

EMPIRICAL MODELING OF PIPING ALONG MISSISSIPPI RIVER LEVEES IN SOUTHWESTERN ILLINOIS

A Thesis
presented in partial fulfillment of requirements
for the degree of Master of Science
in Engineering Science with Emphasis in Geological Engineering
in the Department of Geology and Geological Engineering
The University of Mississippi

By
Rebecca Shields
December 2015

Copyright Rebecca Shields 2015

ALL RIGHTS RESERVED

ABSTRACT

Piping beneath levees within the Middle Mississippi River, MMR, has been well documented for 78 years, when the Mississippi River Commission initiated geological investigations into underseepage following a substantial flood in 1937. The United States Army Corps of Engineers (USACE) defines a levee as an embankment designed to supply flood protection from seasonal high water. Piping is the “*active erosion* of sand or other soil from the top stratum as a result of substratum pressure and concentration of seepage in localized channels” (USACE, 1956a). The geological investigations beginning in 1937, and still continuing today, have consistently listed two conditions necessary for piping to occur: 1) a pervious substratum overlain (2) by a semi- to impervious top stratum (Fisk, 1945; Turnbull and Mansur, 1959). The phrase “conductive environment” is used for this type of environment. Where these factors are present during the time when a levee is subjected to water loading, the force exerted by the weight of water on the riverside of the levee can be transferred through the pervious substratum to the landside, resulting in a hydrostatic imbalance between strata and the surface landside of the levee (Turnbull and Mansur, 1959; USACE, 2000).

Innumerable miles of levee along the MMR and other rivers meet the “conductive environment” susceptible to piping and merit maintenance and piping prevention measures. Several secondary factors were identified in previous studies resulting in detailed geological investigations of all known levee districts meeting the “conductive environment”. However, limited funding complicates data management and therefore, adequate piping prevention measures, along

these levees. Using the Glynn and Kuszmaul (2004) database for PDR and FTC, several datasets were defined for regression analysis to develop a model that improves the efficiency of vulnerability assessments of the vast lengths of levees managed by the USACE. Single-variable regression analysis, to determine significance of each independent variable, and multi-variable regression analysis, to determine the final models for the datasets, were conducted during model building. Several possible models for each dataset were created using a modified forward stepwise regression procedure, also called a stepwise regression procedure, as suggested by Le (2010). Model selection was based on the chi-square statistic value and each models performance under thresholds discussed in subsection 4.3.1.

The model building process presented in this study proved to be a successful method for developing regression models meant to predict the potential for piping given the availability, or lack, of geologic and flood specific data. The final selected model, Limited Previous Model A, significantly predicted areas of high, medium, and low potential for piping along three levee districts; Prairie du Rocher Levee District (PDR), Fort Chartres Levee District (FTC), and East Cape Girardeau Levee District (ECG). The high significance of this model is largely attributed to the inclusion of previous piping events and interaction terms.

DEDICATION

I would like to dedicate this thesis to my son, Cameron Milo.

