**ADVANCEMENTS IN EVALUATING RELIABILITY
OF NONDESTRUCTIVE TECHNOLOGIES FOR THE
DETECTION OF SUBSURFACE FRACTURE
DAMAGE IN R.C. BRIDGE DECKS**

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
ALI ABED SULTAN

Dr. Glenn Washer, Dissertation Supervisor

July 2017

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

**ADVANCEMENTS IN EVALUATING RELIABILITY OF
NONDESTRUCTIVE TECHNOLOGIES FOR THE DETECTION
OF SUBSURFACE FRACTURE DAMAGE IN R.C. BRIDGE
DECKS**

presented by **Ali Abed Sultan**,
a candidate for the degree of **Doctor of Philosophy of Civil Engineering**,
and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Glenn Washer

Dr. Hani Salim

Dr. Brent Rosenblad

Dr. Mahmoud Almassri

Dedication

This dissertation is dedicated

To my Parents,

To my Sisters and Brothers, and

To my Family,

Without their patience and devotion, this dissertation would not be.

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to those who have contributed to this thesis and supported me in one way or the other during my PhD. journey.

First and foremost, I am extremely grateful to my advisor, Professor Glenn Washer, for his guidance, support, and all the useful discussions and brainstorming sessions, especially during the difficult conceptual development stage. His deep insights helped me at various stages of my research. Even during the most difficult times when writing this thesis, he gave me the moral support and all the freedom I needed to move on. I also remain indebted for his empathy and support during the times when I was really down and depressed due to personal family problems. It was a great honor and privilege for me to work with Professor Washer during my PhD. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Professor Brent Rosenblad, Professor Mahmoud Almassri, and Associate Dean of Academic Programs, Professor Hani Salim for their valuable feedbacks and insights.

Very special thanks is reserved for Mrs. Karen Washer for proofreading the technical papers published as part of this research and the final draft of this dissertation. Her valuable comments and suggestions on almost every page of the dissertation were greatly helpful. Further, an enormous thank to my academic professors especially Professor Vellore Gopalaratnam, Professor Steven Neal, and Professor Zhen Chen at the University of Missouri. Special thanks goes to Charles Nemmers, Director of the Transportation Infrastructure Center, for his support and efforts for providing me the opportunity to attend the Transportation Research Board (TRB) conference, Washington D.C. in 2014. The experience I got from attending this conference was very useful and has guided me throughout my research phases.

PhD students often talk about loneliness during the course of their study, but

this is something I never experienced at MU. A heartfelt thanks to the really supportive and active NDE group here in MU and to all my fellow labmates and friends who made the PhD experience something special. In particular, I would like to thank Michael Trial, Alan Jungnitsch, Habibulrahman Hotaki, James Dawson, Pedro Ruiz Fabian, Massoud Nasrollahi, Hammed Mujtaba, and Henry Brown for their help, support, and for all the stimulating discussions we have had.

Special thanks goes to Missouri Department of Transportation (MoDOT) for providing the necessary information regarding the testing bridges and their actual condition as determined from the routine inspection. Also, I greatly appreciate the MoDOT personnels who accompanied the research team during the field testing. Information regarding the testing bridges and the traffic control that MoDOT personnel provided during testing were of great importance for getting the necessary NDE data reliably and safely.

I am also indebted to my friends Falih Alnahwi, Mohannad Salim, and Mohannad Al-Ali for their help in developing the MATLAB script necessary for the analysis of IRT data.

Words cannot express the feelings I have for my parents who, unfortunately, could not live longer to see this achievement. I know that they can feel my happiness while I am writing this part of the dissertation, and they would certainly know that I am almost done.

Finally, I would like to acknowledge and thank the most important person in my life, my wife, Suraa Makki. She has been a constant source of strength and inspiration. There were times during the past five years when everything seemed hopeless and I did not have any hope. I can honestly say that it was only her determination and constant encouragement that ultimately made it possible for me to see this project through to the end.

Contents

ACKNOWLEDGMENTS	ii
LIST OF FIGURES	viii
LIST OF TABLES	xvi
ABSTRACT	xviii
1 INTRODUCTION	1
1.1 Background	1
1.2 Destructive vs. Nondestructive Testing	4
1.3 Nondestructive Inspection (NDI) vs. Nondestructive Testing (NDT) vs. Nondestructive Evaluations (NDE)	6
1.4 Problem Statement/Motivation	6
1.5 Goals and Objectives	8
1.6 Outlines of the Dissertation	9
2 LITERATURE REVIEW	11
2.1 Mechanisms of Reinforced Concrete Bridge Deck Delamination . . .	11
2.1.1 Corrosion of reinforcing steel	11
2.1.2 Improper concrete production	14
2.1.3 Freezing-thawing cycles	14
2.2 NDE Technologies of Subsurface Damage Detection	15
2.2.1 Impact Echo (IE)	15
2.2.2 Infrared Thermography (IRT)	24
2.2.3 Ground Penetrating Radar (GPR)	28
2.3 Earlier Works on the Detection of Delamination	36
2.3.1 Earlier work of IE technology	36
2.3.2 Earlier work of IRT technology	39

2.3.3	Earlier work of GPR technology	41
3	RELIABILITY ASSESSMENT METHOD/REVIEW	46
3.1	Introduction	46
3.2	Factors Affecting the Reliability of NDE	48
3.3	Probability of Detection (POD) Analysis	49
3.3.1	History and applications	49
3.3.2	Statistical analysis	53
3.3.3	Sample size requirements	59
3.4	Receiver Operating Characteristic (ROC) Analysis	60
3.4.1	Measurements outcomes	62
3.4.2	ROC curve construction	65
3.4.3	Area under the ROC curve (AUC)	67
4	DESCRIPTION OF THE EXPERIMENTAL WORK	70
4.1	Fabricated Slab I	71
4.2	Fabricated Slab II	73
4.3	Kansas City Bridge, Bridge A0295	75
4.4	Fulton City Bridge, Bridge A2111	77
4.5	Fulton City Bridge, Bridge A2112	78
5	ANALYSIS AND RESULTS OF IMPACT ECHO (IE) TECHNOLOGY	80
5.1	Introduction	80
5.2	Results and Analysis of Fabricated slab I	82
5.3	Results and Analysis of Fabricated slab II	88
5.4	Results and Analysis of Bridges A0295 & A2111	95
6	ANALYSIS AND RESULTS OF INFRARED THERMOGRAPHY (IRT) TECHNOLOGY	104
6.1	Introduction	104

6.2	Data Acquisition	104
6.3	IRT Data Processing	106
6.4	Results of ROC Analysis	115
7	ANALYSIS AND RESULTS OF GROUND PENETRATING RADAR (GPR) TECHNOLOGY	122
7.1	Introduction	122
7.2	Testing Parameters and Data Analysis	123
7.3	Analysis and Results of Fabricated Slab I	127
7.4	Analysis and Results of Bridge A0295	133
7.5	Analysis and Results of Bridge A2111	140
7.6	Analysis and Results of Bridge A2112	145
8	CONCLUSIONS AND RECOMMENDATIONS	155
8.1	Introduction	155
8.2	Conclusions	156
8.2.1	Impact Echo (IE)	156
8.2.2	Infrared Thermography (IRT)	158
8.2.3	Ground Penetrating Radar (GPR)	160
8.2.4	ROC Analysis	162
8.3	Overall Summary	163
8.4	Recommendations	164
	BIBLIOGRAPHY	165
	APPENDICES	178
A	MATLAB Script & Functions	179
A.1	Main script	179
A.2	Functions	181

B	Details of Mockups and In-service Bridge Decks Images	182
B.1	Images of the fabricated slab I	182
B.2	Images of the in-service bridge decks	184
VITA		186

LIST OF FIGURES

2.1	Corrosion mechanism of steel reinforcement embedded in concrete .	13
2.2	Spalling of concrete due to unrepaired sub-surface delamination. . .	14
2.3	Stress waves in semi-infinite material.	16
2.4	Principles of IE testing.	19
2.5	Portable data acquisition system for multi-nondestructive testing [30].	22
2.6	Time and frequency domains of sound (A & B) and flawed (C & D) concrete determined using IE testing.	23
2.7	Mechanism of infrared radiation in presence of subsurface delamina- tion during daytime (A) and night time (B)	26
2.8	Heat sources commonly used in active thermography experiments: (A) heat gun, (B) thermal blanket, (C)halogen lamp, and (D) flash lamp [37].	27
2.9	Principles of GPR testing.	30
2.10	Polarity of the reflected EM waves at different interfaces.	31
2.11	Ground-coupled GPR system manufactured by Geophysical Survey System, Inc. (GSSI).	34
2.12	GSSI air-lunched, vehicle-mounted GPR system.	35
2.13	"Zig-Zag" GPR testing.	35
3.1	Range interval POD analysis with the 90% lower confidence of an NDE data [7, 88].	50
3.2	Non-overlapping POD analysis with the 90% lower confidence of an NDE data [88].	51
3.3	Overlapping POD analysis with the 90% lower confidence of an NDE data [88].	52
3.4	Main steps of POD analysis for hit/miss and a vs. $\hat{a}$ analyses. . . .	54
3.5	POD analysis of qualitative NDE data showing the detectability at cracks of interest [7].	57

3.6	POD analysis of quantitative NDE data showing the detectability at cracks of interest [7].	59
3.7	Overlapping distributions of arbitrary test results over the actually positive and negative instances.	65
3.8	ROC curves of inspection techniques with different performances. .	67
4.1	Concrete blocks to investigate different materials simulating delami- nation.	72
4.2	IE frequency domains of different materials simulating delamination (A) sound concrete, (B) delaminated: Styrofoam, (C) delaminated: plastic sheets, and (D) delaminated:bubble wrap.	73
4.3	Size, depth, and horizontal distribution of delamination targets of fabricated slab I.	74
4.4	Details of the reinforcement and delamination targets of fabricated slab I.	75
4.5	Size, depth, and horizontal distribution of delamination targets of fabricated slab II.	75
4.6	Details of the reinforcement and delamination targets of fabricated slab II.	76
4.7	Aerial view of Bridge A0295 in Kansas City, Missouri.	77
4.8	A close view of the deck of bridge A0295 showing no symptoms of previous repairs.	77
4.9	Aerial view of Bridges A2111 and A2112 in Fulton City, Missouri. .	78
4.10	A close view of the deck surface of Bridge A2111 showing previous repairs.	78
4.11	A close view of the deck surface of Bridge A2112 showing concrete spalling and some previous repairs.	79
5.1	Components of impact echo instrumentation.	81

5.2	Time and frequency domains of an arbitrary sound test point before (A) and after (B) processing.	82
5.3	Frequency domain of an arbitrary processed test point of shallow delaminated concrete.	84
5.4	A contour map showing the frequency of the reflected P-wave of the first fabricated slab.	84
5.5	Distribution of IE test results over sound and delaminated test points.	85
5.6	ROC curve of IE technology for the detection of shallow delamination.	88
5.7	POD and POFA of IE technology for the detection of shallow delamination as functions of the threshold settings.	88
5.8	Time (A) and frequency (B) domains of an arbitrary deep delami- nated test point.	89
5.9	Contour maps of the frequency of the reflected P-waves of the individual groups in the second slab.	91
5.10	Distribution of IE test results over sound and delaminated test points.	92
5.11	ROC curves of IE for the detection of delamination of different size and depth.	93
5.12	POD and POFA of IE for the detections of delamination of different sizes and depths as functions of the decision threshold settings. . . .	95
5.13	IE tested section on bridge A0295, Kansas City, MO.	96
5.14	A contour map of the frequency of the reflected P-waves.	98
5.15	Distributions of the IE test results over sound and delaminated test points.	98
5.16	ROC curve of IE technology for the detection of real delamination in Bridge A0295, Kansas City, MO.	99
5.17	POD and POFA of IE technology as function of the decision thresh- old for the detection of real delamination in Bridge A0295, Kansas City, MO.	99
5.18	Tested section of bridge A2111 in Fulton, MO.	100

5.19	A contour map of the frequency of the reflected P-Waves.	100
5.20	Distributions of the IE test results over sound and delaminated test points.	101
5.21	The actual condition of bridge A2111, Fulton, MO; (A) results from sounding, (B) results from infrared thermography, and (C) a combination of (A) and (B).	102
5.22	ROC curve of IE technology for the detection of delamination. . . .	103
5.23	POD and POFA of IE technology as functions of the decision threshold setting of bridge A2111.	103
6.1	IRT testing setup: (A) fabricated slabs, (B) in-service bridge decks. . .	105
6.2	Effects of threshold setting in the detection of delamination using IRT technology (A) high Δt , (B) moderate Δt , and (C) low Δt . . .	106
6.3	Process of image registration (A) before, (B) during, and (C) after registration.	108
6.4	Exaggerated views of the reference points of of Figure 4-33.	109
6.5	Transformation of digital image to a binary-coded image: (A) Original, (B) changing the color temperature, and (C) as read by MATLAB script.	110
6.6	Transformation of digital and thermal images into binary-coded format: (A) A digital image, (B) the digital image as in a binary-coded format, and (C) the corresponding thermal image by setting a specific threshold for the thermal contrast.	112
6.7	Work-flow of the proposed analysis method.	113
6.8	Filtering out irrelevant thermal data resulting from out-of-control objects.	114
6.9	Distributions of the thermal contrast of the actually negative and positive pixels before (A) and after (B) filtering out the irrelevant data of the image shown in the previous figure.	115
6.10	Results of ROC analysis of the fabricated slab I thermal images. . .	116

6.11	Results of ROC analysis of the in-service bridge decks thermal images.	117
6.12	POD and POFA as functions of the decision threshold (thermal contrast between deteriorated and sound locations) of the fabricated slab images.	118
6.13	POD and POFA as functions of the decision threshold (thermal contrast between deteriorated and sound locations) of the in-service bridge decks images.	119
6.14	Averaged ROC curve and the upper and lower 95% confidence of all 22 images.	120
6.15	POD and POFA as functions of the decision threshold (thermal contrast between deteriorated and sound locations) of all 22 images.	120
7.1	An arbitrary GPR scan before and after processing.	124
7.2	An arbitrary GPR scan before and after stretching to enhance the quality of GPR data.	125
7.3	Contour map of raw GPR data of the fabricated slab.	128
7.4	GPR test results of the actually sound and delaminated test points of the fabricated RTF slab.	128
7.5	ROC curve of raw GPR data of the fabricated RTF slab.	129
7.6	POD and POFA of raw GPR data as functions of the reflection amplitude of the RTF fabricated slab.	130
7.7	Reflection amplitude versus TWTT of all dataset of the RTF fabricated slab.	131
7.8	Line fitting of the 90 th percentile of the reflection amplitudes at each TWTT bin.	131
7.9	Depth-corrected amplitude versus TWTT of all dataset.	132
7.10	Depth-corrected contour of GPR data of the RTF fabricated slab. .	132
7.11	ROC of depth-corrected GPR data of the RTF fabricated slab. . . .	133
7.12	POD and POFA of depth-corrected GPR data as functions of the reflection amplitude of the RTF fabricated slab.	133

7.13	Contour plot of raw GPR data of bridge A0295.	135
7.14	GPR test results of the actually sound and delaminated test points of bridge A0295.	135
7.15	ROC curve of raw GPR data of A0295 bridge deck.	136
7.16	POD and POFA of raw GPR data as functions of the reflection amplitude of bridge A0295.	136
7.17	Reflection amplitude versus TWTT of all dataset of bridge A0295. .	137
7.18	Line fitting of the 90 th percentile of the reflection amplitudes at each TWTT bin	137
7.19	Depth-Corrected amplitudes versus TWTT of all dataset of bridge A0295	138
7.20	Depth-Corrected contour of GPR data of bridge A0295	138
7.21	ROC curve of depth-corrected data of bridge A0295.	139
7.22	POD and POFA of depth-corrected GPR data of bridge A0295 as functions of the reflection amplitude setting.	139
7.23	Images of different NDE field testing performed on bridge A2111. .	141
7.24	Contour of raw (depth-uncorrected) GPR data of bridge A2111. . .	142
7.25	GPR test results of the actually sound and delaminated test points of bridge A2111.	142
7.26	ROC curve of raw GPR data of A2111 bridge deck.	143
7.27	POD and POFA of raw GPR data of bridge A2111 as functions of the reflection amplitude.	144
7.28	Reflection amplitude versus TWTT of all dataset of bridge A2111. .	145
7.29	Line fitting of the 90 th percentile of the reflection amplitudes at each TWTT bin.	145
7.30	Depth-corrected amplitudes as function of the TWTT of all dataset.	146
7.31	Contour map of the depth-corrected amplitudes.	146
7.32	ROC curve of depth-corrected GPR data of bridge A2111.	147

7.33	POD and POFA of depth-corrected GPR data of bridge A2111 as functions of the reflection amplitude setting.	147
7.34	The actual condition of Bridge A2112 deck.	148
7.35	A contour map of the depth-uncorrected GPR signals of the deck of A2112 bridge.	148
7.36	Results of GPR testing over the actually sound and delaminated concrete areas.	149
7.37	ROC curve of depth-uncorrected GPR data of Bridge A2112. . . .	150
7.38	POD and POFA of depth-uncorrected GPR data of bridge A2112 as functions of the reflection amplitude setting.	150
7.39	Amplitude versus TWTT of GPR data of Bridge A2112.	151
7.40	The best fitting line over the 90 th percentile data.	152
7.41	Amplitude versus TWTT of depth-corrected data of Bridge A2112. .	152
7.42	Depth-corrected contour of GPR data of Bridge A2112.	153
7.43	ROC curve of depth-corrected data of bridge A2112.	153
7.44	POD and POFA of depth-corrected GPR data as function of the reflection amplitude of Bridge A2112.	154
A.1	The developed MATALB script necessary to analyze data of IRT technology	180
A.2	Functions used in the above MATLAB script	181
B.1	Details of the thermal and digital images of the fabricated slab I. .	182
B.2	Distributions of the thermal contrast over the actually negative and positive locations (pixels) of the fabricated slab I thermal images. .	183
B.3	ROC curves of the fabricated slab I thermal images.	183
B.4	Details of the thermal and digital images of the in-service bridge decks.	184

B.5	Distributions of the thermal contrast over the actually negative and positive locations (pixels) of the in-service bridge decks thermal images.	185
B.6	ROC curves of the in-service bridge decks thermal images.	185

LIST OF TABLES

2.1	Speed Estimations of Stress Waves.	17
2.2	Specific Acoustic Impedance of Some Materials and the Reflection Coefficients of P-wave in Concrete [20].	18
2.3	Relative Dielectric Constant of Some Construction Materials. [38] .	32
2.4	Depth of Penetration and Possible Applications of Some Antenna Units.	33
3.1	A Contingency Table of an Arbitrarily Inspection Technique.	63
3.2	A Contingency Table of an Arbitrarily Inspection Technique Using Conditional Probability Notations.	63
5.1	Statistics of P-wave Velocity Estimation (in <i>ft/s</i>) of the Fabricated Slabs.	83
5.2	Estimations of Delaminated and Sound Areas as Functions of The Decision Threshold Setting.	87
5.3	AUC Estimations of IE for the Detection of delamination of Various Sizes and Depths.	94
5.4	Statistics of P-wave Velocity Estimation (in <i>ft/s</i>) of A0295 and A2111 Bridge Decks.	97
6.1	Specifications of Flir T-620 Thermal Camera Used in Field Testing.	105
6.2	Statistics of AUC Estimation of IRT Images.	119
7.1	Reliability Indices of GPR Testing of the RTF Fabricated Slab with Five Arbitrary Threshold Settings.	129
7.2	Reliability Indices of GPR Testing of Bridge A0295 with Five Arbitrary Threshold Settings.	134
7.3	Chloride Ion Concentrations of Four Decks in Missouri State.	140

7.4	Reliability Indices of GPR Testing of Bridge A2111 with Five Arbitrary Threshold Settings.	143
7.5	Reliability Indices of GPR Testing of Bridge A2112 with Five arbitrary Threshold Settings.	149
8.1	Summary of ROC Reliability Analysis of the NDE Technologies Investigated in This Research	163

Advancements in Evaluating Reliability of Nondestructive Technologies for the Detection of Subsurface Fracture Damage in R.C. Bridge Decks

Ali Abed Sultan

Dr. Glenn Washer, Dissertation Supervisor

ABSTRACT

During the last few decades, many efforts have been made to assess the reliability of nondestructive evaluation (NDE) technologies used for the detection of subsurface damage in concrete bridge decks. During these efforts, reliability of NDE technologies has either been described anecdotally, or been solely relegated to the probability of detection (POD) or accuracy estimation. Although these indices are important, most of the previous work did not take into account the probability of false alarm (POFA) of NDE technologies, nor did they investigate the reliability considering multiple threshold settings throughout test results. In addition, the existing body of research has used a limited physical sampling such as coring to validate NDE results. Consequently, the assessments were rather controversial, and there was no general agreement about the reliability of such technologies.

Because most diagnosis systems are characterized by noisy data and less than-perfect detection characteristics, reliability is to be carefully assessed considering all possible diagnosis output with multiple threshold settings within practical range of applications. In other words, when NDE data do not fall into either of the two obviously defined categories: true positive (TP), meaning the NDE data indicates a defect and there is a defect, or true negative (TN), meaning the NDT data indicates no defect and there is no defect, reliability analysis should also include the two types of incorrect indications: failure to give a positive indication in the

presence of a defect (false negative, FN) and giving a positive indication when there is no defect (a false alarm or false positive, FP).

The “three decades of NDI reliability assessments” report developed by Karta Technologies, Inc. in 2000 under supervision of the Air Force NDI Office stated that POD alone cannot describe the reliability of NDE technologies unless the probability of false alarm (POFA) is also considered in the analysis. POFA may be induced by noise with several possible sources: human, nature of phenomenon to be measured, and environmental conditions. The report covered nearly 150 reports and manuscripts from over 100 authors. However, a review of research literature reveals that little theoretical work on the reliability assessment in terms of both POD and POFA has been undertaken since then.

In this research, the reliability of impact echo (IE), infrared thermography (IRT), and ground penetration radar (GPR) technologies for the detecting of subsurface damage in concrete plate-like members is assessed by using a statistical analysis method called receiver operating characteristic (ROC). The proposed analysis method has the capability to integrate POD and POFA indices over a wide range of decision threshold settings in a single curve, which is useful in assessing trade-off in choosing a threshold and for quantitatively comparing the performance of NDE technologies. This methodology for assessing NDE reliability is intended to provide a more effective means of comparing different technologies used in civil engineering applications, to make the evaluation process of a quantitative scheme, to reduce subjectivity and variability in interpreting NDE data, and to improve sensitivity to extract more information from NDE data. Area under ROC curve (AUC), which is interpreted as the probability of correctly classifying an arbitrarily pair of negative and positive test points, can provide for the desired quantitative reliability index, which can be used to compare the performance of one NDE technology to another.

Results of this research obtained from ROC analysis indicate a great ability of IE and IR in detecting subsurface fracture damage such as delamination and debonding. In both technologies, there exist some threshold settings that can

provide for a relatively high POD with very low POFA, and consequently, the areas under their ROC curves were very high. Data obtained from GPR testing, on the other hand, indicates that GPR technology has a very limited ability to detect physical damage such as subsurface delamination. This conclusion contrasts with that been argued by a large body of the previous work. However, GPR showed a good sensitivity to the presence of corrosive environments such as moisture and chloride when the concentrations of these factors are above some threshold values that may facilitate the initiation of steel reinforcement corrosion.

Chapter 1

INTRODUCTION

1.1 Background

Bridges are the backbone of the transportation industry. Concerns about the integrity of in-service bridges, especially after the catastrophic collapses of some bridge structures, have led to the development of several nondestructive evaluation (NDE) technologies. Condition assessment of bridge elements is one of the most significant tasks of the developed NDE technologies. The assessment includes, but is not limited to, the detection of subsurface anomalies such as corrosion, delamination, concrete quality degradations, and fatigue cracks in concrete bridge decks which are normally very difficult to detect through visual inspections.

Furthermore, bridges are often operated beyond their original design life. As these infrastructures age, they become more vulnerable to different kinds of deterioration phenomena, and therefore, their maintenance costs increase. Subsurface deterioration mechanisms such as corrosion, delamination, and fatigue damage are examples of serious threats to aging infrastructures. When this kind of deterioration is combined with defects of design, construction and accidental damage, increasing traffic volume and overloading, risk of loss of the load carrying capacity of the deteriorated bridges is inevitable. At the same time, operation and maintenance of aging infrastructures such as bridges have become more complex both individually and globally.

Visual inspection and sounding are the most widely NDE techniques used for the inspection of the concrete bridge decks. Visual inspection is performed by a trained inspector who evaluates and ranks the condition of the bridge on a predefined scale [1]. Sounding method is performed by either dragging a chain of suitable size against the surface of the bridge deck or the surface of the deck is

tapped by a hammer. Hollow sound indicates presence of delamination, while a clear ringing sound indicates absence of delamination [2].

Inability of these techniques to provide objective and quantitative information has led to the development of numerous sophisticated NDE technologies that can overcome the barriers associated with the highly subjective visual and sounding methods. Impact echo (IE), infrared thermography (IRT), and ground penetrating radar (GPR) are the most widely used NDE technologies for the detection of subsurface damage, such as delamination, due to their ability to provide quantitative data that are necessary for good decision-making. Some other NDE technologies such as half-cell potential, electrical resistivity, ultrasound, and radiography are also available and widely used for other applications such as the detection of corrosion and concrete degradation.

According to these circumstances, efforts in recent years have been paid towards not only developing new NDE technologies, but rather towards inventing and developing more reliable methods which can be operated safely, consistently, and with high classification rate (i.e., accuracy).

With the increasing number and applications of NDE technologies, concerns about the reliability of these techniques are also increased. Reliability of NDE technologies is a growing topic that challenges authors due to the high number of controlling factors that influence the reliability analysis. Examples of such factors can be related to the NDE technology itself such as sensitivity of sensors used, testing methodology, and limitations and strengths of the NDE or related to operation condition such as weather condition, ease of access, materials to be inspected, and testing time limitation. Human factors such as skills and attitude of inspectors, and availability of testing and data analyzing standards can play a significant role in obtaining reliable and consistent results. Although some of these factors have been investigated and overcome, the overall reliability of NDE technologies is still in need for further investigations.

The term of "reliability" has several definitions, but for many applications it

can be defined as the probability of performing, without failure, a specific task under predefined operating conditions for a specified period of time. In other words, there are several components that should all be satisfied without failure to achieve high reliable results. These components are:

- Reliability is a probability, therefore, it is a number between 0 and 1.
- Performing a specific task (function), the test should be intended to provide specific results of interest for the structural component being evaluated.
- Under a certain operating conditions, the “specific task” should be performed under predefined conditions. As discussed above, weather conditions, time restrictions, skills level of inspector, material being tested, etc., are examples of such conditions.
- For a given period of time, the device should achieve all the previous requirement for a certain period of time.

For the purpose of this research, an NDE technology, represented by the instrument and methodology being used, is said to be reliable if it can prove its ability to accurately and consistently detect, identify, and quantify the subsurface damage (i.e., delamination) with high level of certainty compared to the actual condition; actual condition is either to be known priori or to be determined later by some other physical, normally defective, hindsight tools.

Historically, the reliability of NDE technologies is assessed through the probability of detection (POD) analysis; a curve showing the probability of detecting a defect, normally cracks, as a function of defect size [3]. This analysis method, however, does not consider the probability of false alarm (POFA) accompanying the evaluation process, which might be high enough to inversely alter the reliability estimation of test procedure, inspector, and/or the device being used. POFA is a very important index in the NDE application where higher POFA could result in unnecessary repairs that interrupt traffic and consume the already limited resources available for bridge preservation. Consequently, these unnecessary repairs can

negatively influence the structural integrity of the component being maintained where indeed some defective instruments will be used. Similarly, removing and re-installing the “incorrectly-evaluated” damaged parts may reduce the overall load carrying capacity of the bridge structure.

To overcome this obvious limitation of POD analysis and to provide data in terms of both POD and POFA indices, a statistical analysis method called receiver operating characteristic (ROC) analysis was used in this research. As will be discussed in the next chapters, some other advantages of using ROC analysis such as assessing the reliability of the candidate NDE technologies in terms of a quantitative index considering multiple threshold settings (i.e., decision criterion/value that can be obtained from the NDE data, which can be used to classify the data points and consequently the inspected element as good/bad) can be achieved by using ROC analysis.

1.2 Destructive vs. Nondestructive Testing

The traditional, and still most widely used, test methods for concrete and masonry are destructive methods. These methods involve performing some destructive testing procedures such as coring, drilling, removing part of the structure, or loading the member until failure in order to allow for direct examination. Although these methods are highly reliable in bridge inspection, they are time consuming, labor intensive, expensive, and the damage they leave behind often become focal points for further deterioration [4]. For example, subsurface damage inside a concrete bridge deck is normally identified by taken cores from the concrete at certain locations and depths. Even after proper sealing, these locations will become focal points that negatively influence the structural integrity of the bridge.

