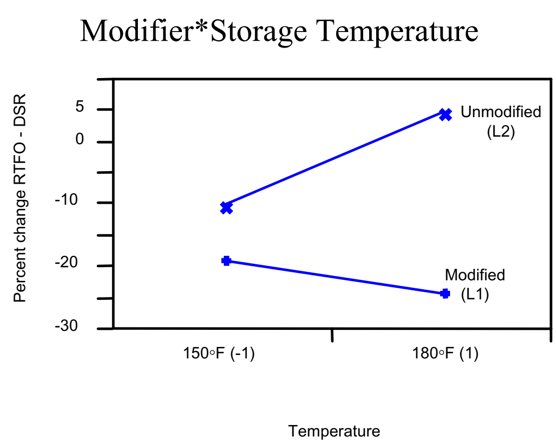


Figure 11. Statistically Significant Interaction Effects for Adjusted Lined Emulsions, Supplier 1.

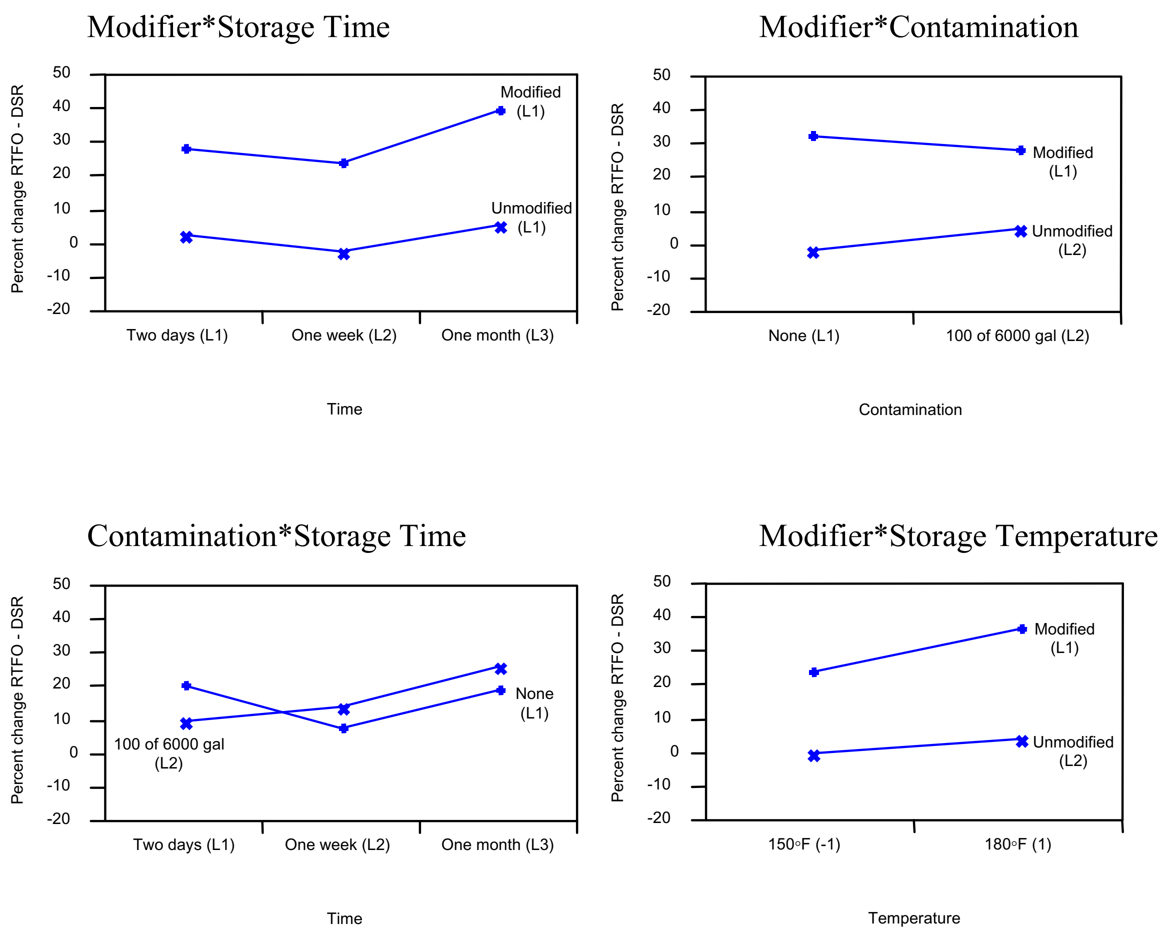


Figure 12. Statistically Significant Interaction Effects for Adjusted Lined Emulsions, Supplier 2.