LIST OF ABBREVIATIONS AND SYMBOLS

USACE----- United States Army Corps of Engineers

USGS----- United States Geological Survey

MRC----- Mississippi River Commission

MMR----- Middle Mississippi River

PDR----- Prairie du Rocher Levee District

FTC----- Fort Chartres Levee District

ECG----- East Cape Girardeau Levee District

i_c ----- Critical gradient

$\tilde{\alpha}'$ ----- Buoyant unit weight of soil

$\tilde{\alpha}_w$ ----- Unit weight of water

G_s ----- Specific Gravity of soil solids

e ----- Void ratio

H ----- Net head

h_0 ----- Factor of the raw head during a flood event

z ----- Top stratum's vertical extent or thickness

z_t ----- Transformed confining layer thickness

\tilde{u} ----- Presence and orientation of swales

L_3 ----- Length of top stratum landside of the levee

k_b ----- Vertical permeability of top stratum

k_n ----- Vertical permeability of soil type

\check{R} ----- Riverside Borrow Pits

\acute{L} ----- Landside Ditches

β ----- Landside Seepage Berms

\acute{R} ----- Relief Wells

\mathcal{P} ----- Previous Piping Events

d_{10} ----- Effective aquifer grain size

ACKNOWLEDGMENTS

I would like to give many thanks to Dr. Joel Kuszmaul for his unwavering attention and encouragement through this process. Dr. Louis Zachos, thank you for your statistical expertise and willingness to explain even the simple questions and thank you Dr. Adnan Aydin for your very helpful questions and encouragement. Thank you Maureen Corcoran and Julie R. Kelley at the United States Engineering Research and Development Center for all of your help and enthusiasm. Finally, I would not have made it this far if my family and friends had not supported me through all this; especially my mother, Nancy Shields; my father, Dennis Shields; my sister, Ellison Shields; and Scott Milo.

TABLE OF CONTENTS

Abstract	ii
Dedication	iv
List of Abbreviations and Symbols	v
Acknowledgments	vii
List of Tables	xiii
List of Figures	xv
1. Introduction	1
1.1 United States Army Corps of Engineers Civil Works Involvement	2
1.2 Levee Failure and Piping	3
1.3 Piping Investigations	3
1.4 Research Objectives	5
1.5 Purpose of Research	7
2. Geologic Setting.....	9
2.1 Middle Mississippi River	10
2.2 Field Areas	12

3. Piping Mechanics and Wilson’s Efforts to Predict Levee Potential	14
3.1 Mechanics of Piping	14
3.1.1 Piping Preventative Measures.....	16
3.1.1.1 Cut Off Trenches	17
3.1.1.2 Pervious Toe Trenches	18
3.1.1.3 Riverside Impervious Blankets	18
3.1.1.4 Landside Seepage Berms	19
3.1.1.5 Pressure Relief Wells	19
3.1.2 Secondary Factors of Piping	20
3.1.2.1 Unfavorable Geologic Condition	21
3.1.2.2 Effectiveness of Top stratum	23
3.1.2.3 Susceptibility of the Substratum	25
3.1.2.4 Severity of Flood	26
3.1.2.5 Effectiveness of Piping Prevention Measures	28
3.1.2.6 Role of Previous Piping Events	28
3.2 Wilson’s Efforts	29

3.2.1 Compilation of Databases	29
3.2.1.1 Analyzed Variables	30
3.2.1.2 Direct Variables	33
3.2.2 Model Building	33
3.2.3 Model Thresholds	37
3.2.4 Conclusions of Wilson's Models	38
4. Methods: Model Building, Selection and Application	40
4.1 The Engineering Applications for Developed Models	41
4.2 Model Building	42
4.2.1 Defining Datasets	42
4.2.2 Stepwise Regression Procedure	45
4.2.2.1 Stepwise Regression Procedure: Defining Criteria	46
4.2.2.2 Stepwise Regression Procedure: Strategy	47
4.3 Model Selection	48
4.3.1 Thresholds	50
4.3.2 Determining Model Utility	54

4.4 Blind Tests	55
5. Results	57
5.1 Significant Independent Variables	57
5.2 Model Utility: Selected Models	63
5.2.1 Comparison of Limited and Unlimited Datasets	63
5.2.2 Comparison of Pure and Previous Datasets	64
5.2.3 Contingency Table Comparisons	67
5.2.4 Geographic Information Systems Applications.....	74
5.3 Model Utility: Blind Testing	79
6. Conclusions	83
6.1 Model Selection: Conclusions	83
6.2 Blind Testing Findings	85
6.3 Recommendations	85
6.4 Recommendations on Future Research	86
List of References	88
Appendix A	94

