

ABSTRACT

Title of Dissertation: SUSTAINABILITY, ACCEPTANCE RISK ANALYSIS
AND MACHINE LEARNING IN ASSESSING
MECHANICAL PROPERTIES AND THE IMPACT OF
HIGHWAY MATERIALS IN TRANSPORTATION
INFRASTRUCTURE

Yunpeng Zhao, Doctor of Philosophy, 2023

Dissertation directed by: Associate Professor, Dimitrios Goulias, Department of Civil
and Environmental Engineering

Improving the performance and extending the service life of transportation infrastructure is a long standing goal of Federal Highway Administration (FHWA) and the transportation community. Accurate prediction of the mechanical properties of highway materials are indispensable for enhancing the sustainability and resilience of transportation infrastructure since it provides accurate inputs for pavement mechanistic-empirical (ME) design and prediction of pavement distresses, helping to optimally allocate the maintenance needs and reduce testing frequencies which account for costly expenditures. Accurate prediction of materials properties can also reduce the acceptance risks during quality

assurance (QA) without conducting extensive testing. Concrete plays an important role in the construction of transportation infrastructure. Developing an empirical and/or statistical model for accurately predicting compressive strength remains challenging and requires extensive experimental work. Thus, the objective of the study was to improve the prediction of concrete compressive strength using ML algorithms. A ML pipeline was proposed in which a two-layer stacked model was developed by combining seven individual ML models. Feature engineering was implemented, and feature importance was evaluated to provide better interpretability of the data and the model. This study promotes a more thorough assessment of alternative ML algorithms for predicting material properties.

In addition, the quality of highway materials and construction translate directly to performance. To develop a statistically sound QA specification, the risks to the agency and contractor must be well understood. In this study, a Monte Carlo simulation model was developed to systematically assess the acceptance risks and the implications on pay factors (PF). The simulation was conducted using typical acceptance quality characteristics (AQC)s, such as strength, for Portland cement (PCC) pavements. The analysis indicated that specific combinations of contractor and agency sample sizes and population characteristics have a greater impact on acceptance risks and may provide inconsistent PF. The proposed methodology aids both agencies and producers to better understand and evaluate the impact of sample sizes and population characteristics on the acceptance risks and PF.

Finally, the use of recycled materials is a key element in generating sustainable pavement designs to save natural resources, reduce energy, greenhouse gas (GHG) emissions and

costs. This study proposed a methodological life cycle assessment (LCA) framework to quantify the environmental and economic impacts of using recycled materials in pavement construction and rehabilitation. The LCA was conducted on two roadway projects with innovative recycled materials, such as construction and demolition waste (CDW) and rock dust. The proposed LCA framework can be used elsewhere to quantify the environmental and economic benefits of using recycled materials in pavements.

SUSTAINABILITY, ACCEPTANCE RISK ANALYSIS AND MACHINE
LEARNING IN ASSESSING MECHANICAL PROPERTIES AND THE IMPACT OF
HIGHWAY MATERIALS IN TRANSPORTION INFRASTRUCTURE

By

Yunpeng Zhao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements of the degree of
Doctor of Philosophy
2023

Advisory Committee:

Professor Dimitrios G. Goulias, Chair

Professor Sherif M. Aggour

Professor Ahmet H. Aydilek

Professor Miroslaw J. Skibniewski

Professor Sung Lee, Dean's Representative

© Copyright by
Yunpeng Zhao
2023

ACKNOWLEDGEMENTS

There are no proper words to convey my deep gratitude and respect for supervisor, Professor Dimitrios G. Goulias. I appreciate your tremendous support, patience and inspiration during my challenging moments. Your immense knowledge and guidance helped me develop my research and write my dissertation.

My sincere thanks must also go to the members of my dissertation committee: Professors Sherif M. Aggour, Ahmet H. Aydilek, Mirosław J. Skibniewski and Sung Lee, for their time to be on my committee and to offer me valuable comments and advice towards improving my work.

I am also grateful to my collaborators for lending me their expertise and intuition to my research: Professors Luca Tefa and Marco Bassani from Polytechnic University of Turin, Italy; Professor Magdalena Dobiszewska from Bydgoszcz University of Science, Poland.

I am extremely grateful to my parents, Limin Zhang, Jianwei Zhao, and my beloved Shih-Huai Cheng for their unconditional trust and support during all my ups and downs. Your love gives me strength and courage.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	vi
LIST OF TABLES	viii
CHAPTER 1: INTRODUCTION	1
1.1 Overview	1
1.2 Research Objectives	6
1.3 Organization of Dissertation	7
CHAPTER 2: PREDICTION OF CONCRETE COMPRESSIVE STRENGTH USING MACHINE LEARNING ALGORITHMS	9
2.1 Background	9
2.2 Machine Learning Algorithms	11
2.2.1 <i>Multiple Linear Regression (MLR)</i>	12
2.2.2 <i>Support Vector Machines (SVMs)</i>	13
2.2.3 <i>Decision Tree (DT)</i>	15
2.2.4 <i>Artificial Neural Networks (ANNs)</i>	16
2.2.5 <i>Ensemble Learning (EL)</i>	17
2.3 Methodology and Implementation	19
2.4 Data Preparation.....	20
2.5 Training Base Models	23
2.6 Training Stacked Model and Generating Predictions.....	24
2.7 Evaluation of Model Performances.....	26
2.7.1 <i>Performance Measures</i>	26
2.7.2 <i>Performance of Base Models</i>	27
2.7.3 <i>Performance of Stacked Model</i>	30
2.7.4 <i>Feature Importance</i>	33
2.7 Comparison with Models from Previous Studies.....	39
2.8 Summary and Conclusions.....	42

CHAPTER 3: EVALUATION OF ACCEPTANCE RISK AND PAY FACTOR IN QUALITY ASSURANCE BY VERIFICATION TESTING	44
3.1 Background	44
3.2 Literature Review	48
3.3 Power of Hypothesis Testing and Acceptance Risks	50
3.4 Numerical Simulation	54
3.2.1 <i>Simulation of AASHTO Verification Procedure</i>	54
3.2.2 <i>Determination of Pay Factor (PF)</i>	56
3.2.3 <i>Power of Statistical tests and PF Probability Curves</i>	57
3.3 Alternative Population Characteristics and Sample Sizes.....	60
3.4 Results and Discussions	63
3.5 Conclusions and Recommendations.....	81
CHAPTER 4: LIFE CYCLE ASSESSMENT OF USING RECYCLED MATERIALS IN PAVEMENT	84
4.1 Introduction	84
4.2 Methodology	85
4.3 Life-Cycle Economic and Environmental Impacts of CDW Recycled Aggregates in Roadway Construction and Rehabilitation.....	90
4.3.1 <i>Literature Review</i>	90
4.3.2 <i>Alternative Pavement Design with CDW Aggregates</i>	93
4.3.3 <i>Life Cycle Assessment (LCA)</i>	97
4.3.4 <i>Life Cycle Cost Analysis</i>	99
4.3.4 <i>Life Cycle Environmental Impacts</i>	100
4.3.5 <i>Sustainability Criteria and Rating</i>	106
4.3.6 <i>Conclusions</i>	110
4.4 Life Cycle Sustainability Assessment of Using Rock Dust as a Partial Replacement of Fine Aggregate and Cement in Concrete Pavements.....	112
4.4.1 <i>Background</i>	112
4.4.2 <i>Materials and Methods</i>	115
4.4.3 <i>Feasible Sustainable Strategies with Rock Dust Addition in Concrete</i>	118
4.4.3 <i>Results and Discussions</i>	122
4.4.4 <i>Sustainability Rating</i>	133

<i>4.4.5 Summary and Conclusions</i>	135
CHAPTER 5: SUMMARY AND RECOMMENDATIONS	137
APPENDIX	141
BIBLIOGRAPHY	142

LIST OF FIGURES

Figure 2.1 Nonlinear transformation.....	14
Figure 2.2 Flow diagram of the proposed methodology	20
Figure 2.3 Relationship between predicted and measured compressive strength: a) linear regression, b) support vector machine, c) decision tree, d) multiple layer perceptron, e) random forest, f) Xgboost, g) Adaboost.....	30
Figure 2.4 Predicted vs. measured strength for stacked model, a) on training dataset, b) on testing dataset	33
Figure 2.5 Relative importance of input variables for each model.....	39
Figure 3.1 Methodological framework of analysis	47
Figure 3.2 Graphical illustration of level of significance, α , β and hypothesis testing power for two-sided case.....	52
Figure 3.3 Monte Carlo based simulation for integrating risk and PF in verification procedures.....	59
Figure 3.4 Alternative distribution scenarios in numerical simulation for pavement thickness	63
Figure 3.5 F-test power for scenario 1 (equal sample and population characteristics).....	65
Figure 3.6 PF probability curves for scenario 1 (equal populations means and standard deviations) with various combinations of sample sizes, n.....	68
Figure 3.7 F-test power for scenario 2 for various combinations of contractor's and agency's sample sizes.....	71
Figure 3.8 PF probability curves for scenario 2 for various combinations of contractor's and agency's sample sizes	71
Figure 3.9 F-test power for scenario 3 for various combinations of contractor's and agency's sample sizes.....	73
Figure 3.10 PF probability curves for scenario 3 for various combinations of contractor's and agency's sample sizes	73
Figure 3.11 t-test power for scenario 4 for various combinations of contractor's and agency's sample sizes	76
Figure 3.12 PF probability curves for scenario 4 for various combinations of contractor's and agency's sample sizes	76
Figure 3.13 F-test power for scenario 5 for various combinations of contractor's and agency's sample sizes	79

Figure 3.14 T-test power for scenario 5 for various combinations of contractor's and agency's sample sizes	79
Figure 3.15 Probability of detecting a difference in means and/or variances for scenario 5 for various combinations of contractor's and agency's sample sizes	81
Figure 3.16 PF probability curves for scenario 5 for various combinations of contractor's and agency's sample sizes	81
Figure 4.1 Methodological framework for generating and evaluating alternative sustainable strategies	87
Figure 4.2 Inputs, outputs of the LCCA and LCA environmental impact analysis	89
Figure 4.3 Schematic representation of alternative strategies and materials; (A) reference design, (B) (C), (D) and (E): alternative strategies.....	96
Figure 4.4 LCCA for alternative strategies. (A) reference design; (B) (C), (D) and (E): alternative strategies	100
Figure 4.5 Life cycle greenhouse gas emissions (CO ₂); (A) reference design, (B) (C), (D) and (E): alternative strategies	102
Figure 4.6 Life cycle energy consumption; (A) reference design, (B) (C), (D) and (E): alternative strategies	103
Figure 4.7 Life cycle water consumption; (A) reference design, (B) (C), (D) and (E): alternative strategies	104
Figure 4.8 Environmental impacts of alternative strategies; (A) reference design, (B) (C), (D) and (E): alternative strategies.....	106
Figure 4.9 Relative weights for sustainability criteria	108
Figure 4.10 Amoeba graphs for strategies C and E	110
Figure 4.11 Comparison of life-cycle costs for alternative strategies.....	125
Figure 4.12 NPV life-cycle cost broken down by materials and processes for sustainable strategy F.....	126
Figure 4.13 Life-cycle greenhouse gas emissions (CO ₂) for alternative strategies.....	129
Figure 4.14 Life-cycle energy consumption; reference design (A), alternative strategies (B–G)	130
Figure 4.15 Life-cycle water consumption; reference design (A), alternative strategies (B–G)	131
Figure 4.16 Sustainability rating for each alternative	134

LIST OF TABLES

Table 2.1 Statistics of concrete mixtures variables.....	22
Table 2.2 ML models with optimized hyperparameters	24
Table 2.3 Comparison of ML model performance on the HPC dataset.....	41
Table 3.1 Means, standard deviations and specification limits (NCHRP 10-79)	62
Table 3.2 Power and average PF in the long run for various combinations of contractor and agency sample sizes	65
Table 4.1 Design features of the paving construction project.....	94
Table 4.2 Alternative Materials and Properties.....	96
Table 4.3 Materials costs.....	98
Table 4.4 Labor, processing cost and overhead rates (PaLATE).....	98
Table 4.5 Environmental factors related to materials production (PaLATE and OpenLCA)	101
Table 4.6 Environmental impacts for alternative strategies.....	105
Table 4.7 Criteria and sustainability targets.....	108
Table 4.8 Points obtained for each parameter and total rating score	110
Table 4.9 Chemical composition of cement.....	116
Table 4.10 Chemical composition of rock dust (basalt origin).....	116
Table 4.11 Concrete mix proportioning	117
Table 4.12 Conventional and alternative sustainable strategies.....	119
Table 4.13 Pavement design input parameters for the roadway site.....	121
Table 4.14 Cost of materials in the region	122
Table 4.15 Emission factors of materials production (after EPA 2022)	127
Table 4.16 Environmental impacts broken down by material production, materials and transportation for strategy F	131
Table 4.17 Environmental impacts for alternative sustainable strategies	132

CHAPTER 1: INTRODUCTION

1.1 Overview

The nation's vast network of road transportation infrastructure is critical to sustain economic development, connect communities and improve the quality of life. Enhancing the durability and extending the life of transportation infrastructure is a strategic priority of the US Department of Transportation (DOT). Accurate estimation and/or prediction of the mechanical properties of transport infrastructure materials is one of the key elements in pavement management systems (PMS) and bridge management systems (BMS). Accurate predictions of materials' properties also provide valuable inputs for pavement mechanistic-empirical (ME) design and prediction of pavement service life. This leads in improving the optimal timing of maintenance and rehabilitation interventions, thus reducing testing frequencies of inspections which account for costly expenditures. Accurate predictions of material properties and strength during mix design and construction may also provide: *(i)* significant savings to producers by reducing testing during mixture design optimization, as well as during placement and construction for assessing material and construction uniformity; *(ii)* lower risks of acceptance during quality assurance (QA); *(iii)* fair and defensible pay factors leading to lower litigations during construction; and finally, *(iv)* more reliable and trustworthy sustainability assessment of roadway infrastructure construction in terms of economic and environmental impact assessment.

Concrete is the most widely used construction material, with an estimated 30 billion tonnes used each year worldwide, and with: a per capita basis that is 3 times as much as 40 years ago (i.e., three tonnes per year used for every person in the world); as well as, a demand growing more steeply than any other construction material such as steel or wood, according to the Global Cement and Concrete Association (GCCA, 2022). Thus, concrete plays an important role in the construction of transport infrastructures such as highways, bridges, airports, and runways. Compressive strength is one of the most critical parameters to assess concrete quality in engineering applications (ACI 318-19); (Kosmatka & Panarese, 2002). To meet compressive strength requirements, concrete mixture proportioning is often based on empirical prescriptive and/or performance-related mixture design methodologies as recommended by the American Concrete Institute (ACI) (ACI 318-19). A concrete mixture consists of various ingredients including cement, water, coarse and fine aggregate, and in several cases additives and admixtures. Understanding the relationship between concrete ingredients and strength is essential for optimizing concrete mixture proportioning and predicting early and long-term compressive strength.

Traditionally, empirical and statistical models, such as linear and nonlinear regression, were developed for predicting concrete compressive strength. However, these models require extensive experimental work and statistical analysis of the data. In addition, since the relationships between ingredients and compressive strength are often complex and highly nonlinear, the conventional models often provide inaccurate results. Therefore, developing a comprehensive model for accurately predicting compressive strength is challenging. In recent years, there has been growing interest in the application of machine

learning (ML) techniques for predicting materials properties. ML provides a data-driven approach that is capable of making predictions based on existing datasets and underlying patterns. ML-based predictions have a significant advantage over the traditional approaches especially for handling nonlinear problems (Harun, 2022); (Miladirad et al., 2021); (Farooq et al., 2021); (Chou et al., 2011); (Young et al., 2019); (Salehi & Burgueño, 2018); (Yeh, 1998). ML algorithms consider the type and quantities of concrete ingredients as input variables to predict compressive strength (i.e., output/target variables). ML algorithms can learn the relationship between the target and input variables without constraints on presumption. This approach also further provides greater flexibility to capture hidden, non-intuitive feature patterns directly from the input data.

This study attempts to improve the prediction of concrete compressive strength using ML algorithms with feature engineering techniques, (Research Objective I). Seven ML models (i.e., base models) of increasing complexity are implemented and compared, including linear regression (LR), support vector machine (SVM), decision tree (DT), multiple layer perceptron (MLP), which is a class of feedforward ANN), and ensemble models (i.e., random forest (RF), Adaboost, and Xgboost). Feature (i.e., input variable) importance is computed to demonstrate the contribution of the synthetic features on prediction accuracy as compared to original features. To further improve the prediction accuracy, this study develops a two-layer stacked model where individually developed base ML models are combined, (Research Objective II). Finally, the model performances from this study are compared with previous studies from the literature to demonstrate the superiority of the proposed methodology.

Quality assurance (QA) plays an important role in delivering long-life pavements. The quality of highway construction and materials translate directly to performance. To develop a statistically sound QA specification, the risks to the agency and contractor must be well understood. The payment provisions and risks associated with sample sizes, specification limits, acceptance quality level (AQL), rejectable quality level (RQL) and pay equations must be fully assessed and understood by the producers and agencies. Modern QA specifications recognize the state highway agencies' (SHAs) responsibility for monitoring the contractor's quality control (QC) activities, conducting agency inspections, and conducting acceptance sampling and testing. However, because of the shortage of personnel at SHAs and intensive construction schedules, many SHAs use the contractor QC data for acceptance and pay decisions. This acceptance procedure is permitted by Federal regulation 23 CFR 637 Subpart B as long as the quality of the material or construction is validated by verification testing using independent samples (FHWA, 1995).

The F- and t- tests are the most used verification procedures in validating contractor test data to determine if the SHA and contractor data are from the same population. These tests are based on sample statistics, which may lead to inadequate verification and statistical validation, and subsequently impose significant acceptance and payment risks to SHAs. Thus, the third objective of this study (Research Objective III) is to develop a Monte Carlo based simulation process to systematically quantify the acceptance risks associated with the verification procedures and assess the implications on pay factor (PF). The simulation is implemented on selected acceptance quality characteristics (AQC), such as strength, thickness, and roughness of Portland cement concrete (PCC) pavement. The statistical power

of the F-test and t-test is determined for various combinations of contractor and agency sample sizes. The acceptance risks and associated PF are quantified by defining the PF probability curves and calculating the average PFs in the long run.

Finally, state department of transportation agencies (DOTs) are now focusing on the implementation of sustainable criteria and practices for “green infrastructure.” To improve the environmental quality and sustainable development of transportation infrastructure, it is important to implement sustainable strategies in pavement construction and rehabilitation. Recently, FHWA launched the Climate Challenge Initiative to quantify the impacts of sustainable pavements and to demonstrate ways to reduce greenhouse gas emissions in highway projects using sustainable construction materials (e.g., recycled materials and by-products). The use of recycled materials is a key element in generating sustainable pavement designs to save natural resources, reduce energy, greenhouse gas emissions, and costs. The fourth objective of this study, (Research Objective IV), is to propose a methodology for systematically assessing the environmental and economic life-cycle benefits when using recycled materials in highway projects. The suggested approach could be potentially implemented and/or integrated in PMS so as to introduce sustainability principles in designing and optimizing alternative rehabilitation strategies. The methodology includes various steps for the analysis, starting with condition assessment of the existing highway, identifying alternative sustainable structural pavement designs, predicting service life, setting up alternative rehabilitation strategies, and conducting life cycle environmental and economic analysis. To demonstrate the value of the methodology worldwide, two case studies are conducted on roadway projects with innovative recycled materials, such as

construction and demolition waste (CDW) and rock dust filler, representing actual field conditions for primary roads in Italy and Poland, complementing thus the previous analysis in the US conducted during the master's degree (Zhao et al., 2021).

1.2 Research Objectives

Thus, the objectives of this dissertation were to:

- Explore the development of ML modeling and the implementation of feature engineering techniques using domain knowledge to improve the prediction of concrete compressive strength, (Research Objective I).
- Enhance ML prediction accuracy by exploring model stacking where individual ML algorithms are combined, (Research Objective II).
- Develop a systematic methodology for quantifying acceptance risks based on Monte Carlo simulation and relate risks to pay rewards for quality, (Research Objective III).
- Propose a methodology for systematically assessing the environmental and economic life-cycle benefits when using recycled materials in highway projects, (Research Objective IV), and assessing response of the suggested approach to real case worldwide real case projects when innovative recycled construction materials are used.

1.3 Organization of Dissertation

Chapter 1 provides an overview of the dissertation and outlines the research objectives.

Chapter 2 presents the development of ML models for predicting concrete compressive strength (Research Objectives I and II). In this chapter, seven alternative ML models of increasing complexity are developed and compared, including linear regression (LR), support vector machine (SVM), decision tree (DT), multiple layer perceptron (MLP), which is a class of feedforward ANN, random forest (RF), Adaboost and Xgboost. To further improve the prediction accuracy, a ML pipeline is proposed in which the feature engineering technique is implemented, and a two-layer stacked model was developed. The implementation process of the proposed ML pipeline includes the following steps/sections: (1) data preparation; (2) training and optimizing individual models; (3) training the stacked model and generating predictions; (4) evaluation of model performance.

Chapter 3 presents the Monte Carlo simulation model which is developed to systematically quantify acceptance risks and assess the implications on PF (Research Objective III). The model is implemented for five scenarios of contractor and agency population characteristics using pavement thickness data reported in a national study. The numerical simulation process and results are presented. For each scenario, the acceptance risks are assessed and the implications in terms of PF to the contractor and the agency are discussed.

Chapter 4 presents a life cycle assessment (LCA) framework for using recycled materials in pavement construction and rehabilitation (Research Objective IV). The LCA is conducted on two pavement projects representative of typical construction practices in Italy and Poland to quantify the environmental and economic impacts.

Chapter 5 summarizes the research findings, provides recommendations and discusses future research.

CHAPTER 2: PREDICTION OF CONCRETE COMPRESSIVE STRENGTH USING MACHINE LEARNING ALGORITHMS

2.1 Background

The most common ML models used for forecasting the compressive strength of concrete can be categorized into four types: SVM, DT, artificial neural networks (ANN), and ensemble learning algorithms (EL). Numerous studies demonstrated that ML techniques provide higher accuracy for concrete strength prediction than traditional statistical analysis (i.e., multivariate regression). For instance, a neural network with backpropagation was adapted to predict the compressive strength of high-performance concrete (HPC) (Yeh, 1998). The model incorporated concrete mixture ingredients and age as input variables. Results demonstrated that ANN exhibited good prediction performance, outperforming regression models in terms of accuracy. Several other studies have also reported the use of ANN in predicting concrete strength. For example, Duan et al., (2013) proposed an ANN model to predict the compressive strength of recycled aggregate concrete (RAC). The performance assessment revealed that ANN offered a fairly high accuracy with a coefficient of determination (R^2) of 0.995. However, this study is limited to an extremely small dataset (i.e., $n=168$) which may not sufficiently represent the predictor variable space. As such, the generalization ability of ANN needs to be validated. Siddique et al., (2011) compared the predictive performance of a simple backpropagation neural network on two concrete compressive strength databases. Their study also presented the relative importance of each input variable in prediction. Gupta, (2007) investigated the potential use of SVM for

predicting concrete compressive strength using two relatively small datasets (i.e., $n=181$ and 190). The results showed that SVM with radial basis function (RBF) can effectively predict compressive strength and provide a correlation coefficient of 0.994 . Ling et al., (2019) proposed an SVM model combined with a k-fold cross-validation technique to predict the compressive strength of concrete in the marine environment. The results showed that the ANN model outperformed ANN and DT. Chou et al., (2011) attempted to optimize the prediction accuracy of compressive strength of HPC by comparing different ML models, including multiple additive regression trees (MART), SVM, bagging, and ANN. The performance comparison indicated MART that was superior in prediction accuracy, computational efficiency, and aversion to overfitting. Young et al., (2019) compared the performance of four ML models (i.e., linear regression, ANNs, SVM, and random forest) for predicting both field and laboratory concrete data. In their study, the random forest provided the best prediction for both field ($R^2=0.60$) and laboratory data ($R^2=0.86$). Feng et al., (2020) proposed an ensemble ML model (i.e., AdaBoost) that employed an adaptive boosting algorithm to establish a strong learner by integrating several weaker learners and reported promising accuracy. Zhang et al., (2019) developed a random forest (RF) to predict the compressive strength of self-compacting concrete. The model achieved high predictive accuracy indicated by a high correlation coefficient ($R = 0.97$). Farooq et al., (2021) compared ensemble approaches with individual machine learners for predicting the compressive strength of HPC. The results revealed that ensemble models with boosting and bagging showed robust performance as compared to the individual approach (i.e., DT, ANN, and SVM).

Since each ML technique has various advantages and drawbacks, the selection of the most suitable model is based on different criteria (i.e., the nature of the data, predictive accuracy, computational time). Model selection usually involves the process of choosing one among many candidate models for a predictive modeling problem. As such, a systemic comparison of all common ML techniques for predicting concrete compressive strength is needed. Additionally, the performance of the proposed ML models for predicting concrete strength from literature was often optimized by extensively tuning up the hyperparameters of each algorithm or developing a hybrid model based on conventional ML algorithms, which is considered time-consuming and computationally expensive as it requires testing numerous combinations before attaining the optimum values or models. Alternatively, this study presents the importance of exploratory data analytics and featuring engineering using domain knowledge to improve the prediction performance of concrete compressive strength. Feature (i.e., input variable) importance was computed to demonstrate the contribution of the synthetic features on prediction accuracy as compared to original features. Finally, the model performance from this study was compared with other previous from the literature to show the superiority of the proposed methodology.

2.2 Machine Learning Algorithms

This section briefly reviews the ML algorithms investigated and compared for optimizing the accuracy of concrete compressive strength.

2.2.1 Multiple Linear Regression (MLR)

Multiple linear regression algorithm establishes a linear relationship between a response variable (quantitative) and several explanatory variables. For multidimensional inputs, this algorithm learns a mapping from D -dimensional inputs to scalar outputs: $\mathbf{x} \in \mathbb{R}^D, y \in \mathbb{R}$ with the following equation:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = \sum_{i=1}^D w_i x_i + b \quad (2.1)$$

where w_i is the D -dimensional coefficient of the input variable x_i . In the proposed regression model, y represents concrete compressive strength, and x_i represents mixture components. The algorithm learns a vector of weights w by minimizing the least-squares cost function given by

$$E(\tilde{\mathbf{w}}) = \sum_{i=1}^N (y_i - \tilde{\mathbf{w}}^T \tilde{\mathbf{x}})^2 \quad (2.2)$$

where

$$\tilde{\mathbf{w}} = \begin{bmatrix} w_1 \\ \vdots \\ w_D \\ b \end{bmatrix}, \tilde{\mathbf{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_D \\ 1 \end{bmatrix} \quad (2.3)$$

2.2.2 Support Vector Machines (SVMs)

SVM is a supervised learning method that can be used for both nonlinear regression and classification analysis based on structural risk minimization (SRM). SVM was first proposed by Cortes & Vapnik (1995). The main concept of the SVM method is using an effective separation by a hyperplane, with the largest distance to the nearest training-data point and the lowest generalization error, to allow the SVM to minimize the error and obtain a better generalization. In support vector regression (SVR), the inputs from lower-dimensional input space are firstly mapped to a higher-dimension feature space using a nonlinear kernel function (e.g., polynomial function, sigmoidal function and Gaussian radial basis kernel functions), Figure 2.1. In this feature space, the SVM attempts to construct a linear objective function so that its output has a maximum deviation of ε from the actual targets, y_i , in the training dataset. The linear objective function in the high-dimensional feature space is expressed as

$$f(x, w) = \sum_{i=1}^n w_i g_i(x) + b \quad (2.4)$$

Where $g_i(x)$ is a set of n nonlinear transformations; w_i represents the weight vector, and b is a bias term.

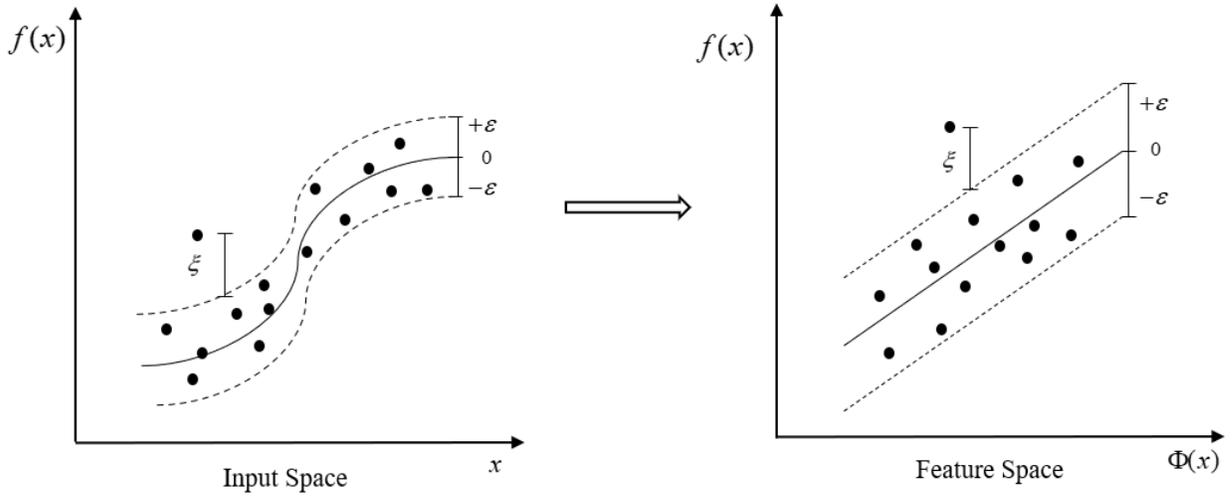


Figure 2.1 Nonlinear transformation

Compared to other machine learning algorithms, SVM has multiple advantages including a unique optimization approach and the effective utilization of high-dimensional feature spaces and computational learning theory. The kernel functions transform the data into a higher dimensional feature space to make it possible to perform linear separation. Thus, various kernel functions (Eq. 5) may generate different support vectors.

The SVR employs the following ε -insensitive loss function that penalizes error only if it is greater than ε .

$$L(\xi) = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \quad (2.5)$$

Linear regression is conducted in the high-dimensional feature space by minimizing $\|w\|^2$ and using ε insensitive loss. This can be achieved by introducing the nonnegative slack

variables ξ and ξ^* to determine training samples that deviate from the ε -insensitive zone (Figure 2.1). The linear regression is formulated following the optimization problem:

$$\begin{aligned} & \min\left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)\right) \\ & \text{subject to } \begin{cases} y_i - f(x_i, w) \leq \varepsilon + \xi_i \\ f(x_i, w) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \quad (i = 1, \dots, n) \end{cases} \end{aligned} \quad (2.6)$$

Compared to other machine learning algorithms, SVM has multiple advantages including a unique optimization approach and the effective utilization of high-dimensional feature spaces and computational learning theory. The kernel functions transform the data into a higher dimensional feature space to make it possible to perform linear separation. Thus, various kernel functions (Equation 2.7) may generate different support vectors.

$$K(X_i, X_j) = \begin{cases} X_i, X_j & \text{Linear} \\ (\gamma X_i \cdot X_j + C)^d & \text{Polynomial} \\ \exp(-\gamma |X_i - X_j|^2) & \text{RBF} \\ \tanh(\gamma X_i \cdot X_j + C) & \text{Sigmoid} \end{cases} \quad (2.7)$$

Where

$$K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j)$$

2.2.3 Decision Tree (DT)

Decision tree algorithms build regression or classification models and represent the data in the form of a tree structure. The algorithm breaks down a dataset into smaller and

smaller subsets while at the same time an associated decision tree is incrementally developed. The condition, or test, is represented as the “leaf” (node) and the possible outcomes as “branches” (edges). This splitting process continues until no further gain can be made or a preset rule is met (i.e., the maximum depth of the tree is reached). The core algorithm, ID3, for building decision trees employs a top-down, greedy search through the space of possible branches with no backtracking. The algorithm creates a multi-way tree in which each node can have two or more edges to detect the categorical feature that will maximize the information gain using the impurity criterion-entropy. The ID3 algorithm can be used to construct a decision tree for regression by replacing information gain with standard deviation reduction. Classification and regression trees (CART) are another widely used decision tree algorithm developed by Breiman (1984). The algorithm creates a binary tree (each node has exactly two outgoing edges) finding the best numerical or categorical feature to split using an appropriate impurity criterion. For regression, CART introduced variance reduction using least mean square error (MSE).

2.2.4 Artificial Neural Networks (ANNs)

Artificial neural networks are models that attempt to determine the relationship between input and output variables by simulating the structure of the biological neural network (human brain). ANN models can efficiently solve multi-dimensional or multi-variable problems. A neural network consists of an input layer, output layer, and hidden layer(s) of neurons. The input-output relationship is contained in the connections between neurons. Neurons in the hidden layers represent the features of input relevant to output. Backpropagation and gradient descent algorithms were employed to update the weights and

minimize the error. To train the neural network, first, the total error, which is the difference between NN output and target, is calculated. Then, the weight contributions (i.e., gradients) to the error are calculated at each connection. The error is calculated from the output, then moving back toward the inputs to calculate the gradient. Each weight keeps being updated and moves in the negative direction of the gradient to minimize the error, and until the desired error level is reached. The main challenge of ANN is that the selection of network size and parameters can be time-consuming.

2.2.5 Ensemble Learning (EL)

Ensemble learning is the approach by which multiple learners (e.g., SVM, linear regression, decision tree, and NN) are strategically generated and combined to one predictive model in order to improve predictive performance or reduce bias and variance (Feng et al., 2020) (Breiman, 1996) (Bühlmann & Yu, 2002)). Bootstrap aggregating (Bagging) and boosting are the most widely used ensemble learners. Bagging was first defined by Breiman as a method to generate different versions of a predictor leading to a more robust prediction (Breiman, 1996). It is an ensemble algorithm that reduces the variance by applying bootstrap sampling to obtain data subsets and using these different versions of datasets to generate multiple models. The final prediction is obtained by averaging the outcomes of these models for regression and using plurality voting for classification. For example, N different trees are trained on different subsets of the data (chosen randomly with replacement) and the ensemble is computed as follows:

$$f(x) = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (2.8)$$

Random Forest is a typical bagging technique capable of performing both regression and classification with the use of decision trees as a base learner. The basic idea is to combine and train multiple decision trees using only a random subset of the input variables to determine the final output rather than relying on an individual decision tree.

In boosting, the base learners are built sequentially such that each subsequent learner aims to reduce the errors of the previous learner. The main principle of boosting is to fit a sequence of weak learners to weighted versions of the data. More weight is given to examples that were misclassified by earlier rounds. The predictions are then combined through a weighted majority vote (i.e., classification) or a weighted sum (i.e., regression) to produce the final prediction. The most widely used form of boosting algorithms is adaptive boosting (Adaboost) proposed by Freund & Schapire (1996). In the Adaboost algorithm, the first base learner (i.e., decision tree) is trained using equal weighting coefficients on the original dataset. Then the weighting coefficients are adjusted according to the error of the current prediction in the subsequent boosting rounds. In this method, more weights are assigned to incorrectly classified or predicted samples. In the end, the weak learners are assembled with different weights to a strong and robust learner.

