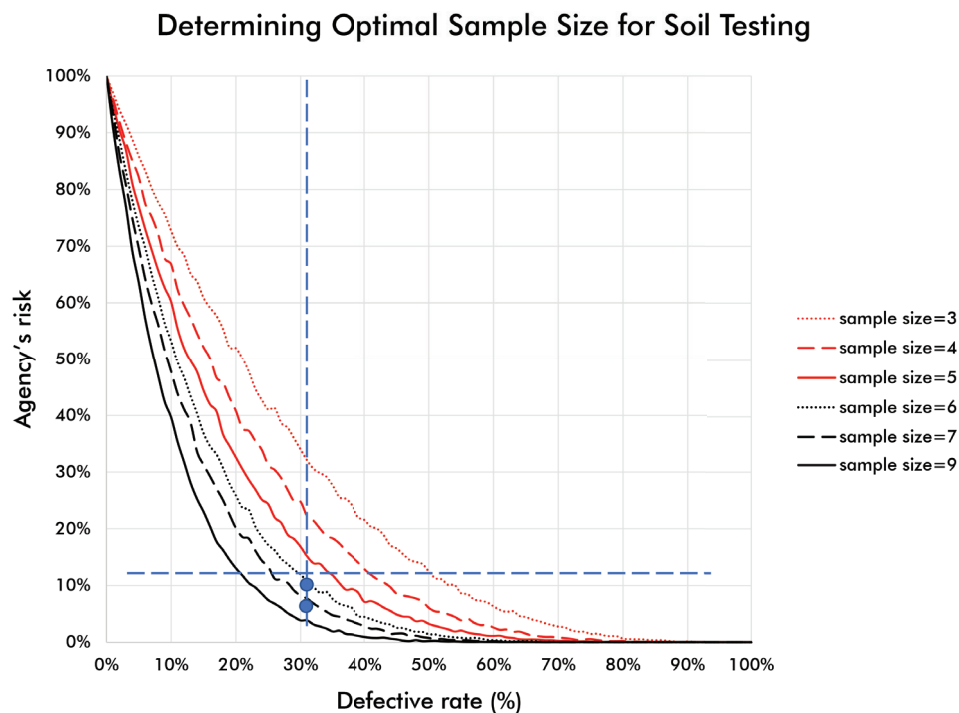


JOINT TRANSPORTATION RESEARCH PROGRAM

INDIANA DEPARTMENT OF TRANSPORTATION
AND PURDUE UNIVERSITY



Pavement Acceptance Testing: Risk-Controlled Sampling Strategy



Jiannan Cai, Qingyi Gao, Hyonho Chun, Hubo Cai

RECOMMENDED CITATION

Cai, J., Gao, Q., Chun, H., & Cai, H. (2019). *Pavement acceptance testing: Risk-controlled sampling strategy* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2019/08). West Lafayette, IN: Purdue University. <https://doi.org/10.5703/1288284316918>

AUTHORS

Jiannan Cai

Qingyi Gao

Graduate Research Assistants
Department of Statistics
Purdue University

Hyonho Chun, PhD

Associate Professor
Mathematics and Statistics
Boston University

Hubo Cai, PhD

Associate Professor
Lyles School of Civil Engineering
Purdue University
(765) 494-5028
hubocai@purdue.edu
Corresponding Author

ACKNOWLEDGMENTS

This project was made possible by the sponsorship of the Joint Transportation Research Program (JTRP) and the Indiana Department of Transportation (INDOT). The authors acknowledge the valuable assistance and technical guidance from the members of the Study Advisory Committee in the course of performing this study.

JOINT TRANSPORTATION RESEARCH PROGRAM

The Joint Transportation Research Program serves as a vehicle for INDOT collaboration with higher education institutions and industry in Indiana to facilitate innovation that results in continuous improvement in the planning, design, construction, operation, management and economic efficiency of the Indiana transportation infrastructure. https://engineering.purdue.edu/JTRP/index_html

Published reports of the Joint Transportation Research Program are available at <http://docs.lib.purdue.edu/jtrp/>.

NOTICE

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views and policies of the Indiana Department of Transportation or the Federal Highway Administration. The report does not constitute a standard, specification or regulation.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA/IN/JTRP-2019/08	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Pavement Acceptance Testing: Risk-Controlled Sampling Strategy		5. Report Date April 2019	
		6. Performing Organization Code	
7. Author(s) Jiannan Cai, Qingyi Gao, Hyonho Chun, and Hubo Cai		8. Performing Organization Report No. FHWA/IN/JTRP-2019/08	
9. Performing Organization Name and Address Joint Transportation Research Program Hall for Discovery and Learning Research (DLR), Suite 204 207 S. Martin Jischke Drive West Lafayette, IN 47907		10. Work Unit No.	
		11. Contract or Grant No. SPR-4101	
12. Sponsoring Agency Name and Address Indiana Department of Transportation (SPR) State Office Building 100 North Senate Avenue Indianapolis, IN 46204		13. Type of Report and Period Covered Final Report	
		14. Sponsoring Agency Code	
15. Supplementary Notes Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.			
16. Abstract <p>Acceptance testing is a critical aspect of the quality control and quality assurance (QC/QA) program to ensure the reliable long-term performance of pavement. A typical acceptance testing specification includes acceptable quality characteristics (AQC's), testing methods, number of samples, sample locations, and acceptance criteria. In the current practice, Indiana Department of Transportation (INDOT) accepts pavement by sampling and testing materials with a pre-determined, very low frequency at random locations, leading to a significant problem: testing results are not truly "representative" of the project because sampling is neither based on a statistical foundation, nor on the reliability concept.</p> <p>This study developed a systematic guideline that has addressed the aforementioned problem of material acceptance testing in four aspects: identifying key material properties for testing, selecting sample locations, designing acceptance criteria, and determining optimal sample size. Key material properties that are critical to the pavement long-term performance are identified by comparing with sensitive material properties in MEPDG. A random sampling mechanism was devised based on two spatial indices to control the spatial pattern of samples to minimize the influence from spatial autocorrelation. Risk-based acceptance criteria was proposed based on statistical methods to control the agency's risk at a desired level given a specific sampling and testing strategy, based on which optimal sample size is determined from a risk perspective. Cost analysis approaches were developed to estimate the total cost of acceptance testing by integrating the risk of making incorrect decision and enable the determination of optimal sample size from a cost perspective. Additionally, quality control chart was exploited as a complementary tool to ensure the consistency of the pavement quality of a project. The results of this study were validated using real data from INDOT projects, and a web tool that incorporates the newly created methods in this study was developed to assist the field pavement QA practice.</p>			
17. Key Words quality assurance, acceptance testing, risk assessment, spatial pattern analysis, optimal sample size, cost analysis, Portland cement concrete pavement, hot mix asphalt		18. Distribution Statement No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 59	22. Price

EXECUTIVE SUMMARY

PAVEMENT ACCEPTANCE TESTING: RISK-CONTROLLED SAMPLING STRATEGY

Introduction

Acceptance testing is a critical aspect of the quality control and quality assurance (QC/QA) program to ensure the reliable long-term performance of pavement. A typical acceptance testing specification includes acceptable quality characteristics (AQC's), testing methods, number of samples, sample locations, and acceptance criteria. In the current practice, the Indiana Department of Transportation (INDOT) accepts pavement by sampling and testing materials with a predetermined very low frequency at random locations. This leads to a significant problem: testing results are not truly "representative" of the project because sampling is not based on a statistical foundation nor on the reliability concept. Therefore, there is a critical need for INDOT to develop a guideline for the testing protocol from a system perspective to minimize risk and ensure the reliability of material acceptance testing.

This study developed a systematic guideline that addresses the aforementioned problem with material acceptance testing in four aspects: (1) identifying key material properties for testing, (2) selecting sample locations, (3) designing acceptance criteria, and (4) determining optimal sample size. Key material properties that are critical to pavement long-term performance are identified by comparing them with sensitive material properties in the Mechanistic-Empirical Pavement Design Guide (MEPDG). A random sampling mechanism was devised based on two spatial indices to control the spatial pattern of samples in order to minimize the influence from spatial autocorrelation. Acceptance criteria were proposed to control the agency's risk at a desired level given a specific sampling and testing strategy, based on which optimal sample size is determined from a risk perspective. Cost analysis approaches were developed to estimate the total cost of acceptance testing by integrating the risk of making incorrect decisions and to enable the determination of optimal sample size from a cost perspective. Additionally, a quality control chart was exploited as a complementary tool to ensure the consistency of pavement quality. The results of this study were validated using real data from INDOT projects, and a web tool that incorporates the newly created methods in this study was developed to assist the field pavement QA practice.

Findings

The main findings and recommendations to control the risk and improve the reliability of pavement and soil acceptance testing are summarized into four aspects, detailed as follows.

- Findings and recommendations related to key material properties.
 - Ten hot mix asphalt (HMA) properties, 10 Portland cement concrete pavement (PCCP) properties, and 9 soil properties have been identified to be very important to the pavement performance, among which 5 in HMA, 4 in PCCP, and 6 in soil are not tested in the current INDOT practice. Most of the missing items in PCCP and HMA do not have certified testing standards or are difficult and not warranted in laboratory testing. They can be substituted by default design value.
 - It is recommended that (1) HMA thickness be tested and measured directly rather than being estimated, considering its effect on pavement performance, and (2) a sampling and testing protocol be established for key material properties of soil that are currently not tested.
- Findings and recommendations related to random sampling.
 - ITM 802-based sampling leads to "gaps"—areas that will never be tested due to the limited random numbers in the published table. For earthwork, gap areas can receive insufficient compaction without being caught. For pavement, although gap areas are not large enough to accommodate a whole truckload, they lead to lower probabilities for certain trucks to be tested, compared with the true random sampling process. Such discrepancy results in a lower risk for contractors to manipulate these trucks, and a higher risk for INDOT to accept inferior materials. Therefore, the ITM 802-based sampling is less reliable than the true random sampling process that generates random numbers in real time.
 - Material properties are spatially auto-correlated. Therefore, the sample spatial pattern must be controlled to avoid spatial clustering and ensure the effectiveness of random sampling. Two spatial pattern indices, i.e., nearest neighbor index (*NNI*) and coefficient of variance (*CV*), are proposed for this purpose. *NNI* assesses the degree of spatial dispersion and is recommended to serve as the primary metrics. *CV* measures the variability of sample locations and is recommended to serve as the secondary metrics.
 - It is recommended to (1) use a real-time random number generator to determine sample locations, and (2) control the spatial pattern indices sample locations to ensure that the sample locations are spatially distributed and truly representative of the project.
- Findings and recommendations related to acceptance criteria.
 - The current acceptance methods based on either percent within limit (PWL) or sample mean at INDOT do not achieve the desired level of risk control.
 - For pavement acceptance, the newly proposed numerical *M*-method and *S_{pk}*-method incorporate sample variance and treat statistical measures of samples as random variables. They are capable of controlling the risk at a desired level and achieving higher testing power compared to the current practice at INDOT.
 - The numerical *M*-method is suitable for all scenarios and is recommended to replace the current practice, while *S_{pk}*-method can be applied when the sample size is relatively large to simplify the computation.
 - For the soil compaction acceptance, a simulation approach is proposed to assess INDOT's risk. It is recommended to conduct the simulation to examine the risk once a sampling strategy is determined, and then make a corresponding adjustment to the sampling strategy according to the desired risk.
 - Plotting QA data on a QC chart helps identify out-of-control process at both project and lot levels. It is recommended that a QC chart be used as an additional quality assurance tool to supplement the numerical *M*-method.

4. Findings and recommendations related to optimal sample size.

- The optimal sample size for pavement acceptance testing can be determined to satisfy the risk expectation of both agency and contractor using the proposed numerical M -method and S_{pk} -method. The optimal sample size for soil acceptance can be determined as the smallest value that satisfies the agency's desired risk level using the proposed simulation approach.
- Two cost analysis methods have been devised to determine the optimal sample size by minimizing the total cost of acceptance testing, one for accept/reject decision only and the other for pay factor decisions.

Implementation

The newly developed methods and corresponding results in this study have been validated using real data from INDOT projects. For practical implementation, a web tool that incorporates the newly created methods in this study has been developed based on R Shiny. It is an effective tool to facilitate the pavement and soil QA practice for the INDOT engineers and has the ability to select random samples, control the agency's risk, and determine optimal sample sizes for both pavement and soil acceptance testing. The web tool consists of seven separate tabs that correspond

to different tasks in this study. Tabs 1 and 2 generate random locations and control their spatial pattern for the random sampling in both pavement and soil. Tabs 3 and 4 control the agency's risk of making incorrect acceptance/rejection decisions and calculate the confidence interval of PWL given a specific strategy for pavement acceptance testing. Tab 5 estimates the agency's risk of not identifying defective material and determines the optimal sample size for soil acceptance testing. Tabs 6 and 7 determine the optimal sample size for pavement acceptance testing from both a risk and a cost perspective.

A phased implementation is suggested as follows:

1. Set up R Shiny website to implement the web tool.
2. Tool rollout and pilot testing by phases.

- Phase I: field crew adopt the random number generator tool to locate sample location.
- Phase II: field crew and material engineers use the risk assessment tool to determine the probability of accepting inferior products given the testing results.
- Phase III: engineers use the web tool to determine the optimal sample size.

A clear understanding of the technical principles and a set of skills are necessary to use the web tool and interpret the results. As such, training for INDOT staff is recommended.

CONTENTS

1. INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	1
1.3 Overall Objective and Work Plan.	1
1.4 Expected Benefits and Deliverables	1
2. KEY MATERIAL PROPERTIES TO PAVEMENT LONG-TERM PERFORMANCE	2
2.1 Introduction	2
2.2 Identification of Key Material Properties in MEPDG	2
2.3 Comparison with INDOT Testing Items.	3
2.4 Findings and Observations	3
3. RANDOM SAMPLING	4
3.1 Introduction	4
3.2 Evaluation of the Random Sampling Practice	5
3.3 Findings, Recommendations, and Implementation.	6
4. SPATIAL PATTERN ANALYSIS.	7
4.1 Introduction	7
4.2 Spatial Autocorrelation in Soil Compaction	7
4.3 Spatial Pattern INDEX.	7
4.4 Findings and Recommendation	8
5. RISK ASSESSMENT FOR GIVEN TESTING STRATEGIES	8
5.1 Introduction	8
5.2 Risk Assessment Methods.	9
5.3 Risk Assessment for Soil Acceptance Testing	13
5.4 Findings and Recommendations	13
6. OPTIMAL SAMPLE SIZE	14
6.1 Introduction	14
6.2 Optimal Sample Size for Pavement	14
6.3 Optimal Sample Size for Soil.	16
6.4 Implementation	16
7. COST ANALYSIS FOR DETERMINING OPTIMAL SAMPLE SIZE	16
7.1 Approach 1: Cost Analysis Considering Accept/Reject Decisions	17
7.2 Approach 2: Cost Analysis Considering Pay Factor Decisions	17
7.3 Case Study	18
7.4 Findings and Recommendations	20
8. QC CHART PATTERN ANALYSIS.	22
8.1 Quality Control (QC) Chart	22
8.2 Interpretation of QC Chart	23
8.3 Case Study	23
8.4 Findings and Recommendations	27
9. VALIDATION AND CASE STUDIES	27
9.1 Validation Using PCCP DATA.	27
9.2 Validation Using HMA DATA	29
9.3 Validation Using Soil DATA	30
9.4 Issues in Pavement Acceptance Testing Implementation Strategy	32
9.5 Findings and Recommendations	34
10. IMPLEMENTATION	35
10.1 TAB 1: Point Pattern Analysis—Pavement	35
10.2 TAB 2: Point Spatial Pattern Analysis—Soil	36
10.3 TAB 3: Testing Power—Pavement	37
10.4 TAB 4: Confidence Interval.	38
10.5 TAB 5: Testing Power-Soil	38
10.6 TAB 6: Optimal Sample Size	39
10.7 TAB 7: Cost/Benefit	39

11. SUMMARY AND RECOMMENDATIONS	46
11.1 Summary	46
11.2 Findings and Recommendations.	46
11.3 Validation and Implementation	47
REFERENCES	47

LIST OF TABLES

Table	Page
Table 2.1 Number of key properties for different material types	3
Table 2.2 Comparison of MEPDG and INDOT specification—key HMA material properties	3
Table 2.3 Comparison of MEPDG and INDOT specification—key PCCP material properties	4
Table 2.4 Comparison of MEPDG and INDOT specification—key soil material properties	4
Table 7.1 Case study settings in Approach 1	19
Table 7.2 Case study settings in Approach 2	20
Table 8.1 Interpretation of most common QC charts patterns	24
Table 8.2 Comparison among QC chart, PWL, and numerical <i>M</i> -method for case study 1	25
Table 8.3 Comparison among QC chart, PWL and numerical <i>M</i> -method for case study 2	26
Table 9.1 Comparison of testing power between single lot and merged lot	34
Table 10.1 Description of the web tool	36

LIST OF FIGURES

Figure	Page
Figure 1.1 Examples of material testing with low frequency	1
Figure 1.2 Overview of the work plan	2
Figure 3.1 Comparison between ITM 802-based sampling and true random sampling	5
Figure 3.2 Largest gap area from ITM 802-based sampling	5
Figure 3.3 Probability of being tested for each truck	6
Figure 3.4 Probability of detecting at least one truck with inferior materials	6
Figure 4.1 Spatial pattern analysis for soil CMV	7
Figure 4.2 Illustration of scenarios with varying NNI and CV	8
Figure 5.1 Hypothesis testing based on yield index	12
Figure 5.2 Testing power of HMA air voids acceptance testing	12
Figure 5.3 Workflow of testing power calculation and risk assessment for soil acceptance	13
Figure 5.4 Testing power-defective rate curves in soil acceptance testing	14
Figure 6.1 Process for determining the optimal sample size in pavement acceptance testing	15
Figure 6.2 Agency's risk with varying sample size	15
Figure 6.3 Determining the optimal sample size for soil based on the agency's risk curves	16
Figure 7.1 Determining the optimal sample size through cost analysis	17
Figure 7.2 Approach 2 workflow for determining the total cost at a given sample size n	18
Figure 7.3 Optimal sample sizes and total costs from Approach 1	19
Figure 7.4 Impact of P_{RQL} in Approach 1	19
Figure 7.5 Impact of $I_{percent}$ in Approach 1	20
Figure 7.6 Optimal sample sizes and total costs from Approach 2	21
Figure 7.7 Impact of $I_{percent}$ in Approach 2	21
Figure 7.8 Impact of the pavement quality in Approach 2	21
Figure 8.1 Typical QC chart	22
Figure 8.2 Music bar chart for case study 1	25
Figure 8.3 QC charts for case study 1	25
Figure 8.4 Music bar chart for case study 2	26
Figure 8.5 QC charts for case study 2	26
Figure 9.1 Histogram of flexural strength	27
Figure 9.2 Comparison of testing power in flexural strength acceptance testing	28
Figure 9.3 Comparison of testing power using averaged samples	29
Figure 9.4 Histogram of air voids	30
Figure 9.5 Comparison of testing power in air voids acceptance testing	31
Figure 9.6 Grid map of CMV data	31
Figure 9.7 Preprocessing of CMV data	32
Figure 9.8 Comparison of testing power	33
Figure 9.9 Discrepancy between simulated and real testing power in dataset 2	33
Figure 9.10 Comparison of random sampling with and without spatial control in dataset 2	34
Figure 9.11 Comparison of testing power between sampling version and subsampling version	35

Figure 10.1 User interface of Tab 1	36
Figure 10.2 User inputs of Tab 1	37
Figure 10.3 Outputs of Tab 1	37
Figure 10.4 User interface of Tab 2	38
Figure 10.5 User inputs of Tab 2	38
Figure 10.6 User interface of Tab 3	39
Figure 10.7 User inputs of Tab 3	39
Figure 10.8 Outputs of Tab 3	40
Figure 10.9 User interface of Tab 4	40
Figure 10.10 User interface of Tab 5	41
Figure 10.11 User interface of Tab 6	42
Figure 10.12 User inputs of Tab 6	43
Figure 10.13 Output table of Tab 6	44
Figure 10.14 Output charts of Tab 6	44
Figure 10.15 User interface of Tab 7	45
Figure 10.16 User inputs of Tab 7	45
Figure 10.17 Outputs of Tab 7	46

1. INTRODUCTION

1.1 Background

Acceptance testing is a critical aspect of the quality control and quality assurance (QC/QA) program to ensure the pavement quality. A typical acceptance testing specification includes acceptable quality characteristics (AQC's), testing methods, number of samples, sample locations, and acceptance criteria. Comparing the testing results to the predefined acceptance criteria, an engineer decides whether or not to accept the pavement, with or without penalty or bonus. Since the quality (central tendency and consistency) of pavement material properties is critical to long-term pavement performance, it is important that the state departments of transportation (DOTs) have reliable acceptance testing procedures to ensure that the samples and their testing results truly represent the whole product.

1.2 Problem Statement

In the current practice, the Indiana Department of Transportation (INDOT) accepts pavement by sampling and testing materials with a predetermined, very low frequency at random locations. For instance, for the US 50 project, (1) on 8/26/2015, only two dynamic cone penetration testing (DCPT) samples clustered in the south-west corner were taken for an entire area of approximately 47,600 ft² (see Figure 1.1a), and (2) on 9/25/2015, only four DCPT samples at two locations (two at each location) were taken for an entire area of approximately 138,000 ft² (see Figure 1.1b). Sparse samples taken at point locations are assumed to truly represent the entire area, leading to the significant problem: *testing results are NOT truly "representative" of the material properties in a project because sampling is based NEITHER on a statistical base NOR on the reliability concept.*

Adding more samples and testing locations and distributing them throughout the entire area could make the testing results more representative. *The question becomes how many locations to add and where should they be.* To answer this question, we must address the sampling frequency and testing interval requirements based on the following:

- The reliability concept—the consistency in measuring pavement and the associated risk of accepting non-consistent/inferior pavement. Low quality pavement is always associated with premature failures and requires excessive and expensive repairs.
- Considering the pavement as a system that is composed of soil/earthwork (sub-base), base course(s), and surface course. Any inferior component could lead to premature failures.
- Treating the calculated average and standard deviation of samples as statistical variants to incorporate the underlying randomness in to the process.

Therefore, there is a need to develop a guideline on the testing protocol from a system's perspective for INDOT to minimize the risk of premature pavement failure due to the lack of materials testing and inspection interval.

1.3 Overall Objective and Work Plan

The overall objective of this project is to enhance pavement quality by improving the materials testing and quality assurance process to (1) ensure testing results are representative of the composing layers of pavement and reflective of the pavement consistency, and (2) minimize INDOT's risk on accepting low quality pavement.

To achieve the ultimate goal, the following seven specific tasks are studied in this project. Figure 1.2 illustrates the overview of the work plan and how individual tasks work together to achieve the objective.

1. Identify key material properties to pavement long term performance
2. Assess the spatial pattern of sample locations
3. Determine the risk for specific testing strategies
4. Determine the optimal testing strategy for a given level of risk
5. Adjust testing strategy based on historical performance
6. Cost analysis
7. Validation and case studies

1.4 Expected Benefits and Deliverables

This project is expected to provide a technical understanding of risk-based pavement acceptance testing for

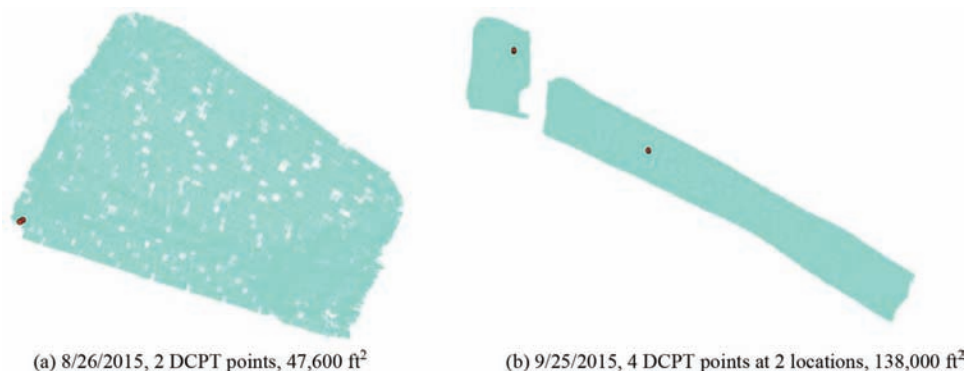


Figure 1.1 Examples of material testing with low frequency.

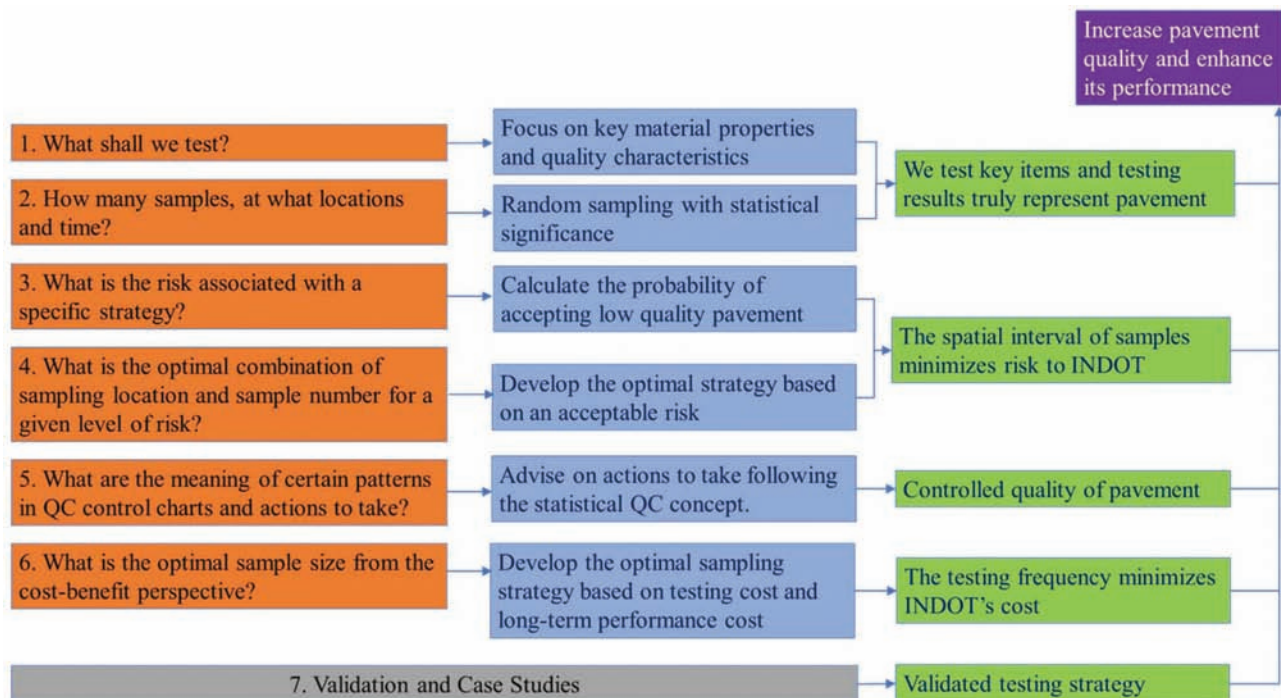


Figure 1.2 Overview of the work plan.

INDOT engineers in terms of key material properties, random sampling (including sample location and sample size), and the risk associated with the acceptance testing. The findings of this project allow INDOT to ensure the representativeness of acceptance testing and minimize its risk in accepting low-quality pavement.

Six deliverables are: (1) a synthesis study regarding the key material properties and construction quality measures that significantly affect pavement long term performance; (2) a statistical analysis procedure for INDOT engineers to calculate the risks associated with pavement given specific testing frequencies and spatial intervals; (3) an analysis procedure for INDOT engineers to determine the appropriate testing intervals and frequencies given a specific level of risk; (4) a cost and benefit analysis procedure for INDOT engineers to optimize the testing procedures and intervals; (5) a guideline on interpreting quality control charts and taking appropriate actions; and (6) a web-based tool that incorporates the newly created methods to assist the field pavement QA practice.

2. KEY MATERIAL PROPERTIES TO PAVEMENT LONG-TERM PERFORMANCE

2.1 Introduction

The long-term performance of pavement depends on the material properties to a large extent. The Mechanistic-Empirical Pavement Design Guide (MEPDG) (AASHTO, 2008) provides a theoretical base for relating the pavement long-term performance to material properties: different distresses such as cracking, rutting, faulting, and smoothness are predicted according to the

input material parameters for both flexible and rigid pavements.

In the current practice, INDOT specification (INDOT, 2016) and material testing manual (INDOT Office of Materials Management, 2015) specify testing items for different types of pavement such as Portland cement concrete pavement (PCCP) and hot mix asphalt (HMA), as well as earthwork. However, the development of INDOT material testing specification was not based on MEPDG, which means the impact of material properties on the pavement long-term performance was not taken into account when selecting testing items. Two critical questions have arisen. First, what are the key material properties to the long-term performance of pavement? Second, does INDOT test all key items in their current acceptance testing? This chapter addresses these two problems by comparing INDOT testing items with key material inputs in MEPDG, to which the predicted pavement performance is sensitive.

2.2 Identification of Key Material Properties in MEPDG

In this study, the identification of key material properties is based on the work of Schwartz, Li, Kim, Ceylan, & Gopalakrishnan (2011), who evaluated the sensitivity of the pavement performance predicted in MEPDG to the change of material properties. In their study, the performance of HMA was captured by five types of distresses: longitudinal cracking, alligator cracking, asphalt concrete (AC) rutting, total rutting, and international roughness index (IRI). The performance of PCCP was characterized by faulting, transverse cracking, and IRI.

TABLE 2.1
Number of key properties for different material types.

