01 Mar 2023

# Integrating Quality Assurance in Balance Mix Designs for Durable Asphalt Mixtures: State-Of-The-Art Literature Review

Yizhuang David Wang

Jun Liu

Jenny Liu
*Missouri University of Science and Technology*, jennyliu@mst.edu

## Recommended Citation

Y. D. Wang et al., "Integrating Quality Assurance in Balance Mix Designs for Durable Asphalt Mixtures: State-Of-The-Art Literature Review," *Journal of Transportation Engineering Part B: Pavements*, vol. 149, no. 1, article no. 03122004, American Society of Civil Engineers (ASCE), Mar 2023.
The definitive version is available at https://doi.org/10.1061/JPEODX.PVENG-957

# Integrating Quality Assurance in Balance Mix Designs for Durable Asphalt Mixtures: State-of-the-Art Literature Review

Yizhuang David Wang, M.ASCE[1]; Jun Liu[2]; and Jenny Liu, M.ASCE[3]

**Abstract:** Delivering durable asphalt concrete within a reasonable cost is one of the great ambitions of pavement material engineers. This state-of-the-art review article documents the efforts spent in the past two decades to ensure the durability and performance of asphalt mixtures in mix design and production. A perspective with the attempt to integrate laboratory mix design, plant production quality assurance, and field place acceptance is applied in the review. The development of the performance specification and performance mix design is summarized. The paper categorizes performance specification into index-based performance specification and predictive performance specification that include performance-related specification and performance-based specification. The approaches to developing index-based performance mix design/balanced mix design and predictive performance mix design are also compared in the review. The challenges and solutions in incorporating performance tests in asphalt productions are documented and discussed. The challenges include selecting performance testing methods, determining index threshold limits, estimating and incorporating testing variability and uncertainty, determining sampling position and testing frequency, and so forth. The paper also provides suggested areas of future research and implementation activities. **DOI: 10.1061/JPEODX.PVENG-957.** *© 2022 American Society of Civil Engineers.*

**Author keywords:** State-of-the-art review; Balanced mix design (BMD); Performance specification; Asphalt mixture; Performance test; Quality assurance; Performance-based and performance-related specification.

## Introduction

Pavement engineers have been seeking methods to improve the durability of asphalt mixtures ever since asphalt pavements were introduced. In the United States, 4.1 million miles of public roads have been paved over the past few decades, and 350 million tons of asphalt mixtures are produced each year (NAPA 2019). Paving with durable materials can prevent premature pavement distresses, improve the road rideability and traffic safety, and benefit the taxpayer economically.

One of the early research efforts was the AASHO Test Road in the 1960s. The findings were incorporated in the 1972 Interim Design Guide and the 1993 AASHTO Guide for Design of Pavement Structures, which are still used in some state highway agencies (SHAs) as the primary pavement design guide. In the late 1980s, Congress approved $150 million to initiate the Strategic Highway Research Program (SHRP). Besides the asphalt binder performance grading system, another major product of the research effort was the Superpave mix design method. It originally envisioned three hierarchical levels: volumetric-based mix design for mixtures targeting low traffic volume (Level 1) and volumetric and performance mix design (Levels 2 and 3) for mixtures targeting over one million equivalent standard axial loads (ESALs) of traffic.

However, due to the limited modeling and testing techniques at that time, only the Level 1 volumetric design method was implemented for all the traffic volumes. However, the efforts to develop performance-based asphalt pavement and mixture design methods continued. In early 2000, the mechanistic-empirical pavement design guide (MEPDG) was introduced. The design guide intended to directly use the predicted pavement performance from the mechanistic-empirical (ME) models based on the project's local traffic and climate information to adjust the designed pavement structure. The MEPDG (now known as the AASHTO Pavement ME Design or the PMED) and other similar ME design methods were later incorporated into quality assurance (QA) methods and construction specifications. In recent years, as various performance testing methods have been introduced to the pavement industry, there has been a growing trend of including simple performance tests to design asphalt mixtures. Those efforts spawned the concepts of the balanced mix design (BMD).

As research studies on the aforementioned topics have been conducted, remarkable review studies have been completed and acknowledged (West et al. 2018; Yin and West 2021; McCarthy et al. 2016). In this state-of-the-art review, the authors aim to organize and present the latest information related to the great efforts in incorporating performance in mix mixture design and mixture production, and most importantly, a new perspective with the attempt to integrate laboratory mix design, plant production, and field place acceptance is applied in the review. The challenges and solutions in the implementations are also summarized in the article.

## Improving Mixture Durability under the Framework of Volumetric Design

### *Past and Current Practices in Mix Design Methods*

Pavement engineers have been trying to understand asphalt mixture since the late 19th century. The early work focused on the

---

[1] Research Associate, Dept. of Civil, Architectural and Environmental Engineering, Missouri Univ. of Science and Technology, Rolla, MO 65409. ORCID: https://orcid.org/0000-0002-5149-9898. Email: y.wang@mst.edu

[2] Research Associate, Louisiana Transportation Research Center, Louisiana State Univ., Baton Rouge, LA 70808. Email: junliu@lsu.edu

[3] James A. Heidman Professor, Dept. of Civil, Architectural and Environmental Engineering, Missouri Univ. of Science and Technology, Rolla, MO 65409 (corresponding author). ORCID: https://orcid.org/0000-0002-3840-1438. Email: jeenyliu@mst.edu

Note. This manuscript was published online on December 14, 2022. Discussion period open until May 14, 2023; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering, Part B: Pavements***, © ASCE, ISSN 2573-5438.

development and usage of the volumetric parameters as well as some pass/fail tests. Some of these concepts were inherited by today's mix design methods. In the 1890s, the first mixture design formula and construction specification were published by F. V. Greene. In 1905, the concept of voids in mineral aggregate (VMA), an important volumetric parameter that is still used in today's mix design specifications, was introduced by Clifford Richard, as well as a pat-paper test to determine the mixture gradation and binder content. Based upon Richardson's design method, in the 1920s, the Hubbard Field design method was introduced, where the binder content was adjusted based on air voids, VMA, and mix stability. The philosophy of the design method was to have sufficient asphalt binder to satisfy the binder absorption in aggregates and form a binder film with a minimum thickness. The design used a test with motor oil to estimate the binder absorption of aggregates and applied the Hveem stabilometer to test the mixture stability. Compared to other mix design approaches (for example, the Marshall design), this method tended to yield lower binder content in mixtures and fewer rutting problems. The method is still used in some western US states. The Marshall mix design method was developed in the late 1930s and early 1940s and adopted by the Corps of Engineers. The compaction method and the stability tests from the Hveem design were upgraded with automated devices. This method is currently widely used outside of the United States, and its philosophy has been inherited from the Superpave mix design method (McDaniel et al. 2011).

In 1987, to deliver more durable mixes, the pavement community initiated the SHRP. Fifty million dollars was allocated, which made SHRP the largest and most highly focused pavement research effort since the AASHO Road Test. As one major product of the SHRP program, the performance-based binder specification (based on rheology tests) was successfully implemented. However, on the mixture side, the Superpave mix design was finalized primarily based on volumetric parameters, which were originally planned mixtures with a design traffic volume of less than one million ESALs. The approaches drafted for higher traffic levels involved performance tests that could not be widely adopted due to the limitations in the modeling and testing technologies in the 1990s. The Superpave mixtures have been shown to perform better than previous mixes, and a benefit-cost study conducted in 1996–1997 quantitively showed that improved pavement performance and the increased pavement service life introduced by the new binder specification and the Superpave design method could bring a direct saving of $637 million per year or over 1.7 billion in 20 years if counting the reduced maintenance-related delays and vehicle maintenance (McDaniel et al. 2011). However, after years of practice, concerns about the Superpave volumetric design method have also been raised by pavement engineers. For example, it was reported that Superpave mixes often lack sufficient binder content for adequate durability (Maupin 2003), and the Superpave design lacks proof tests to ensure the mixture's performance (AAT and LLC 2011).

### *Improving Mix Durability by Adjusting Volumetric Parameters*

Over decades of usage, revisions to the volumetric parameters have been tried to produce more durable mixtures. VMA is one of the most important control parameters. A VMA within the proper range can provide sufficient space for effective binder and air voids as well as the potential for adequate permeability (Asphalt Institute 2015). Some states have proposed to establish the maximum VMA values of 1.5% to 2.0% above the minimum values and remove the upper limit of Voids filled with asphalt (VFA) to simplify the design procedure (Christensen and Bonaquist 2006). While attempting to adjust the VMA limit, one needs to consider the interactions among the volumetric parameters. For example, adjusting the aggregate structure to target a higher VMA may decrease the mixture compatibility, thus potentially leading to a lack of compaction or a low density in the field (Christensen and Bonaquist 2006). VFA and effective binder content ($V_{be}$) are also important factors. Some states attempted to adjust the minimum requirements for VFA to ensure sufficient binders in the mixture (AAT and LLC 2011). Alternatively, some engineers proposed using the concept of binder film thickness to determine the optimum binder content. While some researchers believed that there was no physically existing film in compacted mixtures, others thought it did exist in loose mixtures and the calculated 'apparent film thickness' had higher correlations with mixture performance than the conventional VFA. Researchers found that apparent film thickness between 7 and 9 microns could yield both suitable workability and rut resistance (AAT and LLC 2011). In the National Cooperative Highway Research Program (NCHRP) 9-25 Project, the fineness modulus (FM300) was introduced to quantitatively represent the effective surface area of aggregates. It was believed that the effective surface area can be used to calculate the film thickness and further determine the binder content (Newcomb et al. 2015).

Air voids at the design number of gyrations ($N_{des}$) or the design air voids have also been adjusted. The design air void was fixed at 4% in the Superpave mix design. The Indiana Department of Transportation (INDOT) showed that increasing the design air void to 5% can increase the average mixture density compacted in the field from 93.3% to 95.3% (Montoya et al. 2018). Broadening the design air void content from a fixed value of 4.0% to a range of 3.0% to 5.0% is another common adjustment among SHAs (Christensen and Bonaquist 2006). The associated additional compaction effort and stability are believed to be the major benefits of increasing the design air voids in most cases. However, it is important to know that, like VMA, the design air void is a function of many other factors, i.e., the gradation, binder content, and compaction energy. Some researchers pointed out that increasing the design air void might lead to an aggregate structure that is hard to compact as well as lower effective binder content, which might compromise the expected benefits (AAT and LLC 2011).

In addition to the volumetric variables in asphalt mixtures, engineers have also attempted to adjust other parameters, for example, $N_{des}$. Some studies suggested increasing the $N_{des}$ by one level to request more compaction energy in the field (Christensen and Banaquist 2006). Meanwhile, some others including the Utah DOT proposed to decrease the $N_{des}$ to increase the design binder content (AAT and LLC 2011; Tran et al. 2016). Essentially, the mechanism and the results of adjusting the $N_{des}$ are similar to changing the design of air voids. Increasing or lowering the required values may lead to the expected improved mixes, but the adjustments should be determined carefully with the consideration of the changes and interactions of other volumetric parameters.

The dust-to-binder ratio or the dust ratio is specified with a limit of 0.6–1.2 in the current Superpave design. This parameter is related to the specific aggregate surface area. A survey conducted by the New Jersey DOT showed that most of the states have adjusted the limit in their design specifications (NJDOT 2011). The NCHRP 9-25 Project mentioned that some adjustments in the dust-to-binder should be able to reduce the mixture permeability. In a study conducted for the Colorado DOT, an increase in dust-to-binder ratio was suggested to account for the 1 percent hydrated lime (Scott 2019).

### Quality Assurance under the Framework of Volumetric Design

According to a survey by McCarthy et al. (2016), QA is the most popular type of construction specification. In this system, mix design, mix acceptance, and QA usually share the same set of volumetric-based parameters and have consistent threshold limits. The pay adjustments or the incentives/disincentives are usually determined based on the percentage within limits (PWL). However, since the early 2000s, almost one decade after the Superpave was deployed, the pavement community has realized that using volumetric parameters only is not sufficient to ensure mixture durability and performance, especially when recycled materials and innovative composites are involved. One solution is to incorporate the mixture performance directly in mix design and construction specification. The efforts to develop performance mix design and performance specification are documented in the next section.

## Development of Performance Specifications

A construction specification includes information and methodologies to complete the project (i.e., project initiation, bidding, design, production and placement, and acceptance and payments) (TRB 2005, 2018). The developments and the merits of different types of pavement construction specifications, including the performance specifications, are briefly documented in this section.