Vita	100
------------	-----

LIST OF TABLES

Table 1. Description of independent variables used for regression analysis.....	29
Table 2. Description of PDR-93 Linear model.....	34
Table 3. Description of PDR-93 Logit model.....	34
Table 4. Description of PDR-95 Linear model.....	34
Table 5. Description of PDR-95 Logit model.....	35
Table 6. Performance of Wilson (2003) models under his applied thresholds.....	37
Table 7. Allocation of independent variables for each defined dataset.....	44
Table 8. Number of models created during model building.....	48
Table 9. Top two performing models from each dataset.....	50
Table 10. Significance of final independent variables to piping locations.....	58
Table 11. PDR Pure Limited Model B.....	67
Table 12. PDR_FTC Pure Limited Model A.....	68
Table 13. PDR Previous Unlimited Model CS.....	68
Table 14. PDR Previous Limited Model E.....	69
Table 15. PDR Pure Limited Model B.....	70
Table 16. PDR Pure Limited Model B.....	70
Table 17. PDR Pure Limited Model B.....	70
Table 18. PDR_FTC Pure Limited Model A.....	71
Table 19. PDR_FTC Pure Limited Model A.....	71
Table 20. PDR_FTC Pure Limited Model A.....	71
Table 21. PDR Previous Unlimited Model CS.....	72
Table 22. PDR Previous Unlimited Model CS.....	72
Table 23. PDR Previous Unlimited Model CS.....	72
Table 24. PDR_FTC Previous Limited Model E.....	73

Table 25. PDR_FTC Previous Limited Model E.....	73
Table 26. PDR_FTC Previous Limited Model E.....	73
Table 27. PDR Raw Previous.....	74
Table 28. PDR_FTC Raw Previous.....	74
Table 29. Description of final selected model, Limited Previous: Model A.....	85

LIST OF FIGURES

Figure 1. General study area.....	13
Figure 2. Depiction of underseepage, piping, and sand boil development through a cross-section of a levee.....	15
Figure 3. Best Fit top Unlimited Pure Model.....	52
Figure 4. Best Fit top Unlimited Pure Model.....	52
Figure 5. Best Fit top Unlimited Previous Model.....	53
Figure 6. Best Fit top Unlimited Previous Model.....	53
Figure 7. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Unlimited Pure model, original piping observations, and all observations.....	59
Figure 8. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Unlimited Previous model, original piping observations, and all observations.....	59
Figure 9. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Limited Pure model, original piping observations, and all observations.....	60
Figure 10. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Limited Previous model, original piping observations, and all observations.....	60
Figure 11. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Unlimited Pure model, original piping observations, and all observations.....	61
Figure 12. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Unlimited Previous model, original piping observations, and all observations.....	61
Figure 13. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Limited Pure model, original piping observations, and all observations.....	62

Figure 14. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Limited Previous model, original piping observations, and all observations.....	62
Figure 15. Bar chart for the Unlimited Previous model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	65
Figure 16. Bar chart for the Limited Previous model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	65
Figure 17. Bar chart for the Unlimited Pure model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	66
Figure 18. Bar chart for the Limited Pure model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	66
Figure 19. Map of piping potentials for Prairie du Rocher levee district determined from Limited Pure: Model B.....	75
Figure 20. Map of actual piping locations atop piping potentials for Prairie du Rocher levee district determined from Limited Pure: Model B.....	76
Figure 21. Map of piping potentials for Prairie du Rocher levee district determined from Unlimited Previous: Model CS.....	77
Figure 22. Map of actual piping locations atop piping potentials for Prairie du Rocher levee district determined from Unlimited Previous: Model CS.....	78
Figure 23. Bar chart for the application of the Limited Previous model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	80
Figure 24. Bar chart for the application of the Unlimited Previous model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	80
Figure 25. Bar chart for the application of the Limited Pure model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	81
Figure 26. Bar chart for the application of the Unlimited Pure model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.....	81

CHAPTER ONE

Introduction

Since 1717, earthen levees have helped defend valuable private and public property in the United States, protecting both farmland and major cities, such as St. Louis, Missouri, and New Orleans, Louisiana, from flood events along the Mississippi River (NHRAIC-UCB, 1992). The United States Army Corps of Engineers (USACE) (2000) defines a levee as an embankment designed to supply flood protection from seasonal high water. Seasonal high water events are irregular in intensity and timing, but occur as a result of weather and climate cycles (USACE, 2000). USACE design standards state levees should be designed to withstand water loading only for a few days to weeks per year. Earthen dams are required for circumstances when water loading is more constant (USACE, 2000).