In contrast, nondestructive testing (NDT) technologies are non-intrusive, and therefore, they do not harm or permanently alter the physical properties of the structural member being inspected (i.e., do not cause any surface or internal damage to the item being tested). These methods can also be used in both new

and existing structures for quality control and integrity assessment purposes. They can also be less expensive, less-labor intensive, and can be performed in much faster rate in some applications as compared to the traditional destructive testing. Nondestructive testing (NDT) can be necessary and essential in many applications in almost all industries. In the civil engineering field, the possible applications and advantages of using NDT methods can be summarized as follows:

- In situ quality control of structures.
- Confirming the workmanship involved in concrete production (i.e., batching, mixing, placing, compacting, or curing of concrete).
- Monitoring of strength development of concrete in either reinforced concrete, prestressed, or similar applications.
- Detecting and quantifying of the extent of deterioration phenomena (i.e., cracking, delamination, degradation, etc.) resulting from factors such as overloading, fatigue, chemical attack, or environmental effects.
- Determining the concrete uniformity (quality) in precast, cast in place, or any other similar applications.
- Determining the position, quantity, or condition of reinforcement as well as the consistency of concrete cover.
- Monitoring long term changes in concrete properties such as chloride content, corrosion rate, and resistivity.

Indeed, some of these applications of NDT can also be performed by means of destructive testing, but using NDT is highly recommended due to the non-intrusive natures of almost all NDT. Time and budget restrictions in some applications also promote the use of NDT over the traditional destructive methods. Additionally, NDT technologies, as well as the construction materials, continue to be developed, and many new functionalities are being added to the existing methods; therefore,

the list above is expected to include some other applications, which might be difficult to be performed using traditional destructive testing.

1.3 Nondestructive Inspection (NDI) vs. Nondestructive Testing (NDT) vs. Nondestructive Evaluations (NDE)

Nondestructive inspection (NDI), nondestructive testing (NDT), and nondestructive evaluation (NDE), are all expressions commonly used to describe methods of testing that satisfy the condition of not tentatively or perpetually altering the physical properties of the structural component being inspected. The only apparent difference is that “NDE” has a broader range of applications than other terms. NDE encompasses both the inspection itself and the subsequent statistical and engineering analyses of the inspection data and includes, for example, determining properties of materials and also classifying damage modes and levels of severity [3, 5, 6]. In this research, the term “NDE” is used to describe the nondestructive evaluation as it fits for the intent of the research.

1.4 Problem Statement/Motivation

Many efforts have been made during the past few decades to assess the reliability or examine the performance of NDE technologies used for the detection of subsurface damage in concrete bridge decks. However, in most of the previous work, the reliability of NDE technologies has been discussed anecdotally without implementing any complementary mathematical or statistical analysis methods to describe the assessment process. As a result, a large body of previous work failed to objectively describe the reliability and provide a quantitative index that can be used to compare several NDE technologies against each other to find out which can provide lower risk for decision-making.

The other shortcoming with most of the previous work is that the comparison between NDE results and the actual condition, also known as “ground-truth,” has

been performed using a very limited number of test points. Ground-truth data is normally in the form of cores extracted from the test member. Aside from the obvious disadvantages of coring, especially the damage it leaves behind, the number of cores used for comparison against NDE results is normally very limited and this can negatively influence the reliability assessment. An NDE technology would have been ranked as “reliable” or “accurate” if the cores are extracted, by chance, in locations where an NDE provided correct results regardless of the other possible locations of coring which might result in incorrect correlation against the NDE results. Examples of previous work will be provided throughout the research.

Historically, reliability with performance assessments has frequently been solely relegated to the probability of detection (POD) assessments [7]. Although this is certainly an important component, it is merely the first step in the reliability analysis because most NDE technologies are characterized by noisy data and less than-perfect detection characteristics. In other words, NDE data do not necessarily fall into either of the two obviously defined categories: true positive (TP), meaning the NDT data indicates a defect and there is a defect, or true negative (TN), meaning the NDT data indicates no defect and there is no defect. NDE data can also include two types of incorrect indications: failure to give a positive indication in the presence of a defect (false negative, FN) and giving a positive indication when there is no defect (a false alarm or false positive, FP).

The “three decades of NDI reliability assessments” report developed by Karta Technologies, Inc. in 2000, under the supervision of the Air Force NDI office stated that POD alone is not enough to describe the reliability of NDE technologies unless the probability of false alarm (POFA) is also considered in the analysis [8]. POFA may be induced by noise with several possible sources: limitations of the NDE itself, human error, nature of phenomenon to be measured, and environmental conditions. The report covered nearly 150 reports and manuscripts from over 100 authors. However, a review of research literature suggests that insufficient theoretical work on the reliability assessment in terms of POD and POFA has been

undertaken since then.

The discussed limitations of the previous work, combined with the shortcoming of POD analysis in explicitly reporting POFA of NDE technologies, were the motivations behind this research.

1.5 Goals and Objectives

The main goal of the current research is to provide a more effective means of assessing and comparing the reliability of different nondestructive technologies applied to civil engineering applications. The main objectives of the research can be summarized as follows:

- Reviewing the principles, strengths, limitations, and testing procedures of the most widely NDE technologies used for evaluating concrete structures.
- Reviewing probability of detection (POD) analysis, the most widely used reliability analysis method of NDE technologies
- Extend using the ROC analysis, which is not used as commonly as POD analysis, to assess the reliability of NDE technologies. Impact echo (IE), Infrared thermography (IRT), and ground penetrating radar (GPR) are the candidate NDE technologies that will be used as example technologies during this research.
- Estimation of the probability of false alarm (POFA) and probability of detection (POD) as functions of the decision threshold of each NDE technology.
- Investigate multiple decision threshold settings within a practical range of application to interpret results of the NDE technologies and determine the optimum setting that may maximize the accuracy of NDE data.
- The previous objectives combined can provide and suggest a more comprehensive analysis method to assess the reliability of the nondestructive evaluation (NDE) technologies.

- Compare results of ROC analyses to determine to what extent NDE technologies are reliable and to suggest the best technologies that can provide lowest risk for decision-making.

1.6 Outlines of the Dissertation

This dissertation is comprised of eight chapters. Chapter 1 briefly discusses the motivation behind the development of NDT/E technologies summarizes some of the advantages and applications of these technologies. The problems associated with reliability analysis methods and the motivations for conducting this research are discussed in this chapter. Additionally, the goals and objectives of the current research will be discussed in this chapter.

Chapter 2 discusses in details the mechanism of the concrete bridge deck delamination and the most widely used NDE technologies for the detection of delamination damage. Principles of the IE, IRT, and GPR testing methods will be explained with some necessary details. Some of the previous work for the detection of delamination are cited throughout this chapter.

A review of the reliability analyses and factors that control the reliability assessment will be provided in chapter 3. The focus will be on POD and ROC analysis methods. Reliability indices inferred from such analysis methods will be also discussed in this chapter.

Chapter 4 is devoted to describing, in detail, the experiment work performed to achieve the objectives of the research. Both laboratory work and field testing will be discussed in details.

Results of testing along with ROC analyses of impact echo (IE), infrared thermography (IRT), and ground penetrating radar (GPR) technologies are discussed in chapters 5 through 7. The estimation of the best threshold setting for each technology and the calculations of area under the curve will be also discussed in details in this chapter.

The last chapter is devoted to a summary of the overall conclusions and findings

from this research. Recommendations for future works stemming from the current research will be presented in chapter as well. The appendices will contain the MATLAB manuscripts that have been developed for the purpose of ROC analysis and some other relevant tables and figures.

Chapter 2

LITERATURE REVIEW

2.1 Mechanisms of Reinforced Concrete Bridge Deck Delamination

Concrete is heterogeneous in nature and contains many paste-aggregate interlocks, therefore, it is very vulnerable to different kind of deterioration phenomena. There are many factors including severe weather conditions, freeze-thaw cycles, and exposure to salty water and some chemicals over certain threshold values for an extended period of time that can influence concrete durability and serviceability. Cracking, honeycombing, erosion, spalling, corrosion of reinforcement, and delamination are common problems in concrete resulting from the discussed deterioration factors. These deterioration mechanisms are complex and often prompt one another. Delamination, for example, can be formed as a result of corrosion of steel reinforcement, improper production, or freeze-thaw cycles [9]. Therefore, presence of delamination is a symptom of concrete deterioration.

Delamination in concrete bridge decks are typically characterized by horizontal splitting, cracking, or separation of a slab in a plane roughly parallel to the surface of the concrete. Delamination causes a thin layer (usually about two inches) to flake or blister producing a safety hazards for the motorists. The next section will be devoted to the discussion of the mechanisms of reinforced concrete bridge deck delamination as one of the most critical and frequent subsurface deterioration that affects most concrete bridge decks. [10].

2.1.1 Corrosion of reinforcing steel

Steel reinforcement embedded in the concrete is protected from corrosion by a passivation layer (an oxide film) provided by the high alkalinity environment of

concrete (pH level around 12.5). However, with the assistance of some external factors, thermodynamically unstable materials, such as steel, have a natural tendency to revert to the lowest energy state that is in equilibrium with the environment. This process is called corrosion [11]. Concrete is a porous material and contains many surface microcracks resulting from shrinkage and expansion cycles, overloading, freezing and thawing, improper concrete production, etc. These surface cracks, or pores, can represent focal pathways for two types of corrosion: Chloride-induced corrosion and Carbonation-induced corrosion.

Carbonation-induced corrosion happens when atmospheric carbon dioxide penetrates into the concrete. In the presence of moisture and oxygen, when the carbon dioxide reacts with the calcium hydroxide, which is one of the main components of cement paste, calcium carbonate will be formed and the high alkalinity environments will be diluted (pH level is reduced to around 9 or less) [12]. This process is known as carbonation or depassivation. Carbonation normally starts off at the surface and slowly penetrates into the concrete through the air filled pores leaving it more vulnerable to corrosion. Typical reactions are as follows [12]:

The time required for carbonation, or depassivation, can be estimated with the knowledge of the concrete permeability, k , and concrete cover, d , using the equation 2.3 [13]. It has been reported in the literature that carbonation rate is very slow and does not represent a serious threat for concrete structures [14].

$$t = \left(\frac{d}{k}\right)^2 \quad (2.3)$$

On the other hand, Chloride-induced corrosion happens when Chloride ions, present in seawater or resulting from the application of deicing salt on bridge deck, ingress to the steel layer. Accumulation of the chloride ions beyond certain

concentration levels can reduce the alkalinity level and breakdown the passivation layer, and leaves steel reinforcement vulnerable to corrosion [11, 15]. In either case, once the protective oxide film is destroyed, an electrochemical process begins. This process involves movement of electrons and electrical charges between two distinct regions on the steel bar (or bars). These two regions are the anode and the cathode. The anodic reaction (or oxidation) starts off when electrons starts to migrate from the anode region to the cathode region. The reduction reaction, on the other hand, happens in the cathode region, when the migrated electrons combine with water and oxygen in the concrete to form a rust [11]. Figure 2.1 illustrates the mechanism of electron migration and the formation of rust in steel reinforcement. The reactions are typically as follows [12].

Figure 2.1: Corrosion mechanism of steel reinforcement embedded in concrete

When steel corrodes, it expands up to 4-10 times its original size, producing excessive stresses on the concrete surrounding the corroded steel [16]. Such expansion can eventually break the bond between the steel and the concrete and cause the concrete to delaminate. Delamination, if not detected and repaired as soon as it is initiated, can cause spalling of the concrete and produce a safety hazard to the bridge's users, Figure 2.2. Corrosion by itself is a very serious problem because it reduces both the area and ductility of the steel reinforcement, and therefore, reduces the load carrying capacity of the member. It was also found that corroded steel bars can transfer the failure mode from ductile to brittle [17].

Corrosion damage can be controlled and reduced using several techniques, but these are beyond the scope of this topic.

2.1.2 Improper concrete production

Delamination can also form prematurely during production of the concrete. When the top surface of the concrete is densified before the excess mix water and air in the concrete mixture have a chance to reach the surface, delamination can form. In other words, when fresh concrete is placed and compacted, the solids (cement and aggregate) settle and the lighter materials (excess mix water and entrapped air) flow upward to the surface. This process is called bleeding. When finishing operations start before bleeding is completed, air and/or water are trapped under the densified surface mortar. As concrete hardens, subsurface voids develop where the water or air is trapped. These voids create weakened zones right below the surface that can eventually detach during slab use [18].

2.1.3 Freezing-thawing cycles

Freezing-thawing cycles is another cause for delamination formation. When the temperature values drop below 0°C , water contained in the pore system of the concrete can freeze, causing an increase in volume by 9% and producing stresses which may result in cracks and formation of delamination [15]. However, this process is very slow and depends on the number of freeze-thaw cycles and the diameter of pores where larger pores are more likely to be influenced by the process.

Figure 2.2: Spalling of concrete due to unrepaired sub-surface delamination.

2.2 NDE Technologies of Subsurface Damage Detection

As discussed in the previous chapter, several NDE technologies are used for the purpose of the detection of delamination in concrete structures. NDE technologies range in sophistication from simple ones, where the quality of sound obtained by striking the surface of the material with a hammer indicates the quality of the material, to complicated techniques where different waves traveling through the material are analyzed mathematically. Principles, underlying physics, signal processing, applications, apparatus, and advantages and limitations of the most common NDE technologies of concrete structures are discussed in this section. These are impact echo (IE), infrared thermography (IRT), and ground penetrating radar (GPR).

2.2.1 Impact Echo (IE)

Impact echo (IE) is one of the most widely stress wave propagation methods [19, 20]. Stress waves occur when pressure or deformation is applied suddenly, such as by impact, to the surface of a solid. Elastic wave propagation in heterogeneous solids, such as concrete, is a complex process. Therefore, some key characteristics of wave propagation, such as wave types, speeds, and principles of reflection and refraction, will be discussed first in order to better understand the IE technology.

2.2.1.1 Types of stress waves

When a solid is suddenly stressed by striking, using a hammer for example, three different kinds of stress waves will be generated. These waves are analogous to sound waves traveling through air. The generated waves are: P-, S- and R-waves. “P-waves” (also known as compression, primary, pressure, or longitudinal waves) are those in which the particle motion induced by the disturbance is parallel to the pulse propagation direction. P-waves are associated with normal stress and can propagate in all types of media. “S-waves” (also known as shear, secondary, distortional, or

transverse waves) are those in which the particle motion is perpendicular to the direction of the propagation and associated with shear stresses. “R-waves,” on the other hand, propagate along the surface of the medium and the particle motion is retrograde elliptical [21, 22]. Figure 2.3 illustrates the three kinds of waves.

Figure 2.3: Stress waves in semi-infinite material.

2.2.1.2 Speed of stress waves

In an isotropic infinite elastic solid, the speed of each of these waves is a function of the elastic properties of the medium (here, in concrete). The elastic properties necessary for wave speed estimation are: Young’s modulus of elasticity, E , the mass density, ρ , and Poisson’s ratio, ν [4, 19, 20]. For normal strength concrete, which is normally used in bridge decks construction, Poisson’s ratio is typically in the range of 0.18-0.2, Young’s modulus of elasticity is about $40 \times 10^6 \text{ N/m}^2$ and the density equals 2400 kg/m^3 . Table 2.1 summarizes the equations of the stress waves, the estimations of the speeds of these waves for normal strength concrete, and the relative speeds with respect to the P-wave [20].

2.2.1.3 Reflection and refraction

When a stress wave hits an interface with dissimilar properties, a portion of the wave is reflected and the remainder is refracted. The amplitude of the reflected wave is a function of the angle of incident and is a maximum in normal incident (i.e., perpendicular to an interface). The reflection coefficient is a function of

Table 2.1: Speed Estimations of Stress Waves.

Wave type	Equation	Wave speed; ^a ft/s (m/s)	Relative speed ^b
P-Wave	$C_p = \sqrt{\frac{E(1-\nu)}{\rho(1+\nu)(1-2\nu)}}$	13925(4245)	1.0
S-Wave	$C_s = \sqrt{\frac{E}{2\rho(1+\nu)}}$	8645(2635)	0.62
R-Wave	$C_r = C_s * \frac{0.87+1.12\nu}{1+\nu}$	7880(2400)	0.56

^a *In normal strength concrete.*^b *Relative to the P-wave speed.*

the mismatch of the specific acoustic impedances at an interface. The specific acoustic impedance can be determined by equation 2.5. Table 2.2 lists the acoustic impedance values of some of the most widely used materials in civil engineering applications and the reflection coefficients at interface for a P-wave traveling through concrete [20].

$$Z = \rho C_p \quad \left[\frac{kg}{m^2 s} \right] \quad (2.5)$$

If layer 1 and layer 2, respectively, have acoustic impedances of Z_1 and Z_2 , then the reflection coefficient at the interface is determined by equation 2.6 below [23].

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1} \quad (2.6)$$

Therefore, P-wave, which propagates normal to the surface, is expected to have a total reflection when incident on a concrete-air interface (i.e., delamination) due to the drastic mismatch in the acoustic impedances of concrete and air. This is the key factor for the successful application of stress wave methods in detecting subsurface defects within solids.

When Z_2 is less than Z_1 such as in a concrete-air interface, the reflection coefficient would be negative. This means that the incident compressive waves will be reflected as a tensile wave. The opposite is true when Z_2 is larger than Z_1 (i.e., concrete-steel interface). Distinguishing this phenomena is important in processing the signals (echoes) of the stress waves methods.

Table 2.2: Specific Acoustic Impedance of Some Materials and the Reflection Coefficients of P-wave in Concrete [20].

Material	Specific acoustic impedance ($kg/m^2.s$)	Reflection Coefficient, R
Air	0.4	-1.0
Water	0.5×10^6	-0.65 to -0.75
Soil	0.3 to 4×10^6	-0.3 to -0.9
Concrete	7 to 10×10^6	—
Steel	47×10^6	0.65 to 0.75

2.2.1.4 Principles

IE method was developed at the National Institute of Standards and Technology (National Bureau of standards, NBS), beginning in the mid 1980's for the detection of voids or delamination in concrete structures [23,24]. When properly used, impact echo is capable of determining the location and the extent of subsurface flaws such as crack, delamination, voids, debonding, or honeycombing in concrete and masonry structures. The principles of IE testing are based on the seismic analysis of material. As described previously, when a sudden disturbance is applied to the surface of a test object by means of a mechanical impact (i.e., hammer, solenoid impactor, steel ball etc.), stress waves are generated. Waves of a primary use for IE testing are the P-waves. When P-waves and S-waves propagate into the test object, they undergo multiple reflections between the surface of the test object and any interface of dissimilar acoustic impedance, such as internal flaws and/or the external boundaries of the test object. Arrival of P-wave to the surface where the impact is introduced produces displacements. These displacements are measured and recorded with respect to their arrival times by using a receiving transducer(s) and a digital oscilloscope. When the speed of P-wave (C_p) in the material is known, the round-trip travel time (Δt) of P-wave can be used to determine the depth of the reflecting interface (D or T) using equation 2.7. Figure 2.4 illustrates the

principles of IE testing.

$$D(or T) = \frac{\Delta t \cdot C_p}{2} \quad (2.7)$$

Figure 2.4: Principles of IE testing.

According to Carino, 2001, the first successful applications of IE occurred in geotechnical engineering to evaluate the integrity of concrete piles and caissons. In such applications, the long length of these foundations allow a sufficient time between generation, and receiving the reflected signals and the technique was called a sonic echo [23]. Therefore, using the time domain to process the recorded signal is quite feasible in such applications. In contrast, in thin concrete members, such as bridge decks, slabs, and walls, processing the signals of IE method in the time domain is more complicated than in the deep foundation applications. An alternative approach is the frequency analysis of the displacement waveforms [4]. Successive reflections of the P-waves between the surface and the interface causes the surface to displace each time the signal arrives to the surface. Therefore, the waveform is periodic and the period is equal to the travel path, $2T$, divided by the P-wave speed, C_p . The frequency of each characteristic displacement then is equal to

$$f = \frac{C_p}{2T} \quad (2.8)$$

If P-wave speed of a waveform is priori known, by testing a known thickness point of the test object, the depth (D), or the thickness of the test object (T), can be determined by the knowledge of the frequency content (f).

$$D(orT) = \frac{\beta C_p}{2f} \quad (2.9)$$

β is an empirical correction factor (shape factor) used to obtain thickness mode frequency. Sansalone 1997, found that for Poisson's ratio of 0.18, the P-wave speed in cores taken from the concrete is 4% less than the P-wave speed in an infinite plate, and therefore, β can be assumed as 0.96 [4]. The same factor was considered again by Gibson et al. 2005, based on Guided wave theory [25]. Based on Gibson's interpretation, exact theoretical values of β for plate-like structures are estimated in the range from 0.945 to 0.957 over the normal range of concrete material properties. The shape factor was implemented in ASTM C1383 as 0.96 [26].

This process of transforming the time domain to the frequency domain is usually done using the principles of Fast Fourier transform (FFT) where waveforms can be represented as a sum of sine curves [27]. Each of these sine curves with a particular amplitude, frequency, and phase shift. A distribution of amplitudes and frequencies of the sine waves, when added together, allows a description of the waveform being transformed, and the important frequencies in the time-domain function appear as a distinct peaks in the frequency-domain output. From principles of wave propagation, the low frequency waves penetrate deeper in the test object, but provide less resolution compared to high frequency waves. Frequency distribution is determined by the force-time function of the impact where smaller impactors can generate broader range of useful frequencies for IE testing [4, 23]. Therefore, exciting sufficient range of frequencies suitable for the application (i.e., the expected thickness of the member or the depth of delamination) is crucial. This is often accomplished by using a steel ball as an impactor (or any other impactor) with sufficient diameter to generate the required frequencies. For example, when detecting a shallow delamination (less than 4 inches deep) is of a primary interest,

smaller impactor may be used since this can generate high frequency P-wave, which is preferred to have a higher resolution output. In general, impact sources with about 20-80 μs impact duration, such as spring-loaded spherically-tipped impactors, have been used for detecting flaws within structural members less than one meter thick [19].

2.2.1.5 Instrumentation

Impact echo (IE) test system consists of three main components:

- an impact source,
- a high-fidelity receiving transducer, and
- data acquisition system.

Earlier research by Sansalone 1986 used a set of steel balls (less than 10mm diameter) dropped onto the surface of test object to generate the required stress waves [24]. Subsequent research used either steel balls attached to a high-strength spring to shorten the contact time, or solenoid impactors [10, 20, 26]. As discussed earlier, using a steel ball with sufficient diameter to generate stress waves of a range of frequencies optimum for the application is crucial for successful testing.

The receiving transducer is necessary to measure the surface response. This is normally a piezoelectric transducer and can be a ground-coupled or a pre-polarized air-coupled sensor. Air-coupled transducers allow for less time consuming testing. However, it has been reported in the literature that such transducers are less convenient especially in thick concrete members compared to the conventional ground-coupled sensors [28, 29]. Accelerometers are also an option but interpretation of data provided by such an option is more difficult than data provided by transducers [20].

Data acquisition systems are based on portable computers used to store, analyze, and process the captured data through an accompanying software. The most recent data acquisition systems may be integrated to work with more than one NDE

technology such as the one shown in Figure 2.5, which can be used with most stress wave methods [30].

Figure 2.5: Portable data acquisition system for multi-nondestructive testing [30].

Now that the principles of IE testing have been reviewed, it is important to provide an example showing how the technology can be used to detect subsurface delamination in concrete structures. Figure 2.6 (A & B) illustrates the time and frequency responses obtained from testing a “sound,” 7 in. thick, concrete slab with C_p equals 12000 ft/s. The peak at 10 kHz (0.01 GHz) corresponds to the thickness frequency of the solid slab which can be calculated by equation 2.9. Figure 2.6 (C & D) illustrates the responses of a test over a “deep” void within the same slab. For impact echo purposes, deep voids are voids deeper than 4 in. from the surface of the member [24]. The larger frequency peak at 14000 kHz (0.014 GHz) represents the P-wave frequency of the interface (the defect).

2.2.1.6 Advantages and limitations

One of the main advantages of using IE is that it is not affected by weather conditions, which has been recognized as a main limitation of other NDE technologies such as IRT and GPR. The other advantage is that IE testing can be used to detect

(A) Time domain of sound concrete

(B) Frequency domain of sound concrete

(C) Time domain of flawed concrete

(D) Frequency domain of flawed concrete

Figure 2.6: Time and frequency domains of sound (A & B) and flawed (C & D) concrete determined using IE testing.

a variety of subsurface damage such as delamination, honeycombing, degradation or surface cracks. Also, it can be used to estimate the strength of the concrete and estimate the thickness of the concrete which may be needed for quality assurance applications.

However, IE technology has two major limitations. First, it requires a high level of skill in order to analyze and interpret the data which is in form of either time or frequency domain. Especially when applied on testing a concrete element, analyzing IE data in either of the available domains is often accompanied by high variability among different inspectors or even the same inspector. Due to the heterogeneity of concrete materials, variation of concrete quality from point to point is an issue, as well as production problems, non-uniform thickness, or depth

of delamination; these factors all influence IE results. Secondly, IE testing is a “point-wise” test, meaning that the test is to be done on the basis of point-by-point. A grid is normally required to perform the test and this must be small enough to assure the some deteriorated areas are not missed. As discussed previously, missing such flaws can promote other deterioration mechanisms or even make the missed ones worse. Both of these limitations make IE a very time consuming technology.

2.2.2 Infrared Thermography (IRT)

Infrared thermography (IRT) is a noncontact, remote sensing NDE technology that has been used in civil engineering applications for over 30 years. IRT is based on the measurement of the radiant temperature of a material and converting that temperature measurement into a visual image [31]. It has been reported in the literature that IRT technology is economical and more convenient in testing large areas such as bridge decks, garages, dams, airport taxiways and around sewer lines [21]. IRT has many other applications such as detecting missing insulation, air infiltration or moisture damage. The next sections will be devoted to the review of the principles, instruments, and advantages and limitations of the technology for the condition assessments of bridge structures.

2.2.2.1 Principles

Thermal radiant energy is emitted from all objects that have a temperature greater than absolute zero ($-273.15\text{ }^{\circ}\text{C} = 0\text{ Kelvin}$) as an electromagnetic energy. Radiant temperature of a material is defined by the amount of electromagnetic energy exiting the object in two electromagnetic spectrum windows: $3\text{-}5\text{ }\mu\text{m}$ and $8\text{-}14\text{ }\mu\text{m}$ [31].

There are two basic concepts behind a successful implementation of thermographic inspection in condition assessment applications. First, the wavelength of the electromagnetic waves and the rate of energy emitted per unit surface of a real body depends on the temperature and the physical properties of a body

such as emissivity and conductivity [19]. Higher temperatures produce shorter wavelengths, which might be within the visible spectrum if the temperature is very high. Emissivity is defined as the ability of a surface to emit energy relative to the perfect emitter, the blackbody, in the same condition of temperature, direction, and spectral band of interest. The blackbody absorbs all incident radiation regardless of wavelength and incident direction. Conductivity, on the other hand, is the ability of a material to conduct heat when it is heated up. The speed of conducting heat depends on the material's coefficient of thermal conductivity. The rate of the emitted energy is given by Stefan-Boltzmann law, equation 2.10 [14].

$$Q_{rad} = \varepsilon \sigma T^4 \quad (2.10)$$

where Q_{rad} is the total energy emitted per unit surface of the area W/m^2 , ε is the emissivity of the material, σ is Stefan-Boltzmann constant ($5.67 \times 10^{-8} W/m^2.K^4$) and T is the absolute temperature (K)

As shown in equation 2.10, the radiant energy emitted from a material is proportional to the fourth power of its temperature. This factor is very important for the IR inspection, such that small variation in temperature will result in large changes in the radiant energy. However, such variation in the temperature of the surface may not always be attributed to the presence of subsurface anomalies. For example, presence of dirt, debris, paint, oil residues, localized moisture, high speed wind, or surface with irregular textures may affect surface temperature measurements [19].

The other concept is that flow of heat (conduction) through the material being inspected is disrupted by the presence of a subsurface delamination and voids. Assuming that the anomalies are not too deep, these changes in heat flow cause localized differences in surface temperature and can be easily detected using a thermal camera [32, 33].

Figure 2.7 illustrates two scenarios of heat transfer mechanism in presence of subsurface delamination. In the first scenario, Figure 2.7 (A), when the surface

of the slab is heated during the day time, as the ambient temperature increases, concrete absorbs temperature and radiates thermal energy. Areas above delamination resist heat flow and warm up at a faster rate than the surrounding sound concrete, and therefore, appear as hot spots. In contrast, during the night, areas above delamination cool down in faster rate, and thus, appear as cold spots in the thermal image, Figure 2.7 (B). In both cases, there will be a thermal contrast (ΔT) between delaminated (T_d) and sound (T_s) concrete, which can be used as an indication of presence of subsurface damage [14, 32, 33].

$$\Delta t = T_d - T_s \quad (2.11)$$

where T_d and T_s are the temperatures of delaminated and sound concrete respectively.