### Construction Specification Types and Methods

Fig. 1 lists existing types of construction specifications. From left to right, the risk in a construction project gradually shifts from the agency to the contractor. The earliest specification was the method specification (or the recipe specification), which required the contractor to produce and place a product using the specified materials in definite proportions and specific types of equipment and methods under the direction of the agency. It could not incorporate the construction variability and did not allow any innovations from the contractors. In the 1950s, the end-result specifications were introduced along with the construction of the AASHO Road Test. It required the contractor to take the entire responsibility for producing and placing a product. The agency would either accept or reject the final product or apply a price adjustment commensurate with the degree of compliance with the specification.

Although it provided the contractor with some flexibility, the agency would take the risk of having to reject a large quantity of materials at the end of the project.

Since its introduction in the 1960s, QA specifications have been prevalent among SHAs. According to Ksaibati and Butts (2003), among the 45 SHAs that responded, 40 US state agencies had adopted QA specifications. It has inherited the merits of both method specifications and end-result specifications by requiring the contractor to conduct quality control (QC) and the agency to perform acceptance activities throughout the production and placement of a product. The final acceptance of the product is usually based on a statistical sampling of the measured quality level for key acceptable quality characteristics (AQCs). While the contractors are granted the flexibility for innovations in their products, the agencies have control of the key AQCs and the product quality on a real-time basis during production. Another type of specification is the warranty specifications, which guarantees the integrity of a product and assigns responsibilities for the repair or replacement of defects to the contractor (TRB 2018).

In the 1990s, the component of pavement performance was proposed to be added to the QA specifications, which led to the development of performance-related specifications (PRS) and performance-based specifications (PBS). PBS is defined as "a QA specification that describes the desired levels of fundamental engineering properties that are predictors of performance and appear in primary prediction relationships (i.e., models that can be used to predict stress, distress, or performance from combinations of predictors that represent traffic, environment supporting materials, and structural conditions)" (TRB 2005, 2018). The fundamental engineering properties include but are not limited to dynamic modulus, creep properties, and fatigue properties. In a PBS, the acceptance should be based on the measurement of the fundamental engineering properties of the finished product instead of the AQCs that are indirectly related to performance, and the pay adjustment should be determined based on the difference between the as-designed lifecycle cost (LCC) and the as-built or as-constructed LCC. However, no true PBS for hot mix asphalt (HMA) has been implemented because most of the fundamental engineering properties cannot be measured within the time requirements during production (TRB 2018). Given the challenges in developing PBS, PRS is an alternative solution.

PRS is defined as "a QA specification that uses quantified quality characteristics and LCC relationship that are correlated

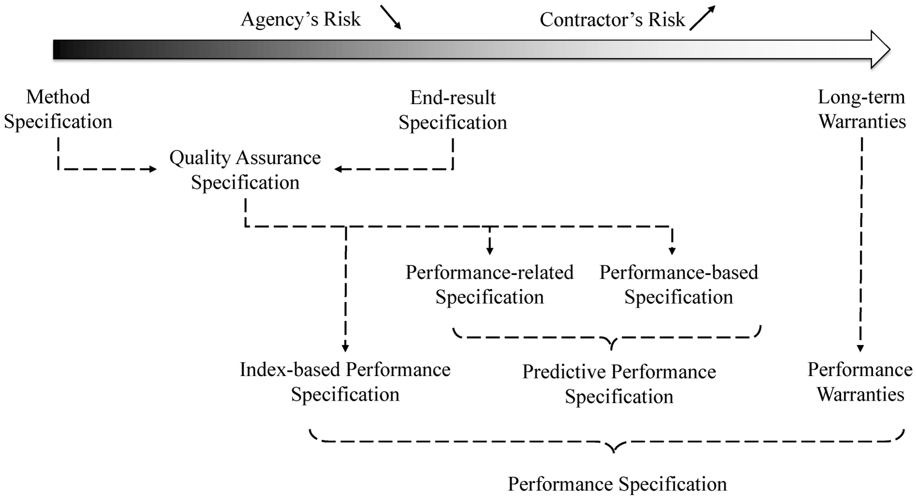

**Fig. 1.** Relationships among different types of construction specifications.

to product performance" (AASHTO 2003; TRB 2005). In other words, PRS can use AQCs to estimate or predict fundamental engineering properties. Like the PBSs, pay factors in PRSs are determined based on pavement life predicted using the estimated fundamental properties. Along with the performance warranty specifications, which are derived from the long-term warranty specifications, the PBSs and PRSs form the performance specification family.

As simple performance tests become prevalent, pavement engineers have been trying to incorporate these tests into QA specifications. Instead of predicting pavement performance using mechanistic models and ME programs, the performance tests use index parameters to evaluate mixture performance and durability. The relevant QA specifications are also written based on the performance indices. In this article, those types of QA specifications are designated as index-based performance specifications (IPS). Some researchers have named their index-based specifications as PBS; to avoid confusion, in this article, they are referred to as IPSs. The integration of mixed design and performance specification is discussed later in this article.

## Development of Performance Specifications

### Predictive Performance Specifications

Predictive performance specifications, including the PRS and PBS for HMAs, are the QA specifications that base the acceptance and the pay adjustments on the predicted pavement performance. The research PRS and PBS starting from the early 2000s intended to utilize the ME performance predictive models. A PRS framework was developed using the WesTrack test road data during the NCHRP 9-20 project. One major product was the software package, *HMA Spec*. It could generate the project-specific PRS and pay adjustments based on the differences between the as-design and the as-constructed pavement performance. The program used stiffness, permanent deformation, and fatigue cracking as the three primary variables to determine the pay factors. The performance predictive models considered factors including the construction material, environment, traffic, and roadbed soil. The cracking and rutting model was developed based on the regression analysis using the collected field data (Epps et al. 2002).

The NCHRP 09-22 project was later launched to conduct further investigation on the PRS framework (Fugro Consultants, Inc. and Arizona State University 2011). The main difference between the proposed PRS in the new project and the previous one was the predictive performance models. The MEPDG program (ARA 2004) was adopted to predict the as-design and the as-constructed pavement lives. A PRS program called the Quality Related Specification Software (QRSS) was developed to generate project-specific PRSs. The predicted pavement distresses, i.e., rutting, fatigue cracking, thermal cracking, and rideability quantified by the international roughness index (IRI) were used to adjust the payments in the specification. The AQCs that were required to estimate fundamental engineering properties included air voids, asphalt content, aggregate gradation, volumetric properties, and the binder viscosity of the AC layer. The program used the probabilistic method considering the variabilities in the construction and laboratory measurements to the pay factors. The method applied the Monte-Carlo simulations for rutting and fatigue cracking, while the Rosenblueth probabilistic point estimate method was used for thermal cracking analysis. The pay factors were determined based on the PF-PLD relationship (where PLD stands for predicted life differences). The final incentive and disincentive were calculated in dollars by considering all individual pay factors.

Due to the limitations in the prediction accuracy of the MEPDG models and the efforts required to calibrate the models, from 2008 to 2021, the Federal Highway Administration (FHWA) funded the Applied Research Associates, Inc. (ARA) and North Carolina State University (NCSU) to develop an improved PRS framework (Kim et al. 2022a, b). The outcomes of the research included the development of three fundamental material models and the corresponding material testing protocols, a structural performance simulation program, a performance mix design framework, and a QA strategy. Among the three material models, both the fatigue cracking model, i.e., the simplified viscoelastic continuum damage (S-VECD) fatigue model, and the low-temperature cracking model stemmed from the VECD theory, and one cyclic fatigue test was able to calibrate the model coefficients for both models. The models used fundamental material properties, such as the describe the relationship between the reduction of material pseudo stiffness ($C$) and the growth of damage ($S$) (Underwood et al. 2012; Wang and Kim 2017; Ashouri et al. 2021), which were independent of loading conditions (i.e., modes of loading, loading amplitudes, frequency, and loading temperatures) and could be used to predict material behaviors under different circumstances. The rutting model was able to the permanent strain formed under different loading amplitudes with various resting periods at different temperatures. The materials models were implemented in the structural performance simulation program, FlexPAVE, that conducted a three-dimensional viscoelastic analysis with the consideration of the in-situ pavement structure, traffic, and climate data. Good agreements have been found between the FlexPAVE predictions and field measurements (Wang et al. 2016, 2018, 2021b). To determine incentive/disincentive based on predicted performance and save the material testing time during QA, a bridge connecting the routine AQC and the predicted performance, i.e., the performance-volumetric relationship (PVR), was developed (Wang et al. 2019). During production, the contractor and the agency only needed to measure the volumetric properties, the same as in the existing QA procedures, and the PLD for each lot could be predicted using PVR. Shadow projects using the PRS framework have been deployed in some US states (Jeong et al. 2020; Kim et al. 2022a, b); however, extensive efforts for training and demonstration would still be needed for future implementation.

In addition to the research efforts supported by the NCHRP and FHWA, other researchers have also developed performance specifications that were suitable for local applications. Since the late 1990s, California has been spending efforts to develop PBS based on the ME methods. The original performance mix design and construction specification incorporated performance tests, i.e., the flexural beam test and the repeated simple shear test. In 2000, the CalME flexible pavement design software was first introduced. It was developed based on incremental-recursive damage models regarding the fatigue and rutting performance (Harvey et al. 2014). The CalME program calculated the pavement fatigue life with the given material and structural conditions. The reliability of the predictions was evaluated using the Monte-Carlo analysis. The CalME calculated the mechanical responses using the linear layered-elastic-based ELSYM5 program. The obtained largest maximum principal tensile strain was then used to evaluate the temperature equivalency factor (TEF) and temperature conversion factor (TCF). The pavement fatigue life was predicted using the three-stage Weibull equation (Tsai et al. 2012) considering the reliability and the effects of pavement structure and climate.

In 2004, Williams et al. (2004) developed a PRS for the Michigan DOT. The specification used predictive equations and relationships between field performance (i.e., rutting and cracking) and mix properties (i.e., air voids and binder content) to determine the pay factors. In the proposed PRS framework, the testing methods included the four-point beam fatigue test and the asphalt pavement analyzer test, and the obtained performance indices were correlated with field performance based on past testing experience. The sampling methods were thoroughly discussed in the study, and the existing mix designs were verified.

In summary, many researchers and organizations have participated to establish the PRS framework in the past two decades. The proposed PRSs utilize either mechanistic or empirical material and structural models to predict pavement lives. These approaches have the merits of both PBSs and conventional QA specifications. On the other hand, the challenges to deploying the PRSs include effectively quantifying and encompassing the risk and reliability in the specifications (Hughes 2005; Hughes et al. 2012), providing training and instructions for SHAs and contractors, developing local material database and local model calibration coefficients, and simplifying the performance tests. Some of these problems have been mitigated in the IPS.

#### Index-Based Performance Specification

Since the late 2000s, various simple mixture performance tests have been introduced to the asphalt community. The tests usually define indices to evaluate the mixture's durability or its resistance to a certain type of pavement distress, and most of the tests can be easily and quickly performed compared to mechanistic-based material characterization tests. The index-based performance specifications or the IPSs have been introduced with the integration of simple performance tests and QA specifications.

In 2016, Mohammad et al. (2016b) developed a standard PBS method and a simplified PBS approach. The standard PBS was based on the AASHTO Pavement ME Design Software and considered the pavement structure, climate, traffic, and material properties (i.e., dynamic modulus). The simplified PBS was an IPS that required the rutting and cracking test results. The comparison between the standard PBS approach and the simplified PBS approach suggested the simplified PBS was recommended because the pavement ME program had not been locally calibrated.

Given the limitations in the volumetric-based mix design and QA specifications and to accommodate the usage of reclaimed asphalt pavement (RAP), the New Jersey DOT (NJDOT) started to develop performance specifications with performance-based acceptance procedures (Bennert et al. 2014). In the proposed IPS, the contractor should design mixtures passing the volumetric criteria, and the agency conducts performance tests with the submitted component materials. Once the mixture is approved, the contractor should construct a test strip where plant-produced mixtures are sampled and tested. If the mixture meets the performance criteria, the production can be continued, and the mixtures should be sampled and tested at a certain frequency during production. The asphalt pavement analyzer (APA) should be used to assess the mixture rutting resistance, and either the flexural beam fatigue test or the overlay test should be adopted for the cracking resistance evaluation depending on the mixture type. The NJDOT has successfully developed five performance-based asphalt mixtures, namely, high-performance thin overlay (HPTO), binder-rich intermediate course (BRIC), bridge deck waterproofing surface course (BDWSC), bottom-rich base course (BRBC), and high RAP (HRAP) (Bennert et al. 2014).