Stacking generalization is an ensemble learning method that learns how to optimally combine multiple ML algorithms using a new algorithm (i.e., meta model). In this study the base models were trained on a complete training dataset, while the meta model was trained

on the final predictions of all the base-level models. The stacking approach considers heterogeneous weak learners whereas bagging and boosting consider mainly homogeneous weak learners. Stacking learns to combine the base models using a meta model while bagging and boosting combine weak learners following deterministic algorithms. As mentioned earlier, the stacking ensemble model approach used in this study learns how to optimally combine the predictions from the base models to improve predictions. It has the advantage that any base model that performs poorly does not harm the performance of the stacking ensemble, since the meta learner would assign a weight of zero to such base learner. Thus, such stacking ensemble model uses combinations of predictions from other learners to produce a superior prediction.

2.3 Methodology and Implementation

Figure 2.2 illustrates the implementation process of the proposed methodology, which consists of three steps: *(i)* data preparation and feature engineering; *(ii)* training and optimizing individual models; *(iii)* training the stacked model and generating predictions. The modeling and analysis are implemented in Python programming.

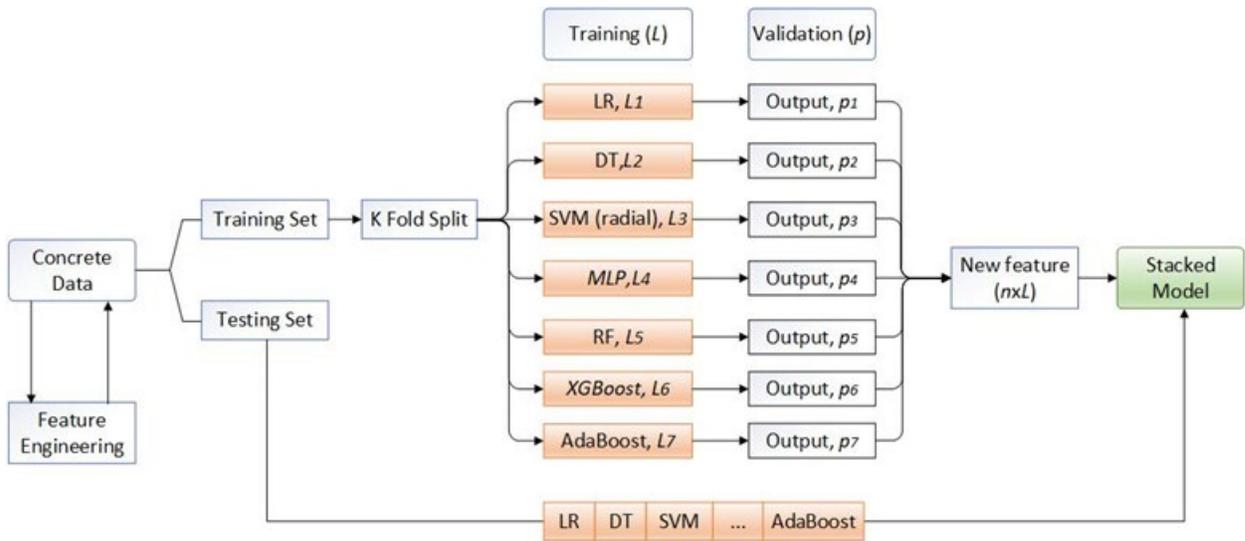


Figure 2.2 Flow diagram of the proposed methodology

2.4 Data Preparation

The database employed in this study includes 1031 laboratory measured HPC compressive strength from various mixtures (Yeh, 1998). Even though this dataset has been used by several researchers for evaluating ML algorithms, a comparison of the model performances, included later in this paper, demonstrates the superiority of proposed models and feature engineering techniques. There are eight independent variables (e.g., cement, blast furnace slag, fly ash, water, superplastic, coarse and fine aggregate, age) and one dependent variable (i.e., compressive strength). All the records are numeric, and no missing values have been observed in the dataset (Table 2.1). Data preparation involves best exposing the unknown underlying structure of the problem to learning algorithms by implementing the following tasks: exploratory data analysis, data cleaning, data transformation and scaling, feature engineering, and selection. The inter-quartile range (IQR)

method was employed to perform outlier detection. Based on the results from IQR, a small number of data points (n=68) were considered as outliers. However, the study decided to keep these data since they may represent the variability inherent in the concrete mixture data. In addition to outlier detection, normalization and transformation are carried out to convert the input variables into the best structural representation for ML algorithms. Data is normalized through scale and center transformation to convert the initial variables to have a mean of 0 and a standard deviation of 1. The input variables are standardized using the following formula:

$$z_i = \frac{(x_i - \mu)}{\sigma} \quad (2.9)$$

Where z_i is the standardized value of the original input variable; μ and σ are mean and standard deviations of the input variables

Power transforms are techniques for transforming numerical input variables to have a Gaussian or more-Gaussian-like probability distribution, which is preferred by many ML algorithms. In this study, a Box-Cox transformation is implemented to transform non-normal dependent variables into normal shapes. The formulation and implementation of such transformation can be found elsewhere (Box & Cox, 1964).

Feature engineering is the process of using domain knowledge to extract features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data. Three synthetic features were created including water to cementitious materials ratio (W/C ratio), fresh density, and aggregate to cement ratio. Table

2.1 summarizes the general statistics of each variable. All the three new variables were extracted from the eight original dependent variables. For instance, the W/C ratio is obtained by dividing water by cementitious materials (i.e., cement, fly ash, and Blast furnace slag), while fresh density is calculated by the summation of seven ingredients in one cubic meter of the mixture. The correlation coefficients between the input and output variables were also presented in Table 2.1. These additional features show a relative strong correlation to the compressive strength and thus are anticipated to be useful for improving the predictive performance as they have been proven to have considerable influences on concrete compressive strength (Ni & Wang, 2000); (Popovics & Ujhelyi, 2008). The feature importance was calculated in the later section to demonstrate how useful they are at predicting the strength as compared to the original variables.

Table 2.1 Statistics of concrete mixtures variables

Features	Mean	Standard deviation	Minimum	Maximum	Correlation coefficient (with strength)
Cement (kg/m ³)	281.17	104.51	102.00	540.00	0.48
Fly ash (kg/m ³)	54.19	64.00	0.00	200.10	-0.05
Blast furnace slag (kg/m ³)	73.90	86.28	0.00	359.40	0.14
Water (kg/m ³)	181.57	21.35	121.80	247.00	-0.37
Superplasticizer (kg/m ³)	6.02	5.97	0.00	32.20	0.40
Coarse aggregate (kg/m ³)	972.92	77.75	801.00	1145.00	-0.17
Fine aggregate (kg/m ³)	773.58	80.18	594.00	992.60	-0.16
Age (days)	45.66	63.17	1.00	365.00	0.52
Water to cementitious materials ratio	0.46	0.12	0.25	0.90	-0.66

Aggregate to cement ratio	7.33	2.88	3.10	17.93	-0.47
Fresh density (kg/m ³)	2345.87	61.05	2194.60	2551.00	0.45
Compressive strength (MPa)	35.82	16.70	2.33	82.60	1.00

2.5 Training Base Models

As shown in Figure 2.2, seven base models were investigated including four conventional modes (i.e., LR, DT, and SVM) and three ensemble learners (RF, Xgboost, and Adaboost). Prior to model training and evaluation, the dataset was randomly portioned into training (85%) and testing (15%) sets. This ensures to examine the generalization ability of predictive models and avoid data leakage as the testing set does not involve model training in any way. The seven base ML models are separately trained and tuned using the training data. K-fold cross-validation (where $k = 5$) is carried out to train the models and avoid overfitting. The hyperparameters were tuned using randomized search with cross-validation to optimize the models' predictions. Randomized search was used in which random combination of hyperparameters were selected to train each model. Firstly, the distribution for each hyperparameter was defined. The randomized search algorithm was then randomly selected sample values for each hyperparameter from the corresponding distribution to train each model using such values. This process was repeated for a specified number of iterations, and the optimal hyperparameters were chosen based on the performance of the models. The root means square error (RMSE) is selected as the primary performance metric during the training process. The optimized hyperparameters for each model are listed in Table 2.2.

Table 2.2 ML models with optimized hyperparameters

Model	Parameters	Setting
LR	polynomial features	2 degrees
DT	max depth	20
	min_samples_leaf	2
	min_samples_split	6
SVM	kernel	RBF
	regularization (C)	100
	gamma	2.0
MLP	hidden layers	2
	neurons	32
	activation	Relu
RF	number of trees	400
	min_samples_split	2
	max features	6
Xgboost	number of trees	600
	learning rate	0.1
	max depth	4
	subsample	0.8
	colsample_bytree	0.8
AdaBoost	number of trees	200
	loss	Square
	learning rate	0.3

2.6 Training Stacked Model and Generating Predictions

Stacked generalization or stacking is an ensemble technique that uses a new algorithm (i.e., meta model) to learn how to best combine the predictions from two or more models trained on the dataset. In this study, a two-layer stacked model is implemented to

combine the base models and model configurations that were investigated. The first layer consists of the 7 base models (Figure 2.2), and the outputs of these base models are used to train the stacked model (i.e., the second layer or meta-model). Firstly, a list of L base models (i.e., $L=7$) with specific model parameters is specified (Table 2.2). To train the stacked model, we need to make a loop for k -fold cross-validation. In each iteration, the training data is randomly divided into k blocks. The k -fold cross-validation is performed on each of the base models where all models use the same k -fold of the data, and the cross-validated predictions are collected. The predicted values represent $p_1, p_2 \dots p_L$ (Equation 2.10). The n cross-validated predicted values from each of the L algorithms can be combined to form a new $n \times L$ feature matrix represented by Z in Equation 2.10.

$$n \left\{ \begin{array}{c} \left[\begin{array}{c} p_1 \\ p_2 \\ \dots \\ p_L \end{array} \right] \end{array} \right\} \rightarrow n \left\{ \begin{array}{c} \overbrace{\left[\begin{array}{c} Z \end{array} \right]}^L \\ \left[\begin{array}{c} y \end{array} \right] \end{array} \right\} \quad (2.10)$$

Where n = number of rows in the training set. y = original response vector (i.e., target variable)

Once the feature matrix is obtained, a meta-learning algorithm needs to be specified. The meta-learning algorithm can be any of the base models investigated, but it is often suggested to use some form of regularization regression (Grolemond, 2014). As such, lasso regression is selected as the meta-model. It should be noted that the cross-validated predictions from each of the base models became new features (i.e., feature matrix Z) for the

meta-model. The feature matrix, along with the original response vector y is used to train the meta-model.

$$y = f(Z) \quad (2.11)$$

To make ensemble predictions, the predictions from each base model on the testing set need to be generated. These predictions are fed into the meta-model to generate the final prediction from the meta-model.

2.7 Evaluation of Model Performances

2.7.1 Performance Measures

The performance and accuracy of each ML model are evaluated by calculating performance measures including, root mean square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE). RMSE quantifies the averaged Euclidean distance between predicted and true strength data in the test set, while R^2 evaluates the accuracy of the predicted strength in terms of how close the data are from the fitted regression line (a perfect fit would provide an $R^2 = 1$). The performance measures are calculated based on the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{pred} - y_{obs})^2}{\sum_{i=1}^n (y_{obs} - \bar{y}_{obs})^2} \quad (2.12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{pred} - y_{obs})^2}{n}} \quad (2.13)$$

$$MAE = \frac{\sum_{i=1}^n |y_{pred} - y_{obs}|}{n} \quad (2.14)$$

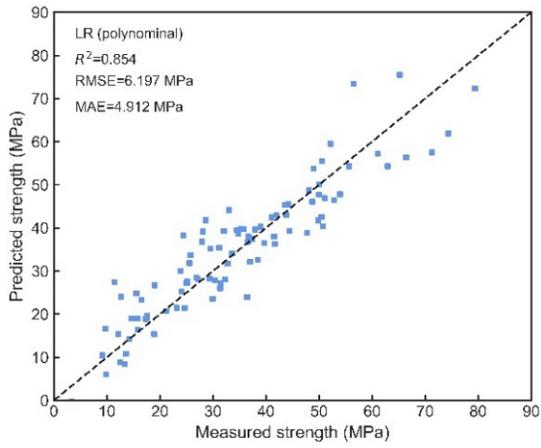
where y_{pred} and y_{obs} are the predicted and observed values, respectively; \bar{y}_{obs} is the mean value of the observed data; n is the total number of samples in the data set.

2.7.2 Performance of Base Models

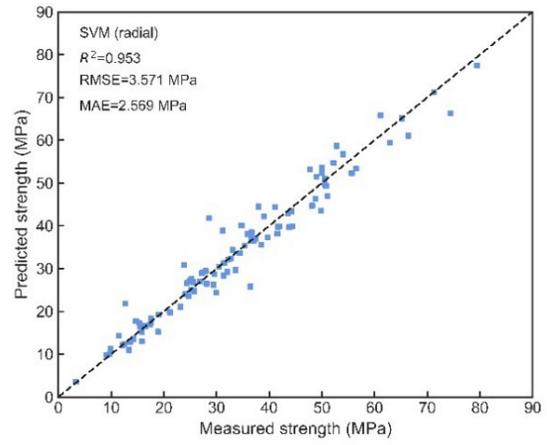
The prediction performance of the seven base models on the testing dataset is presented in Figure 2.3. The scatter plots show the relation between the predicted and measured strength for each model. A good model should have minimal discrepancies between the predicted and the measured values, or, in other words, all the data should be close to the regressed diagonal line. Linear regression is the first base model investigated in this analysis. The performance of linear regression is considered as a baseline for comparing the model performance of other ML models. For the linear regression model, a polynomial feature transform is implemented which raises the input variables to a polynomial power of two. This approach can help to better expose the complexity of interpreting the input variables and their relationships. The linear regression has a testing R^2 , RMSE and MAE of 0.854, 6.197 MPa and 4.912 MPa, respectively, which is significantly better than linear models (i.e., R^2 of 0.611 and 0.66) developed on the same dataset with the original features (Chou et al., 2011); (Young et al., 2019). This demonstrates that polynomial transformation

of the predictor variables improves the performance of the linear regression. Among the three individual ML models, the SVM made the best prediction (i.e., $R^2 = 0.953$, RMSE = 2.569) which outperformed DT and MLP with a R^2 of 0.907 and 0.940 respectively. The performance of SVM with RBF kernels also demonstrates the importance of input variables transformation to linearly separate the patterns that exist among concrete ingredients.

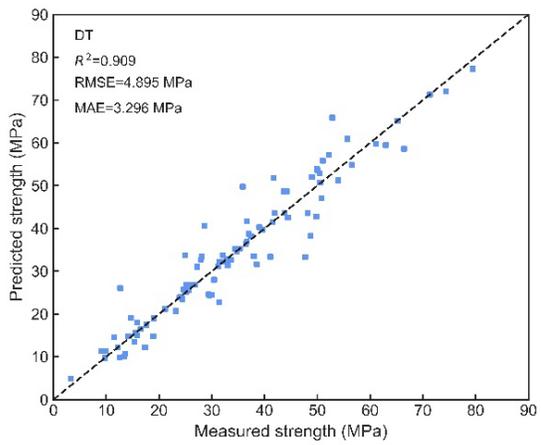
The ensemble learning models show a better performance than the single learning model proposed in this study, except that Adaboost generates a slightly lower accuracy ($R^2 = 0.937$, RMSE = 2.569) than SVM and MLP. This indicates the reliable prediction capabilities of generalization for these ML techniques. In particular, The Xgboost exhibits the best predictive capability, which explains more variance in the data ($R^2 = 0.975$) and achieves higher accuracy (RMSE = 2.565, MAE = 1.835) in concrete strength prediction. The second-best predictive model is RF with an R^2 , RMSE, and MAE of 0.941, 3.946 MPa, and 2.880 MPa, respectively. The superior prediction performance of Xgboost and RF may be attributed to the ability of tree-based methods to learn inconsistent variable importance in the data.



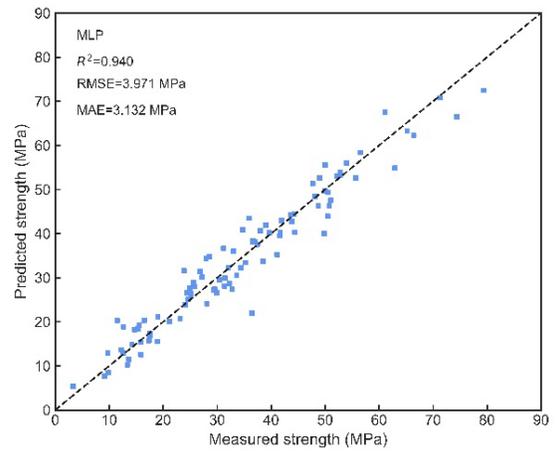
a



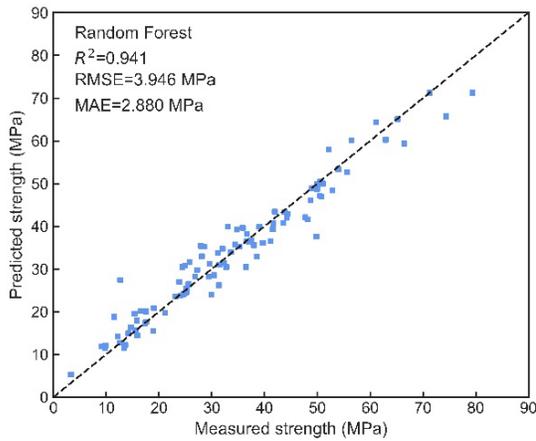
b



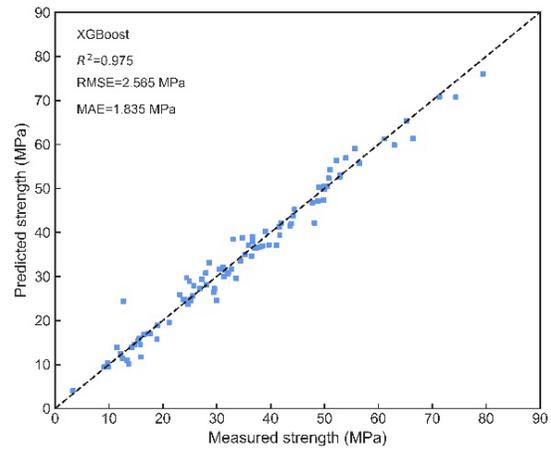
c



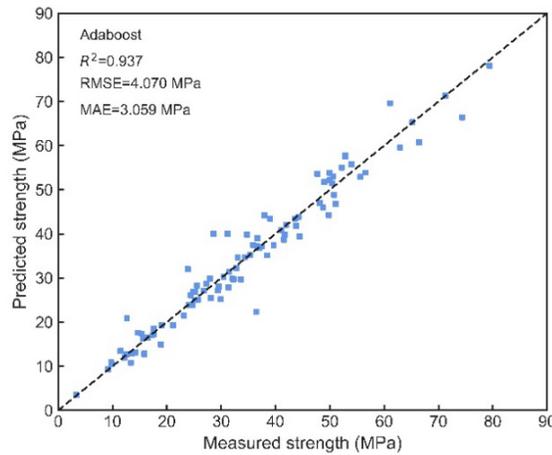
d



e



f



g

Figure 2.3 Relationship between predicted and measured compressive strength: a) linear regression, b) support vector machine, c) decision tree, d) multiple layer perceptron, e) random forest, f) Xgboost, g) Adaboost

2.7.3 Performance of Stacked Model

The predictive performance of the stacked ensemble is presented in Figure 2.4. For both training and testing dataset, a strong linear relation (R^2 of 0.993 and 0.985, respectively)

between the predicted and measured values is observed. The stacked model achieves a considerably high accuracy on the testing data with an RMSE of 1.941 MPa and MAE of 1.135 MPa, which outperforms any of the base models. This indicates that the stacked model learns the optimal combination of the base learners, and thus improves the prediction performance. The theoretical detail on the optimality of stacking is explained by (Van der Laan et al., 2007). As shown in Figure 2.4, most of the points for testing data are either on or close to the diagonal line indicating high accuracy of prediction. However, for extremely large values (i.e., compressive strength >70 MPa), the model slightly underestimates the actual compressive strength. The reason could be that the base models steadily underestimate the actual strength for these data points as shown in Figure 2.3. Further residual or error analysis (i.e., residual normality and heteroscedasticity) could be conducted to check whether these data points are outliers.

In order to assess which learning algorithm (i.e., “learner”) is more suitable for the particular dataset, different types of base algorithms are explored. When the underlying functional form, relating the various material properties to strength in this case, is simple each algorithm may be able to provide good predictions. However, in this case due to the complexity of such relationship, one type of algorithm may be more successful than another. For instance, unlike a main terms parametric model, a tree-based algorithm like random forest inherently considers interactions and is unaffected by monotone transformations of the data. Since the true functional form of the parameters is unknown, there is a need to explore alternative base learners. Therefore, in this study alternative models were considered, from the standard parametric models (e.g., linear regression) to the more complex data-

adaptive models (e.g., SVM, tree-based algorithms, and multiple layer perceptron). Furthermore, in this effort when a base learner performs poorly it does not harm the performance of the stacking ensemble since the meta learner would assign a weight of zero to such learner. The advantage of the stacking ensemble model is that considers a combination of predictions from other learners to produce a superior prediction.

The meta-learning algorithm is often some sort of regularized linear model which provides a smooth interpretation prediction of the predictions generated from the base models (Phillips et al., 2022). Using a simple linear regression as the meta model may also reduce the chance of overfitting the predictions from the base models. Moreover, predictions from the base models are usually strongly correlated, as they are trying to predict the same relationship. Therefore, a lasso regression with regularization parameters was used to deal with the correlations between the predictions of the base models.

Although different combinations of base models may affect slightly the accuracy of the stacked model, the stacking is expected to perform better than or equal to any of the base models. Higher performance gains are usually produced when stacking base learners have high variability and uncorrelated predicted values (Grolemond, 2014). The cross-validated predictions for the base models represent the model's generalization capability in making predictions on data not seen during training (i.e., cross-validation). A 5-fold cross-validation was used to find the optimal weighted combination of predictions from the candidate algorithms (i.e., base models) By training a meta-model with out-of-sample (i.e., testing data) predictions of the base models, the meta-model learns how to both correct and best combine the out-of-sample predictions from multiple models. As such the stacked

model is unlikely to overfit the training data. When comparing to the performance of the base models (i.e., Xgboost), the stacked model increases 0.011 in terms of the R^2 score. The dataset used in this study is relatively small and less complex, which means that it is not easy to capture any of the patterns that the base algorithms (i.e., RF, Xgboost) could capture already. However, the results still demonstrate the superiority of stacking which produces better prediction accuracy. The prediction performance is expected to be optimized to a higher degree for more complex concrete datasets (e.g., field concrete data).

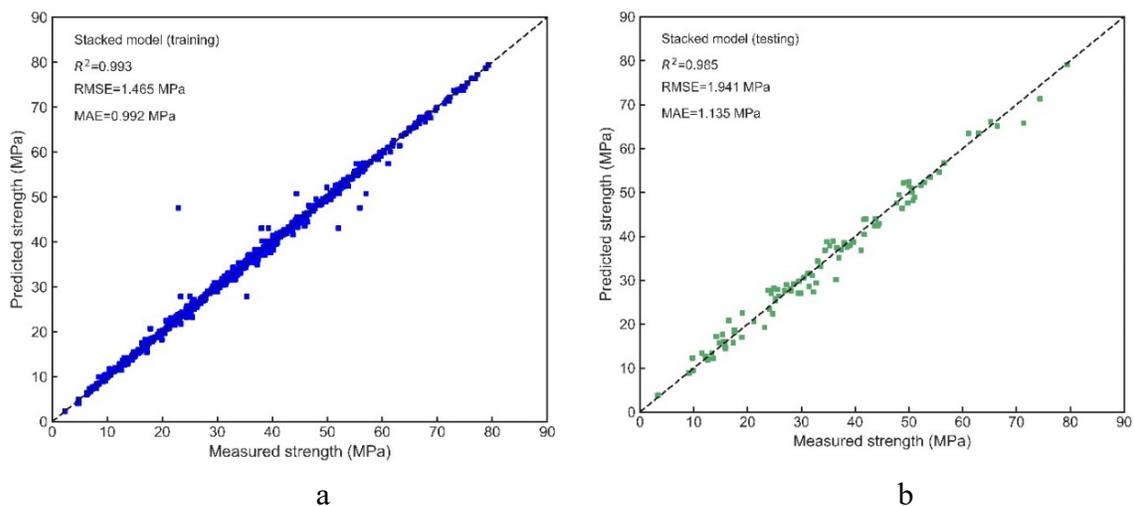


Figure 2.4 Predicted vs. measured strength for stacked model, a) on training dataset, b) on testing dataset

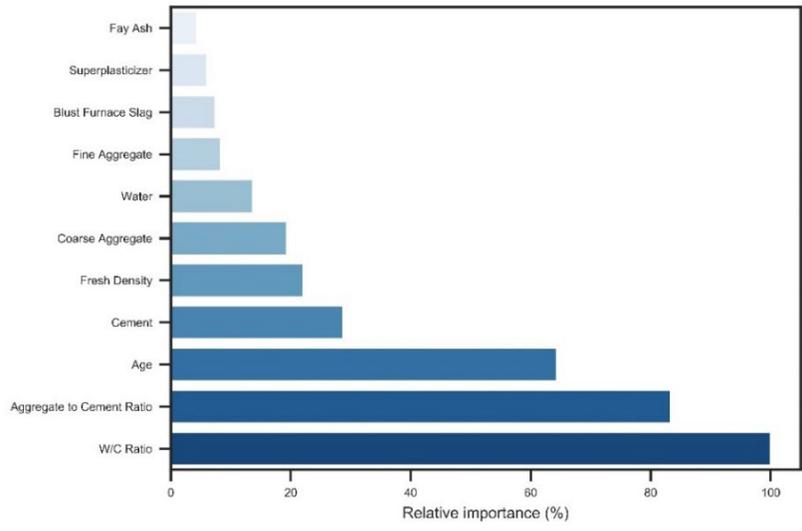
2.7.4 Feature Importance

Feature importance is used to describe how important the feature is at predicting the target variable. More precisely, it is referred to as a measure of the individual contribution of the corresponding feature for a particular model, regardless of the shape (e.g., linear or nonlinear relationship) or direction of the feature effect (Zhang et al., 2021); (Fisher et al.,

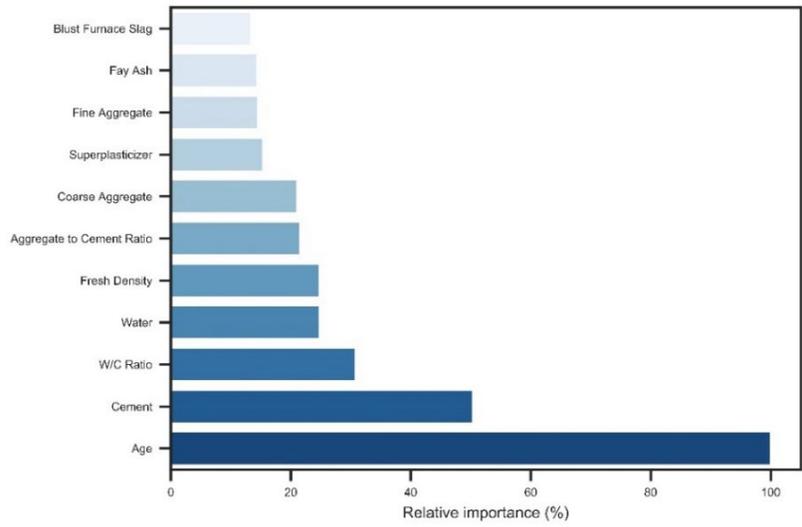
2019). This means that the feature importance of the input data depends on the corresponding ML model. Feature importance provides better interpretability of the data and the model, and sometimes improves the performance of the model by implementing feature selection techniques.

There are three main types of metrics for determining feature importance, including model coefficients, the mean decrease in impurity (MDI), and permutation feature importance. For parametric linear models (e.g., LR, logistic regression, lasso, and ridge), the input variables are evaluated against the absolute normalized values of regression coefficients. Tree-based models (e.g., DT, RF and Xgboost) provide an alternative measure of feature importance based on MDI, which is quantified by the splitting of the decision tree. For example, the input variables of RF are evaluated by the mean square error (MSE) on the out-of-bag data for each tree before and after permuting the variables. The variable importance is determined by the averages and normalized variations of MS. For Xgboost, in each boosting iteration, the decrease in the loss function (i.e., MSE) in association with each input variable at each split is noted and summed (Zhang et al., 2021). Permutation feature importance is a technique for determining a relative importance score that is dependent on the model employed. As such it is used for models (e.g., SVM and MLP) that do not support native feature importance scores. The importance of a feature is measured by calculating the increase in the model's prediction error after permuting the feature (Interpretable machine learning reference). A feature is considered as "important" if shuffling its values increases the model error because in this case, the model relied on the feature for the prediction. The employed algorithm for the permutation feature is based on (Fisher et al., 2019).

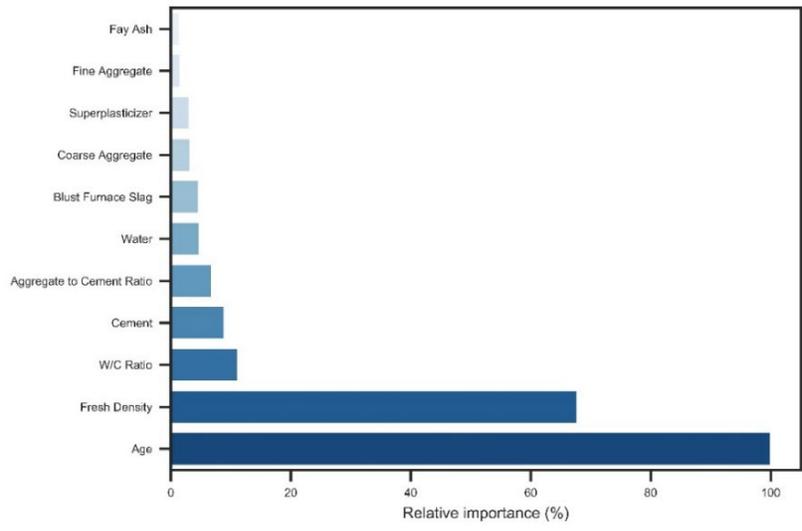
The relative importance of the input variables for each model is presented in Figure 2.5. It can be observed that age, water, fresh density, cement, W/C ratio and aggregate to cement ratio play a more significant role than the rest of the variables for predicting compressive strength. This also demonstrates the value of synthetic features (i.e., fresh density, W/C ratio, and aggregate to cement ratio) and the importance of feature engineering for improving prediction performance. In particular, age is determined to be the most important variable in prediction for nonlinear models, while the W/C ratio made the biggest contribution for linear regression. This may be because there is a strong correlation (-0.66) between W/C and compressive strength, which can be easily captured by linear models. Even though no significant correlation of coefficient is observed between age and strength, the degree of hydration is synonymous with the age of concrete, which significantly increases the compressive strength by 28 days. ML models provide greater flexibility to capture such nonlinear patterns. The other two new features (i.e., aggregate to cement ratio and density) are also proven to be useful in predicting concrete strength as shown in Figure 2.5. These observations tend to be consistent with engineering practice and the physical properties of regular concrete mixtures. Therefore, the ML models and feature importance techniques may be useful and reliable to analyze the relation and interaction between the ingredients and strength of innovative mixtures (e.g., lightweight, self-healing and green concrete mixtures).



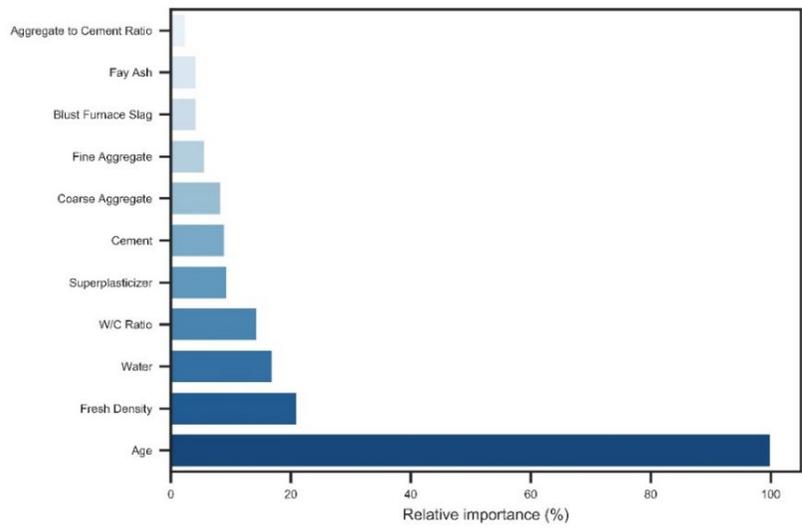
(a) Linear regression



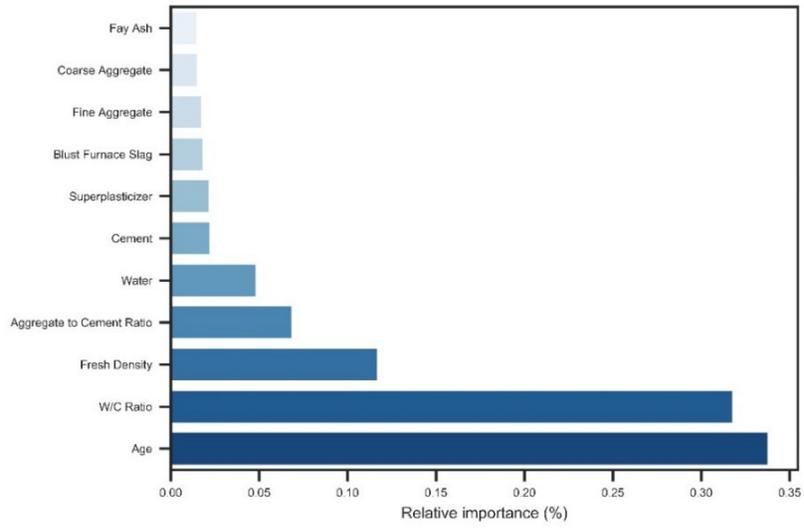
(b) SVM



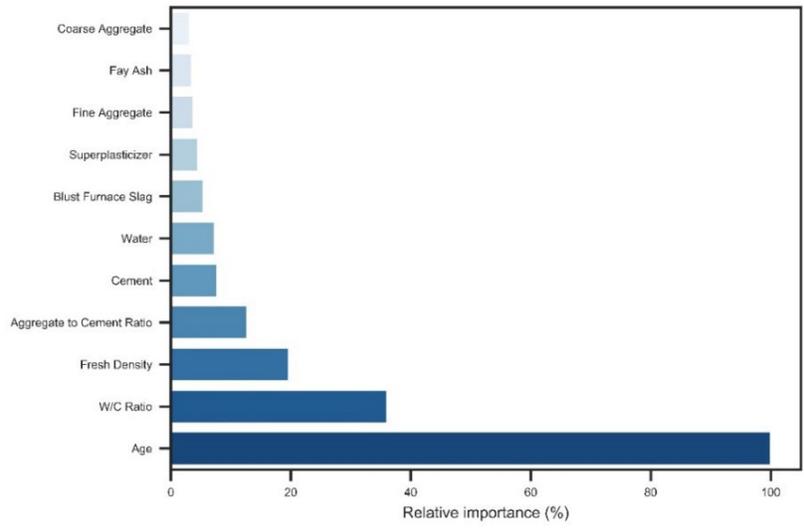
(c) Decision tree



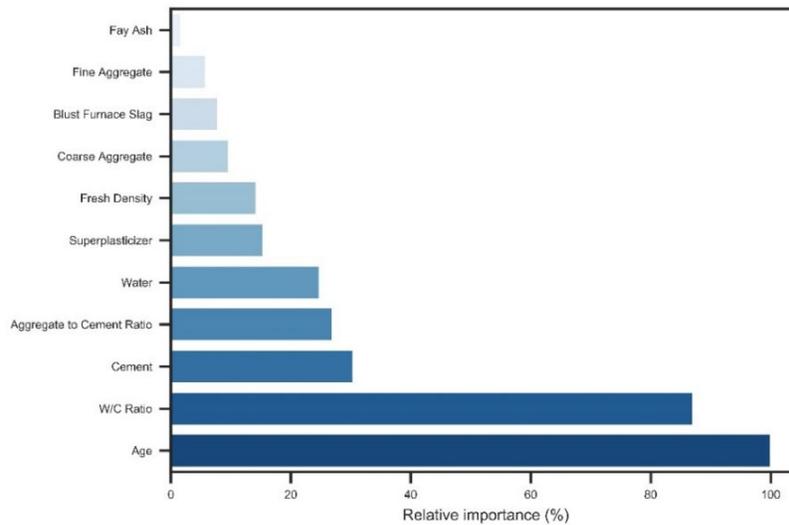
(d) MLP
2.8



(e) Random forest



(f) Xgboost



(f) Adaboost

Figure 2.5 Relative importance of input variables for each model

2.7 Comparison with Models from Previous Studies

Several studies have proposed models to predict concrete compressive strength using the adapted HPC dataset. Table 2.3 Comparison of ML model performance on the HPC dataset provides a comparison of models' performance. For the four individual models, this study significantly improves the prediction performance except for ANN, which may be largely attributed to the newly created variables. For instance, the linear regression achieves an R^2 of 0.859 which is significantly higher than other developed linear models. There are two possible reasons for such an improvement. One is that the power transformation makes the input variables more normally distributed which is preferred in linear models. A second possible reason is that W/C ratio has a higher linear correlation with strength which improves the prediction performance. This is also demonstrated by Yeh's results where an R^2 of 0.779

is obtained using only the W/C ratio and age as input variables (Yeh, 1998). Additionally, this study improves the performance of the SVM model significantly with an R^2 of 0.953 compared to Chou et al. with an R^2 of 0.953 (Chou et al., 2011). Even though the ANN generates slightly lower accuracy as compared to Nguyen et al. and Erdal et al. it should be noted that this study simply employs an MLP with 2 hidden layers while other studies developed much more complex hybrid models which could be time-consuming and computationally expensive (Nguyen-Sy et al., 2020); (Erdal et al., 2013).