Figure 2.7: Mechanism of infrared radiation in presence of subsurface delamination during daytime (A) and night time (B)

IRT experiments involve two different options of heat sources necessary to create the required heat-flow condition in the test object [34]. These are known as “passive” and “active” thermography. Heat in passive thermography can be applied naturally using solar heating provided by the sun in the daytime or ambient cooling during night times [14]. Alternatively, in active thermography experiments, heat can be applied artificially by using an external thermal stimulus, such as halogen lamps, ultrasound exciters, xenon flash lamps, or electrical heaters [34–36]. Figure 2.8 illustrates some heat sources normally used in active thermography

experiments [37].

Practically, in bridge deck condition assessment applications, passive thermography is preferable since it is more economical and can provide an even heat condition for such large areas [32], while active thermography is most widely used in laboratory testing [36].

Figure 2.8: Heat sources commonly used in active thermography experiments: (A) heat gun, (B) thermal blanket, (C) halogen lamp, and (D) flash lamp [37].

2.2.2.2 Instrumentation

An infrared camera is used to detect and image variations in the radiated energy from an object. In general, infrared cameras consist of three main components: scanner, data acquisition system, and image recorder. These components are normally integrated into one device [19]. The scanner is an optical lens that transmits only an infrared radiation with wavelength in the range of (3-14 μm). Data obtained by the scanner unit are stored, digitized, and displayed as shades of gray or color. An image recorder is used to provide a record of the region scanned by the infrared camera where different colors in the image refer to areas with different thermodynamic properties.

2.2.2.3 Advantages and limitations

Advantages of using infrared thermography (IRT) technology in condition assessment of bridge infrastructures can be summarized as follows:

- It has a non-contact, non-invasive nature which allows testing without altering the physical properties of the material being inspected.

- It has the capability to perform remote inspections on large surfaces such as concrete bridge decks, and therefore, bridge closure is avoided.
- IRT is less time consuming than other NDE technologies, especially when a vehicle-mounted system is used.
- Though subjective, analyzing the IRT images is much easier as it can provide for in-situ judgments regarding whether or not subsurface damage is present.

The limitations, on the other hand, is that IRT technology is dependent on weather condition during and prior to testing. Wind speed, presence of moisture (or ice in freezing weather conditions) on or below concrete surface, humidity, cloud cover, and precipitation can affect the amount of infrared radiation, and therefore, negatively influence IRT results [14, 32]. The other limitation is that IRT can detect only shallow (up to 3-4 inches) subsurface anomalies from the surface being inspected. Deeper anomalies cannot be detected easily [32]. If anomalies such as delamination exist deeper than 5 in. from the surface of a concrete bridge deck, they may only be detected by thermal-imaging the deck's soffit which is impractical in many situations. Further, the technology has no ability to estimate the depth or thickness of subsurface anomalies.

2.2.3 Ground Penetrating Radar (GPR)

Ground, or probing, penetrating radar is a technique that has an enormously wide range of applications, ranging from planetary exploration to the detection of buried mines [38]. The early uses of the technology were in military applications [19]. Later on, it is used in a variety of fields and applications. According to V. M. Malhotra, the earliest attempt to using radar technology for the purpose of airfield inspection and void detections was done by Bertram et al. back in 1974 [21]. More recently, GPR has been used for monitoring strength developments of concrete [39], to measure the thickness of concrete [40], and determining the chloride and moisture concentrations in concrete [41, 42], as well as the locations

and depth of steel reinforcement [43]. The technology is also successfully used in other applications such as pavement profiling, mine detection, archaeological investigations, geophysical investigations, and borehole inspection [10]. Theory of EM wave propagation is complex, and no attempt was made here to cover the entire topic. However, to facilitate the understanding of the technology especially for the detection of subsurface damage applications, the next sections will be devoted to briefly discussing the basic principles, instrumentation, and the advantages and limitations of using GPR technology. The interested reader can refer to the following references for much more details [38, 44].

2.2.3.1 Principles

GPR uses electromagnetic (EM) wave propagation to inspect materials. It is analogous to impact echo (IE) technology discussed above, except it uses EM waves instead of stress waves. An antenna, with transmitter and receiver of a specific central frequency sufficient for the application, transmits short pulses of high-frequency EM waves into a material. As illustrated schematically in Figure 2.9, portion of these waves will reflect back to the surface when incidents on an interface with dissimilar dielectric properties. The remainder will continue propagating until it encounters another interface where again, a portion is reflected and the other portion continues to propagate further beyond the interface and so on until eventually the entire wave energy is dissipated [44]. The reflected waves are captured, recorded according to their arrival times, and stored for further processing.

The intensity of the reflected energy at an interface is a function of the mismatch in the dielectric properties of the two materials making an interface which can be described by the reflection coefficient, R , as follows:

$$R = \left(\frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \right) \quad (2.12)$$

Where η_2 and η_1 are the impedances (in Ohms) of materials 2 and 1, respectively.

The intrinsic impedance (the ratio of the electric to the magnetic field) of a nonmagnetic material is

$$\eta = \sqrt{\frac{\mu_0}{\varepsilon}} \quad (2.13)$$

Figure 2.9: Principles of GPR testing.

μ_0 is the magnetic permeability of free space (air), which equals 1.26×10^{-6} henry/meter and ε is the dielectric permittivity of a material in farad/meter. According to equation 2.13, the wave impedance of air, η_0 , can be defined as follows;

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \quad (2.14)$$

where ε_0 is the dielectric constant of air, which is 8.86×10^{-12} farad/meter. If we define the relative dielectric constant, ε_r as,

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \quad (2.15)$$

then equations 2.12 and 2.13 can be rewritten as follows

$$\eta = \frac{\eta_0}{\sqrt{\varepsilon_r}} \quad (2.16)$$

$$R = \frac{\sqrt{\epsilon_{r2}} - \sqrt{\epsilon_{r1}}}{\sqrt{\epsilon_{r2}} + \sqrt{\epsilon_{r1}}} \quad (2.17)$$

Equation 2.17 shows that when material 2 has a dielectric constant larger than the dielectric constant of material 1 (i.e., air-concrete or concrete-steel interfaces), polarity of the reflected waves will be positive. In contrast, when material 2 has a dielectric constant smaller than material 1 such as in concrete-air interface (i.e., delamination), the polarity will be negative as illustrated in Figure 2.10.

Figure 2.10: Polarity of the reflected EM waves at different interfaces.

The speed of the EM waves, V , is a function of the relative dielectric constant of a material. For lossy dielectric material such as concrete, V , can be determined as follows:

$$V = \frac{C}{\sqrt{\epsilon_r}} \quad (2.18)$$

where C is speed of light in free space, which is 1 ft/ns (0.3 m/ns) [21]. With the prior knowledge of the speed of the EM wave in a material and the relative dielectric constant, analyzing the arrival time, t , of the reflected waves permits

determinations of depth, d , of the reflecting interface as follows

$$d = \frac{V.t}{2} \quad (2.19)$$

Dielectric properties govern the ability of EM wave to penetrate a material, speed of the wave, and the intensity of reflected wave compared to the incident wave [10]. Dielectric constants of some typical construction materials used in civil infrastructures are given in Table 2.3.

Table 2.3: Relative Dielectric Constant of Some Construction Materials. [38]

Material	Relative dielectric constant
Air	1
Asphalt, dry	2-4
Asphalt, wet	6-12
Concrete, dry	4-10
Concrete, wet	10-20
Sand, dry	2-6
Sand, wet	10-30
Water	81

As discussed previously, presence of cracks on the surface of concrete can allow moisture and chemicals such as Chloride ions and carbon dioxide to penetrate the concrete. This intrusion can, to some extent, affect the chemical composition of concrete, and thus, can affect the dielectric constant at these areas causing steel to corrode and concrete to delaminate. Since the speed of EM wave is dependent on the dielectric constant, analysis of the time-domain of the reflected EM waves can reveal possible areas of deterioration. Some other relevant information can be obtained by analyzing amplitude, shape, and polarity of the reflected waves [19].

2.2.3.2 Instrumentation

GPR systems normally consist of four main components: antenna, data acquisition device (control unit), a display device, and surveying wheel [19,45,46]. The antenna

transmits short pulses high-frequency EM waves in high repetition. It can be either ground- or air-coupled. Both configurations have extensively been cited in the literature of assessment of concrete bridge decks [47–50]. It is well known that low-frequency (long wavelength) waves can penetrate deeper in a material, and therefore, is sufficient for deep investigations. However, the resolution of low-frequency waves is less than the resolution of high-frequency waves which in turn has limited penetrating ability. The central frequency of the antenna should be carefully chosen to fit the application. In bridge deck condition assessment investigations, high-frequency antennae (1000 KHz or higher) are most commonly used [45]. Table 2.4 shows depth of penetration and possible applications of some antenna choices. The antenna unit also contains a receiver which receives the reflected waves.

Table 2.4: Depth of Penetration and Possible Applications of Some Antenna Units.

Central frequency (MHz)	Depth of penetration ft.(m)	Possible applications
100	60(20)	Geology, environmental
200	25(8)	Geology, environmental
400	9(3)	Shallow, geology, utility, and archeology
900	3(1)	Concrete, shallow soil, and Archeology
1600	1.5(0.5)	Structural, concrete, roadways, and bridge decks

The control unit is the main unit that controls the transmission of the pulses, provides the power to emit the pulses, and acquires and amplifies the received signals [19, 46]. The display device can be an oscilloscope, an oscilloscope recorder, or a facsimile gray-scale graphic recorder which can plot a succession of recorded waveforms [21]. The surveying wheel measures the traveled distance with accuracy of ± 66.5 in./mile (100 mm/km) and a resolution of 1 in. (25 mm) [45]. Figure 2.11 shows a GPR system that has been extensively cited in the literature.

To provide for maximum resolution, it is recommended that GPR testing is to be performed transversely to the uppermost steel bars and by means of one pass (scan) every 2 ft [45]. Also, for a successful application of GPR testing, surface of the deck must be cleaned to remove any debris prior to testing. Testing should be performed when the bridge deck is dry since the EM waves are very sensitive to moisture as they affect the speed of the waves, and therefore, mask the determinations of the travel time and a subsurface feature depth.

Figure 2.11: Ground-coupled GPR system manufactured by Geophysical Survey System, Inc. (GSSI).

2.2.3.3 Advantages and limitations

GPR testing is very fast, especially when an air-launched antenna is used in a vehicle-mounted configuration such as the one shown in Figure 2.12. Ground-coupled antenna also provides for fast testing, but at a lower rate and bridge closure for a short period is often necessary. From a practical point of view, a bridge of about 200 ft long and 40 ft wide requires approximately two hours of testing assuming that the testing is to be performed in the traffic directions where fewer passes are needed to cover the whole area of the deck. Testing in the transverse direction may requires an additional hour or so. Some of the commercial systems allow for a “zig-zag” testing which can save some time through testing comparing to the regular testing configuration, Figure 2.13.

The other advantage of GPR is that the reflection coefficient of EM waves from a delamination (air-filled crack) is about 50% according to equation 2.17, assuming a dry concrete of dielectric constant equals 9. The other 50% of the waves will penetrate further beyond the interface. Therefore, features below the reflecting interface can be seen with GPR testing.

Figure 2.12: GSSI air-lunched, vehicle-mounted GPR system.

Figure 2.13: "Zig-Zag" GPR testing.

However, GPR has some limitations. Among these, it requires a well-experienced operator to operate and analyze the data. Another limitation is that when the GPR data is of low quality, due to nonideal testing parameters or instrumentation limitations, it must undergo some complicated filtering steps to enhance the quality for better interpretation. Further, GPR data is highly affected by the moisture

content of bridge deck, and therefore, testing during or just after a rainy day may result in misleading conclusions regarding the condition of the tested bridge. Consequently, the false (both positive and negative) indications increase in such circumstances. In addition, in a congested steel layout, the size and the distribution of steel bars can be very difficult to determine since the reflections from the steel bars overlap with each other. This last scenario can also impede the propagated EM wave from “passing” through the steel. Implementations of GPR in such applications is very difficult and can result in many false indications as well.

2.3 Earlier Works on the Detection of Delamination

The formal requirements for bridge inspections have been initiated after the collapse of the Silver Bridge, West Virginia, in 1967 [1]. This catastrophic collapse led to the development of several NDE technologies and the performance of these technologies has become a major topic during the past five decades and is continuing to grow.

Many efforts have been made to investigate the applicability of NDE technologies for the detection of subsurface damage in concrete bridge decks. Some of these efforts have considered a single NDE technology; others have considered multiple NDE technologies. The next sub-sections will shed the light on some of the previous works that are greatly relevant to the current topic. Key findings from these efforts will be briefly discussed.

2.3.1 Earlier work of IE technology

In the earlier applications of IE testing, Sansalone M. [51] studied the surface displacement responses produced by point impact on thick plates containing some “disk-shaped” flaws. Several variables were investigated such as, contact time of the impact, flaw size/depth ratio, distance between the impact and the response points, and the length of the time domain. The main conclusion of the study is that larger size/depth ratio can produce a clear response which is easier to detect and is the distance between impact and response points must be kept small to

obtain a maximum displacement.

Another study by Sansalone [52] investigated the reliability of IE in detecting delamination in concrete bridge decks with and without asphalt overlays. A modification on the regular impactor used with bare concrete was suggested to allow for testing soft materials such as asphalt. Delamination with unknown size, depth, and location were all detected using IE testing.

In 1995, Clemena GG [53] investigated the applicability of IE testing as an alternative to the conventional coring procedure for quality assurance of concrete thickness. Two grooved pavements and one un-grooved pavement were used in the study. Results of the study indicated that the wave velocity was inconsistent not only between pavements, but also within the same pavement, and therefore, considering the average velocity in IE testing can produce unacceptable error.

The feasibility of using transient stress waves to detect and quantify cracking caused by distributed damage mechanisms, such as alkali-silica reactivity and delayed ettringite formation, was studied by Kesner et al. [54]. A finite element simulation was used to determine how distributed damage can affect propagation and speed of the stress waves. A concrete slab with artificially distributed cracking, as well as some field studies, were used to verify results of the numerical simulation. The main conclusion is that IE can be used to estimate the amount of cracking damage in concrete members.

The ability of IE, as well as IRT and GPR, to detect different types of defects was investigated and compared against each other by Yehia et al. [55]. Three types of simulated flaws, i.e., surface cracks, delamination, and voids of known dimensions, location, thickness, and geometry were used in the experimental work. Results of each NDE technology were presented in a “hit-miss” format. Comparison results indicated that the three techniques are reliable to a certain degree with some detection variability of different flaws.

Gucunski et al. [56], developed a 3D visualization platform for IE data as an effort to make the interpretation of IE data easier and to save time in the data

interpretation phase. The developed framework also provided analytical software tools that characterize the IE data and provide the overall assessment of the condition of a bridge deck.

In 2012, Thomas et al. [29] evaluated the use of an air-coupled IE system for use in condition assessments as an alternative to the conventional ground-coupled systems. Displacement transducers normally used in ground-coupled IE systems were replaced by microphones held 1-2 cm from the concrete surface. Concrete slabs with some artificial defects of different depths and sizes were tested using the new system. The system was also tested in the field on the R.C. slab utility tunnel covers. It has been concluded that air-coupled IE is a good alternative to the displacement based IE with some limitations in detecting or interpreting returned signals from deep targets.

IE testing is a “point-wise” testing, meaning that the test is performed point-by-point. It has been reported in the literature, and confirmed later during the field and laboratory testing, that IE testing is very time consuming. In an effort to accelerate IE data acquisition, Mazzeo et al. [57] suggested the use of a non-contact acoustic excitation source that uses water droplets and some solid de-icers in combination with an air-coupled acoustic measurement in IE experiments. Experimental work involved using a pump to pump the water up and then down to create the necessary droplet with the required velocity and pressure. Both sound concrete and delaminated concrete were tested to demonstrate the applicability of the new technique with different conditions. The authors claimed that the data clearly shows the feasibility of the proposed technique to detect concrete delamination.

Li et al. [58] presented a method of using a mobile robot equipped with an IE device to automatically collect data and perform automatic classifications. The system has been tested on 100 test points on the surface of an in-service bridge deck. It was found that the proposed system is very effective and time saving in both data acquisition and data interpretation phases.

In the most recent strategic highway research program (SHRP2) report [10], which aimed to identify and describe the effectiveness of NDE technologies for the detection and characterization of deteriorations in concrete bridge decks, IE and several other NDE technologies were evaluated from the prospective of accuracy, precision, ease of use, speed, and cost. The experimental work involved testing an existing bridge and a fabricated concrete deck with some artificial flaws by several NDE technologies, such as impact echo, infrared thermography, ground penetrating radar, and half-cell potential. Participating technologies showed a variable ability in the detection of different deterioration mechanisms and also a variable performance measure, according to the previously mentioned criteria.

Similarly, in 2015, Khan [59] investigated the advantages and limitations of IE and IRT fusion in the detection of shallow delamination. Two concrete slabs with some embedded delamination targets were used in the experimental work. Analysis results indicated the advantages of fusion for a complete subsurface features detection.

2.3.2 Earlier work of IRT technology

One of the earliest work on the detection of delamination in concrete bridge decks using infrared thermography was done in 1969 by Arnold et al. [60]. The report summarized the results of preliminary experiments to attain the feasibility of using IRT for the detection of delamination. Several test blocks with different subsurface damage configurations and with different surface textures were used. Results of the experiments showed thermal emissivity differences and that delamination can be noticed especially during the ambient cooling periods.

In 1980, Manning [61] described the principles of IRT in detection of delamination in concrete bridge decks and investigated the effects of environmental variables, such as ambient temperature, solar radiation, and wind. It was found that delamination can produce a surface temperature contrast up to 3 Celsius degrees compared to sound concrete, and therefore, can be easily detected assuming

good environmental conditions.

K. Maser [62] discussed the physical principles of thermography and radar and proposed physical models for the predications of IRT and GPR responses to various bridge decks conditions. Effects of parameters such as Chloride ions concentrations, moisture content, cover depth of steel reinforcement, and thickness of delamination were investigated. Results showed that thickness of delamination affected the magnitude of the thermal contrast between delaminated and sound regions. The results also showed that the predictive model suggested was able to predict the area of deterioration to within 5% of the deck area.

Clark et al. [63] applied IRT technology to two case studies in Europe. One case study was a concrete bridge deck and the other one was a arch brick bridge. The goal of the study was to determine whether IRT technology can be applied to detect subsurface deteriorations in cold weather climates. Results indicated that the IRT was able to detect delamination in both case studies since the sensitivity of the thermal camera used was 0.08°C and the temperature difference between delaminated and sound regions was about $0.2\text{-}0.3^{\circ}\text{C}$.

Maierhofer et al. [64] used finite difference calculations to perform a systematic study on the influence of multiple factors including environmental conditions, heating power and time, depth and size of targets, and thermal properties of concrete. Results of the study indicated that the bigger and shallower the target is, the greater is the thermal contrast between delaminated and sound regions. Also, the results indicated that the higher thermal conductivity of the concrete, the better heat flow, and therefore, thermal contrast will be reached faster.

Cheng et al. [65] presented a pilot study of the combination of IRT and elastic wave propagations for the detection of delamination. Active thermography, by using a halogen lamps as a heat source, was used in the experimental work to detect small delamination targets in the test blocks. It was found that the temperature difference, or thermal contrast, depends on the depth and area of the defects targets.

The application of active thermography, using a pulsed thermography provided by halogen lamps, was also investigated by Dumoulin et al. [66]. The experiment involved using IRT technology for the detection of delamination targets made of wood and air in asphalt covered samples. In addition to the experimental work, numerical simulations based on the finite volume method of heat transfer inside the investigated sample were carried out. Results indicated that depth of defects and time for maximum contrast calculated by numerical simulation are of good agreement with the experimental results.

Recently, the effects of the environmental conditions such as ambient temperature, wind, humidity, and thermal loading were investigated by Washer [14, 32, 33]. During these efforts, a numerical model using 3D finite element simulations to predict thermal contrasts for subsurface delamination based on a given set of environmental conditions was developed. Also, guidelines for the use and applications of the IRT for the detection of delamination using a hand-held thermal camera was developed. In a separate study, Washer et al. [67] also worked on developing guidelines for use and applications of the technology for the detection of delamination in the soffit areas of the deck, which are not exposed to a direct heating from the sun, but from the ambient temperature.

2.3.3 Earlier work of GPR technology

One of the earliest applications of ground penetrating radar (GPR) for the assessment of integrity of concrete decks was done by Manning [68]. The research was devoted to develop methods for rapid and automatic collection of data using GPR and IRT for the concrete and asphalt-covered deck assessment. During this research, programs were written to automatically process data of both technologies and produce a plans of the bridge showing type and location of deterioration.

Advantages and limitations of GPR, as well as IRT and IE, were studied by Reis in 1998 [69]. The objective of the research was to evaluate the methodologies and the equipment of the aforementioned NDE technologies. Several case studies were

considered and the data were analyzed. Results of analysis indicated that GPR falsely indicated a presence of delamination at actually sound concrete locations. Also, GPR data missed many of the delamination that were detected by chain dragging, and therefore, it was suggested to use GPR in combination with other NDE such as IRT or IE.

In 2001, Steve et al. [70] performed a GPR survey over 11 bridges in the state of Missouri using an air-launched 1.0 GHz antenna and a ground-coupled 1.5 GHz antenna. The objective of the survey was to find out which antenna type can provide most accurate data compared to the available “ground-truth” information. Data provided by both antennas declared that the high frequency ground-coupled antenna provided better estimations for the deteriorated areas than the air-launched antenna. A good correlation with actual condition have been reported by both types of antennae, though.

Scott et al. [71] compared the accuracy of two GPR systems and IE against the actual condition of some case studies as determined by chain dragging and some cores extracted from the tested bridges. In this research, a commercial ground-coupled GPR system of 1.5 GHz central frequency and a prototype system developed by Lawrence Livermore National Laboratory (LLNL) were used. The LLNL system is known as the High Speed Electromagnetic Roadway Mapping and Evaluation System (HERMES). This LLNS system contains an array of 64 air-coupled high frequency antennas configured in an 8 x 8 array that can surveying a deck surface one lane at a time at a scan width of about 6.25 ft (1.9 m) at a traffic speed of about 18.5 mile/h (30 km/h). The two GPR system were not able to provide a consistent response to delamination targets. As a result, the detection of delamination by GPR was unsuccessful as compared with IE and sounding.

Barnes [50] investigated the estimation of the deteriorated area of a bridge deck using GPR compared to chain dragging and half-cell estimation. Twenty-four asphalt covered bridge decks with a wide spectrum of deterioration levels were tested in this study. Results showed that bridges with less than 10% and more

than 50% deterioration have significantly different results than predicted by chain dragging and half-cell potential. According to the results of this study, the authors suggested using GPR in combination with visual inspection in order to improve repair quantities.

In 2008, Barnes et al. [72] proposed a normalization procedure to improve the accuracy of GPR for evaluating Chloride-based corrosion damage in concrete bridge decks by subtracting the effects of signal loss due to inconsistent depth of steel reinforcement. The proposed procedure was applied to six unpaved concrete decks and the results showed a good improvement of the GPR data compared to the uncorrected data. The authors claimed that the proposed procedure could be successfully applied to GPR data obtained from asphalt-covered bridge decks, assuming that the thickness of asphalt is uniform.

To overcome subjectivity in interpreting GPR data, and to save time and effort during data acquisition, Wang et al. [73] proposed a method for automatic detection of deck condition. The proposed method used partial differential equations to detect rebar signature so that delamination can be easily located. The method was tested on synthetic and real GPR images. The results proved the accuracy of the method for the detection of delamination. The proposed method can detect rebar in spite of the low signal-to-noise ratio (SNR) or low contrast situations.

Tarussov et al. [74] discussed limitations and drawbacks of the numerical analysis GPR data processing algorithms and claimed that the results provided by such analyses are ambiguous. The research proposed using visual interpretation in combination with computer assistance to correctly identify regions of deterioration in concrete bridge decks. Several case studies were tested and data were acquired for analysis. The proposed analysis method, visual with computer assistance, was able avoid misclassifications due to the structural features or surface anomalies and therefore, was more accurate than the automated numerical analysis methods proposed by others.

In 2015, Romero et al. [75] extended the previous work of Barnes [72] and

provided three different techniques for correction reflection amplitude values due to inconsistency in steel bars depths. One manual and two automated processes were presented to account for rebar depth variations. The three techniques were applied, using a ground-coupled 1.5 GHz antenna, to three concrete bridge decks, bare concrete surface, and asphalt-overlaid, and the results were presented as contour maps before and after depth-corrections. Results indicated that the three processes can provide almost-identical results in all case studies, and also, depth-correction contours were more accurate in mapping the deteriorated areas. The authors suggested that ASTM D6087-08 [45] must incorporate such correction for more accurate results.

In general, there has been some confusion regarding the capability of GPR for the detection of delamination. Some research indicated that GPR has the ability to only provide a likelihood of presence of delamination at locations of elevated chloride and moisture concentrations. Some other research indicated that GPR is an effective tool for detecting delamination damage directly and accurately. For example, Maser et al. stated that GPR is not capable of directly detecting delamination, but rather provides a measure of moisture and chloride in the concrete [62]. Barnes provided an estimated probability of correctly detecting delamination of 60% in asphalt-covered decks using a 1 GHz air-coupled GPR system [76]. In another study, Barnes et al. provided a very comprehensive study on the correlation of air-coupled GPR results with chain drag (sounding) results that illustrated the limitations of GPR with respect to detecting delamination in concrete [50]. Scott et al. demonstrated that a subsurface air gap simulating delamination could only be detected if the air gap was quite large [77]. Kim et al. indicated that GPR cannot be used to directly detect delamination, but could be used to detect and map areas where EM wave attenuation was increased and velocity decreased, and that areas with these characteristics were typically areas of delamination [78].

In contrast, some other research has suggested that GPR may be an effective

technique for the detection of subsurface delamination. A recent study undertaken in Wyoming indicated that GPR was an effective tool for detecting delamination [79]. The recently published SHRP 2 report on NDE for bridge decks ranked GPR as having approximately the same accuracy for delamination detection as Infrared Thermography (IRT) and chain dragging [10]. GPR was rated as the best overall value for delamination detection, apparently due to factors such as high speed and lower cost as compared with seven other NDE technologies. This rating contributed to a lack of clarity regarding the capabilities of GPR. The ASTM standard test method addressing GPR for the evaluation of bridge decks provides a full paragraph in the Scope section of the standard that describes delamination in concrete to infer that the GPR procedures described in the standard are capable of detecting such damage [80]. Clemena [81] claimed that GPR technology has the capability to detect delamination with an average accuracy of 82%. Yehia et al. [47] estimated that the accuracy of GPR for the detection of delamination is about 77% based on results of 13 cores.

Chapter 3

RELIABILITY ASSESSMENT METHOD/REVIEW

3.1 Introduction

With time, bridge infrastructures deteriorate and accumulate a variety of defects such as corrosion, debonding, delamination, and fatigue cracks. Assuring the integrity of these infrastructures requires periodic inspections with “reliable” technologies. Reliability can be defined and computed in different manners, but there appears to be nearly universal agreement that the reliability of NDE testing is a measure of performance and consistency to determine properties of material, geometrical characteristics, or to locate and size flaws.

Historically, the reliability of NDE technologies has been primarily described through probability of detection (POD) analysis. However, it has been recognized that POD alone is not truly a reliability measure, rather it is an element assessment since the NDE reliability requires a multiple parameter estimation [82]. This is true since probability of detection is often estimated solely as a function of defect’s size or depth neglecting all other factors that might influence the reliability estimation.

As will be discussed later, POD analysis has been primarily used to describe the reliability of NDE technologies for the detection of cracks in metal specimens where material is homogeneous over the entire test member. In contrast, concrete is a heterogeneous material and contains many paste-aggregate interlocks which may vastly alter the outputs of the inspection method used if the concrete quality of the inspected member is inconsistent. In other words, a high percentage of incorrect indication/decision is expected when testing concrete material compared to testing metal such as steel and aluminum. These incorrect indications influence the reliability and accuracy estimation of NDE systems, and therefore, should be considered in the analysis.

The incorrect indications have different negative consequences. A false negative rate (FNR), resulting from failure to detect existing defects, can increase the risk of using deteriorated bridges, and therefore, producing a safety hazard for motorists.

In contrast, the probability of false alarm (POFA), resulting from measured signals at sound locations is not explicitly considered in POD analysis. POFA is a very important index and should not be neglected in the reliability assessment. Repair and maintenance that result from false positive calls not only would be unnecessary and consume the “already limited” funds available for bridge preservation, but might negatively affect the structural integrity of the bridge. High risk of inventory mis-prioritization is another effect of false positive calls when several bridges are inspected using different NDE technologies with different POFA indices.

In order to assure the best budget allocations and minimize the risk of using deteriorated bridges, the incorrect decisions provided by an NDE inspection method must be kept to a minimum. The problem is that POFA and FNR are inversely related. Decreasing the POFA increases the FNR (misses), and this means using a deteriorated bridge under high possibility of risk. On the other hand, decreasing the FNR increases the POFA and this means poor use for the funds available to maintain deteriorated bridges.