The same ANOVA model proposed for adjusted lined emulsion data was used to analyze the post-calibration (unadjusted) data. Conclusions derived from the ANOVA analysis were somewhat different as compared to the adjusted data analysis. For Supplier 1, the adjusted analysis yielded Modifier*Contamination, Modifier*Storage Time, and Modifier*Storage Temperature as statistically significant interaction effects, while the unadjusted analysis showed that only the interaction effect of Modifier*Storage Temperature was statistically significant at a level of α=0.05. For Supplier 2, the adjusted analysis indicated that all interaction factors were statistically significant at a level of α=0.05, while the analysis based on unadjusted post calibration data found that Modifier* Storage Temperature was not statistically significant. The interaction plots showed approximately the same pattern for both sets of data. In addition, none of the statistically significant effects were considered practically significant because, in general, an effect is considered practically significant when the relative change in RTFO-DSR is more than 20 percent. This change corresponds conservatively to the difference in penetration required for an emulsion residue to obtain an "h" designation by ASTM D2397 specification (*15*). Thus, from an engineering viewpoint there was no difference between both results. In comparing the unlined versus the lined results in terms of practically significant effects, all of them yielded scaled estimate values of less than 20 percent, meaning that rust was not a major issue or concern in terms of aging effects.

RTFO-DSR tests are not currently performed on emulsion residues, so Penetration tests were performed to allow for comparison with any field sample results. To regulate and provide QC of the testing process, several control samples were prepared using an AC-10 binder, since the Penetration test was originally developed for unmodified binders. This control sample was measured at the beginning and end of each set of tests on a daily basis. Data collected from the control samples were found inconsistent because the results showed a decreasing tendency with time; decreasing results could be related to sample hardening. Due to the fact that a noticeable pattern was found in the control sample, the tests on the control and all the other samples were considered inconclusive. After evaluating the shortcomings of this test, a change in the testing protocol for emulsion residues was suggested, replacing this test with a more fundamental procedure, especially one that can be reliable when measuring both modified and unmodified types of material and one that can be related to performance.

CHAPTER 6. RECOMMENDED CHANGES AND RESOURCE REQUIREMENTS

TxDOT samples and approves asphalt materials at the source, and these materials are utilized in highway projects without consideration of possible changes in properties that may occur subsequent to approval. Toward the primary goal of evaluating the current TxDOT QA program for binders and recommending revisions as necessary, this report documents an assessment based on (1) an extensive information search and review that included two detailed surveys of TxDOT districts and nine other state DOTs and (2) results and analysis from a comprehensive laboratory testing program. This assessment produced a set of recommendations toward improving the TxDOT binder QA program. This chapter provides a list of these recommendations and the resources required to introduce field sampling into the current TxDOT binder QA program.

RECOMMENDED CHANGES

While it was not possible to offer recommendations for a pilot study due to the relative infancy of the effort completed to date toward establishing an improved binder QA system, the following recommended changes to the current TxDOT binder QA program are provided:

- The appointment of a binder QA program manager is recommended in addition to an education program for all employees on all aspects of the revised binder QA program to ensure maximum benefit at the least cost.
- The binder QA program established by TxDOT is recommended as only one tool in a system aimed at improving quality of the materials utilized during pavement construction and thus prolonging pavement life. Other recommended tools include required QC plans for both binder suppliers and asphalt paving contractors. Training programs for all binder technicians and personnel responsible for taking samples is also suggested. A round-robin program to establish the testing variability for selected binder QA parameters across multiple laboratories is recommended as another tool in the system.