Levee design and maintenance in this country has evolved from construction with minimal standards to engineered structures built using federal assistance by the addition of several flood controls acts written for the specific purpose of flood defense (NCLS, 2009; NHRAIC-UCB, 1992). Extreme damage and loss of life in the early 20th Century prompted the first official federally funded flood control laws, also known as the Flood Control Act of 1917, issued by Congress under prolonged national political and public pressure to do so (Wright, 2000). Also, the Flood Control Act of 1936 officially adopted a national policy of river development for flood control and devoted a total of \$320 million, equal to \$5.5 billion today, to its development (Wright, 2000).

1.1 United States Army Corps of Engineers Civil Works Involvement

Congressional establishment of the U.S. Army Corps of Engineers (USACE) occurred in 1802 with the Act of Mar. 16, 1802, written to recognize a need for a corps of engineers specializing in military knowledge and establish a base for the corps and a respective military academy at West Point, New York (Powers, 1977; Curtis, 2005). The primary responsibilities of the USACE initially focused on the construction and maintenance of military structures with the expansion into coastal fortifications between Maine and New Orleans during the War of 1812 (Power, 1977; Curtis, 2005). The USACE began to establish some lighthouses, jetties, harbors, and other coastal features, a transition from fortification to navigational improvement, post-War of 1812. It was during this time that the USACE initiated their (still ongoing) efforts to improve civil works across the nation (Power, 1977).

As of 2010, the USACE had constructed and/or maintained approximately 383 reservoirs, over 90 coastal storm damage reduction projects, and 2,000 levees, equal to 8,500 river miles (NCLS, 2009; USCAE, 2010). These efforts cost the federal government approximately \$120 billion but were able to prevent an estimated \$706 billion in flood damages (USACE, 2010). A National Levee Safety Program established by the USACE is funded for continued research, development, and implementation of tools, policies, and methods defined by the USACE in the Recommendations for a National Levee Safety Program draft report (NCLS, 2009; USACE, 2010).

1.2 Levee Failure and Piping

Four forms of levee failure are identified in the levee design manual produced by USACE (2000): (1) overtopping, (2) surface erosion, (3) internal erosion (piping), (4) slides within the levee embankment or the foundation soils (USACE, 2000). The third form of failure, internal erosion (piping) was first acknowledged after a damaging flood along the Lower Mississippi River in 1937 (Turnbull and Mansur, 1959). Piping is the “*active erosion* of sand or other soil from the top stratum as a result of substratum pressure and concentration of seepage in localized channels” (USACE, 1956). It can occur during periods of high water along levees constructed atop an environment conducive to piping occurrence: a semi- to impervious top stratum underlain by a pervious substratum (USACE, 2000). This paper will refer to this type of environment as a “conductive environment”

1.3 Piping Investigations

Piping was first acknowledged in 1937 by the Mississippi River Commission (MRC) following a damaging flood along the Lower Mississippi River (Turnbull and Mansur, 1959). An investigation into piping and its controls was commissioned by the MRC in September 1940 (Turnbull and Mansur, 1959). Numerous studies on piping mechanics and prevention were funded by the MRC and conducted by some of the same scientists commissioned in the original investigation, as well as others, throughout the next twenty years, e.g. Fisk, 1945; Fisk, 1947; Turnbull, Krinitzsky, and Johnson 1950; Turnbull and Mansur, 1954; Mansur, Kaufman, and Schultz, 1956; Turnbull and Mansur, 1959.