Receiver operating characteristic (ROC) analysis, which is mainly used in medical, machine learning, and data mining fields, has been recognized as a good alternative to the conventional reliability analyses in these fields. ROC analysis can explicitly consider the probability of false alarm (POFA) in addition to the probability of detection (POD). In contrary to POD analysis, ROC analysis estimates both the POD and POFA considering a multiple threshold settings within practical range of application. ROC analysis has not been employed as commonly as POD analysis for the assessment of NDE reliability. Therefore, it will be interesting if this analysis tool is introduced to the NDE community, because the reliability of NDE technologies has long been a question of great interest in a

wide range of applications.

From the above brief introduction, it seems that both POD and ROC analyses complement each other rather than compete. The next sections will be devoted to discussing both POD and ROC analysis procedures, requirements, limitations, and advantages.

POD analysis has been widely and intensively cited in the literature for NDE analysis. Although a brief discussion will be provided in this chapter, the interested reader can refer to [7, 83, 84] for more details about this analysis method. In contrast, as was discussed earlier, ROC analysis has been very rarely used for the assessment of NDE reliability especially, for the detection of anomalies of three dimensional nature such as delamination. Therefore, ROC analysis will be used in this research to demonstrate the applicability and feasibility of using such statistical analysis in NDE reliability tasks.

3.2 Factors Affecting the Reliability of NDE

Before delving into discussing the approaches used to assess the reliability of NDE technologies, it is important to discuss the factors that might affect the reliability estimation. These factors can be related to the physical properties of the defect itself, such as size, microstructure, depth, or orientation, or related to the inspection system, such as calibration, parameters limitations, or sensitivity. These factors can also be related to the inspection condition, such as weather, or access and time limitations, or related to human factors such as experience, attitude, or attention to duty.

According to this, an NDE system could be divided into an appropriate number of elements (modules). The reliability of an NDE is estimated by combining the reliability of all elements. Mueller et al. [85] suggested the following formula to estimate the reliability of an NDE system as the sum of three individual elements

$$R = f(IC) - g(AP) - h(HF) \quad (3.1)$$

where, IC is the intrinsic capability of NDE system, AP represents application parameters affecting the reliability measure (i.e., access restrictions, weather, defect characteristics, etc.), and HF is devoted to describe the human factor effects (i.e., subjectivity, training, attention to duty, education etc.).

This, in fact, supports the conclusion of that the reliability measure involve multiple parameter and should not be solely estimated through POD curve considering only one variable (often crack size).

3.3 Probability of Detection (POD) Analysis

3.3.1 History and applications

Attempts to quantify NDE capability through POD analysis began in the 1970s with two main studies by the United States National Aeronautics and Space Administration (NASA) and the United States Air Force (USAF) [86]. The main purpose of these studies was to stenograph reliability data to a single measure that would be useful for the comparison purposes.

In the first study by NASA, a total of 118 Aluminum panels containing 328 artificially produced fatigue cracks were used. Cracks ranged in length from 0.018 to 1.27 cm with depths ranging from 0.003 to 0.451 cm. Some intact panels were included in the test set to satisfy the randomness of the test. All panels then were tested by three skilled inspectors. Testing methods were: fluorescent penetrant, eddy current, ultrasonic, and X-radiography [87]. In this study, panels were tested many times by same inspectors as the physical properties altered from “as machined panels,” to either notched panels (to remove the smeared material resulting from the machining process) or proof loading the panels to 85% of the yield strength and to the failure.

Data from multiple tests were used to draw POD curve as function of crack length, crack size, and crack cross-sectional area. Hit/miss analysis was used for all tests. For example, the successes (hits) for each crack length were represented as a

point on the graph, and then a smooth curve was drawn through the plot points using least squares fit procedure. Estimates of variance at the 95% level were then estimated and registered for each data point using a “Chi-Square” distribution analysis procedure [87].

In order to gain a 90% confidence, it was required that at least 29 cracks from the same size group would be tested and detected. This procedure is well known as “29/29,” which has been considered as a standard for the NDE reliability estimations [82, 86]. The reason for using a 90% confidence level instead of the widely accepted level of 95% is due to the low number of cracks available for each size. This is also the reason behind grouping the cracks according to their length as intervals (range interval method, RIM) [88]. The drawback of this analysis method is the variability of the sample size through intervals and this consequently leads to difficulties in estimating the confidence bounds of the POD values. As can be seen in Figure 3.1, which shows the POD estimates ($\hat{p}$) and the lower confidence limits (LCL) for each crack size range, the LCL can vary significantly based on the numbers of samples in each size range.

Figure 3.1: Range interval POD analysis with the 90% lower confidence of an NDE data [7, 88].

Some other methods have been suggested to overcome the problem of variable

size of samples and difficulty of estimating the confidence limits. Non-overlapping and overlapping constant sample size methods were used in some experiments. These two analysis methods allow for variable intervals, but constant sample size. The confidence intervals for the overlapping method were narrowed by allowing some overlaps to exist between intervals (groups) [88]. Figures 3.2 and 3.3 show the results of both methods of the eddy current test. It was realized later that each of the discussed methods has its own limitations and advantages. More detail can be found in [88].

It should be noted that the previous analysis methods demonstrate the capability of NDE methods through the estimation of the POD function and confidence bounds through single inspection of cracks covering a range of lengths.

Figure 3.2: Non-overlapping POD analysis with the 90% lower confidence of an NDE data [88].

Figure 3.3: Overlapping POD analysis with the 90% lower confidence of an NDE data [88].

A more comprehensive reliability analysis study was done by the United States Air Force (USAF). Sections of retired aircrafts and other specimens were inspected for cracks by representative personnel, using various NDE systems in a typical environment. This experiment is known as the “Have-Cracks-Will-Travel” Program, and the data from the program is often referred to as the “Have-Cracks” data [89]. In this program, estimation of the POD function and confidence bounds was accomplished by performing multiple inspections of cracks covering a range of lengths. In this analysis approach, a method based on Logistic regression to analyze hit/miss data was developed [86].

As discussed earlier, there are several other factors that affect the detections of cracks beside the crack size. Cracks with approximately the same size do not always yield the same detectability criterion. Therefore, this analysis method yields an estimate of the detectability for each particular crack size. Best fitting is often required through the means of probability of detections at all crack sizes as well as estimation of confidence bounds.

3.3.2 Statistical analysis

Inspection results are normally recorded in two different formats; the format determines the analysis method to be used in modeling the $POD(a)$ function. When the results of an inspection are expressed in a binary format, i.e., whether or not a crack was detected, the data are called a “hit/miss.” For the purpose of analysis, a binary response of (1) and (0) are often used, corresponding to hit and miss outputs respectively. Examples of such data would be the results of visual inspection, sounding methods, magnetic particles or fluorescent penetrant inspection, or any inspection for which the magnitude of the response to the inspection stimulant is not recorded.

In contrast to the hit/miss capability, some other inspection methods can provide more information regarding the size, depth, shape, severity, or any other property of the flaw being detected. When this is the case, inspection data is normally represented as “ a vs $\hat{a}$.” Examples of such inspections are ultrasonic testing, infrared thermography, impact echo, and spectral analysis of surface waves.

In general, the simplest way to think about $POD(a)$ is through the estimation of the number of successes (detections) out of the total number of cracks of the same size, a , or same interval category

$$POD(a) = \frac{\text{No. of success}}{\text{No. of cracks inspected}} \quad (3.2)$$

Once all of the cracks have been inspected, a continuous curve is fitted either manually or by using an appropriate statistical curve fitting to plot POD curve as a function of crack size. In the United States Air Force (USAF) study, it was realized that cracks of the same size (or same interval) were not detected equally because crack size is not the only factor influencing the detection capability, as discussed above. Figure 3.4 illustrates the main steps towards construction of POD curve from hit/miss and a vs $\hat{a}$ data. In the next subsections, statistical analyses and the required sample size will be discussed in details for both data formats.

Figure 3.4: Main steps of POD analysis for hit/miss and a vs. $\hat{a}$ analyses.

3.3.2.1 Hit/miss analysis

Berens and Hovey [84] proposed using a probabilistic description of the POD where at any particular size, a , the *POD* of a large number of cracks of size a is approximately normally distributed. Once the probability of detections were calculated for all sizes, seven different statistical functions were tested to determine which can provide for the best representation of “Have-Cracks” data and the best fitting criterion. The tested functions were Weibull, modified Weibull, probit, log probit, logistics, log logistics, and the arcsine models [84]. A linear regression analysis was used on a transformed probability of detection and crack size axes. It was found that the log logistic model (log odds) is the best among the tested models for “Have-Crack” data. However, there appears to be a universal agreement that there is no unique transformation (model) is recommended for all POD data. Statistical analysis of log logistic model requires some knowledge of the logistic regression. The next paragraph will be devoted for this purpose.

In contrast to the correlation and regression techniques of numerical data, non-numerical (categorical) data requires a special regression techniques. Logistic

regression can be used for this purpose where the categorical data is transformed to a binary numerical data (i.e., 0 and 1). A logistic regression equation is used to describe how the probability of “success” (for example, probability of flaw detection) changes as a numerical predictor variable, crack size or depth, changes. The general form of the logistic equation is

$$P = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (3.3)$$

In the nondestructive evaluation applications, P and x represent respectively, the probability of detection, POD , and the crack size, a . Although the above equation looks complicated, some tedious, but straightforward algebra can be used to simplify it as follows:

$$\begin{aligned} P &= \frac{e^{\beta_0 + \beta_1 a}}{1 + e^{\beta_0 + \beta_1 a}} \\ e^{\beta_0 + \beta_1 a} &= POD(a)(1 + e^{\beta_0 + \beta_1 a}) \\ e^{\beta_0 + \beta_1 a} &= POD(a) + POD(a) * (e^{\beta_0 + \beta_1 a}) \\ POD(a) &= e^{\beta_0 + \beta_1 a} - POD(a)(e^{\beta_0 + \beta_1 a}) \\ POD(a) &= e^{\beta_0 + \beta_1 a}(1 - POD(a)) \\ e^{\beta_0 + \beta_1 a} &= \frac{POD(a)}{1 - POD(a)} \\ \beta_0 + \beta_1 a &= \ln\left(\frac{POD(a)}{1 - POD(a)}\right) \end{aligned} \quad (3.4)$$

This means that if the logistic regression curve is reasonable to describe the relationship between POD and a , the relationship between $\ln \frac{POD}{1-POD}$ and a is linear. As described previously, Berens and Hovey [84] used this general logistic model and the log logistic model (log odds) with some modification, $\log(a)$ instead of a , on the logistic model as follows:

$$\ln\left(\frac{POD(a)}{1 - POD(a)}\right) = \beta_0 + \beta_1 \ln a \quad (3.5)$$

Given a pair of a_i and P_i , x_i and y_i can be defined as follows:

$$\begin{aligned} x_i &= \ln a_i, \text{ and} \\ y_i &= \ln \frac{\hat{P}_i}{1 - \hat{P}_i} \end{aligned} \quad (3.6)$$

Equation 3.5 can be written in another mathematical form as follows:

$$POD(a) = \frac{1}{1 + e^{\frac{-\pi}{\sqrt{3}}(\frac{\ln a - \mu}{\sigma})}} \quad (3.7)$$

where β_0 , β_1 , σ , and μ are unknown parameters that can be estimated through regression analysis [90]. In contrast to β_0 , β_1 which have no physical meaning, parameters σ , and μ do have a useful physical interpretation. Parameter $\mu = \ln a_{0.5}$, where $a_{0.5}$ is the median flaw size satisfying $POD(a_{0.5}) = 0.5$; σ is related to the steepness of the $POD(a)$ curve. Examining equations 3.5 and 3.7, one can find that the relationship between (β_0, β_1) and $(\sigma, \text{ and } \mu)$ is

$$\mu = -\frac{\beta_0}{\beta_1}, \text{ and } \sigma = \frac{\pi}{\sqrt{3}\beta_1} \quad (3.8)$$

Any measurement process is often accompanied by uncertainty due to some measurement errors (systematic and random). Therefore, probability of detection (POD) is always expressed within the confidence bounds (or interval). In the literature for NDE, POD is normally expressed with 95% confidence. Some crack sizes were of primary interest for both NASA and the USAF, and, it was required that the detectability of such cracks be estimated with at least 95% confidence. These are $a_{10/95}$, $a_{50/95}$, and $a_{90/95}$, see Figure 3.5. For example, $a_{10/95}$, represents the crack size at which the inspection technique has the ability to detect, with 95% confidence, at least 90% of existing cracks of the same size, or interval. The other defect sizes have the same meaning, but with different detectability criteria.

Figure 3.5: POD analysis of qualitative NDE data showing the detectability at cracks of interest [7].

3.3.2.2 a vs. $\hat{a}$ analysis

Some NDE technologies can provide quantitative outputs, often called $\hat{a}$ (pronounced a-hat), such as amplitude, temperature, or travel time. This kind of response requires a preceding analysis step that relates the actual flaw size, a , to the recorded signal, $\hat{a}$. Normally, the relationship between the actual size and the recorded signal is linear or linearizable by some axes transformations [7, 90, 91]. Again, even the measurement of $\hat{a}$ in terms of a accompanying some systematic or/and random errors, and therefore, the following linear regression equation is used to predict the defect size

$$\hat{a} = \beta_0 + \beta_1 a + \varepsilon \quad (3.9)$$

where ε is the error (residual), which is generally assumed normally distributed with zero mean and standard deviation σ_ε [92]. β_0 and β_1 are parameters of regression analysis (intercept and slope respectively). Besides the linear model,

other probabilistic models can also be applied to describe the relationship between the response, $\hat{a}$, and the actual size, a , depending on the actual experiment data of the inspection technique. Berens, A.P., [91] used the natural logarithms of $\hat{a}$ and a to describe the relationship between the actual and response sizes of defects

$$\ln \hat{a} = \beta_0 + \beta_1 \ln a + \varepsilon \quad (3.10)$$

Regardless of the model used, once the relationship between the measured signal and the actual size is established, the $POD(a)$ is estimated depending on a preset value of decision threshold, $\hat{a}_{dec}$, [7, 91]. Every signal above $\hat{a}_{dec}$ is classified as crack detected, otherwise the inspection result is considered noise, and therefore, the crack is classified as missed.

$$POD(a) = P(\hat{a} > \hat{a}_{dec}) \quad (3.11)$$

The number of hits for any particular flaw size divided by the total number of flaws of the same size defines the $POD(a)$. Figure 3.6 summarizes the steps of constructing a POD curve for any particular NDE technique. The following mathematical representation of $POD(a)$ is applied:

$$POD(a) = \phi\left(\frac{\ln a - \mu}{\sigma}\right) \quad (3.12)$$

where ϕ is the cumulative standard normal distribution function. μ and σ can be defined as follows:

$$\mu = \frac{\ln \hat{a}_{dec} - \beta_0}{\beta_1} \quad (3.13)$$

and

$$\sigma = \frac{\sigma_\varepsilon}{\beta_1} \quad (3.14)$$

Figure 3.6: POD analysis of quantitative NDE data showing the detectability at cracks of interest [7].

3.3.3 Sample size requirements

Precision of measurement is vastly driven by sample size. A larger sample size of any particular experiment can, in general, provide greater precision in the estimation of the desired index and vice versa. In addition, a larger sample size reduces the margin of error calculated by the following equation and this leads to narrower confidence bounds in the estimation

$$MOE = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (3.15)$$

However, it is not always possible to produce a high number of test specimens pertaining to the required precision. NDE measurement, for example, is more driven by the economics of specimen fabrication and flaw characterization than by the desired degree of precision [91]. On the other hand, misleading conclusions may be reached when using experimental specimens that contain very few flaws. As discussed earlier, some NDE techniques can provide a quantitative output, $\hat{a}$ in terms of the actual flaw size, a . This numeric output can provide additional useful information regarding the size, depth, shape, etc. of the flaw. If this is the case,

the required flawed locations for such cases is recommended to not be less than 30 [3,84,91].

In contrast, NDE techniques that provide a binary output, a “hit/miss,” cannot provide additional information, and therefore, the required sample size, flawed locations, is larger. It is recommended that the sample size not be less than 60 in these kinds of cases [3,84,91].

Because the very small flaws will be almost certainly missed and the very large flaws will be almost entirely detected, flaws within these intervals should only be added to randomize the test. Otherwise, using such flaws with almost the same percentage of flaws of the useful range can increase the cost of fabrications for no apparent reason.

In addition to this, some unflawed locations are also required to be added and tested to assess the capability of the NDE in discriminating between flawed and unflawed locations. As a rule of thumb recommendation, the required unflawed locations should be in the range of twice to three times the flawed locations [3,91].

All of the above estimations of the sample size and the unflawed locations are subjective and may only be applicable for the case of metallic specimens where introducing cracks may be relatively easier. The experimenter can include the desired number of flaws needed to satisfy the desired confidence; this is mainly driven by the application itself and the overall accuracy of the NDE technique. For example, testing a reinforced concrete culvert in a rural area would not be as important as testing a major reinforced concrete bridge in an urban area, even if they are both made of the same material.

3.4 Receiver Operating Characteristic (ROC) Analysis

In the POD analysis, the probability of false alarm, POFA, accompanying NDE testing method is not explicitly provided. In some circumstances, a signal is incorrectly recorded (noise) when the test point is actually unflawed (i.e., extreme left of the POD graphs above). This measured signal, or noise, in an unflawed

location is the probability of false alarm, POFA. As is the case with POD, POFA also depends on the preset decision threshold, $\hat{a}_{dec}$, of the signal, and therefore, there is a distinct POFA index for every possible decision threshold. Results of POD analysis are often presented as a single POD curve depending on a single preset decision threshold.

For NASA, the USAF, and other similar applications where the major focus is on the crack size at which there is, for example, 90% detection, one single POD curve with 0.05 preset POFA can satisfy the requirements. For some other applications, such as assessing the integrity of bridge infrastructures in civil engineering applications, POFA is very important index and needs to be considered in the analysis. High POFA in concrete bridge deck inspections may commence unnecessary repairs that interrupt traffic and are quite complicated, time consuming, and labor intensive. In addition, such unnecessary repairs might negatively influence the structural integrity of the bridge as compared to the repairs of airplanes or airspace shuttles, which are normally performed inside controlled laboratories.

Receiver operating characteristics (ROC) analysis is a plot of the POD and POFA considering a wide range of threshold settings. The x-axis in ROC graph represents the probability of detecting a signal above a threshold criterion despite the fact that the location is unflawed, POFA. The y-axis, on the other hand, represents the probability of detecting a signal above the same threshold criterion where the location is actually flawed, POD. Therefore, ROC analysis can explicitly provide a solution for both of the previously described shortcomings associated with POD analysis.

ROC analysis was developed by electrical and radar engineers during World War II for detecting enemy objects in battlefields, i.e., radar operators had to decide whether a blip on the screen represented an enemy target, a friendly ship, or just a noise. Since then, it has been adapted in several areas: clinical, data mining, bioinformatics, machine learning, and just recently, in the field of NDE to assess the ability (performance) of diagnostic systems in classifying subjects

into relevant subgroups. ROC analysis is also useful in comparing, ranking, and classifying inspection techniques according to their ability to achieve their intended functions.

Before going further in the discussion of the ROC analysis, some definitions, key concepts, and abbreviations of various indices used in the analysis process must be viewed first. The next section will be devoted to this purpose, as well as to serve as a basis for the analysis of NDE data by ROC approach. For the indices in the next section to be valid, the actual condition, also known as a “ground truth” or “gold standards” of the test object is to be known either prior to performing the test or by means of hindsight through some physical testings that are often destructive or partially destructive procedures.

3.4.1 Measurements outcomes

Because almost all inspection methods are not perfect and error might accompany the inspection process, four possible decisions are expected. These are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Consider for example the detection of delamination in a concrete bridge deck using an NDE technology when several test points across the surface of the deck are to be tested. If the test is performed over a delaminated location (i.e., positive) and the test results indicate “positive,” this output is to be classified as true positive (TP); if test results indicates “negative,” the output is to be classified as false negative (FN). In contrast, if the test is performed over a sound concrete (i.e., negative) and the test results indicate “negative,” this output is to be classified as true negative (TN); if test results indicate “positive,” the output is to be classified as false positive (FP). Table 3.1 uses a two-by-two representation, which is called the contingency table (also known as confusion or error matrix), to summarize the four outputs.

N^- , N^+ , D^- , and D^+ are the total number of actually negative instances, total number of actually positive instances, total number of negative decisions, and the

Table 3.1: A Contingency Table of an Arbitrarily Inspection Technique.

<u>Actual Condition</u>	<u>NDE Results</u>		
	<u>Sound</u>	<u>Delaminated</u>	<u>Total</u>
<i>Sound</i>	<i>TN</i>	<i>FP</i>	<i>n-</i>
<i>Delaminated</i>	<i>FN</i>	<i>TP</i>	<i>n+</i>
<i>Total</i>	<i>D-</i>	<i>D+</i>	<i>n</i>

total number of positive decisions, respectively.

To represent the contingency table in terms of conditional probabilities, let “ F ” represent the presence of flaws in the test object, and let NDE represent the result of an NDE testing method. The prevalence of the flaws in the test object is represented by $P(F^+)$, and the probability that flaws is actually absent in the test object is represented by $P(F^-)$. Then $POFA$, for example, is equivalent to the conditional probability $P(NDE^+|F^-)$, which is read as “the probability of a positive NDE test, given the absence of flaw.” Similarly, POD (or TPR) can be denoted as $P(NDE^+|F^+)$; FNR as $P(NDE^-|F^+)$; and TNR as $P(NDE^-|F^-)$. Table 3.2 illustrates the contingency table in terms of conditional probability notations.

Table 3.2: A Contingency Table of an Arbitrarily Inspection Technique Using Conditional Probability Notations.

<u>Actual Condition</u>	<u>NDE Results</u>		
	<u>Sound</u>	<u>Delaminated</u>	<u>Total</u>
<i>Sound</i>	$P(NDE^- F^-)$	$P(NDE^+ F^-)$	$P(F^-)$
<i>Delaminated</i>	$P(NDE^- F^+)$	$P(NDE^+ F^+)$	$P(F^+)$
<i>Total</i>	$P(NDE^-)$	$P(NDE^+)$	1

Probability of detection, POD (also known as a true positive rate (TPR), hit rate, recall, or sensitivity of the test) is a measure of the ability of the test to correctly identify the positive instances (flawed test points). POD then, is the number of the true positive indications out of the total number of actually positive cases.

$$P(NDE^+|F^+) = TPR = \frac{TP}{TP + FN} \quad (3.16)$$

Probability of false alarm, POFA (also known as false positive rate (FPR)) is the number of the false positive indications out of the total number of the negative cases.

$$P(NDE^+|F^-) = FPR = \frac{FP}{FP + TN} \quad (3.17)$$

True negative rate, TNR (also known as a specificity of the test) is a measure of the ability of the test to correctly distinguish the negative instances (sound test objects). TNR is the number of the true negative indications out of the total number of negative cases.

$$P(NDE^-|F^-) = TNR = \frac{TN}{TN + FP} \quad (3.18)$$

False negative rate, FNR (also known as a miss rate) is the number of the false negative indications out of the total number of positive cases.

$$P(NDE^-|F^+) = FNR = \frac{FN}{FN + TP} \quad (3.19)$$

Because all test objects are classified in a binary fashion, either positive or negative, it is obvious to now show that the following relationships must exist between the above indices for each actual state

$$POD + FNR = 1P(NDE^+|F^+) + P(NDE^-|F^+) = 1 \quad (3.20)$$

and

$$POFA + TNR = 1P(NDE^+|F^-) + P(NDE^-|F^-) = 1 \quad (3.21)$$

Equations 3.20 and 3.21 reveal that only two indices are necessary to fully describe the probabilities of the four possible inspection outcomes. ROC analysis uses POD on the vertical axis and POFA on the horizontal axis for multiple threshold settings. Every threshold setting can define a single point in the ROC graph.

POD and POFA can also be estimated from the probability density functions of the distribution of the results of the actually positive and the actually negative cases. When these two distributions overlap, which is normally the case with most diagnostic systems such as NDE technologies, no single threshold value can clearly separate the populations. Therefore, several threshold values are recommended to be investigated and for every decision threshold setting there will be a unique data set in the contingency table.

Consider for example Figure 3.7 which shows the distributions of the results of an inspection technology over the actually positive and the actually negative instances in an arbitrary study. By setting a high (to the right hand side of the figure) threshold value, both *POD* and *POFA* will be less frequent because almost all the data points will be classified as negatives. Conversely, setting a low threshold (to the left hand side of the figure), *POD* and *POFA* will be more frequent. Any other setting for the threshold value yields a unique pair of *POD* and *POFA*.

Figure 3.7: Overlapping distributions of arbitrary test results over the actually positive and negative instances.

3.4.2 ROC curve construction

ROC graphs are two dimensional graphs in which *POD* (also known as test *sensitivity*) is plotted on the vertical axis and *POFA* (also known as $(1-\text{specificity})$) is plotted on the horizontal axis considering multiple threshold settings. Therefore,

an ROC graph is a representation of the reciprocal relationship between the cost (*POFA*) and the benefits (*POD*) [93, 94]. Point (0, 0) in the ROC graph represents the threshold criterion that never produce a positive classification, whereas, point (1, 1) is the criterion of classifying all test points as positives. All other threshold criteria can produce *POD* and *POFA* values somewhere between those two extremes.

As shown graphically in Figure 3.8, the diagonal line connecting points (0, 0) to (1, 1) is the ROC curve of corresponding to random guessing where the distributions of the positive and negative instances totally overlap, and therefore, the inspection method has no ability to distinguish the positive and negative instances. At any particular threshold setting, *POD* and *POFA* are similar to each other and different threshold settings can eventually produce a 45° line. This line is also known as “the reference line” where the reliability of diagnostic systems is normally compared to.

In contrast to random guessing, when the inspection method has a perfect ability to produce totally separated positive and negative distributions, the ROC curve of such inspection methods can be represented by two line-segments connecting points (0, 0) to (0, 1) and then point (0, 1) to point (1, 1), Figure 3.8. As discussed previously, most inspection methods such as NDE technologies are not perfect, and therefore, their ROC curves lie between those two extremes.

In addition, any inspection method that produces an ROC curve in the lower right triangle (compared to the 45° reference line) performs worse than random guessing. In such a scenario, most of the actually positive test points (i.e., test points over delaminated concrete) are incorrectly classified as negative (i.e., sound concrete), and most of the actually negative test points (i.e., test points over sound concrete) are also incorrectly classified as delaminated. However, negating the classification strategy can enhance the performance of the inspection method and eventually move the ROC curve to the upper right triangle [95].

Figure 3.8: ROC curves of inspection techniques with different performances.

3.4.3 Area under the ROC curve (AUC)

The standard measure that summarizes the accuracy of an inspection method is the area under the ROC curve (often denoted as AUC or AUROC). Estimation of AUC reduces ROC performance to a single quantitative measure necessary to compare the performance of classifiers (herein NDE technologies) [96,97]. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1.0. However, since random guessing produces the diagonal line between (0, 0) and (1, 1), which has an area of 0.5, no realistic classifier should have an AUC less than 0.5. AUC has two interpretations [96,98]:

- The average POD over the entire range of POFA, or vice versa, or
- The probability that a randomly selected positive (i.e., delaminated concrete) test point will have a test result that indicates greater suspicion of being delaminated than a randomly selected negative (i.e., sound concrete) test point.

An ROC, and consequently AUC, can be constructed/estimated either parametrically or nonparametrically depending upon the distribution functions of the

two cases (i.e., delaminated vs. sound concrete) under consideration. The parametric estimation requires that the inspection test results of the actually positive and actually negative test points are independent and normally distributed with different means and standard deviations [99,100].

To elaborate more, let $X = x_1, x_2, \dots, x_n$, and $Y = y_1, y_2, \dots, y_m$, be respectively, two independent variables denoting the test results of the delaminated and sound test points, Figure 3.7. Let also that D (for delaminated) and S (for sound) are the cumulative distribution functions of the X and Y , respectively. Then, at any particular threshold setting, th , POD and $POFA$ can be defined as follows:

$$POD = 1 - D(th), \text{ and } POFA = 1 - S(th) \quad (3.22)$$

The parametric estimation of AUC assumes that the two populations, X and Y , are normally distributed with $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$. If this assumption holds, then the area under the curve, AUC, can be estimated using the following statistical representation:

$$AUC = \phi\left(\frac{a}{1 + b^2}\right) \quad (3.23)$$

where a , b , and ϕ are, respectively, the standardized difference in the means of the test results of the delaminated and sound test points, the ratio of the standard deviations of the test results of the delaminated and sound test points, and the standard normal distribution function of the test results. The factors a and b can be calculated as follows [100–102]:

$$a = \frac{\mu_x - \mu_y}{\sigma_x}, \text{ and } b = \frac{\sigma_y}{\sigma_x} \quad (3.24)$$

Equation 3.23 then can be rewritten as

$$AUC = \phi\left(\frac{\mu_x - \mu_y}{\sigma_x^2 + \sigma_y^2}\right) \quad (3.25)$$

AUC can also be estimated nonparametrically without any assumptions regarding the distributions of the two populations [100,101]. The simplest way to nonparametrically estimate AUC is by connecting data points by straight lines and sum over the areas under the curve [103].