- Data collected in the binder QA program should be stored in a user-friendly database that can be accessed by TxDOT district personnel. In addition, the number of labels for data records should be reduced to three (Pass, Fail, or For Information Only), if possible, to facilitate the production of meaningful statistical results.
- It is strongly recommended that data be organized and analyzed frequently to detect problems or show historical specification compliance for different binders and suppliers. TxDOT may use historical data to set field sampling rates by binder and supplier on an annual basis. Implementation of this recommendation will require time to educate suppliers, contractors, and TxDOT personnel.
- When field samples are taken, contractors or TxDOT personnel must label them with the corresponding acceptance laboratory number based on the supplier sample. With this information and a readily accessible database, statistical analysis can be used to gather further evidence of the potential problem of binder properties changing subsequent to acceptance.
- Based on the results from the laboratory experiment for asphalt cements, statistical analyses indicated that modifier, time, temperature, and contamination produce a significant change in the selected binder property (RTFO-DSR). With this result, the inclusion of special handling requirements in QC plans for both suppliers and contractors is recommended. Contractors may need to check for specification compliance of supplier and/or field samples and total storage time at elevated temperatures.
- Data for a particular binder shipment should include storage times and storage temperatures for both the supplier and contractor locations. This information, along with specification compliance of supplier and/or field samples, should be stored in the same database as pavement performance data throughout the life of the pavement. This may help in forensic investigations and allow future research projects to examine the effect of binder noncompliance on pavement performance. Development of these types of models is urgently needed.
- To improve the supplier/field sample identification system, the supplier bill of lading that already contains the supplier name, date and time of the shipment, number of the truck, storage temperature, project number, and the TxDOT laboratory approval number of the material should also include the storage time at the supplier site. The contractor will use that

information to label the samples pulled at the field site plus the time and temperature of storage the material was subjected to at the construction site before sending them to the TxDOT laboratory in Austin for testing. This system will provide a more reliable database, with improved data storage and recovery procedures.

RESOURCE REQUIREMENTS

The initial workplan included calculation of cost benefit ratios for the recommended changes generated in this project. These ratios could not be calculated because an assessment of benefits was not possible due to the lack of performance models that incorporate increases in rutting stiffness parameters beyond that required in the specification. The models proposed by Charmot are only applicable to binders that do not meet the rutting stiffness specification. With this limitation, resource requirements were estimated for implementing the recommended change of moving to a binder QA system that includes field sampling. This type of system is desired to ensure that the material utilized meets the required specification tied to adequate performance, and other smaller states serve as an example with successful implementation of field sampling systems.

The required resources were estimated for field sampling 100 percent of the 34,000 truckloads of performance graded (PG) binders utilized on an annual basis. Only PG binders were considered, and TxDOT personnel provided asphalt binder data from 2001 that were used for this determination. Based on discussions with TxDOT personnel, binder technicians were assumed to spend 95 percent and 65 percent of their time conducting PG tests in the district and central (Austin) laboratories, respectively. It was also assumed that to conduct a RTFO-DSR test takes only 0.33 the time of a complete set of PG tests. A complete set is currently run on supplier samples and any field samples voluntarily taken by the TxDOT districts.

In all TxDOT laboratories, it was assumed that current complete PG testing is conducted 52 weeks per year, five days per week at a rate of 20 samples per day in June, July, August, September, and October; 15 samples per day in March, April, May, and November; and 10 samples per day in January, February, and December. It was also assumed that the percentage of total HMA letting by district is equivalent to the percentage of total PG HMA letting based on concurrence of these values for four districts. A histogram of mixture testing by month was utilized along with these other data and assumptions to estimate the required number of binder technicians by month and by district.

Using these data and assumptions, resource requirements were estimated for two scenarios as a fraction of the current effort expended on the binder QA program by only the central TxDOT laboratory in Austin. This current effort, according to 2001 data, is 4175 complete PG samples per year with 4.5 binder technicians. This level of effort allows for testing of approximately 12 percent of the total number of PG binder truckloads if field sampling were implemented. If the central laboratory conducted limited testing (RTFO-DSR) for 100 percent of the PG binder truckloads without any assistance from the district laboratories, it would require an increase of funding between 1.5 and 2.5 times the current funding level. This varies by month with 11 binder technicians required at the peak of testing during the summer. At a sampling frequency of 50 percent of the PG binder truckloads, the central laboratory can provide limited binder QA testing at the current funding level, although more binder technicians would still be required for specific times of the year. If only the district laboratories conduct limited testing for 100 percent of the PG truckloads, no increase in funding is required, but some districts would have to send their field samples to other districts for testing because some districts do not have the required laboratory equipment and/or human resources. Another option would be for these districts to send their field samples to the central laboratory in Austin and/or slowly move the binder QA resources to the appropriate districts. According to the 14 responses from the TxDOT district survey, there are 22 binder technicians.