Material group	HMA	PCCP	Soil
Hypersensitive ($NSI > 5$)	2	0	0
Very sensitive ($1 < NSI < 5$)	4	8	3
Sensitive ($0.1 < NSI < 5$)	4	2	6
Total number of key properties	10	10	9

Note: The sensitivity level of one property varies with respect to distress types, and this table lists the highest sensitivity level of the property in each material type, Tables 2.2, 2.3, and 2.4 show the detail categories.

The sensitivity of the predicted performance is quantified using normalized sensitivity index (NSI), which represents the percentage change of predicted distress relative to the design limit due to the percentage change in the material property. Based on NSI, the material properties are divided into four groups: hypersensitive ($NSI > 5$), very sensitive ($1 < NSI < 5$), sensitive ($0.1 < NSI < 1$), and non-sensitive ($NSI < 0.1$). In this study, material properties in the hypersensitive, very sensitive, and sensitive categories are considered as key material properties. Table 2.1 lists the number of key properties for each material type (i.e., HMA, PCCP, and soil).

2.3 Comparison with INDOT Testing Items

The key material properties are grouped by the type of material (e.g., HMA, PCCP, soil) and compared with material testing items extracted from the INDOT specification (INDOT, 2016; INDOT Office of Materials Management, 2015) (see Tables 2.2, 2.3, and 2.4).

TABLE 2.2
Comparison of MEPDG and INDOT specification—key HMA material properties.

Material properties	MEPDG sensitivity analysis					Comparison with INDOT
	Longitudinal cracking	Alligator cracking	AC rutting	Total rutting	IRI	
Time-temperature dependent dynamic modulus (E^*)	HS	HS	HS	HS	HS	(air void content, asphalt mixture)
Poisson's ratio	VS	VS	VS	VS	S	N
Surface shortwave absorptivity	VS	VS	VS	VS	S	N
Thermal conductivity	S	S	S	—	—	N
Heat capacity	S	S	S	S	—	N
Total unit weight	S	S	S	S	—	Y
Asphalt binder content	VS	VS	—	—	S	Y
Asphalt binder grade	S	S	S	S	—	Y
Thickness	HS	HS	VS	VS	—	N
Air voids	VS	VS	—	—	S	Y

Notes: HS, hypersensitive; VS, very sensitive; S, sensitive.

In the last column, “Y” indicates that the corresponding item is tested in INDOT’s practice; “N” indicates that the item is not tested in INDOT’s practice; “()” indicates the item is not directly tested, but can be derived from the testing items enclosed in the parentheses.

Boldface indicates the key material properties not tested in current INDOT practice.

2.4 Findings and Observations

In Tables 2.2, 2.3, and 2.4, the key material properties that are not tested in the current INDOT practice are highlighted, 5 in HMA, 4 in PCCP, and 6 in soil. Possible reasons for the discrepancy are grouped and discussed as follows.

- No current AASHTO certified standards are available—surface shortwave absorptivity for both HMA and PCCP. MEPDG provides default design value.
- Laboratory testing is difficult and sometimes not warranted—Poisson’s ratio for HMA, PCCP, and soil. Using typical values for analysis is considered satisfactory according to MEPDG (AASHTO, 2008).
- No description in INDOT specification—heat capacity for HMA, coefficient of thermal expansion for PCCP, thermal conductivity for HMA and PCCP, and soil properties including seasonally adjusted resilient modulus, base erodibility, parameters to define the soil water characteristic curve, thickness and loss of friction. Among these factors, MEPDG provides default values for the properties of HMA and PCCP, but soil properties are supposed to be tested in the laboratory.
- Not tested in INDOT practice—thickness of HMA. In INDOT’s practice, it is estimated using the quantity of the HMA.

It is recommended that (1) HMA thickness should be tested and measured directly rather than being estimated, considering its effect on pavement performance, and (2) a sampling and testing protocol shall be established for key material properties of soil that are currently not tested, i.e., resilient modulus, Poisson’s ratio, base erodibility, parameters to define the soil water characteristic curve, thickness, and loss of friction.

TABLE 2.3
Comparison of MEPDG and INDOT specification—key PCCP material properties.

Inputs	Sensitivity analysis			Comparison with INDOT
	Faulting	Transverse cracking	IRI	
Static modulus of elasticity (E)	S	VS	S	(flexural strength)
Poisson's ratio	S	S	S	N
Unit weight	VS	VS	VS	Y
Coefficient of thermal expansion	VS	VS	VS	N
Modulus of rupture	S	VS	S	(flexural strength)
Cement content	S	S	S	Y
Water-to-cement (w/c) ratio	S	VS	S	Y
Surface shortwave absorptivity	S	VS	S	N
Thermal conductivity	S	VS	S	N
Layer thickness	S	VS	S	Y

Note: HS, hypersensitive; VS, very sensitive; S, sensitive.

Boldface indicates the key material properties not tested in current INDOT practice.

TABLE 2.4
Comparison of MEPDG and INDOT specification—key soil material properties.

Inputs	Sensitivity analysis								Comparison with INDOT
	HMA					PCCP			
	Longitudinal cracking	Alligator cracking	AC rutting	Total rutting	IRI	Faulting	Transverse cracking	IRI	
Resilient modulus (Mr)	VS	VS	S	S	S	S	S	S	N
Poisson's ratio	S	S	S	—	—	—	—	—	N
Base erodibility	—	—	—	—	—	S	S	S	N
Plasticity index	S	S	—	—	—	—	—	—	(liquid limit)
Effective grain sizes	VS	S	S	S	S	—	—	—	INDOT has some regulations regarding aggregates in 207, 904
Parameters to define the soil water characteristic curve	S	S	—	—	—	—	S	—	N
Liquid limit	S	S	S	—	—	—	—	—	Y
Thickness	VS	VS	S	—	—	S	S	—	N
Loss of friction	—	—	—	—	—	—	S	—	N

Note: HS, hypersensitive; VS, very sensitive; S, sensitive.

Boldface indicates the key material properties not tested in current INDOT practice.

3. RANDOM SAMPLING

3.1 Introduction

In the current practice, INDOT follows ITM 802 (INDOT Office of Materials Management, 2019) and uses a published table of random numbers to determine the sample locations. The main concern is whether using a published table allows guessing sample locations and manipulating the production correspondingly, leading to inferior product that escapes QA.

3.1.1 Random Sampling for Soil

In its current QA practice for soil, INDOT determines the quality of soil compaction by conducting dynamic cone penetrometer test (DCPT) at three random locations (selected in accordance with ITM 802)

for every 2,000 cubic yards of compacted soil (INDOT, 2016). Specifically, three pairs of random numbers are retrieved from the ITM 802 table and multiplied by the length and width of the covered area to determine the longitudinal and offset positions. Figure 3.1a illustrates the possible sample locations within an area in a pavement project. Since the published table has a fixed number (875) of pairs of random numbers, it is straightforward to determine which areas, such as the two large “gaps,” will never be sampled. Contrary, Figure 3.1b illustrates the sample locations using a true random number generator in real time. Increasing the number of simulations will fill the blank spaces in Figure 3.1b, but will not fill the gap areas in Figure 3.1a as the sample locations repeat themselves among the 875 pairs of random numbers. By determining such “gaps” in advance, the compaction can focus on areas where

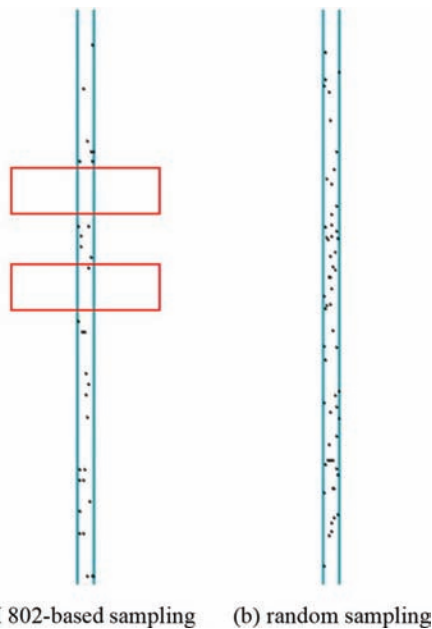


Figure 3.1 Comparison between ITM 802-based sampling and true random sampling.

samples might be taken and neglect the “gaps,” leading to low consistency in the compaction and inferior quality in the gap areas.

3.1.2 Random Sampling for Pavement

For pavement testing, either HMA or PCCP, the sampling locations are also determined by selecting number pairs in the random number table published in ITM 802, leading to similar “gaps” shown in Figure 3.1a. In comparison to earthwork where the compaction is directly performed on the in-situ soil, the pavement is usually constructed using materials delivered by trucks. The main concern is whether the gaps are large enough to be covered by one truck load. If so, the corresponding truck(s) will never be tested and theoretically, inferior materials can be used without being tested. Furthermore, with finite and predetermined random numbers listed in the ITM 802 published table, the effectiveness of such sampling protocol needs to be carefully evaluated by comparing it with the true random sampling using random number generated from uniform distribution in real time.

3.2 Evaluation of the Random Sampling Practice

Three simulation cases were conducted to evaluate the effectiveness of ITM 802-based sampling. Case 1 aims at determining if a gap area is larger than the size covered by one truckload. If so, that specific truck will never be tested, making it possible to manipulate the materials without being detected. Case 2 compares the ITM 802 sampling process to the true random sampling (using random numbers generated in real time) in the

aspect of the probability for each truck to be tested. Case 3 further examines the effectiveness of ITM 802-based sampling in comparison to the true random sampling in terms of the probability of detecting inferior materials.

All three simulation cases assume an HMA pavement subplot of 1,000 tons, pavement mixture of 165 lb/ yd² in density, and pavement width of 12 ft. The typical truckload, 20 tons, is assumed and the subplot involves 50 truckloads.

3.2.1 Case Study 1

Figure 3.2 illustrates the sample locations after 10,000 simulations using the ITM-802 table. The largest gap is identified to have a gap length of 81.81 ft. The gap area is calculated to be 982 square feet (by multiplying this gap length with the pavement width (12 ft)). This area corresponds to 9 tons in quantity ($12 \text{ ft} \times 81.81 \text{ ft} / 9 \text{ ft}^2 \text{ per yd}^2 \times 165 \text{ lb/yd}^2 = 18,000 \text{ lb} = 9 \text{ tons}$), smaller than one truckload (20 t). Therefore, it is infeasible to use a whole truckload of inferior material with zero chance of being tested.

3.2.2 Case Study 2

In Case Study 2, the probability for each truck to be tested using ITM 802-based sampling (Figure 3.3a) is compared to using the true random sampling (Figure 3.3b). Figure 3.3b illustrates that each truck has a 0.02 probability of being tested if the sampling process is truly random. Figure 3.3a, however, illustrates that in the current random sampling practice

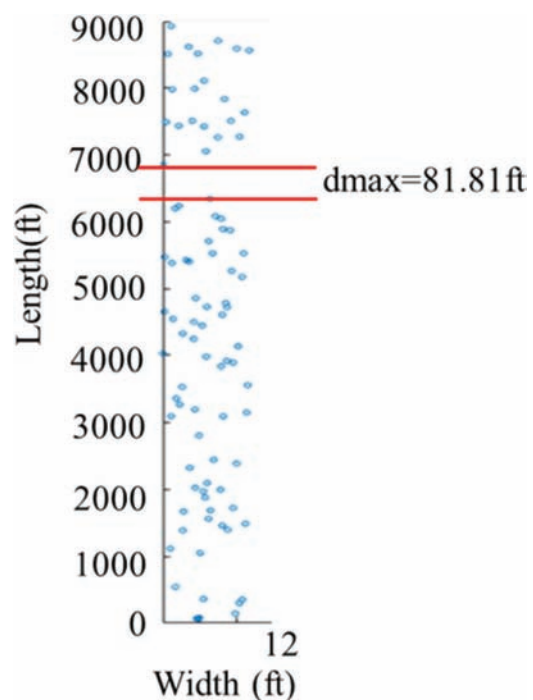


Figure 3.2 Largest gap area from ITM 802-based sampling.

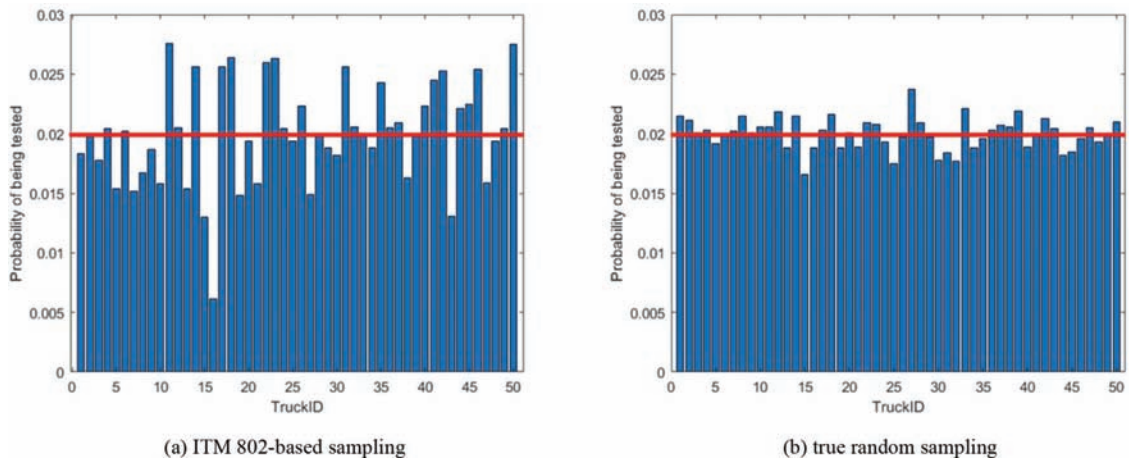


Figure 3.3 Probability of being tested for each truck.

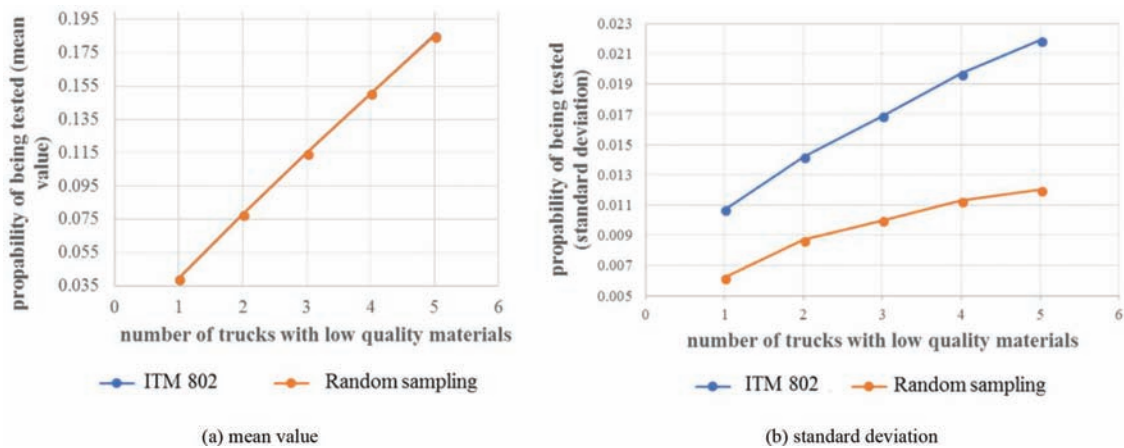


Figure 3.4 Probability of detecting at least one truck with inferior materials.

following ITM 802, the probability of being tested varies significantly among the trucks. Truck 15 and Truck 16 have distinguishable low probabilities of being selected. The results reveal that in ITM 802-based sampling, although every truck has a chance to be tested, the probability for some of them is low, and the result is not as level as the result from the true random sampling process. This discrepancy leads to low risk for contractors to manipulate certain trucks.

3.2.3 Case Study 3

In Case Study 3, the efficacy of the ITM-based sampling was compared with true random sampling in the aspect of detecting inferior materials. Assuming there are 1–5 random trucks with inferior materials, the probability of detecting at least one such truck was computed for both approaches. Figure 3.4 illustrates the results. While the average probabilities of two methods are almost the same, the ITM 802 approach leads to much larger standard deviations, indicating a lower reliability.

3.3 Findings, Recommendations, and Implementation

The main findings are as follows.

- ITM 802-based sampling leads to “gaps”—areas that will never be tested due to the limited random numbers in the published table.
- For earthwork, gap areas can receive insufficient compaction without being caught.
- For pavement, although gap areas are not large enough to accommodate a whole truckload, they lead to lower probabilities for certain trucks to be tested, compared with the true random sampling process. Such discrepancy results in a lower risk for contractors to manipulate these trucks, and a higher risk for INDOT to accept inferior materials.
- The ITM 802-based sampling is less reliable than the true random sampling with random number generated from uniform distribution in real time.

Therefore, it is recommended to use a real-time random number generator instead of a published random number table to perform the sampling for both earthwork and pavement. In order to implement this recommendation, a web-based tool has been developed to assist

field engineers in determining sampling locations in real time. Section 10.1 and 10.2 provide details.

4. SPATIAL PATTERN ANALYSIS

4.1 Introduction

Random sampling is applied in the acceptance testing with the expectation that the random samples are representative of the population. An inherent assumption is that samples are independent, which, however, is not realistic in pavement and earthwork. For soil, Tobler's First Law of Geography—everything is related to everything else, but near things are more related than distant things—applies. For pavement, materials are delivered by trucks and paved continuously. Consequently, pavement at neighboring locations is expected to have similar level of quality. This scenario of nearby locations showing similar level of quality, defined as spatial autocorrelation, leads to a key challenge in QA: if samples are close to each other, the effective number of samples is reduced even though the sample locations are randomly chosen. As a result, samples are biased, no longer truly representative of the population. This chapter aims at developing a method to minimize the impact of spatial autocorrelation by spatially spreading the sample locations.

4.2 Spatial Autocorrelation in Soil Compaction

As an illustration, Figure 4.1a shows the soil compaction meter value (CMV), an indicator of soil stiffness measured via intelligent compaction, for an INDOT project. Intuitively, the higher values (light cells) tend to cluster in the upper area and lower values (dark cells) fall in the lower area. Moran's I is computed to be 0.261 by taking into consideration of both feature locations and feature values, from a statistical perspective. Equation 4.1 illustrates the calculation of Moran's I . Assuming the entire area is divided into N cells, x_i represents the CMV for cell i , and \bar{x} is the average CMV for all cells. $W_{i,j}$ is the spatial weight between cells i and j , and typically, the closer two cells are, the larger

the weight is. Moran's I is the ratio of deviation values for all neighboring features (numerator) to the sum of individual deviations (denominator), normalized by the aggregation of all spatial weights (W) (ArcGIS, 2017). Considering that the reference range for Moran's I is $[-1, 1]$, positive values indicating spatial cluster, negative value indicating spatial dispersion, and near zero values indicating spatially random distribution. The Moran's I for the soil signals the CMV values are spatially clustered. Figure 4.1b illustrates the corresponding clusters.

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (\text{Equation 4.1})$$

4.3 Spatial Pattern Index

Two spatial indices, nearest neighbor index (NNI) and coefficient of variation (CV), are proposed in this study to evaluate the spatial pattern of samples based on their distances to each other. NNI measures the degree of spatial dispersion based on the minimum of inter-feature distances (Chou, 1997). CV measures the co-variation among the samples.

In this study, given a set of samples NNI calculation follows the following steps.

1. The nearest distance of a specific sample location, d_i , is determined as the Euclidean distance between this sample location and its nearest neighbor.
2. The average nearest distance of the set of samples, Ad , is calculated as the average of the nearest distances of all sample locations as $Ad = (\sum_{i=1}^n d_i) / n$, where d_i is the distance from sample i to its nearest neighbor, and n is the sample size.
3. Since a pavement segment or a soil compaction path is linear (i.e., the length is much larger than the width), we focus on spatial dispersion in the longitudinal dimension and simplify the statistically expected value of the average nearest distance, Ed , as $Ed = L / n$, where L is the lot/area length.
4. The nearest neighbor index (NNI) is calculated as $NNI = Ad / Ed$. By definition, the samples are spatially clustered

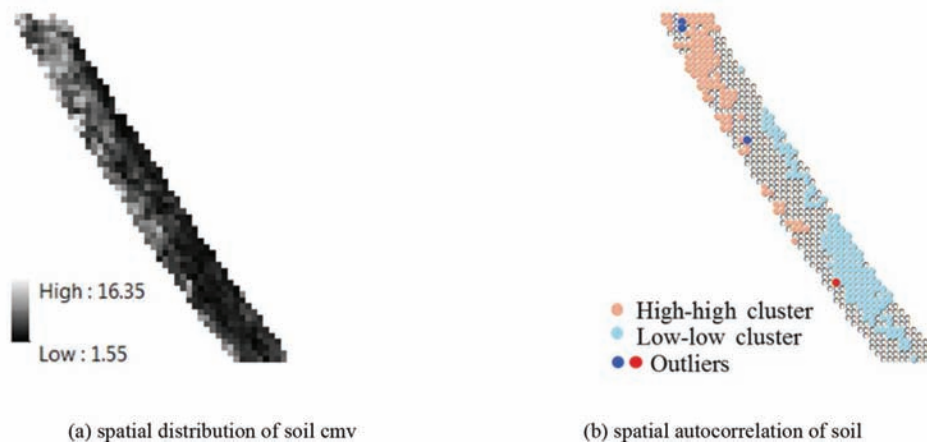


Figure 4.1 Spatial pattern analysis for soil CMV.

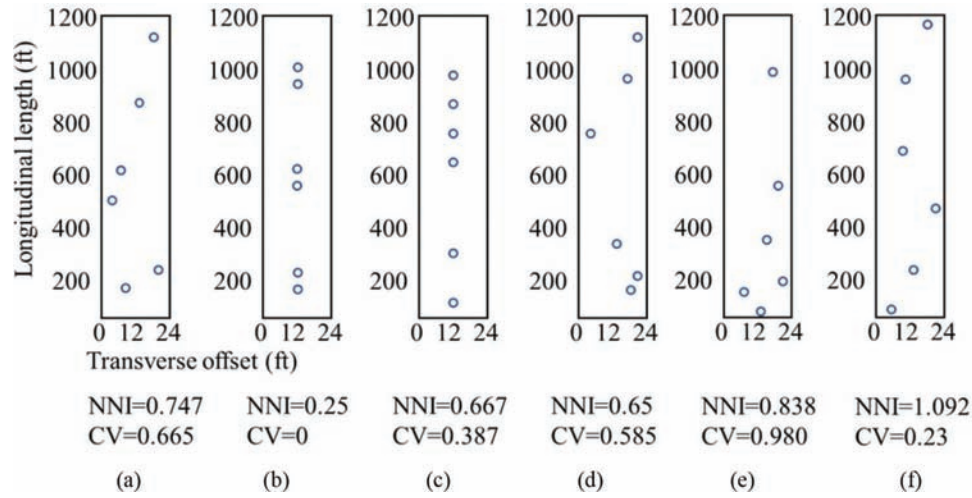


Figure 4.2 Illustration of scenarios with varying *NNI* and *CV*.

when *NNI* is less than one and spatially dispersed when *NNI* is larger than one.

In addition, it is desirable that the samples cover the entire lot to the largest extent, i.e., samples should “evenly” spread throughout the area. This is achieved by minimizing the variability among all nearest distances. In this study, *CV* is used to measure the variability. It is denoted as $CV = Std / Ad$, where *Std* is the standard deviation of nearest distances, computed as

$Std = \sqrt{\sum_{i=1}^n (d_i - Ad)^2 / n}$. Under the ideal scenario, where the *n* samples are exactly evenly distributed, *CV* is 0.

NNI and *CV* work together to assure the spatial distribution of sample locations. Figure 4.2 illustrates six scenarios with varying *NNI* and *CV*. Scenarios (a), (d), (e), and (f) are randomly generated and in scenarios (b) and (c) the sample locations are manually selected. Intuitively, scenario (f) has the best spatial pattern. The large *NNI* and small *CV* values confirm the sample locations spread out almost evenly. Scenario (b) shows that the points are clustered in three groups with a small *NNI* (i.e., 0.25), but a perfect *CV* (i.e., 0). Scenarios (a), (c), (d), and (e) all have similar large *NNI* values, but differ in *CV*. Correspondingly, the sample locations are generally disbursed with some local clustering. The general trends indicate that *NNI* and *CV* shall both be used to assure the spatial distribution of sample locations—*NNI* can serve as the primary index and *CV* can serve as the supplementary index.

In implementation, both *NNI* and *CV* are incorporated in the web tool (Section 10.1 and 10.2) developed in this study by constraining them in reference ranges to prevent the spatial clustering and high variability in sample locations. The reference ranges vary with lot or area size and sample size and are set based on 1,000 simulations. Specifically, 1,000 simulations of the required number of samples are performed, leading to 1,000 *NNI*s and 1,000 *CV*s. These results are ordered ascendingly. The range of the 95% largest *NNI*s and the range of

the 95% smallest *CV*s are the reference ranges to constraint both *NNI* and *CV*. The web tool will continue to generate sets of required sample locations till they fall in the reference ranges for both *NNI* and *CV*.

4.4 Findings and Recommendation

The key findings and recommendations are as follows.

- The material properties are spatially auto-correlated. Therefore, sample spatial pattern needs to be evaluated and controlled to avoid spatial clustering and ensure the effectiveness of random sampling.
- Both *NNI* and *CV* shall be used for evaluating the spatial pattern of sample locations. *NNI* assesses the degree of spatial dispersion and is recommended to serve as the primary metrics. *CV* measures the variability of sample locations and is recommended to serve as the secondary metrics.
- In the implementation through the web tool, both *NNI* and *CV* are constrained in the 95% reference ranges to ensure the sample locations are spatially distributed and truly representative of the population.

5. RISK ASSESSMENT FOR GIVEN TESTING STRATEGIES

5.1 Introduction

In the current pavement QA practice at INDOT, a predetermined number of samples are taken at random locations following ITM-802 and tested, and statistical measures such as mean, standard deviation, and percent within limits (PWL) are calculated and compared to acceptance criteria to determine whether to accept or reject the product. For PCCP pavement, sample mean is the statistical measure and for HMA pavement, sample PWL (calculated based on sample mean and standard deviation) is the statistical measure.

Unfortunately, sample statistical measures are just estimates of the correspondingly population statistical

measures and they vary from one set of samples to another set of samples. Discrepancies between sample and population measures are inevitable. Thus, it is critical to incorporate such discrepancies and sampling variability into the decision-making process to determine the risk in accepting inferior product given a set of samples. This is achieved by treating the sample statistical measures as statistical variables (e.g., the population PWL is not the same as the sample PWL; rather, its boundaries at certain confidence level are estimated based on a critical value). By comparing the sample PWL with the critical value while making the acceptance decision, the risk can be controlled within an acceptable level. Two statistical methods: M-method and S_{pk} -method, were exploited to derive the critical value and make corresponding decisions. Both methods were compared to the current PWL and mean methods at INDOT to reveal the difference in risk. Note in this section, risk is denoted by testing power—the probability of making correct acceptance decision: accept the product that should be accepted and reject the product that should be rejected. Calculation-wise, risk equals to 1 – testing power.

For soil compaction acceptance, the current QA practice is different. DCP value of individual sample, instead of the quality measures computed from a group of samples, is used to make acceptance decision. When a deficiency is identified at a random location, the contractor is required to investigate and correct this deficiency (INDOT, 2014). In such a scenario, the testing power is defined as the probability of catching at least one random sample for defective soil. Risk is still calculated as 1 – testing power. This allows INDOT to determine their risk—failure in detecting an inferior area—for a given testing strategy.

5.2 Risk Assessment Methods

Since PWL measures the proportion of the pavement that falls within the specification limits considering both material average property and the variability, it is chosen as the statistical measure in this study and statistical hypothesis testing procedures that incorporate sampling variability are developed correspondingly to make acceptance decisions and determine the risk. We propose two estimates for PWL, one based on fraction defective p and the other based on yield index S_{pk} , to perform statistical hypothesis testing. These measurements are derived under the normality assumption, following the practice of most state DOTs.

5.2.1 PWL Estimate Based on Fraction Defective

The first approach is based on fraction defective. It is a variant of the classic M -method (Lieberman & Resnikoff, 1955), a commonly adopted method in statistical quality control. The fraction defective refers to the fraction that is outside of acceptable limit and it is calculated as $p=1-PWL$. When estimated from a group of samples, it is denoted as \hat{p} ; \hat{p} is compared with

a critical value M that is derived for a given quality target (set as portion within limits) and confidence level. If \hat{p} is smaller than the critical value M , the pavement will be accepted. Otherwise, the pavement will be rejected. For instance, if the confidence level is 90% and the rejectable fraction defective p is 20% (the agency will reject the pavement if p is larger than 20%), a corresponding critical value M can be derived. Assuming \hat{p} is obtained from the pavement with rejectable quality and is larger than M , we will reject the pavement with a 90% confidence—the risk of incorrectly accepting the inferior product is 10%.