The Northeast Pavement Preservation Partnership (NEPPP) developed a pilot specification for HPTO mixtures. The specification encompassed surface preparation, material properties, mixture design requirements, RAP testing requirements, and mixture performance criteria. Performance tests, i.e., the overlay test, the thermal stress restrained specimen test (TSRST), the four-point flexural beam fatigue test, and the APA test, were required to be conducted on plant-mixed mixture during production and placement. The mixtures must meet both the performance index limits and the Superpave volumetric requirements. After two years of field monitoring, the HPTO mixture from New Hampshire performed well with minimal cracking observed; in contrast, the conventional mixture had 25% cracking returned. No apparent distress was found in the Vermont test sections for both HPTO mixtures with and without RAP (Mogawer et al. 2012).

In summary, the IPSs are QA specifications with simple performance tests incorporated. Compared to the predictive performance specifications, the IPSs are more intuitive and can adopt performance tests that can be completed in a timely manner. However, there are still challenges in implementing the IPSs, including the relatively high cost and uncertainties, lack of communication between the SHA and the contractors, and the delay in production for conducting performance tests (McCarthy et al. 2016).

## Development of Performance Mix Design Methods

### *Performance Mix Design Approaches*

In the 2000s, several years after the Superpave mix design was widely implemented in the United States, multiple simple performance tests were developed and incorporated into mix design methods, including the prevalent BMD. Like the performance specifications, in this article, the performance mix design methods are introduced as index-based performance mix design (IPMD) and predictive performance design. The existing BMD methods are IPMDs as the incorporated performance testing methods are index or tolerance tests.

#### Index-Based Performance Mix Design/Balanced Mix Design

The concept of BMD was first proposed in 2007, as a balanced binder content was expected to be identified for given mix components through index-based performance tests so that the designed mix was neither too lean to form pavement fatigue cracking or too wet to yield deep rut depth (Zhou et al. 2007). In September 2015, the FHWA Expert Task Group (ETD) on Mixtures and Construction founded a BMD Task Force. The BMD Task Force defined BMD as "asphalt mix design using performance tests on appropriately conditioned specimens that address multiple modes of distress taking into consideration mix aging, traffic climate and location within the pavement structure" (West et al. 2018). Three pathways were originally developed by the task force, and later, they were expanded to four approaches when the corresponding provisional AASHTO standards PP105-20 and MP 46-20 were submitted, as presented in Fig. 2 (Yin and West 2021). The details and the highlights of each approach are demonstrated in Table 1. Under the BMD framework, several implementation plans have been developed among SHAs (Paye 2014; Cross and Li 2019; ALDOT 2020; Bennert 2020; Coleri et al. 2020).

Like the IPSs, the advantages of IPMD or BMD include the relatively short turnaround time of the simple performance tests and the intuitive design philosophy. Using both the volumetric and performance criteria provides the pavement engineers confidence in the mixture quality. However, the index threshold limits can only be determined based on empirical relationships between field performance and performance test results. Besides, the indices cannot take the project-specific information (i.e., structural, environmental, and traffic conditions) into account, and neither can the formation of pavement distress as a function of service time be predicted using the index-based approaches, as envisioned in the original SHRP project.

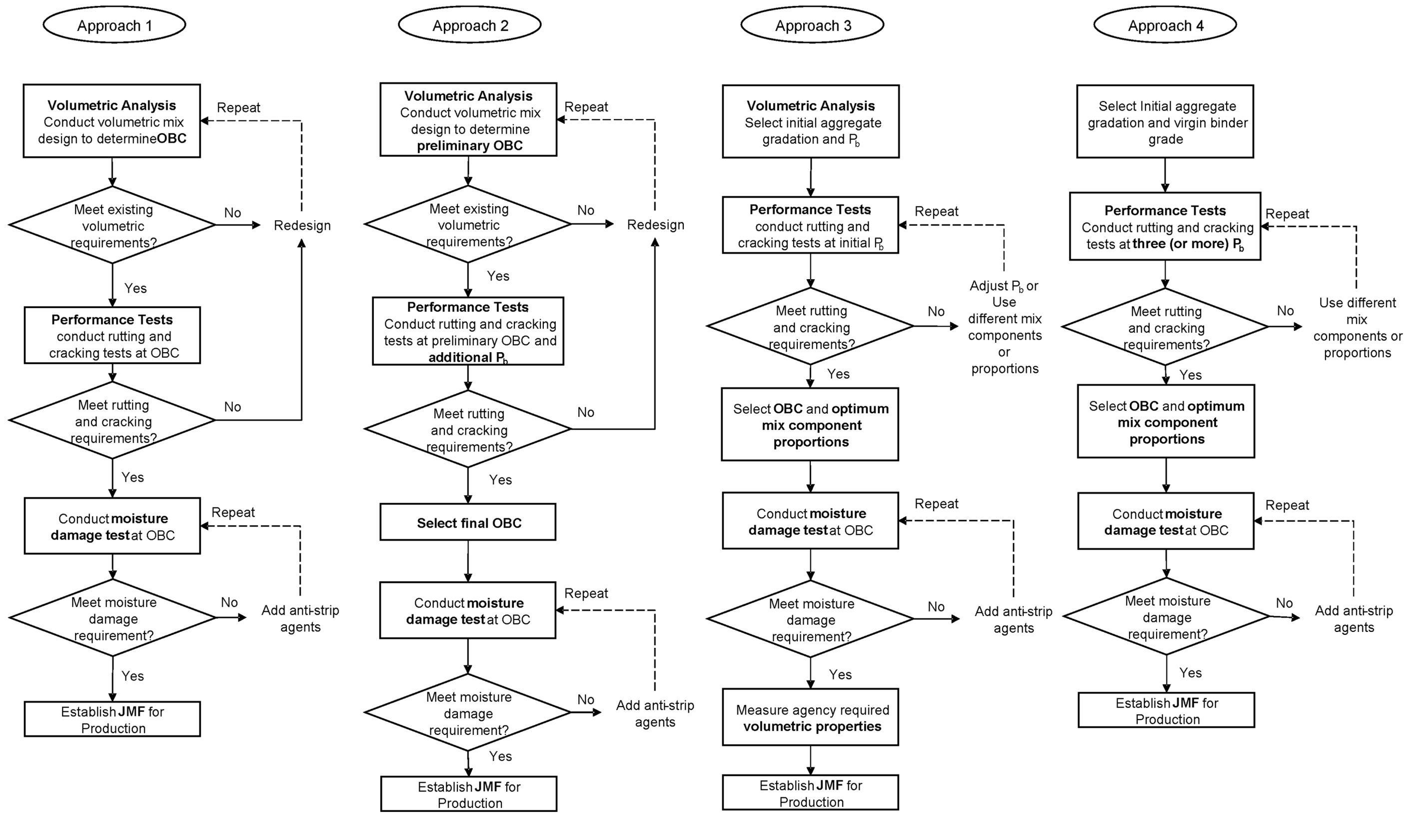

**Fig. 2.** Flowcharts demonstrating BMD approaches based on Yin and West (2021).

**Table 1.** Highlights of different BMD approaches

| Approach | Description | Highlights |
|---|---|---|
| BMDApproach 1 | Volumetric design with performance verification approach | • This approach makes sure that all the mix designs are products of volumetric design methods with performance requirements satisfied.<br>• This approach applies additional constraints for performance requirements to the original volumetric designs. This combination provides the engineers the most confidence but the least design flexibility for the contractors. |
| BMDApproach 2 | Volumetric design with performance optimization approach | • This approach is an expanded version of Approach 1, and it was not included in the original three approaches proposed by the former FHWA BMD Task Force.<br>• This approach allows a potential offset in optimum binder content determined based on the performance test results from volumetric optimum binder content while the mixture gradation and other mix components will remain the same as designed by the volumetric-based method.<br>• When this approach is adopted, the binder contents for performance testing will usually be preliminary OBC − 0.5%, preliminary OBC, preliminary OBC + 0.5%, and preliminary OBC + 1.0%. |
| BMDApproach 3 | Performance-modified volumetric design approach | • This approach and Approach 1 both start with a volumetric design.<br>• Unlike Approach 1, this approach allows adjustments for both the gradation and binder content based on the performance test results. The final combination of the gradation and binder content is not directly obtained from volumetric design, and only some volumetric criteria are required to be met. |
| BMDApproach 4 | Performance design approach | • This approach is similar to the third approach proposed by the former FHWA BMD Task Force, but more details and instructions were provided than the descriptions when it was first introduced by the task force.<br>• This approach is a combination of Approach 2 and Approach 3, and it may not necessarily start with a volumetric design.<br>• After the initial selection of aggregate gradation, recycled asphalt materials, content, and virgin binder grade is determined, the binder contents for performance testing will usually be initial binder content −0.5%, initial binder content, initial binder content +0.5%, and initial binder content +1.0%. |

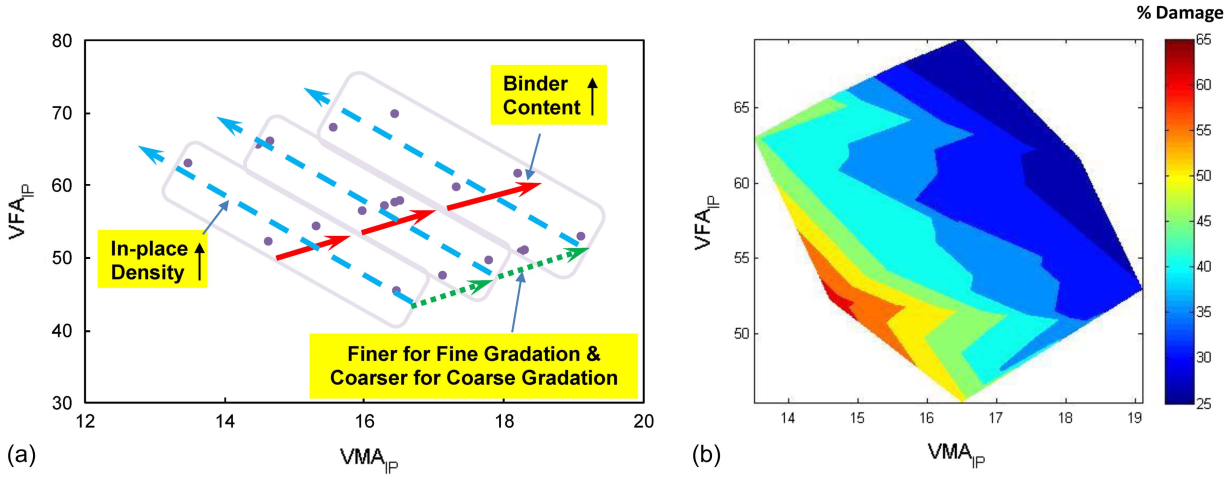

**Fig. 3.** (a) Volumetric space formed by $VMA_{IP}$ and $VFA_{IP}$ and effects of mix design parameters on $VMA_{IP}$ and $VFA_{IP}$; and (b) % damage contour in a pavement structure in the volumetric space. [Y. D. Wang, A. Ghanbari, B. S. Underwood, and Y. R. Kim, “Development of a Performance-Volumetric Relationship for Asphalt Mixtures.” *Transportation Research Record* 2673 (6): pp. 416–430, © 2019 by SAGE, reprinted by permission of SAGE Publications, Ltd.]