In terms of ensemble learning, the accuracy of the RF model in this study is slightly lower compared to (Han et al., 2019). It is worth mentioning that they also created a new variable (i.e., coarse aggregate to binder ratio) to train the RF model and conducted hyperparameter optimization. The XGBoost model developed in the present study provides similar performance in terms of R^2 , RMSE and MAE compared to Chakraborty et al. (2021). The AdaBoost of Feng et al. provided a better prediction performance than the present study by performing a hyperparameter optimization (Feng et al., 2020). Chou et al. (2011) implemented stacking of LR, DT, and SVM, however, resulted in much lower accuracy. Their results showed that the stacking outperformed any of the base learners, however, due to the poor performance of the base models, the stacking did not provide a competitive prediction ability. Notably, the stacked model proposed in this study outperforms other models reported from the literature. For the two-layer stacked model, assembling stronger regressors in the first layer is likely to improve the performance of the model, while a simple, explainable, parametric model (i.e., Lasso) serves better in the second layer. The mechanism behind the more efficient learning ability of the proposed model is related to the elements in

the proposed methodology (Figure 2.2), including data processing, feature engineering, base model development, hyperparameters tuning, and stacking.

Table 2.3 Comparison of ML model performance on the HPC dataset

ML algorithms	R ²	RMSE (MPa)	MAE (MPa)	Remarks	Reference
LR	0.854	6.197	4.912		*
	0.779	N/A	N/A	Input variables: water to binder ratio and age	Yeh (1998)
	0.611	10.428	N/A		Chou et al. (2011)
	0.660	8.800	N/A		Young et al. (2019)
DT	0.909	4.895	3.296		*
	0.911	4.948	N/A	Employed MART (i.e., gradient boosting)	Chou et al. (2011)
	N/A	7.840	5.860		Chou et al. (2014)
	N/A	7.37	4.62		Farooq et al. (2021)
SVM	0.953	3.571	2.569	Radial kernel	*
	0.886	5.572	N/A		Chou et al. (2011)
	N/A	5.590	3.750		Chou et al. (2014)
	0.830	6.400	N/A		Young et al. (2019)
ANN	0.940	3.971	3.132	Simple MLP with two hidden layers	*
	0.942	4.050	2.850	Proposed a high order deep neural network	Nguyen et al. (2019)
	0.953	5.750	4.830	Employed gradient boosted ANN	Erdal et al. (2013)
	0.914	N/A	N/A	Hand tuning for hyperparameters	Yeh et al. (1998)
RF	0.941	3.946	2.880		*
	0.850	5.800	N/A		Young et al. (2019)

	0.965	4.433	3.105	new feature: coarse aggregate to binder ratio	Han et al. (2019)
	0.922	4.6	3.23		Farooq et al. (2021)
Xgboost	0.975	2.565	1.835		*
	0.930	4.640			*
	0.980	2.650	1.890	With feature selection and hyperparameter optimization	Chakraborty et al. (2021)
	0.902	5.170	3.710		Farooq et al. (2021)
Adaboost	0.937	4.070	3.059		Nguyen-Sy et al. (2020)
	0.982	2.200	1.640	With hyperparameter optimization	Feng et al. (2020)
	0.919	5.220	3.690		Farooq et al. (2021)
Stacked model	0.985	1.94	1.135		*
	N/A	5.08	3.520	Stacking of three learners: SVM, DT, and LR	Chou et al. (2014)

Note: * models developed in this study

2.8 Summary and Conclusions

The goal of this study was to explore alternative ML models for enhancing the prediction of compressive strength as a function of mixture ingredients and proportions. Seven individual ML techniques were considered and compared including LR, DT, SVM, MLP, RF, Xgboost, and Adaboost. To improve the prediction accuracy, a methodology was proposed in which the feature engineering technique was implemented, and a two-layer stacked model was proposed. The k-fold cross-validation approach was employed to

optimize model parameters and train the stacked model. The performance of these models was evaluated and compared with R^2 , RMSE, and MAE. Furthermore, the relative importance of the input variables in predicting compressive strength was assessed.

The results showed that, among the seven individual models, the Xgboost exhibited the best predictive performance with an R^2 of 0.975 followed by SVM which generated an R^2 of 0.953. The results for feature importance showed that age, water, fresh density, cement, W/C ratio, and aggregate to cement ratio play a higher role in predicting compressive strength than the rest of the variables. The proposed stacked models in this study outperformed other models reported in the literature with a testing R^2 , RMSE, and MAE of 0.985, 1.941 MPa, and 1.135 MPa. This approach encourages a more thorough consideration of alternative ML algorithms rather than focusing on a single machine learning approach. While the implementation process for the proposed methodology was presented, the study results indicated that the ML models (i.e., SVM, stacked model) can provide fairly high accuracy of concrete strength predictions. Thus, such ML models can be used to accurately predict compressive strength, minimizing the need for time-consuming and costly testing for mix design optimization and quality assurance during production.

CHAPTER 3: EVALUATION OF ACCEPTANCE RISK AND PAY FACTOR IN QUALITY ASSURANCE BY VERIFICATION TESTING

3.1 Background

The most recent quality assurance (QA) stewardship reviews show that over half of the state highway agencies (SHAs) are using contractor testing data for the acceptance and pay decisions of pavement construction and materials (Grogg, 2021). This practice is permissible under Federal regulation 23 CFR 637 Subpart B as long as the quality of the material or construction is validated by verification testing using independent samples ((FHWA, 1995). Although verification procedures vary from agency to agency, the most used statistical verification tests are the F-test and t-test as recommended by AASHTO (AASHTO, 1996). The F-test is used to compare the variances of the contractor and agency test data, whereas the t-test is used to assess the degree of difference in means between the two data sets. These two hypothesis tests are used to statistically determine if the contractor and agency test data are from the same population (i.e., if contractor data from production and acceptance testing data by the owner represent the same population in statistical terms, and thus reflect the same level of quality). Over the years the various benefits and associated risks with the verification procedures have been pinpointed. On one hand, the use of contractor test results for acceptance can reduce inspection personnel and testing resources for SHAs (Schmitt et al., 2001); (Wani & Gharaibeh, 2013). Contractors could be encouraged to pay more attention to their sampling and testing procedures knowing that the test results will be used for payment decisions (Burati et al., 2010); (Carr et al., 2016). On

the other hand, the primary concern is the fact that contractor and SHA data do not always consistently compare well (Schmitt et al., 2001); (Wani & Gharaibeh, 2013); (Burati et al., 2010); (Carr et al., 2016); (Killingsworth & Hughes, 2002). There is also a lack of understanding of the practical implications of the statistically significant differences in test results, which may limit the ability of the QA process to properly identify the risks of lower quality and/or erroneous rewards. It is therefore important to evaluate the acceptance risks and associated payments with the verification procedures and identify any possible improvements.

Figure 3.1 shows a summary of the proposed methodological framework for conducting such analysis. In the first step the current specifications and pertinent QA standards and policies need to be reviewed. This review will identify the acceptance quality characteristics (AQC), such as strength, density, thickness, the specification upper and lower limits, and the pay factor (PF) schedule. Next, the available contractor production quality data will provide the characteristics of the pertinent population distribution. Similarly, the agency acceptance data will lead to the distribution for such population. In the next step the numerical simulation will be conducted using the population distributions reflecting the contractor's and agency's data. The objective of the simulation is to generate lots and sublots for assessing the effects on acceptance risks, which are explained in greater detail in later sections, and pay factors at the long run of production. The lots are generated based on the distribution characteristics of the specific AQC. The simulated lots generated using the contractor and agency distributions are then compared. In this process alternative hypothesis scenarios between the contractor and agency lots are used in order to assess the impact of

sample size, dispersion values, and distance in central tendency. Based on these analyses, the power surface and PF probabilities curves are developed. Such results will thus provide the acceptance risks for each scenario, and the related rewards. Finally, based on these analyses and findings, recommendations on adjusting specifications may be identified. These may include modifications in sample size, specification tolerances, specification limits; pay schedules revisions to balance risks levels to rewards; and recommendations on improving construction quality, in relation to achievable levels by the industry and construction equipment. The methodology is described in detail in the following sections

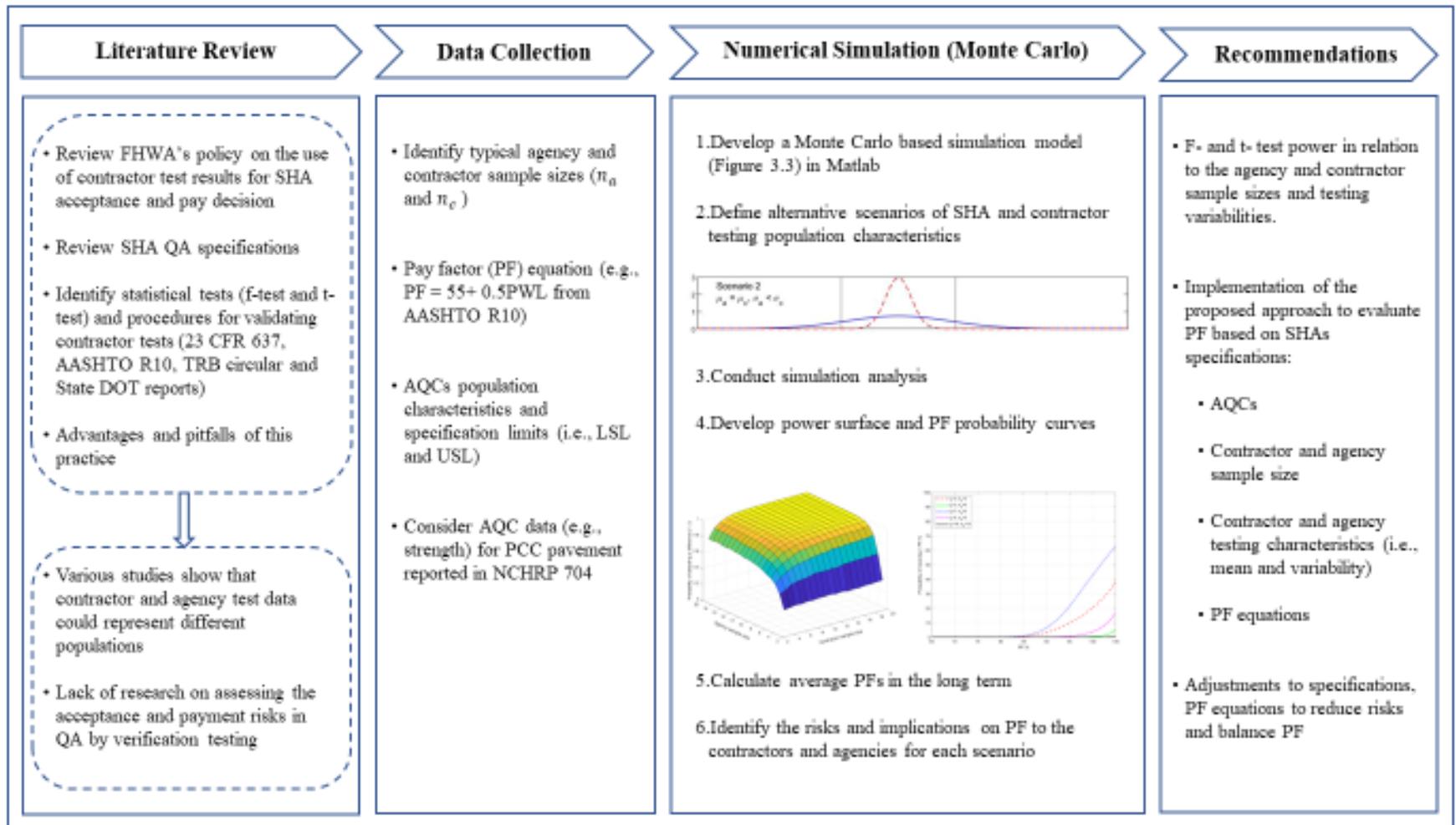


Figure 3.1 Methodological framework of analysis

3.2 Literature Review

A literature review was conducted to examine the various aspects of using contractor production quality data for acceptance, as well as identifying recommended procedures for validating contractor tests results. Killingsworth and Hughes (2002) pointed out that using contractor test data for acceptance and pay decisions has some inherent risks. It was suggested that when using contractor testing for acceptance, (i) the separation of agency quality verification and contractor testing should be maintained, and (ii) quality-based contractor prequalification procedures should be developed. Federal regulation 23 CFR 637 Subpart B also requires the separation of contractor and agency testing data (FHWA, 1995). Schmitt et al. (2001) proposed a statistically based method to perform verification tests by recognizing the importance of sample size, variability, and pertinent risks. The study also identified that there is a lack of risk assessment in current QA practice. Mahboub et al. (2004) conducted paired t-tests to detect any differences between the contractor and agency testing data of key AQC's for hot mix asphalt (HMA) and Portland cement concrete (PCC). These AQC's included air void and asphalt content, volume in the mineral aggregate, VMA, slump and strength. For each AQC, more than 6000 independent samples were examined. It was concluded that there was no statistically significant difference between the two data sets (i.e., contractor production quality and agency acceptance data), yet potential acceptance risk for deviations between the two datasets were not considered.

In another study comparing testing results of contractor and agency data from various states on asphalt mixtures mat density identified that such datasets are statistically different. Furthermore, the contractor data was less variable, producing thus higher values within the

specification limits, and thus providing higher level of acceptance and rewards. Turochy & Parker (2007); LaVassar et al., (2009) examined the contractor and agency test results from the states of California, Minnesota, Texas, using alternative statistical analyses and including F- and t-tests. The study concluded that F- and t-test were effective in validating contractor test results when adequate sample sizes were used. Karimi et al. (2012) analyzed the means and variability of a large set of Maryland contractor and agency testing data for hot mix asphalt (HMA). Through statistical analysis (i.e., F-test and t-test) it was concluded that the contractor and agency data represent different populations. Thus, research is needed to assess how these statistically significant differences in testing populations affect acceptance risks and payments.

Wani et al. (2013) identified that F-test and t-test verification procedures can lead to erroneous pay decisions due to the potential manipulation of contractor tests results. Coenen et al. (2019) examined the implementation of statistical verification testing in percent within limits (PWL) specifications for HMA for the Wisconsin Department of Transportation (WisDOT). It was concluded that such approach has allowed WisDOT to adjust payments in conformity to material quality and consistency, and ultimately relate PF to the anticipated long-term performance. Recent studies evaluated the procedures for validating contractor test data when used for acceptance of construction and materials (National Academies of Sciences and Medicine, 2020); (Nimeri, 2019). The analysis assessed different aspects including sampling methods, sample size, retesting and associated risks.

Despite the vast research on the validation of contractor test results in QA, none of the studies systematically evaluated the power of the statistical tests considering different

contractor and agency population characteristics. Equally important, the majority of SHAs are implementing PWL specifications with pay adjustment provisions. In this context, the combined contractor and agency testing data are used to determine PWL and pay factor (PF) of a lot. Therefore, it is vital to assess how this type of procedure affects the associated risks and PF in the long run of production. Thus, this study proposed the analysis process based on Monte Carlo simulation to: (i) determine the statistical power of the F-test and t-test (i.e., probability of detecting a difference between two populations versus the actual difference) for any combination of contractor and agency sample sizes; and, (ii) evaluate the PF and associated risks when validated contractor production quality testing data, in combination with agency acceptance data, are used for acceptance and pay decisions. To demonstrate the value of the proposed approach, the simulation was conducted on selected AQC (i.e., thickness) of PCC pavements. The testing data reported in a national study were used to define the contractor and agency testing distributions and develop the alternative scenarios (Hughes et al., 2011). An initial implementation of the proposed approach was presented recently (Zhao & Goulias, 2021b); (Zhao & Goulias, 2021a) and the value of such analysis led to this extended study.

3.3 Power of Hypothesis Testing and Acceptance Risks

When the contractor test results are used as part of the material quality acceptance decision, validation testing is required. Hypothesis tests (i.e., F and t-test) are commonly

used in this process. The power of hypothesis tests is introduced herein in order to identify the associated risks with verification testing in QA.

The null (H_0) and the alternative hypothesis (H_1) need to be specified in the analysis. In statistics the assumption is that H_0 is true, and then the sample data are used to determine if there is adequate evidence to reject H_0 . Thus, the null hypothesis can only be either accepted or rejected. Hypothesis testing is performed at a significance level (α). The value of α is typically selected at 0.10, 0.05, or 0.01 significance level. An α of 0.05 means that there is only a 5% chance that H_0 is true and was erroneously rejected. This is the probability of type I error (or α risk). There is also a risk of failing to reject H_0 when it is actually false. This is identified as type II error (or β risk). The statistical power is thus defined as the probability of rejecting H_0 when it is false, and it is equal to $1-\beta$. A graphical illustration of type I and II errors and associated power is shown in Figure 3.2. In terms of material quality for example, α risk means that while the agency and contractor populations are the same H_0 (i.e., equal means and standard deviations), the statistic tests on the averages (t-test) and/or variability (i.e., standard deviation, F-test) between the agency and contractor sample data conclude that they are not. Thus, the sample statistics are used to calculate the P-value. If the P-value is less than or equal to α , then H_0 is rejected in favor of the alternative hypothesis. When the P-value is greater than α , then H_0 is accepted. In validating contractor test data, H_0 is defined as the contractor and agency test data are from the same population. For the t-test, H_0 is that the means of the contractor and agency data are statistically equal, while, for the F-test, H_0 is that the variances for the two datasets are equal.

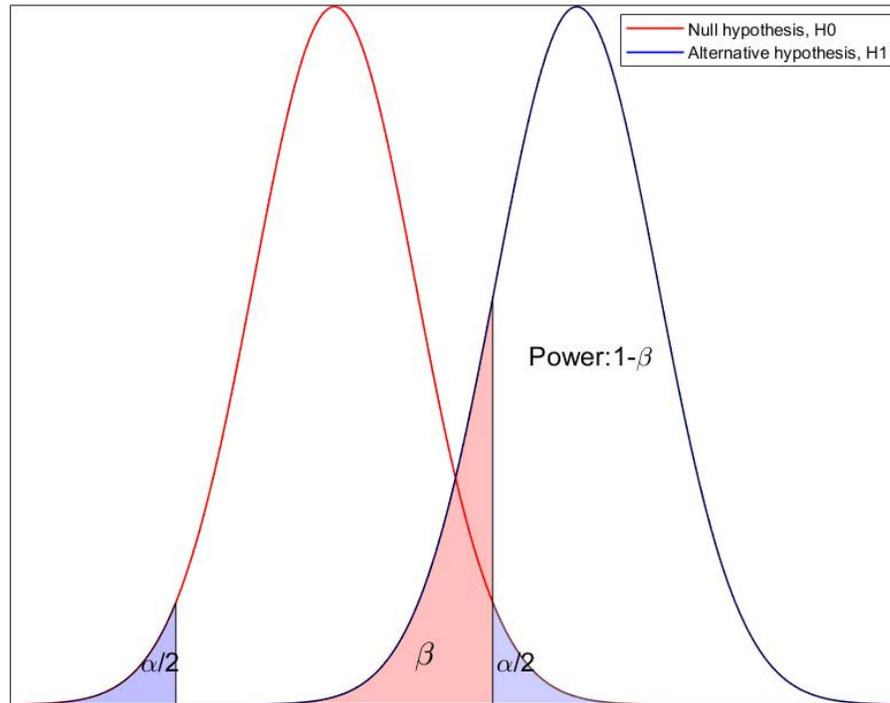


Figure 3.2 Graphical illustration of level of significance, α , β and hypothesis testing power for two-sided case

The power of the hypothesis tests is evaluated with the power curves. These represent the probability of detecting a difference ($1-\beta$) versus the actual difference between the two populations that are compared. There are 4 primary factors that affect the power of a statistical test, including: the α level, the difference in means, variability, and sample size. In reality, the actual difference between the two populations' means and variability are unknown. However, this is reasonably assumed (i.e., statistically inferred) based on the statistical analysis of the agency and contactor data using historical data on actual highway projects that may be available to SHAs. Sample size affects the power by influencing the variability of the sample distribution of the means. Increasing sample size reduces the

variability of sample means and thus increases the power. The required samples to reach a specific power can then be determined through power analysis.

In this study, the power of the F-test and t-test was determined for a wide combination of contractor and agency sample sizes at the significance level of 0.05. It should be noted that these hypothesis tests are conducted to determine if the contractor and agency test results are from the same population. This is an important step in QA since several highway agencies in their QA plans and specifications do consider the option of using contractor data for acceptance. Neither of the tests determine the accuracy of the test data, or whether the samples are representative of the material or construction population. Thus, the risk/power for the verification procedure does not reflect the risk for acceptance plans. The risks associated with verification testing are: the probability of incorrectly detecting a difference that does not exist (α), (i.e., statistically concluding that the contractor and agency test results are from different populations, when in reality they are not); and the probability of not detecting a difference that exists (β) (i.e., concluding that the contractor and agency test results are from the same population, when in reality they are not). On the other hand, the risks for acceptance plans are the probability of erroneously rejecting AQL material or construction (α); and the probability of incorrectly accepting RQL material or construction (β). The risk analysis for acceptance plans has been conducted by many studies (AASHTO, 2010); (Burati et al., 2003); (Burati, Straub, et al., 2004); (Karimi et al., 2015); (Zhao and Goulias, 2021a); (Zhao and Goulias, 2021b). However, there is a lack of research on risk and PF analysis associated with the verification procedures which was the objective of this study.

3.4 Numerical Simulation

This study proposed the necessary analysis process based on Monte Carlo simulation (Fig. 5) to: (i) determine the statistical power of the F-test and t-test for any combination of contractor and agency sample sizes, and (ii) evaluate the PF and associated risks when validated contractor testing data, in combination with agency testing data, are used for acceptance and pay decisions. To demonstrate the value of the proposed approach, the simulation was conducted on selected AQC (i.e., thickness) of PCC pavement. The simulation consists of three parts: (1) simulation of the AASHTO verification procedure; (2) determination of PF for each simulated lot; (3) evaluation of F-test and t-test power and PF.

3.2.1 Simulation of AASHTO Verification Procedure

In each simulation iteration, contractor and agency test results are randomly sampled from the populations, representing independent samples taken from a lot. The simulated contractor and agency test results are used to determine parameters such as mean and standard deviation (or variance). Once the means and variances are calculated for both agency and contractor testing data, the contractor data is validated against the agency data using the AASHTO's verification process (AASHTO, 1996). In this process, F-test is firstly conducted to determine if the variances of the two datasets are statistically different. The F-statistic is calculated as the ratio of the variances of the contractor and the agency test results from a pavement lot.

$$F = \frac{s_1^2}{s_2^2} \quad (3.1)$$

where s_1^2 is the larger variance from either contractor or agency, and s_2^2 is the smaller variance of the two.

The P-value for F-test is determined using the built-in F cumulative distribution function, *fcdf*, in MATLAB. When the P-value is less than the significant level, α , the null hypothesis is not rejected meaning there is no reason to believe that the variances of the two test results are statistically different. In this case, the student's t-test is used to compare the means of the two test results. Student's t statistic is calculated as follows:

$$t = \frac{u_a - u_c}{\sqrt{\frac{(n_a - 1)s_a^2 + (n_c - 1)s_c^2}{n_a + n_c - 2} \left(\frac{1}{n_a} + \frac{1}{n_c} \right)}} \quad (3.2)$$

The P-value associated with t-statistic is calculated using the built-in t cumulative distribution function, *tcdf*. Similarly, when the P-value is less than the significant level, α , it is concluded that the means of contractor and agency test results are not statistically different. When the P-value for the F test is equal to or larger than the significant level, the null hypothesis is rejected, and it is concluded that the two test results have unequal variance. In this scenario, Welch's test is used for comparing the means of the two test results. Welch's test defines the statistic t by the following formula:

$$t = \frac{u_a - u_c}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_c^2}{n_c}}} \quad (3.3)$$

If both the F-test and t-test indicate that the contractor and agency test results are not statistically different, the two sets of data are combined for acceptance and PF determination (Transportation Research Circular E-C235: Glossary of Transportation Construction Quality Assurance Terms, 2018); (MDSHA, Standard Specifications for Construction and Materials, 2022). Combining the two data sets results in a larger sample size and thus reduces the risks of erroneous pay. However, it should be noted that the agency and contractor test results can be combined to estimate PF only if the F-test and t-test show that they are from the same population. Otherwise, only the agency test results are used for estimating the PF for a lot.

3.2.2 Determination of Pay Factor (PF)

After validating the contractor against the SHA data, the next step is to determine the PF for the lot based on the quality of the delivered construction or materials which is measured by PWL in AASHTO's guide specification. The PWL is determined using the quality index, Q, which represents the distance, in standard deviation units, of the mean from the specification limits. Q can be calculated as follows:

$$Q_L = \frac{u - LSL}{s} \quad (3.4)$$

$$Q_U = \frac{USL - u}{s} \quad (3.5)$$

The Q values are used in conjunction with the sample size (i.e., n_a or $n_a + n_c$) to estimate the corresponding PWL_U and PWL_L (one-sided upper and lower specification limits) value using reference tables or normal distribution functions in MATLAB. For two-sided specification limits, the PWL is calculated as follows:

$$PWL = PWL_U + PWL_L - 10 \quad (3.6)$$

Once PWL is determined using Eq. 18-20, the corresponding PF is calculated. In this study, the following pay equation recommended by AASHTO (AASHTO) is used:

$$PF = 55 + 0.5 * PWL; PF = 0 \text{ for } PWL < 50 \quad (3.7)$$

3.2.3 Power of Statistical tests and PF Probability Curves

PF probability curves are developed to evaluate the pay performance in the long term. Such PF is also compared with the PF that is calculated using agency or contractor test results no matter if a difference is detected. The probability of receiving less than a certain pay factor can be determined as follows:

$$P_{(<PF)} = \frac{N_{(<PF)}}{N_T} \quad (3.8)$$

where $N_{(<PF)}$ = number of lots receiving less than certain PF, N_T = the total number of simulated lots. The PF for each simulated lot is calculated, and the PF probability curve

is developed to evaluate the pay performance in the long term. Such PF is also compared with the PF that is calculated using QA or QC test results no matter if a difference is detected.

The power of the statistical tests is estimated as follows:

$$P = \frac{N_D}{N_T} \quad (3.9)$$

where P = probability of detecting a statistical difference, N_D = number of lots detected a difference, N_T = total number of simulated lots.

Through the simulation analysis illustrated in Figure 3.2, the percentage of the lots passing F-test, t-test, or both F- and t-test in the long term can also be estimated by dividing the number of the numbers of lots passing the statistical test (i.e., F-test, t-test, or both F- and t-test) by the total number of lots. The sample size was varied such that the power surface and PF probability curve was developed to evaluate the power and PF for different combinations of contractor's and agency's sample sizes.

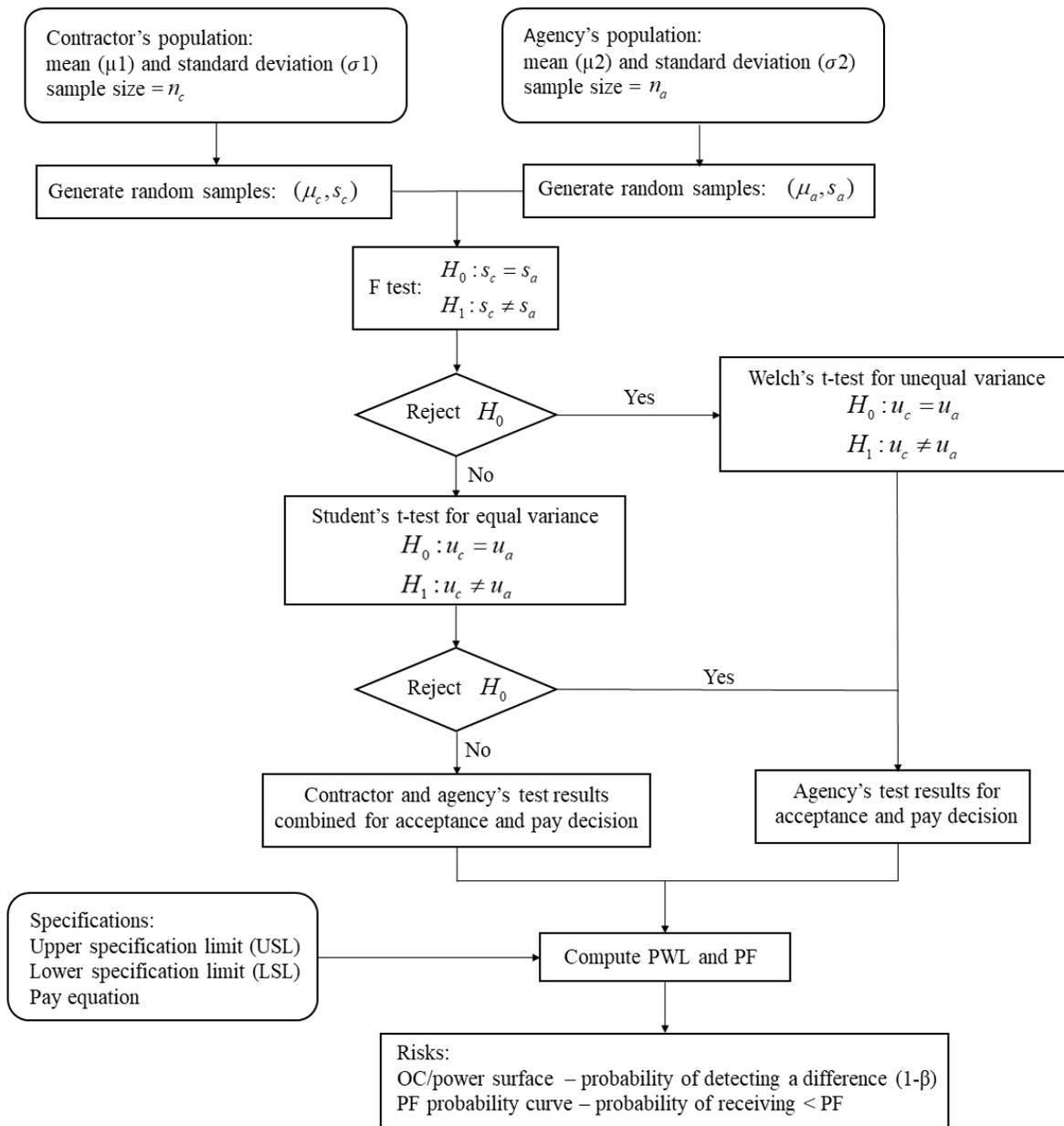


Figure 3.3 Monte Carlo based simulation for integrating risk and PF in verification procedures

3.3 Alternative Population Characteristics and Sample Sizes

To evaluate the impact of (i) sample size and (ii) differences in means and variances on the risk and associated PF the simulation was conducted with five different scenarios between the agency and contractor population distributions. In each scenario, the sample sizes varied from 3 to 20. These scenarios were based on the pavement data reported in NCHRP 10-79, Table 3.1 (Hughes et al., 2011). The data was collected from actual construction projects at a lot level, and with a lot size of two-lane miles and a sample size of 10. Figure 3.4 illustrates the five scenarios of agency and contractor testing populations encountered.

In the first scenario, the contractor and agency data have identical populations ($\mu_c = \mu_a, \sigma_c = \sigma_a$). In this case verification testing should support the null hypothesis (i.e., confirming that the two data sets are not from different populations, and thus contractor data can be used for acceptance and thus pay decisions). Thus, the probability of detecting a difference equals the level of significance, α . In the second scenario, the means of contractor and agency distributions are equal, but the standard deviation of the agency data is smaller than that of the contractor ($\mu_c = \mu_a, \sigma_c > \sigma_a$). The third scenario also considers equal means for the contractor and agency data; however, the agency population has a greater standard deviation than that of the contractor ($\mu_c = \mu_a, \sigma_c < \sigma_a$). This corresponds to the findings of some studies from the literature reporting that lower contractor's variability is often questionable, and thus will be affecting risks and PF.

Under the fourth scenario, the standard deviations are equal for both distributions, however, the mean of the agency distribution is larger than that of the contractor by 1.5 standard deviation units ($\mu_c \neq \mu_a, \sigma_c = \sigma_a$). In the last scenario, the mean of the agency population is less than that of the contractor by one time of contractor's standard deviation, while the standard deviation of the agency is half of the contractor's one ($\mu_c \neq \mu_a, \sigma_c \neq \sigma_a$). In this case verification testing will support the alternative hypothesis H_1 , and thus the contractor testing data cannot be used for acceptance. These five scenarios cover all possible situations between the contractor and agency distributions in the database (Hughes et al., 2011).

In this study, the range of sample sizes selected for the simulation were based on those reported for acceptance in the dataset collected from the national study, as well as those reported from additional SHAs quality assurance studies (Karimi et al., 2012); (Turochy and Parker, 2007); (Hughes et al., 2011). The review of these studies indicated that the sample sizes per lot used by the agencies can be grouped into three categories: (1) one sample per lot, (2) greater than two and less than twenty per lot, (3) greater than 20. However, statistical validation (i.e., F-test and t-test) against the contractor tests cannot be conducted when the agency performs only one test per lot. This puts significant risks on the agency of making erroneous acceptance and pay decisions. The minimum sample size for conducting a statistical test is three (Nimeri, 2019). It is also reported that large sample sizes, category (3) were achieved by pooling all the testing for a project as a single lot such that the statistical tests are too discriminating to be used for validation (Karimi et al., 2012).