A substantial flood in 1973, brought on by an unusually wet winter, broke record flood levels along the Upper/Middle Mississippi River. At that time, w Charles Kolb (1975), former Chief of Engineering Geology Division, U.S. Army Engineer Waterways Experiment Station, believed flooding was the result of river constriction from the levee system between Alton to Gale, IL. Investigations in the performance of piping prevention measures during the flood and updated research on the influence of geologic features on the location of sand boil formation were completed using the empirical data available from the flood (Kolb, 1975; USACE, 1976).

Extensive piping and sand boil formation did not occur along the Mississippi River again until the Great Flood of 1993, “the most costly and widespread natural disaster in Illinois history” (Chrzastowski et al., 1994). The Great Flood of 1993 resulted in unprecedented flooding throughout the Mississippi River watershed and estimates of damage hover around \$1.3 billion for the state of Illinois alone (Chrzastowski et al., 1994). While most levees failed from overtopping, the especially long duration of water loading resulted in significant piping and sand boil formation along several levees in the Middle Mississippi River (MMR) (Chrzatowski et al., 1994). This flood provided a much needed update on levee performance and piping locations along the Mississippi River. Several studies were conducted using the new data: Bhowmik et al. (1994), Li et al. (1996), Mansur, Postol, and Salley (2000), Ozkan (2003), Wilson (2003), and Glynn and Kuszmaul (2004).

These studies, especially those by Fisk (1945), Turnbull and Mansur (1956), Kolb (1975), Mansur (2000), and several by USACE (1956a, 1956b, 1976), were very successful in determining the “conductive environment”, identifying several secondary characteristics, and developing flood control measures designed for piping prevention. However, innumerable miles of levee along the MMR and other rivers are constructed on the “conductive environment”

susceptible to piping and merit maintenance and piping prevention measures. With limited funds available for such maintenance and upgrade projects, the purpose of the research for this thesis is to develop a method to prioritize different levee segments by identifying areas most vulnerable to piping beneath the levees. Several secondary factors were identified in previous studies resulting in detailed geological investigations of all known levee districts meeting the “conductive environment” criterion.

Knowledge of “conductive environment” locations correlates with extensive data on these secondary factors on the hundreds to thousands of miles of levee meeting the “conductive environment” compiled by funded agencies. However, limited funding complicates data management and therefore, adequate piping prevention measures, along these levees. By expanding the research conducted by Wilson (2003), this study seeks to identify quantifiable conditions along levees that influence subsurface erosion and provide a way for the efficient management of the vast quantity of data in order to alert the affected parties to piping potential in their area.

1.4 Research Objectives

Previous researchers (Fisk, 1945; Turnbull and Mansur, 1959; and Kolb, 1975) have observed the defining role that secondary factors play in the development of piping. The objectives of Wilson (2003) were to create a database of influential variables on piping along the MMR levees and use that database in a geographic information system (GIS) to determine the potential for piping through regression analysis. For a detailed explanation of regression analyses, please see Appendix A.

The GIS database included two levee districts: Prairie du Rocher (PDR) Levee District and Fort Chartres (FTC) Levee District located in Prairie du Rocher, Illinois. Data were obtained through USACE, St. Louis District, and included boring logs and flood reports from the 1993 Flood and a 1995 flood, U.S. Geological Survey (USGS) aerial photography and Light Detection and Ranging (LiDAR) data, and “flood fight” notes supplied by local levee inspectors and other involved parties (Wilson, 2003). The methods and results of Wilson’s research will be further discussed in Chapter 3.

Wilson’s work proved that some variables, especially locations of previous piping events, are more significant to piping than others and a levee’s piping potential could be predicted by regression analysis. However, it is possible that model utility could greatly improve by categorizing PDR and FTC data into defined datasets with the addition of interaction terms and using the forward elimination method during logistic regression analysis suggested by Le (2010), as opposed to the backward elimination method suggested by Le (1998) and used by Wilson (2003).