ROC analysis has been very rarely used in the field of nondestructive evaluation. The first study to demonstrate the applicability of ROC analysis for the estimation of the reliability of NDE technologies was done by Nockemann in 1991. The study examined the reliability of radiography for detection defects in welds. The study showed the usefulness of ROC analysis for the performance assessment in nondestructive testing (NDT) experiments. Shin and Grivas [46] used ROC analysis to investigate the reliability of GPR for the detection of subsurface damage in a concrete bridge deck. However, a limited number of core samples was used in the analysis. Reliability of GPR for the detection of corrosion damage was also investigated by Martino et al. using ROC analysis [104].

NDE technologies are improved with time, relied upon more frequently than before in condition assessments of transportation infrastructures, and factors influencing the overall reliability need further investigations as compared to the traditional inspection methods such as visual inspection and sounding. Therefore, the limited studies cited above suggest that more elaboration is needed to investigate the applicability of ROC analysis for the estimation of other case studies or other NDE technologies.

Chapter 4

DESCRIPTION OF THE EXPERIMENTAL WORK

In order to provide suitable data for ROC analysis, two full-scale reinforced concrete slabs were fabricated and tested and three in-service bridge decks were tested. All of the in-service bridge decks tested, as well as the fabricated slabs, are approximately located in the same geographical area in the state of Missouri. This chapter will be devoted to discussing the experimental work in detail.

As part of the experimental work, two reinforced concrete mockups representing bridge decks were fabricated and tested. This part of the experimental work is necessary as the exact location and depth of delamination targets were known prior to testing. The exact location and depth of delamination targets allow for an accurate reliability analysis as these slabs can provide a bench mark for the reliability of the NDE technologies used. The other part of the experimental work involved the evaluation of three in-service highway bridge decks using the same NDE technologies used in the laboratory work. Assessing the reliability of the candidate NDE technologies in real world applications is also important to determine the actual capability of these technologies under practical operating conditions where high noise from traffic flow and environmental conditions might exist.

Once the data from an individual NDE technology is collected and processed to extract the features and information necessary for condition assessment, ROC analysis was used to analyze the processed data and to quantitatively estimate its overall reliability. Details of ROC analysis of each NDE technology will be provided in the next chapter.

4.1 Fabricated Slab I

This fabricated slab was 20 ft x 10 ft x 7 in. thick. The slab was reinforced with two mats of steel reinforcement with top and bottom concrete cover (from the nearest concrete surface) of about 2.5 in. and 2 in., respectively. Steel bars #3 spaced at 8 in. in both directions were used in both layers. As per Missouri Department of Transportation (MoDOT) requirements for bridge decks material, type B-1 concrete with 28-day design compressive strength of 4000 psi was used in the slab. The slab was built over an existing concrete pad; 1 in. Styrofoam sheets were placed underneath the new slab as a separation between the old and the new concrete which is necessary to avoid intervention from the old concrete pad. To better simulate the case of testing an actual bridge deck, the surface of the slab was finished with light brush strokes in the transverse direction of the slab.

In an effort to accurately simulate the actual condition of subsurface delamination, three different materials representing delamination had been investigated priori using small concrete blocks. The purpose of this work was to determine which material could provide similar results compared to an actual delamination. Four small concrete blocks (2 ft X 2 ft X 5 in. thick) were fabricated at the Civil & Environmental Engineering Department (CEE) laboratories using a 5000 psi ready mixed concrete. In addition to the reference block which contained no delamination target, each of the other three blocks contained either open-celled Styrofoam, bubble wrap, or open-celled plastic sheets; these were placed in the mid-height of the block. The three materials representing delamination targets were approximately the same size and thickness (≈ 0.2 in. thick). The concrete blocks were water-cured for two weeks and then tested using impact echo. The three materials used to simulate delamination in these blocks were intentionally chosen because each has a relative acoustic impedance close to the relative acoustic impedance of air, assuming that the delamination is an air-filled. Figure 4.1 shows the blocks before pouring the concrete and the concrete type used (5000 Psi ready

mixed concrete). Analysis of the frequency domains of IE testing indicated that both Styrofoam and plastic sheets provided relatively similar results to each other and also similar to what is expected from testing air-filled delaminated concrete. On the other hand, the frequency domain of the bubble wraps also provided the expected results, but the frequency domain was less distinct as compared to the other two materials. This noisy output might be due to the surface texture of bubble wrap material suggesting it would be suitable to simulate honeycombing damage, but not delamination. Figure 4.2 illustrates the IE output of the reference and the three materials used and one can clearly notice that both Styrofoam and plastic sheet materials gave almost identical output as compared to the bubble wraps. Therefore, as will be discussed later, delamination damage was introduced in the first and second slabs by using an open-celled Styrofoam material and plastic sheets, respectively. Because a large portion of delamination stem from corrosion

Figure 4.1: Concrete blocks to investigate different materials simulating delamination.

of steel reinforcement and these are within relatively 2-3 in. deep from the surface of the concrete, only shallow delamination (2.5 in. deep) of various sizes were considered in the first slab. Delamination targets ranged in size from 6 in. x 6 in. to 24 in. x 24 in. It was assumed that delaminated areas smaller than 6 in. x 6 in. will be excluded from repair plans even if they have been detected, and therefore, they might practically be outside the target size of interest. On the other hand, reviewing the available literature reveals that shallow delamination larger than 24

Figure 4.2: IE frequency domains of different materials simulating delamination (A) sound concrete, (B) delaminated: Styrofoam, (C) delaminated: plastic sheets, and (D) delaminated: bubble wrap.

in. x 24 in. have been mostly detected using sounding techniques such as hammer or chain drags, and therefore, there is no need to implement such an advanced NDE techniques to detect such a large delamination. Additionally, this size range of delamination is most frequently cited in the literature, see for example [10, 105]. Figure 4.3 shows a schematic of the size, depth, and horizontal distribution of the delamination targets in the first slab. Figure 4.4 shows the formed slab before pouring concrete.

4.2 Fabricated Slab II

The dimensions of this slab was 11 ft x 19 ft x 7 in. thick that contains delamination targets of various sizes and depths using an open-celled plastic sheets of 0.15 in. thick. The strategy for building this slab was to investigate the effect of delamination size and depth. As stated earlier in chapter two, delamination can also form as a result of some other “non-corrosion” factors, and therefore, might form in different depths than the cover of steel reinforcement. The size of delamination targets were 3 in. x 3 in., 6 in. X 6 in., and 9 in. x 9 in. The

Figure 4.3: Size, depth, and horizontal distribution of delamination targets of fabricated slab I.

size of delamination targets has been changed in this case study for two reasons. First, analysis of IE and IRT output from the larger delamination in the first slab was almost 100% and 90%; this brought to our attention that delaminated areas larger than 12 in. X 12 in. will be almost entirely detected using these NDE technologies, especially if the delamination is shallower than 2-4 inches. The other reason is to investigate the detectability of the candidate NDE techniques in detecting delamination of different depths and sizes for a comprehensive reliability assessment. Therefore, each group size of delamination was embedded at a depth of 2 in., 4 in., and 5 in. from the surface of concrete.

This slab was also reinforced with two mats of uncoated #5 bars spaced at 12 in. in both directions for both layers. Figure 4.5 shows the horizontal distributions and size of delamination targets. Styrofoam sheets of 1 in. thick were put underneath this slab to create a separation between the base materials and the new concrete.

Figure 4.4: Details of the reinforcement and delamination targets of fabricated slab I.

As in the first slab, type B-1 concrete with 28-day design compressive strength of 4000 psi was also used in the slab. Details of reinforcement and delamination targets can be seen in Figure 4.6.

Figure 4.5: Size, depth, and horizontal distribution of delamination targets of fabricated slab II.

4.3 Kansas City Bridge, Bridge A0295

Bridge A0295 was built in 1959 in Kansas City, MO during the construction of I-70 to serve as a ramp between I-70 and US-24; however, US-24 was never constructed at this location and the bridge was abandoned in place. The bridge deck is of a

Figure 4.6: Details of the reinforcement and delamination targets of fabricated slab II.

tapered shape with a length of 216 ft and a width ranging from 39 ft to 21.5 ft. The total deck area is 6540 sq. ft. Chain drag, performed by MoDOT personnel, and infrared thermography (IRT), performed by the research team, indicated a number of areas with subsurface delamination in the deck. Limited physical sampling in the form of drilling a 1/2 in. hole in the deck and observing the delamination with a borescope was conducted in one area of delamination.

The bridge has never been exposed to traffic loading or deicing chemicals, and therefore, it was believed to be an excellent candidate for the investigation of the reliability of the candidate NDE technologies, especially GPR, to detect actual delamination in concrete that has been resulted from natural processes over the course of the last 50+ years. These areas of delamination may have been produced from the ingress of water through cracks in the deck, resulting in corrosion of the reinforcing steel. Freeze-thaw damage may have also played a role. Regardless, the areas of delamination are naturally-occurring, but without increased levels of chlorides. Figure 4.7 is an aerial view of the bridge. Figure 4.8 offers a closer view of the bridge deck, showing absolutely no symptoms of distress or previous repairs.

Figure 4.7: Aerial view of Bridge A0295 in Kansas City, Missouri.

Figure 4.8: A close view of the deck of bridge A0295 showing no symptoms of previous repairs.

4.4 Fulton City Bridge, Bridge A2111

Bridge A2111 was built in 1968 and is located on Rt. 54 near Fulton, MO. The bridge is about 175 ft long. The bridge deck has two 12 ft-wide lanes and two 10 ft-wide shoulders, making the total deck area approximately 7,700 ft². A visual inspection of the bridge deck indicated that a number of patches had been applied to repair damage in the deck. Figure 4.9 shows an aerial image of the location of bridge A2111, as well as bridge A2112. In addition, the deck surface contains a number of exposed steel bars as a result of severe concrete spalling. Figure 4.10 is a close view of the deck's surface showing previous repairs.

Figure 4.9: Aerial view of Bridges A2111 and A2112 in Fulton City, Missouri.

Figure 4.10: A close view of the deck surface of Bridge A2111 showing previous repairs.

4.5 Fulton City Bridge, Bridge A2112

Bridge A2112 is a three-span bridge located in close proximity of bridge A2111 in Fulton, MO. The bridge is 152 ft long, with a roadway of 44 ft curb-to-curb width. The reinforced concrete deck is 7.5 in. thick. Visual inspection of the deck surface indicated a number of concrete repairs in the form of concrete and asphalt patches have been performed on the bridge. A close view of the surface of the deck can be seen in Figure 4.11.

Figure 4.11: A close view of the deck surface of Bridge A2112 showing concrete spalling and some previous repairs.

In the next chapters, ROC analysis will be used to analyze the reliability of the most widely NDE technologies implemented by DOTs and bridge owners for the detection of delamination in concrete bridge decks. As stated earlier, this analysis method can provide for a quantitative index, i.e., the area under the ROC curve (AUC) necessary to compare the performance of one NDE technology to another.

Chapter 5

ANALYSIS AND RESULTS OF IMPACT ECHO (IE) TECHNOLOGY

5.1 Introduction

The impact echo instrument used in this research consists of a compact test head with automatic solenoid impactor and a piezoelectric displacement transducer, and a data acquisition system. The solenoid impactor has a movable anvil attached to the base of the test head, which allows for testing very thin concrete sections such as sections over shallow delaminated areas. Figure 5.1(A) shows the IE system with the extension arm attached to estimate the P-wave speed, whereas Fig. 5.1(B) shows a close look to the compact test head.

IE testing has been performed on the two fabricated slabs and two large sections of the decks of bridges A0295 and A2111 using the following parameters and filtering algorithms as recommended by the instrument manufacturer:

- Sampling Rate of 7 microseconds: How often (in the time domain) the system will acquire data points within a given data trace.
- Points Per Record of 512: The number of sampling points applied for each waveform.
- Number per records of 2: The total number of IE data records that will be saved and averaged for each individual data point.
- Solenoid time of 12 microseconds: This is the time applied to drive the solenoid; this controls the amount of energy and the wave length of the generated stress waves.

Once all datasets were collected and stored, they were transferred to a PC with

the necessary software for analysis had already been installed and tested. As per the manufacturer recommendations, a high-pass filter with a cutoff frequency of 6000 Hz was applied to all suspected sound and deep delaminated test points. On the other hand, a bandpass filter between 10000-50000 Hz was applied on all suspected shallow delaminated points. Filtering was necessary to filter out the effect of R-waves captured by the receiving transducer during the data acquisition process and to relatively refine the reflected signals. In addition, the datasets were windowed to include only reflections of the P-waves during the first $1500\mu s$ as recommended by the manufacturer to reduce the effects of noise for better interpretations. Figure 5.2 illustrates an arbitrarily IE response of a sound area before and after processing.

Figure 5.1: Components of impact echo instrumentation.

Figure 5.2: Time and frequency domains of an arbitrary sound test point before (A) and after (B) processing.

5.2 Results and Analysis of Fabricated slab I

A total of 684 test points were tested, of which 104 test points were over delaminated concrete and the other 580 over sound concrete. Since the test is a point-wise test and different points might result in different output, IE testing was performed three times over an eight months timespan. The first and second tests spaced two months apart and the third test was performed six months after the second test. As stated earlier, testing was performed on a grid of 6 in. by 6 in. to provide a higher resolution contour map for the condition assessment of the slab, i.e., to precisely map the extent of delamination.

The velocity of P-wave had been estimated prior to each test in order to precisely estimate the frequency of the reflected waves and also to evaluate the consistency of concrete quality over the entire slab area. Minimal variation was found in the speed of P-wave in all three measurements performed; variation might have resulted from inconsistency of the compaction process of concrete during pouring. Table 5.1 shows some statistics regarding the estimation of P-wave speed

for the fabricated slabs I and II. The average P-wave speed of 13250 *ft/s* was used in the analysis of data of this case study.

Insignificant variation was reported in the results obtained from the three IE testing trials; this variation might be resulted due to the variation of the P-wave velocity with time as the velocity is a function of concrete properties such as Poisson's ratio, modulus of elasticity, and density. Therefore, results obtained from the last trial were used in the analysis of IE reliability as the P-wave velocity should be constant nine months after pouring of the concrete.

Table 5.1: Statistics of P-wave Velocity Estimation (in *ft/s*) of the Fabricated Slabs.

Parameter	Slab I	Slab II
Minimum	12899	11500
Median	13257	12323
Average	13242.5	12289.2
Maximum	13645	13020
Standard Deviation	204.47	495.93

As expected, testing over a shallow delaminated spot resulted in a very low P-wave frequency for most test points as compared to testing over a sound spot. Figure 5.3 shows the frequency domain of an arbitrarily processed test point of shallow-delaminated concrete. In this figure, the higher-amplitude peak at 2000 Hz represents the flexural mode of vibration frequency, while the other lower-amplitude higher frequency peak represents noise resulting from multi-flexural mode of vibration resulting from testing off-center of a shallow delaminated point.

Once the dataset has been processed successfully, a contour map representing the dominant frequencies of data points has been drawn, as shown in Figure 5.4. The dashed-line boxes represent the actual locations of delamination targets. Examining this figure, one can clearly see that IE technology was capable of detecting all of the delamination targets embedded in the slab with minimal incorrect (both positive and negative) indications; this reveals the excellent ability of IE for the detection of fracture planes such as delamination in plate-like members. However,

Figure 5.3: Frequency domain of an arbitrary processed test point of shallow delaminated concrete.

this statement is still subjective and needs to be confirmed through a quantitative analysis. Therefore, ROC analysis has been performed to quantitatively estimate the reliability of IE for the case study presented herein; in this way, the index can be used in comparison against other NDE technologies used for the same purpose.

Figure 5.4: A contour map showing the frequency of the reflected P-wave of the first fabricated slab.

The first step of ROC analysis is the construction of probability density functions of the test results over the delaminated and sound test points. Although this step is optional, it can be beneficial to visually estimate the overlap between the results of both groups and to determine the practical range of threshold settings that can be used in the subsequent analysis steps. Figure 5.5 shows the histograms of the test results over the delaminated and sound test points with the Kernel smoothing probability density function superimposed on each group. Kernel smooth function has been used since the datasets of both groups do not fit any known parametric function, such as normal, gamma, or lognormal. As stated earlier, this step is optional, and therefore, Kernel smooth function is satisfactory to visually estimate the overlap between the results of both groups.

Figure 5.5: Distribution of IE test results over sound and delaminated test points.

Figure 5.5 also reveals the capability of IE testing to distinguish sound and delaminated test points; no significant overlap exists between the results of both groups. However, this figure does not provide the quantitative reliability measure necessary to compare one technology to another.

In order to determine the exact reliability index of IE technology, ROC analysis has been performed. As stated earlier, ROC draws the probability of detection (POD) versus the probability of false alarm (POFA) of the technology as functions

of the decision threshold. Therefore, the first step in ROC analysis is to set a value for the decision threshold (here the frequency of the reflected P-wave) and then calculate the resulting POD and POFA indices. Every test point with a frequency value less than the preset threshold will be classified as positive (i.e., delaminated) and every test point with a frequency value larger than the threshold will be classified as negative. Other threshold settings are investigated and the resulting POD and POFA indices associated with each threshold setting are calculated. Once all the desired threshold settings have been considered, the resulting pairs of POD and POFA are drawn in ROC space to construct ROC curve.

The recommended threshold settings for ROC analysis, as determined from the IE data, ranged from about 3800 Hz to 13000 Hz, see Figures. 5.4 and 5.5. As discussed earlier, the extreme values of 3800 Hz and 13000 Hz must be included in the analysis since these settings can provide for points of 0, 0 (i.e., all data points will be classified as negatives) and 1, 1 (i.e., all data points will be classified as positives), which are necessary to calculate the area under the curve (AUC). The final ROC curve is shown in Figure 5.6. Threshold settings have been varied in a small interval to produce a smooth ROC curve.

Figure 5.7 shows the POD and POFA as functions of threshold settings. In this figure, the optimum threshold value of 8118 Hz has been determined to produce the best balance between POD and POFA (i.e., higher POD at lower POFA), and therefore, increases the accuracy of the dichotomization process. In fact, as can be seen in this figure or Figure 5.5, any threshold setting between about 7000 Hz and 9000 Hz would produce a close estimate for POD and POFA, but the setting of 8118 Hz has been found as the best in this range.

To confirm the estimation of the best threshold setting, the total delaminated area determined from using various threshold settings have been compared against the actual delaminated area of the slab. For example, Table 5.2 lists the estimations of delaminated areas from five arbitrary threshold settings, including the extremes and the optimum settings, and the percentage error between the estimation of IE

and the actual area. As can be seen in this table, the threshold setting of 8000 Hz has the lower error as compared to other settings.

Area under the ROC curve (AUC) has been nonparametrically estimated as the normality of test results have not been confirmed. The nonparametric estimation of AUC involves connecting every two adjacent data points in the ROC space by a straight line, and then calculating the resulting area under the curve using any mathematical method. As can be seen in Figure 5.6, AUC has been estimated as 0.994. In contrast to previous statements regarding the capability of IE for the detection of delamination, this last statement (i.e., AUC) is quantitative considering a high resolution analysis process. Such a quantitative index is of extreme importance to compare the reliability of different NDE technologies.

Table 5.2: Estimations of Delaminated and Sound Areas as Functions of The Decision Threshold Setting.

Index	Frequency of the Reflected P-Wave (Hz)				
	12640	10000	8000	5500	3940
POFA	1	0.25	0.02	0	0
POD	1	1	0.97	0.45	0
Defective area from FP (sq. ft)	145	36.25	2.9	0	0
Defective area from TP (sq. ft)	26	26	25.30	11.7	0
IE total defective area (sq. ft)	171	62.25	28.2	11.7	0
Actual total defective area (sq. ft)	26	26	26	26	26
% Error	557	139	8	-55	-100
FNR	0	0	0.027	0.55	1
TNR	0	0.75	0.98	1	1
Sound area from FN (sq. ft)	0	0	0.702	14.3	26
Sound area from TN (sq. ft)	0	108.75	142.1	145	145
IE total sound area (sq. ft)	0	108.75	142.80	159.3	171
Actual total sound area (sq. ft)	145	145	145	145	145
% Error	-100	-25	-1	9	18

Figure 5.6: ROC curve of IE technology for the detection of shallow delamination.

Figure 5.7: POD and POFA of IE technology for the detection of shallow delamination as functions of the threshold settings.

5.3 Results and Analysis of Fabricated slab II

Results from fabricated slab I suggest an investigation of the reliability of IE for the detection of smaller and deeper delamination since the reliability of the technology for the detection of relatively large shallow delamination was about 95%. Therefore,

a second slab was fabricated to include delamination targets of smaller size and located on different depths. In order to ensure not missing a delamination target and to precisely map the extent of delamination, IE testing was performed using different grid sizes suitable for each group, i.e., the grid size ranged from 3 in. x 3 in. to 9 in. x 9 in. Figure 5.8 illustrates the time and frequency domains of an arbitrary test point over a deep delamination target. Other output of testing sound concrete or shallow delaminated concrete have already been discussed.

Figure 5.8: Time (A) and frequency (B) domains of an arbitrary deep delaminated test point.

Velocity of the P-wave has also been estimated by testing multiple locations over the surface of the slab using the extension arm provided with the instrument. Some statistics are shown in Table 5.1. The average P-wave velocity of about 12300 ± 200 ft/s was used in the analysis of this case study. As can be seen in the table that the standard deviation is larger for this case study. However, since

this slab has multiple groups of delamination that were analyzed individually, the relative speed associated with each group was used, but in general these were within ± 200 ft/s as per the estimation above. As discussed earlier, these steps were performed only to ensure the consistency of the quality of concrete and for the estimation of the depth of delamination.

ROC analysis was performed individually for each group of delamination targets. a total of 49 test points were tested in each group, of which, 13 points were over delamination targets. A unique range of threshold values was determined for each group depending on the results obtained from testing. Figure 5.9 illustrates the contour maps of the dominant frequency of the reflected P-waves of each group. Figure 5.10 illustrates the histograms with Kernel smooth function superimposed on each graph as the data were not normally distributed as was the case with the first fabricated slab. The resulting ROC curves of all individual groups were categorized per size of delamination targets and are shown in Figure 5.11.

Figure 5.9: Contour maps of the frequency of the reflected P-waves of the individual groups in the second slab.

Figure 5.10: Distribution of IE test results over sound and delaminated test points.

Figure 5.11: ROC curves of IE for the detection of delamination of different size and depth.

The area under the ROC curve (AUC) of each of the nine groups discussed were nonparametrically estimated, and they are listed in Table 5.3. Examining this table, the reader can easily notice that IE has an excellent ability to detect shallow and large delamination targets of less than 4 in. deep and larger than 36 in.², respectively. Deeper and smaller delamination targets were not detected

accurately, and therefore, a large number of false decisions were reported.

Table 5.3: AUC Estimations of IE for the Detection of delamination of Various Sizes and Depths.

Group	Depth (in.)	AUC
3 in. x 3 in.	2	0.67
	4	0.46
	5	0.26
6 in. x 6 in.	2	0.98
	4	0.99
	5	0.83
9 in. x 9 in.	2	1.00
	4	1.00
	5	0.99

As with the first fabricated slab, there is a wide range of possible threshold values that can be used to maximize the accuracy of the technology for the detection of large or shallow delamination. For example, for targets of 4 in. deep and an area of larger than 36 in.², any value within the range of 7000 Hz to 9000 Hz would be ideal to detect almost all targets with no-to-minimal false decisions. As discussed earlier, however, the AUC is independent of the threshold setting, and therefore, can be an ideal index to describe reliability of NDE technologies regardless of threshold settings. Figure 5.12 shows the POD and POFA as functions of the decision threshold, i.e., frequency of P-waves, for the nine groups of delamination. Inspectors can use such graphs to establish guidelines to interpret IE signals in order to overcome the challenges of subjective interpretations.

Figure 5.12: POD and POFA of IE for the detections of delamination of different sizes and depths as functions of the decision threshold settings.

5.4 Results and Analysis of Bridges A0295 & A2111

Due to the point-wise nature of IE testing, and the fact that the test may be repeated many times in some test points until a stable output is obtained, IE testing is a very time consuming and might require a few days to test an entire bridge deck. According to this, IE was only performed on specified sections of the

decks of bridges A0295 and A2111. Testing was performed using a grid of 1 ft x 1 ft in both case studies. The tested section in each bridge deck was chosen so that it contains both sound area and delaminated area(s). Prior knowledge of the location of delamination targets as determined from the combination of infrared thermography (IRT), and an expert-level sounding was used to choose the tested sections.

Dimensions of the tested section on the deck of bridge A0295 were 12 ft x 15 ft. IE test was performed on a total of 180 test points. A test point is defined at the center of each square in the grid shown in Figure 5.13. Out of these, 40 test points were above delamination targets. Areas of delamination, as determined from the combination of IRT and sounding, were encircled in the figure; the shape of delaminated targets were approximated to rectangular shapes. As is the case with the fabricated slabs, P-wave velocity was determined from testing ten locations over the deck's area. Some statistics of the P-wave velocity estimation are provided in Table 5.4. The average P-wave velocity was used in the analysis of signals, and consequently, in the ROC analysis of IE data.

Figure 5.13: IE tested section on bridge A0295, Kansas City, MO.

A contour map of the processed IE data is shown in Figure 5.14. The actual locations of delamination targets are superimposed on the contour. Distributions of test results of the actually negative and positive test points with Kernel probability distribution functions are shown in Figure 5.15. Although qualitative, these figures reveal the capability of IE to recognize areas of delamination. ROC analysis was performed using the range of threshold values shown in Figs. 5.14 and 5.15. The

Table 5.4: Statistics of P-wave Velocity Estimation (in ft/s) of A0295 and A2111 Bridge Decks.

Parameter	A0295 Bridge	A2111 Bridge
Minimum	10221	11340
Median	11292.5	11774
Average	11176.1	11886.6
Maximum	12311	12455
Standard Deviation	641.9	379.24

same testing parameters with the exception of grid interval and signal filtering algorithms, which were used in the previous case studies, were also used in this case study to enhance the quality of IE signals for better interpretation. ROC curve resulted from investigating multiple thresholds of the return signal frequency is shown in Figure 5.16. AUC was nonparametrically estimated as 0.975, which reveals the excellent ability of IE technology to detect delamination. Finally, the POD and POFA of IE as functions of the decision threshold were plotted as shown in Figure 5.17. This figure can be used to determine the probability of detection or the probability of false alarm with each threshold setting and the optimum setting for the application, whether to operate towards the left-hand side of Figure 5.15 for minimal false alarm rate, but a minimal detection rate, or to the right-hand side for larger detection, but a larger false alarm should be anticipated.

Figure 5.14: A contour map of the frequency of the reflected P-waves.

Figure 5.15: Distributions of the IE test results over sound and delaminated test points.

Figure 5.16: ROC curve of IE technology for the detection of real delamination in Bridge A0295, Kansas City, MO.

Figure 5.17: POD and POFA of IE technology as function of the decision threshold for the detection of real delamination in Bridge A0295, Kansas City, MO.

Similarly, a rectangular section on the deck of bridge A2111 in Fulton, MO, was tested by means of 1 ft intervals in both directions. The tested section was 14 ft long and 9 ft wide, two feet away from the curb. Data was processed using the discussed smoothing and filtering parameters. Figure 5.18 shows the tested section, whereas Figure 5.19 shows a contour map of the processed IE data.

Figure 5.18: Tested section of bridge A2111 in Fulton, MO.

Figure 5.19: A contour map of the frequency of the reflected P-Waves.

In order to visually investigate the accuracy of the data, Figure 5.20 illustrates the histograms of the test results over the actually delaminated and sound test points. As expected, there was no significant overlap between the results, which reveals the good ability of IE technology for the delineation of delaminated from sound concrete.

Figure 5.20: Distributions of the IE test results over sound and delaminated test points.

ROC analysis was performed using threshold values within the range of the frequency of the reflected P-waves shown in Figs. 5.19 and 5.20. The actual condition of the entire bridge, including the tested section, had been determined from the combination of sounding and infrared thermography. Figures 5.21(A and B) show the actual condition status obtained from sounding and IRT individually, where areas of delamination were encircled in both figures; Figure 5.21(C) represents the actual condition of the bridge deck as determined from the combination of sounding and IRT. This combination was necessary to accurately defining the actual condition for an effective reliability analysis since both sounding and IRT have missed some of the delamination targets.

ROC curve resulting from the analysis of IE data is shown in Figure 5.22. AUC was nonparametrically estimated to be 0.946 as shown in the figure. This high AUC reveals the excellent ability of IE for the detection of delamination considering multiple threshold settings. Figure 5.23 illustrates both the POFA and POD as functions of the threshold settings.