5.2.1.1 Mathematical Model. The statistical distribution of the fraction defective is derived as follows. Given a population that follows a normal distribution $N(\mu, \sigma^2)$, where μ is population mean and σ is the population standard deviation, the fraction defective p refers to the total proportion of the pavement that is outside the lower specification limit (LSL) and upper specification limit (USL) (see Equation 5.1).

$$p = P(X < LSL | \mu, \sigma) + P(X > USL | \mu, \sigma) \quad (\text{Equation 5.1})$$

By definition, $p = 1 - PWL$. Population p is unknown and can only be estimated (as \hat{p}) from samples of the population. \hat{p} is a variable that follows specific statistical distribution, illustrated in Equations 5.2–5.4, where x_1, \dots, x_n are random samples from the population with unknown mean and standard deviation, sample mean is calculated as $\bar{x} = \sum_{i=1}^n x_i / n$, sample standard deviation is calculated as $s = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}$, and $B(x; n/2-1, n/2-1)$ is the Beta probability density function (PDF) with parameters $n/2-1$ and $n/2-1$ for \hat{p}_U and \hat{p}_L . The detailed derivation of \hat{p}_U and \hat{p}_L can be found in Lieberman and Resnikoff (1955).

$$\hat{p} = \hat{p}_L + \hat{p}_U \quad (\text{Equation 5.2})$$

$$\hat{p}_U = \int_0^{\max(0, \frac{1}{2} - \frac{1}{2} \frac{USL - \bar{x}}{s} \frac{\sqrt{n}}{n-1})} B(x; \frac{n}{2}-1, \frac{n}{2}-1) dx \quad (\text{Equation 5.3})$$

$$\hat{p}_L = \int_0^{\max(0, \frac{1}{2} - \frac{1}{2} \frac{\bar{x} - LSL}{s} \frac{\sqrt{n}}{n-1})} B(x; \frac{n}{2}-1, \frac{n}{2}-1) dx \quad (\text{Equation 5.4})$$

Knowing the statistical distribution of \hat{p} is important because it is the base for determining the critical value and performing hypothesis testing, and making accept/reject decisions at a controlled risk level.

5.2.1.2 Critical Value M and Hypothesis Testing. For pavement acceptance testing, there are two hypotheses, null hypothesis $H_0 : p \leq c$ and alternative hypothesis $H_1 : p > c$, where c is the target value of fraction

defective and is regulated in the specification. For instance, if the rejectable quality level is 20% in the specification, or c equals 0.2, the two hypotheses become $H_0 : p \leq 0.2$ and $H_1 : p > 0.2$.

The fraction defective estimate \hat{p} , calculated from a set of random samples, is compared to a critical value (its derivation is provided in the rest of this section) and if the fraction defective estimate is larger, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted. Otherwise, we fail to reject the null hypothesis H_0 . Failure in rejecting the null hypothesis H_0 means there is not enough evidence to reject the null hypothesis with the current samples; however, when more samples are collected, we may be able to reject the null hypothesis. Contrarily, rejecting hypothesis H_0 means the fraction defective is statistically larger than the target and the pavement is rejected.

Since the sample fraction defective \hat{p} is a random variable to estimate the population fraction defective, a critical value M is required for hypothesis testing. If $\hat{p} \leq M$, we accept the null hypothesis at the specific confidence level; that is, we accept the pavement with the confidence of $1-\alpha$. The statistical meaning is that the probability of the population fraction defective being greater than the target value c is less than α . This critical value M is determined by solving Equation 5.5, where (μ_0, σ_0) is the population mean and standard deviation.

$$P(\hat{p} < M | \mu_0, \sigma_0) = P(\hat{p}_U + \hat{p}_L < M | \mu_0, \sigma_0) = 1 - \alpha \quad (\text{Equation 5.5})$$

It is extremely difficult to derive the critical M value using an analytical approach, which is recognized as a drawback of the M -method. In this study, we developed a two-step process as a practical approach to determining the critical value M , referred as numerical M -method. Step 1 aims at converting the derivation of the estimated fraction defective, \hat{p} , into a statistically computable format by substituting the random components with computable components to facilitate Step 2 simulations. Step 2 aims at simulating the \hat{p} and determining the critical value using a Monte-Carlo approach, which can consider all populations that satisfy the target fraction defective instead of estimating from samples and leads to more reliable results.

In Step 1, the inputs to the conversion process are Equation 5.3 and Equation 5.4 that illustrate the derivation for \hat{p}_U and \hat{p}_L . The random components of $\sqrt{n}(USL - \bar{X})/S$ and $\sqrt{n}(\bar{X} - LSL)/S$ in the equations make the computation impossible. These two random variables follow non-central t distributions with degrees of freedom $n-1$ and the non-centrality parameter $ncp_1 = \sqrt{n}(USL - \mu_0)/\sigma_0$ and $ncp_2 = \sqrt{n}(\mu_0 - LSL)/\sigma_0$, respectively. Due to the distributional equivalence of the non-central t distribution, i.e., $\frac{\sqrt{n}(USL - \bar{X})^d}{S}$

$$= \frac{ncp_1 - Z}{\sqrt{\chi/(n-1)}} \text{ and } \frac{\sqrt{n}(\bar{X} - LSL)^d}{S} = \frac{ncp_2 + Z}{\sqrt{\chi/(n-1)}}, \text{ where } Z$$

is a standard normal random variable and χ is an independent chi-square random variable with $n-1$ degrees of

freedom, they substitute the random components in Equation 5.3 and Equation 5.4, leading to Equation 5.6 and Equation 5.7. These two new equations are the outputs of Step 1. They mathematically describe the computation of \hat{p}_U and \hat{p}_L .

$$\hat{p}_U = \int_0^{\max(0, \frac{1}{2} - \frac{1}{2} \frac{ncp_1 - Z}{\sqrt{\chi/(n-1)}} \frac{1}{n-1})} B(x; \frac{n}{2} - 1, \frac{n}{2} - 1) dx \quad (\text{Equation 5.6})$$

$$\hat{p}_L = \int_0^{\max(0, \frac{1}{2} - \frac{1}{2} \frac{ncp_2 + Z}{\sqrt{\chi/(n-1)}} \frac{1}{n-1})} B(x; \frac{n}{2} - 1, \frac{n}{2} - 1) dx \quad (\text{Equation 5.7})$$

In Step 2, based on Equation 5.6 and Equation 5.7, a Monte-Carlo simulation approach is developed to determine the critical value M . The process is detailed as follows.

1. Determine 20 pairs of $(\mu_{0,j}, \sigma_{0,j})$ that corresponds to a given target defective fraction c , where $j=1, \dots, 20$. Note that an infinite number of pairs of (μ_0, σ_0) can be found to correspond to the same c as far as they satisfy $P(X < LSL | \mu_0, \sigma_0) + P(X > USL | \mu_0, \sigma_0) = c$, and this study takes 20 pairs for the computation purpose. For example, when $LSL=2.6$, $USL=5.4$, and $c=0.1$, both $\mu_0=3.37$, $\sigma_0=0.6$ and $\mu_0=4.29$, $\sigma_0=0.8$ satisfy the requirement $P(X < 2.6 | \mu_0, \sigma_0) + P(X > 5.4 | \mu_0, \sigma_0) = 0.1$. In this study, these 20 pairs of (μ_0, σ_0) are determined through numerical computation.
2. For each pair of $(\mu_{0,j}, \sigma_{0,j})$,
For $i=1, \dots, B=10000$,
 - a. Select a random variable $Z_{i,j}$ from standard normal distribution, and a random variable $\chi_{i,j}$ from chi-square distribution with degree of freedom $n-1$. These two random variables are used to compute the statistical parameters in step b.
 - b. Compute $\frac{ncp_{1,j} - Z_{i,j}}{\sqrt{\chi_{i,j}/(n-1)}}$ and $\frac{ncp_{2,j} + Z_{i,j}}{\sqrt{\chi_{i,j}/(n-1)}}$, where
 $ncp_{1,j} = \sqrt{n}(USL - \mu_{0,j})/\sigma_{0,j}$ and $ncp_{2,j} = \sqrt{n}(\mu_{0,j} - LSL)/\sigma_{0,j}$.
 - c. Compute $\hat{p}_{i,j} = \hat{p}_{L,i,j} + \hat{p}_{U,i,j}$.
3. For the population with $(\mu_{0,j}, \sigma_{0,j})$, 10,000 estimated fraction defective $\hat{p}_{i,j}$ result from Step 2, and the critical value for this population, M_j , is determined as the $\hat{p}_{i,j}$ that is larger than $(1-\alpha) \times 100\%$ of all 10,000 $\hat{p}_{i,j}$. It is equivalent to finding a value such that the probability of $\hat{p}_{i,j}$ exceeding M_j equals α , as defined in Equation 5.5.
4. Since the true population mean and standard deviation are unknown, these 20 pairs of $(\mu_{0,j}, \sigma_{0,j})$ are used as an approximation, leading to 20 different critical values, M_j . For different pairs of $(\mu_{0,j}, \sigma_{0,j})$, the tails of the histogram of $\hat{p}_{i,j}$ are different, either symmetric or asymmetric. In the risk calculation, it is desirable that the probability of \hat{p} exceeding M is no larger than α when considering all possible situations. Therefore, the largest value among M_j is chosen as the critical value (i.e., $M = \max\{M_1, \dots, M_{20}\}$) to cover the most asymmetric cases.

5.2.2 PWL Estimate Based on Yield Index

The second approach is based on yield index S_{pk} . It was devised by Boyles (1994) to assess the compatibility of the process performance with the specification, and has been successfully adopted to the QA/QC practice in the manufacturing industry. The yield index S_{pk} has one-to-one correspondence to PWL, i.e., $PWL = 2\Phi(3S_{pk}) - 1$, or $S_{pk} = \frac{1}{3}\Phi^{-1}\left(\frac{PWL+1}{2}\right)$. The estimate of S_{pk} - from a group of samples, follows an asymptotically normal distribution. This normal distribution affords the simplification of the hypothesis testing as a standard one-tail Z-test.

5.2.2.1 Mathematical Model. Equation 5.8 shows the mathematical definition of the yield index S_{pk} , where μ and σ are the population mean and standard deviation, and Φ is the cumulative distribution function (CDF) of the standard normal distribution.

$$S_{pk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL-\mu}{\sigma}\right) + \frac{1}{2}\Phi\left(\frac{\mu-LSL}{\sigma}\right)\right\} \quad (\text{Equation 5.8})$$

Since μ and σ of the pavement population are unknown, the S_{pk} needs to be estimated from n samples by substituting the μ and σ in Equation 5.8 with sample mean \bar{x} and sample standard deviation s , denoted as \hat{S}_{pk} . Lee, Hung, Pearn, and Kueng (2002) and Pearn and Cheng (2007) proved that \hat{S}_{pk} is asymptotically normally distributed with mean S_{pk} and variance. $(a^2 + b^2) / \{36n[\phi(3S_{pk})]^2\}$ Equations 5.9–5.11 illustrate its probability density function.

$$f_{\hat{S}_{pk}}(x) = \sqrt{\frac{18n}{\pi}} \frac{\phi(3S_{pk})}{\sqrt{a^2 + b^2}} \exp\left[-\frac{18n[\phi(3S_{pk})]^2}{a^2 + b^2}(x - S_{pk})^2\right] \quad (\text{Equation 5.9})$$

$$a = \frac{1}{\sqrt{2}}\left\{\frac{USL-\mu}{\sigma}\phi\left(\frac{USL-\mu}{\sigma}\right) + \frac{\mu-LSL}{\sigma}\phi\left(\frac{\mu-LSL}{\sigma}\right)\right\} \quad (\text{Equation 5.10})$$

$$b = \phi\left(\frac{USL-\mu}{\sigma}\right) + \phi\left(\frac{\mu-LSL}{\sigma}\right) \quad (\text{Equation 5.11})$$

5.2.2.2 Hypothesis Testing. For pavement acceptance, there are two hypotheses using S_{pk} (Lee et al., 2002; Pearn & Cheng, 2007), null hypothesis $H_0 : S_{pk} \geq c$ and alternative hypothesis $H_1 : S_{pk} < c$, where c is the target value that is derived from the PWL in specification. For instance, if the specification regulates the PWL for acceptance is 0.9 or larger, the target value is calculated as $c = \frac{1}{3}\Phi^{-1}\left(\frac{PWL+1}{2}\right) = 0.55$. The two hypotheses become $H_0 : S_{pk} \geq 0.55$ and $H_1 : S_{pk} < 0.55$. The testing

statistic T calculated using the yield index estimate \hat{S}_{pk} , is compared to a critical value (the derivation of testing statistic and critical value is provided in the rest of this section) and if the testing statistic is smaller, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted. Otherwise, we fail to reject the null hypothesis H_0 . Failure in rejecting the null hypothesis H_0 means there is not enough evidence to reject the null hypothesis with the current samples; however, when more samples are collected, we may be able to reject the null hypothesis. Contrarily, rejecting hypothesis H_0 means the yield index is statistically smaller than the target and the pavement is rejected.

Due to the asymptotically normal distribution of \hat{S}_{pk} , the hypothesis testing is simplified to be a standard Z-test, where the testing statistic is derived as

$$T = \frac{6(\hat{S}_{pk} - c)\sqrt{n}\phi(3\hat{S}_{pk})}{\sqrt{\hat{a}^2 + \hat{b}^2}}, \text{ where } \hat{a} \text{ and } \hat{b} \text{ are computed}$$

from Equation 5.10 and Equation 5.11 by replacing μ and σ with \bar{x} and s , respectively. The critical value (Z) is computed from standard normal distribution. Figure 5.1 illustrates that given a specific confidence level $1-\alpha$, Z_α is the lower α quantile of the standard normal distribution. If $T \geq Z_\alpha$, we fail to reject the null hypothesis; that is, we accept the pavement with the confidence of $1-\alpha$. The statistical meaning is that the probability of the population yield index being smaller than the target value is less than α .

5.2.3 Case Study

Air voids acceptance in HMA pavement is used as an example to demonstrate the advantage of the proposed methods in risk assessment by comparing them with the current practice at INDOT. According to INDOT specification for HMA (INDOT, 2016), the USL and LSL for air voids (%) are 5.4 and 2.6 respectively, five samples are taken per lot, and pavement with PWL lower than 0.5 is regarded as failed material—the target PWL is 0.5. It is assumed that the desirable risk level of INDOT is 0.1 ($\alpha=0.1$). To evaluate the testing power of different population, the population mean is set to vary between 2 and 6, and the standard deviation is 0.6 or 1.2. To reveal the influence from sample size on the risk, the sample size is set to be 5, 10, and 100.

Figure 5.2 illustrates the comparisons of the testing power—the probability of rejecting failed materials among INDOT's current practice (black line), and numerical M - (red line) and S_{pk} -methods (blue dash line), using numerical simulations (10,000 times). The horizontal axis is the population mean (μ) varying from 2 to 6. All diagrams are symmetrical, centered at $\mu=4$ which is the average of USL and LSL and corresponds to the largest PWL. The farther away μ is from 4, the smaller the population PWL is. The two vertical dash lines represent the boundary of μ . The left and right range outside these two lines represent failed materials that should be rejected, referred as rejectable regions. The vertical axis is the testing power, or the probability

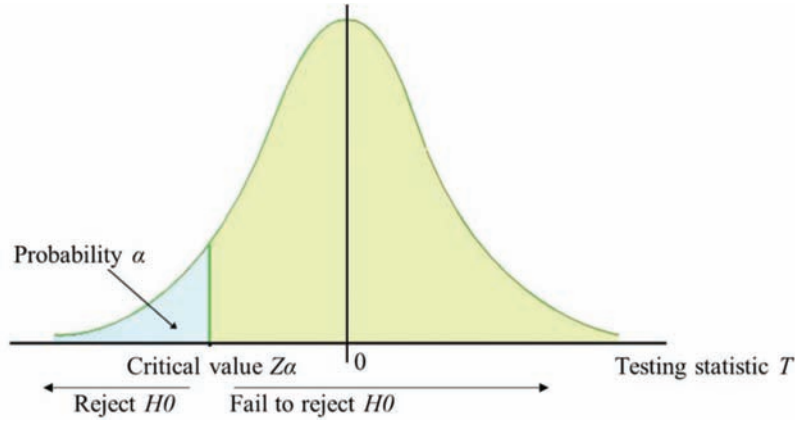


Figure 5.1 Hypothesis testing based on yield index.

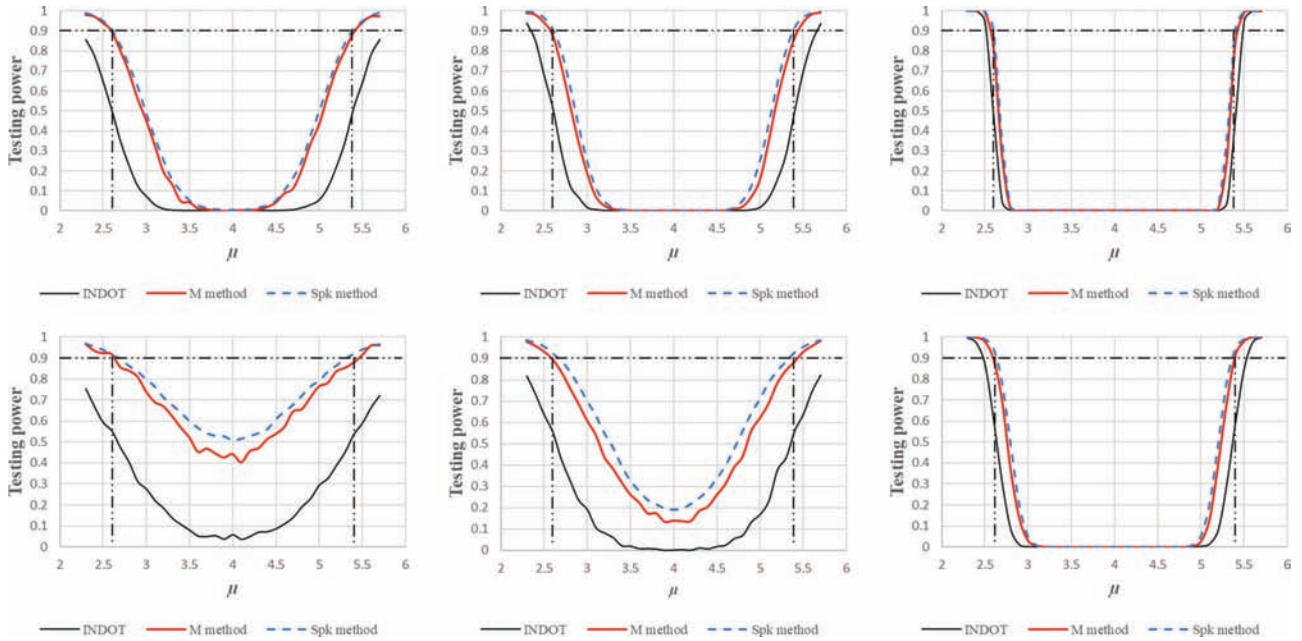


Figure 5.2 Testing power of HMA air voids acceptance testing.

of rejecting the materials. The horizontal dash line represents the testing power at 0.9. The desired outcome is to control the risk of INDOT in accepting rejectable materials to be 0.1 or less, which is equivalent to achieving a larger-than-0.9 testing power in the rejectable regions. In all charts, red lines represent the results of numerical M -method, blue dash lines represent the results of S_{pk} -method, and black lines represent the results of the current PWL method at INDOT.

In every chart, the performance curve of the numerical M -method intersects with the boundary lines and the 0.9 testing power line at the same two points. The numerical M -method succeeded in rejecting failed materials with a testing power is 0.9 in all scenarios. INDOT's risk of accepting failed materials is controlled to be below 0.1. The performance curve of the current PWL method intersects the two vertical boundary

lines around 0.6 testing power. The risk of INDOT in accepting failed materials is 0.4 or larger and is larger when the sample size is smaller. The S_{pk} -method, in general, aligns with the numerical M -method. However, the risk controlled by S_{pk} -method is slightly biased from the target of 10% level when the sample size is small ($n=5, 10$) and standard deviation is large ($\sigma=1.2$). The discrepancy stems from two reasons. First, the theoretical foundation of hypothesis testing is the asymptotically normal distribution of \hat{S}_{pk} , which is slightly deviated from its true distribution when the sample size is smaller than 40 (Pearn, Lin, & Wang, 2004). Second, the critical value is approximated using samples from a specific population instead of considering different populations as the numerical M -method does, which may lower the robustness when the standard deviation is large.

Based on the results, we recommend to use numerical M -method in all scenarios and use S_{pk} -method when the sample size is relatively large to simplify the computation. Overall, the numerical M -method is recommended to replace the current PWL practice.

5.3 Risk Assessment for Soil Acceptance Testing

The decision process in accepting soil compaction is different from pavement. In soil acceptance testing, DCP value of individual samples, instead of the statistical measures computed from a group of samples, is used to make acceptance decisions. When a deficiency is identified at one random location, the contractor needs to investigate and correct this deficiency (INDOT, 2014). Therefore, the testing power for soil QA testing is defined as the probability of detecting at least one random samples from defective soil.

In this study, a simulation approach has been developed to assess the risk to INDOT in not catching failed materials following the current practice. Figure 5.3 illustrates the overall process. Given a specific defective rate ($dRate$), the percentage of defective soil, we divide the entire area into 1,000 cells and randomly designate $1000 * dRate$ cells as defective ones. When taking n random samples from these 1,000 cells for acceptance testing, if at least one of them is in a defective cell, we succeed in identifying the defective soil. This process is repeated 1,000 times; the testing power is calculated as (the number of successes / 1,000), the probability of successfully identifying at least one defective cell.

Figure 5.4 illustrates the testing power-defective rate curves for sample sizes of 3 to 7. The horizontal

axis represents the defective rate of soil, varying from 0% to 100%. The vertical axis represents the testing power, the probability of detecting at least one samples from the defective material. Two main observations are as follows. First, for all sample sizes, the testing power increases sharply as the defective rate increases at the low end, but levels out once the defective rate reaches 50%. Second, the testing power increases as the sample size increases. These curves allow the determination of the risk in accepting defective soil given a specific sample size and an estimated soil defective rate. For instance, if the sample size is three and the defective rate of the target area is expected to be around 25% according to the historical performance of the contractor, the testing power is 60% and the risk is 40%. The probability of defective soil escaping acceptance testing is 40%. These curves also allow the determination of the appropriate sample size if the risk is to be controlled at a specific level. For instance, for the above example, if the risk is to be controlled at 15%, the appropriate sample size shall be 7. Thus, the chart in Figure 5.4 allows INDOT to determine its risk at a given sample size and the appropriate sample size at a given risk. This simulation method has been validated using real data, and details are in Section 9.3.

5.4 Findings and Recommendations

The main findings and recommendations are as follows.

- The current acceptance testing methods at INDOT do not achieve the desired level of risk control.

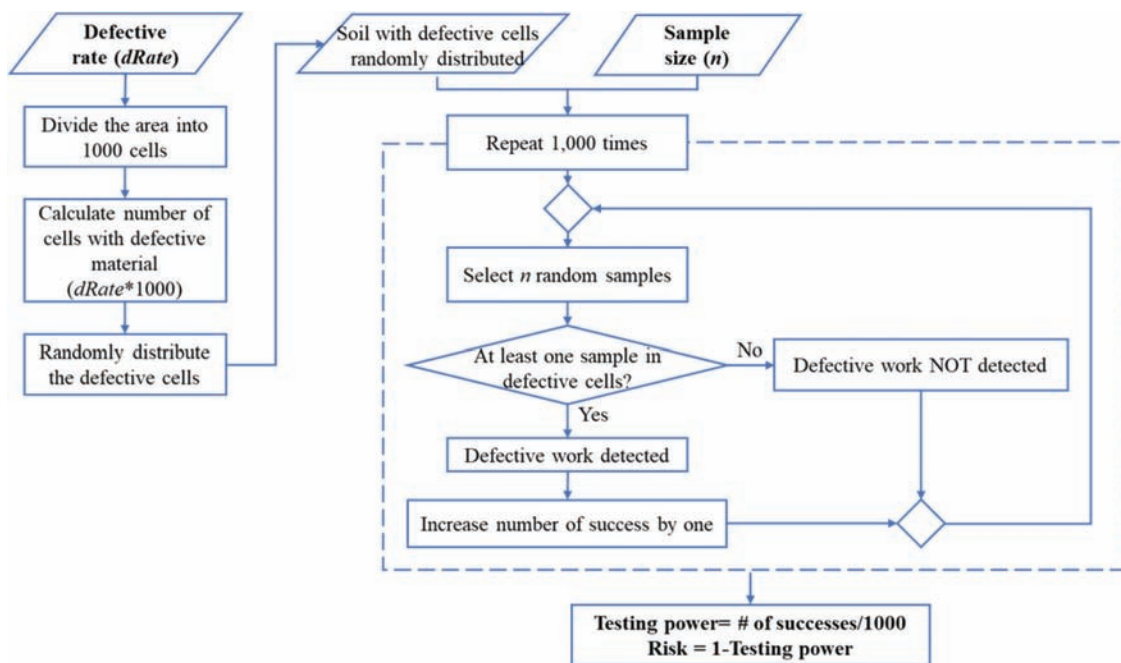


Figure 5.3 Workflow of testing power calculation and risk assessment for soil acceptance.

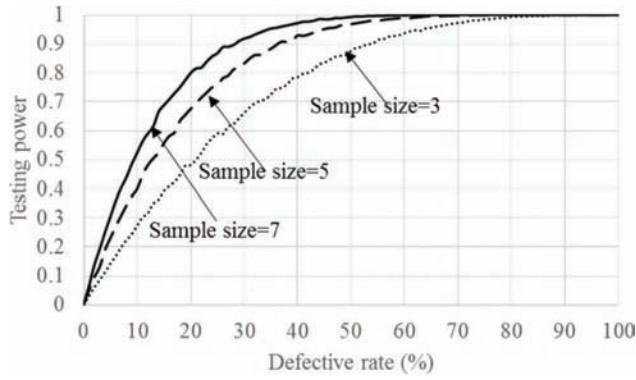


Figure 5.4 Testing power-defective rate curves in soil acceptance testing.

- The newly proposed numerical M -method and S_{pk} -method incorporate sample variance and treat statistical measures of samples as random variables. They are capable of controlling the risk at a desired level and achieving higher testing power compared to the current practice at INDOT.
- In general, the numerical M -method and S_{pk} -method aligns with each other. The numerical M -method is suitable for all scenarios and is recommended to replace the current practice, while S_{pk} -method can be applied when the sample size is relatively large to simplify the computation.
- For the soil compaction acceptance, a simulation approach is proposed to assess the risk of INDOT. It is recommended to conduct the simulation to examine the risk once a sampling strategy is determined, and then make corresponding adjustment on the sampling strategy according to the desired risk.

The proposed methods for pavement and soil acceptance have been implemented as two tabs in the newly developed web tool, detailed in Section 10.3 and 10.5. For pavement testing, INDOT engineers can estimate their risk of accepting failed materials ($1 - \text{testing power}$) using the proposed methods and compare the new methods to INDOT's current practice (PWL for HMA and mean value for PCCP). The users can also compare the testing power of different sample sizes by changing the sample size in the input. For soil acceptance testing, INDOT engineers can compare their risk in accepting defective soil work at different sample sizes and determine the appropriate sample size to control the risk at any desired level.

6. OPTIMAL SAMPLE SIZE

6.1 Introduction

A critical question in QA practice is the optimal sample size for controlling the risk at a specific level. In the previous section statistical methods were developed to assess the risk given a sample size. The connection between risk and sample size for both pavement and soil has been established. This connection forms the base to determine the optimal sample size to control risk at a specific level, which is the aim of this chapter.

The risk for INDOT is defined as the probability of accepting a product that is of rejectable quality level (RQL); the risk for contractor is defined as the probability of a product that is of acceptable quality level (AQL) being rejected. Since the acceptance testing for pavement and soil is different, the determination of the optimal sample size for pavement and soil is treated separately in this chapter.