Table 3.1 Means, standard deviations and specification limits (NCHRP 10-79)

Quality characteristics	Strength (psi)	Thickness (in.)	Roughness (in./mile)
Mean	4536.00	12.10	59.69
Standard deviation	509.90	0.54	8.69
Specification limits	LSL = 3500	LSL = 11.5, USL = 12.7	USL = 75

For each distribution the power was calculated for the sample sizes under consideration using equation 9, and the power surface was plotted. The USL and LSL illustrated in Figure 3.4 are used to calculate the quality index and PWL using Equations 3.4 to 3.6. The PF was calculated subsequently for each simulated lot. The PF probability curves were then plotted for selected combinations of sample sizes.

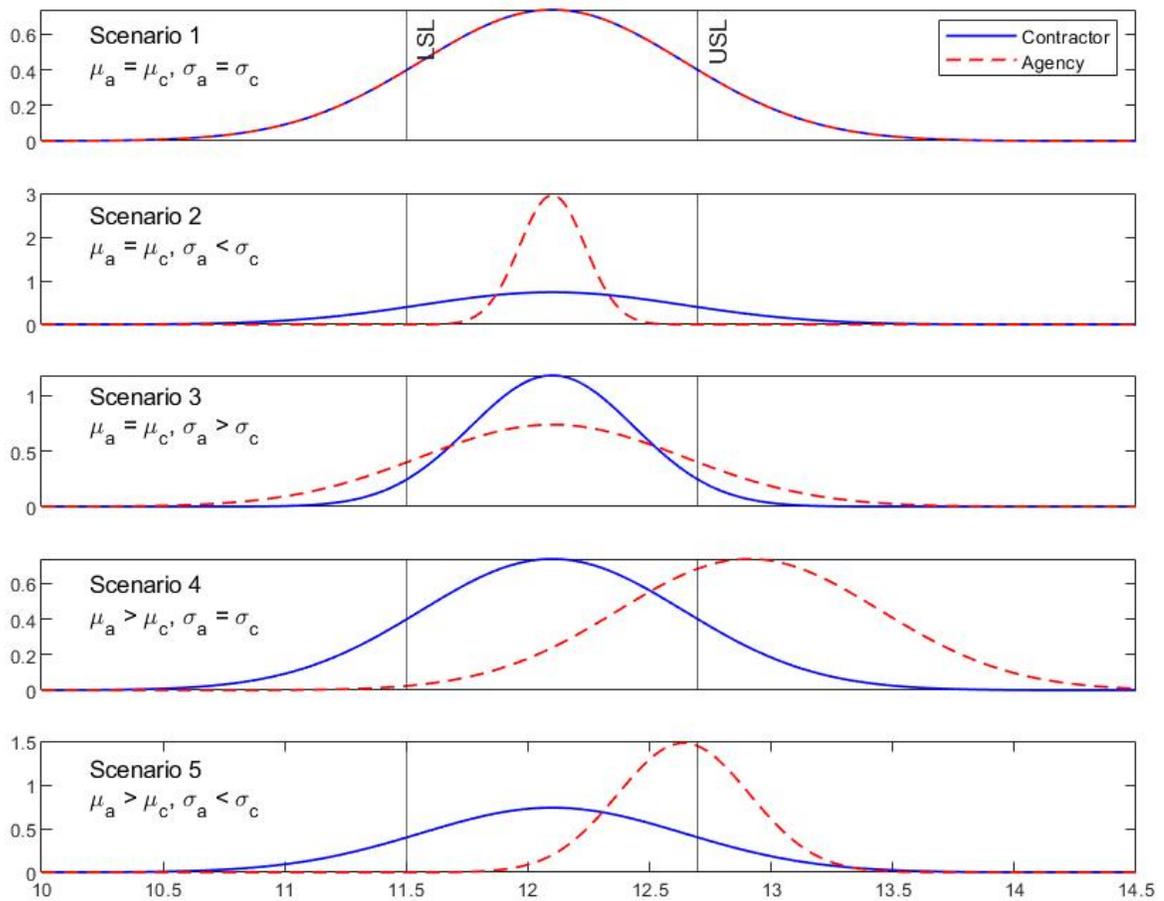


Figure 3.4 Alternative distribution scenarios in numerical simulation for pavement thickness

3.4 Results and Discussions

The simulation was performed for the five alternative distributions at a significant level, α , of 0.05. For each scenario, the power surface was developed. The power surface presents the probability of detecting a difference $(1-\beta)$ for various combinations of contractor

and agency samples sizes (i.e., n from 3 to 20). The PF probability curve was also plotted for each case, representing the probability of receiving a PF lower than the one indicated.

Figure 3.5 and Figure 3.6 show the power surface for the F-test and PF probability curves, respectively, for scenario 1. It can be observed that the F statistical power remains almost constant at 0.05 when the agency and contractor population characteristic are the same, Figure 3.5. This is expected since there is no difference in population means and variances. In this case the two hypotheses shown in Figure 2 will overlap and thus the power is equal to the selected level of significance, α , of 0.05. However, the PF probability curves show that as the sample size increases the curves become steeper, Figure 6. For instance, as the contractor and agency sample size increase from 3 to 5, the probability of receiving a PF less than 100% increases from 80.0% to 89.4% approximately, while the probability of receiving a PF less than 85% drops from 17.8% to 10.0%. As the sample size increases, the variability of the sample data decreases and thus the change in PF is reduced.

The power for the F-test and t-test, and average PFs in the long run, are summarized in Table 3.2. In this analysis, “in the long run” implies for all simulated lots. For a highway agency such analysis may be focused on all the data from projects in a construction season for which it is of interest to assess risks and PFs. In this scenario, increasing contractor or agency sample sizes increases the average PF associated with verification procedures. The PFs that are calculated using the agency or contractor test results, no matter if a difference is detected, are also reported in Table 3.2. The PFs calculated using the agency test results represent the cases where no contractor data are used in acceptance and pay decisions. In this situation, the contractor appears to be exposed to a greater payment risk than the agency.

For instance, when the contractor and agency have the same sample size of 3, the PF associated with verification procedure is 88.276% while the PF that is calculated using only the agency data is 79.710%. Combining agency and validated contractor test results for acceptance and payment determination, reduces the risk of incorrect decision and erroneous pay to both parties due to the resulting larger sample size. However, the contractor has always to take a 0.05 risk of incorrectly having the means or standard deviations declared different when in fact they are equal. One way to reduce the contractor's risk is to use a smaller α value, for example, $\alpha = 0.01$.

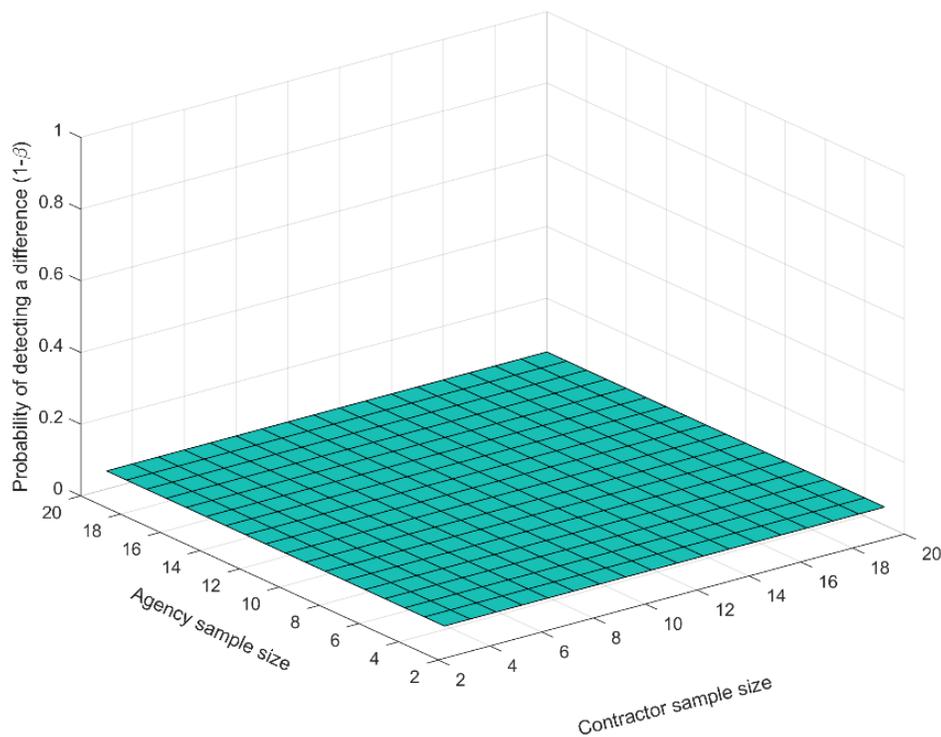


Figure 3.5 F-test power for scenario 1 (equal sample and population characteristics)

Table 3.2 Power and average PF in the long run for various combinations of contractor and agency sample sizes

Scenarios	Sample size		Probability of detecting a difference (1-β)			Average PWL, % (verification)	Average PF, % (verification)	Average PF, % (agency)	Average PF, % (contractor)
	Contractor (n_c)	Agency (n_a)	F-test	t-test	F or t-test				
$\left(\begin{array}{l} u_a = u_c \\ \sigma_a = \sigma_c \end{array} \right)$	3	3	0.050	0.050	0.099	74.062	88.276	79.710	79.340
	3	6	0.050	0.049	0.101	74.147	88.900	89.206	79.538
	7	3	0.049	0.050	0.099	74.170	90.470	80.016	90.550
	5	5	0.048	0.049	0.100	74.201	91.309	87.403	87.753
	10	10	0.050	0.050	0.103	74.334	92.009	91.705	91.698
$\left(\begin{array}{l} u_a = u_c \\ \sigma_a < \sigma_c \end{array} \right)$	3	3	0.290	0.050	0.352	94.500	102.219	105.000	79.340
	3	6	0.620	0.050	0.748	99.049	104.525	105.000	79.538
	7	3	0.319	0.049	0.368	89.465	99.719	105.000	90.550
	5	5	0.679	0.050	0.726	98.629	104.314	105.000	87.753
	10	10	0.974	0.051	0.990	99.976	104.988	105.000	91.698
$\left(\begin{array}{l} u_a = u_c \\ \sigma_a > \sigma_c \end{array} \right)$	3	3	0.072	0.050	0.157	80.942	91.790	80.017	100.870
	3	6	0.067	0.050	0.128	79.056	93.954	90.310	100.965
	7	3	0.151	0.051	0.304	79.425	86.286	79.741	101.577
	5	5	0.115	0.050	0.225	79.826	92.411	87.203	101.314
	10	10	0.257	0.050	0.338	78.545	93.911	91.601	101.633
	3	3	0.049	0.283	0.320	44.473	41.662	20.234	79.340

4 $\left(\begin{array}{l} u_a > u_c \\ \sigma_a = \sigma_c \end{array} \right)$	3	6	0.051	0.444	0.472	39.443	26.199	13.354	79.538
	7	3	0.050	0.437	0.489	45.505	43.248	20.438	90.550
	5	5	0.050	0.539	0.585	39.206	30.883	14.841	87.753
	10	10	0.050	0.886	0.895	35.439	13.313	8.565	91.698
5 $\left(\begin{array}{l} u_a > u_c \\ \sigma_a < \sigma_c \end{array} \right)$	3	3	0.094	0.235	0.337	64.135	73.749	59.441	79.340
	3	6	0.224	0.403	0.612	61.672	68.999	62.203	79.538
	7	3	0.092	0.303	0.340	64.758	75.185	59.081	90.550
	5	5	0.215	0.410	0.573	62.300	69.330	61.285	87.753
	10	10	0.499	0.754	0.902	59.759	67.270	66.123	91.698

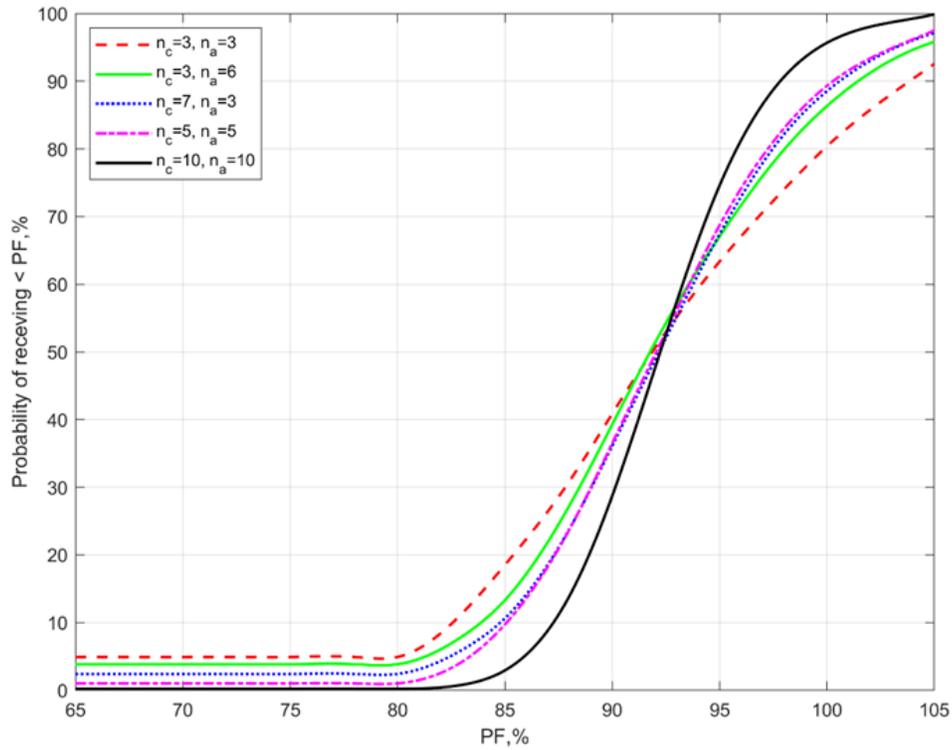


Figure 3.6 PF probability curves for scenario 1 (equal populations means and standard deviations) with various combinations of sample sizes, n

In scenario 2, since the contractor and agency mean are equal, the t-test statistical power remains constant at 0.05 despite an increase in sample size. Figure 3.7 shows the F-test power surface, while Figure 3.8 shows the PF probability curves for various combinations of contractor and agency sample sizes. The F-test power increases with increasing agency's sample size. In this scenario it can be observed that the limiting factor in how well the F-test can identify differences depends on the number of agency tests, Figure 3.7. For example, when the contractor sample size increases from 3 to 7, the F-test power increases from 0.290 to 0.319 with an agency sample size of 3. However, the F-test power increases significantly from 0.290 to 0.62 as the agency samples increase from 3 to 6 with a contractor sample size of 3 (Table 3.2). The probability of detecting a difference for the F-

test or t-test (i.e., the probability of detecting a difference in means or variances) is also reported in Table 3.2. Similarly, such probability increases as the sample size increases, and is more affected by the agency's sample size. Since the agency usually has fewer tests than the number of contractor tests, the number of agency tests will determine the F-test power to identify variability differences when they do exist. Therefore, in this scenario, the agency's risk of not detecting a difference that exists is significant when the agency sample size is small (e.g., $n_a=3$). For example, the β risks is 0.681 for an agency sample size of 3 and a contractor sample size of 7.

In terms of PF, Figure 3.8 shows that the probability of receiving a PF less than 100% decreases as the agency sample size increases, while such probability increases as the contractor sample size increases. The average PF increases as the agency sample size increases (Table 3.2). For instance, it can be observed that the average PF increases from 102.219% to 104.525% as the agency sample size increases from 3 to 6 while the contractor sample size remains constant at 3. Such a relatively large increase in PF is caused by the reduction of variability in sample data due to an increase in sample size. As the sample size increases, the F-test power increases meaning more lots are detected with a difference in standard deviation. The PFs for these lots are calculated using agency data which produces a greater PWL, compared to contractor data, and thus resulting in a larger PF. This poses a financial risk to the agency because it seems to pay for more than the actual quality. Such risk is reduced slightly as the contractor sample size increases. For example, when the contractor sample size increases from 3 to 7, the PF associated with the verification

procedure decreases from 102.219% to 99.719%. Thus, in this scenario the agency would seem to be exposed to a greater payment risk than the contractor.

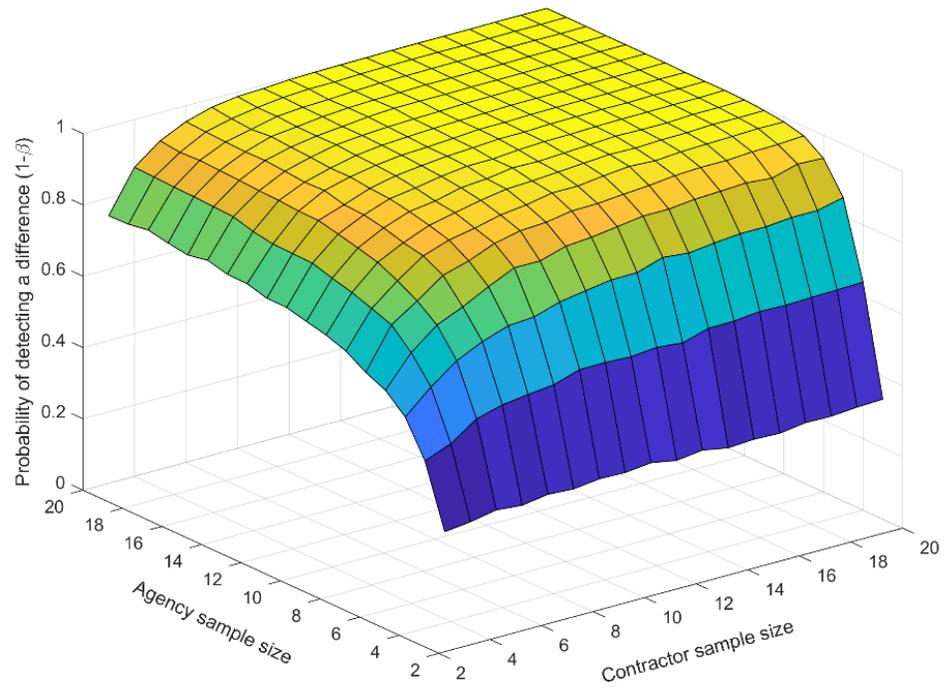


Figure 3.7 F-test power for scenario 2 for various combinations of contractor's and agency's sample sizes

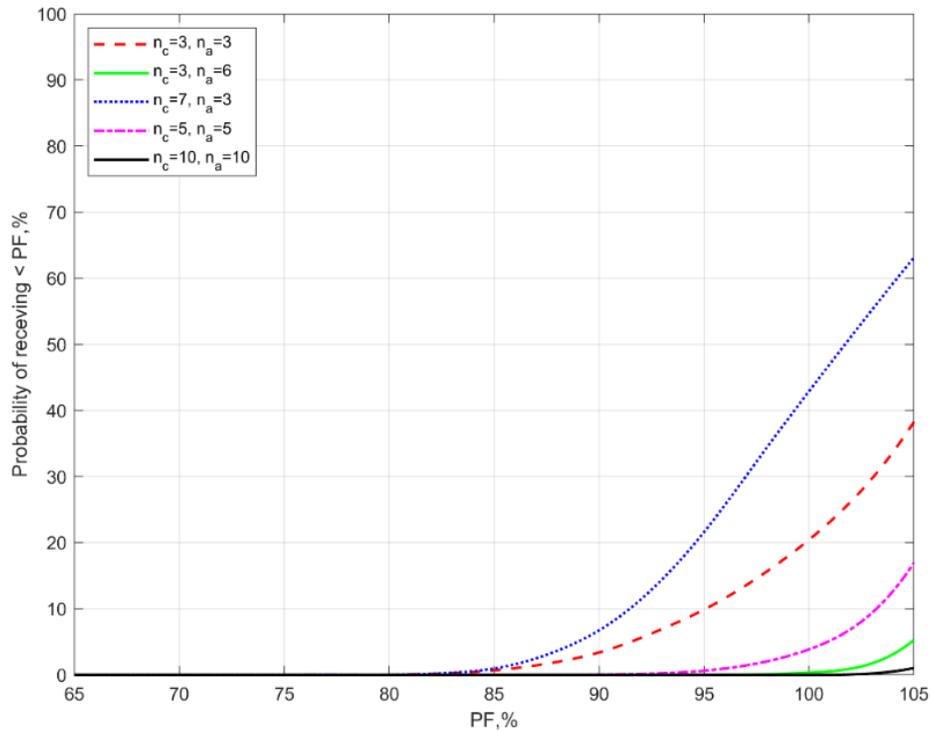


Figure 3.8 PF probability curves for scenario 2 for various combinations of contractor's and agency's sample sizes

Figure 3.9 presents the F-test power surface for scenario 3 in which the agency's testing data have greater variability (i.e., standard deviation) than that of the contractor. Typically, when contractor data is used as part of the acceptance, contractor testing would be more frequent than that of the agency. If the variability in agency's data is larger (i.e., scenario 3), then the contractor sample size will determine how well the F-test will be able to identify differences. However, when the agency's data variability is smaller than the contractor's (i.e., scenario 2), the F-test power is limited by the agency sample size. Therefore, one of the potential weaknesses of the F-test is that the F-test power does not

improve much as the ratio of n_c and n_a increases for the case when the agency's variability is smaller than that of the contractor.

Figure 3.10 presents the PF probability curves for scenario 3. In this case, the probability of receiving PF less than 100% increases as the agency sample size increases. However, as the contractor sample size increases from 3 to 7, the probability of receiving a PF less than 100% remains the same. It can be observed from Table 2 that the average PF increases from 91.790% to 93.954% as the agency sample size increases from 3 to 6 while the contractor sample size remains constant at 3. On the contrary, the average PF decreases from 91.790% to 86.286% when the contractor sample size increases from 3 to 7 with the same agency sample size of 3. This indicates that PFs are influenced to a greater extent by the test data (contractor or agency) that has a greater sample size.

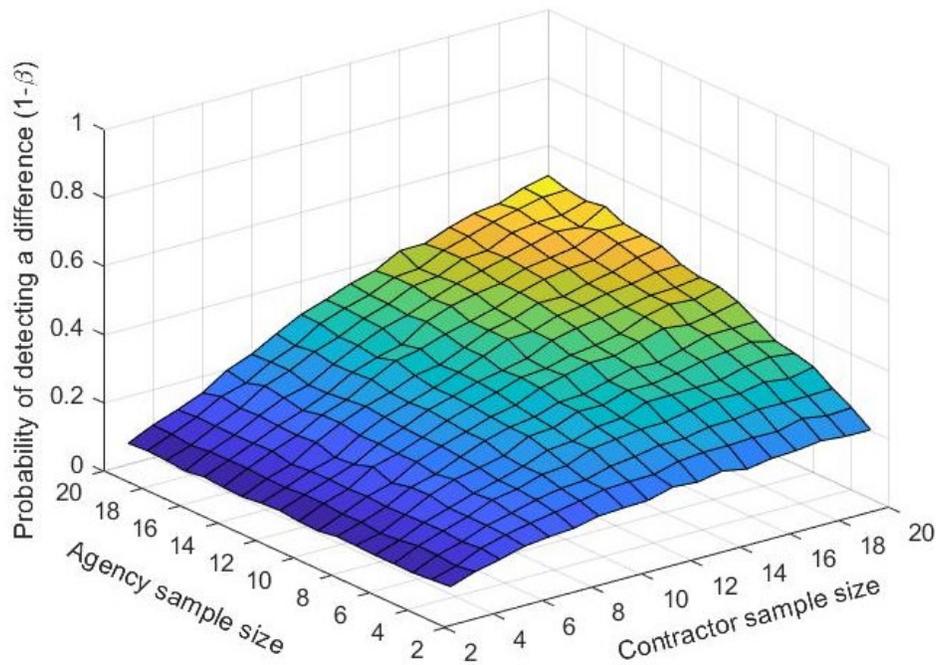


Figure 3.9 F-test power for scenario 3 for various combinations of contractor's and agency's sample sizes

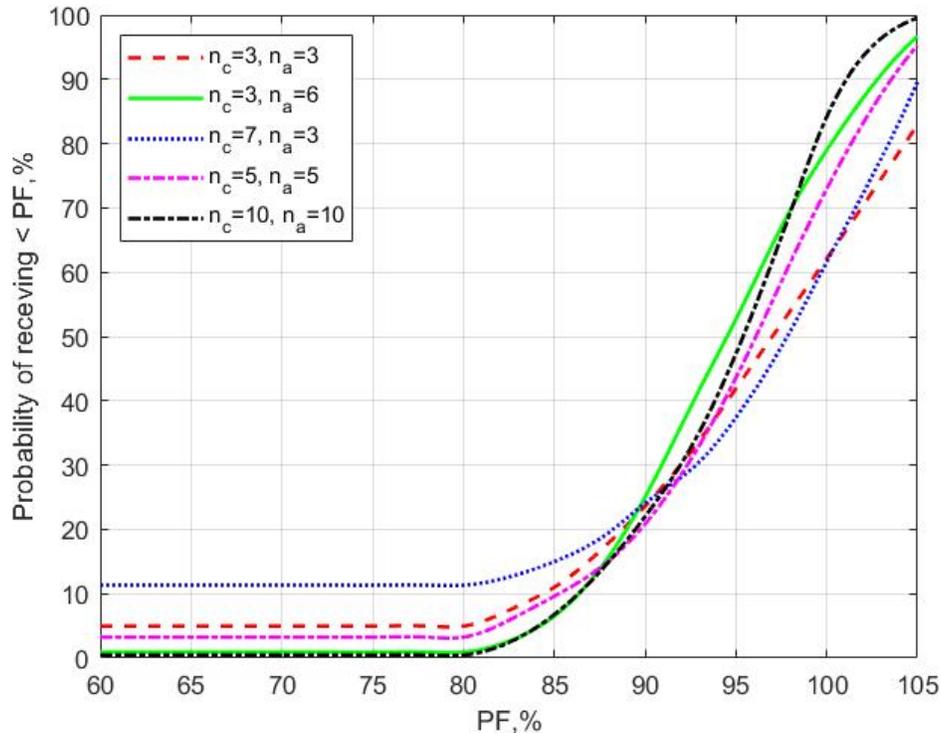


Figure 3.10 PF probability curves for scenario 3 for various combinations of contractor's and agency's sample sizes

When the variability of the contractor sample is smaller than that of the agency (i.e., scenario 3), the PF associated with the verification procedure tends to be favorable to the agency. For example, when the agency and contractor sample sizes are 3 and 7, respectively, the PF associated with the verification procedure is 86.286%. This is closer to the PF (79.741%) calculated using agency's data than when using contractor data (with a PF of 101.577%). This poses a great financial threat to the contractor because it could ultimately get paid for less than the actual quality. More importantly, in this case the contractor has a higher testing frequency which should more accurately reflect the actual population quality.

Figure 3.11 shows the t-test power surface for scenario 4. It can be observed that both agency and contractor sample sizes have a significant effect on the t-test power. The power of the F-test remains almost constant around 0.05 regardless of the contractor and agency sample size. It can be observed that the t-test power is limited to the larger sample size. For example, with a constant agency sample size of 3, the power test increases to 0.8016 when the contractor sample size increases to 20. Similarly, equal contractor and agency sample sizes will maximize the t-test power. For example, as shown in Table 3.2, the t-test power is estimated to be 0.489 with a contractor and an agency sample size of 7 and 3 respectively. The power increases to 0.585 when the contractor and agency have the same sample size of 5.

In scenario 4, the PWL based on the agency population is much smaller than that when the contractor population is considered. Thus, the calculated PFs are smaller than those in the other scenarios, Table 3.2. As shown in Figure 3.12, the probability of receiving less than a certain pay factor increases as the agency sample size increases. However, the probability of receiving less than a certain pay factor decreases slightly when the contractor sample size increases from 3 to 7 (i.e., $n_c=3, n_a=3$ and $n_c=7, n_a=3$). As the contractor sample size increases, the PF associated with the verification procedure increases. For instance, when the contractor sample size increases from 3 to 7, the average PF from the verification procedure increases from 41.662% to 43.248%, (Table 3.2), when the agency sample size remains equal to 3. On the other hand, the PF associated with the verification procedure decreases when the agency sample size increases. For example, the average PF decreases from 41.662% to 26.199% as the agency sample size increases from 3 to 6 while the

contractor sample size remains constant at 3. In this scenario, the contractor would seem to be exposed to a greater payment risk than would the agency because the PF associated with the verification procedure is always in favor of the agency despite sample sizes.

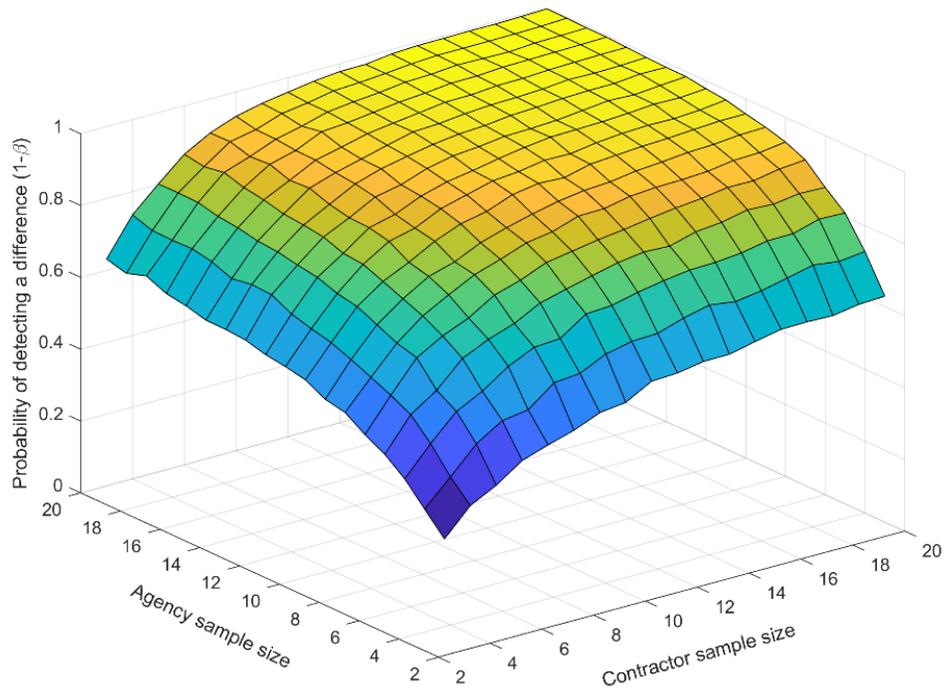


Figure 3.11 t-test power for scenario 4 for various combinations of contractor's and agency's sample sizes

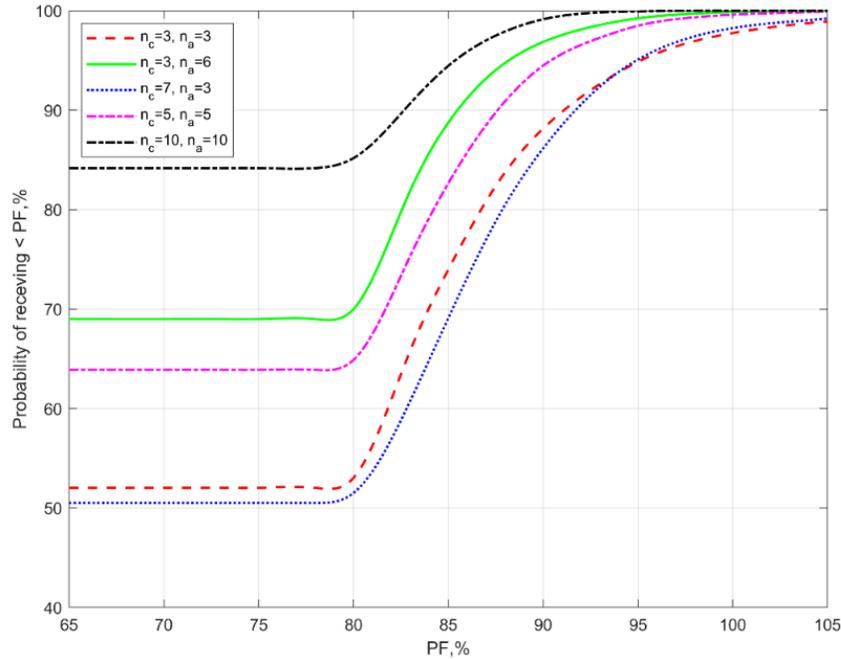


Figure 3.12 PF probability curves for scenario 4 for various combinations of contractor's and agency's sample sizes

In many projects it may be possible that the contractor and agency populations have different means and standard deviations, which is represented in scenario 5. In this case, the standard deviation ratio between agency and contractor is 2, while the means' difference is 1 (i.e., in standard deviation units). Figure 3.13 and Figure 3.14 present the F-test and t-test power, respectively. Figure 3.13 indicates that the number of agency tests determines the effectiveness of the F-test regardless of the means difference. For a contractor sample size of 3, the F-test power is only 0.094 in relation to an agency sample size of 3. The power increases to 0.313, and 0.344 when the agency sample size increases from 10 to 20

respectively. The highest power observed is 0.838 when both contractor and agency sample sizes are equal to 20.

As presented in Figure 3.14, the t-test power is influenced by both contractor and agency sample sizes. When the sample size for the contractor increases from 3 to 20, and for the agency's sample size of 3, the power increases from 0.245 to 0.359. The power surfaces discussed above were focused to either F-test or t-test. However, SHAs usually combine both F and t tests in their verification procedure as shown in Figure 3.3. Thus, the probability of detecting a difference in means and/or variances was evaluated using simulation. This probability is estimated with Equation 3.9. In this case, N_D is the number of lots that a difference in means and/or variances is detected. Thus, such probabilities for scenario 5 are plotted in Figure 3.15 for various combinations of contractor and agency sample sizes. In this scenario the power of detecting a difference largely depends on the number of agency tests. The simulation results show that the power in this case is not simply the sum of the F-test and t-test power. For example, the F-test and t-test power are 0.215 and 0.410 when the contractor and agency sample sizes are equal to 5, while the power of detecting a difference in means or variances is only about 0.573, Table 3.2. Figure 3.15 can be used to evaluate the power of the verification procedure that includes both F-test and t-test.

From the analyses of the five scenarios considered in this study, it can be observed how the statistical power is affected in relation to the contractor and agency population distributions (i.e., means and variability). For a specific combination of contractor and agency sample sizes, the further apart are the two populations from each other, in both means and standard deviations, the greater the power. In statistics such effect is identified as the

“effect size” when assessing how meaningful the relationship between two groups in this case is. Thus, when the two distributions representing the contractor and agency data are further apart fewer number of tests are needed to detect the statistical difference between them. Highway agencies need to conduct such analysis based on their population characteristics to make informed decisions regarding the verification testing frequency for a given power and level of significance. Such analysis could eventually identify the level of testing needed in each scenario, and thus reduce testing time and cost pertinent to acceptance.