The analysis of multiple case studies ranged from fabricated R.C. slabs with synthetic delamination targets to a number of real delamination targets in in-service bridge decks reveals the great ability of IE for the detection of delamination with

Figure 5.21: The actual condition of bridge A2111, Fulton, MO; (A) results from sounding, (B) results from infrared thermography, and (C) a combination of (A) and (B).

a very small rate of false alarm in each case study. This conclusion has been reached considering the very high areas under the resulting ROC curves for all case studies. As discussed earlier, such an index can be of great importance when bridge owners and transportation agencies seek information about the most reliable NDE technology, i.e., maximum POD with minimal POFA, to eliminate or reduce the unnecessary repair resulting from incorrect data.

Figure 5.22: ROC curve of IE technology for the detection of delamination.

Figure 5.23: POD and POFA of IE technology as functions of the decision threshold setting of bridge A2111.

In the next chapter, the analysis of the reliability of infrared thermography (IRT) for the detection of delamination for the same case studies will be discussed.

Chapter 6

ANALYSIS AND RESULTS OF INFRARED THERMOGRAPHY (IRT) TECHNOLOGY

6.1 Introduction

As discussed earlier, IRT technology uses the radiant temperature of a material and converts that temperature measurement into a visual image. The technology offers faster testing and quicker analysis as compared to the discussed IE technology. Passive thermography was used in the entire course of testing of all case studies considered in this research as this approach is the most convenient one for testing in-service bridge decks.

Thermal data analyzed in this research were all from the first fabricated slab and from the in-service bridge decks. Data from the second fabricated slab was not included in the analysis since the surface of slab was of inconsistent texture.

6.2 Data Acquisition

Figure 6.1(A) shows the testing setup used to capture thermal images of the first fabricated slab. A similar set-up was used for field testing, with the exception that the thermal camera was hand-held to capture images of different portions of the deck, Fig 6.1(B). This arrangement was used to model an inspector using a hand-held camera to inspect a concrete bridge deck from a standing position.

A hand-held camera, Flir T-620, was used for capturing thermal and digital images necessary for analysis from all cases studies discussed in this research. Specifications of the used thermal camera are summarized in Table 6.1. Most thermal images used in the reliability analysis were captured between 10 am and 4 pm as this time interval provides for suitable thermal contrast between the area

Figure 6.1: IRT testing setup: (A) fabricated slabs, (B) in-service bridge decks.

above shallow delamination and a sound area [33]. It should be noted also that the digital images produced by Flir T-620 camera have a much higher resolution than the thermal images.

A total of 22 thermal images were analyzed using ROC analysis to assess the reliability of the IRT technology. A group of 12 thermal images was from the fabricated slab I, where each individual image covered only a small portion of the slab. Another group of 10 thermal images from different in-service bridge decks (A0295, A2111, and A2112 bridges) was also analyzed; each image covered only a small portion of the decks' surface. Although the same analysis procedure was used in both groups, these two groups were analyzed individually to find out whether the analysis method can be successfully applied on in-service bridge decks.

Table 6.1: Specifications of Flir T-620 Thermal Camera Used in Field Testing.

Features	Remark
Field of view	25°x19°
Minimum focus distance	0.82ft(0.25m)
Thermal sensitivity	$\leq 50mK@ + 86^{\circ}F(\leq 0.04^{\circ}C@30^{\circ}C)$
Focal Plane Array/Spectral range	Uncooled microbolometer/7.8 – 14 μm
Thermal image resolution	640 x 480 pixel
Digital image resolution	2592 x 1944 pixel
Digital zoom and pan/focus	1-4X continuous, digital zoom, including panning
Image frequency (frame rate)	30Hz

6.3 IRT Data Processing

In order to best understand the ROC analysis procedure of IRT data, consider Figure 6.2, which shows the thermal signature of pixels along an arbitrary line on the surface of a concrete test object. The line passes over several delamination targets as well as over sound concrete. By setting a high threshold for the thermal contrast between delaminated and sound pixels, as shown in Figure 6.2(A), a very low number of those pixels will be classified as positives (i.e., delaminated) and consequently, a very low TPR and FPR will be reported. If the threshold is lowered slightly as shown in Figure 6.2(B), both TPR and FPR will be increased since there will be some actually positive and actually negative cases that will be classified as positives. By further lowering the threshold value, Figure 6.2(C), a higher TPR and FPR will be reported. An image analysis process was developed

Figure 6.2: Effects of threshold setting in the detection of delamination using IRT technology (A) high Δt , (B) moderate Δt , and (C) low Δt .

to allow for the thermal images and the digital visual images to be compared on a pixel by pixel basis. Due to the mismatch between the view-angles of any pair of thermal and digital images, it was necessary to spatially coincide each pair of thermal and digital images using a suitable image registration algorithm;

digital and thermal images are captured using different optics, i.e., lenses. Tools+ software, developed by the camera manufacturer and released in 2014 (Available from: <http://www.flir.com>), was used to spatially coincide images using a three-points image registration algorithm; another software with similar features is also available [106].

Image fusion is necessary to precisely perform image registration. Image fusion is the process of combining two or more images into a single image while retaining the important features from the original images [107]. Figure 6.3 shows an example of the process of spatially coinciding a pair of thermal and digital images. Figure 6.3(A), shows an arbitrarily unfused image (contains information from the original thermal and digital images) before the registration process was applied and the mismatch in the image is evident. Figure 6.3(B) illustrates both the thermal and the digital image showing the process of choosing three reference points in both images to perform image registration; coins were used to provide for additional distinct points in both images. Figure 6.3(C) illustrates the fused image and one can clearly notice that the mismatch shown in Figure 6.3(A) has been overcome. Figure 6.4 illustrates exaggerated views for the reference points used in Figure 6.3. Once the step of registration is completed, thermal signatures of all pixels in the thermal image were transferred to Excel and saved as a comma separated values (CSV) format for further processing.

On the other hand, digital images were transformed to binary-coded images by changing the color temperature of the delaminated areas using a suitable image manipulation software. For example, Figure 6.5(A) shows an arbitrary original digital image whereas, Figure 6.5(B) shows the same image after changing the color temperature of the delaminated spots. Then, a code of “0” was assigned to sound pixels and of “1” for delaminated pixels. This transformation is necessary to allow the developed MATLAB script (discussed next) to read color temperature of each pixel of the digital image as either 0 or 1 (i.e., sound pixel or delaminated pixel, respectively). Figure 6.5(C) shows the same image as read by MATLAB script.

Figure 6.3: Process of image registration (A) before, (B) during, and (C) after registration.

Figure 6.4: Exaggerated views of the reference points of of Figure 4-33.

Figure 6.5: Transformation of digital image to a binary-coded image: (A) Original, (B) changing the color temperature, and (C) as read by MATLAB script.

In an effort to automate the subsequent analysis steps, A MATLAB script was developed to perform the following tasks (See "Appendix A"):

- (a) Resize the resolution of the digital images to match the resolution of the corresponding thermal images as these are of different number of pixels as shown in Table 6.1. This step is mandatory to allow for a pixel-by-pixel analysis.
- (b) Calculate the thermal contrast between each pixel and a "well-known" sound

pixel using equation 2-12. The temperature of the well-known sound pixel was chosen as the median of the temperature of all negative pixels. However, any alternative practical value can be set instead.

- (c) Construct the histograms and the probability density functions (PDFs) of the thermal contrast (Δt) of pixels over the actually delaminated (i.e., positive) and sound (i.e., negative) areas of concrete. This step is optional in this stage of the analysis, but the construction of the histograms and PDFs would be beneficial to estimate, at least visually, the amount of overlap between the results of the actually positive and negative cases.
- (d) Transform thermal images to a binary-coded images of “0” and “1” by setting a threshold value for the thermal contrast (Δt). A pixel having Δt larger than the pre-set threshold value will be classified as “1” and a pixel having Δt less than the threshold value will be classified as “0.”
- (e) Estimate the contingency table indices by comparing each pixel from the thermal image with the corresponding pixel from the digital image.
- (f) Calculate the resulting pair of TPR and FPR for the given threshold setting.

Figure 6.6 illustrates the process of changing an arbitrary digital image and the corresponding thermal image into binary-coded format and the resulting correct and incorrect decisions (i.e., TP, TN, FP, and FN) for a single threshold setting of the thermal contrast. After the pixel-by-pixel comparison for one particular threshold setting is completed and the resulting TPR and the FPR were calculated, the script will assign another threshold value within a practical range of application which might be determined from the histograms. The steps from (d) to (f) in the procedure above are repeated until all practical threshold settings are investigated. As many threshold settings as desired may be investigated in the analysis. Extreme threshold settings that yield both TPR and FPR to be 0 and 1 must be included also. Finally, as shown in Figure 6.7, which illustrates the analysis work-flow, the

script will then use all the resulting TPR and FPR pairs to construct the ROC curve and calculate the area under the curve (AUC).

Figure 6.6: Transformation of digital and thermal images into binary-coded format: (A) A digital image, (B) the digital image as in a binary-coded format, and (C) the corresponding thermal image by setting a specific threshold for the thermal contrast.

In an effort to accurately simulate the actual condition where thermal data from dirt, debris, tire residue, paint, etc. affects temperature distribution of the deck's surface, the MATLAB script has included a special function to filter out any irrelevant thermal data results from such factors where these are normally in form of outliers (i.e., more/less than three standard deviation from the dataset mean). Since these factors are often of irregular shape and difficult to be precisely detected and excluded from both thermal and digital images, the outliers data has been replaced by the median of the negative or positive pixels (i.e., halfway value when the data is arranged in order). When these irrelevant data come from artifacts located over sound concrete, which is mostly the case since delamination are localized, the

Figure 6.7: Work-flow of the proposed analysis method.

irrelevant data will be replaced by the median of sound pixels temperatures. On the other hand, if the majority of a captured deck's area is delaminated, the irrelevant data will be replaced by the median of the delaminated pixels temperatures. This function is also useful when testing a member of composite material such as concrete-steel or concrete with asphalt patches where data from irrelevant material must be filtered out for an accurate analysis. For example, Figure 6.8(A) shows an arbitrary pair of digital and thermal image before filtering out the thermal data resulting from placing a foil tape on the surface of slab. In this figure, one

can visually recognize the foil tape in the digital image, and thermal data from the tape in the thermal image is much lower than the data of the sound concrete (approximately 92 degree as compared to about 98 degree). Because the tape was placed over a sound area, data resulting from the tape has assigned a value of “0” (i.e., negative) in the digital image and assigned the median value of the sound thermal data in the thermal image, Figure 6.8(B). However, this function can only be executed when irregular thermal data presents in the captured image and does not execute otherwise.

As has been discussed earlier, it would be very beneficial to draw histograms for the test results over the sound and delaminated locations. Figure 6.9 (A and B) illustrates, respectively, the histograms of the thermal contrasts (in Fahrenheit degrees) of the actually negative and positive pixels before and after filtering out irrelevant data for the image shown in Figure 6.8. It can be seen in this figure that the irrelevant data of thermal contrast less than -4 Fahrenheit that is a result of the placement of the foil tape has successfully been filtered out using the impeded function in the developed MATLAB script.

Figure 6.8: Filtering out irrelevant thermal data resulting from out-of-control objects.

Figure 6.9: Distributions of the thermal contrast of the actually negative and positive pixels before (A) and after (B) filtering out the irrelevant data of the image shown in the previous figure.

6.4 Results of ROC Analysis

As stated earlier, two groups of 12 and 10 images were individually analyzed using the proposed analysis method. Images of the first group were all from different locations of the first fabricated slab. Images of the second group were all from in-service bridge decks (namely A0295, A2111 and A2112 bridges). “Appendix B” contains information of the images on both groups, including thermal and digital images, histograms, and the ROC curve of each individual pair of images. In order to represent the reliability of IRT technology through a unique ROC curve, the individual ROC curves of each group were vertically averaged. Vertical averaging takes the average of PODs at each fixed POFA [108]. Since the size of both datasets are relatively small, the 95% confidence band of POD averaging is also calculated for each group using “student-t distribution.” Figure 6.10 (A & B)

shows the ROC curves of the fabricated slab images and the averaged ROC curve with its 95% confidence limits, respectively. Figures 6.11 (A & B) shows the ROC curves of the in-service bridge decks images and the averaged ROC curve with the 95% confidence limits, respectively.

(A) ROC curves of 12 images

(B) Vertical averaging with its 95% confidence

Figure 6.10: Results of ROC analysis of the fabricated slab I thermal images.

Since the ROC curves do not show the desired indices of POD and POFA as functions of the decision threshold used in the analysis, it would be beneficial as discussed previously to establish such curves that show how POD and POFA

(A) ROC curves of 10 images

(B) Vertical averaging with its 95% confidence

Figure 6.11: Results of ROC analysis of the in-service bridge decks thermal images.

correlate with the decision threshold. In order to do so, data obtained from the averaged ROC curve was used in combination with decision threshold to draw both POD and POFA as functions of the decision threshold for both groups of images. Figures 6.12 and 6.13 show such graphs for the images of the fabricated slab and the in-service bridge decks. Using these figures can allow analysts to precisely determine the expected POD and POFA by setting any individual thermal contrast to define sub-surface deterioration, such as delamination. For example, in Figure 6.12 when a thermal contrast of 0 °F is set, POD and POFA can be estimated as 0.92 and 0.51 respectively. Increasing the thermal contrast to 2 °F, can decrease POFA to about 0.05, but also decreases the POD to 0.48. Final dichotomization of the result should be left to the decision-maker.

Figure 6.12: POD and POFA as functions of the decision threshold (thermal contrast between deteriorated and sound locations) of the fabricated slab images.

The previous set of figures and the figures of Appendix B illustrate that IRT is approximately equally reliable for the detection of synthetic and real delamination damage in R.C. plate-like members, such as slabs and bridge decks. However, more images are recommended to be tested in order to narrow the confidence band in the estimation of the area under the averaged ROC curve for both groups. In order to do this since the analysis of the images depends on the thermal contrast, not

Figure 6.13: POD and POFA as functions of the decision threshold (thermal contrast between deteriorated and sound locations) of the in-service bridge decks images.

the actual temperature of the pixels, and the simulation should match the actual condition of an in-service bridge deck, data from both groups were pooled, and the same analysis for vertical averaging was applied for the resulting 22 pooled images. Figures 6.14 and 6.15 show, respectively, the averaged ROC curve for all images and the desired indices of POD and POFA as functions of the thermal contrast between damaged and sound concrete.

Details of the AUC estimations and the upper and lower confidence intervals of images of both groups are shown in Table 6.2. The pooled estimation of the AUC and the confidence intervals of all 22 images are also shown in same table.

Table 6.2: Statistics of AUC Estimation of IRT Images.

Group	Size	Mean	St. Dev.	95% LCL	95% UCL	Min.	Max.
Fabricated Slab	12	0.871	0.090	0.813	0.928	0.707	0.979
In-service Decks	10	0.834	0.079	0.777	0.890	0.648	0.892
All Images	22	0.854	0.086	0.816	0.892	0.648	0.979

The standard deviation of the fabricated slab is slightly larger than the standard deviation of the in-service decks which might have resulted from edge effects of

Figure 6.14: Averaged ROC curve and the upper and lower 95% confidence of all 22 images.

Figure 6.15: POD and POFA as functions of the decision threshold (thermal contrast between deteriorated and sound locations) of all 22 images.

some images. Regardless, as can be seen in Figures 6.14 and 6.15 and Table 6.2, IRT technology has an excellent ability (mean AUC of all 22 images = 0.854) to detecting delamination in both laboratory and in-service case studies, and that the best threshold range for the thermal contrast is somewhat between 0.5-1.5 °F as this range can provide for larger POD with lower POFA.

In the next chapter, the detection of delamination damage using ground penetrating radar (GPR) technology will be discussed in details.

Chapter 7

ANALYSIS AND RESULTS OF GROUND PENETRATING RADAR (GPR) TECHNOLOGY

7.1 Introduction

Ground Penetrating Radar (GPR) uses electromagnetic (EM) wave propagation to image the subsurface condition of the targeted material. Reflections occurs when the propagated EM pulses encounter a medium (interface) with dissimilar dielectric proprieties than the transporting medium. Successive arrivals of the reflected waves are monitored through a receiver or an array of receivers located close to the excitation source. Analysis of the arrival times and amplitudes of the reflected waves can allow for determining the subsurface condition of the tested member.

Elevated levels of moisture and chloride ions in the concrete resulting from rain, snow, and the applications of de-icing chemicals during winter sessions can cause localized variations in the dielectric properties of the concrete. Chlorides and moisture are also two of the key factors that contribute to the corrosion of the steel reinforcement [109,110]. Areas of elevated chlorides and moisture typically result in greater attenuation of the EM wave than areas of normal levels. Decreased velocity of the EM wave may also be exhibited in these areas. These localized variations in the reflected signal amplitude form the basis of the data analysis described in the ASTM standard, with areas of greater signal attenuation being correlated with deteriorated concrete [80].

As discussed earlier in chapter two, after several years of implementation, confusion remains regarding the actual reliability of GPR for the detection of delamination damage in concrete. To clarify this confusion, multiple case studies, ranging from fabricated slab with some synthetic delamination targets of different

sizes to three concrete decks of in-service highway bridges were evaluated using ground-coupled GPR system. The next sections will be devoted to discuss the characteristics of the chosen case studies, GPR testing parameters and setup, analysis procedure to extract the reflection amplitudes, and the ROC analysis of GPR data obtained using the procedure described in the current ASTM standard that employs signal reflections from uppermost steel bars.

7.2 Testing Parameters and Data Analysis

All GPR tests were performed using a ground-coupled GPR system consisting of an SIR-3000 data acquisition controller device, a 1600 Hz antenna, and encoder to measure the traveled distance. The following data acquisition parameters were set in all case studies as recommended by the instruments manufacturer and previous work for shallow applications:

- Scan/Second = 120
- Scan/Unit (ft) = 48
- Samples/Scan = 512
- Scans intervals = 6 in. for laboratory specimen, 2 ft for the in-service bridge decks
- Test direction = Perpendicular to the uppermost steel bar

Once the required data was collected and transferred to a PC, the individual 2D scans for each case study were combined in a single 3D file. Data in the 3D file was then processed for each case study. Data processing included time-zero correction, noise reduction, and migration. Time-zero correction was performed manually by adjusting the vertical profiles to match the actual condition in order to precisely determine the depth of the subsurface targets such as steel bars and delamination. Noise reduction was performed to enhance the quality of the subsurface image.

Migration was performed to collapse the exact position of the steel bars and provide accurate determination of the reflection amplitude, two-way travel time, and depth associated with each embedded steel bar. Figure 7.1(A) illustrates an arbitrary GPR scan before processing. RADAN-7 software, provided by GSSI Inc., was used to process the data for all case studies. The deterioration mapping function available in RADAN 7 software was used to extract the reflection amplitudes of steel bars as shown in Figure 7.1(B). Some manual corrections were needed to process the data. Datasets were then transferred to a third-party software for further analyses and to produce plan-view color-coded images representing the GPR results.

(A) before processing

(B) after processing

Figure 7.1: An arbitrary GPR scan before and after processing.

Sometimes, when the the distribution, i.e., layout, of steel bars is very dense, it

is necessary to stretch the profiles (scans) in order to overcome the existing overlap between rebar reflections for better interpretations. Stretching the scans involves creating additional scans in between the original scans by using the simple average of every two adjacent scans (or the specified number of scans), and therefore, the number of scans per unit distance would be doubled if the stretching scale is set to “double.” For example, Figure 7.2(A) shows an arbitrary profile with raw result where it is difficult to recognize reflections from individual bars. Figure 7.2(B) illustrates stretching the profile by a scale of 2 and one can clearly see the benefits of stretching where steel bars are clearly separated from each other and visualizing horizontal reflectors has been improved.

(A) before stretching

(B) after stretching

Figure 7.2: An arbitrary GPR scan before and after stretching to enhance the quality of GPR data.

One important thing should be discussed before moving ahead to the next sections. Since EM waves attenuate with traveled distance (i.e., depth of propagation),

steel bars of different depths can produce variable results even if they are of the same condition. For example, a steel bar of 2.5 in. deep is expected to produce lower reflection amplitude than a steel bar of 2 in. deep even if they are of the same exact condition. This depth variation if not corrected before interpretation, some areas of the slab would be classified incorrectly as either delaminated or sound. Therefore, depth correction intended to address signal losses and gains due to depth variations of steel reinforcement. Depth correction of the amplitudes due to steel depth variation was completed as recommended by Barnes et al. [72]. The general hypothesis behind the correction procedure listed below is that in any given bridge deck, at least 10% of the deck's area is sound area, and therefore, variation of measurements of the 90th percentile as a basis for correction can be reasonably assumed due to only Chloride and moisture variation. The procedure can be summarized as follows [75]:

1. Plot amplitudes (on the vertical axis) against the associated two-way travel time (TWTT) on the horizontal axis of all steel bars. Exclude any outlying values outside three standard deviation (σ) or three interquartile range (iqr) from the mean of the dataset or the nearest quartile value respectively
2. Section the TWTT of all dataset into short bins of at least 25 data points [104] or 30 data points [75].
3. Determine the amplitudes 90th percentile of each bin; 90th is less sensitive to presence of outliers than the maximum amplitudes [72].
4. Construct the best fitting line between the 90th percentile of each bin and the center of the TWTT of that particular bin.
5. Subtract the intercept and slope from the amplitude of steel bars at each bins.
6. Plot the depth-corrected amplitude versus TWTT of the entire dataset and determine the best-fitting line to confirm the results.

It should be noted here that this procedure it is not considered in the current ASTM standard [80]. Additionally, depth-correcting GPR data is not needed if the steel bars of the tested deck are of approximately the same depth. However, in the next sections, ROC analysis will be individually applied onto bare and depth-corrected GPR data in order estimate the potential improvement, if any, in the reliability of GPR technology for the detection of subsurface delamination.

7.3 Analysis and Results of Fabricated Slab I

As discussed earlier, slab I was fabricated with some air voids of different sizes using an open-cell Styrofoam material of 0.22 in. thick. These air voids simulate the presence of “non-corrosion” delamination which might result from poor production of concrete, carbonation, or from over stressing. Details of the horizontal distribution of the targets are shown in Figure 4.3. As discussed above, GPR testing was performed by means of two scans every foot in the longitudinal direction of the slab, perpendicular to the uppermost steel bars. This fine scan interval was intended to provide for higher resolution data in order to ensure not missing the small delamination targets in the slab.

Once the data has been collected, transferred to a PC machine, compiled in a single 3D file, and successfully processed, the resulting reflection amplitudes of steel reinforcement were contour mapped as shown in Figure 7.3. In this figure, the horizontal and vertical axes represent the lateral dimensions of the slab, and the depth axis represents the processed amplitude values of the reflected signals from the upper steel layer where larger and lower reflection amplitudes are to the left and right of the colored-bar respectively. The exact locations of the delamination targets were superimposed on this figure for visual interpretation purpose.

As discussed earlier, in order to perform ROC analysis, it is advisable to construct the histograms of the test results of the actually negative and positive test points as shown in Figure 7.4. This figure clearly shows that the reflection amplitudes of both groups significantly overlap; any particular threshold setting

Figure 7.3: Contour map of raw GPR data of the fabricated slab.

will result in an approximately similar true and false positive rates. However, this is a subjective conclusion and ROC analysis is needed to examine the adequacy of this conclusion and to provide a quantitative measure of the amount of overlap, and consequently, the exact reliability of the technology.

Figure 7.4: GPR test results of the actually sound and delaminated test points of the fabricated RTF slab.

As is the case with IE testing, multiple threshold settings for the test results, herein the reflection amplitude, were investigated. With each threshold setting,

indices (counts of the true and false indications) of the contingency table, and consequently, the POD and POFA were estimated. Table 7.1 is an example showing the output from five individual threshold settings. As shown in Figure 7.5, the POD and POFA indices resulting from 13 threshold settings were then plotted in the ROC space and the AUC was nonparametrically calculated. Figure 7.6 illustrates POD and POFA as functions of the threshold setting.

Table 7.1: Reliability Indices of GPR Testing of the RTF Fabricated Slab with Five Arbitrary Threshold Settings.

Index	Reflection Amplitude (dB)				
	-30.28	-27.8	-24.8	-21.8	-18.8
TP	0	2	48	111	125
FP	0	17	131	353	395
TN	395	378	264	42	0
FN	125	123	77	14	0
POD	0	0.016	0.384	0.888	1
POFA	0	0.043	0.331	0.893	1

Figure 7.5: ROC curve of raw GPR data of the fabricated RTF slab.

As can be seen in Figures 7.5 and 7.6, at any threshold setting the probability of correctly identifying delamination is very close, if not exactly equal at some

Figure 7.6: POD and POFA of raw GPR data as functions of the reflection amplitude of the RTF fabricated slab.

points, to the probability of incorrectly identifying sound areas as delaminated areas. This scenario clearly indicates the inability of GPR to detect such kind of delamination that might have been formed as a result of some factors other than corrosion of steel reinforcement.

The reliability assessment performed in the previous section was performed considering the raw (uncorrected) reflection amplitudes. In order to perform depth-correction, the procedure discussed above has been applied to the raw data. Figure 7.7 is a scatter plot representing the reflection amplitude versus the TWTT of all data points (raw data). The entire dataset then was sectioned into short bins (≈ 25 data point) of TWTT. The best fitting line through the top 10% of the data in each bin was obtained using regression analysis to find the exact trend of the decreasing amplitude with respect to TWTT, Figure 7.8. The equation of the resulting line was used to adjust the reflection amplitude of all dataset at each TWTT bin. Figure 7.9 shows the depth-corrected amplitudes versus TWTT. Examining Figure 7.9, it can be seen that the depth-related trend of signals has been overcome and the remaining amplitude variation is due to solely the chloride and moisture content in the concrete cover. Figure 7.10, on the other hand, represents the resulting

depth-corrected contour.

Figure 7.7: Reflection amplitude versus TWTT of all dataset of the RTF fabricated slab.

Figure 7.8: Line fitting of the 90th percentile of the reflection amplitudes at each TWTT bin.

ROC analysis of the depth-corrected data was performed as in the previous cases. Figure 7.11 illustrates the ROC curve of the depth-corrected data while Figure 7.12 illustrates the POD and POFA as functions of the depth-corrected amplitudes. A very slight improvement in the detectability criterion was noted

Figure 7.9: Depth-corrected amplitude versus TWTT of all dataset.

Figure 7.10: Depth-corrected contour of GPR data of the RTF fabricated slab.

as the GPR data was corrected; AUC of depth-corrected data was slightly larger than AUC of raw data. However, the overall reliability of GPR for the detection of non-corrosion induced delamination is still relatively low.

Figure 7.11: ROC of depth-corrected GPR data of the RTF fabricated slab.

Figure 7.12: POD and POFA of depth-corrected GPR data as functions of the reflection amplitude of the RTF fabricated slab.

7.4 Analysis and Results of Bridge A0295

For this case study, GPR data was collected, processed, and analyzed in the same manner followed in the RTF fabricated slab with the exception of the scan interval, which was 2ft in the longitudinal direction of the bridge (i.e., parallel to traffic direction). Ground-truth data had been already determined from the combination

of sounding (i.e., chain drag and hammer sounding) and infrared thermography (IRT) to confirm the level of agreement between these technologies in locating delaminated areas. IE testing was also performed on some of different areas of the deck to confirm the results obtained by sounding and IRT. Sounding tests had been performed previously by skilled personnel from MoDOT. Though time consuming, this process of determining the ground-truth by three different NDE technologies was necessary to develop a consensus of the delaminated areas and to provide for a more accurate basis of comparison.

Figure 7.13 shows the contour map of the raw GPR data. Figure 7.14 illustrates the histograms of the test results over the actually sound and delaminated locations. The histograms indicate a significant overlap exists between data of both groups as in the RTF slab case study; this means that GPR was not able to discern areas of delamination from the sound area of the deck. ROC analysis was performed by investigating several threshold settings of GPR raw (before depth-correction) data within the range shown in the previous two figures. Table 4.6 shows the decision indices and the resulting POD and POFA corresponding to five arbitrary threshold settings. Figure 7.15, on the other hand, shows the ROC curve of GPR testing resulting from the investigations of the threshold listed in Table 4.6 and some other settings. The AUC was nonparametrically estimated as 0.475 indicating again that GPR has no ability to distinguish areas of delamination in this bridge deck.

Table 7.2: Reliability Indices of GPR Testing of Bridge A0295 with Five Arbitrary Threshold Settings.

Index	Reflection Amplitude (dB)				
	-27.18	-24	-21	-18	-14
TP	0	31	142	307	341
FP	1	216	2690	4905	5358
TN	5357	5142	2668	453	0
FN	341	310	199	34	0
POD	0	0.09	0.41	0.90	1
POFA	0	0.04	0.50	0.91	1

In order to determine the optimum threshold (i.e., the threshold that maximizes

Figure 7.13: Contour plot of raw GPR data of bridge A0295.