### Predictive Performance Mix Design

Predictive mix design utilizes the mixture/pavement performance predicted from the mechanistic models to determine the optimum mixture design. One predictive design approach has been proposed by NCSU. The method is also known as the performance-engineered mix design (PEMD) (Wang 2019; FHWA 2019; Kim et al. 2022a, b; Wang et al. 2021a). One feature of the method is that instead of using the trial-and-error approach (creating one trial design and using volumetric or performance criteria to determine pass or fail), the PEMD identifies the performance-optimum design directly from the infinite numbers of combinations of the given material components. The identification of the optimum design can be achieved by using the PVR. PVR characterizes a given mix design with two variables, the in-place VMA ($VMA_{IP}$) and in-place VFA ($VFA_{IP}$), and forms a two-dimensional volumetric space. Each point in the space corresponds to one combination of gradation, binder content, and compaction level, as presented in Fig. 3(a). Previous research (Wang et al. 2019) found that when the climate, traffic, and structural conditions are known, the predicted pavement performance (% fatigue damage and permanent deformation) is a bilinear function of $VMA_{IP}$ and $VFA_{IP}$. The PVR function, therefore, provides a spectrum of performance as a function of the change of gradation, binder content, and compaction level, as presented in Fig. 3(b). The performance-optimum design can be acquired by combining the predicted pavement life determined by fatigue life and rutting failure (Wang 2019; Wang et al. 2021a). It will be the SHA’s discretion to require all or some of the volumetric limits to be met, and the moisture susceptibility of the design mixture can be tested afterward. The design procedure requires four sets of performance tests to calibrate the PVR function coefficients. The performance tests used in PEMD are the cyclic fatigue test and the SSR test. If the design candidate fails the volumetric or moisture tests, the designer can select another combination from the volumetric spectrum without conducting additional performance tests. Therefore, unlike the unknown number of iterations that the IPMD methods may require, with the fixed number of performance tests and testing time, the design timeline can be planned by the SHAs and contractors. Twelve days are expected to complete the design including the specimen preparation and testing time (Wang et al. 2021a).

## *Comparison of Performance Mix Design Approaches*

Table 2 summarized several important criteria to evaluate the performance design approach. Among the four BMD approaches, the involvement of the performance increases from Approach 1 to Approach 4 as the volumetric restrictions for binder contents and gradations are gradually released. As a result, Approach 1 provides the agency highest confidence and lowest risks when BMD is first implemented. However, Approach 1 meanwhile grants the contractor the least design flexibility. As for the predicative performance mix design methods, the performance involvement and design flexibility would be the highest among all the methods; however, the agency may have the least confidence since the design will primarily be based on the predicted performance instead of the conventional volumetric parameters.

In terms of the design effort needed in each approach, the predictive performance mix design method may consume the longest time because of the complexity of the required performance tests. However, the BMD approaches may take longer time if multiple iterations are needed for redesign or adjustments; in contrast, the number of performance tests is fixed for the predictive performance design method. As for the development of performance specifications, in each approach, the same performance requirements can be consistently applied in mix design and in QA as long as the specifications are well engineered.

## *Integrating Performance Mix Design with QA*

As both mix design and QA are important steps on the asphalt production chain, it is necessary to develop holistic methodologies to accommodate performance mix design and QA methods. Consistent evaluation strategies and criteria would be preferred for the system. As BMD methods are implemented, great efforts have been made by researchers and SHAs to form such coherent systems. For example, Zhou et al. (2020, 2021) proposed a framework. It started with a mix design method where the volumetric criteria are used to obtain the initial candidate and performance tests are conducted at multiple binder content levels to determine the final binder content. The agency was required to conduct performance tests on the submitted mix design for verification and acceptance. During production, the plant-mixed mixtures would be sampled for QA testing.

**Table 2.** Factors to consider for comparing performance design approaches

| Evaluation criteria | Description |
|---|---|
| Performance involvement | The extent of performance involved in the design procedure. If the performance tests are only used for pass/fail decisions after volumetric design, the approach will be evaluated with low-performance involvement. High involvements are granted to the methods where gradation, binder content, and other design parameters are determined based on the mix performance. |
| Confidence and risk | The confidence level that SHA has when they first switch from volumetric mix design to this approach. For example, if volumetric limits are not required in a new mix design, the SHA would have less confidence and take a higher risk on the mix than on the mixtures that meet the volumetric requirements. |
| Design flexibility | The flexibility that the contractor has while conducting the mix design. For example, if the volumetric requirements are inherited, the additional constraints added for performance will further limit the design flexibilities for contractors. Higher design flexibility can provide contractors incentives to apply innovative materials and technologies. |
| Design effort | The design effort indicates the time and resources that the performance design approach costs. This criterion should be evaluated based on the number of gyratory specimens required to be fabricated, the number of performance tests, the testing turnaround time, and the estimated design time in days. |
| Performance prediction capacity | The performance prediction capacity indicates how well the mixture performance is incorporated in the mix design approach. Does the mix design rigorously consider the target pavement structure, climate, traffic volume, and other factors? Can the mix design approach predict the pavement distress deterioration with time as initially expected in the Superpave Level III design? Are the performance threshold limits determined by empirical or mechanistic methods? |
| Compatibility with QA | The compatibility with the QA index indicates the difficulty level to develop a Quality Assurance method and a performance construction specification with pay adjustments using the same testing methods and/or performance threshold limits. |

The same performance tests and performance index limits as in the mix design were suggested for QA. To meet the time requirement, simple tests, such as the IDEAL-CT, IDEAL-RT, and HWTT, were proposed to assess the mixture cracking resistance, rutting resistance, and moisture susceptibility, respectively. Meanwhile, several states have specified performance requirements during production, as presented in Table 3.

As for the predictive performance designs and specifications that rely on performance testing results to predict mix performance and determine pay factors, the testing turnaround time during production would be the main challenge to developing the coherent framework. To integrate the PRS and PEMD, one strategy that was applied was to use the PVR (Jeong et al. 2020). While the volumetric parameters were measured as AQCs during production, with PVR, the performance of the mixes in each lot could also be estimated. To accommodate the variabilities in materials, the performance tests were suggested to be conducted once every several thousands of tons to calibrate the PVR coefficients. The challenges for implementing and integrating the performance mix design and performance construction specifications are discussed in the following section.

## Challenges and Solutions in Development of Holistic Methods

Introducing the performance specifications/mix design methods will make changes to the process of bidding, mix design, acceptance, production, and payment/reward and penalty. Though some simple performance tests have been developed, one major challenge is still their turnaround time compared to the conventional volumetric tests. Also, depending on the type of performance specification/mix design method that the SHA uses, some extent of communication and training is also necessary at the initial stage of the deployment. Selecting the type of performance specification and mix design method that is suitable for the local area is usually the SHA's responsibility. If an index-based mix design method will be applied, the SHA also needs to determine the design approach

**Table 3.** Summary of state-of-the-practice on BMD implementation

| BMD approach | State | Rutting test | Cracking test | Performance testing for production acceptance? |
|---|---|---|---|---|
| Approach 1 | Illinois | HWTT | I-FIT | Yes, HWTT for "Pass/Fail" |
| | Louisiana | HWTT | SCB-Jc | Yes, "Pass/Fail" |
| | New Jersey | APA | OT, BBF | Yes, "Pass/Fail" or Pay Adjustment |
| | Texas | HWTT | OT,IDEAL-CT | Yes, "Pass/Fail" |
| | Vermont | HWTT | I-FIT | Yes, PWL |
| Approach 1 and 4 | Virginia | APA | Cantabro, IDEAL-CT | Yes, "Pass/Fail" |
| Approach 3 | California | FN, HWTT | BBF, I-FIT | Yes, HWTT for "Pass/Fail" |
| | Missouri | HWTT | I-FIT, IDEAL-CT | Yes, HWTT for "Pass/Fail", I-FIT, and IDEAL-CT for Pay Adjustment |
| | Oklahoma | HWTT | IDEAL-CT | No |
| Approach 4 | Alabama | HT-IDT | AL-CT | Yes, "Pass/Fail" |
| | Tennessee | HWTT | IDEAL-CT | To be determined |

Source: Data from Yin and West (2021).
Note: OT = overlay test.

**Table 4.** List of performance test methods

| Distress | Test method | Testing standard | Type of test result |
|---|---|---|---|
| Intermediate-temperature cracking | Direct Tension Cyclic Fatigue Test | AASHTO TP 107-18, AASHTO TP 133-21 | Performance index; predicted performance |
| | Flexural Bending Beam Fatigue Test | AASHTO T 321-17, ASTM D8237-21 | Performance index; predicted performance from empirical relationships |
| | Indirect Tensile Asphalt Cracking Test (IDEAL-CT) | ASTM D8225-19 | Performance index |
| | Illinois Flexibility Index Test (I-FIT) | AASHTO TP 124-20 | Performance index |
| | Semi-Circular Bend Test (Louisiana method) | LADOTD TR 330, ASTM D8044-16 | Performance index |
| | Overlay Test | NJDOT B-10Tex-248-F | Performance index; predicted performance from empirical relationships |
| | $N_{flex}$ Factor | AASHTO TP 141-20 | Performance index |
| Rutting | Asphalt Pavement Analyzer | AASHTO T 340-10 | Performance index |
| | Flow Number Test | AASHTO T 378-17 | Performance index; predicted performance from empirical relationships |
| | Hamburg Wheel-Tracking Test | AASHTO T 324-19 | Performance index |
| | Stress Sweep Rutting | AASHTO TP 134-19 | Performance index; predicted performance |
| | High Temperature Indirect Tension | None | Performance index |
| | Rapid Shear Rutting Test (IDEAL-RT) | Draft ASTM Work Item (WK 71466) | Performance index |
| Low-temperature cracking | IDT Creep Compliance and Strength Test | AASHTO T 322-07 | Performance index; predicted performance |
| | Disc-Shaped Compact Tension (DCT) Test | ASTM D7313-20 | Performance index |
| Mixture toughness | Cantabro Test | AASHTO TP 108-14 | Performance index |
| Moisture susceptibility | Tensile Strength Ratio | AASHTO T 283-21 | Performance index |

(Approaches 1, 2, 3, or 4), the performance testing methods, and the corresponding threshold values. Some other potential challenges in developing and deploying performance specification/mix designs are listed in the following (Harvey et al. 2014; McCarthy et al. 2016; Diffenderfer and Bowers 2019; Lee et al. 2020; Kim et al. 2022a, b):

- lack of experience and confidence in developing and using the new specification/design methods;
- lack of historical data to develop material testing database, calibrate prediction models, and determine index threshold limits;
- needs to quantitatively estimate the risk and reliability in the specification and design method in a multilayer pavement system;
- needs to specify the details in the specifications/mix design methods, such as the sampling position and sampling frequency, type of mixture for performance testing (laboratory mix versus plant mix), and aging conditions in performance tests (short-term aging versus long-term aging); and
- needs to evaluate and account for the effects of the testing variabilities in the specification/mix design.

The following discussion focuses on the details and solutions to some of the common challenges in developing performance specification/mix design.

## Selection of Performance Testing Methods

Selecting the performance testing methods is one major step to develop an IPS or IPMD. The performance tests should correspond to the major types of pavement distresses in the region (Diffenderfer and Bowers 2019; Yin and West 2021). There are multiple cracking and rutting tests that can be used in an IPS or IPMD [AASHTO TP 107-18 (AASHTO 2021a); AASHTO TP 133-21 (AASHTO 2021b); AASHTO T 321-17 (AASHTO 2021c); ASTM D8237-21 (ASTM 2021a); ASTM D8225-19 (ASTM 2019); AASHTO TP 124-20 (AASHTO 2021d); LADOTD TR 330 (LADOTD 2014); ASTM D8044-16 (ASTM 2017); NJDOT B-10 (NJDOT 2007); Tex-248-F (Tex 2007); AASHTO TP 141-20 (AASHTO 2021e); AASHTO T 340-10 (AASHTO 2021f); AASHTO T 378-17 (AASHTO 2021g); AASHTO T 324-19 (AASHTO 2021h); AASHTO TP 134-19 (AASHTO 2021i); WK 71466 (ASTM 2022); AASHTO T 322-07 (AASHTO 2021j); ASTM D7313-20 (ASTM 2021b); AASHTO TP 108-14 (AASHTO 2021k); AASHTO T 283-21 (AASHTO 2021l)]. Some of the tests may require longer testing time but provide the material fundamental engineering properties along with performance indices, such as the cyclic fatigue test and the disc-shaped compact tension (DCT) test. Some of the tests are tolerance tests performed under a single testing condition. A parameter obtained from the test is usually used as the index representing the material's resistance to a certain pavement distress. When selecting the testing methods, SHAs should consider key factors such as users' experience, availability of existing data, availability of testing equipment, cost of the test, turnaround time, and the effectiveness of the test. Table 4 presents some commonly used performance testing methods. West et al. (2018) provided a nine-step guide for determining the testing methods in BMD. The features of each testing method have been documented in articles (McCarthy et al. 2016; Lee et al. 2020; Yin and West 2021) and AASHTO MP 46-20.