The specific objectives of this research are:

1. Create several datasets using Wilson’s data for PDR and FTC to develop functional models for piping potential along the levees based on 1993 piping occurrences. The method of forward elimination described by Le (1998) will be used for this model building. This type of model is applicable to districts where previous piping events have not been observed and directly indicates the direct influence of secondary geologic factors on piping.
2. Use datasets from step 1 with the addition of 1993 piping locations to develop functional models for piping potential

3. along PDR and FTC levees based on 1995 piping occurrence. This will be accomplished by following the same methods as step 1. This type of model is applicable to districts where previous piping events have been observed; however, it is not able to show the direct influence geologic factors.
4. Determine selected models for each dataset by applying high, medium, and low piping potential thresholds, developed in analysis of the best fit model for each respective dataset. Best fit is determined by the Chi^2 -value of the model (Le, 2010; Davis, 2002).
5. Create new dataset of all variables used in selected models from step 1 and step 2, for East Cape Girardeau (ECG) Levee District in East Cape Girardeau, Illinois. Model selection is accomplished following step 3.
6. Apply ECG dataset (from Step 4) to selected models (from Step 3) to assess model utility.

1.5 Purpose of Research

Limited funding for flood control restricts levee maintenance and installation of piping prevention measures. The objective of this research is to develop improved methods for the *efficient* management of large datasets on geologic variables and flood events for levees for the purpose of identifying the conditions along levees that make them most susceptible to subsurface erosion.

For piping prevention, detailed geologic investigations of all levee districts meeting the “conductive environment” must be compiled to understand the influence of secondary factors. The

result is large quantities of data for hundreds to thousands of miles of levees along the Mississippi River and other rivers across the country.

CHAPTER TWO

Geologic Setting

The Mississippi River, a meandering stream type, stretches approximately 3,770 river kilometers from Lake Itasca, Minnesota to the Gulf of Mexico (Alexander et al., 2012; Mac et al., 1998). The river's watershed spans 31 States and covers 41% of the continental United States (Alexander et al., 2012). It is confined by the Rocky Mountain Belt to the west and the Appalachian Mountains to the east (Alexander et al., 2012).

Natural levees form on the outside of channel bends while point bars form on the inside of channel bends (Fisk, 1945). The geological evolution of a point bar has been extensively studied and more detailed specifications are available through other texts such as Fisk (1945) and Kolb (1975). However, in general, point bars are formed on the inside of the bends where river's velocity is slowest and deposition of sediment occurs most rapidly (Kolb, 1975). Channel migration results in sandy ridges adjacent to clayey depressions, known as ridge and swale topography (Fisk, 1945). Abandoned channels may also account for swale-type behavior due to thick deposits of clay within the channel but this is dependent on the type of abandoned channel (Fisk, 1945). Meandering stream deposition is still active and occurring today.

Establishment of the Mississippi River is a direct consequence of glacial advancement and retreat during the Pleistocene Age, which concluded approximately 10,000 years ago (Anfinson, 2003). The most recent major glacial event, the Wisconsin Glaciation from 85,000 to 10,800 years ago, describes the general advancement of the North American ice sheet during

which several minor glacial retreats occurred (Anfinson, 2003). A minor retreat of the Des Moines Lobe and Superior Lobe north of the continental divide resulted in the formation of Lake Agassiz and Lake Duluth (Anfinson, 2003). Sediment free drainage from the two lakes began to incise river valleys to the south, forming the River Warren, a precursor to the Mississippi River named after G. K. Warren, first commander of the St. Paul District, Corps of Engineers (Anfinson, 2003). After one final advancement, glacial retreat north of the continental divide established active downcutting through the Mississippi River Valley until the conclusion of the Pleistocene (Anfinson, 2003).

From approximately 30,000 years ago to the present sea level establishment approximately 6,000 years ago, the Mississippi River Valley was defined by frequent flooding of shallow braided streams carrying large amounts of sediment throughout the river valley (Fisk, 1945). Aggradation in the valley ensued, lowered valley slopes, and decreased the sediment load in the tributary streams (Anfinson, 2003; Fisk, 1945). Decreased load in the streams led to the formation of a main channel and the current position of the Mississippi River was established approximately 2,000 years ago (Fisk, 1945).