Figure 7.14: GPR test results of the actually sound and delaminated test points of bridge A0295.

the accuracy of the output), POD and POFA were plotted against the threshold setting as shown in Figure 7.16. From this figure, it can be clearly seen that POD and POFA are approximately the same regardless of threshold setting and not any particular threshold setting can produce accurate results.

GPR reflection amplitudes were also depth-corrected to account for the gain and loss of GPR signal due to steel of variable depth and to determine whether depth-correction can improve the reliability of GPR in this case study. The same

Figure 7.15: ROC curve of raw GPR data of A0295 bridge deck.

Figure 7.16: POD and POFA of raw GPR data as functions of the reflection amplitude of bridge A0295.

procedure described earlier was used to perform depth-correction of amplitudes. Figure 7.17 illustrates the reflection amplitudes as function of TWTT of the raw data. On the other hand, Fig 7.18 shows the 90th percentile of data at the center of each TWTT bin and the the resulting best fitting line among these data points. The equation of the line was used to correct the amplitudes of the reflected signals as previously described. Depth-corrected amplitudes of all dataset is shown in

Figure 7.19. The contour map of the depth-corrected amplitudes of all data points is shown in Figure 7.20 where, again, the amplitude variation is supposed to be solely due to chloride and moisture content of concrete cover.

Figure 7.17: Reflection amplitude versus TWTT of all dataset of bridge A0295.

Figure 7.18: Line fitting of the 90th percentile of the reflection amplitudes at each TWTT bin

ROC analysis was also performed onto depth-corrected data to quantify the improvement, if any, in the reliability of GPR technology when data is depth-corrected and to determine whether it is worthy to implement the depth-correction

Figure 7.19: Depth-Corrected amplitudes versus TWTT of all dataset of bridge A0295

Figure 7.20: Depth-Corrected contour of GPR data of bridge A0295

algorithm. In the same fashion, the threshold value necessary to classify test results as positive and negative was varied, and the resulting reliability indices described earlier were calculated with each setting. The resulting POD and POFA pairs of each threshold setting were then plotted against each other in the ROC space as shown in Figure 7.21. Additionally, POD and POFA were plotted against the threshold setting of the depth-corrected amplitudes as illustrated in Figure 7.22.

By examining these figures, one can clearly see that depth-correction did not

Figure 7.21: ROC curve of depth-corrected data of bridge A0295.

Figure 7.22: POD and POFA of depth-corrected GPR data of bridge A0295 as functions of the reflection amplitude setting.

significantly improve the overall reliability of the technology and the capability of GPR for the detection of subsurface delamination is still limited.

As discussed previously, corrosion activity of steel reinforcement normally arises from the intrusion of chlorides and moisture above some threshold values to the steel layer. After a certain amount of time (many factors are involved), corrosion begins and consequently, delamination can form. In order to find out the actual

chloride levels in the in-service bridge decks, determine whether the threshold value for chloride had been reached, and that the corrosion has possibly occurred, it was necessary to estimate the chloride ions concentration in the decks of bridge A0295 and bridge A2111, as well as two other bridge decks in the same geographic area has been estimated. A total of six chloride samples were taken from each deck and then tested by a specialist personnel from MoDOT. Some of the samples were taken from delaminated concrete and others were taken from sound concrete. The concentrations of chloride ions of the tested decks are provided in Table 4.7. As can be seen from this table, chloride concentration in bridge A0295 is lower than the other bridges, and also much lower than the threshold that is required to initiate corrosion. This possibly explains the limited capability of GPR in detecting the delaminated areas in bridge A0295 and the fabricated slab since the RTF fabricated slab is relatively chloride-free.

Table 7.3: Chloride Ion Concentrations of Four Decks in Missouri State.

Structure	Average Cl^- of Six Samples % (ppm ^a)
Bridge 1	0.0905(905)
Bridge 2	0.2238(2238)
Bridge A0295	0.0272(272)
Bridge A2111	0.2011(2011)

^a *Part per Million*

7.5 Analysis and Results of Bridge A2111

As discussed earlier, bridge A2111 is located on Rt. 54 near Fulton, MO. A visual inspection of the bridge deck indicated a number of patches had been applied to repair damage in the deck. Chloride ion results shown in Table 1 indicated a high level of chlorides in the deck; this suggests that the damage in the bridge deck is corrosion damage. The deck was assessed using IRT (hand held camera and ultra time domain (UTD) system) and sounding to provide ground-truth data necessary for reliability assessment of GPR. A delamination map was then

produced combining the results of these two technologies to allow for point-by-point comparison, see Figure 5.21. The contour map showed large areas of delamination in the deck. Some delaminated areas were further evaluated using impact echo (IE) to confirm results of IRT and sounding. One large area of delamination was identified through sounding, but not through IRT. Investigation showed that this area of delamination was saturated with water, resulting in reduced thermal contrast. This was confirmed by drilling into the delamination; saturated cementitious material was evident in the drill waste from the level of the delamination. Figure 7.23 shows images of different NDE testing performed on the bridge deck.

Figure 7.23: Images of different NDE field testing performed on bridge A2111.

GPR testing was performed as previously described using the same parameters used in the other two case studies. GPR scans (data) were acquired in the longitudinal direction of the bridge (with traffic directions as shown in Figure 7.23 D). Data was then processed to determine reflection amplitudes, TWTT, and the horizontal and vertical coordinates of each data point necessary to perform ROC analysis. The raw results of GPR testing of bridge A2111 was contoured as shown in Figure 7.24. On the other hand, Figure 7.25 shows the histograms of the test

results (i.e., reflection amplitudes) over the actually positive and negative locations. Examining this figure reveals that some overlap exists between the distributions, but it is not as extensive as the other two case studies. This relative improvement in the overall detectability of delamination using GPR might have stemmed from the high chloride concentrations in the deck, as can be seen in Table 4.7 above.

Figure 7.24: Contour of raw (depth-uncorrected) GPR data of bridge A2111.

Figure 7.25: GPR test results of the actually sound and delaminated test points of bridge A2111.

ROC analysis was performed by estimating the POD and POFA with several threshold settings within the range of reflection amplitudes shown in the previous two figures. Table 4.8 lists the POD and POFA with five arbitrary threshold settings where one can clearly see that both indices increase or decrease simultaneously with threshold setting. The same conclusion can be reached from the area under the resulting ROC curve (AUC), as shown in Figure 7.26. However, even with

this relative improvement, a limitation still exists in the overall reliability of GPR for the detection of delamination as the AUC is not significantly larger than the reference value of 0.5 (i.e., pure guessing). Figure 7.27 illustrates the POD and POFA as functions of the decision threshold setting. It is clear that there is actually some improvement in the detectability of GPR for this case study, as the two curves are slightly separate from each other which indicates that for most threshold settings, POD is larger than POFA as compared to the previous two case studies.

Table 7.4: Reliability Indices of GPR Testing of Bridge A2111 with Five Arbitrary Threshold Settings.

Index	Reflection Amplitude (dB)				
	-26.15	-22.15	-18.15	-14.15	-10.15
TP	0	127	570	1061	1107
FP	0	405	2069	6326	7495
TN	7495	7090	5426	1169	0
FN	1107	980	537	46	0
POD	0.000	0.115	0.515	0.958	1.000
POFA	0.000	0.054	0.276	0.844	1.000

Figure 7.26: ROC curve of raw GPR data of A2111 bridge deck.

Amplitude values were also depth-corrected to assess the reliability of GPR when only variation in amplitude due to moisture and chloride variations are

Figure 7.27: POD and POFA of raw GPR data of bridge A2111 as functions of the reflection amplitude.

considered. Figure 7.28 shows the reflection amplitude as a function of the TWTT of raw GPR data. On the other hand, Figure 7.29 illustrates the fitting line among the 90th percentile of amplitudes at the center of each TWTT bin, as discussed previously. The resulting equation of the best fitting line was used to depth-correct Amplitude values of all dataset. Figure 7.30 shows the depth-corrected amplitudes versus TWTT and one can clearly see that the obvious trend in the raw data has been overcome. The contour map of the depth-corrected amplitudes is shown in Figure 7.31.

Depth correction of amplitudes resulted in slight changes in the contour map, especially at the center line of the bridge in the traffic direction. In order to confirm whether depth correction has improved the quality of the data, ROC analysis was repeated using the depth-corrected data. Threshold values were ranged from 3.2 to -13.8 dB as shown in Figure 7.31. The resulting ROC curve of depth-corrected data is shown in Figure 7.32 along with the AUC, which indicates an insignificant improvement in the overall reliability (i.e., $AUC = 0.683$ as compared to $AUC = 0.672$). The resulting POD and POFA as functions of the threshold setting (i.e., depth-corrected amplitudes) are shown in Figure 7.33.

Figure 7.28: Reflection amplitude versus TWTT of all dataset of bridge A2111.

Figure 7.29: Line fitting of the 90th percentile of the reflection amplitudes at each TWTT bin.

7.6 Analysis and Results of Bridge A2112

This bridge deck was tested to validate the incipient conclusion that GPR is sensitive to only the presence of corrosive environment. If GPR testing showed a good detection for corrosion induced delamination with minimal false alarm rate, then the conclusion regarding the sensitivity would be validated as the surface of

Figure 7.30: Depth-corrected amplitudes as function of the TWTT of all dataset.

Figure 7.31: Contour map of the depth-corrected amplitudes.

the decks has numerous signs of concrete deterioration. If not, then GPR testing would be of limited ability for the detection of a corrosive environment as well.

As discussed previously, ROC analysis requires comparing the results of the candidate technology with a form of benchmark data. Infrared thermography (IRT) and sounding of the entire bridge deck had been performed previously to provide the ground truth condition necessary for the analysis of reliability. It should be mentioned here that there was a great spatial correlation between results of IRT and sounding with very minimal mismatch. Figure 7.34 shows the actual condition

Figure 7.32: ROC curve of depth-corrected GPR data of bridge A2111.

Figure 7.33: POD and POFA of depth-corrected GPR data of bridge A2111 as functions of the reflection amplitude setting.

with the locations of delamination areas across the surface of the deck are encircled.

GPR testing was performed in a similar way to the other case studies in terms of testing parameter, scan intervals, and direction of scans. Once the data had been collected, it was processed using the same algorithms previously used in the other case studies to get the exact reflection amplitudes of rebar signals so as to perform

Figure 7.34: The actual condition of Bridge A2112 deck.

ROC analysis. Figure 7.35 is the contour map of the raw (depth-uncorrected) GPR signals of steel bars where areas of possible delamination tend to show lower reflection amplitudes (i.e., the right hand side of the color bar), as compared to the reflections from sound concrete areas. Figure 7.36, on the other hand, shows the histograms of test results over the actually sound and delaminated areas. As expected, the histograms of both groups overlap, indicating a non-perfect reliability scenario.

Figure 7.35: A contour map of the depth-uncorrected GPR signals of the deck of A2112 bridge.

In order to perform ROC analysis, multiple threshold values within the range of reflection amplitudes, shown in Figures 7.35 or 7.36, have been investigated and the contingency table of each setting has been defined. Once the full range of threshold setting had been investigated, ROC curve was constructed and the area under the curve (AUC) was estimated accordingly. Table 7.5 gives an example of the estimation of POD and POFA indices corresponding to some arbitrary threshold

Figure 7.36: Results of GPR testing over the actually sound and delaminated concrete areas.

settings. A full range of POD and POFA indices can be seen in the ROC curve in Figure 7.37. In order to be consistent with other case studies, the area under the curve (AUC) was non-parametrically estimated using the finite difference method. AUC was estimated to be 0.751, indicating an improvement in the detection and false alarm rates, as compared to the other case studies. This improvement in the detection rate might be due to the fact that delamination targets in the deck are corrosion induced delaminations that are a result of high chloride and moisture concentrations as was thought.

Table 7.5: Reliability Indices of GPR Testing of Bridge A2112 with Five arbitrary Threshold Settings.

Index	Reflection Amplitude (dB)				
	-32	-25	-20	-15	-10
TP	0	60	260	431	553
FP	0	128	755	3228	7126
TN	7126	6998	6371	3898	0
FN	553	493	293	122	0
POD	0	0.108	0.470	0.779	1
POFA	0	0.018	0.106	0.453	1

In an attempt to combine POD and POFA in one graph as functions of the

Figure 7.37: ROC curve of depth-uncorrected GPR data of Bridge A2112.

decision threshold, Figure 7.38 was constructed. This figure, in addition to Figure 7.37, can be used to determine the optimum threshold setting that might maximize the accuracy of the output. According to this figure, the optimum threshold setting can be (-17 dB). Larger amplitude settings can increase the POFA at a much higher rate than POD, while lower amplitudes can decrease POD at a higher rate than POFA.

Figure 7.38: POD and POFA of depth-uncorrected GPR data of bridge A2112 as functions of the reflection amplitude setting.

Although it has been shown that depth correction did not improve the quality of GPR data from other case studies, it was decided that depth correction should be performed on the GPR signal of this deck since the original data has been more reliable than other case studies and implementing depth correction might improve the quality of data even more. As has been discussed earlier, depth correction involves plotting reflection amplitudes versus TWTT of each data point, Figure 7.39, and then considering only the 90th percentile of amplitudes at each TWTT bin center, Figure 7.40. The equation of the best fitting line is used to correct the raw GPR data. Figures 7.41 and 7.42 show the depth corrected amplitudes versus TWTT and the depth corrected contour of GPR data.

Figure 7.39: Amplitude versus TWTT of GPR data of Bridge A2112.

Visual examination of the contour maps before and after depth correction of GPR signal indicates no significant improvement in the overall data has been achieved, except that the entire dataset has been normalized to an apparent value. However, consideration of only a visual comparison would be highly subjective. In order to tell exactly whether depth correction had really improved the overall reliability and to what extent if it did, ROC analysis was performed on depth-corrected data. The same procedure of ROC analysis was performed by considering the estimation of POD and POFA at multiple depth-corrected signal amplitudes

Figure 7.40: The best fitting line over the 90th percentile data.

Figure 7.41: Amplitude versus TWTT of depth-corrected data of Bridge A2112.

and AUC was then calculated as has been done previously. Figure 7.43 shows the depth-corrected ROC curve of GPR data along with the estimation of AUC. A slight improvement in the overall reliability has been achieved as the value of AUC has increased from 0.751 to 0.777.

Finally, POD and POFA has been plotted versus the depth-corrected signals in order to find the optimum setting that can optimize the output. Figure 7.44 illustrates POD and POFA as functions of the decision threshold. As has been

Figure 7.42: Depth-corrected contour of GPR data of Bridge A2112.

discussed previously, the final decision regarding the optimum threshold setting that can be used to classify test results depends on many factors beside the accuracy of the output. In some circumstances it can be set away from the optimum criteria in order to get maximum probability of detection or minimum probability of false alarm when the consequences of these outputs are of overriding importance. Cost-benefit analysis of diagnostic output can play an important role in the final decision as well.

Figure 7.43: ROC curve of depth-corrected data of bridge A2112.

From the previous discussion, it appears that GPR technology has a limited

ability to detect a physical damage such as delamination as compared to the IE and IRT that showed an excellent detection with minimal false calls. The results so far stand in contrast to some previous research that has suggested GPR as a good tool for the detection of subsurface delamination. However, GPR has the potential to detect the corrosive environment (the earlier symptoms of concrete deterioration), specifically moisture and chloride concentrations in concrete. Such important data offered by GPR technology can be of great importance to transportation agencies and bridge owners to prioritize maintenance plans and schedule inspection intervals of in-service bridge decks.

Figure 7.44: POD and POFA of depth-corrected GPR data as function of the reflection amplitude of Bridge A2112.

Chapter 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Introduction

Bridge infrastructures are the backbone of nations transportation systems. The aggressive environment and increasing volume of traffic loading that daily pass over these infrastructures make them highly prone to deterioration. Deterioration can take several forms such as degradation of concrete quality, surface cracking, steel corrosion, and formation of fracture planes (e.g., delamination). These deterioration mechanisms can promote each other if not detected and fixed as soon as they are initiated.

Delamination is recognized as one of the most critical damage modes that most bridge decks suffer from. As was discussed in the previous chapters, delamination can be formed either as a result of steel corrosion, improper concrete pouring, freeze-thaw cycling, or as a result of overloading. Regardless, development of internal delamination can reduce the load carrying capacity of bridge decks, and consequently, the potential of deck failure can be high if the internal delamination had not been detected and repaired.

In the previous chapters, the reliability of the most widely nondestructive evaluation technologies used for the assessment of integrity of concrete bridge decks was evaluated. These technologies were impact echo (IE), infrared thermography (IRT), and ground penetrating radar (GPR). The purpose of reliability evaluation was to determine to what extent these technologies can be safely, accurately, and reliably used for condition assessment of concrete bridge decks with respect to the delamination detection and also to find out which NDE technology can provide lower risk of decision for optimum maintenance and repair plans.

Historically, the reliability of NDE technologies has been described and evaluated

through probability of detection (POD) analysis by estimating the true detection rate at some specific defect sizes. Among the drawbacks of this analysis method is that it cannot provide an explicit estimation of the false alarm rate and also it can be very difficult to be used for the detection of defects of 3D nature such as delamination. In addition, the method requires a high number of defects to be tested in order to provide confident results.

In this research, an alternative analysis method was used to assess the reliability of the aforementioned NDE technologies. This method is known as receiver operating characteristics (ROC) analysis. As will be discussed later in this chapter, ROC analysis can be efficiently used for the purpose of this research due to its ability to provide for a quantitative reliability index, which allows for the comparison of one technology to another, considering all possible decision outputs such as correct and incorrect testing output.

In this chapter, the summary of the findings of the research is presented in detail and the recommendation for future work is made.

8.2 Conclusions

In this study, two reliability analysis methods were used to assess the performance of three of the most widely used NDE technologies for condition assessment of bridge decks. In contrast to the analysis method used historically for this purpose, these methods can provide quantitative indices necessary to describe the overall reliability of diagnostic systems, considering all possible decision outputs such as true positive, false positive, true negative, and false negative rates. The main findings of the research were as follow:

8.2.1 Impact Echo (IE)

Impact echo (IE) technology was found to be very reliable and has an extraordinary ability to detect delamination damage in concrete bridge decks when the delamination is less than 4 inches deep and the area of the delaminated portion is

larger than 9 in.². This extraordinary ability can be verified by the high detection rate and the very low probability of false alarm rate. Consequently, areas under the resulting ROC curves (AUC) were larger than 90% in almost all case studies indicating the IE can be very reliable in such circumstances.

This ability of IE technology was diminished for the delamination targets of 3 x 3 in. lateral dimensions in the second fabricated slab. However, it is believed that delamination targets of this small size will be of little impact on the maintenance and repair plans even if they had been detected, providing these are not located adjacent to each other.

Since reinforcing steel are placed relatively at depth of 2-3 inches from the surface of concrete, and due to the fact that most delaminations can be formed as a result of steel corrosion, it was found that there exists a wide range of threshold settings can be used to accurately distinguish sound from delaminated concrete with very low probability of false alarm (POFA). In most cases, threshold settings within the range of 6000 Hz to 9000 Hz can be used to reliably recognize sound concrete from delaminated concrete (i.e., frequencies of the reflected P-waves less than 6000 Hz can be accurately assigned to delaminated concrete while frequencies larger than 9000 Hz can be assigned to sound concrete). According to this scenario, inaccurate P-waves' speed assessment can have little influence on the the final decisions. In contrast, deep delamination (i.e., larger than 5 inches deep) can have very close frequency range to the thickness frequency, making it difficult to distinguish sound concrete from deep delaminated concrete, especially when P-wave speed had not been obtained at testing locations.

Steel reinforcement density has not explicitly been investigated. However, the analysis of IE data has shown no effect of steel bars density on the output of IE. This conclusion was reached as the density of steel in the first fabricated slab was much less than the density of steel in the second fabricated slab and also less than the steel density in both of the in-service bridge decks. This, in fact, suggests that IE technology can be a good option regardless of the steel density in concrete.

Edge effect (i.e., testing in close proximity of curbs or edge of decks) has shown to have no influence on the reliability of IE testing, especially when the test is performed at a distance equal to or larger than the deck thickness from the edge of deck or the curb.

However, according to the point-wise nature of IE testing, acquiring and analyzing IE data can be very time consuming, as well as requiring a considerable level of knowledge of the physical background of the technology. Difficulties in IE testing and data analysis can be summarized as follows:

- The estimation of the speed of P-waves is, by itself, a time consuming process and requires the generation of waves within practical range of frequencies using impactors of multiple sizes.
- Testing might be repeated several times until a stable reflected signal is reached. After this point, the test is recommended to be repeated for at least three times to confirm the results and consider the average.
- Presence of dense gravel layer in concrete cover can influence the results; speed of stress waves in gravel is different from concrete.
- Signals (echoes) obtained from testing can be very difficult to be accurately interpreted without a proper smoothing. Applying smoothing algorithms can further complicates and lengthens the analysis process.
- Inconsistent concrete quality can influence the determination of the exact depth of delamination if the speed of P-waves has not been estimated at the exact testing location.

8.2.2 Infrared Thermography (IRT)

A novel analysis method was developed to analyze data of infrared thermography. The developed analysis method relies on a pixel-by-pixel comparison of thermal contrasts with the actual condition of each pixel using image processing algorithms.

The analysis method was used to analyze data from the first fabricated slab (with shallow artificial delamination targets of different size) and in-service bridge decks containing delamination targets of irregular shapes.

In terms of overall reliability, infrared thermography (IRT) has also shown a very good ability to detect delamination targets in both fabricated slabs and in-service bridge decks. The average AUC indices of a total of twelve thermal images of the fabricated slab and ten images of the in-service bridge decks were about 87% and 83%, respectively. The average AUC of all combined twenty-two thermal images was determined to be equal to 85%, indicating the great ability of infrared thermography to detect delamination damage.

Considering the 95% confidence intervals shown in Table 6.2, the results indicate that if two pixels were arbitrarily chosen, one above an area of delamination and the other above sound concrete, there would be no less than an 80% probability (with 95% confidence) that these pixels would be correctly identified as positive and negative, respectively. This high AUC reveals IRT's ability to discern the areas of shallow delamination from the areas of sound concrete in concrete bridge decks. This value may be affected by several experimental factors as follows:

- Weather conditions: Data was captured in the afternoon during good weather, resulting in “good” IRT data. Different weather conditions may provide different results. For example, accumulation of water in concrete bores resulting from rain or ice can create internal artifacts that mask the temperature distribution and might consequently result in a very large number of incorrect positive or negative decisions.
- Edge effects: POFA was adversely affected by edge effects in the slab. This effect would be reduced in a large concrete bridge deck. If thermal images are to be subjectively interpreted, inspectors need to be careful from calling such regions delaminated regions unless they show a very large thermal contrast. However, in a large concrete deck there may be debris and color variations that produce the same effect. Results indicated similar AUC values for each

case.

The above quantitative indices can provide for the information needed by bridge owners and transportation agencies regarding the actual reliability of Infrared thermography as compared to the previous subjective assessment of “good versus bad” image analysis basis.

Infrared thermography is an area-wise testing, therefore, it can be an excellent alternative to the very time consuming IE technology with minimal loss of overall reliability. The benefit of using IRT instead of IE is the time saved in the closure of the tested structure. For example, in bridge structures, closure time can be among the main factors that prioritize the use of NDE technologies.

Since the analysis of thermal data relies on the the estimation of thermal contrast between each pixel and a “well-known” sound pixel, it was found for both image groups that a thermal contrast within the range of 0.5-1.5 °F can provide for the most optimum results that can increase the accuracy of the classification. Threshold setting within this range can provide for maximum probability of detection (POD) at minimum probability of false alarm (POFA).

8.2.3 Ground Penetrating Radar (GPR)

Data obtained from GPR testing using 1600 Hz ground-coupled GPR antenna have been analyzed twice for every case study discussed so far (i.e., without and with depth-correction necessary to remove gains and losses of the reflected signals due to inconsistent steel depth). In both analyses, GPR has shown a limited ability to detect delamination damage. This can be clearly confirmed by the resulted AUC indices. In fact, in the RTF slab and bridge A0295 case studies where delamination targets were “non-corrosion induced,” GPR has proven to be nothing more than random guessing (possibility of positive or negative is about 50% for all data points).

For “corrosion-induced delamination” of bridge decks A2111 and A2112, the reliability of GPR has increased to about 67% to 77.7%, respectively. This simply

means that at this reliability level, there at least 22% of the data points have been classified incorrectly. The relative improvement in the detection rate of the in-service bridge decks suggests that GPR is more sensitive to the presence of corrosive environment (the early symptoms of deterioration of concrete), but not the presence of physical damage (i.e., delamination) resulting from the corrosive environments.

Implementation of depth-correction algorithms has had very little effect on the overall reliability of GPR technology in all case studies. For example, reviewing AUC indices of all case studies with and without the implementation of depth-corrections, no significant improvement in the overall reliability was determined. However, this can be because the tested bridges have minimal steel depth variability; depth-correction might result in significant difference in bridges with high steel depth variability.

Data analyzed so far from the aforementioned cases studies show that there is no unique threshold setting that can be set to optimize the decision output of all case studies, and the optimum threshold setting can be site-specific. In other words, the reflection amplitude of steel bars might be different from structure to structure as the properties of concrete, depth and size of reinforcement, and weather condition prior to testing might all result in a variable reflection amplitude. In addition, intensity of reinforcement might influence the final results.

The previous conclusions stand in contrast to some of the previous work that clearly suggest GPR as an effective tool for the detection and localization of delamination. After almost four decades of implementation, confusion remains regarding the actual reliability of GPR for the detection of physical damage such as delamination. However, data obtained from this research so far suggests that GPR technology has a limited ability to detect delamination by itself unless the delamination is formed as a result of corrosion of steel reinforcement.

In terms of speed of the test, GPR testing is much faster than IE testing according to its line-wise nature and also the ability to transmit and receive EM

signals at a very fast rate. For example, testing bridge decks of about 180 ft long and 44 ft wide took the research team about three hours to complete the test using two feet intervals between individual scans. Using air-coupled GPR antenna can accelerate the testing even further as compared to the used ground-coupled systems.

However, data analysis can be very time consuming, especially when the depth-correction algorithm is implemented in the analysis. It should be noted that the software provided by the GPR system's manufacturer that was used in the analysis has some limitations regarding the detection of reflections from actual steel bars, and therefore, manual adjustment was performed in order to obtain the actual reflection amplitudes from steel bars.

8.2.4 ROC Analysis

ROC analysis was found as an excellent alternative to the previously discussed POD analysis for the purpose of analyzing the reliability of NDE data. This analysis method is being used extensively in other disciplines, but has been very rarely used in the field of nondestructive evaluation. The strengths of this analysis method are that it can explicitly consider all possible decision output considering a wide range of threshold settings, and also, produce a quantitative index that can be used to describe the reliability of NDE technologies. In addition, the method can provide for visual assistance for the trade-off between the possible decision output as the threshold setting is changed. Therefore, this analysis method can fulfill the requirements of departments of transportation and bridge owners in the most efficient and cost effective manner

The other advantage of using ROC analysis is that it is simple, but comprehensive, and can be easily understood by the civil engineering community and implemented for the purpose of assessing the reliability of the existing or newly developed NDE technologies.

8.3 Overall Summary

The overall results of ROC analysis of the reliability of the candidate NDE technologies investigated in this research are summarized in Table 8.1. The Table also shows the recommended threshold settings of each technology in each case study discussed in this research.

Table 8.1: Summary of ROC Reliability Analysis of the NDE Technologies Investigated in This Research

Technology	Case Study		AUC	Optimum Threshold
IE	Slab I		0.99	7500-9250 Hz
	Slab II, 3 x 3 in. ²	D = 2 in.	0.67	9350 Hz
		D = 4 in.	0.46	—
		D = 5 in.	0.26	—
	Slab II, 6 x 6 in. ²	D = 2 in.	0.98	6700-8100 Hz
		D = 4 in.	0.99	7000-8200 Hz
		D = 5 in.	0.83	11000-12000 Hz
	Slab II, 9 x 9 in. ²	D = 2 in.	1	4500-8500 Hz
		D = 4 in.	1	5500-9000 Hz
		D = 5 in.	0.99	11250 Hz
	Bridge A0295		0.975	8000-8500 Hz
	Bridge A2111		0.946	6500-7000 Hz
IRT	Slab I	12 images	0.87	0.75-1.5 F ^o
	In-service Decks	10 images	0.83	0.50-1.0 F ^o
GPR	Slab I	Uncorrected	0.508	—
		Corrected	0.56	—
	Bridge A0295	Uncorrected	0.475	—
		Corrected	0.5	—
	Bridge A2111	Uncorrected	0.672	—
		Corrected	0.683	—
	Bridge A2112	Uncorrected	0.751	-17 dB
		Corrected	0.777	-4 dB

8.4 Recommendations

It is recommended that further research be undertaken in the following areas:

- It would be interesting to use the same analysis methods to assess the reliability of other NDE technologies and for other deterioration types. In this way, a reliability chart could be established that could be of extreme benefit to bridge owners and transportation agencies so that they can consult to prioritize the available resources for bridge evaluation.
- In order to establish a more reasonable estimate of the AUC with a stricter confidence interval, a larger sample size is recommended to be considered for the analysis of IRT data.
- Because most of the area under an ROC curve comes from regions of impractical threshold settings, it is recommended to use the data of this research to compare one NDE technology to another using partial area under the curve within practical range of application. For example, instead of calculating the AUC in Figure 6.10 considering threshold settings of the thermal contrast from -14 to 14 degree, it is suggested that the reliability of IRT technology to be estimated based on practical contrast settings (e.g., from 3 to -3 degree).
- As is the case with IRT technology, it is recommended that GPR data is analyzed, considering the contrast in the reflection amplitude rather than the amplitude values themselves in order to determine a unique practical range of threshold settings that can be used to increase the accuracy of the technology.
- Further research could also be conducted to validate the effectiveness of implementation of depth-correction algorithm of GPR data by analyzing data from a larger number of bridge decks.