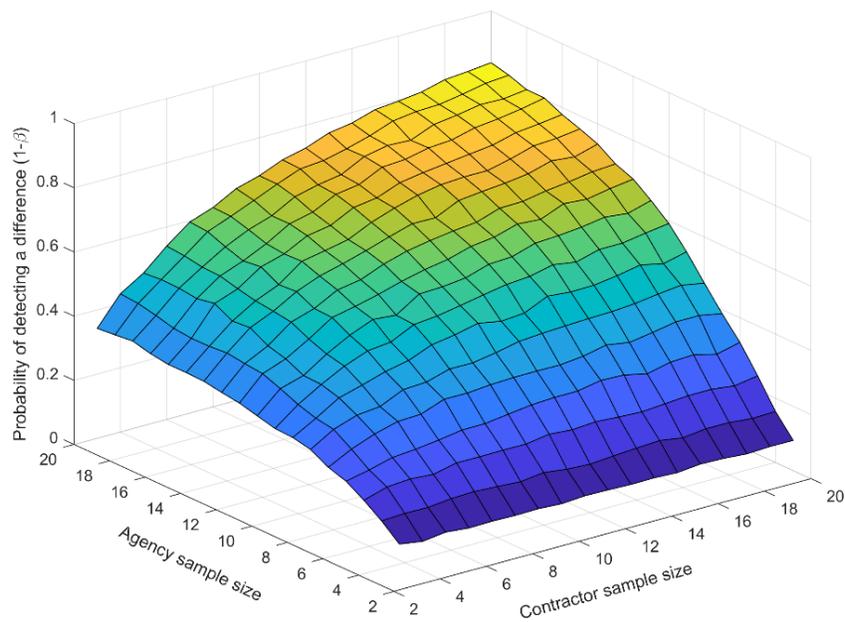


Figure 3.13 F-test power for scenario 5 for various combinations of contractor's and agency's sample sizes

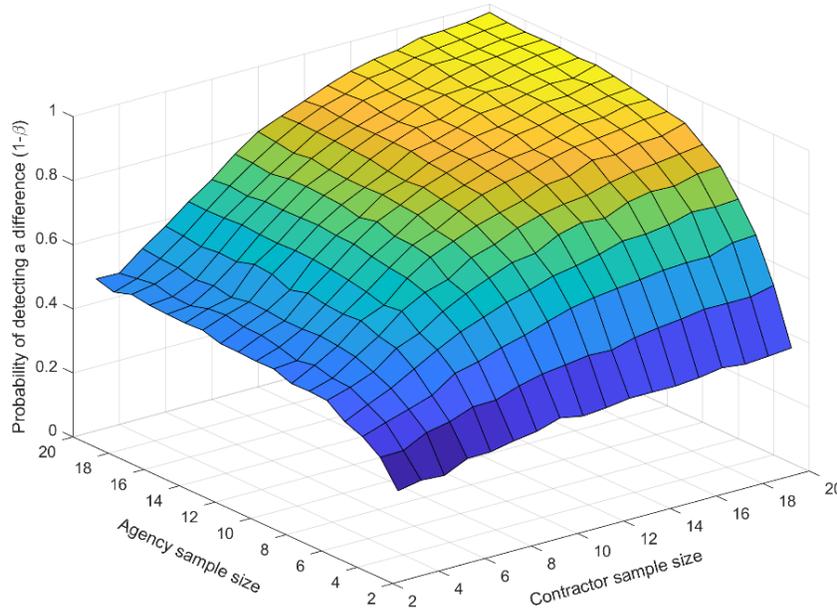


Figure 3.14 T-test power for scenario 5 for various combinations of contractor's and agency's sample sizes

The PF probability curves are plotted in Figure 3.16, while the average PFs for various combinations of sample sizes are summarized in Table 3.2 for scenario 5. As shown in Table 3.2, when contractor and agency have the same number of tests, equal to 3 for example, the average PWL is estimated to be about 64.135% and the average PF is 73.749%. However, in this case there is a 11.70% probability that the contractor receives a PF equal to or greater than 100% (i.e., 1- probability of receiving >100%). The PF probability curves and the average PFs in Table 2 clearly show how the contractor and agency sample sizes influence PF when validated contractor tests, in combination with the agency verification tests, are used for pay decisions. A small number of contractor and agency tests (i.e., $n=3$)

result in a greater difference between the PF associated with the verification procedure and the PF calculated using only the agency data. Such a difference decreases with increasing sample size of both agency and contractor. Since most of the highway agencies have fewer verification tests than the number of contractor tests, in these cases if the contractor test results are not verified and the agency data is used to determine payment, the risk of incorrect decision and erroneous pay is higher due to smaller sample sizes. For instance, for the case of $n_c = 7, n_a = 3$ the PF associated with the verification procedure is 75.185% while the PFs calculated from the agency and/or the contractor data are 59.081% and 90.550%, respectively. If the agency uses its own data to determine PF, the contractor will be exposed to a significant risk because could ultimately be paid for less than the actual quality.

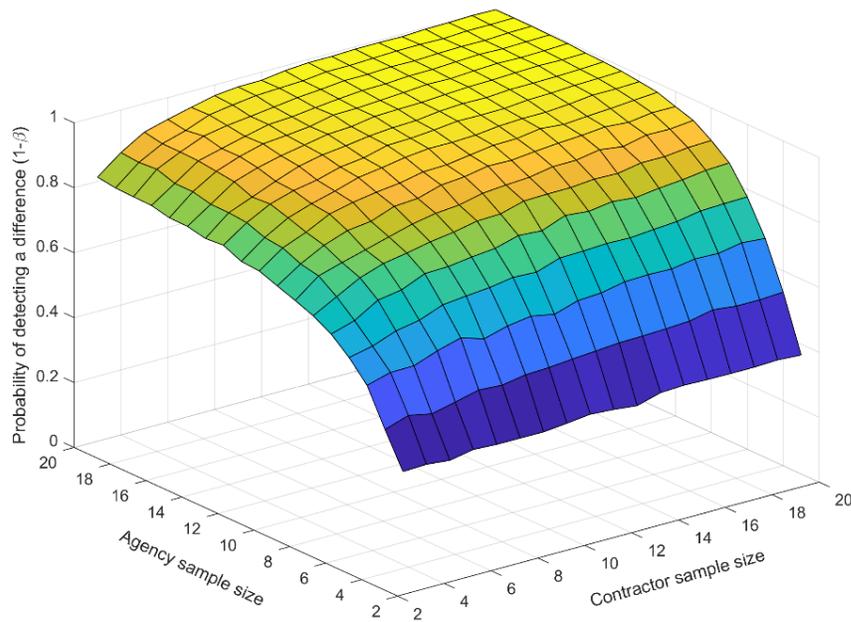


Figure 3.15 Probability of detecting a difference in means and/or variances for scenario 5 for various combinations of contractor's and agency's sample sizes

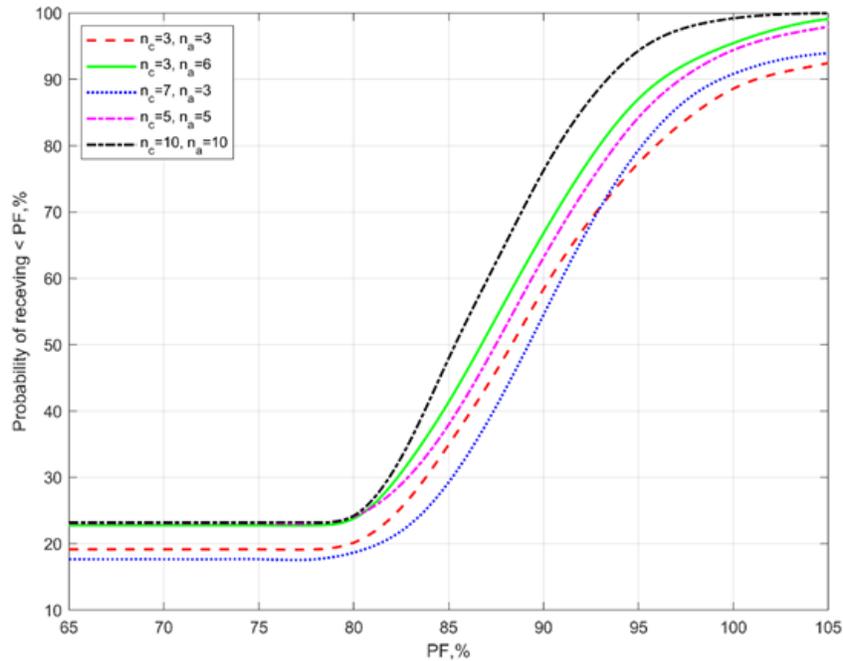


Figure 3.16 PF probability curves for scenario 5 for various combinations of contractor's and agency's sample sizes

3.5 Conclusions and Recommendations

A Monte Carlo simulation process was developed to assess acceptance risks and associated PFs when contractor test results are used in combination with the agency data in acceptance of pavement materials and construction. Alternative scenarios of contractor and agency data populations were considered for evaluating such effects. In each scenario, the testing statistical power was determined, and the power surfaces were developed considering alternative sample sizes. The PF probability curves were developed as well and the average PFs were calculated. Since the proposed analysis quantifies risks and PF in relation to the

population characteristics of the quality acceptance data of interest, such approach could be eventually used in identifying (i) level of risks with current acceptance specifications, and (ii) level of testing needed in each scenario, and thus reduce testing time and cost pertinent to acceptance.

The overall study results indicated that if the variability of the agency test data is larger than that of the contractor (i.e., scenario 3), the contractor sample size will determine how well the F-test will be able to identify differences between the samples. On the contrary, when the variability of the agency test results is smaller (i.e., scenario 2), the probability of detecting a difference in variance (F-test) depends on the number of agency verification tests. Both agency and contractor sample sizes have a significant effect on the t-test power. The limiting factor in how well the t-test can identify differences in means depends on the sample size difference between the contractor and the agency data. Thus, agencies will need to consider this fact when making the decision concerning the balance between acceptable risk levels and sample size.

The PF analysis in this study reveals the impact of the verification procedure on risks and rewards. In scenario 1, combining agency and validated contractor test results for acceptance and payment determination reduces the risk of incorrect decision to both parties. However, as shown from the analysis, the impact of using larger sample size (i.e., combining contractor and agency data) produces a higher risk to the contractor (i.e., at a 5% level). This is associated with incorrectly concluding, in statistical terms, that the means and/or standard deviations between the agency and contractor data are different when in fact they are equal. One way to reduce the contractor's risk is to use a smaller α value, for example, $\alpha = 0.01$.

When the variability in agency data is smaller than that of the contractor (i.e., scenario 2), the agency would seem to be rewarding a greater payment than that calculated when the contractor quality data is used. Such effect is reduced as the contractor sample size increases since the calculated sample variance is decreasing. In the case where the variability of the agency data is greater than that of the contractor, the PF associated with the verification procedure tends to be lower. In this case the agency may question the level of reward since using contractor data will produce a higher PF. While in most cases the contractor has a higher testing frequency, such data may more accurately reflect the actual population characteristics of production, assuming there are no concerns in regards to testing data quality and integrity.

The overall and specific findings of this study could provide useful input to SHAs in evaluating acceptance risks and associated PF. Therefore, it is suggested that SHAs currently implementing verification procedures in their QA program strongly consider performing similar analysis using the proposed simulation analysis approach in order to: (i) evaluate the PF and associated risks; and (ii) make necessary adjustments on sample size and pay equations to balance the risks and properly award pertinent payments. For contractors the analysis approach presented here in can be useful in identifying strategies on how to reduce risk pertinent to rejection of good quality material in terms of adjusting frequency of testing, sample size, production quality and uniformity; and how to adjust production quality for increasing PF. Finally, the methodology developed herein is transferable to agencies where contractor data are used in acceptance and pay decisions

CHAPTER 4: LIFE CYCLE ASSESSMENT OF USING RECYCLED MATERIALS IN PAVEMENT

4.1 Introduction

The United States' national highway system requires new construction and extensive rehabilitation of highways to meet the growing traffic demand and guarantee the safety of drivers (Lee et al., 2013). The maintenance and rehabilitation of this extensive highway network system consumes large amounts of natural resources and energy, produces large quantities of waste and generate significant amounts of greenhouse gas emissions (Lee et al., 2010). Thus, state Departments of Transportation (DOTs) have been aiming to adopt an ever-increasing amount of recycled materials in pavement construction and rehabilitation. The benefits that can be achieved by using large amounts of recycled materials include reducing the use of natural resources; eliminating waste materials generated for disposal; reducing energy and water consumption; and reducing greenhouse gas emissions (National Academies of Sciences, Engineering, and Medicine, 2011).

Recycled asphalt pavement (RAP) is the most common recycled material used in hot mixed asphalt (HMA) and to some degree as aggregate in base layers. The engineering properties of using RAP in highway construction applications have been explored by several studies (Chesner et al., 1998). However, its contributions to sustainability, in terms of greenhouse gas emissions reduction (GHG), energy and water demand reduction, and economic benefits using life-cycle cost analysis (LCCA), have been examined to a lesser degree. Furthermore, the potential widespread use and the sustainability benefits of RAP and

other recycled materials need to be considered within the pavement management system (PMS) analysis when identifying the best rehabilitation strategy for each project. Therefore, the objective of this chapter was to estimate the potential economic and environmental impacts of using recycled materials in pavement construction and rehabilitation through life-cycle analysis (LCA). Firstly, we proposed a methodological framework for assess the life cycle economic and environmental benefits of using recycled materials/by-products. The proposed methodology considers all life cycle stages of pavements including material production, transportation, construction, maintenance, rehabilitation and end of life. To demonstrate the value of the suggested approach, this study conducted LCA of using two recycled materials in pavement. The first case study considered construction and demolition waste (CDW) recycled aggregates as an alternative to natural aggregates for roadway construction, while the second case conducted life cycle sustainability assessments of using rock dust as partial replacement of sand and cement in concrete pavements.

4.2 Methodology

The proposed methodology for assessing the life cycle environmental and economic impacts with the use of recycled materials/by-products in pavement construction and rehabilitation includes the steps of Figure 4.1. The primary objective of encouraging the use of recycled materials in the construction of highways is to reduce economic cost and minimize environmental impacts without comprising the performance. Thus, the mechanical properties, such as compressive strength and elastic modulus, as well durability of concrete

with recycled materials need to be examined. Once the engineering properties requirements are evaluated, the next step is to conduct a site-specific survey or condition assessment: for a new roadway construction project, the survey will require information on project traffic and climate inputs, construction materials and processes; while for a rehabilitation project, condition assessment of existing roadway needs to be performed. This step provides information for selecting the best materials and construction techniques and/or identifying what level of existing materials can be recycled along with the recycling method, e.g., cold in-place recycling (CIR), hot in-place recycling (HIR), full-depth reclamation, use of ex-situ recycling (Gschösser et al., 2012).

The objective of step 3 is to identify the reference conventional and the alternative sustainable strategies. The reference strategy considers conventional (virgin) materials throughout the life cycle of the pavement structure, while alternative strategies use recycled materials or recycling techniques. The reference strategy is used for comparative analysis (step 7) in assessing and comparing against, and in between them, the alternative sustainable strategies in terms of cost and environmental impacts. The next step is related to pavement structural design for both reference and alternative strategies. The pavement structures (layers and thicknesses) are first determined. Since the concrete mix with recycled materials may have different mechanical properties compared to virgin materials, the equivalent layer thicknesses for the alternative strategies need to be determined using pavement analysis tools such as the 1993 AASHTO pavement design guide (AASHTO, 1993), the mechanical-empirical pavement design guide (MEPDG) (AASHTO, 2020), or local agency design procedures. Furthermore, in order to identify appropriate rehabilitation strategies, the service

life needs to be estimated depending on the initial design quality and the minimum acceptable performance condition, considering (i) material properties, (ii) layer characteristics, (iii) traffic load, and (iv) climatic conditions (Zhao et al., 2021). In this study, the 1993 AASHTO pavement design guide was employed for the structural design and performance prediction of the pavements with different materials, considering a minimum present serviceability index (PSI) of 2.5 as the lower acceptable condition (Stroup-Gardiner, 2011).

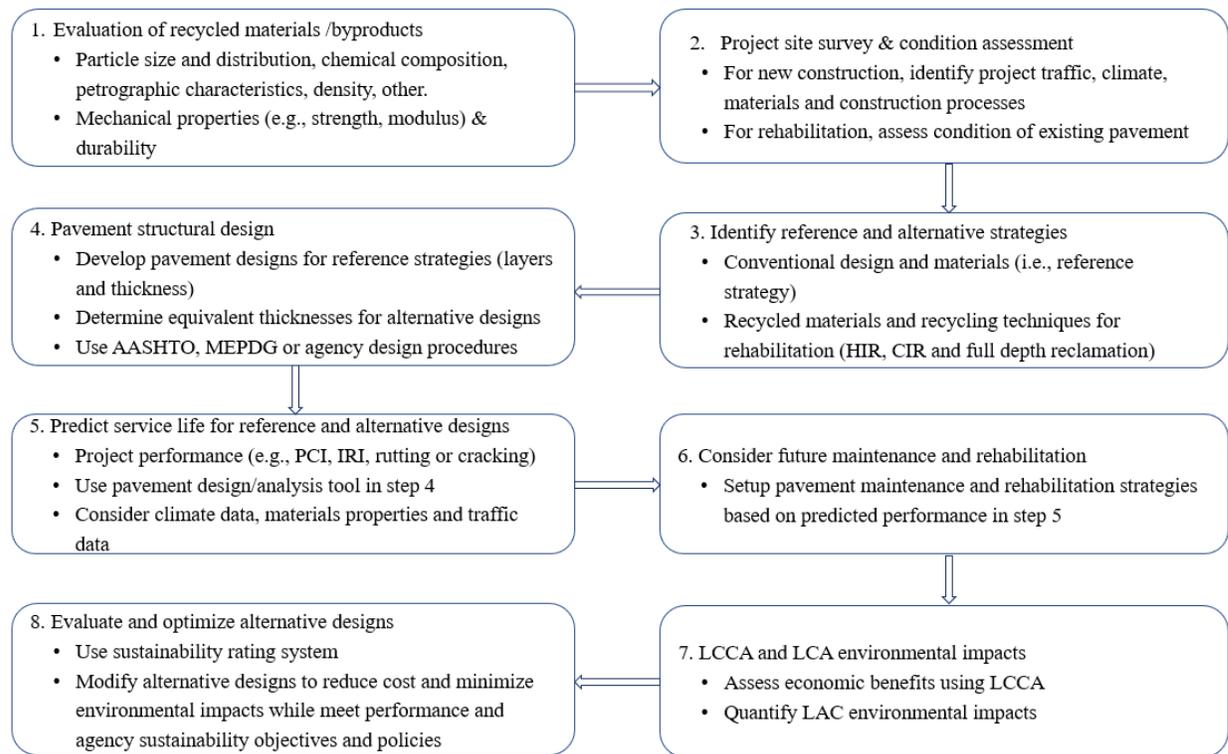


Figure 4.1 Methodological framework for generating and evaluating alternative sustainable strategies

The following step is to conduct the life cycle economic and environmental assessment. LCA models can be used as a sustainability tool given the flexibility of the

methodology in providing a holistic analysis of the environmental and economic impacts of different recycled materials and processes on pavement construction (Nathman et al., 2009). While any LCA tool available may be used, in this study the pavement lifecycle assessment tool for environmental and economic analysis (PaLATE) was used to evaluate the economic and environmental impacts of construction materials and processes for the specific roadway project (PaLATE, 2022). PaLATE is a project-level LCA tool that considers all life cycle phases of pavements (e.g., materials processing, transportation, construction, maintenance and end of life). Three categories of data are used in PaLATE: environmental related data (e.g., emissions factors, energy and water consumption associated with material production, equipment and processes), cost data (materials and processes), design related data (e.g., layers, thicknesses, transport distances, etc.). Figure 4.2 presents the data used in each stage of the life cycle analysis (materials production, transportation, initial construction and maintenance, and end life phases). The input data and calculations within this LCA tool are easily updated to reflect current emission models, local costs and site conditions. Thus, in this study updated emission parameters were used following the Environmental Protection Agency (EPA) in-put-output model, USEEIO, (2022). The design parameters represent typical roadway construction practices for average traffic volumes in Poland. The material costs, labor costs and overhead rates were collected from local contractors, while typical construction, maintenance and transportation costs were based on typical construction projects in the region. PaLATE outputs relate to the life cycle inventory (cost, energy, water consumption, emissions etc.) as shown in Figure 4.2. Such LCA analysis provides an understanding of where environmental impacts are created in the

life cycle of pavements, as well as how and to what extent various sustainability strategies reduce those environmental impacts and identifies potential unintended consequences that can result in increased environmental impacts. Once the LCA analysis is completed, a sustainability rating system should be used to evaluate each alternative in terms of its effectiveness to meet sustainability targets (BE²ST-in-Highways, 2012).

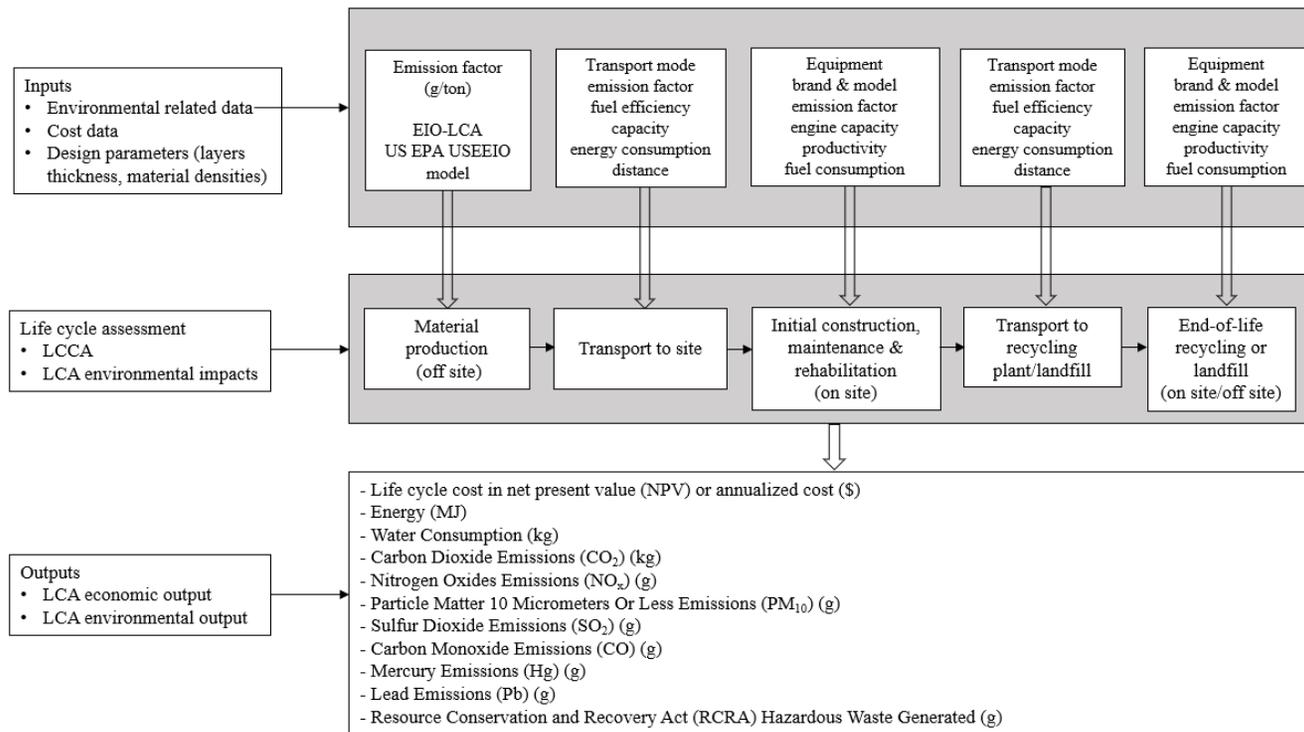


Figure 4.2 Inputs, outputs of the LCCA and LCA environmental impact analysis

4.3 Life-Cycle Economic and Environmental Impacts of CDW Recycled Aggregates in Roadway Construction and Rehabilitation

4.3.1 Literature Review

The construction and rehabilitation of road pavements involve large amounts of natural resources such as raw materials and energy (Amiril et al., 2014); (Gambatese & Rajendran, 2005); (Muench, 2010). However, when recycled materials are used in roadway constructions potential environmental and economic benefits can be assessed (Lee et al., 2010); (Celauro et al., 2015); (Hasan et al., 2020); (Del Ponte et al., 2017); (Blankendaal et al., 2014). A remarkable variety of studies have investigated several recycled materials to be used in road pavements: recycled aggregates (from different sources), clayey materials, industrial by-products (slags, pulverized fuel ash, etc.), plastic and rubber wastes, etc. (Sherwood, 2001); (Abukhattala, 2016); (Kaseer et al., 2019); (Thives & Ghisi, 2017); (Siddique et al., 2008); (Tam et al., 2018); (Reis et al., 2021). Among these materials, aggregates recycled from debris of construction and demolition waste (CDW) play a key role in the sustainability of road infrastructures (del Río Merino et al., 2010); (Ossa et al., 2016); (Kumbhar et al., 2013). CDW aggregates have been recognized as a valid alternative to natural aggregates (NA) for road pavement applications (Silva et al., 2019); (Jimenez et al., 2012). A great number of studies and practical applications recognize that recycled CDW materials can be employed in embankments (Cristelo et al., 2016); (Zhang et al., 2019) and trenches of roadworks (Rahman et al., 2014); (Vieira et al., 2017) as well as in the construction of subgrade layers (Zhang et al., 2020); (Cabalar et al., 2019). However, when employed in base and subbase layers, CDW aggregates are usually stabilized to meet

mechanical and durability of specification and standard requirements (Esfahani, 2020); (Tefa et al., 2021). To reduce ecological footprint deriving from the use of ordinary Portland cement (OPC), some studies recommended to stabilize CDW materials with alternative binders (Arulrajah et al., 2017b) ; (Bassani et al., 2019). Blended cements with supplementary cementitious materials (Agrela et al., 2012), industrial by-products (Arulrajah et al., 2017a); (Camargo et al., 2013) and clinker-free cementitious binders (i.e., those deriving from the alkali-activation of by-products and/or waste) (Mohammadinia et al., 2018); (Bassani et al., 2019) are the current alternative binders for stabilization purposes. In this context, (Bassani et al., 2017) showed that CDW aggregate mixtures can be stabilized with cement kiln dust (CKD) in alternative to OPC, reaching comparable results of unconfined compressive strength (UCS) and resilient modulus (RM).

In addition to the proven feasibility in using CDW aggregates in substitution of NA for road pavement formation, their implementation is strongly promoted by well-recognized environmental benefits (Dahlbo et al., 2015). Several life-cycle assessment (LCA) studies have demonstrated that the recycling of CDW (*i*) reduces emission of environmentally harmful substances, (*ii*) avoids the exploitation of natural resources, and (*iii*) decreases the consumption of energy in comparison with the production of virgin NA (Blengini & Garbarino, 2010); (Hossain et al., 2016); (Simion et al., 2013). Researchers agree that this advantage mainly derive from the landfilling avoidance (Carpenter et al., 2013); (Rosado et al., 2017); (Marzouk & Azab, 2014); (Ram et al., 2020). Economic savings deriving from the adequate management of CDW in the civil sector are estimated as well (Lee et al., 2010); (Yuan et al., 2011); (Rodríguez et al., 2015); (Coelho & De Brito, 2013).

LCA studies on CDW materials are mostly focused on the evaluation of environmental impacts of different recycling strategies in comparison to landfilling (Coelho & De Brito, 2013); (Bovea & Powell, 2016). These studies are limited to the analysis of recycling processes from the demolition stage to the production of the recycled (end-of-waste) product (e.g., cradle-to-gate LCA approach). More work is effectively needed to extend the environmental impact assessment to real applications in which CDW material is included in substituting NA. Some LCA analyses have investigated the environmental benefits of using recycled CDW aggregate in concrete production (Borghetti et al., 2018); (Doshio, 2007); (Ding et al., 2016a); (Serres et al., 2016). Almost all the road pavement LCA-related studies consider the inclusion of recycled and/or alternative materials in substitution of traditional ones for asphalt and concrete layers of flexible and rigid pavements respectively (Shi et al., 2019); (Chiu et al., 2008); (Anastasiou et al., 2015); (Celauro et al., 2015); (Aurangzeb et al., 2014); (Vidal et al., 2013). Only a limited number of studies focused on the environmental assessment deriving from the use of CDW aggregate as granular material in base/subbase layers (Carpenter et al., 2013); (Farina et al., 2017); (Gschösser et al., 2012). Thus, there is a need to extend the LCA analyses to alternative granular materials including stabilized-CDW aggregates with traditional and alternative binders. The previous studies on the LCA analysis using CDW materials focused on the economics and/or environmental impacts during the material production process (Bassani et al., 2019); (Dahlbo et al., 2015); (Doshio, 2007); (Ding et al., 2016a); (Serres et al., 2016); (Gschösser et al., 2012); (Hoxha et al., 2020). Thus, there is a need to consider all stages in the roadway life-cycle performance phases (i.e., construction, maintenance, rehabilitation)

in order to address all potential impacts and benefits of using CDW aggregates in the LCA analysis of roadway projects. This study addresses this need with the proposed novel methodology that quantifies the LCA environmental benefits and economic savings throughout the entire performance period of alternative sustainable strategies considering both construction and rehabilitation stages. This study addresses this need through the analysis of a pavement project representative of typical construction practices for average traffic volumes in Northern Italy. The life cycle economic and environmental benefits of using both natural- and CDW-stabilized aggregates as road base layer material were assessed. CDW aggregates stabilized with different binder types (i.e., cement and CKD) and contents were considered in the comparative analysis of alternative sustainable strategies. For each strategy, the pavement structure was designed to meet the structural requirements in the function of the materials used. The LCA analysis quantitatively assessed the economic and environmental impacts during the materials production, transportation, construction and rehabilitation phases.

4.3.2 Alternative Pavement Design with CDW Aggregates

A pavement project representative of typical construction practices for average traffic volumes in Northern Italy was considered as a case study for the LCA. The project characteristics are presented in Table 4.1 and are for a two-lane pavement with a width of 7.32 meters and a length of 1.6 kilometers (equivalent to one mile). The analysis period considered was 40 years with minor rehabilitation (i.e., overlay) every 10 years as estimated from the deterioration rate of the pavement structure. RAP material is considered for an onsite process and reuse. HMA, NA, CDW aggregates and cementitious materials were

supposed to be delivered from a plant 40 km away from the construction site. The distance between the construction site and landfill was 32 km. These distances are representative of paving projects in the region of the construction project.

Table 4.1 Design features of the paving construction project

Design Considerations	Value
Width (two travel lanes)	7.32 m
Road length	1.6 km
Wearing course depth	100 mm
Asphalt content	4.5%
Performance period of analysis	40 years
Rehabilitation (50 mm mill and overlay)	every 10 years
Distance from plant to site	40 km
Distance from site to landfill	32 km

Figure 2.1 shows the different scenarios considered in this study. The reference strategy includes a conventional road pavement entirely made with virgin materials (design A), while the sustainable alternatives consider stabilized-CDW aggregates in lieu of NA for base layer formation (designs B, C, D, and E). The conventional design consisted of 100 mm HMA over a 200 mm NA base treated with ordinary Portland cement (3%). All the alternative design strategies maintained the 100 mm of HMA composed with new construction materials due to stringent requirements for the quality of the surface layer, with the exception of design C where a 20% RAP was permitted for comparative purposes with option B. The inclusion of such a low content of RAP did not produce changes in the HMA properties. For the base layer, alternative formulations of CDW aggregates were considered

stabilized with different cementitious binders (e.g., CEM-II and CKD). The properties of such stabilized CDW materials are reported in Table 4.2, together with the different structural layer coefficients determined in relation to the properties of the materials. The equivalent thickness for the base layer for each case was determined in order to provide the same structural capacity (i.e., structural number SN, Equation 4.2). The 7-day unconfined compressive strength was used for estimating the structural coefficient (i.e., a_1, a_2, a_3) of each material.

$$\log(W_{18}) = Z_R \cdot S_0 + 0.36 \cdot \log(SN + 1) - 0.20 + \frac{\log[(\Delta PSI) / (4.2 - 1.5)]}{0.4 + 1094 / (SN + 1)^{5.19}} + 2.32 \cdot \log(M_R) - 8.07 \quad (4.1)$$

where

W_{18} = accumulated 18-kip equivalent single axle load for the design period

Z_R = reliability factor

S_0 = standard deviation

ΔPSI = initial PSI–terminal PSI

M_R = subgrade resilient modulus

M_R = structural number:

$$SN = a_1 D_1 + a_2 D_2 m_2 + a_3 D_3 m_3 \quad (4.2)$$

where

a_1, a_2, a_3 = structural layer coefficients for surface, base and subgrade layers

D_1, D_2, D_3 = thicknesses for surface, base and subgrade layers

m_2, m_3 = drainage coefficients for base and subgrade layers.

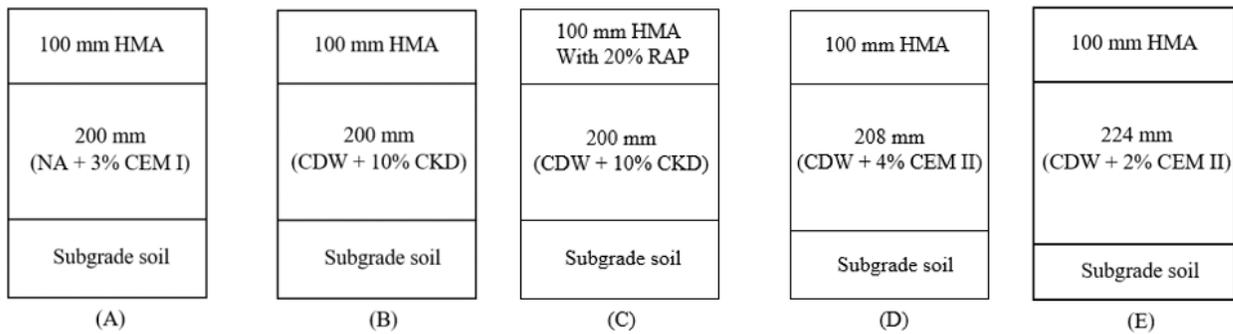


Figure 4.3 Schematic representation of alternative strategies and materials; (A) reference design, (B) (C), (D) and (E): alternative strategies

Table 4.2 Alternative Materials and Properties

Design	Aggregate	Binder	Binder Content	Water Content	Compacted Density	7-Day UCS*
			(%)	(%)	(kg/m ³)	(MPa)
A	NA	CEM-I	3	6.5	2363	2.33
B and C	CDW	CKD	10	12.3	2138	2.32
D	CDW	CEM-II	4	11.7	2141	2.00
E	CDW	CEM-II	2	11.2	2152	1.59

Note: CDW = construction and demolition waste; CKD = cement kiln dust; CEM-I = ordinary Portland cement; CEM-II = cement type II; NA = natural aggregate. Note: * represents for Unconfined compressive strength

The equivalent thicknesses for the base layer of each alternative design are presented in Figure 4.3. Comparatively, lower material strength corresponds to thicker base layer thickness, and vice versa.