6.2 Optimal Sample Size for Pavement

In pavement acceptance testing, state DOTs designate AQL and RQL. AQL represents the low boundary of the quality level the agency will accept the product, and RQL is the high boundary of the quality level that the agency will reject the product. In this study, AQL and RQL are described in terms of fraction defective p (i.e., $1 - \text{PWL}$). For instance, if AQL is 10%, the agency will accept the pavement with fraction defective no greater than 10% (or the PWL no less than 90%). If RQL is 40%, the agency will reject the pavement with fraction defective equal to or larger than 40% (or PWL no greater than 60%). Due to the uncertainty in sampling, there is no guarantee that samples taken from pavements with acceptable quality level will lead to the decision to accept the product. The probability of pavement with AQL being rejected is referred to as contractor's risk, denoted by α . The probability of pavement with rejectable quality level being accepted is referred to the agency's risk, denoted by β (Chang & Hsie, 1995). According to the statistical quality control theory, an acceptance sampling plan is designed to meet the risk requirements of both parties: (1) the probability of the pavement with AQL being accepted is $1 - \alpha$, and (2) the probability of pavement with RQL being accepted is β . Equation 6.1 and Equation 6.2 illustrate this concept.

$$P(\text{acceptance} \mid p = AQL) = 1 - \alpha \quad (\text{Equation 6.1})$$

$$P(\text{acceptance} \mid p = RQL) = \beta \quad (\text{Equation 6.2})$$

In this study, both probabilities in Equation 6.1 and 6.2 are computed using the numerical M -method or S_{pk} -method proposed in Chapter 5. By solving these two equations simultaneously, the sample size n and corresponding critical value are determined (Montgomery, 2009). Unfortunately, there is no simple close-form solution. In this study, we propose a numerical approach to derive the optimal sample size to control the agency's risk while assuming a fixed risk level for the contractor. Figure 6.1 illustrates the process. For a specific AQL and contractor's risk α , the sample size n is initialized to be 1 and the corresponding critical value M is computed using either the numerical M -method or S_{pk} -method. To evaluate the agency's risk (β), the probability of accepting RQL pavement, we run 1,000 simulations, i.e., repeating the following process 1,000 times: taking n samples from an RQL pavement, calculating the sample statistic for 1,000 times, and compare the statistical measure of each set of the n samples to the critical

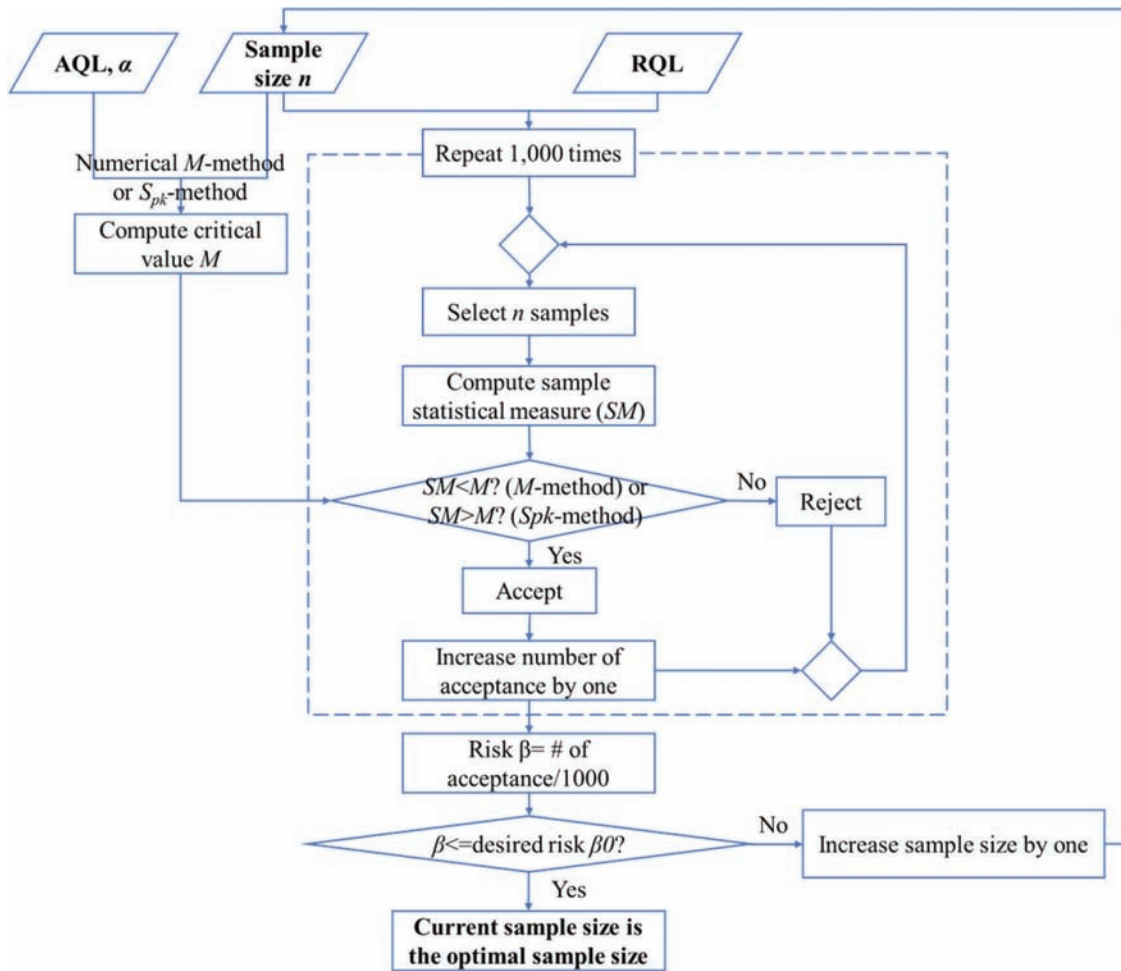


Figure 6.1 Process for determining the optimal sample size in pavement acceptance testing.

value to make accept or reject decision. The agency's risk (β) is calculated as the number of accept decisions divided by 1,000. If this risk is larger than the desired level of risk (β_0), the sample size n is increased by 1 and the whole process is repeated till the calculated risk is no larger than the desired level of risk (β_0). The sample size at the stopping point is the optimal (the smallest) sample size to control the agency's risk at the desired level (β_0). This process works because of the general trend: larger sample size, smaller risk and the starting sample size being very small.

To illustrate the process numerically, HMA air voids acceptance testing is used as an example to determine the optimal sample size that corresponds to the agency's risk. According to INDOT's specification, LSL and USL for air voids are 2.6% and 5.4%, respectively, and there are five sublots per lot. AQL and RQL are set to be 10% and 50% in terms of fraction defective, following the common practice. The contractor's risk is set to be 5% and the standard deviation of the pavement is assumed to be 0.8 (this value can be obtained from historical records). Figure 6.2 displays the INDOT's risk with respect to sample size. The horizontal axis represents the number of

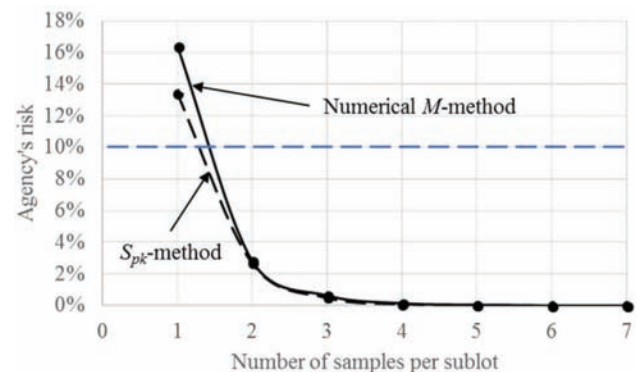


Figure 6.2 Agency's risk with varying sample size.

samples taken at each subplot, and the total sample size is computed as the product of subplot number and the sample size per subplot. The vertical axis represents the agency's risk in accepting RQL materials. Furthermore, the horizontal dash line indicates the 10% risk, which is common in the current practice. Unsurprisingly, agency's risk decreases with the increase of sample size, and the curves obtained from numeric- M method (black line) and S_{pk} -method

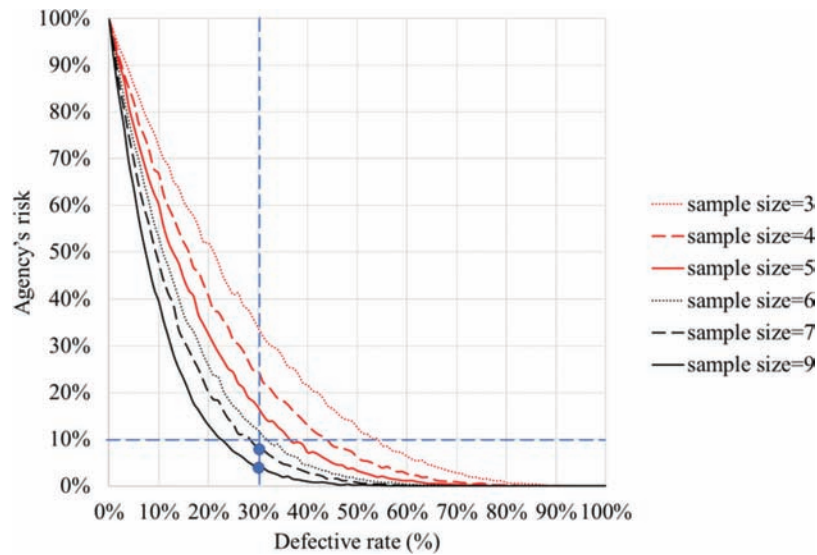


Figure 6.3 Determining the optimal sample size for soil based on the agency's risk curves.

(green line) are almost the same. To ensure the risk does not exceed 10% (shown as dash horizontal line), the optimal sample size is 2 samples/sublot * 5 sublots/lot = 10 samples/lot.

6.3 Optimal Sample Size for Soil

In Section 5.3, a set of testing power—defective rate curves for varying sample sizes have been established. Since the risk to INDOT in soil acceptance testing is defined as the probability of accepting defective soil, it is calculated as $1 - \text{testing power}$. Consequently, these testing power curves are converted into risk curves to illustrate the agency's risk with respect to defective rate. Figure 6.3 illustrates the risk curves for sample size of 3, 4, 5, 6, 7, and 9. To determine the optimal sample size, the user will start with a vertical line corresponding to the estimated defective rate and a horizontal line corresponding to the level of risk. The vertical line intersects the risk curves. If an intersection is below the horizontal risk line, the corresponding sample size ensures the risk is smaller than the desired level and thus, is controlled as desired. The smallest sample size among this group is the optimal sample size. For instance, given an area with an estimated defective rate of 30% (indicated as the vertical dash line), if the desired risk level is 10% (represented as the horizontal dash line), sample size 7 and 9 both satisfy this requirement as the intersections of their corresponding curves and the vertical line are below the horizontal line. 7 is the optimal sample size to control the risk to be below 10%.

6.4 Implementation

Two separate tabs in the web tool have been designated to determine the optimal sample size for pavement and soil acceptance testing, respectively, following

the methods described. Details are in Sections 10.5 and 10.6.

7. COST ANALYSIS FOR DETERMINING OPTIMAL SAMPLE SIZE

In Chapter 4 and 5, we have proven that more samples in acceptance testing leads to less risk to the agency, and developed methods to determine the optimal sample size that satisfies the risk expectation of the agency and contractor. However, a rigorous testing strategy with more samples increases the cost for QA practice, which may be difficult to implement due to resource constraints of the agency. Therefore, there is a critical need to take both cost and risk into consideration when determining the optimal sample size in order to maximize the benefit of acceptance testing to the agency.

A recent NCHRP report (National Academies of Sciences, Engineering, and Medicine, 2017) pointed out that the cost of quality consists of two compartments: cost of conformance and cost of non-conformance. In pavement acceptance testing, the cost of conformance refers to the detection cost including all costs associated with determining the level of conformance of the product, such as sampling and testing cost. The cost of non-conformance refers to the impact caused by defective or failed materials, such as repair or rework cost, and the reduced life of the pavement. Agency's risk in incorrectly accepting defective materials has a direct connection with the cost of non-conformance. The lower the agency's risk, the lower probability of accepting defective materials and less expected cost of non-conformance attributed to repair or rework of defective products. Consequently, when more samples are taken, the conformance cost is higher and the non-conformance cost is lower. The key challenge is to

determine the optimal sample size to minimize the total cost in acceptance testing.

In this chapter we aim at deriving the optimal sample size that corresponds to the lowest total cost from an economic perspective. Inspired by Gharaibeh, Garber, and Liu (2010) and Cho, Najafi, and Kopac (2011), two cost analysis approaches are developed to evaluate the impact of sample size on the total cost of acceptance testing by integrating the risk of making wrong acceptance decisions. In the first approach, the acceptance decision choice is either to accept or reject the pavement, and the risk of making wrong decision is computed using numerical *M*-method. In the second approach, the acceptance decision is extended to pay factor, and the risk of accepting higher pay factor is computed through a numerical method. In both approaches, the sample size that corresponds to the lowest total cost is the optimal sample size.

7.1 Approach 1: Cost Analysis Considering Accept/Reject Decisions

In Approach 1, three cost components are considered and the decision choice is limited to either accept or reject. The first two cost components are adopted from the sample size optimization model developed by Gharaibeh et al. (2010), namely cost for sampling and testing and cost of making incorrect acceptance decisions. The third cost component is relevant future cost (e.g., maintenance cost) caused by the defective or failed materials.

Equations 7.1–7.4 illustrate the formulas to calculate the total cost and individual costs. Equation 7.1 illustrates that the total cost is sum of three cost components: the sampling and testing cost ($cost_T$), the cost of making incorrect decisions ($cost_D$), and the future cost ($cost_F$) of the defective materials. Equation 7.2 computes the sampling and testing cost as the product of sample size n and the unit cost c of sampling and testing. Equation 7.3 computes the cost associated with making the incorrect decision of accepting defective materials. This cost component is the expected cost of the defective portion. It is calculated as the product of (1) the risk (β) of INDOT in accepting RQL pavement using proposed numerical *M*-method (illustrated in Figure 6.1), (2) the proportion of pavement with RQL (P_{RQL}) according to the historical records, and (3) the contract amount of the pavement, calculated as the bid unit price (B) of the pavement times the lot size (S). Equation 7.4 computes the future maintenance cost due to the defective pavement as a fraction of the expected cost of the defective portion, where $I_{percent}$ is the cost impact coefficient of defective materials that represents the fraction.

$$cost_{total} = cost_T + cost_D + cost_F \quad (\text{Equation 7.1})$$

$$cost_T = c \times n \quad (\text{Equation 7.2})$$

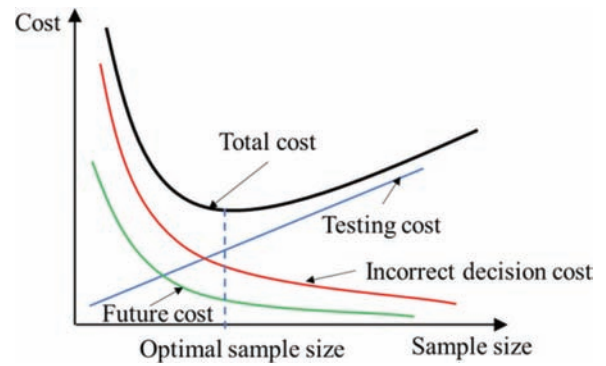


Figure 7.1 Determining the optimal sample size through cost analysis.

$$cost_D = \beta \times P_{RQL} \times (B \times S) \quad (\text{Equation 7.3})$$

$$cost_F = \beta \times P_{RQL} \times (B \times S) \times I_{percent} \quad (\text{Equation 7.4})$$

Figure 7.1 conceptually illustrates the curves of these cost components with varying sample sizes. The straight line for the sampling and testing cost illustrates that this cost component increases linearly with the increase of sample size. Both curves for cost of making incorrect acceptance decision and relevant maintenance cost decrease as the sample size increases due to the risk factor—larger sample size, lower risk. The total cost decreases as the sample size increases till reaching the lowest total cost at the optimal sample size and then increases from this point on as the sample size continues to increase. The sample size corresponding to the lowest total cost is the optimal sample size.

7.2 Approach 2: Cost Analysis Considering Pay Factor Decisions

In Approach 2, we extend the acceptance/rejection decision to pay factor decision—making payment adjustment (either penalty, full payment, or bonus) according to the quality level of the product. We still consider the same three cost components, but the computations for the cost of making incorrect decision and the future cost associated with defective materials are different.

Equations 7.5–7.8 illustrate the computation in detail. Considering that the pay factor corresponds to different quality levels, the risk to the agency in Approach 2 is defined as the probability of using a higher pay factor and the cost for this type of incorrect decisions is the overpay portion, not the entire contract price. Consequently, the cost of choosing incorrect pay factor is calculated as the difference between the payment as suggested by the samples and the true payment for the population. Note that pf_{sample} represents pay factor as suggested by the samples and $pf_{population}$ represents the true pay factor for the population. The future cost of defective materials is still computed as a fraction of the cost of making

incorrect decisions. The assumption is that the agency faces loss and incur cost only when the sample pay factor is higher than the population pay factor. If the sample pay factor is smaller, the cost of incorrect decision and the future cost are both zero.

$$cost_{total} = cost_T + cost_D + cost_F \quad (\text{Equation 7.5})$$

$$cost_T = c \times n \quad (\text{Equation 7.6})$$

$$cost_D = (pf_{sample} - pf_{population}) \times (B \times S) \quad (\text{Equation 7.7})$$

$$cost_F = (pf_{sample} - pf_{population}) \times (S \times B) \times I_{percent} \quad (\text{Equation 7.8})$$

The numerical *M*-method used in Approach 1 is no longer applicable to computing the risk in Approach 2 because it only considers the acceptance/reject decision but does not distinguish different pay factors. Therefore, a numerical method is developed to compute the costs in Approach 2, illustrated in Figure 7.2. The process starts with a pair of μ_0 and σ_0 , which are the mean and standard deviation of the pavement estimated using a large number of samples from the historical records. For all possible populations that contributed to this pair of mean (μ_0) and standard deviation (σ_0), their population means follow the normal distribution $N(\mu_0, \sigma_0^2)$.

Based on this observation, we generate 100 populations by randomly selecting population mean μ from $N(\mu_0, \sigma_0^2)$ and setting the population standard deviation σ to be σ_0 . For each of the generated 100 populations, we run 1,000 simulations, i.e., repeat the process—taking n random samples from the population and computing the cost using Equations 7.5–7.8 1,000 times. The average of the resulting 1,000 total costs is the estimate of the total cost of this particular population. The average of the total costs of the 100 populations is the expected cost of sample size n . By varying the sample size, the cost-sample size curve is derived, and the minimal total cost and the corresponding optimal sample size are identified.

7.3 Case Study

Air voids acceptance in HMA pavement is used as an example to demonstrate the two cost analysis approaches. The pavement property is estimated using air voids QA data from INDOT projects in the Crawfordsville District in 2016. The average and standard deviation of air voids are 3.89% and 0.8%. The bid price is about \$86/t. According to the specification, a typical lot of HMA pavement consists of 3,000t and is divided into 5 sublots. The sampling and testing cost for one sample is assumed to be \$200. The cost impact of the defective materials $I_{percent}$ is set to be 0%, 50%, and 115% of the bid price.

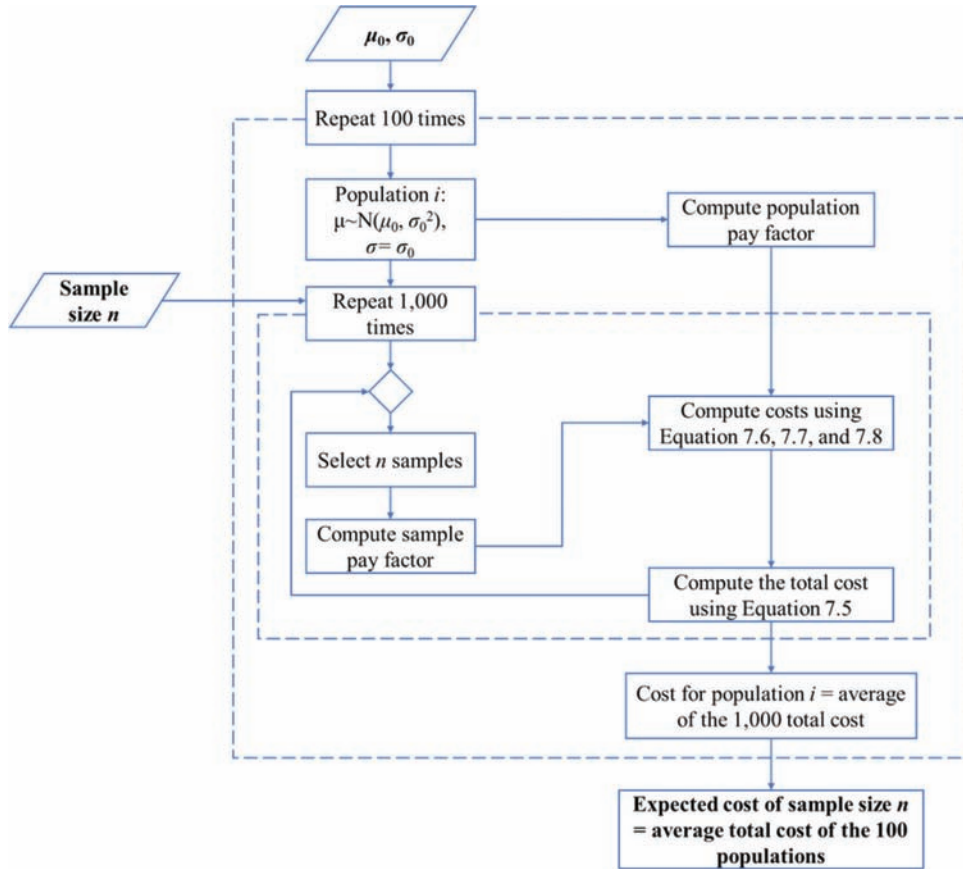


Figure 7.2 Approach 2 workflow for determining the total cost at a given sample size n .

7.3.1 Results from Approach 1

Six scenarios are considered in Approach 1 with different sets of P_{RQL} and $I_{percent}$, listed in Table 7.1. In all scenarios, AQL and RQL are set to be 10% and 50% and the contractor's risk is set to be 0.1.

Figure 7.3 illustrates the cost curves for the six scenarios, where the vertical dash lines intersect the horizontal axis at the optimal sample size. For instance, in Scenario 4 (see Figure 7.3d), the optimal sample size is 10 and the corresponding total cost is \$2,930. These curves can be cross-compared to evaluate the impact on the optimal sample size from the parameters of P_{RQL} and $I_{percent}$. Figure 7.4 illustrates the impact of $I_{percent}$ given a specific P_{RQL} . When $P_{RQL}=0.1$, the cost impact coefficient $I_{percent}$ does not have a significant impact on the sample size, which remains 10 for all three scenarios. When $P_{RQL}=0.2$, the optimal sample size increase as $I_{percent}$ increases. In all scenarios, a larger $I_{percent}$ leads to a higher total cost. Figure 7.5 illustrates the impact

TABLE 7.1
Case study settings in Approach 1.

Scenario	P_{RQL}	$I_{percent}$
1	0.1	0
2		50
3		115
4	0.2	0
5		50
6		115

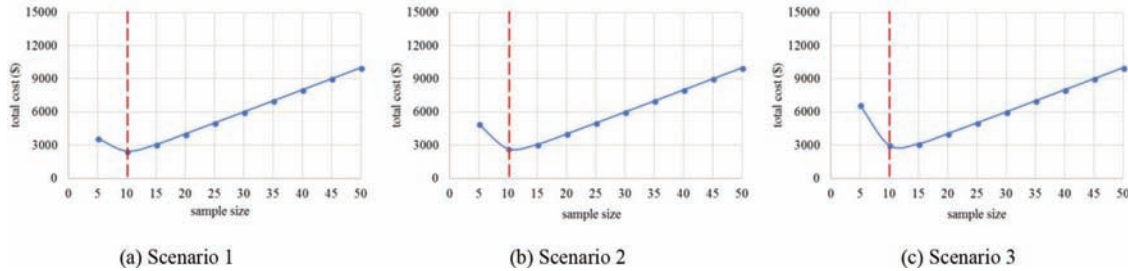


Figure 7.3 Optimal sample sizes and total costs from Approach 1.

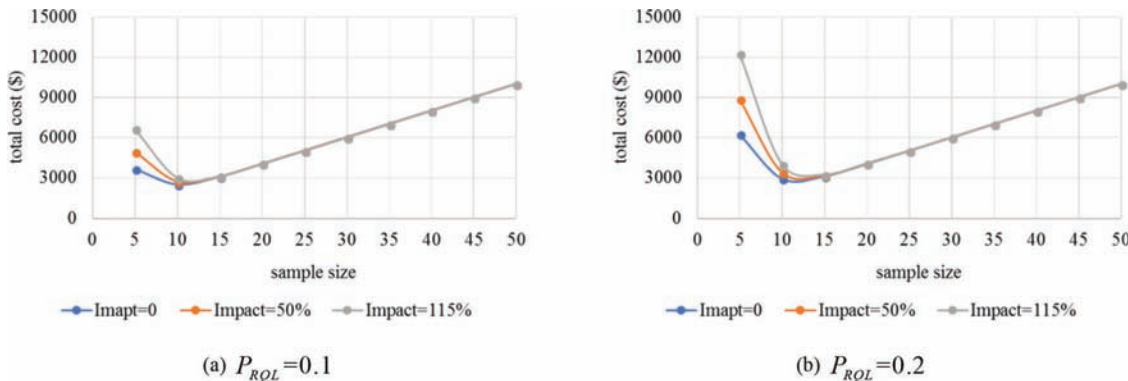


Figure 7.4 Impact of P_{RQL} in Approach 1.

of P_{RQL} given a specific $I_{percent}$. In general, the optimal sample size and total cost both increase with the increase of P_{RQL} , except that the optimal sample size remain unchanged when the cost impact of defective materials is not considered. It is noted that when the sample size is relatively large (>20), the total cost is consistent in all scenarios and increases linearly with the increase of sample size. This is because when the sample size is large, the risk of accepting defective material becomes very small, and the cost associated with erroneous decisions is negligible compared to the sampling and testing cost.

7.3.2 Results from Approach 2

In Approach 2, the pay factor (PF) of air voids is calculated using Equation 7.9 and 7.10 according to the INDOT specification (INDOT, 2016). Note that when $PWL < 50$, the material is considered as failed material and the pay factor is zero.

$$PF = (105 - 0.5 \times (100 - PWL)) / 100 (PWL > 90) \quad (\text{Equation 7.9})$$

$$PF = (100 - 0.000020072 \times (100 - PWL)^{3.5877}) / 100 (50 \leq PWL \leq 90) \quad (\text{Equation 7.10})$$

Nine scenarios are considered in Approach 2 with different combinations of population and cost impact coefficient, listed in Table 7.2. For population 1 and 2,

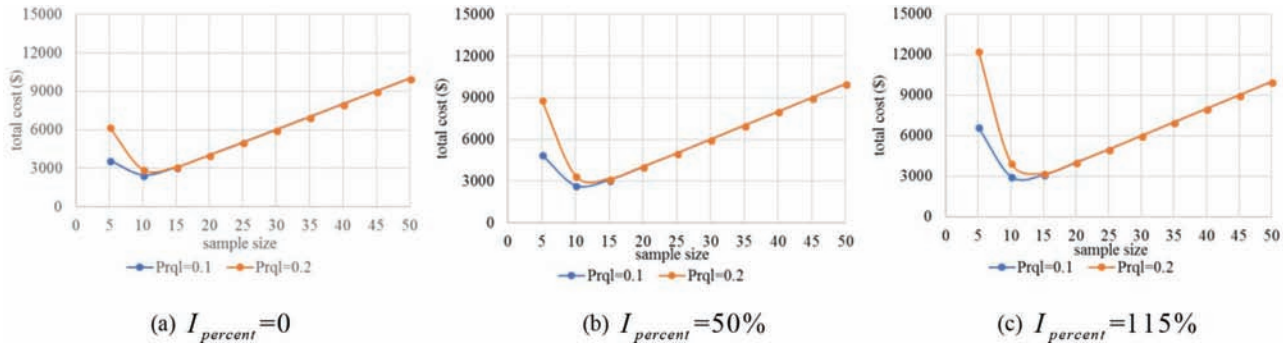


Figure 7.5 Impact of $I_{percent}$ in Approach 1.

TABLE 7.2
Case study settings in Approach 2.

Scenario	Population	(μ_0, σ_0)	$I_{percent}$
1	1	(4, 0.47)	0
2			50
3			115
4	2	(4, 0.7)	0
5			50
6			115
7	3	(3.89, 0.8)	0
8			50
9			115

their mean value is the center of the specification limits, i.e., $\mu_0 = \frac{2.6 + 5.4}{2} = 4$ and the standard deviation is set as $\sigma_0 = \frac{USL - \mu_0}{3} = \frac{5.4 - 4}{3} = 0.47$ and $\sigma_0 = \frac{USL - \mu_0}{2} = 0.7$. Population 3 is estimated from the INDOT historical data with a mean value of 3.89 and standard deviation of 0.8. As a result, Population 1 has the highest quality level, with only 0.28% of the pavement in the RQL, and Population 3 has the lowest quality level, with 10.68% of the pavement in RQL. The proportion of Population 2 in RQL is 4.56%.

Figure 7.6 illustrates the cost curves for the nine scenarios, where the vertical dash lines intersect the horizontal axis at the optimal sample size. For instance, in Scenario 4 (see Figure 7.6d), the optimal sample size is 10 and the corresponding total cost is \$6,332. Similar to Approach 1, cost curves in Approach 2 can also be cross-compared to evaluate the impact on the optimal sample size from the parameters of P_{RQL} and $I_{percent}$. Figure 7.7 illustrates the impact of $I_{percent}$ given a specific population, and the results suggest that the optimal sample size and total cost both increase as the cost impact of defective materials increases. Figure 7.8

illustrates the impact of the pavement quality given a specific cost impact coefficient, which indicates that the optimal sample size and total cost both decrease as the quality of the pavement increases.

Compared with Approach 1, Approach 2 results in larger optimal sample size and higher total cost when the same historical data are used. This is because the acceptance decision in Approach 1 is either acceptance or rejection; INDOT only overpays for the failed material (i.e., the fraction defective is larger than 50%). In Approach 2, however, the pay factor is incorporated, the decision choices are all possible pay factors ranging from 0 (for failed materials being rejected) to over 1 (bonus pay), and INDOT might overpay the contractor for acceptable materials when incorrectly using a higher pay factor.

7.4 Findings and Recommendations

The main findings and recommendations are as follows.

- The cost analysis methods are capable of determining the optimal sample size to minimize the total cost including the cost attributed to the risk.
- Since Approach 1 considers Accept/Reject as the only decision choices but Approach 2 considers all pay factors ranging from 0 (Reject) to over 1.0 (bonus), the risk is defined differently in two approaches. Results show that Approach 2 results in larger optimal sample size under the same setting as Approach 1.
- As expected, the optimal sample size and the corresponding minimal total cost are sensitive to the values of the input parameters and therefore, it is very important to objectively estimate the input parameters using historical records.

As a practical implementation, the cost-benefit analysis has been incorporated in the newly developed web tool as a standalone tab, detailed in Section 10.7 in Chapter 10.