To develop a BMD protocol for Virginia, the state DOT considered the Cantabro test (for its uniqueness in testing durability and simplicity), the APA test (due to its historical application in the state), the overlay test (to test resistance to reflective cracking), the I-FIT test (based on previous research results), the $N_{flex}$ factor test (based on available research results in West et al. (2018) and its simplicity), and the IDEAL-CT test (for its simplicity) as candidate performance testing methods. During the evaluation, factors including the test effectiveness (by comparing test results with known mixture performance), cost of running each test, state-wide equipment distribution, training requirements, specimen preparation time, and test repeatability were considered. The Cantabro test, APA test, and IDEAL-CT tests were eventually selected (Diefenderfer and Bowers 2019; Diefenderfer et al. 2021).

Climate is another factor in the determination of performance testing methods. In Minnesota, the most common pavement distress was found to be low-temperature cracking due to its cold weather. Therefore, the Minnesota DOT included the DCT fracture energy test in the specification provision for BMD (Johanneck et al. 2015). Likewise, in addition to the HWTT and the SCB test, the DCT test was also required by the Wisconsin DOT (Paye 2014).

SHAs usually validate the performance testing results with their local mixtures. One method is to compare the ranking of the mixtures obtained from the testing results with the known mixture performance. Another approach is to conduct performance tests on local mixtures with systematically varying design parameters. The candidate test should be able to distinguish mixtures with different properties and design variables (such as different binder contents, air voids, gradation, RAP contents, aging levels, binder type, and so forth). In the research conducted by the Indiana DOT, the I-FIT SCB test and the HWTT were evaluated as the primary candidate testing methods for BMD (Lee et al. 2020). The effectiveness of the tests was assessed by their sensitivity to different design variables, which included the specimen air void, specimen geometry, and binder performance grade. The results indicated that the I-FIT did not follow the engineers' intuition that asphalt specimens with lower air void contents or modified binders should yield higher FI values. Therefore, the I-FIT test was determined not to be implemented in a BMD in Indiana.

## *Determination of Performance Failure Criteria*

The determination of the performance index threshold values is another important step in developing IPS or IPMD. In this article, methods to determine index limits in existing research are categorized into two primary approaches: using the performance of existing mixtures and using the existing performance index threshold from other testing methods.

### Correlation with Performance of Existing Mixtures

One practical way to determine the performance index criteria is to use the mixtures with known field performance. Meanwhile, comparing the testing results with field performance can validate the effectiveness of the test. In the study conducted by FHWA (Golalipour et al. 2021), the same set of asphalt mixtures was tested using multiple cracking testing methods, i.e., cyclic fatigue, I-FIT, IDAL-CT, overlay, the $\mathrm{N_{flex}}$ indirect tension test, and the Cantabro abrasion loss test. The testing results were compared with the observed cracking on the accelerated lane facility (ALF). The asphalt mixture on each test lane varied in binder performance grades and contents of RAP and reclaimed asphalt single (RAS). The comparison results showed that most of the cracking tests had good correlations with the amount of cracking observed on test lanes. Among the testing methods, the $S_{app}$ parameter from the cyclic fatigue test, $\mathrm{CT_{Index}}$ from the IDEAL-CT test, and the $\mathrm{N_{flex}}$ parameter from the indirect tension test showed the highest correlation (approximately 0.6).

Buttlar et al. (2020) suggest using field performance to determine performance criteria in a study to support the development of BMD specifications for the Missouri DOT. The study showed that DCT, I-FIT, and IDEAL-CT testing results had good correlations with the observed field performance, and the DCT test results yielded the highest $R^2$. The results also indicated that the scores from the pavement condition rating system (PASER) had higher correlations than the IRI. The PASER deterioration rate was then used to determine the thresholds for the performance tests. To calculate the threshold values for roads with different criticality levels, the relationship between the PASER deterioration rate was first obtained. The deterioration rate varied between 1.0 and 0, and the lower rate indicated superior materials were used. The values of the performance indices corresponding to the PASER deterioration rate at 0.4, 0.3, and 0.25 were used as the initial thresholds for mixtures targeting low, medium, and high criticality levels, respectively. The final values could be obtained after initial thresholds are adjusted with the consideration of aging and testing variability. A similar method was also applied in a study conducted for the development of BMD for Illinois (Buttlar et al. 2021).

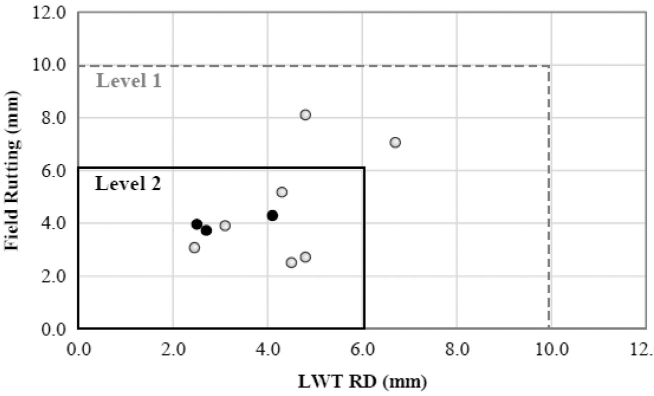

**Fig. 4.** Tentative guidelines of laboratory rutting performance indicators. (LWT RD = Loaded Wheel Tracking Rut Depth.) (Reprinted from Mohammad et al. 2016a.)

In a study conducted in Louisiana, the threshold values for the HWTT rutting test and the SCB cracking test were determined based on the correlation between the testing results and field performance (Mohammad et al. 2016a). To set the rutting limit, an enclosed area under 6 mm of rutting in both field measurement and laboratory test was created for Level 2 traffic, as presented in Fig. 4, and a 10 mm by 10 mm area was made for Level 1 traffic volume. The rutting limits of 6 mm and 10 mm were determined for Level 2 and Level 1 traffic, respectively.

SHAs sometimes do not have sufficient field performance records for the tested mixtures when a new test method is introduced. In this case, Diffenderfer and Bowers (2019) proposed three approaches to determine performance criteria based on testing results from commonly used mixtures. Method 1 adopted the minimum value from the testing results as the threshold, assuming all the mixtures would provide satisfactory performance since they had passed the existing volumetric criteria. This method would have the lowest risk for SHAs to initiate the BMD implementation. Method 2 used the average value of the testing results from all tested mixtures, which might lead to an immediate improvement by introducing BMD and yet might result in half of the existing mixtures being rejected or redesigned. Method 3 considered the average value and the testing variability. The SHA can decide on a suitable strategy to determine the criteria. For example, in Virginia, since historically no severe rutting was reported, Method 1 for APA was adopted. However, because the cracking on pavements was a major concern, Method 2 was used to determine the IDEAL-CT index limit.

Lee et al. (2020) developed performance criteria for Indiana mixtures using methods similar to the aforementioned Method 2. The Indiana DOT intended to use the thresholds to exclude the poorest quality mixtures. The density function and cumulative distribution function of the reported values of the performance indices were plotted. The 10th and 20th percentile of the testing result of a

**Table 5.** Recommended threshold values for $S_{app}$ based on GDOT mixture selection criteria assuming 5% truck traffic and 1.17 ESAL factor

| Two-way ADT | Traffic level (million ESALs) | $S_{app}$ | Mix type | Remarks |
|---|---|---|---|---|
| 10,000–25,000 | >4 and ≤10 | >12 | 12.5-mm Superpave mix with PG 64-22 or PG 67-22 binder | For state routes and shoulders of interstate routes |
| 25,000–50,000 | >10 and ≤20 | >15.5 | 12.5-mm Superpave mix with polymer-modified binder | For high-ADT state routes, interstate routes when recommended by GDOT, all flexible pavement interstate ramps, and all flexible pavement roundabouts |
| >50,000 | >20 | N/A | 12.5-mm stone matrix asphalt | For interstate routes and for state routes when recommended by GDOT |

Source: Reprinted from Etheridge et al. (2019), © ASCE.

series of mixtures were hereby calculated. The final performance threshold values considered the binder performance grade and the nominal maximum aggregate size (NMAS). For example, the FI value for a 9.5 mm mixture containing PG 64-22 binder was 2.6, which was the 10th percentile limit among all the tested mixtures with the same NMAS and binder PG.

Etheridge et al. (2019) conducted a study to determine the threshold limits for the fatigue resistance index parameter, $S_{app}$, obtained from the cyclic fatigue test for the Georgia DOT. Local mixtures with different binder grades, NMAS, and polymer modifiers were tested. The testing results were consistent with the observed performance of these mixtures, and the ranking of the mixtures with different design parameters was in line with the engineers' intuition. Good correlations were also observed between the $S_{app}$ values and predicted pavement fatigue cracking from FlexPAVE and the AASHTO Pavement ME Design Software. Based on the performance of the existing mixtures, the predicted pavement performance, and the SHA's requirements for mixtures at each traffic level, the threshold values of $S_{app}$ for the Georgia DOT were determined, as presented in Table 5. The values are different from the recommendations for nationwide applications (Wang et al. 2020).

## Correlation with Other Testing Methods

Another way to evaluate the effectiveness of the candidate performance tests is to compare the testing results from different tests when there is no sufficient field performance data. A high correlation between the testing results from two or more test methods can provide some confidence. One can determine the preliminary threshold index values for a new test based on the limits of an existing test.

There are multiple cracking tests available for performance construction specifications or mix design methods. Many research studies have compared the test methods and attempted to develop correlations among the testing results. Table 6 presents the correlations between testing results from different cracking test results reported in recent studies.

In some recent studies (Zhou et al. 2020, 2021), a correlation with $R^2$ of 0.97 between the $CT_{Index}$ and the FI was presented. The limits for the $CT_{Index}$ obtained from the IDEAL-CT test were determined based on the existing threshold values of FI. As a reference, $CT_{Index} = 90$ was believed to be equivalent to $FI = 8$.

As good correlations between the results from the IDEAL-CT test and the I-FIT test had been reported in many research studies, Al-Qadi et al. (2021) studied the mechanisms behind the two tests and presented a different opinion. The researchers pointed out that a good correlation between the $CT_{Index}$ and the FI only exists in some conditions. With a notch induced in the SCB cracking test, the energy dissipates into the cracking propagation. In contrast, in the IDEAL-CT test, the energy is distributed between plastic deformation and cracking propagation. When plasticity is not dominating at low temperatures and/or high loading rates, a good correlation between the two tests can be observed. The study also used the digital image correlation results to verify the mechanisms. The test results in the study showed that under different testing conditions, the $R^2$ between the $CT_{Index}$ and the FI varied between 0.56 and 0.04.

**Table 6.** Correlations between cracking tests in recent studies

| Tests | $R^2$ | Remark | Research |
|---|---|---|---|
| I-FIT and IDEAL-CT | 0.61 and 0.81 | • Ohio coarse asphalt concrete base mixture<br>• Field cores from two projects<br>• Over 10 datapoints in each project | Garcia-Ruiz and Sargand (2021) |
| IDEAL-CT and CPR from overlay | 0.96 | • Texas Superpave mixtures<br>• limited number of tests and a relatively high margin of error | Al-Khayat and Epps Martin (2021) |
| I-FIT and IDEAL-CT | 0.74 | • 12 field projects in Wisconsin | Abdalla et al. (2021) |
| IDEAL-CT and $S_{app}$ from cyclic fatigue | 0.86 | • 6 mixtures with different binder grades and contents of recycled materials<br>• ALF test lanes | Golalipour et al. (2021) |
| I-FIT and IDEAL-CT | 0.89 | • 6 mixtures with different binder grades and contents of recycled materials<br>• FHWA ALF test lanes | Golalipour et al. (2021) |
| I-FIT and IDEAL-CT | 0.87 | • 36 mixtures from Missouri | Buttlar et al. (2020) |
| I-FIT and IDEAL-CT | 0.97 | • 9 data points from Texas mixtures | Zhou et al. (2020) |
| IDEAL-CT and CPR from Overlay | 0.90 | • 18 data points from Texas mixtures | Zhou et al. (2020) |
| I-FIT and IDEAL-CT | 0.56 | • 8 mixtures tested at low temperature and high loading rate | Al-Qadi et al. (2021) |
| I-FIT and IDEAL-CT | 0.04 | • 5 mixtures tested at intermediate temperature | Al-Qadi et al. (2021) |

Note: CPR = cracking progression rate.