4.3.3 Life Cycle Assessment (LCA)

The life cycle assessment for both environmental and economic impacts consider the entire supply chain (i.e., material production, construction, transportation, and maintenance activities) over the 40-year analysis period. Material resources, energy use, water consumption, emissions, costs, and other pertinent parameters were included in the analysis. Costs of material, transportation and construction operations, labor, overhead, and profit were included in the LCCA. Material costs were collected from local contractors (Table 4.3), while typical construction, maintenance (i.e., mill and overlay) cost were used. Similarly, labor costs and overhead rates were based on typical construction projects in the region and reported in Table 4.4. Consumption and emission generation in the production and transportation of materials during initial constructions and maintenance were considered to estimate the environmental effects. The environmental impact (energy consumption, water consumption, CO₂, CO, PM₁₀, NO_x, SO₂, and hazardous waste) due to CDW aggregate and CKD production were previously modeled using the software OpenLCA (Lee et al., 2013). The total environmental impacts were calculated as the sum of materials production, transportation and construction equipment. The LCA sustainability analysis was conducted over an analysis period of 40 years with scheduled 50 mm overlay every ten years for both conventional and alternative designs. The time intervals were determined based on the estimated traffic level and deterioration rates using the rehabilitation design principles of the

1993 AASHTO design guide (AASHTO, 1993). The CDW used in the alternative designs were tested for both short- and long-term performance assessment (Tefa et al., 2021). These materials show equal, or better, performance than conventional and alternative recycled materials. Thus, the long-term life of CDW materials is expected to match or exceed the performance of alternate materials, thus providing comparable or conservative values of performance life.

Table 4.3 Materials costs

Material	Cost (USD/ton)
HMA	80.0
RAP	15.0
CDW aggregate (0–40 mm)	2.4
CDW aggregate (0–8 mm)	3.6
NA (for base)	12.0
CEM-I	92.4
CEM-II B-P	96.0
CKD	1.2

Table 4.4 Labor, processing cost and overhead rates (PaLATE)

Process	Cost
Mill and overlay	USD 33/m ² /50 mm
Labor	USD 16,000/1.6 km
Equipment	USD 12,000/1.6 km
Overhead & profit	USD 11,000/1.6 km

4.3.4 Life Cycle Cost Analysis

The life cycle cost associated with each alternative is calculated and reported in terms of net present value (NPV) based on a discount rate of 4%. Figure 4.4 provides a comparison of the economic savings between the reference and the alternative strategies. Cost savings vary in relation to the type and percent of stabilizer used. Despite the relatively high percentage of CKD for stabilization of CDW aggregates in scenarios B and C, a higher level of cost saving is observed. Compared to the conventional case (scenario A), the use of CDW stabilized with 10% CKD (design B) in the base layer provides a cost reduction of up to 17%. This is related to the significantly lower price of CDW aggregates and CKD as compared to that of NA and ordinary cement respectively (Table 4.3). The use of RAP in the surface HMA (scenario C) led to additional savings with respect to references that are associated with the reduction in transportation and landfilling. Alternatives B and C have the same base layers (i.e., CDW aggregate with 10% CKD), however, alternative C presents a 5% lower cost, in relation to B, since 20% RAP is used in the surface HMA layer. The asphalt binder used in asphalt mixtures is the most expensive material in roadway projects. By using 20% RAP in HMA the new binder needed for HMA is reduced. For alternatives D and E, even though a higher amount of CDW aggregate and cement is needed to meet the structural requirements (i.e., thicker base layer according to the structural design) the associated costs were reduced by 11% and 13% respectively in relation to the reference strategy. Overall, the quantified cost savings for these strategies are attributed to the reduction of material costs. For alternative C additional cost savings are associated with the reduction in transportation and landfilling since 20% RAP was used in HMA.

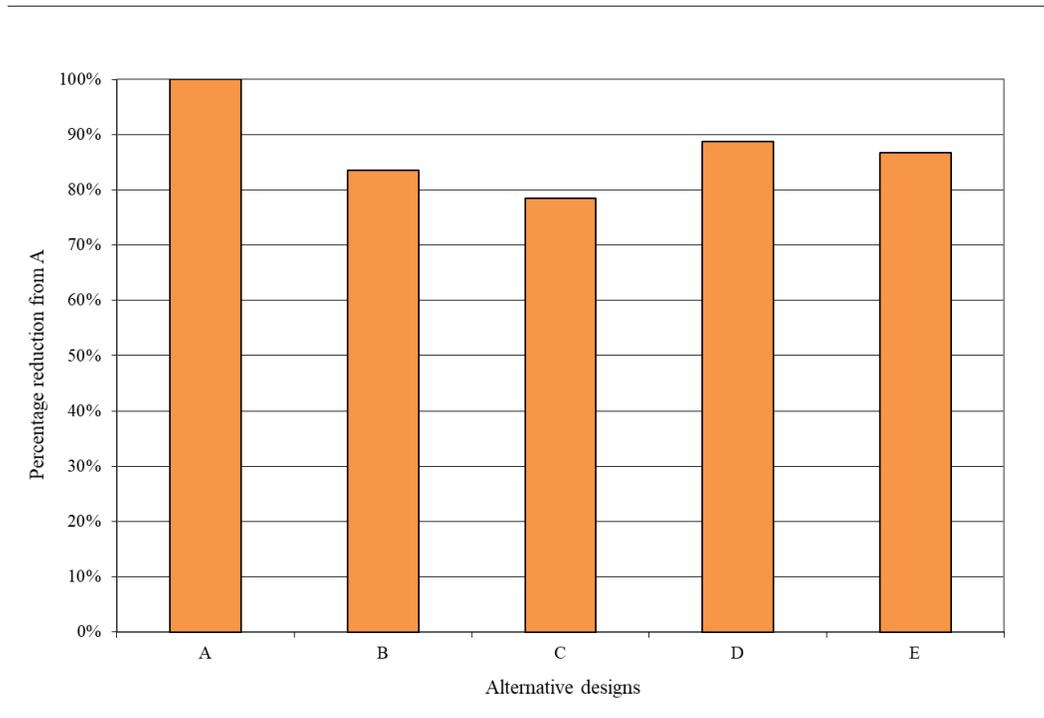


Figure 4.4 LCCA for alternative strategies. (A) reference design; (B) (C), (D) and (E): alternative strategies

4.3.4 Life Cycle Environmental Impacts

The environmental impacts were examined in relation to the resources and equipment used during all processing phases (i.e., production of materials, transportation and construction processes). Three major environmental impact components were reported including greenhouse gas emissions (CO₂), water consumption and energy consumption. Five pollutants that have a direct impact on human health, as identified by the Environmental Protection Agency, EPA, were also considered and include (i) hazardous waste generation, (ii) SO₂, (iii) CO, (iv) PM₁₀, and (v) NO_x (Lee et al., 2013). As shown in Figure 4, the life cycle CO₂ emissions for both conventional and alternative designs are dominated by materials production. The emission factors related to each material production are shown in

Table 4.5. The processes (i.e., equipment for construction and maintenance) and transportation generated a similar amount of greenhouse gas emission for all strategies. This is because a similar level of activities and equipment are used during these construction operations. In terms of materials production, overall, the greenhouse gas emissions are reduced significantly by substituting NA with CDW aggregate in base layers formation. By comparing strategy D to the conventional option, the replacement of virgin aggregates with CDW aggregate decreases approximately 20% the CO₂ emission despite the 1% increase in cement. The main sources of CO₂ emissions during material production include heavy equipment operations and transportation. However, substantial environmental benefits are achieved by using CDW aggregate due to the reduction of virgin materials needed and landfill disposal. In the case of option E, an additional 15% reduction in CO₂ was observed by limiting the amount of cement from 4% (option D) to 2% (option E) despite the higher amount of CDW aggregates needed to address the increased base layer thickness. This reflects the high amount of CO₂ associated with cement production as compared to CDW aggregate production. In case of design B and C, which employ 10% CKD to replace Portland cement, CO₂ emissions were reduced by 56% and 63%, respectively. Since CKD is a by-product of the cement manufacturing process, a significant reduction in CO₂ emissions is observed. In strategy C, CO₂ emissions from material production and transportation were further reduced due to the use of RAP (20%) in HMA.

Table 4.5 Environmental factors related to materials production (PaLATE and OpenLCA)

Materials	Energy MJ/ton	Water g/ton	Hazardous Waste g/ton	CO ₂ g/ton	CO g/ton	PM ₁₀ g/ton	SO ₂ g/ton
-----------	------------------	----------------	--------------------------	--------------------------	-------------	---------------------------	--------------------------

HMA	1968	96	3560	183,016	42.0	48.0	27.0
GAB	49	34,117	179	2718	6.6	2.0	9.2
Cement	4342	2,725,606	1636	879,729	661.9	189.4	783.9
CDW	-123	31,677	0	-5864	-25.2	-7.6	-13.8
CKD	19	12,632	0	4631	3.1	0.8	3.7

Note: The negative environmental effects representing the avoidance of landfill

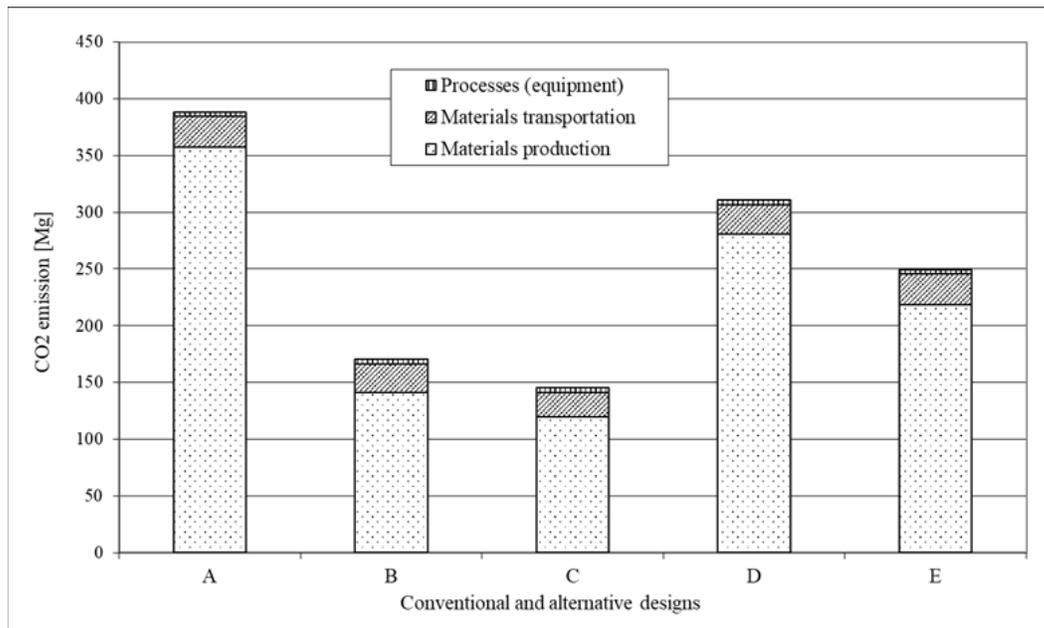


Figure 4.5 Life cycle greenhouse gas emissions (CO₂); (A) reference design, (B) (C), (D) and (E): alternative strategies

Figure 4.6 presents the energy consumption results. The energy savings are analogous to the reductions in CO₂ emissions associated with material production. It can be observed that construction processes consume the least amount of energy compared to material production and transportation. A maximum energy saving (equivalent to 44%) was achieved by using 20% RAP in HMA and considering a base layer with CDW with 10% CKD (scenario C). The substantial energy savings from options B and C reflect the fact that

cement production is an extremely high energy and emission intensive process. By comparing alternative D to the reference design, the energy consumption was reduced by 21% by substituting virgin aggregates with CDW ones. This indicates that the manufacturing of CDW aggregates is more energy efficient than that of NA.

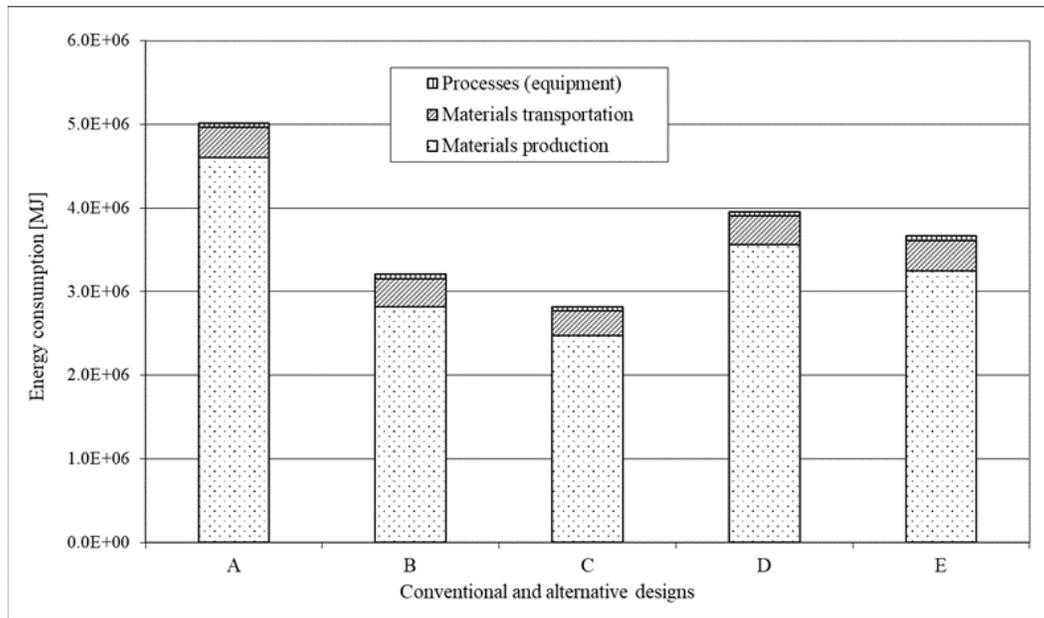


Figure 4.6 Life cycle energy consumption; (A) reference design, (B) (C), (D) and (E): alternative strategies

The life cycle water consumption is shown in Figure 4.7. Material production, especially cement, requires a large amount of water. Since cementitious materials were used in the mixture, water needs to be added to develop the hydration process. The optimum water content for each mixture is shown in Table 4.2. The water consumption is mainly from material production and processes. As higher water contents are needed for stabilized CDW, the water consumption increases for the alternative strategies. However, the total water consumption was reduced by about 15% for strategy C since a lower amount of water is

needed as compared to the reference case for material production of virgin aggregate and cement. In the case of option E, the total water consumption increased dramatically since a high water content is needed combined with the increased amount of material needed for the thicker base layer.

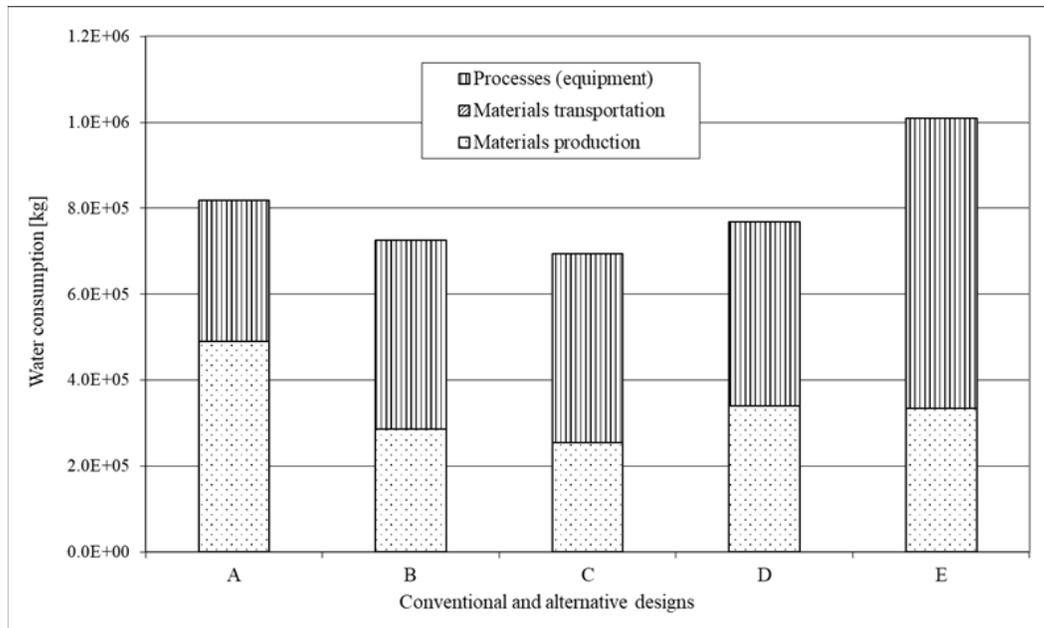


Figure 4.7 Life cycle water consumption; (A) reference design, (B) (C), (D) and (E): alternative strategies

Table 4.6 summarizes the quantities for each environmental parameter considered in this study, while Figure 4.8 presents the comparison between the reference and the alternative sustainable strategies (the latter expressed as relative results with respect to the reference). Hazardous waste generation primarily comes from producing materials such as asphalt emulsion, bitumen and concrete additives, and disposal of these materials to landfill. Aggregate and cement production generates very little hazardous waste compared to these materials. This reflects that only around 6% hazardous waste reduction was observed in

options B, D, and E. On the contrary, hazardous waste was further reduced by 17% when 20% RAP was used in HMA (strategy C). SO₂ emissions are analogous to hazardous waste generation associated with materials production. Additional pollutants (i.e., CO, PM₁₀, and NO_x) were also quantified (Table 4.6). According to the results of Table 4.6, considerable environmental savings for all alternative strategies can be deduced. Design C outperformed other alternatives in terms of all environmental impacts, particularly in energy and water consumption, and CO₂ emissions.

Table 4.6 Environmental impacts for alternative strategies

Designs		Energy Consumption (MJ)	Water Consumption (kg)	Hazardous Waste (kg)	CO ₂ (Mg)	CO (kg)	PM ₁₀ (kg)	NO _x (kg)	SO ₂ (kg)
Reference	A	5,018,259	818,330	35,514	388	804	1954	3040	52,288
Alternatives	B	3,207,405	725,435	33,287	170	466	1605	2483	49,012
	C	2,816,530	694,111	26,885	145	379	1416	2238	48,701
	D	3,956,628	767,996	33,664	311	575	1707	2761	49,109
	E	3,663,825	1,009,706	33,710	250	525	1773	2711	50,017

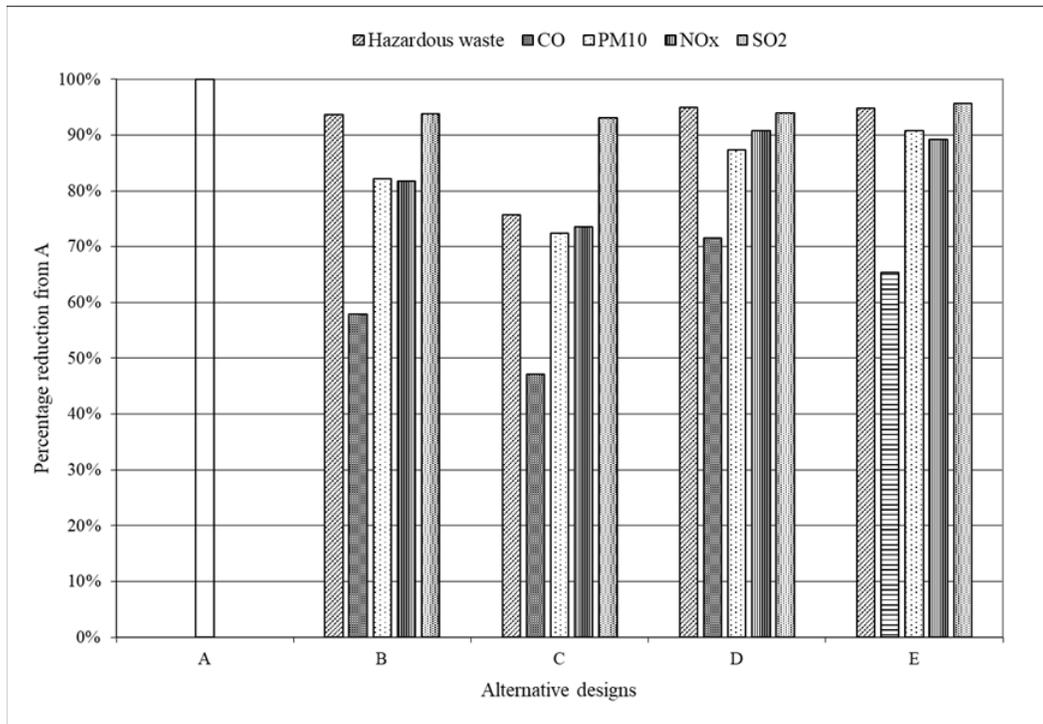


Figure 4.8 Environmental impacts of alternative strategies; (A) reference design, (B) (C), (D) and (E): alternative strategies

4.3.5 Sustainability Criteria and Rating

An overall assessment of each alternative strategy in regard to sustainability was conducted using BE²ST in-Highways (Lee et al., 2013). This sustainability metrics tool evaluates each alternative strategy using a comparative assessments method and rating based on the LCA results. Eight criteria are used in this assessment and include (i) energy use, (ii) global warming potential (GWP), (iii) recycling content, (iv) water consumption, (v) life cycle carbon costs, (vi) social carbon costs (SCC), (vii) traffic noise, and (viii) hazardous waste. Each alternative strategy is compared in relation to the reference one (strategy A). The SCC represents the cost needed to eliminate or address issues caused by carbon

emissions (i.e., USD/Mg of CO₂ emissions) and is associated with the cost of reducing global warming issues (e.g., GWP). Highway agencies often incorporate SCC for evaluating sustainable pavement construction and rehabilitation. As mentioned, in this study the alternative strategies were compared with the reference (i.e., conventional) option where new virgin materials were used for all processes and pavement construction stages. As mentioned earlier, weighting factors are assigned for each criterion to reflect their relative importance based on local conditions and policies for the construction projects. For instance, in some regions greenhouse emission or energy reduction may be more critical than cost savings, and so on. Therefore, higher weights are assigned to such critical parameters. The sum of weights should be equal to 100 (Figure 4.9). For this study, the sustainability criteria and targets, and the relative weights assigned to these parameters are as follows: 15% for energy consumption, global warming potential (i.e., CO₂ emission), recycling content and water consumption, 10% for hazardous waste and social carbon cost, and 5% for traffic noise. These parameters were selected to reflect current construction practice and policies with recycled materials for the specific region of the construction project. These parameters can be modified to reflect construction practices and policies elsewhere. Table 4.7 shows the sustainability target for each criterion. For instance, two points are rewarded if the energy consumption is reduced by more than 20%. While both targets and relative weights were selected for this region, such factors can be modified for roadway projects elsewhere.

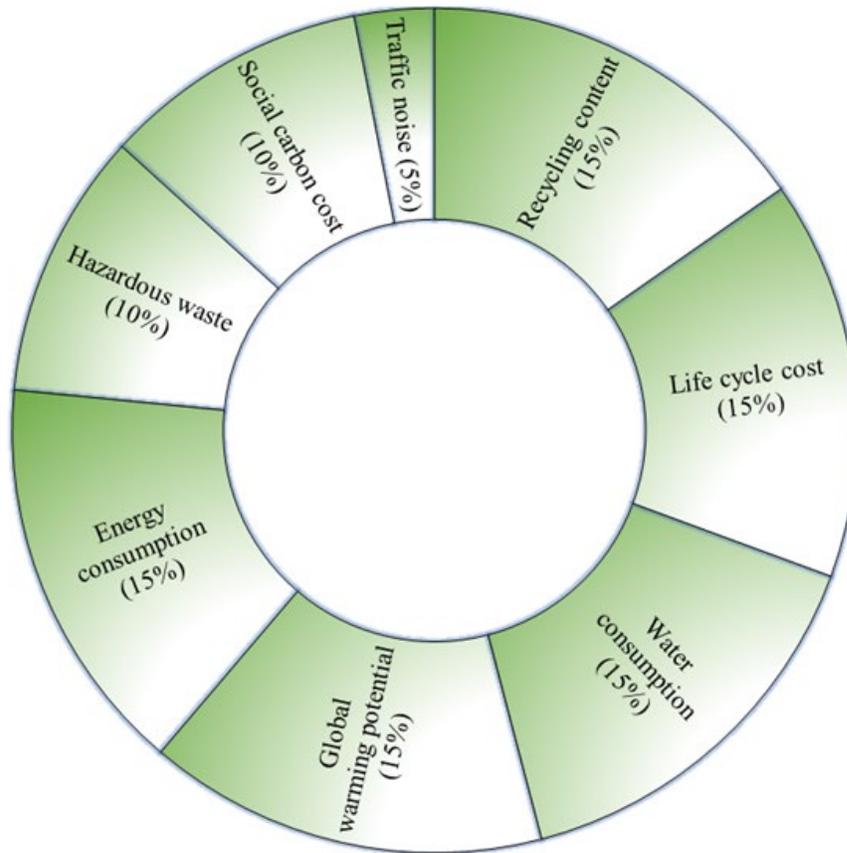


Figure 4.9 Relative weights for sustainability criteria

Table 4.7 Criteria and sustainability targets

Criteria	Unit	Target	
Energy consumption	MJ	≥10% reduction ≥20% reduction	(1 pt)
GWP	Mg		(2 pt)
Life cycle cost	USD		
Recycled content	%	≥10% recycling rate	(1 pt)
		≥20% recycling rate	(2 pt)
Water consumption	kg	≥5% reduction	(1 pt)
Hazardous waste		≥10% reduction	(2 pt)
Social carbon cost	USD	≥USD 12,344/km saving	(1 pt)
		≥USD 24,688/km saving	(2 pt)

Traffic noise	no unit	HMA	(1 pt)
		SMA or OGFC	(2 pt)

Note: SMA = stone mastic asphalt, OGFC = open graded friction course

The sustainability assessments for each strategy, both in terms of reward points pertinent to each criterion and total rating score, are summarized in Table 4.8. A weighted point (i.e., production of obtained point and weighting factor) was computed for each criterion. The total score was then calculated by dividing the total weighted point (i.e., sum of the weighted points for each criterion) into the target (i.e., 2). Strategy C represents the most sustainable option among the four proposed alternatives, with a total score of 92%. The total score is achieved by a 21% reduction in life cycle cost, a 15% reduction in water consumption, a 44% reduction in energy, and 63% reduction in CO₂ emissions. The Amoeba graphs for strategies C (best) and E (worst) are shown in Figure 4.10 as an example. Alternative D achieved a total score of 67% which outperformed alternative E (i.e., total score of 47%) in terms of sustainability even though E used a higher amount of cement (i.e., 4%). This is because strategy E requires 20 mm more layer thickness than D due to the low material strength, and thus more CDW aggregates, and water are needed. The impact of each strategy on such criteria is evident and could be used in further improving each strategy. Significant differences are observed between the two strategies in terms of water consumption, life cycle cost, social carbon cost and hazardous waste. The use of cement stabilization for CDW aggregate is attributed to good part to such effects. Thus, the results could eventually be used to further modify such alternatives for better sustainability scores.

Table 4.8 Points obtained for each parameter and total rating score

Strategy	Energy consumption	GWP	Recycled content	Water consumption	Life cycle cost	Social carbon cost	Traffic noise	Hazardous waste	Total weighted points	Total score
B	2.00	2.00	2.00	2.00	1.79	1.00	1.00	0.63	1.66	83%
C	2.00	2.00	2.00	2.00	2.00	0.85	1.00	2.00	1.83	92%
D	2.00	1.98	2.00	0.62	1.53	0.27	1.00	0.51	1.34	67%
E	2.00	2.00	2.00	0.00	1.63	0.48	1.00	0.51	0.94	47%

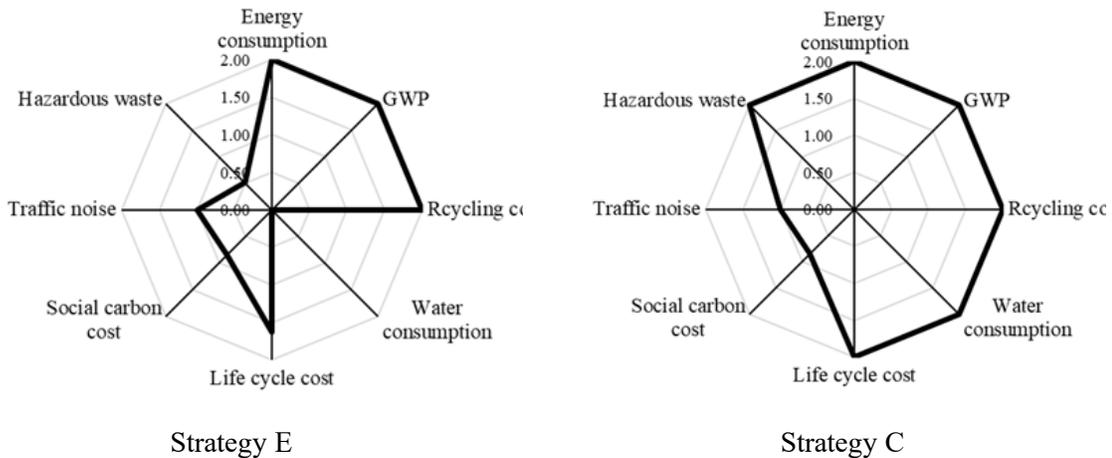


Figure 4.10 Amoeba graphs for strategies C and E

4.3.6 Conclusions

This study examined the life cycle cost savings and environmental benefits of using stabilized CDW recycled aggregates for base layers of roadway pavements. The proposed analysis approach for developing and accessing alternative sustainable strategies was presented in this process. The economic and environmental implications were quantified by

comparing the results of the alternative recycled materials (i.e., CDW, CKD, and RAP) strategies with those of the reference case where new construction materials are used. In the analysis resources and equipment used during the production of materials, construction processes (i.e., equipment used for construction and rehabilitation), and transportation were considered. The alternative strategies were developed based on the laboratory-obtained strength parameters of different stabilized CDW recycled aggregates. The analysis indicated that the alternative strategy employing CDW aggregates stabilized with 10% CKD in the base layer combined with a 20% RAP in the HMA surface layer provided the best sustainable option. This resulted in significant reductions in life cycle cost, energy consumption, water consumption and greenhouse gas emissions. The results also showed that the replacement of Port-land cement with CKD (i.e., alternatives B and D) stabilization of CDW aggregates further enhanced the environmental benefits. The LCCA indicated that cost savings were primarily attributed to the lower costs for CDW and CKD compared to conventional materials, while the LCA results indicated that the production of CDW and CKD requires less energy and generates lower emissions. The economic and environmental benefits quantified in this study could encourage the wider adoption of stabilized CDW aggregates in sustainable roadway construction. While the absolute values of the economic and environmental LCA are related to the inputs considered for this project, the relative benefits of using CDW in base and subbase layers are transferable, in scale, to any other projects where similar uses of these recycled materials are intended. Thus, the suggested approach for LCCA and LCA can be adopted elsewhere for quantifying the sustainability benefits CDW and other alternative recycled materials on roadways.

In conclusion, this study provided a tangible method for assessing the sustainability and contribution of CDW materials on roadways that can be expanded to other recycled materials. While the specific values of the economic and environmental LCA are related to the inputs considered in this project, the relative benefits of using CDW are transferable to other construction projects where similar uses and materials are used. Thus, the suggested approach for LCCA and LCA can be adopted elsewhere. Further research in this area should consider the potential adoption and implementation of sustainability criteria, and the proposed analysis in PMS. This will permit the generation of optimal sustainable alternative construction and rehabilitation strategies at the project and network level.

4.4 Life Cycle Sustainability Assessment of Using Rock Dust as a Partial Replacement of Fine Aggregate and Cement in Concrete Pavements

4.4.1 Background

Over the last decade, there has been an increase in the use of recycled materials/by-products as alternative materials in pavement construction such as RAP, recycled concrete materials (RCM), construction and demolition waste, fly ash, rock dust, glass and crumb rubber and others (Harvey et al., 2016); (Tam et al., 2018); (Williams et al., 2020); (Huang et al., n.d.). In recent years, the technical feasibility of using rock dust as a partial replacement for fine aggregate and/or cement in concrete has been extensively investigated. This stems from the fact that there is growing global concern about the depletion of sand deposits and the highly intensive energy consumption and CO₂ emissions associated with

cement production (Meisuh et al., 2018b). Rock dust is a byproduct obtained from the production process of crushed stone aggregates. A large amount of waste material is produced in the form of rock dust during the quarrying and aggregate processing (Dobiszewska et al., 2022). Thus, the use of rock dust as a partial replacement for sand and/or cement in concrete may have promising potential environmental and economic benefits in terms of reducing construction costs, energy consumption, GHG emissions and saving natural resources.

Rock dust has been found to produce concrete with equivalent or improved mechanical and durability properties when used as fine aggregate and/or cement (Meisuh et al., 2018a); (Kankam et al., 2017a); (Ilangovana et al., 2008); (Kankam et al., 2017b); (Zhao et al., 2022). The effect of rock dust on the mechanical properties of concrete significantly depends on the stone dust specific surface area and the percentage of replacement (Dobiszewska & Barnes, 2020). The use of rock powder as a partial replacement for fine aggregate generally leads to the improvement of mechanical properties and durability of cement composites (Dobiszewska et al., 2022; Dobiszewska & Barnes, 2020). Due to the fact that rock dust is an inert material, replacing cement with dust usually leads to the deterioration of the properties of cement composites. However, with a slight substitution of cement (i.e., in an amount of about 10–15%), a slight increase in strength and improvement in the durability of concretes and mortars is observed (Dobiszewska et al., 2022; Dobiszewska & Barnes, 2020). Beneficial rock dust interaction is attributed to the filler effect, which is the most important and dominant mechanism (Celik & Marar, 1996); (Kankam et al., 2017b); (Soroka & Setter, 1977). Very fine particles of rock powder fill the

space between the cement and the aggregate particles which leads to the reduction in cement matrix porosity. A decrease in the large capillary pores and an increase in small pores content is observed at the same time. This results in the densification of the hardened cement paste microstructure and less permeable structure. As a result, cement composites with a rock dust additive feature higher strength and durability (Dobiszewska & Barnes, 2020). Thus, there is a potential to use rock dust as a partial replacement of fine aggregate or/and cement in Portland concrete cement (PCC) pavement construction.

While the mechanical performance of a concrete mix with rock dust can be met, the potential economic and environmental benefits associated with its implementation need to be assessed, especially in the context of PCC pavement construction which requires large quantities of concrete. Therefore, the objective of this study was to estimate the potential economic and environmental impacts related to the use of rock dust in rigid concrete pavements through life-cycle analysis (LCA). The analyses considered conducting a quantitative assessment of different sustainable PCC pavement designs with rock dust and identifying the best sustainable alternative(s). The existing studies on the LCA of recycled materials and/or industrial by-products mainly focused on the economics and/or environmental impacts during the material production process (Del Ponte et al., 2017); (Ding et al., 2016). Thus, there is a need to consider all stages in pavement life-cycle performance (i.e., roadway design, construction, maintenance, rehabilitation and end of life) in order to address all the potential impacts and benefits of using recycled materials and industrial by-products in the LCA of roadway projects. This study addresses this need by proposing a holistic methodology that quantifies the life-cycle environmental benefits and economic

savings of using recycled materials/by-products in pavement construction and rehabilitation throughout the entire performance period of alternative sustainable strategies. To demonstrate the suggested approach and quantify the potential benefits, this study analyzed a roadway project representative of typical construction practices for average traffic volumes in Poland. The life-cycle economic assessment (LCCA) and life-cycle environmental analysis of using both conventional concrete and concrete with rock dust as a partial replacement of sand and/or cement in PCC pavements were conducted. The LCA considers all stages in the life cycle of pavements, including material production, construction, maintenance and rehabilitation, as well as the end-of-life phase (i.e., landfill or recycling). It is worth mentioning that while there are studies that have looked at LCA of concrete with fly ash (FA) and granulated blast furnace slag (GBFS), there are no studies that have examined the life-cycle economic and environmental impacts of concrete with the addition of rock dust.