Bibliography

- [1] Glenn A. Washer. Developments for the non-destructive evaluation of highway bridges in the USA. *NDT & E International*, 31(4):245–249, 1998.
- [2] ASTM-D4580-86. Standard practice for measuring delaminations in concrete bridge decks by sounding. 1986 (Reapproved 1997).
- [3] MIL-HDBK-1823A. Nondestructive evaluation system reliability assessment. 2009.
- [4] Mary Sansalone and William B. Streett. *Impact-echo non-destructive evaluation of concrete and masonry*. Bullbrier Press, Ithaca, N.Y., 1997.
- [5] Bryan D Olin and William Q Meeker. Applications of statistical methods to nondestructive evaluation. *Technometrics*, 38(2):95–112, 1996.
- [6] J Hellier Charles. Handbook of nondestructive evaluation, 2001.
- [7] Luca Gandossi and Charles Annis. Probability of detection curves statistical best-practices. 2010.
- [8] Ripudaman Singh. Three decades of NDI reliability assessment. *Contract*, 41608(99-C):404, 2000.
- [9] url = <http://www.nrmca.org/> type = Electronic Article NRMCA, title = Concrete in Practice- What, why & how?
- [10] Nenad Gucunski, Arezoo Imani, Francisco Romero, Soheil Nazarian, Deren Yuan, Herbert Wiggensauser, Parisa Shokouhi, Alexander Taffe, and Doria Kutrubes. Nondestructive testing to identify concrete bridge deck deterioration. Technical report, 2013.
- [11] Tri Huu Miller. *Nondestructive inspection of corrosion and delamination at the concrete-steel reinforcement interface*. Thesis, 2010.

- [12] NJ Carino. Methods to evaluate corrosion of reinforcement. *Handbook on Nondestructive Testing of Concrete*, pages 11–1, 2004.
- [13] TRAINING COURSE SERIES No.17. Guidebook on non-destructive testing of concrete structures, 2002.
- [14] Rilya Rumbayan and Glenn A. Washer. Modeling of environmental effects on thermal detection of subsurface damage for concrete bridges. 2013.
- [15] Elena Redaelli, Luca Bertolini, and Bernhard Elsener. *Corrosion of Steel in Concrete : Prevention, Diagnosis, Repair (2nd Edition)*. John Wiley & Sons, Somerset, NJ, USA, 2013.
- [16] Institute American Concrete. *ACI manual of concrete inspection*, volume SP-2. American Concrete Institute, Farmington, MI, 2007.
- [17] Yingshu Yuan, Yongsheng Ji, and Surendra P Shah. Comparison of two accelerated corrosion techniques for concrete structures. *ACI Structural Journal*, 104(3), 2007.
- [18] Portlan Cement Association (PCA). Types and causes of concrete deterioration, 2002.
- [19] AG Davis, F Ansari, RD Gaynor, KM Lozen, TJ Rowe, H Caratin, FD Heidbrink, VM Malhotra, BP Simons, and NJ Carino. Nondestructive test methods for evaluation of concrete in structures. *American Concrete Institute, ACI*, 228, 1998.
- [20] Nicholas J CARINO. Impact echo: The fundamentals. 2015.
- [21] V Mohan Malhotra and Nicholas J Carino. *Handbook on Nondestructive Testing of Concrete Second Edition*. CRC press, 2003.
- [22] Rasoul Moradi, Michael L McCoy, and Hamid M Lankarani. Impact analysis of mechanical systems using stress wave propagation methodology.

- [23] N. J. Carino, Standards National Institute of, and Technology. Impact-echo method an overview. 2001.
- [24] M Sansalone and NJ Carino. Impact-echo: A method for flaw detection in concrete using transient stress waves', NBSIR 86-3452, national bureau of standards, washington, DC. Report, PB 87-10444/AS (National Technical Information Service, Springfield, MA, 1986), 1986.
- [25] Alexander Gibson and John S Popovics. Lamb wave basis for impact-echo method analysis. *Journal of Engineering mechanics*, 131(4):438–443, 2005.
- [26] ASTM C1383-15. *Standard test method for measuring the p-wave speed and the thickness of concrete plates using the impact-echo method*. ASTM International, West Conshohocken, Pa., 2015.
- [27] Jiunn-ming Lin, Mary Sansalone, and William B Streett. Procedure for determining p-wave speed in concrete for use in impact-echo testing using a p-wave speed measurement technique. *ACI Materials Journal*, 94(6), 1997.
- [28] Tae Keun Oh. *Defect characterization in concrete elements using vibration analysis and imaging*. Thesis, 2012.
- [29] Thomas E. Boothby, Paul Kremer, Nicole Trujillo, and Brian Riewestahl. Historic concrete and masonry assessment by the air-coupled impact-echo method. Report, Department of Architectural Engineering, The Pennsylvania State University, 2012.
- [30] Inc. Olson Instruments. Freedom data PC platform for nondestructive testing, <http://www.olsoninstruments.com/data-ndt.php>.
- [31] Khatereh Vaghefi, Henrique A de Melo e Silva, and Devin K Harris. Application of thermal IR imagery for concrete bridge inspection. 2011.
- [32] Naveen Kumar Bolleni. *Environmental effects on subsurface defect detection in concrete structures using infrared thermography*. Thesis, 2009.

- [33] Glenn A. Washer, Richard Fenwick, and Naveen Bolleni. *Development of hand-held thermographic inspection technologies*. Missouri S & T Center for Transportation Infrastructure and Safety, [Rolla, Mo.], 2009.
- [34] B. Wiecek. Review on thermal image processing for passive and active thermography. *Conference proceedings : .. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 1:686, 2005.
- [35] P Rao. Infrared thermography and its applications in civil engineering. *The Indian Concrete Journal*, 82(5):41–50, 2008.
- [36] M. Mieloszyk, M. Krawczuk, P. Malinowski, T. Wandowski, and W. Ostachowicz. *Active thermography method for delamination detection and localisation in composite structures*, volume 2, pages 1261–1268. 2012.
- [37] Rubén Usamentiaga, Pablo Venegas, Jon Guerediaga, Laura Vega, Julio Molleda, and Francisco G Bulnes. Infrared thermography for temperature measurement and non-destructive testing. *Sensors*, 14(7):12305–12348, 2014.
- [38] David J Daniels. *Ground penetrating radar*, volume 1. Iet, 2004.
- [39] WL Lai, Shi Cong Kou, WF Tsang, and Chi Sun Poon. Characterization of concrete properties from dielectric properties using ground penetrating radar. *Cement and Concrete Research*, 39(8):687–695, 2009.
- [40] Antonis Giannopoulos and Nectaria Diamanti. A numerical investigation into the accuracy of determining dielectric properties and thicknesses of pavement layers using reflection amplitude GPR data. In *Ground Penetrating Radar, 2004. GPR 2004. Proceedings of the Tenth International Conference on*, pages 655–658. IEEE, 2004.
- [41] Johannes Hugenschmidt and Roman Loser. Detection of chlorides and moisture in concrete structures with ground penetrating radar. *Materials and Structures*, 41(4):785–792, 2008.

- [42] SF Senin and Roszilah Hamid. Ground penetrating radar wave attenuation models for estimation of moisture and chloride content in concrete slab. *Construction and Building Materials*, 106:659–669, 2016.
- [43] MR Shaw, SG Millard, TCK Molyneaux, MJ Taylor, and JH Bungey. Location of steel reinforcement in concrete using ground penetrating radar and neural networks. *NDT & E International*, 38(3):203–212, 2005.
- [44] AP Annan. *Ground penetrating radar principles, procedures and applications*. 2004.
- [45] ASTM-D6087-08, Standard Test Method for Evaluating Asphalt-Covered Concrete Bridge Decks Using Ground Penetrating Radar, ASTM International, West Conshohocken, PA. 2008.
- [46] Hee Jeong Shin. *Probabilistic bridge deck condition analysis using ground penetrating radar (GPR)*. Thesis, 2004.
- [47] Sherif Yehia, Osama Abudayyeh, Ikhlas Abdel-Qader, and Ammar Zalt. Ground-penetrating radar, chain drag, and ground truth: correlation of bridge deck assessment data. *Transportation Research Record: Journal of the Transportation Research Board*, (2044):39–50, 2008.
- [48] Amos Wamweya. Application of ground penetrating radar (GPR) for bridge deck condition assessment: using a 1.5 ghz ground-coupled antenna. 2009.
- [49] Aleksandra V Varnavina, Aleksey K Khamzin, Evgeniy V Torgashov, Lesley H Sneed, Brandon T Goodwin, and Neil L Anderson. Data acquisition and processing parameters for concrete bridge deck condition assessment using ground-coupled ground penetrating radar: Some considerations. *Journal of Applied Geophysics*, 114:123–133, 2015.
- [50] Christopher L Barnes and Jean-Francois Trottier. Effectiveness of ground penetrating radar in predicting deck repair quantities. *Journal of infrastructure systems*, 10(2):69–76, 2004.

- [51] Mary Sansalone and Nicholas J Carino. Transient impact response of plates containing flaws. *Journal of Research of the National Bureau of Standards*, 92(6):369–381, 1987.
- [52] Mary Sansalone and Nicholas J Carino. Detecting delaminations in concrete slabs with and without overlays using the impact-echo method. *ACI Materials Journal*, 86(2):175–184, 1989.
- [53] GG Clemena. Use of the impact-echo method in nondestructive measurements of the thickness of new concrete pavements. final report. Report, 1995.
- [54] Keith Kesner, Mary Sansalone, and Randall Poston. Use of the impact-echo method to evaluate damage due to distributed cracking in concrete plate members: Theory and field trials. *Transportation Research Record: Journal of the Transportation Research Board*, (1893):61–69, 2004.
- [55] Sherif Yehia, Osama Abudayyeh, Saleh Nabulsi, and Ikhlas Abdelqader. Detection of common defects in concrete bridge decks using nondestructive evaluation techniques. *Journal of Bridge Engineering*, 12(2):215–225, 2007.
- [56] Nenad Gucunski, Greg Slabaugh, Zhe Wang, Tong Fang, and Ali Maher. Visualization and interpretation of impact echo data from bridge deck testing. *Transportation Research Record*, 2050:11–21, 2008.
- [57] Anjali N. Guthrie W. Spencer Mazzeo, Brian A. Patil. Acoustic impact-echo investigation of concrete delaminations using liquid droplet excitation. *NDT & E International*, 51:41–44, 2012.
- [58] Bing Li, Jing Cao, Jizhong Xiao, Xiaochen Zhang, and Hongfan Wang. Robotic impact-echo non-destructive evaluation based on FFT and SVM. pages 2854–2859. IEEE, 2014.
- [59] Fuad Khan and Ivan Bartoli. Detection of delamination in concrete slabs combining infrared thermography and impact echo techniques: A comparative experimental study. volume 9437.

- [60] RH Arnold, HL Furr, and JW Rouse. *Infrared Detection of Concrete Deterioration*. Texas A & M University, Remote Sensing Center, 1969.
- [61] David G Manning and Frank B Holt. *Detecting delamination in concrete bridge decks*. Policy Planning and Research Division, Ontario Ministry of Transportation and Communications, 1980.
- [62] Kenneth R Maser and WM Kim Roddis. Principles of thermography and radar for bridge deck assessment. *Journal of transportation engineering*, 116(5):583–601, 1990.
- [63] MR Clark, DM McCann, and MC Forde. Application of infrared thermography to the non-destructive testing of concrete and masonry bridges. *NDT & E International*, 36(4):265–275, 2003.
- [64] Ch Maierhofer, H. Wiggensauser, A. Brink, and M. Röllig. Quantitative numerical analysis of transient IR-experiments on buildings. *Infrared Physics and Technology*, 46(1):173–180, 2004.
- [65] Chia-Chi Cheng, Tao-Ming Cheng, and Chih-Hung Chiang. Defect detection of concrete structures using both infrared thermography and elastic waves. *Automation in Construction*, 18(1):87–92, 2008.
- [66] Jean Dumoulin, L Ibos, Clemente Ibarra-Castanedo, A Mazioud, M Marchetti, Xavier Maldague, and A Bendada. Active infrared thermography applied to defect detection and characterization on asphalt pavement samples: comparison between experiments and numerical simulations. *Journal of modern optics*, 57(18):1759–1769, 2010.
- [67] Glenn Washer, Richard Fenwick, Seth Nelson, and Rilya Rumbayan. Guidelines for thermographic inspection of concrete bridge components in shaded conditions. *Transportation Research Record: Journal of the Transportation Research Board*, (2360):13–20, 2013.

- [68] DG Manning and FB Holt. The development of deck assessment by radar and thermography. *Transportation Research Record*, (1083), 1986.
- [69] Henrique LM dos Reis and Matthew D Baright. Evaluation of bridge deck delamination investigation methods. Technical report, 1998.
- [70] Steve Cardimona, B Willeford, J Wenzlick, and N Anderson. Bridge deck condition studies in missouri utilizing ground penetrating radar. Report, 2001.
- [71] M Scott, A Rezaizadeh, A Delahaza, CG Santos, M Moore, B Graybeal, and G Washer. A comparison of nondestructive evaluation methods for bridge deck assessment. *NDT & E International*, 36(4):245–255, 2003.
- [72] Christopher L Barnes, Jean-François Trottier, and Dean Forgeron. Improved concrete bridge deck evaluation using GPR by accounting for signal depthâamplitude effects. *NDT & E International*, 41(6):427–433, 2008.
- [73] Zhe Wendy Wang, Mengchu Zhou, Greg Slabaugh, Jiefu Zhai, and Tong Fang. Automatic detection of bridge deck condition from ground penetrating radar images. *Automation Science and Engineering, IEEE Transactions on*, 8(3):633–640, 2011.
- [74] Alexander Tarussov, Marc Vandry, and Aldo De La Haza. Condition assessment of concrete structures using a new analysis method: Ground-penetrating radar computer-assisted visual interpretation. *Construction and Building Materials*, 38:1246–1254, 2013.
- [75] Francisco A Romero, Christopher L Barnes, Hoda Azari, Soheil Nazarian, and Carl D Rascoe. Validation of benefits of automated depth correction method: Improving accuracy of ground-penetrating radar deck deterioration maps. *Transportation Research Record: Journal of the Transportation Research Board*, (2522):100–109, 2015.

- [76] CL Barnes and JF Trottier. Effectiveness of ground penetrating radar for preparing pre-tender deterioration estimates on asphalt covered reinforced concrete bridge decks. *Technomic Publishing Co, Inc, Structural Materials Technology IV(USA)*,, pages 47–52, 2000.
- [77] Michael Scott, Ali Rezaizadeh, and Mark Moore. Phenomenology study of hermes ground-penetrating radar technology for detection and identification of common bridge deck features. Technical report, 2001.
- [78] W Kim, A Ismail, NL Anderson, EA Atekwana, and A Buccellato. Non-destructive testing (NDT) for corrosion in bridge decks using gpr. In *The 3rd International Conference on Applied Geophysics. Orlando, Florida, USA*, 2003.
- [79] Tyler W Robison and Jennifer E Tanner. Bridge deck evaluation using non-destructive test methods. Technical report, 2011.
- [80] ASTM D6087 08(2015)e1. Standard test method for evaluating asphalt-covered concrete bridge decks using ground penetrating radar, 2015.
- [81] GG Clemena. The use of ground-penetrating radar in the surveying of overlaid bridge decks. final report. Technical report, 1993.
- [82] W. D. Rummel. Probability of detection (POD) is not NDT/E reliability. *Review of Progress in Quantitative Nondestructive Evaluation, Vols 32a and 32b*, 1511:1809–1816, 2013.
- [83] George A Matzkanin and HT Yolken. Probability of detection (POD) for nondestructive evaluation (NDE). Report, DTIC Document, 2001.
- [84] A P Berens and P W Hovey. Statistical Methods for Estimating Crack Detection Probabilities. *Probabilistic Fracture Mechanics and Fatigue Methods: Applications for Structural Design and Maintenance*, pages 79–94, 1983.

- [85] Christina Müller, Thomas Fritz, Gerd-Rüdiger Tillack, Carsten Bellon, and Martina Scharmach. Theory and application of the modular approach to NDT reliability. *Materials evaluation*, 59(7):871–874, 2001.
- [86] Jeremy Knopp, Ramana Grandhi, Li Zeng, and John Aldrin. Considerations for statistical analysis of nondestructive evaluation data: hit/miss analysis. *E-Journal of Advanced Maintenance, Japan Society of Maintenology*, 4(3):105–115, 2012.
- [87] Ward D Rummel. Nondestructive inspection reliabilityâhistory, status and future path. In *18th World Conference on Nondestructive Testing, Durban, South Africa*.
- [88] Alan P Berens and Peter W Hovey. Evaluation of NDE reliability characterization. volume i. Report, DTIC Document, 1981.
- [89] William H Lewis, William H Sproat, Bruce D Dodd, and James M Hamilton. Reliability of nondestructive inspections. Report, DTIC Document, 1978.
- [90] Bruce R. Ellingwood, B. Bhattacharya, R. Zheng, U. S. Nuclear Regulatory Commission, Research Office of Nuclear Regulatory, Technology Division of Engineering, and Laboratory Oak Ridge National. *Reliability-based condition assessment of steel containment and liners*. Division of Engineering Technology, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission : Supt. of Docs., U.S. G.P.O. [distributor], Washington, D.C., 1996.
- [91] Alan P Berens. Probability of Detection (POD) Analysis for the Advanced Retirement for Cause (RFC)/Engine Structural Integrity Program (ENSIP) Nondestructive Evaluation (NDE) System-Volume 1: POD Analysis. Report, DTIC Document, 2000.

- [92] Ruoxue Zhang and Sankaran Mahadevan. Fatigue reliability analysis using nondestructive inspection. *Journal of Structural Engineering*, 127(8):957–965, 2001.
- [93] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [94] Scott B Cantor, Charlotte C Sun, Guillermo Tortolero-Luna, Rebecca Richards-Kortum, and Michele Follen. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of clinical epidemiology*, 52(9):885–892, 1999.
- [95] Tom Fawcett. ROC graphs: Notes and practical considerations for researchers. *Machine learning*, 31(1):1–38, 2004.
- [96] Nancy A Obuchowski. Receiver operating characteristic curves and their use in radiology 1. *Radiology*, 229(1):3–8, 2003.
- [97] T. Cai and L. E. Dodd. Regression analysis for the partial area under the ROC curve. *Stat. Sin. Statistica Sinica*, 18(3):817–836, 2008.
- [98] Xiao-Hua Zhou, Donna K McClish, and Nancy A Obuchowski. *Statistical methods in diagnostic medicine*, volume 569. John Wiley & Sons, 2009.
- [99] Karim O Hajian-Tilaki, James A Hanley, Lawrence Joseph, and Jean-Paul Collet. A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic tests. *Medical Decision Making*, 17(1):94–102, 1997.
- [100] S. H. Park, J. M. Goo, and C. H. Jo. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol*, 5(1):11–8, 2004.
- [101] Ertugrul Colak, Fezan Mutlu, Cengiz Bal, Setenay Oner, Kazim Ozdamar, Bulent Gok, and Yuksel Cavusoglu. Comparison of semiparametric, parametric, and nonparametric ROC analysis for continuous diagnostic tests using

- a simulation study and acute coronary syndrome data. *Computational and Mathematical Methods in Medicine*, 2012:7, 2012.
- [102] Luzia Gonçalves, Ana Subtil, M Rosário Oliveira, and PZ Bermudez. ROC curve estimation: an overview. *RevStat-Statistical JI*, 12(1), 2014.
 - [103] M Greiner, D Pfeiffer, and RD Smith. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine*, 45(1):23–41, 2000.
 - [104] Nicole Martino, Ken Maser, Ralf Birken, and Ming Wang. Determine ground penetrating radar amplitude thresholds for the corrosion state of reinforced concrete bridge decks. 2013.
 - [105] Hoda Azari, Deren Yuan, Soheil Nazarian, and Nenad Gucunski. Sonic methods to detect delamination in concrete bridge decks: Impact of testing configuration and data analysis approach. *Transportation Research Record: Journal of the Transportation Research Board*, (2292):113–124, 2012.
 - [106] R Vardasca, P Plassmann, J Gabriel, EFJ Ring, and IDMEC-FEUP LABIOMEPE. Towards a medical imaging standard capture and analysis software. In *12th International Conference on Quantitative InfraRed Thermography, Bordeaux, France*, pages 162–168, 2014.
 - [107] K Senthil Kumar, G Kavitha, G Ramesh, and R Subramanian. *Visual and thermal image fusion for UAV based target tracking*. INTECH Open Access Publisher, 2011.
 - [108] Sofus Macskassy and Foster Provost. Confidence bands for ROC curves: Methods and an empirical study. Proceedings of the First Workshop on ROC Analysis in AI. August, 2004.
 - [109] Henry G Russell. *Concrete bridge deck performance*, volume 333. Transportation Research Board, 2004.

- [110] Xian-Qi He, Zi-Qiang Zhu, Qun-Yi Liu, and Guang-Yin Lu. Review of GPR rebar detection. In *PIERS Proceedings*, pages 804–813.

Appendices

Appendix A

MATLAB Script & Functions

A.1 Main script

```
clear all
close all
clc
load File name.dat; % Read the thermal data %
IR = File name;
X = imread('image1.png'); % Read the digital image
X = rgb2gray(X); % Convert the digital image that was cut
previously in the locations of delamination to a gray scale
X = imresize(X,size(IR)); % Change the size of the last gray picture
to match the size of IR image [480,640]
Y = zeros(size(X));
for I = 1:size(X,2) % The next step is to convert the gray
image to a binary coded picture as either 0 or 1
    for ii = 1:size(X,1)
        if x(ii,I)>= 255 % white color code=255
            y(ii,I)= 1;
        end
    end
end
% the next four steps are to delineate the temp. of delaminated pixels from
sound pixels, (y==1 for flawed points) or (y==0 for sound points)
Intact = find(y==0);
NegPix = IR(Intact);
Flawed = find(y==1);
PosPix = IR(Flawed);
meanNeg = mean(NegPix(:)); % Mean of neg. pixels (reference temperature)
meanPos = mean(PosPix(:)); % Mean of pos. pixels
stdNeg = std(NegPix(:));
stdPos = std(PosPix(:));
medianNeg = median(NegPix(:));
medianPos = median(PosPix(:));
% ----- %
% To filter out outliers (Optional in case of presence of outliers) %
% ----- %
limit1 = meanPos+3*stdPos;
limit2 = meanNeg-3*stdNeg;
index1 = find(IR>limit1);
index2 = find(IR<limit2);
for k=1:size(IR,1)
    for j=1:size(IR,2)
        if (IR(k,j)>=limit1 && y(k,j) == 1) || (IR(k,j)<=limit2 && y(k,j) == 1)
            IR(k,j) = medianPos;
        end
        if (IR(k,j)>=limit1 && y(k,j) == 0) || (IR(k,j)<=limit2 && y(k,j) == 0)
            IR(k,j) = medianNeg;
        end
    end
end
NegPix = IR(Intact); % Neg. pixels after filtering irrelevant data
PosPix = IR(Flawed); % Pos. pixels after filtering irrelevant data
ContrastNeg = NegPix-meanNeg; % Thermal contrast between the neg. pixels and
the avg of the neg pixels (reference)
ContrastPos = PosPix-meanNeg; % Thermal contrast between the pos. pixels and
the avg of the neg pixels (reference)
```

```

%----- %
% Plotting histograms (Optional) %
%----- %
nbins = 100;
figure(1)
hist(ContrastNeg,nbins);
h = findobj(gca,'Type','patch');
set(h,'FaceColor','r','EdgeColor','w','facealpha',0.75)
hold on
hist(ContrastPos,nbins);
h = findobj(gca,'Type','patch');
set(h,'facealpha',0.75);
ylabel('Counts','FontSize',16);
xlabel('Thermal Contrast (\Delta T)','FontSize',16);
h_legend = legend('Sound Pixels','Delaminated Pixels');
set(h_legend,'FontSize',16,'Box','on')
xt = get(gca, 'XTick');
set(gca, 'FontSize', 14)
%----- %
% ROC analysis and AUC calculation %
%----- %
minthre = min(ContrastNeg) - 0.5;
maxthre = max(ContrastPos) + 0.5;
th=minthre:0.15:maxthre; % Threshold range

Z=temp(IR,th, medianNeg); % Function 1 to transform the IR image to a binary
output of(0 and 1) for different decision threshold settings (See below)

re=com(Z,y); % Function 2 to compare pixels from a set of
images and calculated the decision output (See below)

TPR=re(:,2)./(re(:,2)+re(:,1));
FPR=re(:,3)./(re(:,3)+re(:,4));
AUC = trapz(FPR, TPR);
AUC = abs (AUC);
th = th';
%----- %
figure (2)
hold on;
imshow (y)
impixelinfo
figure(3)
imagesc (IR)
colorbar
impixelinfo
figure(4)
P = plot(FPR,TPR,'-'); % ROC curve
set (P, 'LineWidth', 1.5);
grid on
axis([-0.02,1.02,-0.02,1.02])
xlabel('Probability of False Alarm (POFA)','FontSize',14);
ylabel('Probability of Detection (POD)','FontSize',14);

IR = IR(:); % to convert IR matrix to one column, necessary for WMW analysis%
y = y(:); % to convert the binary digital matrix to one column, necessary
for WMW analysis%

```

Figure A.1: The developed MATALB script necessary to analyze data of IRT technology

A.2 Functions

```
% Function 1 %

function ZZ = temp(IR,th, medianNeg)
Tsound = medianNeg; % Tsound is supposed to be equal to the median of the
sound pixels %
DeltaT = IR-Tsound; % delta t between the pixel temprature and the mode or
the reference temp %

for j = 1:length(th)
for I = 1:size(IR,1)
    for ii = 1:size(IR,2)
        if DeltaT(i,ii)>=th(j)
            Z(i,ii) = 1;
        else
            Z(i,ii) = 0;
        end
    end
end
ZZ{j} = Z;
end
end

% Function 2 %

function re = com(Z,y)
for I = 1:size(Z,2)
    a = Z{i};
    b = y-a;
    for ii = 1:size(y,2)
        for iii = 1:size(y,1)
            if y(iii,ii)==1
                if b(iii,ii)==-1
                    c(iii,ii) = 1; % False Negative
                else
                    c(iii,ii) = 2; % True Positive
                end
            else
                if b(iii,ii)==-1
                    c(iii,ii) = 3; % False Positive
                else
                    c(iii,ii) = 4; % True Negative
                end
            end
        end
    end
end
for j = 1:4
    re(i,j) = length(find(c==j));
end
end
end
```

Figure A.2: Functions used in the above MATLAB script

Appendix B

Details of Mockups and In-service Bridge Decks Images

B.1 Images of the fabricated slab I

Figure B.1: Details of the thermal and digital images of the fabricated slab I.

Figure B.2: Distributions of the thermal contrast over the actually negative and positive locations (pixels) of the fabricated slab I thermal images.

Figure B.3: ROC curves of the fabricated slab I thermal images.

B.2 Images of the in-service bridge decks

Figure B.4: Details of the thermal and digital images of the in-service bridge decks.

Figure B.5: Distributions of the thermal contrast over the actually negative and positive locations (pixels) of the in-service bridge decks thermal images.

Figure B.6: ROC curves of the in-service bridge decks thermal images.

VITA

Ali A. Sultan was born in south of Iraq in 1978. He completed his BSc. and MSc. degrees from University of Technology - Iraq in 2000 and 2003, respectively. After graduation, Sultan has occupied several engineering positions in some worldwide human organizations and construction companies such “Save the children” organization and Parsons company. In 2007, he joined University of Thi-Qar as a faculty member in civil Engineering Department. During the period between 2007 to 2012, Sultan taught several engineering courses, in particular, structural analysis, steel structures analysis and design, and project managements. He pursued his PhD at University of Missouri-Columbia starting at 2012. Sultan’s PhD research has focused towards assessing the reliability of nondestructive evaluation technologies applied to assess the integrity of reinforced concrete bridge infrastructures.