**Table 7.** Threshold values of $\text{CT}_{\text{Index}}$ in existing research and current practices

| SHA | Criterion | Testing protocol | Air void (%) | Aging level |
|---|---|---|---|---|
| Missouri (MoDOT 2021a, b) | $\text{CT}_{\text{index}}$ between 32 and 60 for 100% pay, Superpave<br>CTindex between 80 and 159 for 100% pay, SMA | ASTM D8225-19 | $7 \pm 0.5$ | Short-term (AASHTO R 30) |
| Alabama (ALDOT 2020) | $\text{CT}_{\text{index}} \geq 55$ (ESAL < 1M, mix design)<br>$\text{CT}_{\text{index}} \geq 50$ (ESAL < 10M, production acceptance)<br>$\text{CT}_{\text{index}} \geq 83$ (ESAL < 30M, mix design)<br>$\text{CT}_{\text{index}} \geq 75$ (ESAL < 1M, production acceptance)<br>$\text{CT}_{\text{index}} \geq 110$ (ESAL < 10M, mix design)<br>$\text{CT}_{\text{index}} \geq 100$ (ESAL < 30M, production acceptance) | ALDOT 459 | $7 \pm 1$ | Short-term (AASHTO R 30) |
| Oklahoma (Cross and Li 2019) | CTindex ≥ 80 | ASTM D8225-19 | $7 \pm 0.5$ | Short-term (AASHTO R 30) |
| Tennessee (Yin and West 2021) | $\text{CT}_{\text{index}} \geq 50$ (state routes, <10,000 ADT)<br>$\text{CT}_{\text{index}} \geq 75$ (state routes, >10,000 ADT)<br>$\text{CT}_{\text{index}} \geq 100$ [interstates and state routes (controlled access), <10,000 ADT] | ASTM D8225-19 | $7 \pm 0.5$ | Short-term (AASHTO R 30) |
| Texas (TxDOT 2019) | Correlation between IDEAL-CT and overlay test | Tex-250-F | $7 \pm 0.5$ | Short-term |

In terms of rutting test methods, new tests have also been introduced to the family, for example, the IDEAL-RT test. In the study conducted by Al-Khayat and Epps Martin (2021), the testing results showed that the IDEAL-RT test and the HWTT could both effectively evaluate the rutting resistance of asphalt mixtures. A correlation with $R^2$ of 0.92 between the measured APA rut depth and the shear strength from the IDEAL-RT test was reported by Zhou et al. (2020, 2021). To develop a BMD framework in Nebraska (Nsengiyumva et al. 2020), a new rutting test called the gyratory stability test was proposed. A good correlation between the tested G-stability from the gyratory stability test and the flow number was presented. The performance limit was then determined based on the relationship between the two test results and the known flow number limits. The threshold values of prevalent performance tests in some current practices are presented in Table 7.

## Testing Variability and Uncertainty

Variability exists in all performance tests, and the uncertainty derived from the testing variability needs to be accounted for in the performance specification/mix design. The testing variability of different performance tests has been reported in many studies. In the study conducted by Golalipour et al. (2021), mixtures from FHWA ALF lanes were tested using six performance testing methods: the Cantabro mass loss test, the $N_{\text{flex}}$ test, the overlay test, the I-FIT test, the IDEAL-CT test, and the cyclic fatigue test. The coefficients of variation (COV) are presented in Table 8. The test result showed that the $S_{app}$ parameter from the cyclic fatigue test and the critical fracture energy parameter from the overlay test had the lowest COV. The FI parameter from the I-FIT showed the highest COV among all the tests. Garcia-Ruiz and Sargand (2021) compared the testing variability of the I-FIT test and the IDEAL-CT test. The CT Index from the IDEAL-CT test showed 30% less variability than the FI from the I-FIT test in the study.

To understand and estimate the testing variabilities in different laboratories, before the full deployment of the BMD, the Virginia DOT started round robin to evaluate the testing repeatability. Forty-one laboratories from the agency, contractor, and independent testing labs participated. Two different mix designs were used, and 46 sets of test specimens for each design were distributed. The Phase 1 study aimed to evaluate the IDEAL-CT test results from different labs, followed by the Phase 2 study starting in 2021 (Diefenderfer et al. 2020; VAA 2021).

Another way to estimate the testing uncertainty was demonstrated by Ding et al. (2020). The researchers computed the uncertainty in the cyclic fatigue testing results given the testing variabilities in the precedent dynamic modulus measurements and the fatigue test. The Bayesian inference-based Markov Chain Mont-Carlo method was adopted. In the analysis, the uncertainties from the previous step (i.e., dynamic modulus test) propagated into the following step (i.e., the fatigue test, in this case), and the interactions from different steps were accounted for in the final calculation. The 95% credible interval and 95% prediction interval were calculated when a different number of testing replicates were used in the analysis. Four replicates of the cyclic fatigue test were recommended considering the balance between the acceptable uncertainty of fatigue prediction and the testing efficiency.

A completed performance specification/mix design method should consider the testing variabilities and uncertainties. In the PRS developed by Caltrans (Harvey et al. 2014), the specification limits were determined based on the 95% confidence interval for

**Table 8.** Coefficient of variation (COV) of performance tests

| Mixture | Cantabromass loss (%) | $N_{\text{flex}}$ (%) | OT: crack resistance index (%) | OT: critical fracture energy (%) | FI (%) | $\text{CT}_{\text{Index}}$ (%) | $S_{app}$ (%) |
|---|---|---|---|---|---|---|---|
| Lane 3 | 20 | 18 | 17 | 4 | 28 | 30 | 3 |
| Lane 5 | 5 | 20 | 19 | 2 | 29 | 7 | 3 |
| Lane 6 | 18 | 22 | 6 | 5 | 30 | 14 | 3 |
| Lane 7 | 18 | 31 | 11 | 1 | 29 | 40 | 6 |
| Lane 8 | 1 | 31 | 7 | 4 | 34 | 1 | 1 |
| Average | 12 | 24 | 12 | 3 | 30 | 18 | 3 |

Source: Data from Golalipour et al. (2021).

**Table 9.** Sampling position and testing frequency in performance specifications

| State | Sampling position | Testing frequency |
|---|---|---|
| Minnesota (Newcomb 2018) | Plant | Every 2,000 tons (wearing course) |
| Missouri (MoDOT 2021a) | Plant | Every 10,000 tons |
| Alabama (ALDOT 2020) | Plant | Once per day (agency) every 700 tons (contractor) |
| Virginia (Diefenderfer and Bowers 2019) | Plant | Every 1,000 tons (agency) every 500 tons (contractor) |
| California (Yin and West 2021) | Plant | Every 10,000 tons or once per project (HWTT)<br>three specimens per day (flow number and I-FIT) |
| Illinois (IDOT 2021) | Behind paver | Once per day or every 1,000 tons |
| Louisiana (LDOTD 2016) | Plant | Every 2,000 tons |
| Louisiana (Mohammad et al. 2016a) | Field core | 25 cores per lot |
| Vermont (VTrans 2019) | Plant | Every 3,000 tons |

the given measured material properties. Similarly, Buttlar et al. (2020) performed cracking tests (i.e., DCT test, I-FIT test, and the IDEA-CT test) to develop a local BMD method. It was reported that while the average DCT COV was 19.5% for field section evaluations, the I-FIT test had a much higher average COV (52.2%). The recommended threshold values for the performance tests included two times of standard deviation to accommodate the testing variabilities.

### Aging Conditioning and Sampling for Design and QA

Aging is one important factor for mix performance evaluation during design and QA. The short-term and long-term aging conditioning should be performed to simulate the aging effects during construction and service. However, the conditioning is time consuming. Short-term aging may take 2 to 4 h, while long-term aging usually requires at least eight hours (Zhou et al. 2021). Based on the binder chemistry and rheology study in the NCHRP 9-54 project, depending on the climate conditions, one may need to oven condition loose mixes for 24 to 696 h at 95 °C (Kim et al. 2018) to simulate the long-term field aging. Moreover, during production, the aging level of the mix samples may vary depending on the sampling time, storage time in the silo, hauling time to the construction site, and so forth. Thus, it is critical to developing methodologies to incorporate aging factors to implement performance mix design and construction specification.

Researchers have provided several potential strategies to consider the aging effects. For example, Zhou et al. (2020, 2021) proposed a short-term aging protocol with a loose mixture conditioned at compaction temperature for 2 h for both cracking and rutting tests during mix design and QA and a midterm aging procedure (20 h at 100°C) following the short-term aging treatment before the loose mixture were reheated and compacted to fabricate cracking testing samples for performance verification during mixture acceptance. To minimize the inconsistencies of aging levels in mix samples, the Missouri DOT (MoDOT) requires the QC and QA samples for IDEAL-CT cracking tests to be compacted in the plant laboratory without reheating the loose mixes in their recent specification (Balanced Mix Design Performance Testing for Job Mix Approval NJSP-21-08A).

In addition, since the performance tests usually have a longer turnaround time than the conventional volumetric tests, the sampling method and the testing frequency should be well-planned. Williams et al. (2004) conducted thorough research for the development of a performance specification, and a study about sampling methods was included. Different sampling methods, including sampling in a haul truck, sampling in a haul truck from a platform, and sampling on a roadway behind a paver, were compared, and the sampling procedure for each method was documented. It was found that the mixture test results correlated better with tank-sampled binder than with the truck-recovered samples. The testing results also indicated that binder aging was still occurring during the transport and laydown of the mixture.

Tremendous research effort has been spent to understand the difference between asphalt samples prepared by different methods (lab produced versus plant produced). Mohammad et al. (2016b) found that the process-based factors (i.e., return of baghouse fine, delay in specimen fabrication, aggregate absorption, aggregate hardness, and stockpile moisture content) did not have significant effects on the volumetric properties and mechanical properties. However, the sample preparation methods [i.e., plant-mix field-compact (PF), lab-mix lab-compact (LL), and plant-mix lab-compact (PL)] had impact on the mechanical properties of the mixtures. The PF samples were significantly softer than the LL and PL samples at the same air void level, and PF samples consistently yielded higher rut depth in the loaded-wheel test. The study proposed conversion factors to convert the tested rut depth with one sample preparation method to rut depth with other preparation methods. The finding was confirmed by Liu et al. (2017). Liu et al. (2017) also concluded that the variability in volumetric property measurements was found generally lower than in the mechanical property measurements. Daniel et al. (2018) found that the reheating process during the preparation of PL samples had a great impact on the mixture properties. A similar trend was also reported by Al-Khayat and Epps Martin (2021). In addition, several research studies have indicated that the measured fatigue cracking resistance of LL mixtures was consistently higher than that of plant-mixed mixtures (Johanneck et al. 2015; Newcomb 2018; Lee et al. 2020). The sampling position and testing frequency in existing research and current practices are presented in Table 9.

## Summary and Recommendations

In this state-of-the-art review, the efforts in ensuring the durability and performance of asphalt mixture in mix design and production are documented. The development of the performance specification and performance mix design is summarized. The challenges and solutions in incorporating performance tests in asphalt productions are discussed. The paper aims to provide hints to the asphalt community in developing holistic methodologies and integrating QA in the performance mix design. The summary and future recommendations are stated as follows.

### Under the Framework of Volumetric Design

- Some states have proposed to increase the VMA by 1.5 to 2.0% and remove the upper limit of VFA. Some agencies increased the design air void or increased the $N_{des}$ aiming to gain more compaction effort during field placement. Some others decreased the design air void or decreased the $N_{des}$ to create more space for the binder.

- SHAs should be careful when adjusting the volumetric requirements. The interactions between volumetric variables should be thoroughly considered. Increasing the VMA can theoretically generate more room for binder; however, the adjustment may lead to changes in gradation and mixture compactibility.
- It is difficult to use the volumetric-based design method only to accommodate the usage of recycled materials and support contractors to apply new technologies in asphalt mixtures.

### *Development of Performance Specification and Performance Mix Design*

- The definition and history of different types of construction specifications are summarized in this article. The relationships among all types of construction specifications are illustrated in Fig. 1.
- The paper categorizes performance specification into index-based performance specification and predictive performance specification that include PRS and PBS.
- Predictive performance specifications utilize fundamental material engineering properties to predict pavement performance and use the difference between the as-design and as-built pavements to determine the pay adjustments. While no true PBS exists in practice due to difficulties in measuring the fundamental material properties, PRS is one solution to the concern, which uses AQCs to estimate fundamental material properties and predict pavement performance. Challenges to deploying PRSs include effectively quantifying and encompassing the risk and reliability, providing training and instructions for SHAs and contractors, developing local material database and local model calibration coefficients, and simplifying the performance tests.
- IPSs are QA specifications with simple performance tests incorporated. Compared to predictive performance specifications, IPSs are more intuitive and can adopt performance tests that can be completed in a timely manner. However, selecting suitable testing methods and determining appropriate index threshold limits are major challenges in IPSs. Besides, IPSs have limited capability to consider project-specific information such as traffic, structure, and climate conditions.
- Four approaches in balanced mix design/IPMD and one predictive performance mix design method are demonstrated. The five approaches are compared based on six evaluation criteria proposed in this paper.