4.4.2 Materials and Methods

4.4.2.1 Characteristics of Rock Dust and Cement

Ordinary Portland cement, OPC, CEM I 42.5R was used in the concrete mixtures. The cement specific gravity is 3.13 while the specific surface determined by the Blaine method was 3500 cm²/g. The chemical and mineral composition of the OPC is presented in Table 4.9 The gravel of the group of fractions 2/16, and river sand of the group of fractions 0/2 were used as a coarse aggregate and a fine aggregate, respectively.

Table 4.9 Chemical composition of cement

Chemical Composition [%]	
SiO ₂	19.33
Al ₂ O ₃	5.15
Fe ₂ O ₃	2.90
CaO	64.59
MgO	1.25
SO ₃	3.23
K ₂ O	0.47
Na ₂ O	0.21
Cl ⁻	0.05

The chemical composition of the rock dust (basalt origin) used in this study is presented in Table 4.10. The rock dust particle diameters are in the range of 0.5–200 μm . The average particle size of rock dust is 20 μm in diameter. The specific surface area of rock dust determined by the Blaine method was 3500 cm^2/g with a specific gravity of 2.99.

Table 4.10 Chemical composition of rock dust (basalt origin)

Chemical Composition [%]	
SiO ₂	42.61
Al ₂ O ₃	12.90
Fe ₂ O ₃	14.05
CaO	13.00
MgO	7.82
Na ₂ O	1.76
K ₂ O	1.15
P ₂ O ₅	1.80
SO ₃	0.07

MnO	0.25
Cl ⁻	0.10

4.4.2.2 Concrete Mix Design

The mix proportioning for the concrete to be used when generating the feasible alternative strategies is presented in Table 3. These mixtures were developed during the experimental study for meeting the C30/37 class compressive strength value according to the European Standards (The Polish Committee for The Polish Committee for Standardization: PN-EN 206-1+A1:2016-12: Concrete—Requirements, Properties, Production and Compliance; The Polish Committee for Standardization, 2016).

Table 4.11 Concrete mix proportioning

Concrete	Cement (kg/m ³)	Water (kg/m ³)	Rock dust (kg/m ³)	Fine Aggregate (kg/m ³)	Coarse Aggregate (kg/m ³)
A	350	155	0	533	1400
B	350		53.3	479.7	
C	350		106.6	426.4	
D	332.5		70.8	479.7	
E	315		88.3	479.7	
F	332.5		124.1	426.4	
G	315		141.6	426.4	

Six cube specimens of 150 mm × 150 mm × 150 mm were prepared for each concrete mix (i.e., the reference concrete and concretes with rock dust). All the specimens were cured

in water at a temperature of 20 ± 2 °C. Compressive strength test was conducted at 28 days according to European Standards EN 12390-3:2019-07 (EN 12390-3:2019-07; Testing Hardened Concrete -Part 3: Compressive Strength of Test Specimens. European Committee for Standardization, 2019).

4.4.3 Feasible Sustainable Strategies with Rock Dust Addition in Concrete

This study quantified the potential economic and environmental impacts related to the use of rock dust in rigid concrete pavements through the proposed methodology. For this purpose, a typical rural pavement section in Poland consisting of a 1.6 km (1 mi) length with two lanes, each lane 3.65 m (12 ft) wide, 40 km for the transport of materials to and from the plant to the project site and 32 km for the transport of waste materials to landfill/recycling plants. Embankment and shoulders were not considered in this case.

The alternative sustainable strategies included different alternative PCC pavement designs with rock dust in concrete. Table 4.12 shows the various sustainable strategies considered in this study in relation to the concrete strength and stiffness properties when rock dust is added as a fine aggregate and/or cement replacement. As mentioned earlier, concrete strength values for each mixture were obtained from the concrete properties' experimental study, while the corresponding modulus was estimated using Equation 4.3. The concrete slab and base layer thicknesses were obtained from the pavement design structural analysis. The reference strategy is a conventional PCC pavement entirely made with new materials (design A), while the sustainable alternatives consider the partial replacement of sand and/or cement by rock dust in the concrete layer (designs B, C, D, E, F and G). The

conventional design with new raw materials consisted of 203 mm (8 ft) PCC slab over a 150 mm (6 ft) granular base to meet the traffic and climatic conditions of the project. The feasible sustainable alternatives considered a maximum of 20% sand and 10% cement replacement with rock dust since larger amounts result in significant concrete strength reduction (Dobiszewska et al., 2022). Since the impact of various contents of rock dust in concrete strength was kept to comparable levels, the concrete slab thickness did not change significantly according to the structural analysis. Thus, the 150 mm granular base layer was also used for the alternative design strategies, as seen in Table 4.12. The analysis period considered was 40 years with mi-nor rehabilitation (i.e., overlay of 75 mm) at the 20th year as estimated from the deterioration rate of the pavement structure and as identified by the AASHTO design equation (4.3).

Table 4.12 Conventional and alternative sustainable strategies

Strategies	Reference	Sustainable Alternatives					
	A	B	C	D	E	F	G
Concrete mixture	Conventional concrete	10% rock dust in FA	20% rock dust in FA	10% rock dust in FA + 5% in cement	10% rock dust in FA + 10% in cement	20% rock dust in FA + 5% in cement	20% rock dust in FA + 10% in cement
Strength [MPa]	43.5	46.5	48	45	43	46	44
E [GPa]	34	35	35	34.5	34	35	34
Concrete slab thickness (mm)	203	195	190	200	201	200	205
Granular base (mm)	150	150	150	150	150	150	150

The design parameters used for the conventional and alternative strategies are shown in Table 4.13. The equivalent single axle load ESAL was equal to 7000 kg (15,500 lb.) and an annual average daily traffic flow (AADT) of 5000 was considered with 4% trucks. The equivalent thickness for the PCC slab for each alternative was determined for the same performance period using the AASHTO 1993 rigid pavement design guide (AASHTO, 1993). The concrete modulus of rupture and elastic modulus were obtained from the laboratory experimentation, as seen from Table 4.12, and Equations 4.4 and 4.5.

$$\log_{10}(W_{18}) = Z_R S_o + 7.35 \log_{10}(D+1) - 0.06 + \frac{\log_{10} \left[\frac{\Delta PSI}{4.5-1.5} \right]}{1 + \frac{1.64 \times 10^7}{(D+1)^{8.46}}} + (4.22 - 0.32 p_t) \log_{10} \left[\frac{S'_c C_d (D^{0.75} - 1.132)}{215.63 J \left[D^{0.75} - \frac{18.42}{(E_c / k)^{0.25}} \right]} \right] \quad (4.3)$$

W_{18} = design traffic (18-kip ESALs)

Z_R = standard normal deviate

S_o = combined standard error for reliability

D = thickness of concrete pavement slab

ΔPSI = initial and difference between terminal serviceability indices

P_t = terminal serviceability value

S'_c = modulus of rupture for Portland cement concrete

J = load transfer coefficient

C_d = drainage coefficient

E_c = modulus of elasticity for Portland cement concrete

k = modulus of subgrade reaction

Concrete properties based on compressive strength

$$S'_c = 6.7\sqrt{f'_c} \quad (4.4)$$

$$E_c = 57,000\sqrt{f'_c} \quad (4.5)$$

Table 4.13 Pavement design input parameters for the roadway site

Rigid Pavement Parameters	Values
Standard normal deviate, Z_R	-1.645
Overall standard deviation, S_o	0.3
Modulus of rupture, S'_c	As per Equation (4.4)
Difference between initial and terminal serviceability indices, ΔPSI	2.0
Terminal serviceability value, P_t	2.5

Elastic modulus, E_c	As per Table 4.12
Modulus of subgrade reaction, k	5.5 kg/cm ³
Load transfer coefficient, J	2.8
Drainage coefficient, C_d	1.0

4.4.3 Results and Discussions

4.4.3.1 LCCA Results

The life-cycle assessment for both environmental and economic impacts considered all life-cycle stages (i.e., material production, construction, transportation and maintenance/rehabilitation activities) over the 40-year analysis period. The LCCA disaggregates calculations over materials, transportation, landfill tipping fee, labor, process and equipment (including PCC demolition and paving) and overhead rates. Since recycling concrete pavement has been a common practice in recent years, this study considered that the existing PCC pavement was demolished and transported to a recycling plant instead of the landfill. Thus, there was no landfill tipping fee. The material costs were collected from local contractors (Table 4.14), while transportation, labor and equipment costs were based on typical construction projects in the region. The overhead rate was equal to 7% of the total cost, representing construction practice in the region. Since rock dust is a waste material from aggregate production, there was no production cost for this material.

Table 4.14 Cost of materials in the region

Materials	Cost (USD/ton)
Sand	6

Aggregates	26
Cement	115
Concrete additives	1918

The life-cycle cost associated with each alternative was calculated and reported in terms of net present value (NPV, Equation 4.6) based on a discount rate of 4%, representative of the discount rate of transport infrastructure investment in the US and the EU (Lawrence et al., 2015); (Disa, 2018).

As identified in the first two rows of Figure 4.2, all pertinent costs accounted for in LCCA include expenditures related to material production (such as labor, equipment, overhead), cost pertinent to material transportation to site, cost associated with initial construction and future maintenance and rehabilitation activities, costs pertinent to material transport from and to the plant and/or to the landfill and end-of-life recycling or landfill disposal. Equipment performance, energy cost, labor and overhead are all accounted for in each of these construction stages.

Figure 4.11 provides a comparison of the economic savings between the reference and the alternative strategies. Compared to the conventional case (strategy A), 10% sand replacement by rock dust (alternative B) provides a life-cycle cost reduction of 1.6%. This is mainly contributed to the cost savings in materials (i.e., fine aggregate). Furthermore, the concrete mix with a lower rock dust content provides slightly higher compressive strength than the conventional strategy with new raw materials. Thus, the PCC thickness is reduced without compromising performance, which also leads to cost savings. Alternative C and D

have similar life-cycle cost reductions, equivalent to 3.3%. Despite the relatively small quantity of cement used in PCC in relation to sand, 5% cement replacement with rock dust produces the same economic benefits as the 20% sand replacement. This is related to the significantly higher prices of cement as compared to sand (Table 6). Alternative strategy E has a slightly higher cost reduction (i.e., 5.2%) than option F (i.e., 5.0%) also attributed to the higher cement replacement (i.e., 10%) with rock dust. As can be observed, sustainable strategy G provided the highest overall life-cycle cost reduction of 6.8%. It should be noted that the construction and rehabilitation of 1 km of roadway may cost millions of dollars, and thus, a 6.8% reduction in cost could contribute to significant economic benefits for the entire project.

$$NPV = \frac{R_t}{(1+i)^t} \quad (4.6)$$

where

NPV = net present value

R_t = net cash flow at time t

i = discount rate

t = time of cash flow

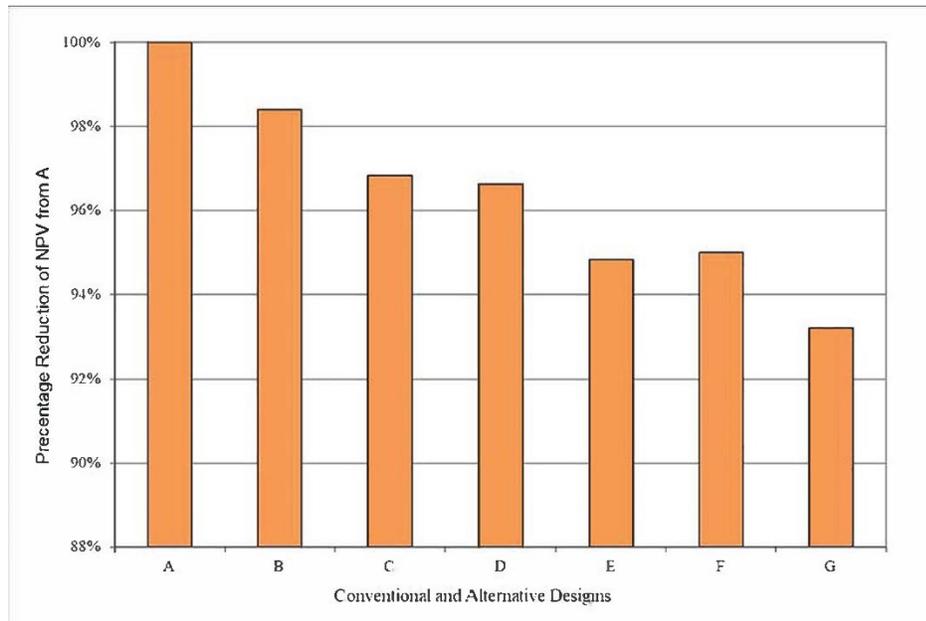


Figure 4.11 Comparison of life-cycle costs for alternative strategies

Figure 4.12 shows the NPV life-cycle cost broken down by materials and processes for alternative F, which is included here as an example in order to provide some insights on which components have the higher impact on the total LCCA. While the actual cost breakdown for each alternative sustainable strategy is different, the relative impact of these components on the total cost are comparable, yet not the same.

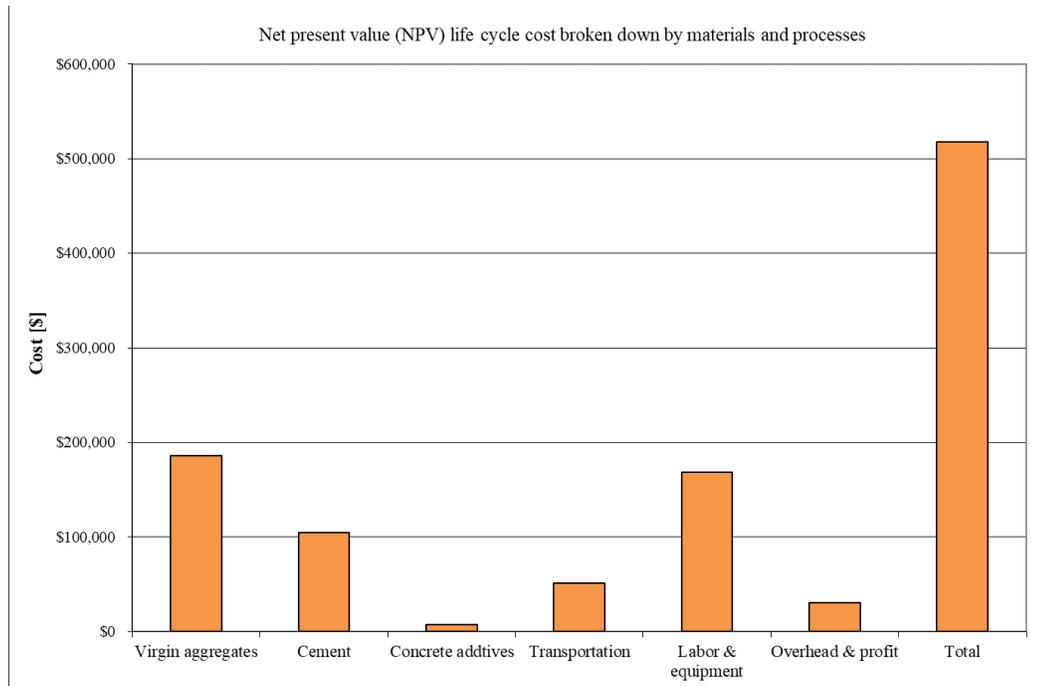


Figure 4.12 NPV life cycle cost broken down by materials and processes for sustainable strategy F

As can be seen from this Figure, the total project cost per two-lane km was calculated to be USD 518,533 in which 57% was associated to the cost of materials and 32% to labor, equipment and processes. Such cost is equivalent to USD 38.5 per square meter of installed PCC pavement (for a 200 mm thick slab), USD 14.5 per square meter of in-place granular base (150 mm thick layer) in regard to the initial instruction and USD 10.8 per square meter of PCC overlay (75 mm thickness) for maintenance. These costs reflect the typical construction projects in the region. While this reflects the construction and cost data in the region of the project site, as mentioned earlier the cost data in PaLATE can be easily revised to reflect the practices for a project at any specific region of interest. It is also expected that since changes in material and processes unit costs proportionally affect the alternative

strategies, the relative economic benefits from their comparison will still be valid in regard to the findings presented herein.

4.4.3.1 Life Cycle Environmental Impacts

The environmental impacts were examined in relation to the resources and equipment used during all processing phases (i.e., production of materials, transportation, construction and maintenance/rehabilitation). Three major environmental impact components include greenhouse gas emissions (CO₂), water consumption and energy consumption. Five pollutants that have a direct impact on human health, as identified by the Environmental Protection Agency (EPA), are also included: (i) hazardous waste generation, (ii) SO₂, (iii) CO, (iv) PM10 and (v) NO_x. The emission factors related to each material production are shown in Table 7. As mentioned earlier, these factors were obtained from the updated EPA emissions data (EPA 2022). Since rock dust is a byproduct/waste material that has already been processed, and since the emissions are accounted for during the aggregate production, no environmental loads were considered in this case so as to not double count such effects. The environmental impacts associated with transportation and processes (e.g., PCC paving, installing base and demolition of existing pavement) were obtained using the available equipment and data in PaLATE.

Table 4.15 Emission factors of materials production (after EPA 2022)

Materials and Processes	Energy [g/ton]	Water Consumption [g/ton]	CO ₂ [g/ton] GWP	NO _x [g/ton]	PM10 [g/ton]	SO ₂ [g/ton]	CO [g/ton]	Hg [g/ton]	Pb [g/ton]	RCRA Hazardous Waste
-------------------------	----------------	---------------------------	-----------------------------	-------------------------	--------------	-------------------------	------------	------------	------------	----------------------

										Generated [g/ton]
Aggregates	309	43	21,884	44	189	21	29	0	0	359
Cement	5168	2561	362,695	4362	817	4324	1550	0	0.4	2240
Concrete additives	23,784	22,190	1,423,609	5796	2084	4285	7299	0.1	3	354,745
Concrete mixing	536	932	37,099	551	172	484	337	0	0	169

Note: Resource Conservation and Recovery Act (RCRA)

As shown in Figure 4.13, the life-cycle CO₂ emissions for both conventional and alternative designs are dominated by materials production. The processes (i.e., equipment for construction and maintenance) and transportation generated a similar amount of greenhouse gas emission for all strategies. This is because a similar level of activities and equipment are used during these construction operations. Comparing strategy D to the conventional option, the replacement of sand with rock dust produces a CO₂ decrease of approximately 3%. The main sources of CO₂ emissions during material production include heavy equipment operations and transportation. In the case of strategy D, an additional 4% reduction in CO₂ was observed by replacing 5% of cement by rock dust. This reflects the high amount of CO₂ associated with cement production as compared to sand production. In the case of designs E and F, a similar reduction (i.e., 7.5%) in CO₂ emissions was observed. Alternative G produced approximately a 10% reduction (100 Mg) in CO₂ emission.

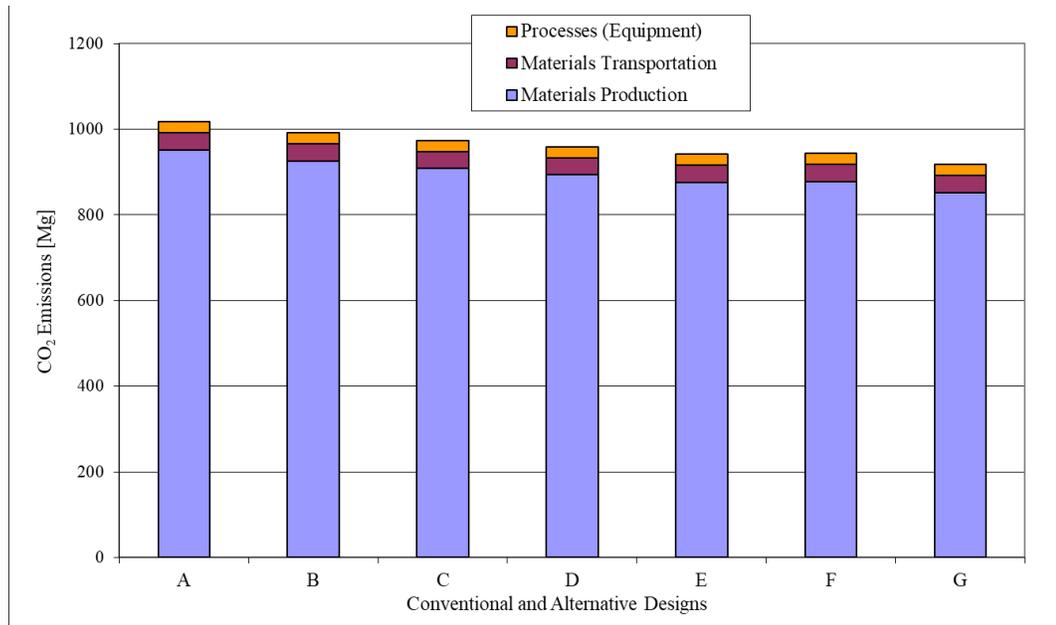


Figure 4.13 Life cycle greenhouse gas emissions (CO₂) for alternative strategies

Figure 4.14 presents the life-cycle energy consumption for each alternative. The energy consumptions are analogous to the reduction in CO₂ emissions associated with material production. It can be observed that the construction and maintenance processes consumed the least amount of energy compared to materials production and transportation. A maximum of 11% (1,488,033 MJ) reduction in energy consumption was achieved by replacing 20% of sand and 10% of cement with rock dust (alternative G). This reflects the fact that cement and aggregate productions are high-energy and emission-intensive processes. The life-cycle water consumption results are presented in Figure 8. Since water consumption is primarily affected by the production of the concrete mix, no significant changes were observed between the sustainable strategies.

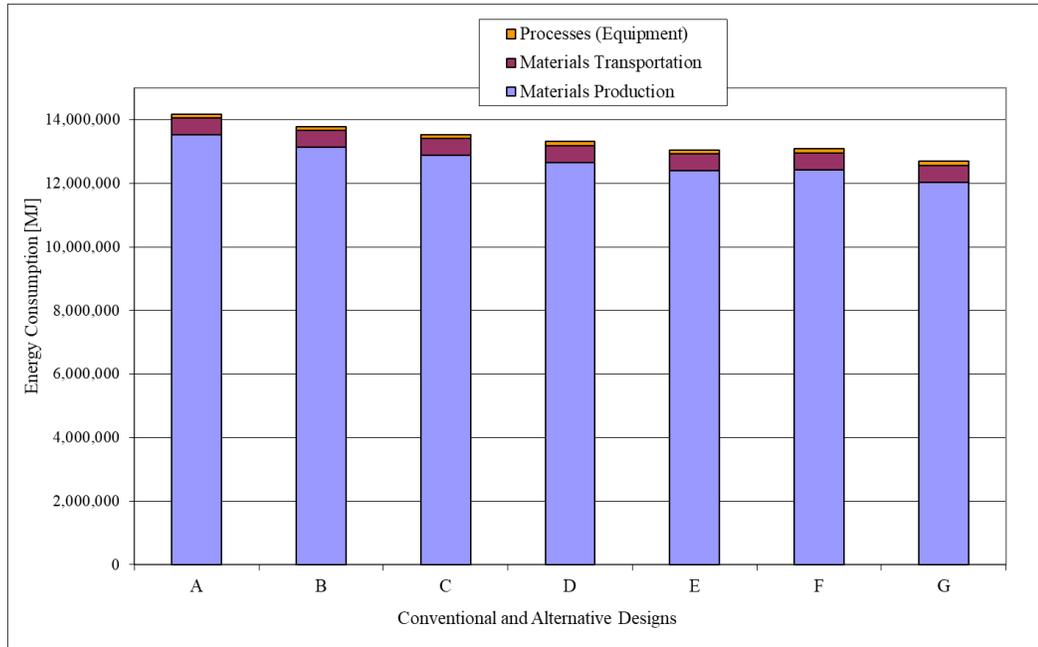


Figure 4.14 Life cycle energy consumption; reference design (A), alternative strategies (B–G)

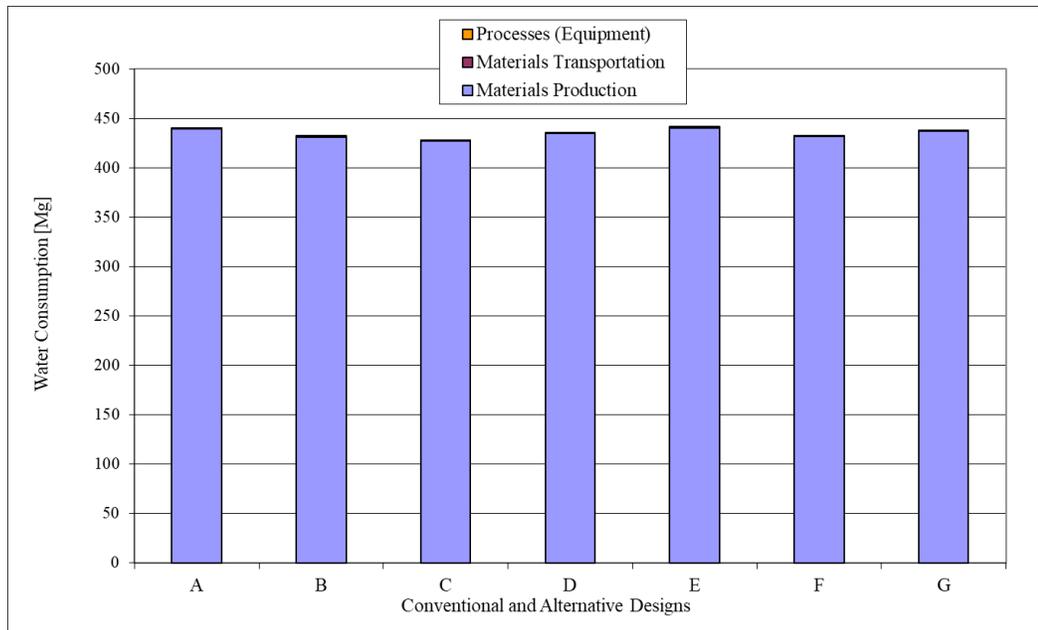


Figure 4.15 Life cycle water consumption; reference design (A), alternative strategies (B-G)

Table 4.16 summarizes the environmental impacts broken down by materials production, transportation and processes for strategy F as an example. It can be observed that these environmental parameters are dominated by materials production, especially by cement, aggregate and concrete mix production. This reveals where the environmental impacts are generated in the life-cycle analysis, as well as how and to what extent various sustainability strategies do in fact reduce them. Table 4.17 provides further details on the comparative assessment of the environmental impacts for each strategy. Overall, the environmental impacts decrease with an increase in the percentage of rock dust in concrete. The energy consumption and CO₂, SO₂, CO, Hg and Pb emissions are dominated by materials production, while transportation significantly contributes to NO_x (approximately 17.2% associated with transportation), PM₁₀ (about 9.2% from transportation) and hazardous waste generation (15.4% from transportation).

Table 4.16 Environmental impacts broken down by material production, materials and transportation for strategy F

Phases		Energy [MJ]	Water Consumption [kg]	CO ₂ GWP [Mg]	NO _x [kg]	PM10 [kg]	SO ₂ [kg]	CO [kg]	Hg [g]	Pb [g]	RCRA Hazardous Waste Generated [kg]
Material production	Aggregates	2,692,434	368,398	200,386	458,881	1,625,638	189,268	246,631	0	56	3416
	Cement	5,299,818	2,626,233	371,960	4,473,129	837,777	4,434,189	1,589,180	5	433	2298

	Additives	128,777	120,145	7708	31,383	11,286	23,201	39,522	0	18	1921
	Concrete mixing	4,310,516	1,362,680	298,275	4,431,384	1,383,974	3,887,396	2,706,768	9	471	7493
Transportation	Concrete from plant to site	246,521	41,972	18,430	981,871	191,393	58,912	81,823	0	8	1776
	Waste materials from site to landfill/recycling plant	280,395	47,739	20,962	1,116,787	217,795	67,007	93,066	0	9	2020
Processes	Base construction	16,796	1633	1261	27,248	4295	1802	5872	0	1	121
	PCC paving	24,865	2418	1866	38,473	13,790	2544	8290	0	1	179

Table 4.17 Environmental impacts for alternative sustainable strategies

Strategy	NOx [kg]	PM10 [kg]	SO ₂ [kg]	CO [kg]	Hg [g]	Pb [g]	RCRA Hazardous Waste Generated [kg]
A	12,368	4540	9271	5106	15	1096	24,879
B	12,124	4421	9079	4982	15	1063	23,150
C	11,996	4340	8980	4907	15	1041	21,665
D	11,801	4369	8748	4846	14	1017	21,197
E	11,671	4361	8598	4782	14	989	19,364
F	11,703	4298	8674	4785	14	998	19,761
G	11,462	4267	8417	4681	14	959	17,854

4.4.4 Sustainability Rating

A sustainability rating system, in this case BE²ST in-Highways, can be used to assess each alternative strategy (BE2ST-in-Highways | Recycled Materials Resource Center). The selected sustainability rating system evaluates each alternative strategy using a comparative assessment and rating method based on the LCA results. Seven criteria are used in this assessment and include: (i) energy consumption, (ii) global warming potential (GWP), (iii) recycling, (iv) water consumption, (v) life-cycle carbon costs, (vi) social carbon costs (SCCs) and (vii) hazardous waste. Each alternative strategy is compared in relation to the reference one (strategy A). SCCs represent the costs needed to eliminate or address issues caused by carbon emissions (i.e., USD/Mg of CO₂ emissions) and are associated with the cost of reducing global warming issues (e.g., GWP). Highway agencies often incorporate SCCs for evaluating sustainable pavement construction and rehabilitation. As mentioned earlier, in this study the alternative strategies were compared with the reference (i.e., conventional) option in which new raw materials are used for all processes and pavement construction stages.

For each alternative, a normalized score was calculated based on the percentage reduction in emissions, cost, consumption or percentage of recycled materials (rock dust) used. If the percentage reduction equals to or is larger than 10%, a score of 1 is assigned. Such method of calculating scores can be modified to encourage more sustainable solutions (e.g., 20% reduction corresponding to a score of 1). A comparison of alternatives is shown in Figure 4.16. The score (i.e., 0–1) for each alternative strategy was calculated based on the percentage reduction in each sustainability criterion of Figure 4.16. A higher score represents

a higher reduction for that economic environmental impact parameter. It can be observed that alternative G outperformed other strategies in all sustainability criteria, except for water consumption. The impact of each strategy on such criteria is evident and could be used in further improving each specific strategy. As it can be observed as more sand and/or cement is replaced by rock dust, a higher reduction in energy consumption, hazardous waste, GWP, life-cycle cost and social carbon cost can be achieved. However, the alternative strategies do not have a significant impact on water consumption. Thus, the results could be eventually used to further modify such alternatives for better sustainable scores.

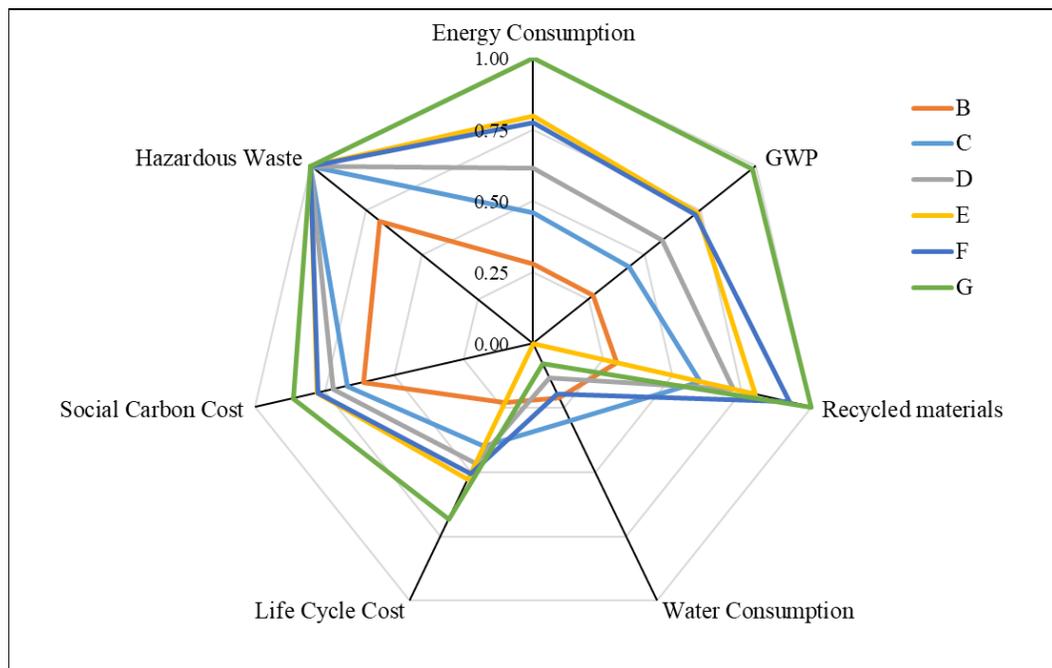


Figure 4.16 Sustainability rating for each alternative

4.4.5 Summary and Conclusions

This section examined the life-cycle economic savings and environmental benefits of using rock dust for the partial replacement of fine aggregate and/or cement in concrete for roadway pavement construction. The proposed methodological approach for developing and accessing alternative sustainable strategies was presented in this process. The life-cycle economic and environmental impacts were quantified by comparing the results of sustainable alternative strategies (i.e., with rock dust use in concrete) with the reference design where new raw construction materials are used. In the proposed holistic analysis approach, the LCCA and LCA environmental analysis considered all stages in the life cycle of pavements, rather than just the material production that past studies have focused on. Thus, the suggested methodology includes analyses and inputs pertinent to material production, construction, maintenance and rehabilitation and end of life (landfill or recycling). The feasible alternative strategies were developed based on the laboratory experimentation results on using rock dust in concrete and providing acceptable strength properties. The analysis indicated that the alternative strategy with 20% fine aggregate and 10% cement replacement with rock dust provided the best sustainable option. This sustainable alternative provides a reduction in life-cycle cost, energy consumption, greenhouse gas emissions, hazardous waste and other environmental parameters. The LCA analysis indicated that cost savings and environmental benefits were primarily attributed to materials production. The economic savings and environmental benefits quantified in this study may encourage the wider adoption of rock dust for sustainable PCC roadway construction. While the reported values of LCA analysis are related to the specific inputs considered for this project, the

relative comparison between such alternative strategies is expected to be maintained, since changes in unit costs and environmental parameters proportionally affect the various materials and construction phases in each option. The methodology and analysis presented in this study can be adopted elsewhere for quantifying the sustainability benefits of rock dust or other recycled materials in roadway construction. Furthermore, such analysis could be integrated into PMSs that agencies currently use for identifying optimal allocation of resources in maintaining their highway network.

CHAPTER 5: SUMMARY AND RECOMMENDATIONS

While the specific findings and research contributions of this dissertation have been included and summarized in each of the previous chapters, this section outlines how the current research effort could be useful and/or transferable elsewhere and/or further expanded.

Among the objectives of this study was to explore alternative ML models for enhancing the prediction of compressive strength as a function of mixture ingredients and proportions. Seven alternative ML models were considered (LR, DT, SVM, MLP, RF, Adaboost, and Xgboost) of increasing complexity for concrete compressive strength prediction. Three synthetic features were created using domain knowledge including water to cementitious materials ratio (W/C ratio), fresh density, and aggregate to cement ratio. The improved model performance demonstrated that these features better represent the underlying patterns, providing better interpretability of the data and the model. A two-layer stacked model was developed to further enhance the prediction accuracy. The proposed stacked model showed superior performance with an R^2 of 0.985 which outperforms the models from literature.