With these estimates, implementation of an exclusive field sampling program in the TxDOT binder QA system appears feasible. Specification verification for supplier samples can possibly be required of the supplier, and more responsibility for checking the supplier samples can possibly be shifted to the contractor as part of a recommended and required QC plan.

CHAPTER 7. CONCLUSIONS AND FUTURE RESEARCH

While the initial work plan could not be exactly executed due to difficulties obtaining data for corresponding field and supplier samples, the effort associated with this project produced meaningful insight into a potential problem that is seldom studied. TxDOT seeks a binder QA program to ensure materials used during construction meet specifications required to control performance and preclude premature failure that ultimately harms the public through inefficient use of funds, delay, and possibly unsafe traffic facilities. Based on an extensive information search and review and a comprehensive laboratory simulation experiment documented in this report, the following conclusions are offered:

- The current binder QA data storage system is inadequate to assess and/or address any potential problem associated with changes in binder properties from production to use during construction.
- Other states are successfully utilizing field samples in their binder QA systems. Although they are not analyzing binder QA data to its full potential, they are collecting useful data and storing it in a readily accessible format.
- The use of cluster analysis, or CART, was demonstrated as a methodology to analyze binder data to identify suppliers with a historical record of specification compliance or noncompliance by product. This type of analysis shows promise for estimating field sampling frequencies by supplier/product combination to reduce resource requirements to a reasonable level within current budget limits.
- Some mistakes were made in this initial effort to study factors that cause a change in binder properties from production to use during construction in a laboratory simulation experiment. Improvements in sealing the storage containers and the use of lined containers for emulsion samples are needed. Additional DSR tests on the original (non RTFO aged) recovered emulsion residue are suggested in order to discard any masking effects the elevated temperatures, that are not representative for emulsions, might have caused.
- Asphalt cements are substantially more susceptible to changes in properties due to storage at elevated temperatures as compared to emulsions. With this conclusion from

the laboratory simulation experiment, handling and storage guidelines are strongly recommended for asphalt cements.

This project was only an initial effort to study the factors associated with changes in binder properties prior to use. The following is suggested for future research:

- There exists a definite need to conduct a comprehensive evaluation of the current TxDOT binder QA program. This type of evaluation would require funding and time resources beyond the scope of this project. TxDOT may utilize a future research project or statistical support contract to accomplish this substantial task.
- Additional chemical engineering studies are required to better understand the reactions in asphalt cements and emulsions during prolonged storage at elevated temperatures. The mechanisms of these reactions and the resulting changes in physical properties have not been studied previously. The hypothesis of the production of numerous free radicals during storage that then readily undergoes oxidation during RTFO aging needs to be verified. A methodology for sealing asphalt cement samples to better simulate field conditions in storage tanks also needs to be developed.
- A significant national study is suggested to establish an upper limit on the RTFO-DSR binder parameter to preclude using material that has been stored for long periods of time or contaminated. This study would involve binder and mixture testing because the upper limit on the stiffness of the binder is controlled by the cracking behavior of the mixture at intermediate and low temperatures.
- Collection and analysis of historical binder data is recommended to establish field sampling frequencies by supplier/product combination. These frequencies would reduce the effort required to implement a binder QA system that includes only field sampling.

REFERENCES

(1) Epps, A., C. Spiegelman, E-S Park, E. Arambula, T. Ahmed, and T. Apanasovich, *Initial Assessment of TxDOT Binder Quality Assurance Program*, Research Report 4047-1, 2001, College Station, Texas: Texas Transportation Institute.

(2) Mitra, A., *Fundamentals of Quality Control and Improvement*, 1998, New Jersey: Prentice Hall.

(3) Summers, D. C. S., *Quality*, 1997, New Jersey: Prentice Hall.

(4) Aguayo, R., *Dr. Deming: The American Who Taught the Japanese about Quality*, 1990, New York: Simon & Schuster.

(5) Gabor, A., *The Man Who Discovered Quality: How W. Edwards Deming Brought the Quality Revolution to America: The Stories of Ford, Xerox, and GM*, 1990, New York: Times Books/Random House.

(6) Deming, W. E., *Out of the Crisis*, 1986, Cambridge, Massachusetts, Massachusetts Institute of Technology.

(7) American Association of State Highway and Transportation Officials (AASHTO), *AASHTO Provisional Standards*, April Edition, 2000, Washington, D.C.: American Association of State Highway and Transportation Officials.

(8) Northeast Center of Excellence for Pavement Technology (NECEPT), *Northeastern Superpave Center News*, Vol. 2, No. 2, Summer 2000.