### *Challenges and Solutions in Developing and Deploying Performance Specification/Mix Design*

- Selecting the testing method is one important step to develop a BMD. SHAs should consider key factors such as users' experience, availability of existing data, availability of testing equipment, cost of the test, turnaround time, and the effectiveness of the test when selecting a performance test.
- One major challenge to developing an index-based performance specification is to determine the appropriate index threshold limit. This article provided two approaches: using the performance of existing mixtures and using the existing performance index threshold from other testing methods.
- When there is no sufficient field performance data, preliminary threshold values can be determined using testing results from existing mixtures. Three approaches to selecting the limits are listed in the paper.
- Many researchers have reported good correlations between different performance tests. The testing results from the I-FIT test and the IDEAL-CT test were found to have high correlations in several studies. However, some researchers (Al-Qadi et al. 2021) pointed out that a good correlation could only exist under limited conditions. The relationships between two tests can be used to determine the threshold values for a new test method.
- Different performance tests have different levels of COV. The testing variability should be accounted for in the determination of the threshold limits.
- Researchers have found the lab-mix lab-compacted samples generally performed better than the plant-mix samples. Reheating the plant-mix loose mixes during sample preparation can have a great impact on the testing results. Sampling position and testing frequency in current practices are presented in this paper.

### *Future Research Directions*

- Predictive performance specification/mix design is more complicated but may have long-term benefits compared to index-based methods. Researchers may continue to simplify the testing methods and the model calibration process to further practicalize the framework.
- The risk and reliabilities in using the index-based performance/mix design should be quantitatively evaluated. The pay adjustments determined based on the index values should be further justified with systematic research studies.
- Comprehensive material testing database and pavement performance database should be established. The performance of the mixtures with performance testing results should continue to be monitored. The long-term benefits of using performance specification/mix design can be evaluated.
- To further simplify the performance specification/mix design, one way is to accurately predict the performance testing results from simple measurable variables. New technologies like the machine learning algorithm can contribute to this direction.

## Data Availability Statement

All data, models, and code that support the findings of this study are available from the corresponding author upon reasonable request.

## Acknowledgments

This study was funded by the National Center for Transportation Infrastructure Durability & Life-Extension (TriDurLE). The authors gratefully acknowledge the TriDurLE for their financial support.

## References

AASHTO. 2003. *Major types of transportation construction specifications: A guideline to understand their evolution and application*. Washington, DC: AASHTO.

AASHTO. 2021a. *Standard method of test for determining the damage characteristic curve and failure criterion using the asphalt mixture performance tester (AMPT) cyclic fatigue test*. AASHTO TP 107-18. Washington, DC: AASHTO.

AASHTO. 2021b. *Standard method of test for determining the damage characteristic curve and failure criterion using small specimens in the asphalt mixture performance tester (AMPT) cyclic fatigue test*. AASHTO TP 133-21. Washington, DC: AASHTO.

AASHTO. 2021c. *Standard method of test for determining the fatigue life of compacted asphalt mixtures subjected to repeated flexural bending*. AASHTO T 321-17. Washington, DC: AASHTO.

AASHTO. 2021d. *Standard method of test for determining the fracture potential of asphalt mixtures using the Illinois flexibility index test (I-FIT)*. AASHTO TP 124-20. Washington, DC: AASHTO.

AASHTO. 2021e. *Standard method of test for determining the indirect tensile Nflex factor to assess the cracking resistance of asphalt mixtures*. AASHTO TP 141-20. Washington, DC: AASHTO.

AASHTO. 2021f. *Standard method of test for determining rutting susceptibility of hot mix asphalt (HMA) using the asphalt pavement analyzer (APA)*. AASHTO T 340-10. Washington, DC: AASHTO.

AASHTO. 2021g. *Standard method of test for determining the dynamic modulus and flow number for asphalt mixtures using the asphalt mixture performance tester (AMPT)*. AASHTO T 378-17. Washington, DC: AASHTO.

AASHTO. 2021h. *Standard method of test for Hamburg wheel-track testing of compacted asphalt mixtures*. AASHTO T 324-19. Washington, DC: AASHTO.

AASHTO. 2021i. *Standard method of test for moisture–density relations of soil–cement mixtures*. AASHTO TP 134-19. Washington, DC: AASHTO.

AASHTO. 2021j. *Standard method of test for determining the creep compliance and strength of hot mix asphalt (HMA) using the indirect tensile test device*. AASHTO T 322-07. Washington, DC: AASHTO.

AASHTO. 2021k. *Standard method of test for abrasion loss of asphalt mixture specimens*. AASHTO TP 108-14. Washington, DC: AASHTO.

AASHTO. 2021l. *Resistance of compacted asphalt mixtures to moisture-induced damage*. AASHTO T 283-21. Washington, DC: AASHTO.

AAT (Advanced Asphalt Technologies) and LLC. 2011. *Manual for design of hot mix asphalt with commentary*. NCHRP Rep. No. 673. Washington, DC: Transportation Research Board.

Abdalla, A., A. Faheem, A. Hosseini, and H. Titi. 2021. "Performance related asphalt mixtures characterization." In *Proc., Transportation Research Board 100th Annual Meeting*. Washington, DC: Transportation Research Board.

ALDOT (Alabama DOT). 2020. *Balanced asphalt mix design for local roads*. Special Provisions 18-0806(2). Birmingham, AL: ALDOT.

Al-Khayat, H., and A. Epps Martin. 2021. "Balanced mix design approach for Superpave hot-mix asphalt (HMA) mixtures with RAP." In *Proc., Transportation Research Board 100th Annual Meeting*. Washington, DC: Transportation Research Board.

Al-Qadi, I. L., I. M. Said, U. M. Ali, and J. R. Kaddo. 2021. "The truth and myth about the cracking mechanism of asphalt concrete." In *Proc., Transportation Research Board 100th Annual Meeting*. Washington, DC: Transportation Research Board.

ARA (Applied Research Associates). 2004. *Guide for mechanistic-empirical design of new and rehabilitated pavement structures*. Final Rep. No. of NCHPR 1-37A Project. Washington, DC: Transportation Research Board of the National Academies.

Ashouri, M., Y. D. Wang, Y. Choi, and Y. R. Kim. 2021. "Development of healing model and simplified characterization test procedure for asphalt concrete." *Constr. Build. Mater.* 271 (Feb): 121515. https://doi.org/10.1016/j.conbuildmat.2020.121515.

Asphalt Institute. 2015. *MS-2 asphalt mix design methods*. 7th ed. Lexington, KY: Asphalt Institute.

ASTM. 2017. *Standard test method for evaluation of asphalt mixture cracking resistance using the semi-circular bend test (SCB) at intermediate temperatures*. ASTM D8044-16. West Conshohocken, PA: ASTM.

ASTM. 2019. *Standard test method for determination of cracking tolerance index of asphalt mixture using the indirect tensile cracking test at intermediate temperature*. ASTM D8225-19. West Conshohocken, PA: ASTM.

ASTM. 2021a. *Standard test method for determining fatigue failure of asphalt-aggregate mixtures with the four-point beam fatigue device*. ASTM D8237-21. West Conshohocken, PA: ASTM.

ASTM. 2021b. *Standard test method for determining fracture energy of asphalt mixtures using the disk-shaped compact tension geometry*. ASTM D7313-20. West Conshohocken, PA: ASTM.

ASTM. 2022. *Standard test method for determination of rutting tolerance index of asphalt mixture using the ideal rutting test*. Work Item 71466. West Conshohocken, PA: ASTM.

Bennert, T. 2020. "New York's State's balanced mixture design (BMD) research efforts." In *Proc., Conf. Northeast Asphalt Users Producer Group (NEAUPG)*. Storrs, CT: Univ. of Connecticut.

Bennert, T., E. Sheehy, R. Blight, S. Gresavage, and F. Fee. 2014. *Implementation of performance-based specifications for asphalt mix design and production quality control for New Jersey: Circular number E-C189*. Washington, DC: Transportation Research Board of the National Academies.

Buttlar, W. G., B. Jahangiri, P. Rath, H. Majidifard, L. Urrar, J. Meister, and H. Brown. 2021. *Development of performance-related asphalt mix design specification for the Illinois Tollway*. Final Report. Downers Grove, IL: Illinois State Toll Highway Authority.

Buttlar, W. G., L. Urra-Contreras, B. Jahangiri, P. Rath, and H. Majidifard. 2020. *Support for balanced asphalt mixture design specification development in Missouri*. CMR 20-010. Springfield, MO: Missouri DOT.

Christensen, D. W., Jr., and R. F. Bonaquist. 2006. *Volumetric requirements for Superpave mix design*. NCHRP Rep. No. 567. Washington, DC: Transportation Research Board.

Coleri, E., S. Sreedhar, and I. A. Obaid. 2020. *Development of a balanced mix design method in Oregon*. FHWA-OR-RD-21-03. Salem, OR: Oregon DOT.

Cross, S., and J. Li. 2019. *Implement balanced asphalt mix design in Oklahoma*. FHWA-OK-19-01. Oklahoma City: Oklahoma DOT.

Daniel, J. S., M. Corrigan, C. Jacques, R. Nemati, E. Dave, and A. Congalton. 2018. "Comparison of asphalt mixture specimen fabrication methods and binder tests for cracking evaluation of field mixtures." *Road Mater. Pavement Des.* 20 (5): 1059–1075. https://doi.org/10.1080/14680629.2018.1431148.

Diffenderfer, S. D., and B. Bowers. 2019. "Initial approach to performance (balanced) mix design: The Virginia experience." *Transp. Res. Rec.* 2673 (2): 335–345. https://doi.org/10.1177/0361198118823732.

Diffenderfer, S. D., I. Boz, and J. Habbouche. 2021. *Balanced mix design for surface asphalt mixtures: Phase I: Initial roadmap development and specification verification*. FHWA/VTRC 21-R15. Virginia Beach, VA: Virginia DOT.

Ding, J., Y. D. Wang, S. Gulzar, Y. R. Kim, and B. S. Underwood. 2020. "Uncertainty quantification of simplified viscoelastic continuum damage fatigue mode using the Bayesian inference-based Markov chain Mont-Carlo method." *Transp. Res. Rec.* 2674 (4): 247–260. https://doi.org/10.1177/0361198120910149.

Epps, J. A., A. Hand, S. Seeds, T. Schulz, S. Alavi, C. Ashmore, C. L. Monismith, J. A. Deacon, J. T. Harvey, and R. Leahy. 2002. *Recommended performance-related specification for hot-mix asphalt construction: Results of the Westrack project*. NCHRP Rep. No. 455. Washington, DC: Transportation Research Board of the National Academies.

Etheridge, R. A., Y. D. Wang, S. S. Kim, and Y. R. Kim. 2019. "Evaluation of fatigue cracking resistance of asphalt mixtures using apparent damage capacity." *J. Mater. Civ. Eng.* 31 (11): 04019257. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002870.

FHWA (Federal Highway Administration). 2019. *Performance engineered pavements*. FHWA-HIF-20-005. Washington, DC: FHWA.

Fugro Consultants, Inc. and Arizona State University. 2011. *A performance-related specification for hot-mixed asphalt*. NCHRP Rep. No. 704. Washington, DC: Transportation Research Board.

Garcia-Ruiz, J., and S. Sargand. 2021. "Rational approach to evaluate asphalt concrete base course design for improving construction quality and performance." In *Proc., Transportation Research Board 100th Annual Meeting*. Washington, DC: National Academy of Science.

Golalipour, A., V. Veginati, and D. Mensching. 2021. "Evaluation of asphalt mixture performance using cracking and durability tests at a full-scale pavement facility." In *Proc., Transportation Research Board 100th Annual Meeting*. Washington, DC: Transportation Research Board.

Harvey, J., R. Wu, J. Signore, I. Basheer, S. Holikatti, P. Vacura, and T. J. Holland. 2014. *Performance-based specification: California experience* *to date: Circular number E-C189*. Washington, DC: Transportation Research Board of the National Academies.