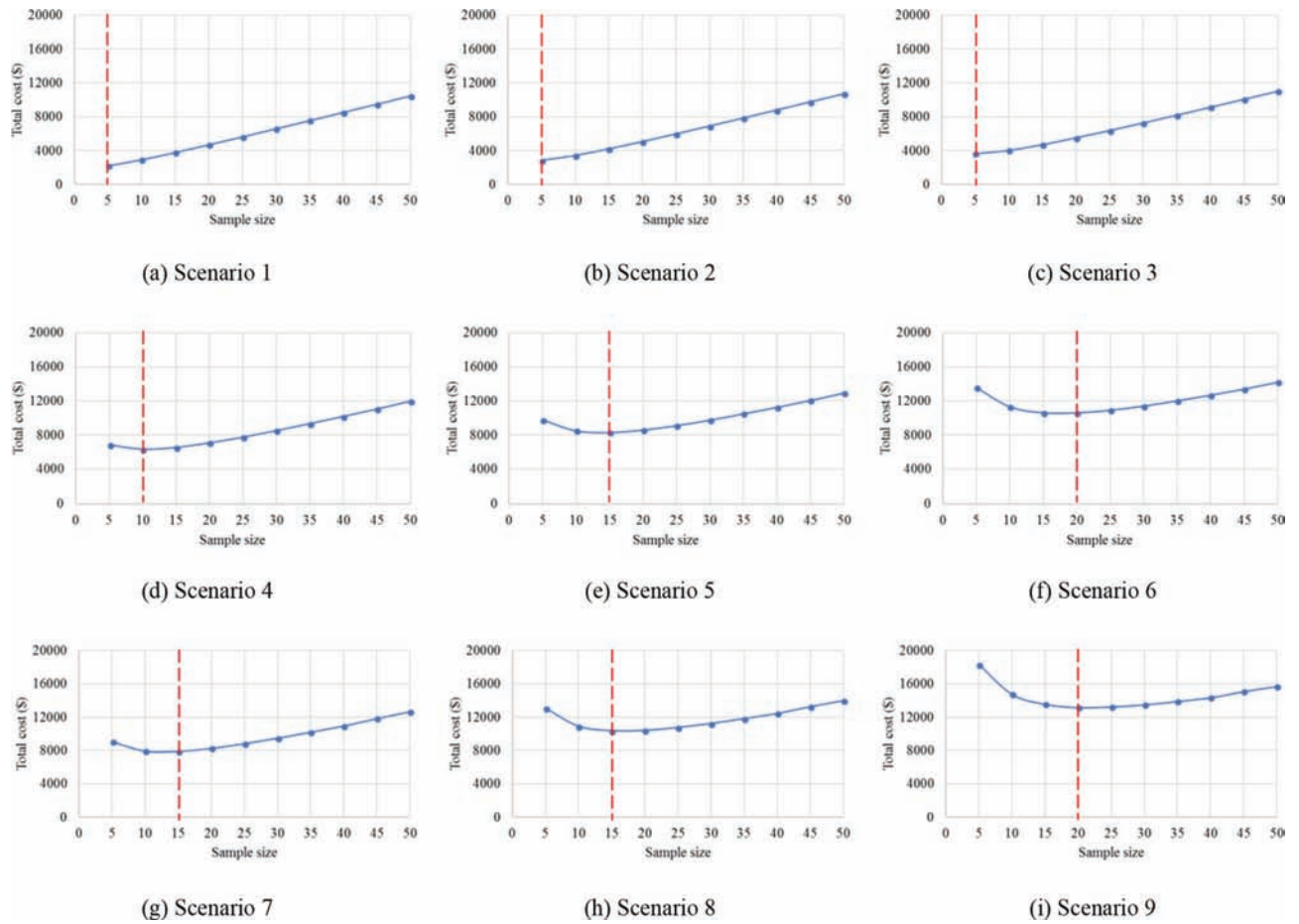


Figure 7.6 Optimal sample sizes and total costs from Approach 2.

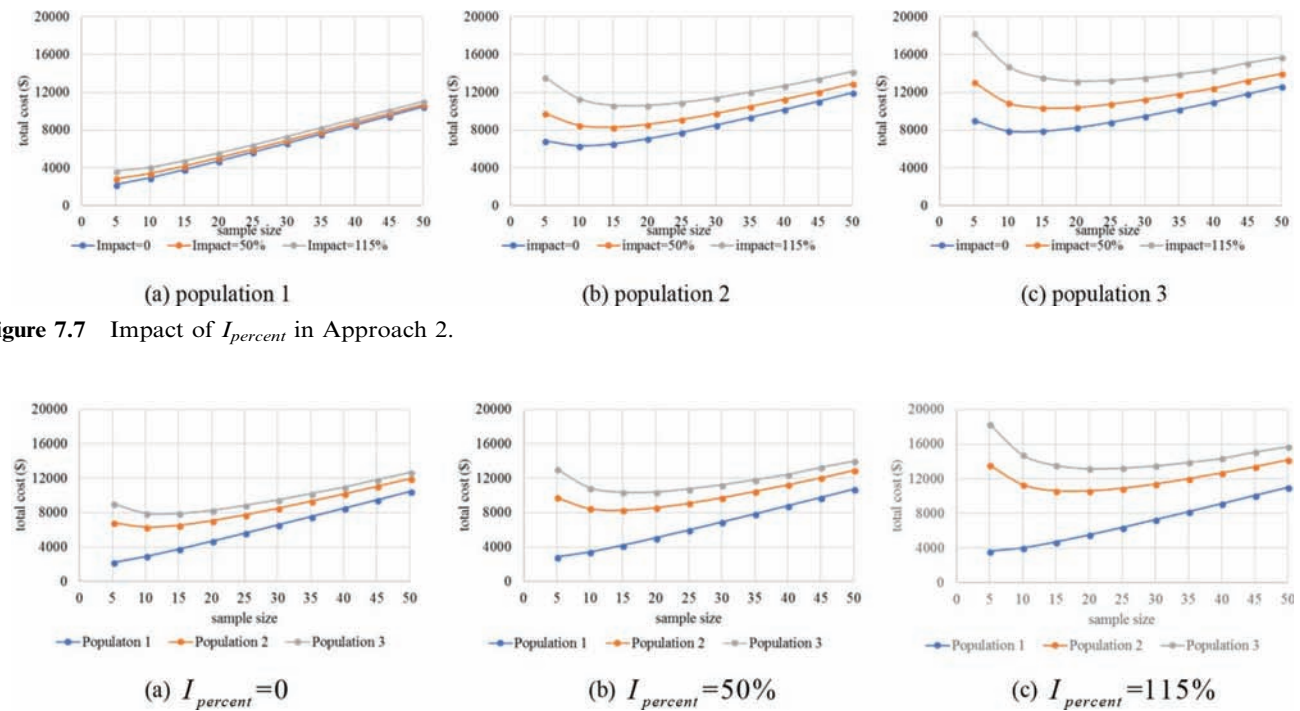


Figure 7.8 Impact of the pavement quality in Approach 2.

8. QC CHART PATTERN ANALYSIS

In the current practice, the acceptance testing in pavement QA practice relies on a lot-by-lot sampling and testing strategy, i.e., INDOT takes random samples from each lot, makes decision to whether accept or reject the lot, and moves on to the next lot. Variance among lots is not considered. While such practice ensures the compatibility of each lot with the specification requirement, it does not guarantee the consistency of multiple lots in a project. Motivated by the promise of QC charts in controlling the consistency among multiple lots, this chapter focuses on applying QC chart to QA data as a complementary tool in acceptance testing, aiming at ensuring the consistency of the production process at a project level.

8.1 Quality Control (QC) Chart

QC chart is a statistical tool to determine if a process is in a state of control. A process is in statistical control when it only involves chance causes of variation, and is out-of-control when assignable causes exist (Montgomery, 2009). The QC chart is an effective tool to control the production process and detect potential abnormalities.

8.1.1 Typical QC Chart

Figure 8.1 displays a typical QC chart, which plots the quality characteristic (e.g., air voids and flexural strength) computed from a group of samples versus the group number or time. It consists of three levels: center line, upper control limit (UCL), and the lower control limit (LCL), represented as horizontal lines. The center line represents the average quality characteristic in the process. The UCL and LCL represent the limits of the quality characteristics if the process is in control. Normally, most points should fall between UCL and LCL, and when points fall outside the UCL or LCL, the process is regarded as statistically out-of-control.

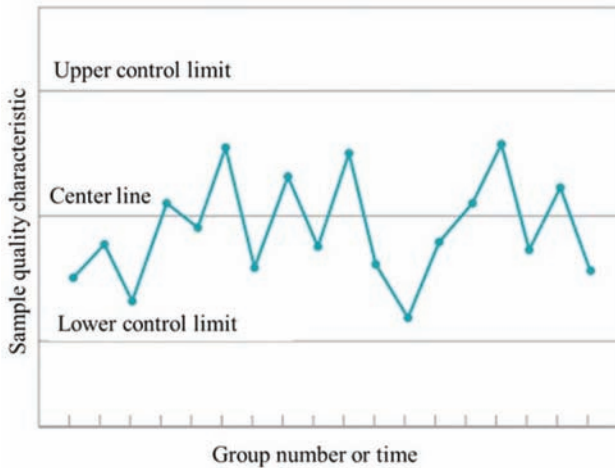


Figure 8.1 Typical QC chart (adapted from Montgomery, 2009).

8.1.2 Construction of QC Chart

According to the measurement of quality characteristic being applied, there are two common types of QC chart: \bar{x} chart and R chart. The \bar{x} chart uses the mean value of the quality characteristic to control the average of the process; R chart uses the range of the quality characteristic to control the variability of the process. These two types of QC charts are adopted in this study to evaluate the quality of the pavement. The construction process of both types of charts is briefly introduced herein. Readers may refer to most quality control textbooks (e.g., Montgomery, 2009) for details.

Assuming the quality characteristic follows normal distribution, but its mean μ and standard deviation σ are unknown, to plot the center line and the control limit lines, m groups of samples at the size of n will be used to estimate their values. Specifically, for a group of samples with values of x_1, x_2, \dots, x_n , the average of this group is calculated as $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$. The center line of the \bar{x} chart is derived as $\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m}$, where \bar{x}_m is the average of m^{th} group of samples. In the case of pavement acceptance testing, for instance, one lot can be viewed as a group and the number of samples in a lot as the sample size. If a pavement project consists of 10 lots, and 5 random samples are taken per lot, then $m = 10$ and $n = 5$, and the center line of the \bar{x} chart is computed using the above procedure.

Three-sigma rule is applied to construct the control limits of UCL and LCL. This rule ensures that for a process that is under control, the probability of the quality characteristic outside the area bounded by UCL or LCL is 0.001. In practice, the standard deviation σ of the process is estimated from the ranges of m groups of samples. Given x_1, x_2, \dots, x_n , the range of this sample group is computed as $R = x_{\max} - x_{\min}$. With m groups, the average range is $\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m}$. As a result, the \bar{x} chart is constructed as follows, where A_2 is a constant varying with the sample size and can be looked up in the standard factor table for QC charts. Note that to ensure the reliability of the QC charts, it should be established using a relatively large number of groups (around 20) (Montgomery, 2009).

$$UCL = \bar{\bar{x}} + A_2 \bar{R}$$

$$\text{Center line} = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - A_2 \bar{R}$$

Similarly, the center line and control limits for R charts are computed as follows, where D_3 and D_4 are constants varying with the sample size, which can also be looked up in the standard factor table for QC charts.

$$UCL = D_4 \bar{R}$$

$$\text{Center line} = \bar{R}$$

$$LCL = D_3\bar{R}$$

8.1.3 Control Charts for Individual Measurements

A special type of control chart, referred as music bar chart, plots individual measurements (i.e., sample size = 1) according to the order of test (Willenbrock, 1976). This type of chart can be used to identify trends during the production as well as to evaluate if the work satisfies the specification by comparing the test results with the target value and specification limits regulated by the agency. In the case of pavement acceptance testing, the measurement of each sample of the subplot can be plotted in the music bar chart. By setting control limits using all samples in one lot, it is possible to evaluate the production process at a lot level.

8.2 Interpretation of QC Chart

A control chart can identify out-of-control conditions when one or more points fall outside the control limits (i.e., UCL and LCL) or points formulate certain patterns. This study reviews some typical patterns of QC chart that can be used to detect out-of-control conditions. Table 8.1 lists the most common patterns of QC charts, their recognition criteria, and the possible causes based on the review of existing studies (Pyzdek & Keller, 2003; Stapenhurst, 2013; Montgomery, 2009; McNeese, 2004). For the QC charts included in Table 8.1, 2-sigma and 1-sigma lines are also included, separating the area bounded by the UCL and LCL lines into two A zones, two B zones, and two C zones to assist further in identifying certain patterns. Note that both \bar{x} chart and R chart should be used to ensure that both the average and variability of the process are controlled simultaneously.

8.3 Case Study

In this section, two case studies were conducted to demonstrate the efficacy of QC chart using HMA air voids QA data from two INDOT projects. The result of QC chart was compared with acceptance decisions derived from the numerical M -method and the current PWL method.

8.3.1 Case Study 1

The first case study is the RS-31500 project in Crawfordsville. This project consists of 17 lots and each lot contains 5 sublots. One air void sample is taken at a random location for each subplot. The average and range is computed for each lot.

Figure 8.2 illustrates the music bar chart for case study 1, with values of individual samples plotted along the horizontal axis. In both charts, specification limits (USL and LSL) are represented as black dash lines.

In Figure 8.2a, the control limits (UCL and LCL) and center line (CL) are established using the first 25 samples, which serve as the base line to evaluate if the following process is under control. Samples from some lots tend to get clustered (e.g., sample 15–20, sample 55–60) and some of them tend to increase or decrease (e.g., sample 25–30, sample 45–50). In Figure 8.2b, the control limits and center line are established on a lot basis, i.e., the five samples of one lot are used to compute the UCL, LCL and CL for this specific lot. For each lot, the red dot line represents the center line, and the blue shaded area represents the area bounded by the UCL and LCL, which implies the variation among samples in one lot. From the results, the control limits vary a lot across different lots, while central lines are relatively stable with slight fluctuation. Additionally, some of the control limits fall outside the specification limits (e.g., sample 5–10 for lot 1), indicating the large variation within the lot.

Figure 8.3 illustrates the R chart and \bar{x} chart for the 17 average values, constructed following the method described in Section 8.1.2. In both charts, the red lines represent the control limits, the green lines represent the center line, and the red dot lines represent the two-sigma and one-sigma lines. The specification limits (USL and LSL) are indicated using blue dot lines in the \bar{x} chart. Each point, denoted as either “A” or “R”, plots the corresponding statistical value for a lot. Letter “A” means the lot is accepted and letter “R” means the lot is rejected, using our numerical M -method.

There are two main observations: (1) in general, the average and range of the air voids varies randomly between UCL and LCL, and (2) Pattern 1 is detected in the \bar{x} chart, revealing the process is out of control and lot 1 is the outlier.

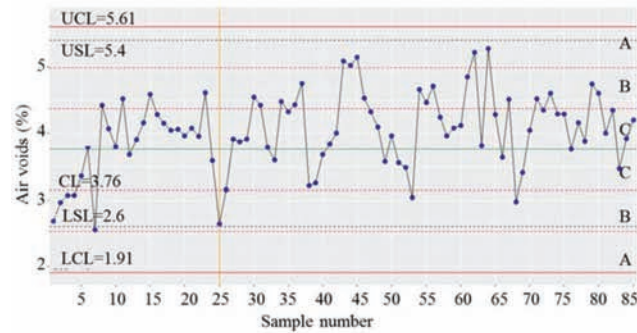
Table 8.2 compares the analysis results using QC chart, current PWL method, and numerical M -method. For lot 9 and lot 13, the QC chart results align with numerical M -method and PWL method. In QC chart, the average values are almost at the boundary, which is Pattern 1 in Table 8.1. The numerical M -method suggests to reject the lot and the PWL value is small (considering 0.9 being the acceptance threshold). For lots 1, 2, and 5, QC chart reveals Pattern 1 in the \bar{x} chart for lot 1 and large range values for lots 2 and 5, the numerical M -method rejects the three lots when RQL is set at 0.2 and 0.1, but accepts the lot when RQL is set at 0.3, and the current PWL method fails to detect the abnormal pattern and suggests full payment for all three lots. Table 8.2 results confirm the advantage of the numerical M -method over PWL in the current practice and suggest that QC charts can serve as a supplementary tool to detect abnormal conditions of the entire process and ensure the consistency at both lot and project levels.

8.3.2 Case Study 2

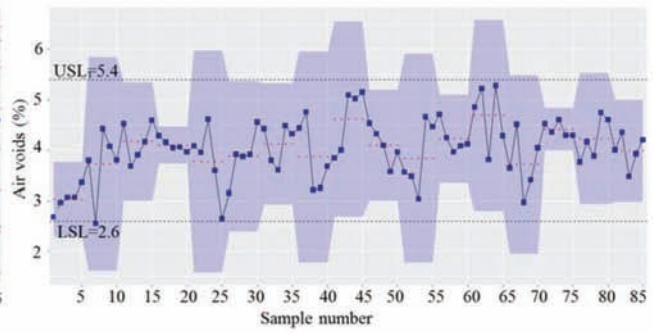
The second case study is the RS-34919 project in Fort Wayne. This project consists of 13 lots and each lot contains 5 sublots. One air void sample is taken at a

TABLE 8.1
Interpretation of most common QC charts patterns.

No.	Typical pattern	Recognition criteria	Possible causes
1		One or more points fall outside the control limits.	A special cause is present in the process.
2		Two out of three consecutive points fall in zone A or beyond. (Note: the two points must be on the same side of the average, i.e., either upper Zone A or lower Zone A).	A special cause exists.
3		Four out of five consecutive points fall in zone B or beyond. (Note: the four points must be on the same side of the average, i.e., either upper Zone B or lower Zone B).	A special cause exists.
4		Fifteen or more consecutive points fall in zone C, either above or below the center line.	It happens when samples are drawn from two or more distributions and have been combined, resulting in unusually high presence of points near the centerline. This pattern infer that the control chart is not valid for this process, user must first correct errors in the methods of sampling.
5		Eight or more consecutive points lie on both sides of the center line with none of the points in zone C.	There is more than one processes present but sampling is done for each process separately. Such pattern may occur if different shifts are operating at different averages.
6		(1) Seven points in a row above the center line, (2) Seven points in a row below the center line, (3) Seven points in a row trending up, or (4) Seven points in a row trending down.	There is an out-of-control situation if one of the pattern is true.

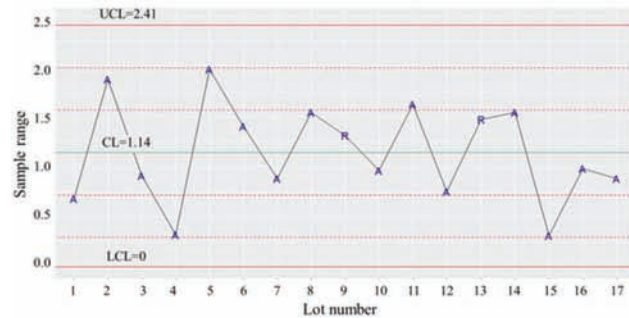


(a) control limits computed from first 25 samples

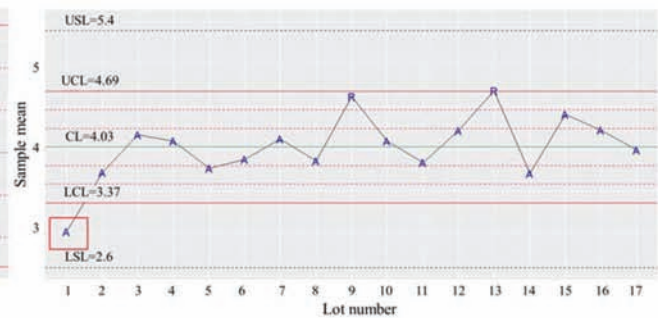


(b) control limits computed from each lot

Figure 8.2 Music bar chart for case study 1.



(a) R chart



(b) \bar{x} chart

Figure 8.3 QC charts for case study 1.

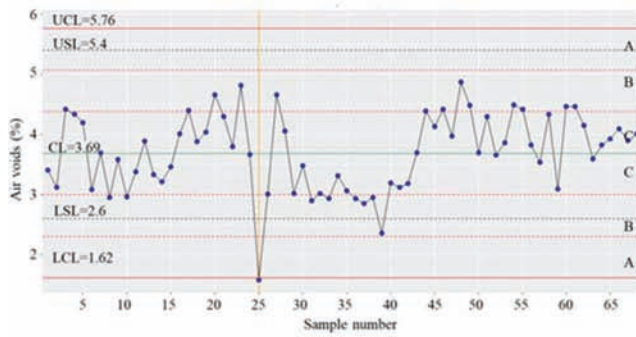
TABLE 8.2
Comparison among QC chart, PWL, and numerical *M*-method for case study 1.

Lot	QC chart		Numerical <i>M</i> -method (risk=0.1)			PWL
	R chart	\bar{x} chart	RQL=0.3	RQL=0.2	RQL=0.1	
1		Pattern 1	Accept	Reject	Reject	1
2	Large value		Accept	Reject	Reject	0.98
3			Accept	Accept	Accept	1
4			Accept	Accept	Accept	1
5	Large value		Accept	Reject	Reject	0.98
6			Accept	Accept	Accept	1
7			Accept	Accept	Accept	1
8			Accept	Accept	Accept	1
9		Close to the UCL	Reject	Reject	Reject	0.9
10			Accept	Accept	Accept	1
11			Accept	Accept	Accept	1
12			Accept	Accept	Accept	1
13		Falls on the UCL	Reject	Reject	Reject	0.87
14			Accept	Accept	Accept	1
15			Accept	Accept	Accept	1
16			Accept	Accept	Accept	1
17			Accept	Accept	Accept	1

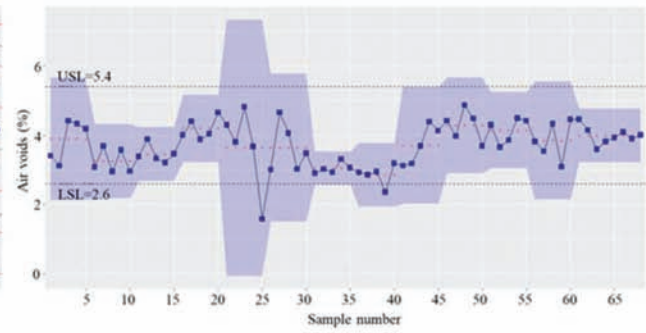
random location for each subplot. The average and range is computed for each lot.

Figure 8.4 illustrates the music bar chart for case study 2. In Figure 8.4a, the control limits and center line are established using the first 25 samples. Sample 30–45 are below the center line, which may be caused

by the introduction of some disturbances that affect a group of nearby samples. In Figure 8.4b, the control limits and center line for each lot are computed using samples of that lot. The process stabilizes after sample 45 because the UCL and LCL tend to be steady, indicating a more consistent process. There are some large

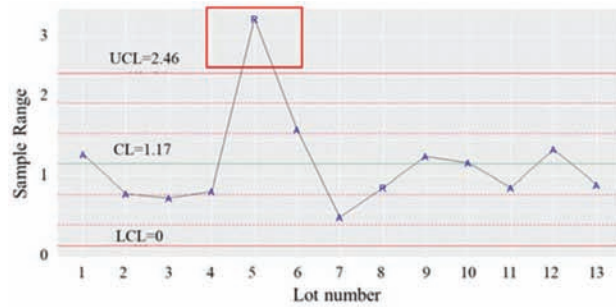


(a) control limits computed from first 25 samples

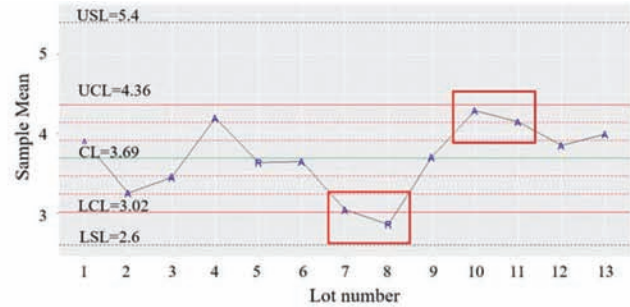


(b) control limits computed from each lot

Figure 8.4 Music bar chart for case study 2.



(a) R chart



(b) \bar{x} chart

Figure 8.5 QC charts for case study 2.

TABLE 8.3
Comparison among QC chart, PWL, and numerical M -method for case study 2.

Lot	QC chart		Numerical M -method (risk=0.1)			PWL
	R chart	\bar{x} chart	RQL=0.3	RQL=0.2	RQL=0.1	
1			Accept	Accept	Accept	1
2			Accept	Accept	Accept	1
3			Accept	Accept	Accept	1
4			Accept	Accept	Accept	1
5	Pattern 1 Large value		Reject	Reject	Reject	0.74
6			Accept	Reject	Reject	0.96
7		Pattern 2	Accept	Accept	Accept	1
8		Pattern 1 and 2	Reject	Reject	Reject	0.8
9			Accept	Accept	Accept	1
10		Pattern 2	Accept	Accept	Accept	1
11		Pattern 2	Accept	Accept	Accept	1
12			Accept	Accept	Accept	1
13			Accept	Accept	Accept	1

fluctuations during the process from lot 5 to lot 8 (i.e., sample 20 to 40) and the variation within a lot decreases, which might be caused by some actions being taken to control the construction quality.

Figure 8.5 illustrates the QC charts for case study 2. Pattern 1 is identified in the R chart: the range of lot 5 exceeds the UCL. Two patterns are recognized in the \bar{x} chart. Lot 8 exhibits Pattern 1: falls outside the LCL. Lots 7 and 8 and lots 10 and 11 exhibit

Pattern 2: two out of three consecutive points fall in zone A or beyond.

Table 8.3 compares the analysis results using QC chart, current PWL method, and numerical M -method. For lot 5 and lot 8, QC charts aligns with numerical M -method and PWL method. The QC charts reveals Pattern 1 in the R chart for lot 5 and Pattern 1 and 2 in the \bar{x} chart for lot 8. The numerical M -method suggests to reject both lots, and their PWL values are

small (considering 0.9 being the acceptance threshold). For lot 6, QC chart reveals the large range and numerical M-method rejects the lot when $RQL = 0.2$ and 0.1 . For lot 7, 10, and 11, QC chart detects pattern 2, while numerical M-method accepts the lots, and PWL method fails to identify abnormal conditions and suggests full payment for all three lots.

The discrepancy between out-of-control patterns identified using QC charts and the acceptance decisions made by the numerical M-method and PWL method is because both numerical M-method and PWL method rely on measurements of individual lot and therefore, they are more sensitive to the pattern where an individual lot is outside the control limits. On the contrary, QC charts evaluate the pattern formed by consecutive lots and detect out-of-control process even all lots are within the limits. Thus, QC chart and numerical M-method can be used as complementary tools to ensure not only the compatibility of individual lots with the specification but also the consistency at the entire project level.

8.4 Findings and Recommendations

The main findings and recommendations are as follows.

- QC chart constructed using pavement QA data is an effective tool to evaluate the consistency of pavement and ensure that the project is under control.
- QC chart and the numerical M-method complement each other to ensure the quality of both individual lot and consistency of the entire project, which is missing in the current practice.

9. VALIDATION AND CASE STUDIES

In this chapter, historical data from INDOT are used to validate (1) the testing power using numerical M-method in PCCP pavement, (2) the testing power using numerical M-method in HMA pavement, and (3) the testing power using simulation approach in soil compaction, and answer questions in implementation on (1) the influence of merging two lots on the testing power,

and (2) the influence of averaging samples in one subplot on the testing power.

9.1 Validation Using PCCP Data

In the development of numerical M-method, we assume the pavement population follows normal distribution and compute the testing power in acceptance testing using simulated population. In this section, historical PCCP data are used to check the normality of the population and validate the simulated testing power.

Both normality validation and testing power validation were performed using three sets of PCCP flexural strength data. The first dataset includes all historical flexural strength data from 60 projects, with a total of 2,469 sublots and 4,938 flexural strength measurements (2 samples per subplot). The second dataset is from project IR-31230, with 236 sublots and 472 flexural strength measurements. The third dataset is from project IR-33045, with 139 sublots and 278 flexural strength measurements.

9.1.1 Validation of Normality

Figure 9.1 illustrates the histograms of the three sets of flexural strength data. The black curves represent the empirical density distribution curve and the blue curves represent the theoretical normal distribution with the same mean and standard deviation as the dataset. The mean values of the three datasets are 650.04, 622.08, and 704.07, respectively. The standard deviations are 69.41, 44.22, and 60.38.

All three charts indicate that the empirical density distribution is slightly skewed from the normal distribution. The Shapiro–Wilk test (Shapiro & Wilk, 1965) was applied to check the normality of the data. The p-value for all three cases are less than 0.01, resulting in rejecting the null hypothesis (i.e., the population is normally distributed), which indicates that the flexural strength of the pavement does not follow normal distribution. The correlation between two samples taken in the same subplot was also examined. The correlation coefficients are 0.66, 0.69, and 0.88, indicating that they are strongly correlated. This might be a possible cause of the non-normality.

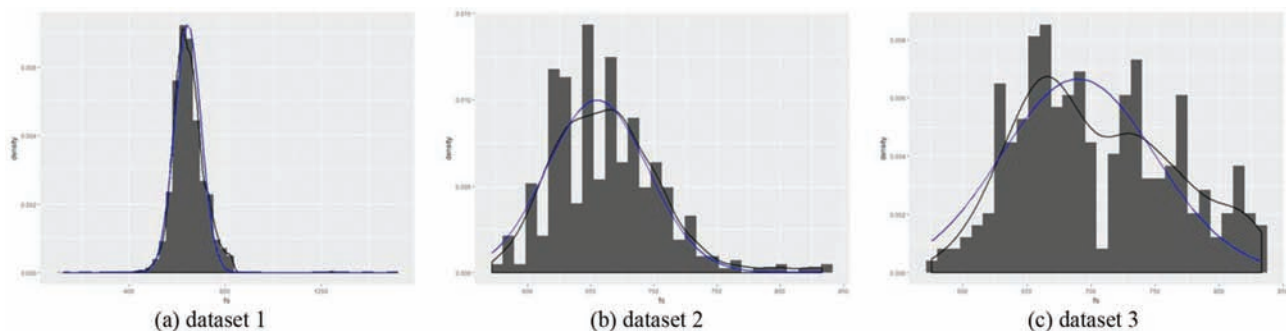


Figure 9.1 Histogram of flexural strength.

9.1.2 Validation of Testing Power

The process to compute the testing power, or the probability of rejecting the pavement, using real data is detailed as follows.