2.1 Middle Mississippi River

The Mississippi River is divided into two main geographic and geologic sections: Upper and Lower (Mac et al., 1998). The Upper portion runs from St. Anthony Falls in Minneapolis, MN to the Ohio River in Cairo, IL and stretches 1,462 km (Mac et al., 1998). The Lower portion runs from the Ohio River to the Gulf of Mexico and stretches 2,243 km (Mac et al., 1998).

These two sections are geomorphically diverse with bluffs defining much of the Upper portion and alluvial and coastal plain sediments defining the Lower portion (Fisk, 1945). A 314 km long segment of the Upper Mississippi River, located between the tail of the Illinois River north of St. Louis, MO and the mouth of the Ohio River, is oftentimes separately identified as the Middle Mississippi River (MMR) (Mac, 1998). USACE, St. Louis district monitors 89 levees, over 700 miles long, within the Middle Mississippi River watershed.

The MMR alluvial valley ranges from 3 to 10 miles wide with a floodplain east of the river in Illinois and resistant rock bluffs on the western side in Missouri (USACE, 1956b). Alluvial deposit depth averages 125 feet and ranges from 75 to 200 feet thick (USACE, 1956b). The valley consists of a pervious substratum and a semi-impervious top stratum, both of varying thickness (USACE, 1956b). The upward gradation to finer-grained sediments is consistent with the river's evolution from a braided stream type to a meandering stream type (USACE, 1956b).

Fisk (1945) categorizes the valley into three types of deposition: braided stream deposition, flood basin deposition, and meandering stream deposition. Braided streams in the MMR valley resulted in the deposition of poorly-sorted silts and sands with small amounts of clay (Fisk, 1945; USACE, 1956b). Evidence of braided streams can still be observed at the north end of the MMR valley but becomes buried under the floodplain further south (Fisk, 1945).

Flood basin deposition is characterized by almost no variation in elevation, also known as a lack of local surface relief (Fisk, 1945). Deposition during this type of environment occurs during flooding in which floodwaters spread far and wide through ancient channels, formed by braided stream topography during the evolution of the MMR (Fisk, 1945). Extended deposition may overtake trees and other plant life, elevating the organic content in the deposit and forming typical

“buckshot” clays by intermittent cycles of oxidation (Fisk, 1945). This type of environment is called a “backswamp” environment and is characterized by silts, silty clays, and clays (Fisk, 1945). For the alluvial valley, the thickness of “backswamp” deposits increases with increasing proximity to the Gulf of Mexico (Fisk, 1945). Natural levees are typically well drained and largely made up of fine sandy-silts and silty clays in the MMR valley (Fisk, 1945).

2.2 Field Areas

Data on two levee districts, Prairie du Rocher (PDR) Levee District and Ft. Chartres (FTC) Levee District, compiled by Jon Wilson (2003) were used in model building and selection for this research. A third levee district, East Cape Girardeau (ECG) Levee District, was characterized for blind testing on selected models. PDR and FTC lie adjacent to one another in Prairie du Rocher, IL (see Figure 1). ECG lies approximately 80 miles to the south in East Cape Girardeau, IL.

The field areas are geographically and geologically similar, and located along the MMR (USACE, 1956a). The width of the MMR is approximately the same in all three districts, 0.35 miles, and in general, a semi-impervious top stratum of variable thickness overlays a pervious substratum of variable grain size (Fisk, 1941). They adhere to previously mentioned generalizations on the evolution of the MMR valley. However, there is variation in more specific

geologic variables such as top stratum thickness for any given levee segment, pervious substratum grain size for any given levee segment, and location and orientation of swales for any given levee district. Also, the location of relief wells, landside seepage berms, and riverside seepage berms varies.