While the ML models proposed herein have shown to be very effective in predicting concrete strength, model generalization to a wider set of data and concrete mixtures will be beneficial for their widespread calibration and use. The superiority of the proposed methodology and stacking is expected to be further optimized with more complex concrete data. Thus, future model development and validation should consider datasets from field-produced concrete where higher noise and variance caused by uncontrolled or unreported

production processing and construction variables may exist. Such datasets may eventually include concrete mixtures such as self-consolidated concrete, fiber-reinforced concrete, and others. In a similar direction, the proposed ML models, and/or modeling techniques, could be calibrated to other regions by using concrete datasets reflecting concrete materials and mixtures elsewhere. In terms of transferability, this study promotes applying and validating the proposed methodology in predicting other mechanical properties of concrete, such as tensile strength and elastic modulus, as well as performance of transportation infrastructures including pavement performance and bridge conditions.

Another important contribution of this study was the development of a Monte Carlo simulation process to systematically quantify acceptance risks and assess the implications on pay factors (PF). The simulation was performed using typical AQCs, such as strength, thickness, and roughness for PCC pavements. The statistical power of the F-test and t-test was determined for various combinations of population characteristics and sample sizes of the contractor and agency data. The analysis indicated that specific combinations of contractor and agency sample sizes and population characteristics have a greater impact on acceptance risks and may provide inconsistent PF. Thus, the findings of this study and the proposed approach can assist both agencies and producers to better assess and understand the impact of sample sizes and population characteristics on the acceptance risks and rewards. Therefore, the proposed methodology can be adopted by highway agencies to develop statistically valid verification procedures and thus more rational and defensible quality assurance (QA) specifications. Producers may use such analysis to identify the level of risks and rewards associated with the current production and identify potential improvements in

quality. In future work, the proposed simulation model can be further applied to evaluate the acceptance risks of other highway materials and construction, such as HMA and bridges, and identify the appropriate sample sizes, PF equations, AQL, RQL and specification limits for QA based on desired risk levels for each SHA agency. Further research should also consider the feasibility of adopting the developed Monte Carlo simulation for reliability analysis in pavement design and performance prediction software.

Finally, this study presented an approach for quantifying the life cycle economic and environmental impacts of using recycled materials in pavement construction. The selected case studies included, (i) CDW aggregates in pavement base layers and (ii) the use of rock dust as a partial replacement of fine aggregate and cement in PCC pavements. A LCA framework, which considers all life cycle stages such as material production, transportation, construction, maintenance, rehabilitation and end of life, was proposed to assess such impacts. This study also provides significant insights on the specific contribution of material production, construction processes and the transportation of materials to the overall environmental benefits and cost savings. The suggested approach for pavement LCA can be adopted elsewhere for quantifying the sustainability benefits of using alternative recycled materials in roadways. The proposed methodology is easily custom tailored to consider off-the-shelf tools used in other regions, and flexible enough to accommodate economic and environmental impact analysis models of interest. Since LCA involves data, assumptions and predictive models, updated inputs should be used in such analysis. When the methodology is transferred to a different scenario, different sources of uncertainty may be introduced. Thus future research should integrate uncertainty assessment methods and

advance statistical considerations into LCA to enable perhaps a probabilistic approach to economic and environmental impacts. Eventually, the LCA should be integrated to PMS and multiple phases of decision-making in a manner that promotes economically efficient environmental impacts reduction.

APPENDIX

The codes for this study are available at <https://github.com/zyp1015/PhD-Dissertation> per request.

BIBLIOGRAPHY

1. AASHTO. AASHTO Guide for Design of Pavement Structures; AASHTO: Washington, DC, USA, 1993; Volume 1.
2. AASHTO. Mechanical-Empirical Pavement Design Guide, MEPDG, 3rd ed.; AASHTO: Washington, DC, USA, 2020.
3. Aboutalebi Esfahani, M. (2020). Evaluating the feasibility, usability, and strength of recycled construction and demolition waste in base and subbase courses. *Road Materials and Pavement Design*, 21(1), 156-178.
4. Abukhattala, M. (2016, May). Use of recycled materials in road construction. In *Proceedings of the 2nd international conference on civil, structural and transportation engineering*, Ottawa, Canada (pp. 138-1).
5. ACI Committee 318, (2019). Building code requirements for reinforced concrete, American Concrete Institute, ACI 318.
6. Agrela, F., Barbudo, A., Ramírez, A., Ayuso, J., Carvajal, M. D., & Jiménez, J. R. (2012). Construction of road sections using mixed recycled aggregates treated with cement in Malaga, Spain. *Resources, Conservation and Recycling*, 58, 98-106.
7. American Association of State Highway and Transportation Officials. AASHTO. Highway Subcommittee on Construction, 1996. Implementation manual for quality assurance.
8. American Association of State Highway and Transportation Officials. 2010a. Standard specifications for transportation materials and methods of sampling and testing, 30, Washington, DC.
9. American Association of State Highway and Transportation Officials. Guide for Design of Pavement Structures; AASHTO: Washington, DC, USA, 1993.

10. Amiril, A., Nawawi, A. H., Takim, R., & Latif, S. N. F. A. (2014). Transportation infrastructure project sustainability factors and performance. *Procedia-Social and Behavioral Sciences*, 153, 90-98.
11. Anastasiou, E. K., Liapis, A., & Papayianni, I. (2015). Comparative life cycle assessment of concrete road pavements using industrial by-products as alternative materials. *Resources, Conservation and Recycling*, 101, 1-8.
12. Arulrajah, A., Mohammadinia, A., D'Amico, A., & Horpibulsuk, S. (2017). Cement kiln dust and fly ash blends as an alternative binder for the stabilization of demolition aggregates. *Construction and Building Materials*, 145, 218-225.
13. Arulrajah, A., Mohammadinia, A., D'Amico, A., & Horpibulsuk, S. (2017). Effect of lime kiln dust as an alternative binder in the stabilization of construction and demolition materials. *Construction and Building Materials*, 152, 999-1007.
14. Aurangzeb, Q., Al-Qadi, I. L., Ozer, H., & Yang, R. (2014). Hybrid life cycle assessment for asphalt mixtures with high RAP content. *Resources, conservation and recycling*, 83, 77-86.
15. Bassani, M., Khosravifar, S., Goulias, D. G., & Schwartz, C. W. (2015). Long-term resilient and permanent deformation behaviour of Controlled Low-Strength Materials for pavement applications. *Transportation Geotechnics*, 2, 108-118.
16. Bassani, M., Riviera, P. P., & Tefa, L. (2017). Short-term and long-term effects of cement kiln dust stabilization of construction and demolition waste. *Journal of Materials in Civil Engineering*, 29(5), 04016286.
17. Bassani, M., Tefa, L., Coppola, B., & Palmero, P. (2019). Alkali-activation of aggregate fines from construction and demolition waste: Valorisation in view of road pavement subbase applications. *Journal of Cleaner Production*, 234, 71-84.

18. Bassani, M., Tefa, L., Russo, A., & Palmero, P. (2019). Alkali-activation of recycled construction and demolition waste aggregate with no added binder. *Construction and Building Materials*, 205, 398-413.
19. Blankendaal, T., Schuur, P., & Voordijk, H. (2014). Reducing the environmental impact of concrete and asphalt: a scenario approach. *Journal of cleaner production*, 66, 27-36.
20. Blengini, G. A., & Garbarino, E. (2010). Resources and waste management in Turin (Italy): the role of recycled aggregates in the sustainable supply mix. *Journal of Cleaner Production*, 18(10-11), 1021-1030.
21. Borghi, G., Pantini, S., & Rigamonti, L. (2018). Life cycle assessment of non-hazardous Construction and Demolition Waste (CDW) management in Lombardy Region (Italy). *Journal of cleaner production*, 184, 815-825.
22. Bovea, M. D., & Powell, J. C. (2016). Developments in life cycle assessment applied to evaluate the environmental performance of construction and demolition wastes. *Waste management*, 50, 151-172.
23. Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, B*, 26(211-234).
24. Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123-140.
25. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. (1984). *Classification and Regression Trees*. Taylor & Francis.
26. Bühlmann, P., & Yu, B. (2002). Analyzing bagging. *The Annals of Statistics*, 30(4), 927-961.
27. Burati, J.L., Straub, D., Delk, A. and Zhou, X., (2010). Validation of Contractor HMA Testing Data in the Materials Acceptance Process (No. FHWA-SC-10-02). South Carolina. Department of Transportation.

28. Burati, J.L., Weed, R.M., Hughes, C.S. and Hill, H.S., (2003). Optimal procedures for quality assurance specifications (No. FHWA-RD-02-095). Turner-Fairbank Highway Research Center.
29. Cabalar, A. F., Zardikawi, O. A. A., & Abdulnafaa, M. D. (2019). Utilisation of construction and demolition materials with clay for road pavement subgrade. *Road Materials and Pavement Design*, 20(3), 702-714.
30. Camargo, F. F., Edil, T. B., & Benson, C. H. (2013). Strength and stiffness of recycled materials stabilized with fly ash: A laboratory study. *Road Materials and Pavement Design*, 14(3), 504-517.
31. Carnegie Mellon University. Economic Input-Output Life Cycle Assessment. (2022). Available online: <http://www.eiolca.net/> (accessed on 26 July 2022).
32. Carpenter, A., Jambeck, J. R., Gardner, K., & Weitz, K. (2013). Life cycle assessment of end-of-life management options for construction and demolition debris. *Journal of Industrial Ecology*, 17(3), 396-406.
33. Carr, E.A., Shuler, M.G. and Burati Jr, J.L., (2016). Validation of contractor HMA testing data in the materials acceptance process-phase II (No. FHWA-SC-16-03). South Carolina. Department of Transportation.
34. Celauro, C., Corriere, F., Guerrieri, M., & Casto, B. L. (2015). Environmentally appraising different pavement and construction scenarios: A comparative analysis for a typical local road. *Transportation Research Part D: Transport and Environment*, 34, 41-51.
35. Celik, T., & Marar, K. (1996). Effects of crushed stone dust on some properties of concrete. *Cement and Concrete research*, 26(7), 1121-1130.
36. Chakraborty, D., Awolusi, I., & Gutierrez, L. (2021). An explainable machine learning model to predict and elucidate the compressive behavior of high-performance concrete. *Results in Engineering*. 11, 100246.

37. Chiu, C. T., Hsu, T. H., & Yang, W. F. (2008). Life cycle assessment on using recycled materials for rehabilitating asphalt pavements. *Resources, conservation and recycling*, 52(3), 545-556.
38. Chou, J.-S., Chiu, C.-K., Farfoura, F., & Al-Taharwa, I. (2011). Optimizing the Prediction Accuracy of Concrete Compressive Strength Based on a Comparison of Data-Mining Techniques. *Journal of Computing in Civil Engineering*, 25(3).
39. Coelho, A., & De Brito, J. (2013). Economic viability analysis of a construction and demolition waste recycling plant in Portugal—part I: location, materials, technology and economic analysis. *Journal of Cleaner Production*, 39, 338-352.
40. Coenen, A. R., Pforr, J. E., Hefel, S. A., & Paye, B. C. (2019). State DOT Implementation of Statistical Analysis and Percent Within Limits. *Transportation Research Record*, 2673(2), 583-592.
41. Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods*, 250, 126-136.
42. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297.
43. Cristelo, N., Vieira, C. S., & de Lurdes Lopes, M. (2016). Geotechnical and geoenvironmental assessment of recycled construction and demolition waste for road embankments. *Procedia Engineering*, 143, 51-58.
44. da Conceição Leite, F., dos Santos Motta, R., Vasconcelos, K. L., & Bernucci, L. (2011). Laboratory evaluation of recycled construction and demolition waste for pavements. *Construction and building materials*, 25(6), 2972-2979.

45. Dahlbo, H., Bachér, J., Lähtinen, K., Jouttijärvi, T., Suoheimo, P., Mattila, T., ... & Saramäki, K. (2015). Construction and demolition waste management—a holistic evaluation of environmental performance. *Journal of Cleaner Production*, 107, 333-341.
46. Del Ponte, K., Madras Natarajan, B., Pakes Ahlman, A., Baker, A., Elliott, E., & Edil, T. B. (2017). Life-cycle benefits of recycled material in highway construction. *Transportation Research Record*, 2628(1), 1-11.
47. del Río Merino, M., Izquierdo Gracia, P., & Weis Azevedo, I. S. (2010). Sustainable construction: construction and demolition waste reconsidered. *Waste management & research*, 28(2), 118-129.
48. Ding, T., Xiao, J., & Tam, V. W. (2016). A closed-loop life cycle assessment of recycled aggregate concrete utilization in China. *Waste management*, 56, 367-375.
49. Asplund, D. (2018). Discounting transport infrastructure investments. VTI.
50. Dobiszewska, M., Bagcal, O., Beycioğlu, A., Goulias, D., Köksal, F., Niedostatkiewicz, M., & Ürünveren, H. (2022). Influence of Rock Dust Additives as Fine Aggregate Replacement on Properties of Cement Composites—A Review. *Materials*, 15(8), 2947.
51. Dobiszewska, M., & Barnes, R. W. (2020). Properties of mortar made with basalt powder as sand replacement. *ACI Materials Journal*, 117(2), 3-9.
52. Reis, G. S. D., Quattrone, M., Ambrós, W. M., Grigore Cazacliu, B., & Hoffmann Sampaio, C. (2021). Current applications of recycled aggregates from construction and demolition: A review. *Materials*, 14(7), 1700.
53. Dosho, Y. (2007). Development of a sustainable concrete waste recycling system-Application of recycled aggregate concrete produced by aggregate replacing method. *Journal of Advanced Concrete Technology*, 5(1), 27-42.

54. Duan, Z.H., Kou, S.C., & Poon, C.S. (2013). Prediction of compressive strength of recycled aggregate concrete using artificial neural networks. *Construction and Building Materials*, 40, 1200-1206.
55. EN 12390-3:2019-07; Testing Hardened Concrete—Part 3: Compressive Strength of Test Specimens. European Committee for Standardization: Brussels, Belgium, 2019.
56. EN 206 + A1:2016-12; Concrete—Requirements, Properties, Production and Compliance. Polish Committee for Standardization: Warsaw, Poland, 2016.
57. Environmental Protection Agency. US Environmentally—Extended Input—Output (USEEIO) Models. 2022. Available online: <https://www.epa.gov/land-research/us-environmentally-extended-input-output-useeio-models> (accessed on 10 August 2022).
58. Erdal, H.I., Karakurt, O., & Namli, E. (2013). High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform. *Engineering Applications of Artificial Intelligence*. 26, 1246-1254.
59. Farina, A., Zanetti, M. C., Santagata, E., & Blengini, G. A. (2017). Life cycle assessment applied to bituminous mixtures containing recycled materials: Crumb rubber and reclaimed asphalt pavement. *Resources, Conservation and Recycling*, 117, 204-212.
60. Farooq, F., Ahmed, W., Akbar, A., Aslam, F., & Alyousef, R. (2021). Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *Journal of Cleaner Production*. 292, 126032
61. Federal Highway Administration (FHWA). (1995). Quality assurance procedures for construction. Final rule, 23 CFR 637, Federal Register, 60(125), 33712.
62. Feng, D.C., Liu, Z.T., Wang, X.D., Chen, Y., Chang, J.Q., Wei, D.F., & Jiang, Z.M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, 230, 117000.

63. FHWA, Quality Assurance in materials and Construction: Observations and Recommendations, Washington, D.C.
<<https://www.fhwa.dot.gov/construction/cqit/qamc0607/05.cfm>> accessed October 10, 2021.
64. Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.*, 20(177), 1-81.
65. Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*. San Francisco.
66. Gambatese, J. A., & Rajendran, S. (2005). Sustainable roadway construction: Energy consumption and material waste generation of roadways. In *Construction research congress 2005: Broadening perspectives* (pp. 1-13).
67. Ghanbari, M., Abbasi, A. M., & Ravanshadnia, M. (2018). Production of natural and recycled aggregates: the environmental impacts of energy consumption and CO₂ emissions. *Journal of Material Cycles and Waste Management*, 20, 810-822.
68. Goulias, D. G., & Ntekim, A. (2001). Durability evaluation of asphalt mixtures modified with recycled tire rubber. *Journal of Solid Waste Technology and Management*, 27(3), 169-173.
69. Goulias, D., Zhang, Y., & Aydilek, A. (2018). Sustainability Assessment of Roadways through Economic and Environmental Impact Life Cycle Analysis. *Int. J. Comput. Eng. Res*, 8, 1-8.
70. Goulias, D., Zhao, Y., & Peterson, D. (2020). Holistic Approach for Generating and Assessing Sustainable Rehabilitation Strategies in Roadway Construction. *International Journal of Computational Engineering Research (IJCER)*, 10(6), 2250-3005.
71. Goulias, D. G., & Ali, A. H. (1998). Evaluation of rubber-filled concrete and correlation between destructive and nondestructive testing results. *cement, concrete and aggregates*, 20(1).

72. Grogg, M. G. (2021). Quality Assurance Stewardship Review, Summary Report 2013 through 2018 (No. FHWA-HIF-20-059).
73. Grolemond, G. (2014). Hands-On Programming with R; OReilly Media: Sebastopol, CA
74. Gschösser, F., Wallbaum, H., & Boesch, M. E. (2012). Life-cycle assessment of the production of swiss road materials. *Journal of materials in civil engineering*, 24(2), 168-176.
75. Gupta, S. (2008). Support vector machines based modeling of concrete strength. *World Academy of Science, Engineering and Technology*, 36, 305-311.
76. Han, Q., Gui, C., Xu, J., & Lacidogna, G. (2019). A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Construction and Building Materials*, 226, 734-742.
77. Harvey, J., Meijer, J., Ozer, H., Al-Qadi, I. L., Saboori, A., & Kendall, A. (2016). Pavement life cycle assessment framework (No. FHWA-HIF-16-014). United States. Federal Highway Administration.
78. Hasan, U., Whyte, A., & Al Jassmi, H. (2020). Life cycle assessment of roadworks in United Arab Emirates: Recycled construction waste, reclaimed asphalt pavement, warm-mix asphalt and blast furnace slag use against traditional approach. *Journal of Cleaner Production*, 257, 120531.
79. Hossain, M. U., Poon, C. S., Lo, I. M., & Cheng, J. C. (2016). Comparative environmental evaluation of aggregate production from recycled waste materials and virgin sources by LCA. *Resources, Conservation and Recycling*, 109, 67-77.
80. Hoxha, E., Vignisdottir, H. R., Barbieri, D. M., Wang, F., Bohne, R. A., Kristensen, T., & Passer, A. (2021). Life cycle assessment of roads: Exploring research trends and harmonization challenges. *Science of the Total Environment*, 759, 143506.

81. Huang, Y., Bird, R. N., & Heidrich, O. (2007). A review of the use of recycled solid waste materials in asphalt pavements. *Resources, conservation and recycling*, 52(1), 58-73.
82. Hughes, C. S., Moulthrop, J. S., Tayabji, S., Weed, R. M., & Burati, J. L. (2011). Guidelines for quality-related pay adjustment factors for pavements. NCHRP Project, (10-79).
83. Ilangovana, R., Mahendrana, N., & Nagamanib, K. (2008). Strength and durability properties of concrete containing quarry rock dust as fine aggregate. *ARP Journal of Engineering and Applied Sciences*, 3(5), 20-26.
84. Jimenez, J. R., Ayuso, J., Agrela, F., López, M., & Galvín, A. P. (2012). Utilisation of unbound recycled aggregates from selected CDW in unpaved rural roads. *Resources, Conservation and Recycling*, 58, 88-97.
85. Kankam, C. K., Meisuh, B. K., Sossou, G., & Buabin, T. K. (2017). Stress-strain characteristics of concrete containing quarry rock dust as partial replacement of sand. Case studies in construction materials, 7, 66-72.
86. K Karimi, S. S., Goulias, D. G., & Schwartz, C. W. (2012). Evaluation of Superpave HMA mixture properties at the plant versus behind the paver: Statistical comparison of QC and QA data. *Journal of transportation engineering*, 138(7), 924-932.
87. Karimi, S. S., Goulias, D. G., & Schwartz, C. W. (2015). Risk and expected pay factor analysis for assessing gap and dense-graded Superpave mixture specifications. *International Journal of Pavement Engineering*, 16(1), 69-79.
88. Kaseer, F., Martin, A. E., & Arámbula-Mercado, E. (2019). Use of recycling agents in asphalt mixtures with high recycled materials contents in the United States: A literature review. *Construction and Building Materials*, 211, 974-987.

89. Killingsworth, B. M., & Hughes, C. S. (2002). Issues related to use of Contractor quality control data in acceptance decision and payment: Benefits and pitfalls. *Transportation research record*, 1813(1), 247-252.
90. Kim, H., & Goulias, D. G. (2015). Shrinkage behavior of sustainable concrete with crushed returned concrete aggregate. *Journal of Materials in Civil Engineering*, 27(7), 04014204.
91. Kosmatka, S. H., & Wilson, M. L. (2011). *Design and Control of Concrete Mixtures*. Portland Cement Association.
92. Kou, S. C., & Poon, C. S. (2009). Properties of concrete prepared with crushed fine stone, furnace bottom ash and fine recycled aggregate as fine aggregates. *Construction and Building Materials*, 23(8), 2877-2886.
93. Kumbhar, S., Gupta, A., & Desai, D. (2013). Recycling and reuse of construction and demolition waste for sustainable development. *OIDA International Journal of Sustainable Development*, 6(7), 83-92.
94. LaVassar, C. J., Mahoney, J. P., Willoughby, K. A., & Willoughby, K. (2009). Statistical assessment of quality assurance-quality control data for hot mix asphalt. Report No. WA-RD, 686.
95. Lawrence, M., Nguyen, P., Skolnick, J., Symoun, J., Hunt, J., & Alfelor, R. (2015). *Transportation systems management and operations benefit-cost analysis compendium* (No. FHWA-HOP-14-032). United States. Federal Highway Administration.
96. Lee, J., Edil, T. B., Benson, C. H., & Tinjum, J. M. (2013). Building environmentally and economically sustainable transportation infrastructure: green highway rating system. *Journal of Construction Engineering and Management*, 139(12), A4013006.

97. Lee, J. C., Edil, T. B., Tinjum, J. M., & Benson, C. H. (2010). Quantitative assessment of environmental and economic benefits of recycled materials in highway construction. *Transportation research record*, 2158(1), 138-142.
98. Ling, H., Qian, C., Kang, W., Liang, C., & Chen, H. (2019). Combination of Support Vector Machine and K-Fold cross-validation to predict compressive strength of concrete in marine environment. *Construction and Building Materials*, 206, 355-363.
99. Mahboub, K. C., Hancher, D. E., & Wang, Y. (2004). Contractor-performed quality control: Is the fox guarding the henhouse?. *Journal of Professional Issues in Engineering Education and Practice*, 130(4), 255-258.
100. Maryland State Highway Administration. *Standard Specifications for Construction and Materials*, section 504, 2021. MDOT. Hanover, MD.
101. Marzouk, M., & Azab, S. (2014). Environmental and economic impact assessment of construction and demolition waste disposal using system dynamics. *Resources, conservation and recycling*, 82, 41-49.
102. Meisuh, B. K., Kankam, C. K., & Buabin, T. K. (2018). Effect of quarry rock dust on the flexural strength of concrete. *Case studies in construction materials*, 8, 16-22.
103. Miladirad, K., Golafshani, E. M., Safehian, M., & Sarkar, A. (2021). Modeling the mechanical properties of rubberized concrete using machine learning methods. *Computers and Concrete*, 28(6), 567-583. [https://DOI: 10.12989/cac.2021.28.6.567](https://doi.org/10.12989/cac.2021.28.6.567)
104. Mohammadinia, A., Arulrajah, A., D'Amico, A., & Horpibulsuk, S. (2018). Alkali-activation of fly ash and cement kiln dust mixtures for stabilization of demolition aggregates. *Construction and building materials*, 186, 71-78.
105. Muench, S. T. (2010). Roadway construction sustainability impacts: review of life-cycle assessments. *Transportation research record*, 2151(1), 36-45.

106. Nathman, R., McNeil, S., & Van Dam, T. J. (2009). Integrating environmental perspectives into pavement management: Adding the pavement life-cycle assessment tool for environmental and economic effects to the decision-making toolbox. *Transportation Research Record*, 2093(1), 40-49.
107. National Academies of Sciences, Engineering, and Medicine. (2007). *Using the Results of Contractor-Performed Tests in Quality Assurance*.
108. Hand, A. J., Nimeri, M. A., Hajj, E. Y., Sebaaly, P. E., West, R. C., Heitzman, M. A., ... & Tayabji, S. (2020). *Procedures and Guidelines for Validating Contractor Test Data* (No. Project 10-100).
109. Stroup-Gardiner, M. (2011). *Recycling and reclamation of asphalt pavements using in-place methods* (No. Project 20-05 (Topic 40-13)).
110. National Cooperative Highway Research Program (NCHRP). *Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures*. Available online: <http://www.trb:mepdg/software.html> (accessed on 25 May 2021).
111. Nguyen-Sy, T., Wakim, J., To, Q.D., Vu, M.N., Nguyen, T.D., & Nguyen, T.T. (2020). Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. *Construction and Building Materials*. 260, 119757.
112. Ni, H.G. & Wang, J.Z. (2000). Prediction of compressive strength of concrete by neural networks. *Concrete and Cement Research*. 30, 1245-1250.
113. Nicodème, C.; Diamandouros, K.; Diez, J.; Durso, C.; Arampidou, K.; Nuri, A.K. *Road Statistics Yearbook*. 2017. Available online: www.erf.be (accessed on 15 August 2022).
114. Nimeri, M. A. (2019). *Statistical Validation Implementation in Pavement Construction Projects* (Doctoral dissertation, University of Nevada, Reno).

115. Oner, E. (2021). Frictionless contact mechanics of an orthotropic coating/isotropic substrate system. *Computers and Concrete*, 28(2), 209-220. <https://doi.org/10.12989/cac.2021.28.2.209>
116. OpenLCA. 2021. Available online: www.openlca.com (accessed on 24 May 2021).
117. Organization for Economic Cooperation and Development (OECD). The Environmental Effects of Freight. Available online: www.oecd.org/environment/envtrade/2386636.pdf (accessed on 25 May 2021).
118. Ossa, A., García, J. L., & Botero, E. (2016). Use of recycled construction and demolition waste (CDW) aggregates: A sustainable alternative for the pavement construction industry. *Journal of Cleaner Production*, 135, 379-386.
119. PaLATE. Pavement Life-Cycle Assessment Tool for Environmental and Economic Effects Version 2.0. Consortium on Green Design and Manufacturing, University of California, Berkeley. 2022. Available online: <https://rmrc.wisc.edu/palate/> (accessed on 14 May 2022).
120. Pavement Life-Cycle Assessment Tool for Environmental and Economic Effects Version 2.0. Consortium on Green Design and Manufacturing, University of California, Berkeley. Available online: rmrc.wisc.edu/palate/ (accessed on 14 May 2021).
121. Popovics, S., Ujhelyi, J. (2008). Contribution to the concrete strength versus water cement ratio relationship, *Journal of Materials in Civil Engineering*. 20 (7), 459–463.
122. Pourkhorshidi, S., Sangiorgi, C., Torreggiani, D., & Tassinari, P. (2020). Using recycled aggregates from construction and demolition waste in unbound layers of pavements. *Sustainability*, 12(22), 9386.
123. Ram, V. G., Kishore, K. C., & Kalidindi, S. N. (2020). Environmental benefits of construction and demolition debris recycling: Evidence from an Indian case study using life cycle assessment. *Journal of Cleaner Production*, 255, 120258.

124. Recycled Materials Resource Centre. Building Environmentally and Economically Sustainable Transportation-Infrastructure-Highways. BE2ST-in-Highways User's Manual; University of Wisconsin-Madison: Madison, WI, USA, 2012.
125. Rodríguez, G., Medina, C., Alegre, F. J., Asensio, E., & De Rojas, M. S. (2015). Assessment of construction and demolition waste plant management in Spain: in pursuit of sustainability and eco-efficiency. *Journal of Cleaner Production*, 90, 16-24.
126. Rosado, L. P., Vitale, P., Penteado, C. S. G., & Arena, U. (2017). Life cycle assessment of natural and mixed recycled aggregate production in Brazil. *Journal of cleaner production*, 151, 634-642.
127. Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. *Engineering Structures*, 171, 170-189.
128. Schmitt, R. L., Hanna, A. S., Russell, J. S., & Nordheim, E. V. (2001). Statistically based methods for verification testing. *Transportation research record*, 1761(1), 86-92.
129. Serres, N., Braymand, S., & Feugeas, F. (2016). Environmental evaluation of concrete made from recycled concrete aggregate implementing life cycle assessment. *Journal of Building Engineering*, 5, 24-33.
130. Sherwood, P.T. (2001). *Alternative Materials in Road Construction: A Guide to the Use of Waste, Recycled Materials and By-Products*, 2nd ed.; Thomas Telford.
131. Shi, X., Mukhopadhyay, A., Zollinger, D., & Grasley, Z. (2019). Economic input-output life cycle assessment of concrete pavement containing recycled concrete aggregate. *Journal of cleaner production*, 225, 414-425.
132. Siddique, R., Khatib, J., & Kaur, I. (2008). Use of recycled plastic in concrete: A review. *Waste management*, 28(10), 1835-1852.

133. Silva, R. V., De Brito, J., & Dhir, R. K. (2019). Use of recycled aggregates arising from construction and demolition waste in new construction applications. *Journal of Cleaner Production*, 236, 117629.
134. Simion, I. M., Fortuna, M. E., Bonoli, A., & Gavrilescu, M. (2013). Comparing environmental impacts of natural inert and recycled construction and demolition waste processing using LCA. *Journal of Environmental Engineering and Landscape Management*, 21(4), 273-287.
135. Soroka, I., & Setter, N. (1977). The effect of fillers on strength of cement mortars. *Cement and concrete research*, 7(4), 449-456.
136. Stroup-Gardiner, M. (2011). Recycling and reclamation of asphalt pavements using in-place methods (No. Project 20-05 (Topic 40-13)).
137. Tahwia, A. M., Heniegal, A., Elgamal, M. S., & Tayeh, B. A. (2021). The prediction of compressive strength and non-destructive tests of sustainable concrete by using artificial neural networks. *Computers and Concrete, An International Journal*, 27(1), 21-28.
138. Tam, V. W., Soomro, M., & Evangelista, A. C. J. (2018). A review of recycled aggregate in concrete applications (2000–2017). *Construction and Building materials*, 172, 272-292.
139. Tanyildizi, H. (2022). Predicting bond strength of corroded reinforcement by deep learning. *Computers and Concrete*, 29(3), 145-159.
140. Tefa, L., Bassani, M., Coppola, B., & Palmero, P. (2021). Strength development and environmental assessment of alkali-activated construction and demolition waste fines as stabilizer for recycled road materials. *Construction and Building Materials*, 289, 123017.
141. Thives, L. P., & Ghisi, E. (2017). Asphalt mixtures emission and energy consumption: A review. *Renewable and Sustainable Energy Reviews*, 72, 473-484.
142. Transportation Research Circular E-C235: Glossary of Highway Quality Assurance Terms. Transportation Research Board of the National Academies, Washington, D.C., 2018.

143. Turochy, R. E., & Parker, F. (2007). Comparisons of contractor and state transportation agency quality assurance test results on mat density of hot-mix asphalt concrete: Findings of multistate analysis. *Transportation research record*, 2040(1), 41-47.
144. Upadhyaya, S., Goulias, D., & Obla, K. (2015). Maturity-based field strength predictions of sustainable concrete using high-volume fly ash as supplementary cementitious material. *Journal of Materials in Civil Engineering*, 27(5), 04014165.
145. Van der Laan, Mark J.; Polley, Eric C.; and Hubbard, Alan E., "Super Learner" (2007). U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 222.
146. Vidal, R., Moliner, E., Martínez, G., & Rubio, M. C. (2013). Life cycle assessment of hot mix asphalt and zeolite-based warm mix asphalt with reclaimed asphalt pavement. *Resources, Conservation and Recycling*, 74, 101-114.
147. Vieira, C. S., Lopes, M. L., & Cristelo, N. (2017). Geotechnical characterization of recycled C&D wastes for use as trenches backfilling. In *WASTES—Solutions, Treatments and Opportunities II* (pp. 175-181). CRC Press.
148. Wani, S. S., & Gharaibeh, N. G. (2013). Evaluating the use of contractor-performed test results in highway construction and material acceptance decisions. *Journal of infrastructure systems*, 19(1), 67-73.
149. Williams, B. A., Willis, J. R., & Shacat, J. (2020). Asphalt pavement industry survey on recycled materials and warm-mix asphalt usage: 2019 (No. IS 138 (10e)).
150. Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, 28(12), 1797-1808.
151. Young, B. A., Hall, A., Pilon, L., Gupta, P., & Sant, G. (2019). Can the compressive strength of concrete be estimated from knowledge of the mixture proportions? New insights from

- statistical analysis and machine learning methods. *Cement and Concrete Research*, 115, 379-388.
152. Yuan, H. P., Shen, L. Y., Hao, J. J., & Lu, W. S. (2011). A model for cost–benefit analysis of construction and demolition waste management throughout the waste chain. *Resources, conservation and recycling*, 55(6), 604-612.
153. Zhang, J., Ma, G., Huang, Y., Sun, J., Aslani, F., & Nener, B., (2019) Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. *Construction and Building Materials*, 210, 713-719.
154. Zhang, J., Ding, L., Li, F., & Peng, J. (2020). Recycled aggregates from construction and demolition wastes as alternative filling materials for highway subgrades in China. *Journal of Cleaner Production*, 255, 120223.
155. Zhang, J., Gu, F., & Zhang, Y. (2019). Use of building-related construction and demolition wastes in highway embankment: laboratory and field evaluations. *Journal of cleaner production*, 230, 1051-1060.
156. Zhang, X., Akber, M.Z., & Zhang, W. (2021). Prediction of seven-day compressive strength of field concrete. *Construction and Building Materials*. 305, 124604.
157. Zhao, Y., & Goulias, D. (2021). Evaluation of Acceptance Risks through Percent within Limit for Highway Materials and Pavement Construction. *Transportation Research Record*, 2675(9), 1175-1184.
158. Zhao, Y., & Goulias, D. (2021). Pay Factor Analysis for Highway Pavements through Percentage of Material within Specification Limits and Monte Carlo Simulation. *Transportation Research Record*, 2675(11), 1258-1271.

159. Zhao, Y., & Goulias, D. (2019). Systematic Assessment of Sustainability Strategies in Highway Projects through Parametric Analysis. *International Journal of Computer Engineering and Research*, 9, 13-20.
160. Zhao, Y., Goulias, D., & Peterson, D. (2021). Recycled Asphalt Pavement materials in transport pavement infrastructure: Sustainability analysis & metrics. *Sustainability*, 13(14), 8071.
161. Zhao, Y., Goulias, D., & Peterson, D. (2021). Recycled Asphalt Pavement materials in transport pavement infrastructure: Sustainability analysis & metrics. *Sustainability*, 13(14), 8071.
162. Zhao, Y., Goulias, D., Tefa, L., & Bassani, M. (2021). Life cycle economic and environmental impacts of CDW recycled aggregates in roadway construction and rehabilitation. *Sustainability*, 13(15), 8611.