1. Preset desired RQL and agency's risk.
2. Calculate critical value M following method described in Section 5.2.1.2.
3. For each lot of the project, compute its defective rate p equals to $1 - \text{PWL}$. If $p < M$, we accept the lot, otherwise, we reject the lot.
4. The proportion of lots being rejected is the estimation of testing power.

According to the specification, the pavement with flexural strength less than 570 psi will receive penalty. To evaluate the effectiveness of the simulation approach for pavement with varying quality levels, this study sets the lower specification limit (LSL) to be 570 psi,

600 psi, and 650 psi, and the RQL is set to be 20%, 30%, and 40%. The desired risk level is 10%, and the sample size is 6 (2 samples/sublot * 3 sublots/lot).

Figure 9.2 illustrates the comparison of testing power obtained using simulated population and real data under various settings. The green curves represent the simulated testing power with respect to defective percentage of the pavement. The black dots represent the testing power computed from real data when the LSL equals 570, the red dots represent the testing power of real data when LSL equals 600, and the green dots represent the testing power of real data when LSL equals 650. Apparently, given the same dataset, the defective percentage of the pavement increases as the LSL increases. For instance, in case study 2 (Figure 9.2 d, e, and f), when LSL is 570, the defective percentage is almost zero. When LSL is 600, the defective percentage is less than 10%, while it increases to about 45% when LSL is 650.

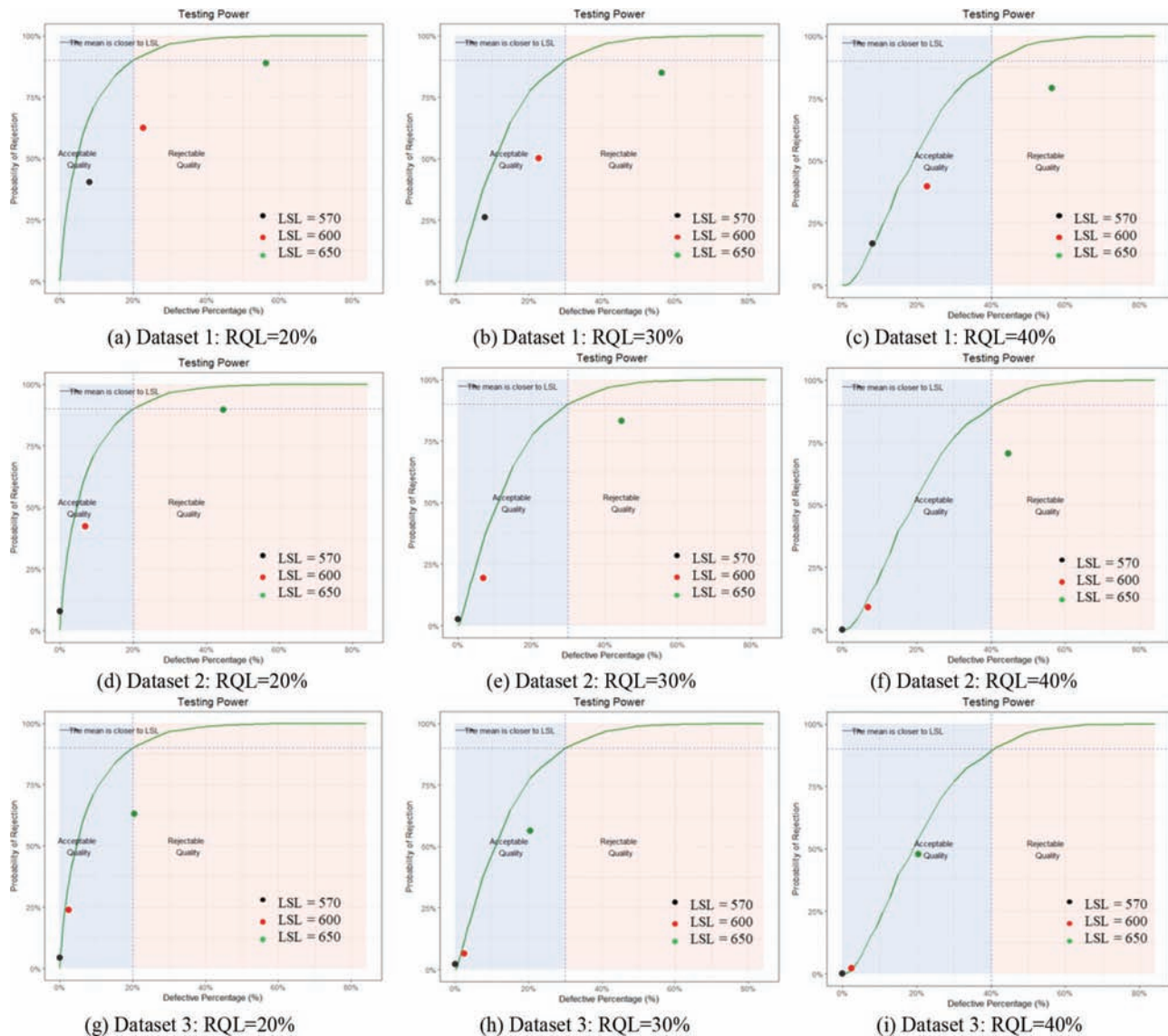


Figure 9.2 Comparison of testing power in flexural strength acceptance testing.

In general, the discrepancy between simulated testing power and the real one increases as the LSL increases for all cases. For instance, in case study 2, when the LSL is 570, the simulated testing power aligns well with the real one. When products are in rejectable region (e.g., LSL is 650), however, the real testing power is much lower than simulated one, indicating INDOT may take more risk than the theoretical risk. In addition, for all cases, larger RQL will lead to smaller discrepancy between the simulation and real data.

It is argued that one possible reason for the discrepancy between simulated testing power and the real one is the correlation between two samples of the same subplot (e.g., the correlation coefficient is 0.69 in case study 2). To mitigate such impact, we combine the two samples and treat their average as the flexural strength of the subplot. Consequently, the sample size for each lot becomes 3.

Figure 9.3 illustrates the comparison of testing power with averaged flexural strength. In general, averaging

two samples in one subplot reduces the discrepancy between simulated and real testing power. For instance, in case study 2, when LSL is 570 and 600, the testing power computed from simulation and real data is almost the same for all RQLs, and the discrepancy decreases compared to the original result when LSL is 650 (green dots in Figure 9.2 d, e, and f). Averaging the two samples in one subplot improves the testing power of acceptance testing as the averaged result is closer to the normality. However, such approach will decrease the sample size from 6 to 3, resulting in zero critical value in all scenarios and even the simulation approach cannot control the agency's risk at the desired level. Therefore, it is only preferred when there are relatively large number of sublots in one lot.

9.2 Validation Using HMA Data

In this section, the normality of the population and the testing power of simulated population computed via

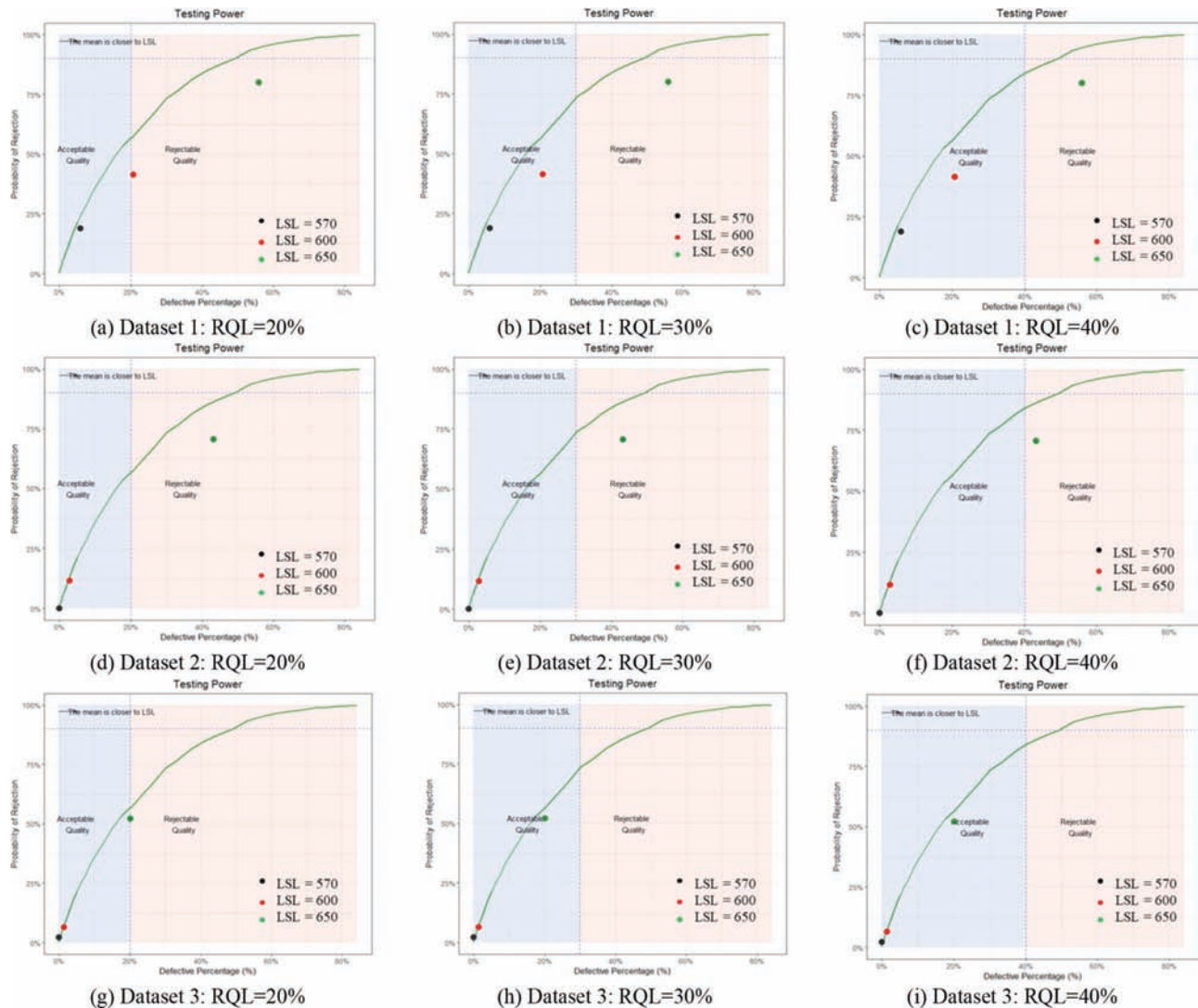


Figure 9.3 Comparison of testing power using averaged samples.

numerical *M*-method are validated using HMA data. Two sets of HMA air voids data were used to test the normality and validate the testing power. The first dataset includes all historical air voids data from 17 projects in Crawfordsville, with a total of 232 sublots and 232 air void measurements (one sample per subplot). The second dataset is from project IR-31500, with a total of 85 sublots and 85 air voids measurements.

9.2.1 Validation of Normality

Figure 9.4 illustrates the histogram of air voids, where the horizontal axis represents the air voids value and the vertical axis represents the density. In both charts, the black curves represent the empirical density of the measurements and the blue curves represent the density of normal distribution which has the same mean and standard deviation with the dataset. The mean values of the two cases are 3.89 and 4.02, and the standard deviations are 0.76 and 0.60. For case study 1, the p-value of Shapiro–Wilk test is 0.001 (less than 0.01), indicating the air voids in Crawfordsville does not follow normal distribution. For case study 2, the p-value of Shapiro–Wilk test is 0.35 (larger than 0.01), indicating the air voids in project IR-31500 follows normal distribution.

9.2.2 Validation of Testing Power

The validation of testing power using HMA data was conducted in a similar way to that using PCCP data. According to the specification, the USL and LSL for air voids acceptance testing are 5.4 and 2.6. The sample size is 5 (one sample/sublot * 5 sublots/lot). The empirical PWL computed from the real data is 93.1% and 98.82% for two case studies, respectively. The desired risk level is set to be 10% and RQL is set to be 30% and 40%.

Figure 9.5 illustrates the comparison of testing power obtained using simulation approach and real data. In both charts, the green curves represent the simulated testing power, and the green dots represent the testing power computed from the real data. In general, the real

testing power aligns well with the simulated one expect that in case study 1, the simulation result leads to higher probability of rejecting the pavement when RQL equals 30% (Figure 9.5a). Such discrepancy might be caused by the non-normality of the population.

9.3 Validation Using Soil Data

This section aims at (1) validating the effectiveness of the proposed simulation approach in determining the testing power of soil compaction, (2) proving the impact of spatial autocorrelation on the testing power, and (3) validating the effectiveness of proposed spatial indices in mitigating the impact of spatial autocorrelation, using soil CMV data. Three types of testing power were computed and compared: (1) the testing power was computed using simulation approach (described in Section 5.3), which assumes the CMV of the entire area is spatially random distributed, (2) the testing power was computed using real CMV data, where CMV of the entire area is spatially auto correlated (proved in Section 4.2), and (3) the testing power was computed using real CMV data but the sample locations were determined through a random sampling process that controls the sample spatial pattern (described in Section 4.3).

9.3.1 Data Description

Four datasets were used, each of which contains CMV data (measured via Intelligent Compaction, or IC technology) in one construction area in the U.S. Highway 31 Kokomo project. Figure 9.6 illustrates the grid map (grid size is 2m × 2m) of CMV value converted by averaging the CMV measurements within each grid.

Given a random sample location, its CMV value is derived through bi-linear interpolation based on the sample location, shown as Figure 9.7. The red point represents the random sample, and the blue points are the center of its neighboring grids. The CMV value of the random sample is estimated through linear interpolation in both x and y directions. The data processing was performed on a commercial geographical information system (GIS) platform.

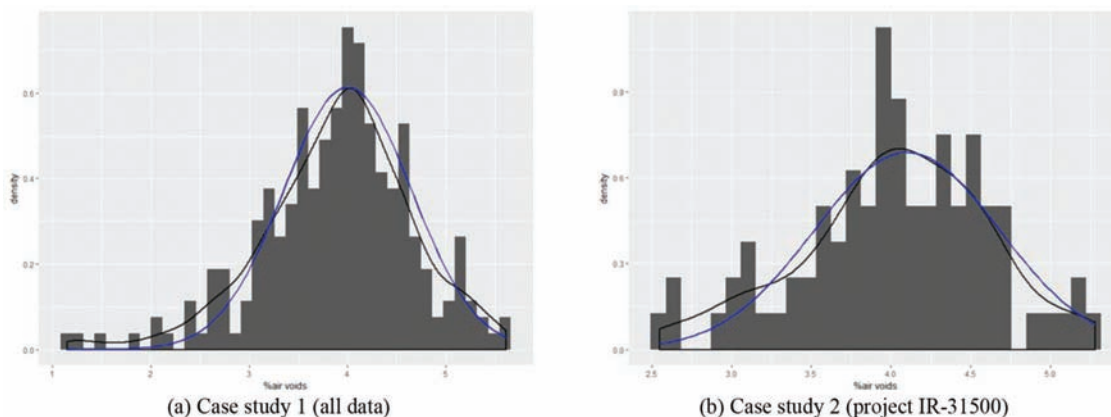


Figure 9.4 Histogram of air voids.

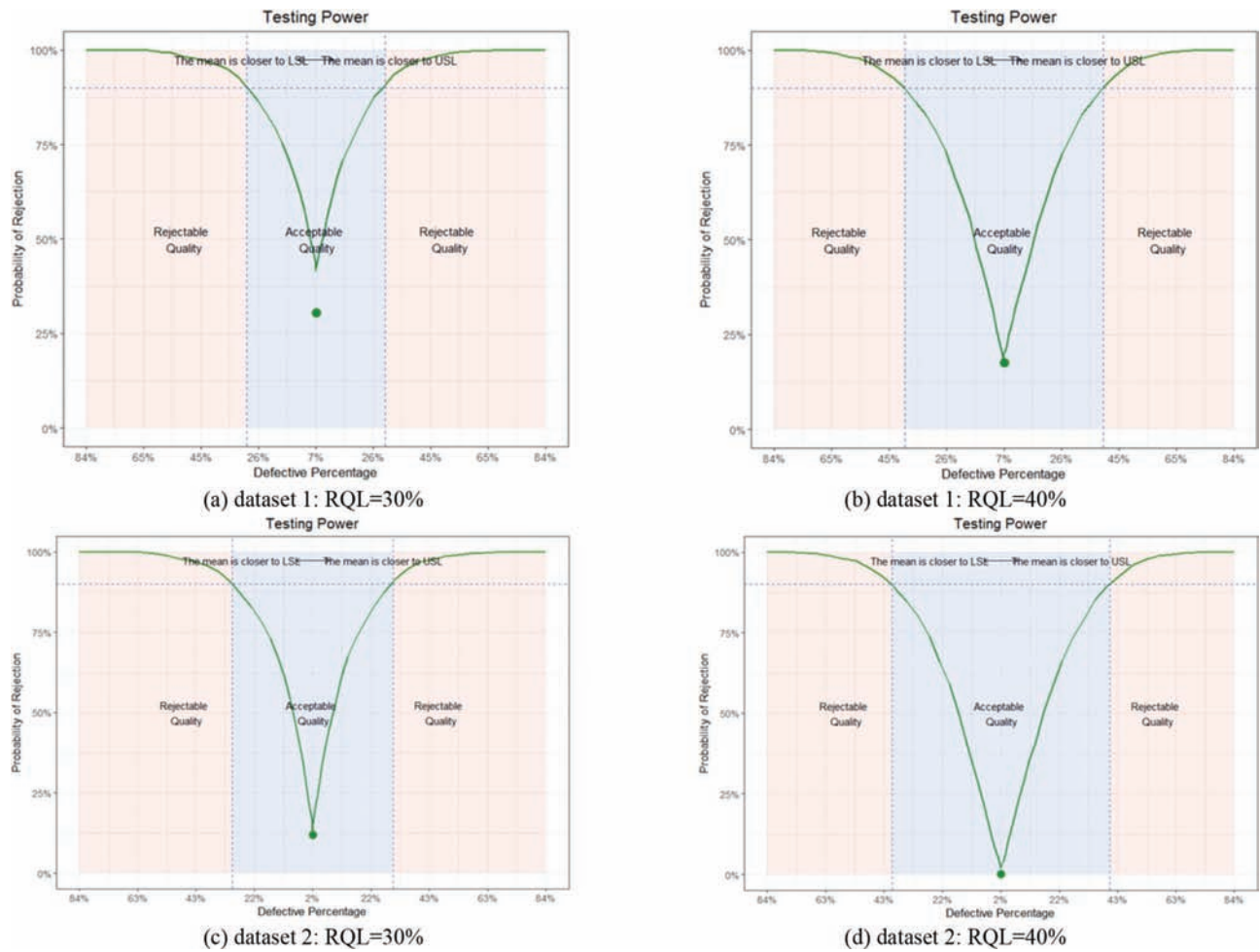


Figure 9.5 Comparison of testing power in air voids acceptance testing.

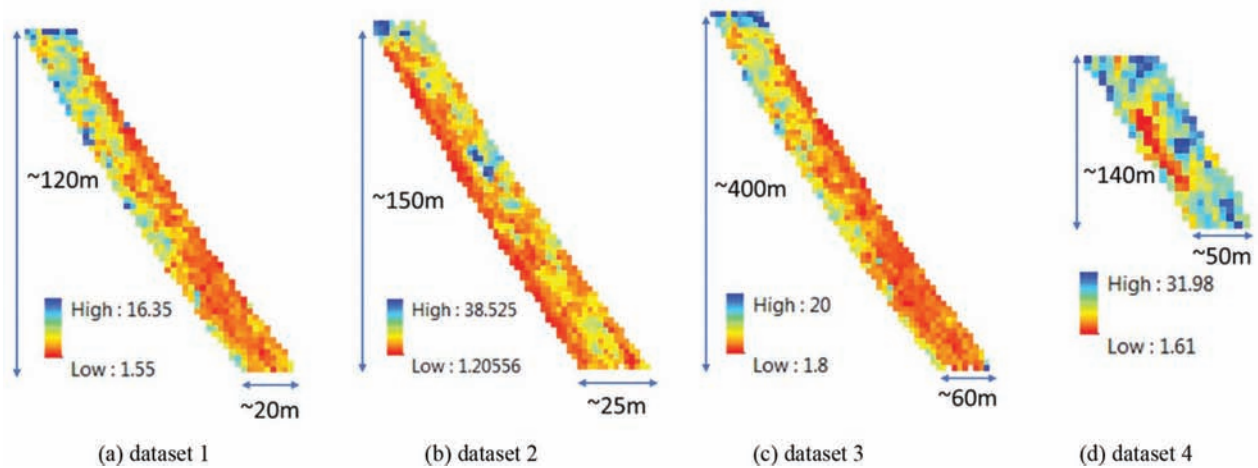
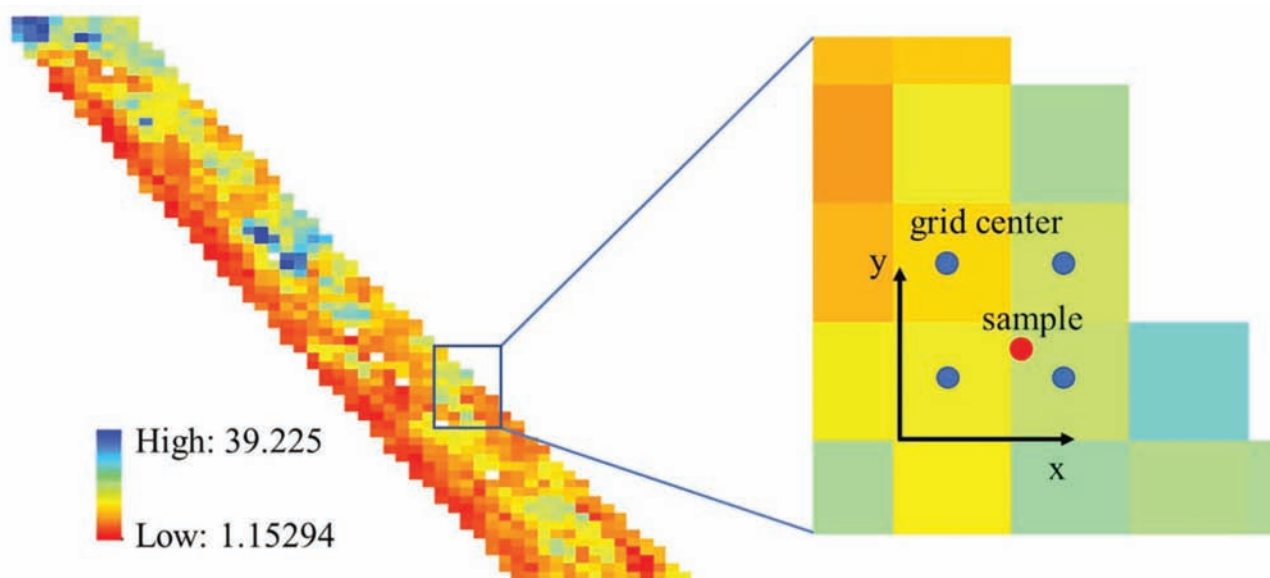


Figure 9.6 Grid map of CMV data.

9.3.2 Validation of Testing Power

This study uses the CMV data as the measurement to decide whether a sample is defective or not. That is, if the CMV of soil is smaller than the target value, the soil

is regarded to be defective. Figure 9.8 illustrates the comparison between simulated and real testing power with the sample size being 3, 6, and 9. The horizontal axis represents the defective rate of the area varying from 0 to 1, which is achieved by setting different target



(a) grid map for CMV value

(b) derivation of CMV value for random sample

Figure 9.7 Preprocessing of CMV data.

CMV values. The vertical axis is the testing power of soil acceptance testing, or the probability of identifying defective soil. Black curves represent the testing power computed from simulation approach, and blue curves represent the testing power computed using real data.

In general, the testing power computed through simulation aligns with that obtained from real data, especially when the defective rate is relatively large. For dataset 1–3, when the defective rate is small (e.g., less than 0.3), the simulated testing power is larger than the real one, and the discrepancy peaks when the defective rate is around 0.1.

Figure 9.9 illustrates the discrepancy between simulated and real testing power in dataset 2. The discrepancy increases as the sample size gets larger, which might be contrary to common sense—larger sample sizes should lead to more reliable decisions and the testing power should be closer to the simulated one. In fact, such pattern actually reveals the influence of spatial autocorrelation. As the sample size increases, samples are more likely to be spatially clustered, leading to decreased effective sample size, which may have a negative impact of the testing power.

Figure 9.10a illustrates the comparison of the testing power of the random sampling process that controls the sample spatial indices within the 80% reference range (red line) to that of the original random sampling (blue line) that does not control the spatial indices, based on dataset 2 with the sample size of 12. Figure 9.10b plots the discrepancy between simulated testing power and real one using these two random sampling approaches. The results suggest that the testing power of spatial-controlled random sampling is closer to the simulated one, which indicates that the use of spatial indices can mitigate the impact of spatial autocorrelation.

Note that, with only one dataset, the improvement is not significant. But when more datasets are involved, the proposed method is expected to be effective in ensuring the reliability of the proposed random sampling process in addressing the spatial autocorrelation issue.

9.4 Issues in Pavement Acceptance Testing Implementation Strategy

In Chapter 6, we have shown that, given a specific RQL, AQL and contractor's risk, agency's risk is only influenced by sample size regardless of the lot size of the pavement. Two questions arose regarding the implementation strategy: (1) Is it possible to increase the size of each lot (e.g., merge two populations) so that a larger area can be tested with the sample size? (2) Since more than one samples may be taken for each subplot (e.g., two samples are taken per subplot in flexural strength testing of PCCP), is it equivalent to treat each sample separately or average the samples of one subplot and treat the averaged one as a single sample? In the separate approach, if two samples are taken per subplot and each lot includes three sublots, the sample size would be $2 * 3 = 6$, while in the average approach, the sample size would be $1 * 3 = 3$ since the two samples of the subplot is substituted by their average. Note that in INDOT's current practice, the separate approach is used. In this section, these two questions are answered by comparing the testing power of different scenarios through simulations.

9.4.1 Influence of Merging Populations on Testing Power

Intuitively, the testing power of a population that is merged from two populations is between the two

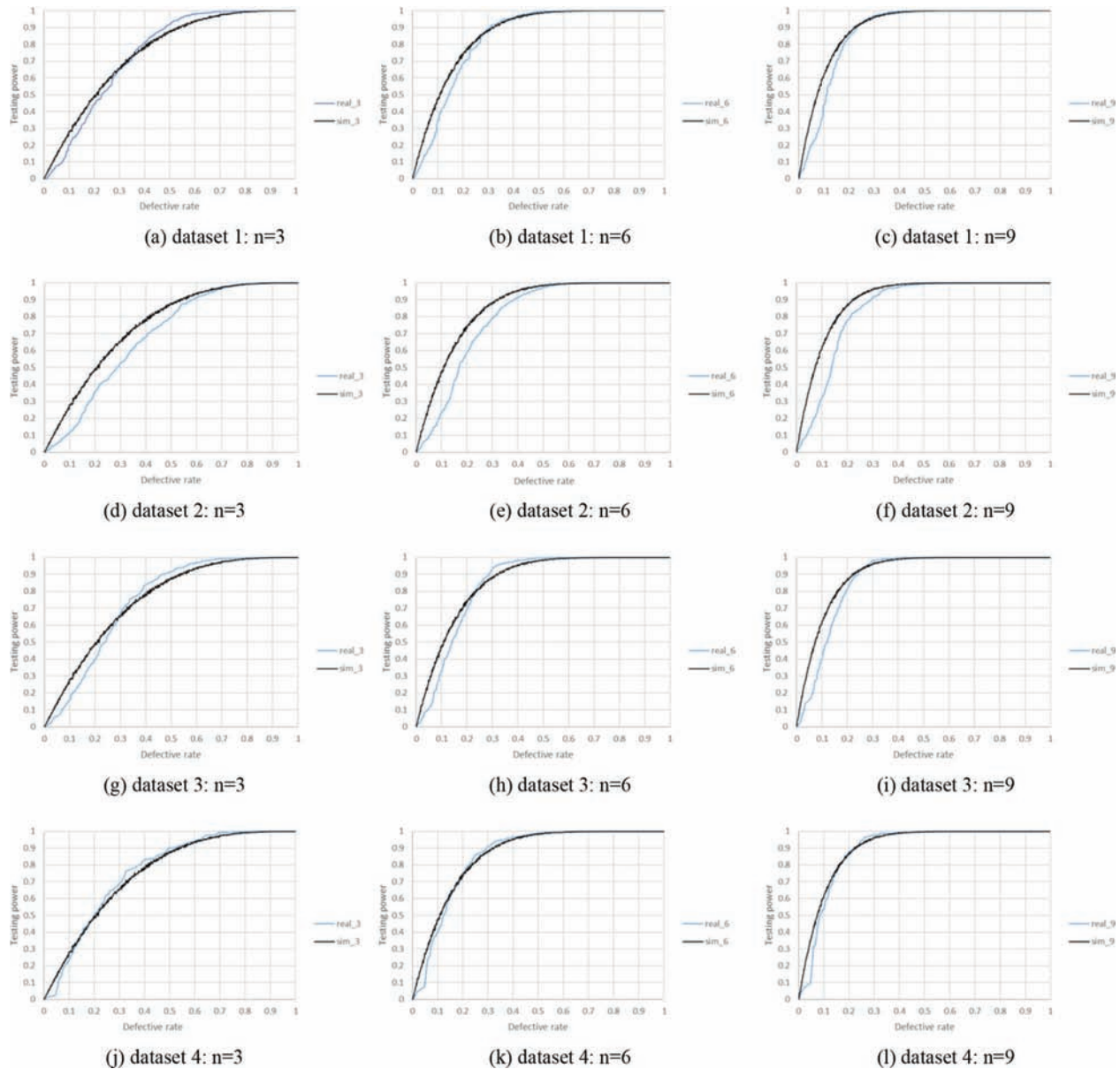


Figure 9.8 Comparison of testing power.