(9) Anderson, D., "QA Specification for Binders: An Update," presentation at NEAU/PG Meeting, Washington, D.C., February 14, 2001, http://www.superpave.psu.edu.

(10) Northeast Center of Excellence for Pavement Technology (NECEPT), *NEAU/PG Split Sampling Program* – Version as of June 13, 2001, http://www.superpave.psu.edu.

(11) Northeast Center of Excellence for Pavement Technology (NECEPT), *PGAB Specification* – Version as of March 22, 2001, http://www.superpave.psu.edu.

(12) The Asphalt Institute, *Performance Graded Asphalt Binder Specification and Testing*, Superpave Series No. 1 (SP-1), 1994, Lexington, Kentucky: The Asphalt Institute.

(13) Bahia, H. U., D. I. Hanson, M. Zeng, H. Zhai, M. A. Khatri, and R. M. Anderson, *Characterization of Modified Asphalt Binders in Superpave Mix Design*, NCHRP Report 459, 2001, Washington, D.C.: National Cooperative Highway Research Program, Transportation Research Board, National Research Council.

(14) University of California at Berkeley, Oregon State University, Austin Research Engineers, Inc., and SWK Pavement Engineering, *Stage 1 Validation of the Relationship between Asphalt Properties and Asphalt-Aggregate Mix Performance*, SHRP-A-398, 1994, Washington, D.C.: Strategic Highway Research Program, National Research Council.

(15) American Society for Testing and Materials (ASTM), *1999 Annual Book of ASTM Standards, Vol. 04.03: Road and Paving Materials; Pavement Management Technologies,* 1999, Philadelphia: American Society for Testing and Materials.

(16) Roberts, F., P. Kandhal, E. R. Brown, D. Lee, and T. Kennedy, *Hot Mix Asphalt Materials, Mixture Design, and Construction*, 1991, Lanham, Maryland: National Asphalt Pavement Association Education Foundation.

(17) Charmot, S., *Pay Adjustment Factors for Superpave Performance Graded Asphalt Binders*, M.S. Thesis, University of Nevada, Reno, 1997.

(18) Monismith, C., J. Epps, F. Finn, R. G. Hicks, and R. Leahy, *Design, Construction, and Performance of Asphalt Paving Mixes*, Course Notes, University of California, Institute of Transportation Studies and University Extension, Richmond, California, December 9-11, 1993.

(19) Texas Department of Transportation, Construction Division, Materials Section, *Procedure for Asphalt Inspection at the Source*, October 1998.

(20) Texas Department of Transportation, Construction Division, Materials Section, *Manual of Testing Procedures*, 2001.

(21) Behrens, M., Quality Assurance Program for Asphalt Binders, Memo to All District Engineers, May 12, 1999.

(22) Texas Department of Transportation, Construction Division, Materials Section, Draft Version 1 of Special Specification for Quality Control / Quality Assurance of Asphalt Binders for Bituminous Mixtures, February 1996.

(23) Texas Department of Transportation, Construction Division, Materials Section, Draft Version 2 of Special Specification for Field Acceptance of Performance Graded Asphalt Binders for Bituminous Mixtures, October 1997.

(24) Texas Department of Transportation, Construction Division, Materials Section, Draft Version 3 of Special Specification for Field Acceptance of Performance Graded Asphalt Binders for Bituminous Mixtures, June 1998.

(25) Sharma, S., *Applied Multivariate Techniques*, 1996, New York: Wiley & Sons.

(26) Texas Department of Transportation, Construction Division, Materials and Pavement Section, Forensic Investigation of the Premature Failure of LP 375, El Paso County, El Paso District, September 2001.

(27) Epps, A. L., C. J. Glover, and R. Barcena, *A Performance-Graded Binder Specification for Surface Treatments*, Research Report 1710-1, 2001, College Station, Texas: Texas Transportation Institute.

(28) *JMP Version 5 English Documentation Set*, 2002, Kary, North Carolina: SAS Publishing.

(29) Glover, C. J., R. Davidson, N. Y. Vassiliev, *A New Method for Simulating Hot-Mix Plant Aging*, College Station, Texas: Texas Transportation Institute.

(30) Brown, P, C. Spiegelman and M. Denham, *Chemometrics and Spectral Frequency Selection*, 1991, Phil. Trans. R. Soc. Lond.