Hughes, C. 2005. *State construction quality assurance programs*. NCHRP Synthesis 346. Washington, DC: Transportation Research Board of the National Academies.

Hughes, C. S., J. S. Moulthrop, S. Tayabji, R. M. Weed, and J. L. Burati. 2012. *Quality-related pay adjustment factors for pavements*. NCHRP Research Results Digest 371. Washington, DC: Transportation Research Board of the National Academies.

IDOT (Illinois DOT). 2021. *Special provision for hot-mix-asphalt—Mixture design verification and production (modified for I-FIT)*. Chicago: IDOT.

Jeong, J., Y. D. Wang, A. Ghanbari, C. Nash, D. Nener-Plante, B. S. Underwood, and Y. R. Kim. 2020. "Pavement performance predictions using performance-volumetric relationship and evaluation of construction variability: Example of MaineDOT shadow project for the development of performance-related specification." *Constr. Build. Mater.* 263 (2020): 120150. https://doi.org/10.1016/j.conbuildmat.2020.120150.

Johanneck, L., J. Geib, D. Van Deusen, J. Garrity, C. Hanson, and E. Dave. 2015. *DCT low Temperature fracture testing pilot project*. MN/RC 2015-20. Saint Paul, MN: Minnesota DOT.

Kim, Y. R., C. Castorena, M. Elwardany, F. Fousefi Rad, S. Underwood, A. Gundla, P. Gudipudi, M. Farrar, and R. Glaser. 2018. *Long-term aging of asphalt mixtures for performance testing and prediction*. NCHRP Rep. No. 871. Washington, DC: Transportation Research Board.

Kim, Y. R., M. Guddati, Y. Choi, D. Kim, A. Norouzi, Y. D. Wang, B. Keshavarzi, M. Ashouri, A. Ghabari, and A. D. Wargo. 2022a. *Hot-mix asphalt performance-related specification based on viscoelastoplastic continuum damage (VEPCD) models*. FHWA-HRT-21-093. McLean, VA: Federal Highway Administration.

Kim, Y. R., B. S. Underwood, M. N. Guddati, A. Ghanbari, Y. D. Wang, B. Keshavarzi, J. Jeong, D. Mocelin, F. Pivetta, and N. Saleh. 2022b. *Development and deploy performance related specifications (PRS) for asphalt pavement construction*. McLean, VA: Federal Highway Administration.

Ksaibati, K., and N. Butts. 2003. *Evaluating the impact of QC/QA programs on asphalt mixture variability*. MPC Rep. No. 03-146. Fargo, ND: Mountain-Plains Consortium.

LADOTD. 2014. *Evaluation of asphalt mixture crack propagation using the semi-circular bend test (SCB)*. LADOTD TR 330. West Conshohocken, PA: LADOTD.

LDOTD (Louisiana DOT and Development). 2016. *Louisiana standard specifications for roads and bridges*. Baton Rouge, LA: LDOTD.

Lee, J., J. E. Haddock, D. Alvarez, and R. R. Rastegar. 2020. *Quality control and quality assurance of asphalt mixture using laboratory rutting and cracking test*. FHWA/IN/JTRP-2019/19. West Lafayette, IN: Purdue Univ.

Liu, J., S. Zhao, and S. Saboundjian. 2017. "Variability of composition, volumetric, and mechanic properties of hot mix asphalt for quality assurance." *J. Mater. Civ. Eng.* 29 (2): D4015004. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001481.

Maupin, G. W. 2003. *Additional asphalt to increase the durability of Virginia's Superpave surface mixes*. VTRC 03-R15. Charlottesville, VA: Virginia Transportation Research Council.

McCarthy, L. M., J. Callans, R. Quigley, and S. V. Scott III. 2016. *Performance specifications for asphalt mixtures: A synthesis of highway practice*. NCHRP Synthesis 492. Washington, DC: Transportation Research Board of the National Academies.

McDaniel, R. S., R. B. Leahy, G. Huber, J. S. Moulthrop, and T. Ferragut. 2011. *The Superpave mix design system: Anatomy of a research program*. Washington, DC: Transportation Research Board. https://doi.org/10.17226/22812.

MoDOT (Missouri DOT). 2021a. *Balanced mix design performance testing and increased density*. NJSP-20-01B. Jefferson City, MO: MoDOT.

MoDOT (Missouri DOT). 2021b. *Balanced mix design performance testing for job mix approval*. NJSP-21-08. Jefferson City, MO: MoDOT.

Mogawer, W. S., A. J. Austerman, R. Kluttz, and M. Roussel. 2012. "High-performance thin-lift overlays with high reclaimed asphalt pavement content and warm-mix asphalt technology: Performance and workability characteristics." *Transp. Res. Rec.* 2293 (1): 18–28. https://doi.org/10.3141/2293-03.

Mohammad, L., M. Kim, and H. Challa. 2016a. *Development of performance-based specification for Louisiana asphalt mixtures*. FHWA/LA.14/558. Baton Rouge, LA: Louisiana Transportation Research Center.

Mohammad, L. N., M. A. Elseifi, S. B. Cooper, C. S. Hughes, J. W. Button, and L. D. Ervin Jr. 2016b. *Comparing the volumetric and mechanical properties of laboratory and field specimens of asphalt concrete*. NCHRPT Rep. No. 818. Washington, DC: Transportation Research Board.

Montoya, M. A., M. R. Pouranian, and J. E. Haddock. 2018. "Increasing asphalt pavement density through mixture design: A field project." *J. Assoc. Asphalt Paving Technol.* https://doi.org/10.12783/aapt2018/33800.

NAPA (National Asphalt Pavement Association). 2019. "Market facts." Accessed August 8, 2019. https://www.asphaltpavement.org/index.php?option=com_content&view=article&id=891.

Newcomb, D. 2018. *Balanced design of asphalt mixtures*. MN/RC 2018-22. Saint Paul, MN: Minnesota DOT.

Newcomb, D. E., E. Arambula, F. Yin, J. Zhang, A. Bhasin, W. Li, and Z. Arega. 2015. *Properties of foamed asphalt for warm mix asphalt applications*. NCHRP Rep. No. 807. Washington, DC: Transportation Research Board of the National Academies.

NJDOT (New Jersey Department of Transportation). 2007. *Overlay test for determining crack resistance of HMA*. NJDOT B-10. Ewing Township, NJ: NJDOT.

NJDOT (New Jersey DOT). 2011. "NJDOT survey on dust to binder ratio for Superpave." Accessed June 16, 2019. https://www.pavementinteractive.org/wp-content/uploads/2011/01/SuperpaveDBRatioSurvey.pdf.

Nsengiyumva, G., Y. Kim, and J. Hu. 2020. *Feasibility and implementation of balanced mix design in Nebraska*. SPR-PI(19) M080. College Station, TX: Texas A&M Univ.

Paye, B. 2014. "WisDOT pilot high recycle mixes." Accessed September 1, 2015. https://www.wispave.org/wp-content/uploads/dlm_uploads/Breakout-2014-Paye-High-Recycle.pdf.

Scott, S. 2019. *Improving durability of asphalt mixtures*. CDOT-2019-06. Denver: Colorado DOT.

Tran, N., F. Yin, F. Leiva, C. Rodezno, G. Huber, and B. Pine. 2016. *Adjustments to the Superpave volumetric mixture design procedure for selecting optimum asphalt content*. Final Rep. No. of NCHRP 20-07/Task 412. Washington, DC: Transportation Research Board.

TRB (Transportation Research Board). 2005. *Glossary of highway quality assurance terms: Transportation research circular, number E-C074*. Washington, DC: TRB.

TRB (Transportation Research Board). 2018. *Glossary of transportation construction quality assurance terms: Transportation research circular, number E-C235*. 7th ed. Washington, DC: TRB.

Tsai, B. W., R. Wu, J. Harvey, and C. Monismith. 2012. *Development of fatigue performance specification and its relation to mechanistic–empirical pavement design using four-point bending beam test results*. Leiden, Netherlands: CRC/Balkema.

TxDOT (Texas DOT). 2007. *Test procedure for overlay test*. Tex-248-F. Austin, TX: TxDOT.

TxDOT (Texas DOT). 2019. *Superpave mixtures—Balanced mix design*. Special Specification 3074. Austin, TX: TxDOT.

Underwood, B. S., C. Baek, and Y. R. Kim. 2012. "Simplified viscoelastic continuum damage model as platform for asphalt concrete fatigue analysis." *Transp. Res. Rec.* 2296 (1): 36–45. https://doi.org/10.3141/2296-04.

VAA (Virginia Asphalt Association). 2021. "BMD round robin: Phase 2 begins." Accessed August 10, 2021. https://vaa.mynewscenter.org/bmd-round-robin-phase-2-begins/.

VTrans (Vermont Agency of Transportation). 2019. *Superpave bituminous concrete pavement, performance engineered method*. Montpelier, VT: VTrans.

Wang, Y. D. 2019. "Development of the framework of performance-engineered mixture design for asphalt concrete." Ph.D. dissertation, Dept. of Civil, Environmental, and Construction Engineering, North Carolina State Univ.

Wang, Y. D., A. Ghanbari, B. S. Underwood, and R. Y. Kim. 2021a. "Development of framework of the predictive performance-engineered mix design procedure for asphalt mixtures." *Int. J. Pavement Eng.* (Jun): 1–16. https://doi.org/10.1080/10298436.2021.1938044.

Wang, Y. D., A. Ghanbari, B. S. Underwood, and Y. R. Kim. 2019. "Development of a performance-volumetric relationship for asphalt mixtures." *Transp. Res. Rec.* 2673 (6): 416–430. https://doi.org/10.1177/0361198119845364.

Wang, Y. D., A. Ghanbari, B. S. Underwood, and Y. R. Kim. 2021b. "Development of preliminary transfer functions for FlexPAVE™." *Constr. Build. Mater.* 266 (Part B): 121182. https://doi.org/10.1016/j.conbuildmat.2020.121182.

Wang, Y. D., B. Keshavarzi, and Y. R. Kim. 2018. "Fatigue performance predictions of asphalt pavements using FlexPAVETM with the S-VECD model and DR failure criterion." *Transp. Res. Rec.* 2672 (40): 217–227. https://doi.org/10.1177/0361198118756873.

Wang, Y. D., and Y. R. Kim. 2017. "Development of a pseudo strain energy-based fatigue failure criterion for asphalt mixtures." *Int. J. Pavement Eng.* 20 (10): 1182–1192. https://doi.org/10.1080/10298436.2017.1394100.

Wang, Y. D., A. Norouzi, and Y. R. Kim. 2016. "Comparison of fatigue cracking performance of asphalt pavements predicted by pavement ME and LVECD programs." *Transp. Res. Rec.* 2590 (1): 44–55. https://doi.org/10.3141/2590-06.

Wang, Y. D., B. S. Underwood, and Y. R. Kim. 2020. "Development of a fatigue index parameter, $S_{app}$, for asphalt mixes using viscoelastic continuum damage theory." *Int. J. Pavement Eng.* 23 (2): 438–452. https://doi.org/10.1080/10298436.2020.1751844.

West, R., C. Rodezno, F. Leiva, and F. Yin. 2018. *Development of a framework for balanced mix design*. Final Rep. No. of Project NCHRP 20-07/Task 406. Washington, DC: Transportation Research Board of the National Academies.

Williams, R. C., D. Hill, M. Zelenock, and J. Bausano. 2004. *Development of laboratory performance test procedures and trial specifications for hot mix asphalt: Final report*. RC-1410. Detroit: Michigan DOT.

Yin, F., and R. West. 2021. *Balanced mix design resource guide*. Greenbelt, MD: National Asphalt Pavement Association.

Zhou, F., S. Hu, and D. Newcomb. 2020. "Development of a performance-related framework for production quality control with ideal cracking and rutting tests." *Constr. Build. Mater.* 261 (Nov): 120549. https://doi.org/10.1016/j.conbuildmat.2020.120549.

Zhou, F., S. Hu, T. Scullion, M. Mikhail, and L. F. Walubita. 2007. "A balanced HMA mix design procedure for overlays." *J. Assoc. Asphalt Paving Technol.* 76 (Nov): 823–850.

Zhou, F., R. Steger, and W. Mogawer. 2021. "Development of a coherent framework for balanced mix design and production quality control and quality acceptance." *Constr. Build. Mater.* 287 (Jun): 123020. https://doi.org/10.1016/j.conbuildmat.2021.123020.