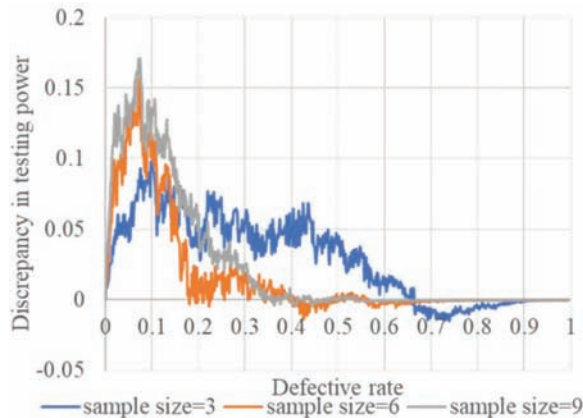


Figure 9.9 Discrepancy between simulated and real testing power in dataset 2.

individual testing powers of the two populations. In this section, the testing power for the merged population (i.e., combining two lots to form a new lot) is compared with individual testing powers obtained for separate populations. Setting RQL to be 0.3, agency's desired risk 10%, and sample size 5, five scenarios are crafted with different combinations of PWLs and the testing powers are computed. Table 9.1 illustrates the results.

The results suggest that (1) the PWL for the merged of is the average of the PWL for lot 1 and 2, and (2) by merging two lots, we compromise on the testing power, which means we may reject (or accept) both good and bad lots since the decision is made for the entire merged lot. Specifically, if the quality of one lot is much better than RQL and another lot is slightly lower than RQL

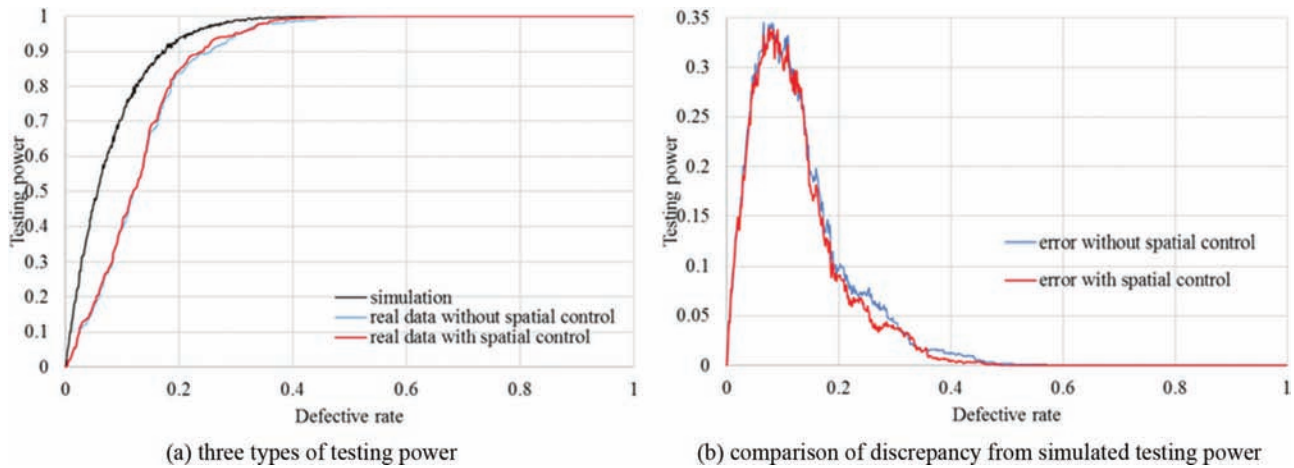


Figure 9.10 Comparison of random sampling with and without spatial control in dataset 2.

TABLE 9.1
Comparison of testing power between single lot and merged lot.

Scenario	Lot 1 (PWL) (%)	Testing power 1 (%)	Lot 2 (PWL) (%)	Testing power 2 (%)	Merged lot (PWL) (%)	Testing power (%)
1	90	56	60	96	75	85
2	90	56	80	78	85	70
3	80	78	60	96	70	90
4	80	78	50	98	65	94
5	60	96	50	98	55	97

Note: Testing power is the probability of rejecting the lot.

(i.e., Scenario 1), the merged lot might lead to lower probability of rejecting the lot with slightly rejectable quality. If the quality of one lot is much lower than RQL and another lot is slightly better than RQL (i.e., Scenario 4), the merged lot might result in higher probability of rejecting the good lot. If the quality of two lots are both acceptance or rejectable (i.e., Scenario 2 and 5), the merged lot does not have a significant influence on the acceptance decision.

9.4.2 Influence of Averaging Samples on Testing Power

In this section, the testing power obtained using all samples (referred as sampling version) is compared with that obtained using averaged samples (referred as subsampling version). Assuming there are 5 sublots in one lot and two samples are taken per subplot, the samples size of sampling version is 10, while that of subsampling version is 5. Given RQL being 0.3 and 0.5, six scenarios are considered with different levels of correlation (represented as correlation coefficient r) between two samples of one subplot, i.e., $r = 0, 0.16$, and 0.7 , where $r = 0$ indicates the samples of one subplot are independent, while $r = 0.7$ indicates the samples are highly correlated.

Figure 9.11 illustrates the testing power of sampling and subsampling. In all charts, M1 and Spk1 represent the testing power of sampling version computed using

numerical M -method and S_{pk} -method, and M2 and Spk2 represent the testing power of subsampling version. The results suggest that subsampling results in higher probability of rejection when samples are independent or slightly correlated, since their sample size is smaller than that of sampling version. When samples are highly correlated, subsampling succeeds in controlling the agency's risk at desired level of 10% at the boundary of acceptable and rejectable regions, while sampling version leads to higher risk.

9.5 Findings and Recommendations

The main findings and recommendations are as follows.

- For PCCP pavement, the simulated testing power aligns well with the real testing power when the pavement is in good quality. However, when the defective percentage of the pavement is relatively large, the simulation approach results in higher testing power compared with that obtained using real data, leading to higher agency's risk of accepting defective material. For HMA pavement, the simulated testing power matches the real one.
- The discrepancy between the simulated testing power and real one may be caused by three reasons. First, most populations are not exactly normal distributed, which violates the normality assumption in the numerical M -method. Second, samples from one subplot are correlated,

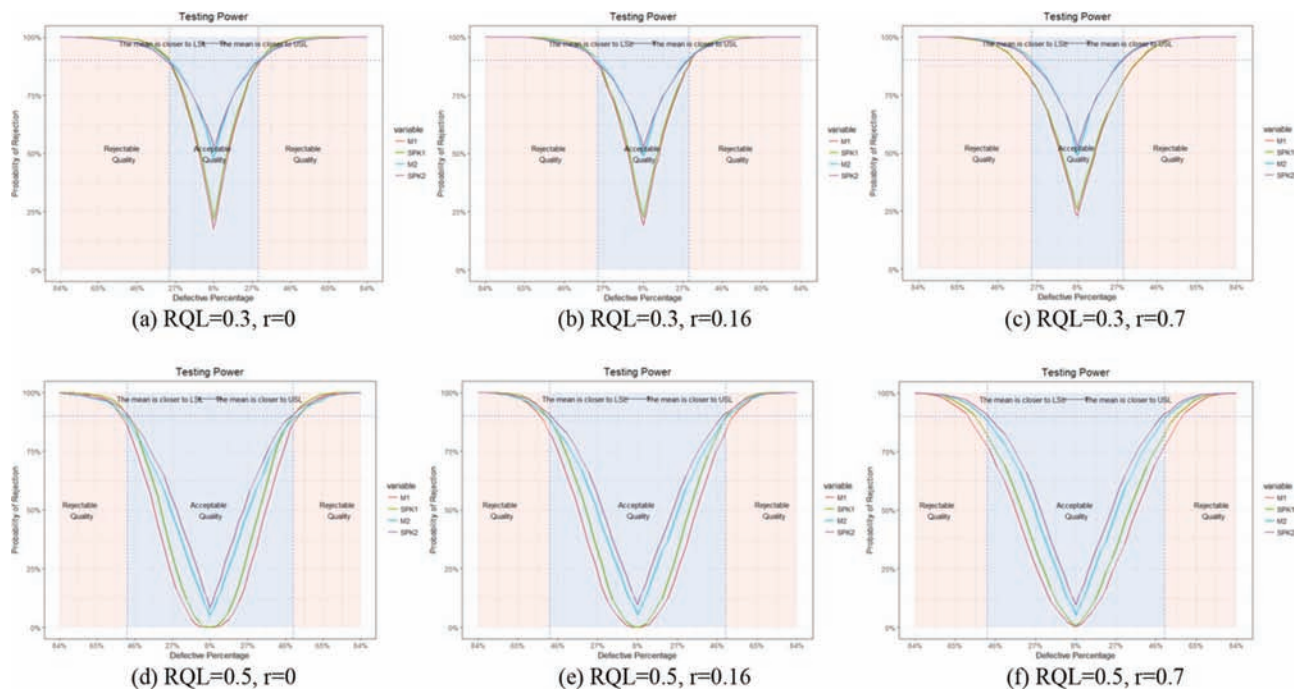


Figure 9.11 Comparison of testing power between sampling version and subsampling version.

against the independence assumption in the random sampling. Third, the testing power is computed using limited lots of the project, which may result in high variability compared with simulation approach which considers thousands of lots to compute the testing power.

- The simulated testing power for soil compaction aligns well with the real one when the defective rate of the area is relatively high. However, the existence of spatial autocorrelation has a negative impact on the testing power when the defective rate is low, leading to higher risk for the agency. It is recommended to adopt the spatial indices developed in Section 4.3 to mitigate the impact by avoiding the spatial clustering of random samples.
- Making acceptance decision based on merged population leads to compromised testing power when the quality of two populations are different. If the quality of two lots are both acceptance or rejectable, the merged lot does not have a significant influence on the acceptance decision.
- Averaging the samples in one subplot and treating it as the new sample leads to higher probability of rejecting the pavement. It ensures the reliability of acceptance testing when the samples are highly correlated.

10. IMPLEMENTATION

For practical implementation, a web tool that incorporates the newly created methods in this study has been developed based on R Shiny. It is an effective tool to facilitate the pavement and soil QA practice for the INDOT engineers with the capability to select random samples, control agency's risk in accepting defective materials, and determine optimal sample sizes for both pavement and soil acceptance testing. The web tool consists of 7 separate tabs which corresponds to

different tasks in this study. Table 10.1 describes the function of each tab and connects them to the chapters where the corresponding methods are introduced. This chapter demonstrates the use of the developed web tool, including the inputs and outputs for each tab.

10.1 Tab 1: Point Pattern Analysis—Pavement

Tab 1 assists the random sampling process and measures the spatial patterns of sample locations in pavement acceptance testing. It generates and visualizes sample locations, and compute the spatial indices of the generated random locations. Figure 10.1 illustrates the user interface, which contains four compartments: user inputs, resulting sample locations, location visualization, and spatial indices.

Figure 10.2 illustrates the user inputs, including basic information on the pavement (lot width, length of each subplot, and the start station of the pavement) and the specification regarding the random sampling (sample size per subplot and the minimum distance to the edge).

Figure 10.3 illustrates the outputs of Tab 1. Specifically, Figure 10.3a illustrates the sample locations including the longitudinal location and the offset, which are determined based on the random numbers generated from uniform distribution. The random locations are converted to station numbers given the specific start station to better facilitate the field practice. Figure 10.3b illustrates the spatial indices of generated random locations. Note that the reference range of spatial indices are listed for users to avoid the spatial clustering and high variability of the random samples, and its rationale is

TABLE 10.1
Description of the web tool.

Tab	Tab heading	Function	Corresponding chapter
Tab 1	Point spatial pattern analysis—pavement	Generate random samples Visualize sample locations Evaluate sample spatial pattern (for HMA and PCCP)	Chapter 3 and 4
Tab 2	Point spatial pattern analysis—soil	Generate random samples Visualize sample locations Evaluate sample spatial pattern (for soil)	Chapter 3 and 4
Tab 3	Testing power—pavement	Compare testing power of numerical M -method, S_{pk} -method, and current practice (PWL for HMA pavement and mean value for PCCP pavement), given a specific testing strategy	Chapter 5
Tab 4	Confidence interval	Compute the confidence interval of PWL	Chapter 5
Tab 5	Testing power—soil	Calculate testing power for different sample sizes Determine the optimal sample size for soil acceptance testing	Chapter 5 and 6
Tab 6	Optimal sample size	Determine the optimal sample size given a desired risk level for pavement acceptance testing	Chapter 6
Tab 7	Cost/benefit	Determine the optimal sample size based on cost-benefit approach for pavement acceptance testing	Chapter 7

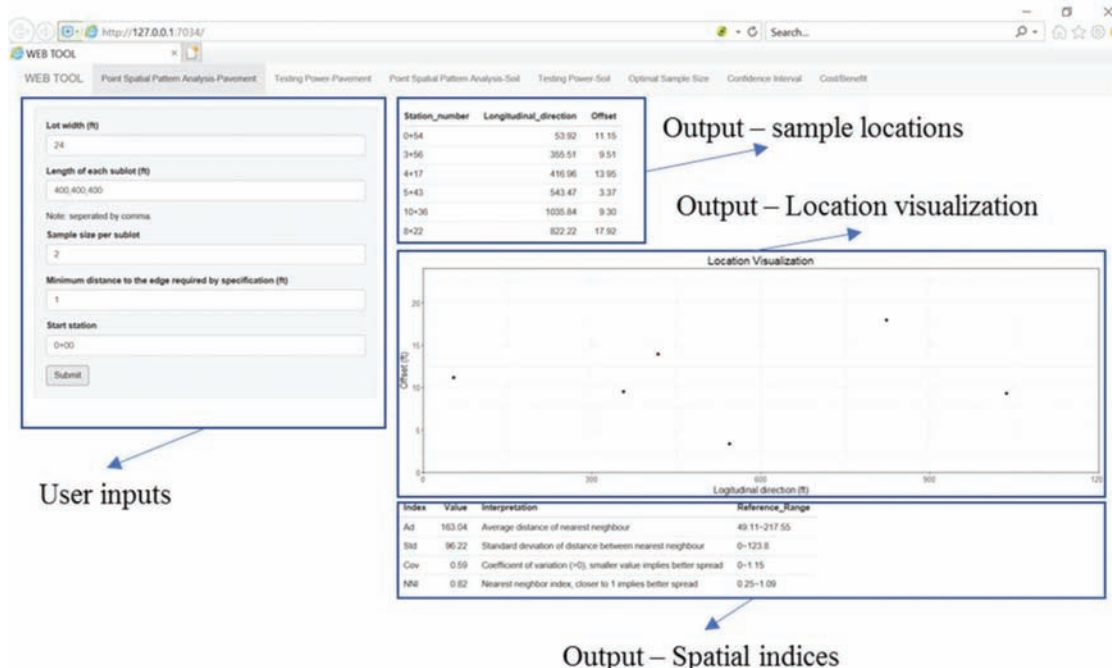


Figure 10.1 User interface of Tab 1.

detailed in Section 4.3. Figure 10.3c depicts the spread of the random samples throughout the pavement.

This tab serves as an effective tool for the field engineers to determine the locations for random samples. The spatial indices are controlled in reference ranges to avoid clustering and high variability.

10.2 Tab 2: Point Spatial Pattern Analysis—Soil

Tab 2 facilitates the random sampling for soil acceptance testing, with the same functionality as Tab 1. The only difference is that the random locations for

soil sampling are not taken based on sublots. Instead, the entire area is divided into subareas with the number equal to the sample size, and one random sample is taken for each subarea. Figure 10.4 illustrates the user interface of Tab 2, which contains four compartments: user inputs, sample locations, location visualization, and spatial indices. Figure 10.5 illustrates the user inputs, including basic information on the construction area (area length and width) and the specification regarding the soil random sampling (sample size and the minimum distance to the edge). The outputs are identical with Tab 1 (Figure 10.3).

Lot width (ft)

Length of each subplot (ft)

Note: separated by comma.

Sample size per subplot

Minimum distance to the edge required by specification (ft)

Start Station

Figure 10.2 User inputs of Tab 1.

10.3 Tab 3: Testing Power—Pavement

Tab 3 compares the testing power of pavement acceptance testing using numerical M -method, S_{pk} -method, and current practice (PWL for HMA pavement and mean value for PCCP pavement), given a specific testing strategy. Figure 10.6 illustrates the user interface of Tab 3, which contains three compartments: user inputs, definition, and comparison of testing power.

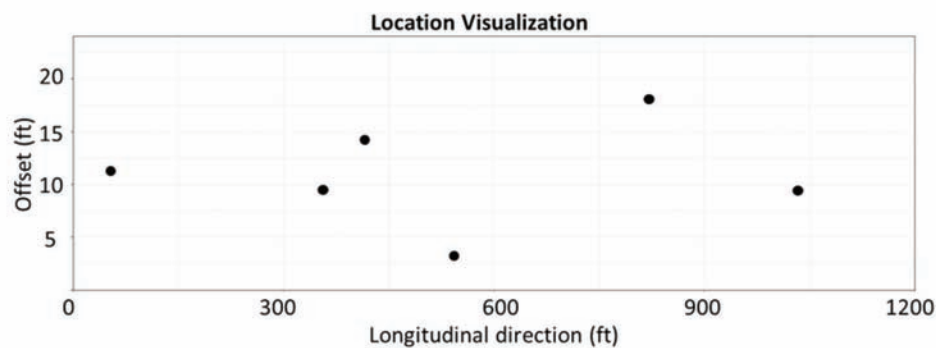
Figure 10.7 illustrates the user inputs of Tab 3, including pavement type (i.e., HMA or PCCP), specification limit type (i.e., double specification limits or single specification limit) and the corresponding specification limit(s), sample size per lot, estimated standard deviation of the pavement, agency's risk level and the rejectable defective percentage (RQL). Note that the difference between two types of pavement is that in current practice at INDOT, PWL is used as statistical

Station_number	Longitudinal_direction	Offset
0+54	53.92	11.15
3+56	355.51	9.51
4+17	416.96	13.95
5+43	543.47	3.37
10+36	1035.84	9.30
8+22	822.22	17.92

(a) Sample locations

Index	Value	Interpretation	Reference_Range
Ad	163.04	Average distance of nearest neighbour	49.11~217.55
Std	96.22	Standard deviation of distance between nearest neighbour	0~123.8
Cov	0.59	Coefficient of variation (>0), smaller value implies better spread	0~1.15
NNI	0.82	Nearest neighbor index, closer to 1 implies better spread	0.25~1.09

(b) Spatial indices



(c) Location visualization

Figure 10.3 Outputs of Tab 1.

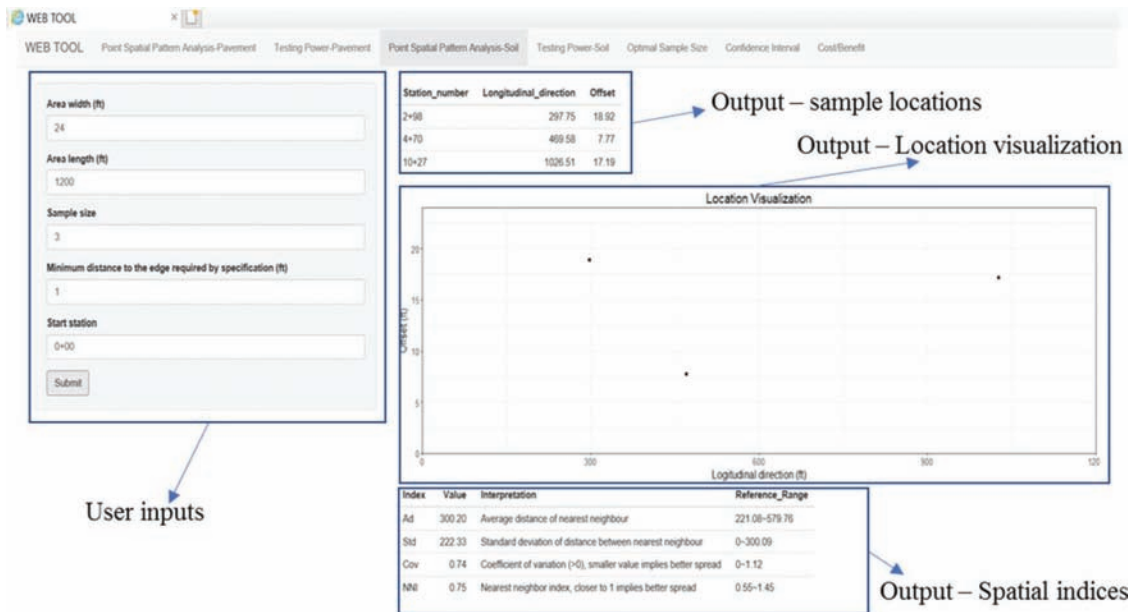


Figure 10.4 User interface of Tab 2.

Area width (ft)
24

Area length (ft)
1200

Sample size
3

Minimum distance to the edge required by specification (ft)
1

Start station
0+00

Submit

Figure 10.5 User inputs of Tab 2.

measure for HMA pavement, while mean value serves as statistical measure for PCCP pavement. Regarding the specification type, double specification limits means that both USL and LSL are specified for the specific property characteristic (e.g., air voids), whereas single specification limit means that either USL or LSL is specified for the given property characteristic (e.g., flexural strength). The compartment of definition explains three important terms, i.e., testing power, agency's risk level, and rejectable quality level. It is provided before users submit the task so that they can refer to the definitions to determine their inputs.

Figure 10.8 illustrates the output of Tab 3. The "case study" section converts the numerical inputs regarding the testing strategy into text explanation so that users

can ensure the consistency of their inputs with the desired strategy. The chart plots the testing power obtained using numerical M-method, Spk-method, and current PWL method given the specific testing strategy, which proves that the proposed method can control the agency's risk at the desired level. The "conclusion" section summarizes main findings from the chart so that users can quickly check for the testing power at the boundary of acceptable and rejectable quality.

10.4 Tab 4: Confidence Interval

Tab 4 computes the 90% confidence interval of population PWL given specific sample measurements. Figure 10.9 illustrates the user interface, which contains two compartments: user inputs and output confidence interval. The user inputs include specification limit type and the corresponding specification limit(s), and the value of each sample. Once the task is submitted, the platform computes the PWL from the input sample values based on the current practice and the 90% confidence interval based on the proposed numerical M-method. The 90% confidence interval indicates the probability of the population PWL falling in the interval is 90%. This tab proves that it is not reliable for the current practice to rely on the sample PWL to make acceptance decision as it is a random variable. The users can have a better understanding of the sampling randomness by referring to the confidence interval.

10.5 Tab 5: Testing Power-Soil

Tab 5 computes the testing power of soil acceptance testing for different sample sizes and allows the users to determine the optimal sample size for soil acceptance

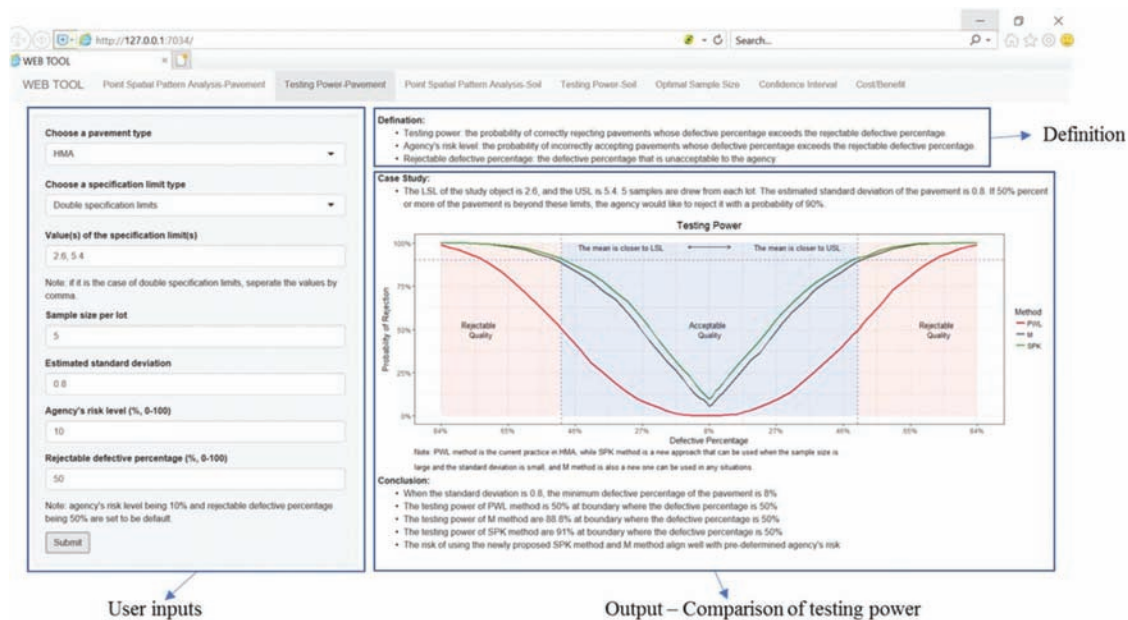


Figure 10.6 User interface of Tab 3.

Choose a pavement type
HMA

Choose a specification limit type
Double specification limits

Value(s) of the specification limit(s)
2.6,5.4
Note: if it is the case of double specification limits, separate the values by comma.

Sample size per lot
5

Estimated standard deviation
0.8

Agency's risk level (% 0-100)
10

Rejectable defective percentage (% 0-100)
50
Note: agency's risk level being 10% and rejectable defective percentage being 50% are set to be default.

Submit

Figure 10.7 User inputs of Tab 3.

testing. Figure 10.10 illustrates the user interface, which contains two compartments: user inputs, and the output of testing power for various samples. The only inputs are the minimum and maximum sample size that the user would like to consider. The probability of defecting defective area, or the testing power, with respect to defective rate (varying from 0% to 100%) is plotted for various sample sizes that are within the range of inputs. The users can assess its risk (i.e., $1 - \text{testing power}$) for

a specific sample size, as well as determine an optimal sample size given a specific desired risk level (explained in Section 6.3).

10.6 Tab 6: Optimal Sample Size

Tab 6 computes the agency's risk for different sample sizes and allows users to determine the optimal sample size according to their desired risk level in pavement acceptance testing. Figure 10.11 illustrates the user interface, which contains three compartments: user inputs, output table of agency's risk, and output charts of agency's risk.

Figure 10.12 illustrates the user inputs, including specification limit type and corresponding limits, number of subplot, estimated standard deviation, RQL, AQL, and contractor's risk. Note that users can input at most three combinations of RQL, AQL, and contractor's risk to compare the optimal sample size under different scenarios.

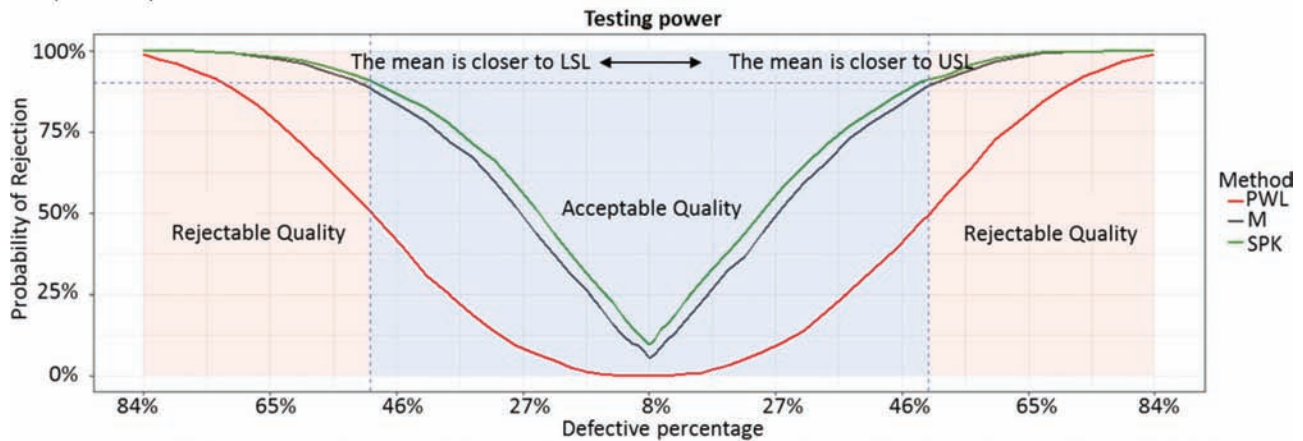
Figure 10.13 illustrates the output table, where agency's risk for different sample sizes computed using both numerical M-method and Spk-method, given the three combinations of RQL, AQL, and contractor's risk. Figure 10.14 plots the agency's risk for each scenario. Referring to the computed values of agency's risk, the user can choose the smallest sample size that achieve agency's desired risk as the optimal sample size. The three charts can be cross-compared to evaluate the optimal sample size for different settings of parameters.

10.7 Tab 7: Cost/Benefit

Tab 7 computes the total cost of pavement acceptance testing for different sample sizes based on the two cost analysis approaches proposed in Chapter 7.

Case Study:

- The LSL of the study object is 2.6, and the USL is 5.4. 5 samples are drawn from each lot. The estimated standard deviation of the pavement is 0.8. If 50% percent or more of the pavement is beyond these limits, the agency would like to reject it with a probability of 90%.



Note: PWL method is the current practice in HMA, while SPK method is a new approach that can be used when the sample size and the standard deviation is small, and M method is also a new one that can be used in any situation.

Case Study:

- When the standard deviation is 0.8, the minimum defective percentage of the pavement is 8%
- The testing power of PWL method is 50% at boundary where the defective percentage is 50%
- The testing power of M method are 88.8% at boundary where the defective percentage is 50%
- The testing power of SPK method are 91% at boundary where the defective percentage is 50%
- The risk of using the newly proposed SPK method and M method align well with the pre-determined agency's risk

Figure 10.8 Outputs of Tab 3.

The screenshot shows a 'WEB TOOL' interface with tabs for 'Point Spatial Pattern Analysis-Pavement', 'Testing Power-Pavement', 'Point Spatial Pattern Analysis-Soil', 'Testing Power-Soil', 'Optimal Sample Size', and 'Confidence Interval'. The 'Confidence Interval' tab is active. On the left, 'User inputs' are provided: 'Choose a specification limit type' (Double specification limits), 'Value(s) of the specification limit(s)' (2.6, 5.4), and 'Value of each sample' (3,4,5,6,3,5). On the right, the 'Conclusion' box displays: 'INDOT's PWL: 75.4%' and '90% confidence interval of PWL (using M method): 44.5% ~ 93.9%'. An arrow points from the 'Conclusion' box to the text 'Output – confidence interval'.

Figure 10.9 User interface of Tab 4.

It helps the users to determine the optimal sample size that minimizes the total cost. Figure 10.15 illustrates the user interface, which contains three compartments: user inputs, output table of total cost and agency's risk with respect to sample size, and the output chart of total cost with respect to sample size.

Figure 10.16 illustrates the user inputs, including the information related to the testing strategy, such as

specification limit type and corresponding limits, RQL, AQL, and contractor's risk, as well as basis information of the pavement, i.e., number of subplot, estimated mean value and standard deviation, proportion of prior lots of rejectable quality, unit price, and lot size. The user also needs to estimate the cost information, such as the testing cost and cost impact coefficient of the defective rate, which can be obtained from expert knowledge.

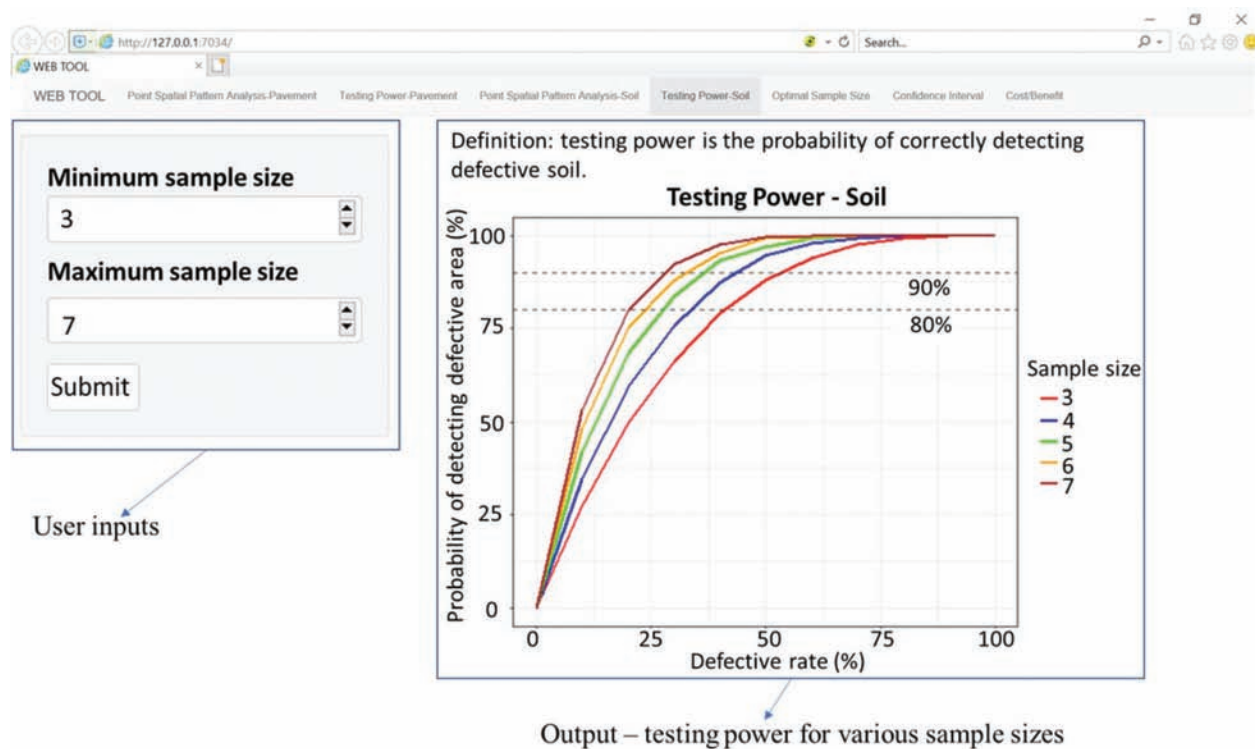


Figure 10.10 User interface of Tab 5.

The meanings of each input and their function in the cost computation are explained in Section 7.1. Once the user confirms the approach they would like to use, either Approach 1 or Approach 2, the web tool will automatically disable the inputs that is not required for the selected approach. Note that, for Approach, since the pay factor is computed using different formulas or tables for different property characteristics, this web tool only applies air voids of HMA pavement as an example.

Figure 10.17 illustrates the outputs of the Tab 7. The agency's risk and corresponding total cost for varying sample sizes are listed in the output table and the total cost is also plotted in the output chart. The users can evaluate their total cost for any sample size from the output and determine the optimal sample size by minimizing the total size. After finishing the analysis, the user needs to reset the tool to do another round of analysis.

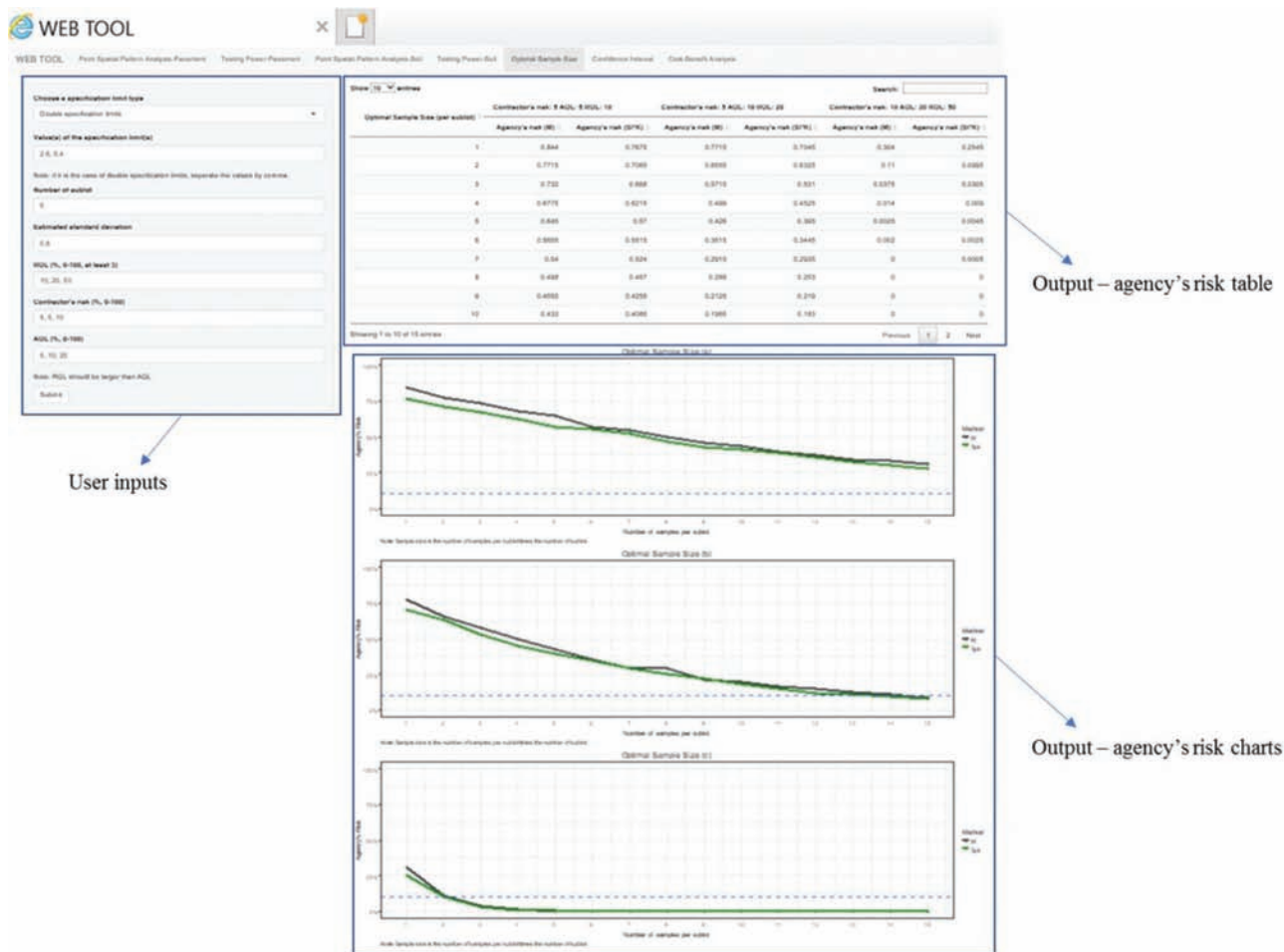


Figure 10.11 User interface of Tab 6.

Choose a specification limit type

Double specification limits

Value(s) of the specification limit(s)

2.6, 5.4

Note: if it is the case of double specification limits, separate the values by comma.

Number of subplot

5

Estimated standard deviation

0.8

RQL (% , 0-100, at least 3)

10, 20, 50

Contractor's risk (% , 0-100)

5, 5, 10

AQL (% , 0-100)

5, 10, 20

Note: RQL should be larger than AQL

Submit

Figure 10.12 User inputs of Tab 6.

Show 10 entries

Search:

Optimal Sample Size (per subplot)	Contractor's risk: 5 AQL: 5 RQL: 10		Contractor's risk: 5 AQL: 10 RQL: 20		Contractor's risk: 10 AQL: 20 RQL: 50	
	Agency's risk (M)	Agency's risk (SPK)	Agency's risk (M)	Agency's risk (SPK)	Agency's risk (M)	Agency's risk (SPK)
1	0.844	0.7675	0.7715	0.7045	0.304	0.2545
2	0.7715	0.7065	0.6555	0.6325	0.11	0.0995
3	0.732	0.668	0.5715	0.531	0.0375	0.0305
4	0.6775	0.6215	0.499	0.4525	0.014	0.009
5	0.645	0.57	0.426	0.395	0.0025	0.0045
6	0.5655	0.5515	0.3515	0.3445	0.002	0.0025
7	0.54	0.524	0.2915	0.2935	0	0.0005
8	0.498	0.467	0.288	0.253	0	0
9	0.4555	0.4255	0.2125	0.219	0	0
10	0.433	0.4085	0.1965	0.183	0	0

Showing 1 to 10 of 15 entries

Previous 1 2 Next

Figure 10.13 Output table of Tab 6.

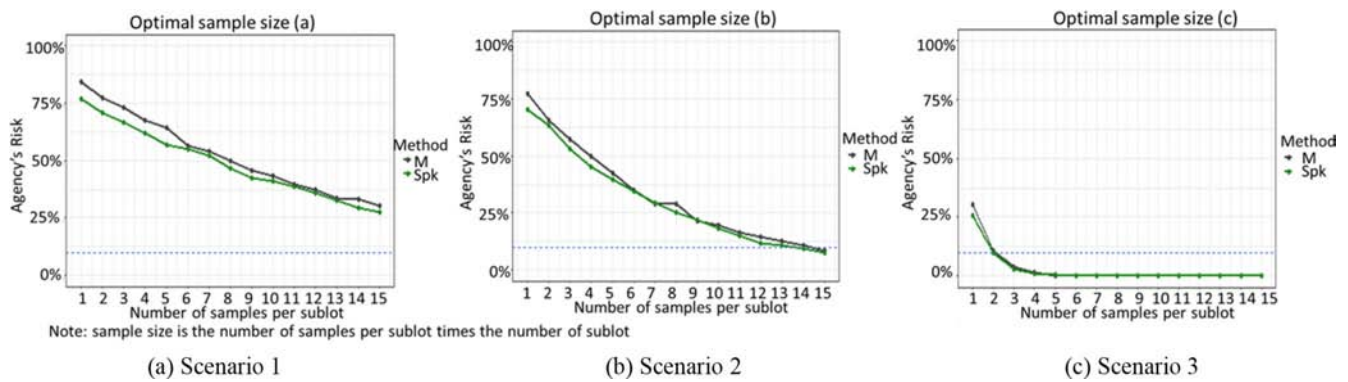


Figure 10.14 Output charts of Tab 6.

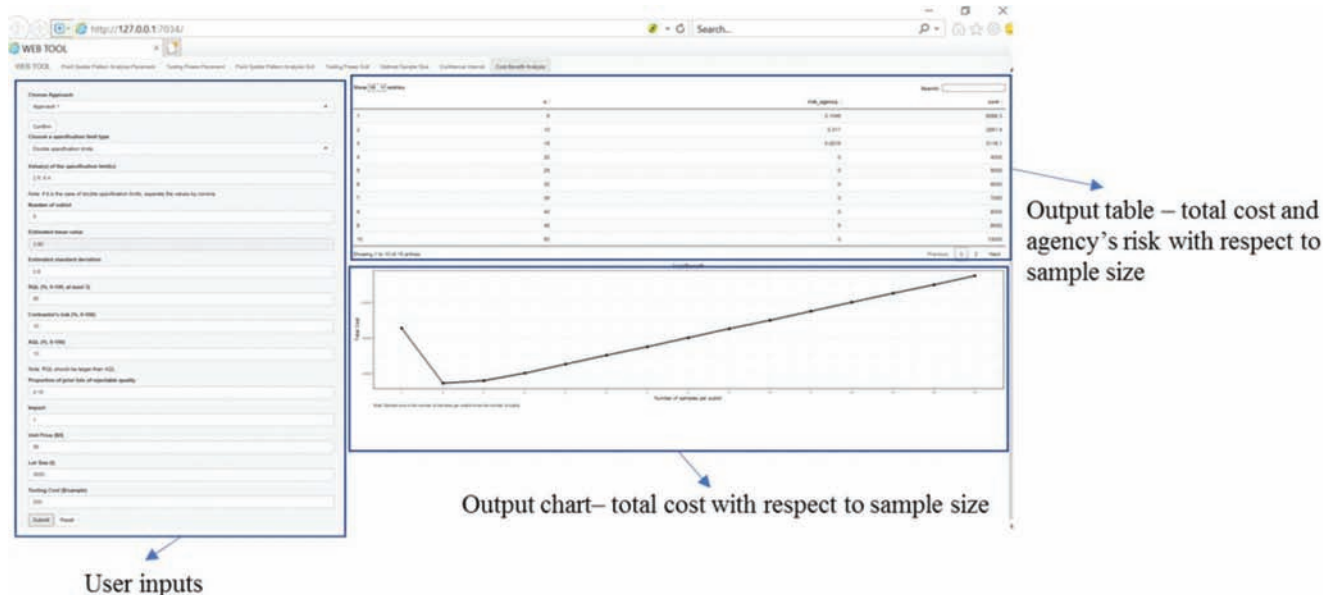


Figure 10.15 User interface of Tab 7.

Choose Approach
Approach 1

Confirm
Choose a specification limit type
Double specification limits

Value(s) of the specification limit(s)
2.6,5.4

Note: if it is the case of double specification limits, separate the values by comma.

Number of subplot
5

Estimated mean value
3.89

Estimated standard deviation
0.8

RQL (% , 0-100, at most 3)
50

Contractor's risk (% , 0-100)
10

AQL (% , 0-100)
10

Note: RQL should be larger than AQL

Proportion of prior lots of rejectable quality
0.15

Impact
1

Unit Price (\$/t)
86

Lot Size (t)
3000

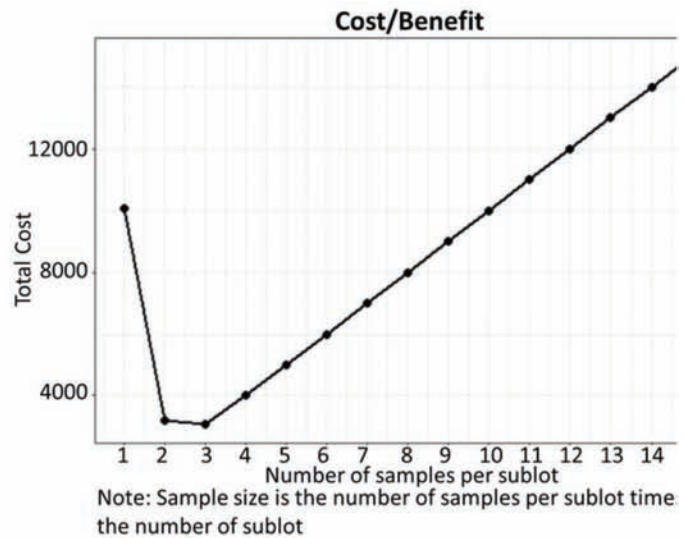
Testing Cost (\$/sample)
200

SubmitReset

Figure 10.16 User inputs of Tab 7.

Show	10	entries	Search:	
	n	risk_agency	cost	
1	5	0.117	10055.8	
2	10	0.0115	3199.7	
3	15	0.001	3077.4	
4	20	0	4000	
5	25	0	5000	
6	30	0	6000	
7	35	0	7000	
8	40	0	8000	
9	45	0	9000	
10	50	0	10000	
Showing 1 to 10 of 15 entries		Previous	1	2 Next

(a) table of total cost and agency's risk



(b) chart of total cost

Figure 10.17 Outputs of Tab 7.

11. SUMMARY AND RECOMMENDATIONS

11.1 Summary

This study addresses four critical aspects of pavement and soil acceptance testing, i.e., identifying key acceptable quality characteristics (AQC)s, selecting sample locations, designing a risk-based acceptance criteria, and determining optimal sample size. Material properties that are tested in current INDOT practice were compared with material properties that are critical to the pavement long-term performance according to MEPDG, and the missing items were identified and analyzed. The pavement property was proved to be spatially auto correlated, and two spatial pattern indices were proposed to evaluate and control the spatial pattern of random samples in order to ensure the effectiveness and reliability of random sampling with the existence of spatial autocorrelation of pavement properties. Two statistical methods: numerical M -method and S_{pk} -method, were proposed to control the agency's risk in accepting defective materials in pavement acceptance testing given a specific testing strategy, and a simulation approach was developed to measure the agency's risk in soil acceptance testing. Then, the above methods were incorporated to determine the optimal sample size for both pavement and soil acceptance testing given a desired risk level. Furthermore, two cost analysis approaches were developed to determine the optimal sample size by integrating total cost of acceptance testing and the agency's risk. Additionally, QC chart was studied as a complementary tool to ensure the consistency of the pavement quality of a project. Finally, a web tool that incorporates the newly created methods in this study was developed to assist the field practice of QA acceptance testing.

11.2 Findings and Recommendations

Key findings and recommendations have been summarized in the Executive Summary section and are repeated as follows.

- Findings and recommendations related to key material properties.
 - There are 10 hot mix asphalt (HMA) properties, 10 Portland cement concrete pavement (PCCP) properties, and 9 soil properties that are very important to the pavement performance, among which 5 in HMA, 4 in PCCP, and 6 in soil are not tested in the current INDOT practice. For the missing items in PCCP and HMA pavement, most of them do not have certified testing standards or are difficult and not warranted in laboratory testing and can be substituted by default design value. For the missing items in soil, most of them are neglected in the current INDOT specification.
 - It is recommended that (1) HMA thickness should be tested and measured directly rather than being estimated, considering its effect on pavement performance, and (2) a sampling and testing protocol shall be established for key material properties of soil that are currently not tested.
- Findings and recommendations related to random sampling.
 - ITM 802-based sampling leads to "gaps"—areas that will never be tested due to the limited random numbers in the published table. For earthwork, gap areas can receive insufficient compaction without being caught. For pavement, although gap areas are not large enough to accommodate a whole truck-load, they lead to lower probabilities for certain

trucks to be tested, compared with the true random sampling process. Such discrepancy results in a lower risk for contractors to manipulate these trucks, and a higher risk for INDOT to accept inferior materials. Therefore, the ITM 802-based sampling is less reliable than the true random sampling with random number generated from uniform distribution in real time.

- The material properties are spatially auto-correlated. Therefore, sample spatial pattern needs to be evaluated and controlled to avoid spatial clustering and ensure the effectiveness of random sampling. Two spatial pattern indices, i.e., NNI and CV , are proposed to evaluate the spatial pattern of sample locations. NNI assesses the degree of spatial dispersion and is recommended to serve as the primary metrics. CV measures the variability of sample locations and is recommended to serve the secondary metrics.
- It is recommended to use a real-time random number generator to determine sample locations, and the spatial pattern indices of random samples should be constrained to ensure the sample locations are spatially distributed and truly representative of the population.

3. Findings and recommendations related to acceptance criteria.

- The current acceptance testing methods at INDOT do not achieve the desired level of risk control.
- For pavement acceptance, the newly proposed numerical M -method and S_{pk} -method incorporate sample variance and treat statistical measures of samples as random variables. They are capable of controlling the risk at a desired level and achieving higher testing power compared to the current practice at INDOT. The numerical M -method is suitable for all scenarios and is recommended to replace the current practice, while S_{pk} -method can be applied when the sample size is relatively large to simplify the computation.
- For the soil compaction acceptance, a simulation approach is proposed to assess the risk of INDOT. It is recommended to conduct the simulation to examine the risk once a sampling strategy is determined, and then make corresponding adjustment on the sampling strategy according to the desired risk.
- Out-of-control conditions in the pavement project can be identified by matching the QC chart established from pavement QA data with typical abnormal patterns. The QC chart and the numerical M -method complement each other to ensure the quality of both individual lot and the entire project, which cannot be achieved only using current practice.

4. Findings and recommendations related to optimal sample size.

- The optimal sample size for pavement acceptance testing can be determined to satisfy the risk expectation of both agency and contractor using proposed numerical M -method and S_{pk} -method. The optimal sample size for soil acceptance can be determined as the smallest value that satisfies the desired risk level of agency using proposed simulation approach.

- In addition to the risk perspective, the optimal sample size for pavement acceptance testing can be determined using proposed two cost analysis methods by minimizing the total cost of acceptance testing. The optimal sample size and the corresponding minimal total cost are sensitive to the values of the input parameters and therefore, it is very important to objectively estimate the input parameters using historical records.
- In the case studies, the optimal sample size is larger than the specification requirements. As such, it is expected that an appropriate increase of the sample size will reduce the risk to INDOT.

11.3 Validation and Implementation

The newly developed methods in this study have been validated using real data from different INDOT projects, including PCCP project, HMA project, and soil compaction project. In general, the results of simulation study align with those obtained using real data. The discrepancy mainly stems from the correlation of samples, the spatial autocorrelation of the material property and the non-normality of the population, which is biased from the normality and independence assumption in the random sampling and acceptance testing and deserves more research efforts in the future.

For practical implementation, a web tool that incorporates the newly created methods in this study has been developed based on R Shiny. It is an effective tool to facilitate the pavement and soil QA practice for the INDOT engineers with the capability to select random samples, control agency's risk, and determine optimal sample sizes for both pavement and soil acceptance testing.

A phased implementation is suggested as follows:

1. Set up R Shiny website to implement the web tool.
2. Tool rollout and pilot testing by phases.
 - Phase I: field crew adopt the random number generator tool to locate sample location,
 - Phase II: field crew and material engineers use the risk assessment tool to determine the probability of accepting inferior products given the testing results.
 - Phase III: engineers use the web tool to determine the optimal sample size.

A clear understanding of the technical principles and a set of skills are necessary to use the web tool and interpret the results. As such, training for INDOT staff is recommended.

REFERENCES

- AASHTO. (2008). *Mechanistic-empirical pavement design guide: A manual of practice*. Washington, DC: American Association of State Highway and Transportation Officials.
- ArcGIS. (2017). How spatial autocorrelation (Global Moran's I) works. Retrieved from <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>

- Boyles, R. A. (1994). Brocess capability with asymmetric tolerances. *Communications in Statistics-Simulation and Computation*, 23(3), 615–635. <https://doi.org/10.1080/03610919408813190>
- Chang, L.-M., & Hsie, M. (1995). Developing acceptance-sampling methods for quality construction. *Journal of Construction Engineering and Management*, 121(2), 246–253. [https://doi.org/10.1061/\(ASCE\)0733-9364\(1995\)121:2\(246\)](https://doi.org/10.1061/(ASCE)0733-9364(1995)121:2(246))
- Cho, D., Najafi, F. T., & Kopac, P. A. (2011). Determining optimum acceptance sample size: Second look. *Transportation Research Record: Journal of the Transportation Research Board*, 2228(1), 61–69. <https://doi.org/10.3141/2228-08>
- Chou, Y.-H. (1997). *Exploring spatial analysis in geographic information systems*. Santa Fe, NM: OnWord Press.
- Gharaibeh, N. G., Garber, S. I., & Liu, L. (2010). Determining optimum sample size for percent-within-limits specifications. *Transportation Research Record: Journal of the Transportation Research Board*, 2151(1), 77–83. <https://doi.org/10.3141/2151-10>
- INDOT. (2014). *Quality controllquality assurance, QC/QA, soil embankment*. Indianapolis, IN: Indiana Department of Transportation. Retrieved from <http://www.intelligentcompaction.com/downloads/Spec/INDOT-Soils-IC-Spec-2014.pdf>
- INDOT. (2016). *Indiana Department of Transportation standard specifications*. Indianapolis, IN: Indiana Department of Transportation. Retrieved from <https://www.in.gov/dot/div/contracts/standards/book/sep15/2016Master.pdf>
- INDOT Office of Materials Management. (2015). *Manual for frequency of sampling and testing and basis for use of materials*. Indianapolis, IN: Indiana Department of Transportation. Retrieved from <https://www.in.gov/indot/files/FreqOfSamplingAndTestingSM.pdf>
- INDOT Office of Materials Management. (2019). *Random sampling* (ITM No. 802-16P). Indianapolis, IN: Indiana Department of Transportation. Retrieved from https://www.in.gov/indot/div/mt/itm/pubs/802_testing.pdf
- Lee, J. C., Hung, H. N., Pearn, W. L., & Kueng, T. L. (2002). On the distribution of the estimated process yield index S_{pk} . *Quality and Reliability Engineering International*, 18(2), 111–116. <https://doi.org/10.1002/qre.450>
- Lieberman, G. J., & Resnikoff, G. J. (1955). Sampling plans for inspection by variables. *Journal of the American Statistical Association*, 50(270), 457–516. <https://doi.org/10.1080/01621459.1955.10501276>
- McNeese, B. (2004). *Interpreting control charts*. Katy, TX: BPI Consulting. Retrieved from <https://www.spcforexcel.com/knowledge/control-charts-basics/interpreting-control-charts>
- Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed.). Hoboken, NJ: John Wiley & Sons.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Guidelines for optimizing the risk and cost of materials QA programs* (Research Report 838). Washington, DC: National Academies Press. <https://doi.org/10.17226/23691>
- Pearn, W. L., & Ching Cheng, Y. A. (2007). Estimating process yield based on S_{pk} for multiple samples. *International Journal of Production Research*, 45(1), 49–64. <https://doi.org/10.1080/00207540600600122>
- Pearn, W. L., Lin, G. H., & Wang, K. H. (2004). Normal approximation to the distribution of the estimated yield index S_{pk} . *Quality and Quantity: International Journal of Methodology*, 38(1), 95–111. <https://doi.org/10.1023/B:QUQU.0000013245.13104.1d>
- Pyzdek, T., & Keller, P. A. (2003). *Quality engineering handbook* (2nd ed.). Boca Raton, FL: CRC Press.
- Schwartz, C. W., Li, R., Kim, S., Ceylan, H., & Gopalakrishnan, K. (2011). Sensitivity evaluation of MEPDG performance prediction. *InTrans Project Reports*, 62. Retrieved from https://lib.dr.iastate.edu/intrans_reports/62
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>
- Stapenhurst, T. (2013). *Mastering statistical process control*. Abingdon, UK: Routledge.
- Willenbrock, J. H. (1976). *A manual for statistical quality control of highway construction* (Vols. 1–3). Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

About the Joint Transportation Research Program (JTRP)

On March 11, 1937, the Indiana Legislature passed an act which authorized the Indiana State Highway Commission to cooperate with and assist Purdue University in developing the best methods of improving and maintaining the highways of the state and the respective counties thereof. That collaborative effort was called the Joint Highway Research Project (JHRP). In 1997 the collaborative venture was renamed as the Joint Transportation Research Program (JTRP) to reflect the state and national efforts to integrate the management and operation of various transportation modes.

The first studies of JHRP were concerned with Test Road No. 1 — evaluation of the weathering characteristics of stabilized materials. After World War II, the JHRP program grew substantially and was regularly producing technical reports. Over 1,600 technical reports are now available, published as part of the JHRP and subsequently JTRP collaborative venture between Purdue University and what is now the Indiana Department of Transportation.

Free online access to all reports is provided through a unique collaboration between JTRP and Purdue Libraries. These are available at <http://docs.lib.purdue.edu/jtrp>.

Further information about JTRP and its current research program is available at <http://www.purdue.edu/jtrp>.

About This Report

An open access version of this publication is available online. See the URL in the recommended citation below.

Cai, J., Gao, Q., Chun, H., & Cai, H. (2019). *Pavement acceptance testing: Risk-controlled sampling strategy* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2019/08). West Lafayette, IN: Purdue University. <https://doi.org/10.5703/1288284316918>