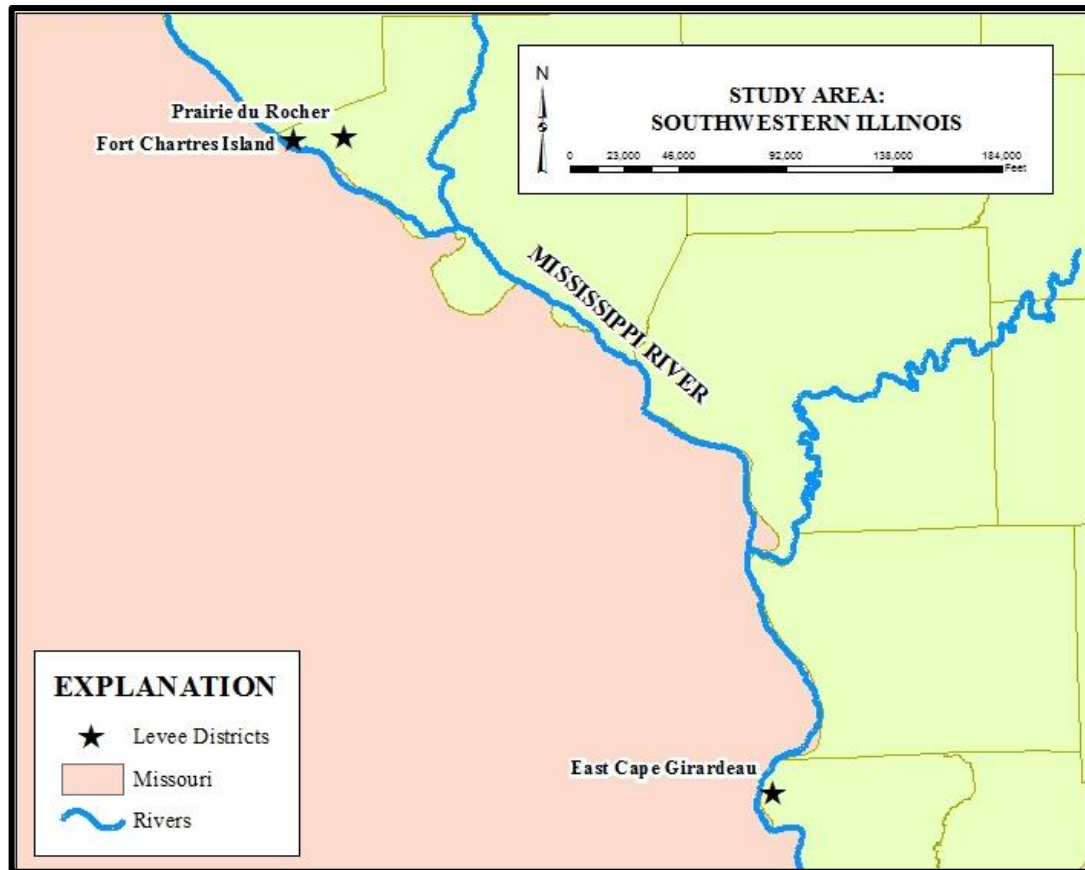


Figure 1. General study area. Southwestern Illinois. Field area locations for Prairie du Rocher, Fort Chartres, and East Cape Girardeau are identified by black stars.

CHAPTER THREE

Piping Mechanics and Wilson's Efforts to Predict Levee Potential

The USACE places great importance on the specific differences between general underseepage and piping. Underseepage, defined as the flow or seepage of water from the riverside to the landside under the levee, may be normal or expected at some locations along a levee (Fisk, 1945; Turnbull and Mansur, 1959). Whereas, piping, a form of underseepage, is not “confirmed” until the formation of sand boils are observed (USACE, 1956b). Occurrences of non-localized, typical underseepage may be expected during times of highwater and pose no threat to levee stability (USACE, 2000). By definition, piping weakens the levee's foundation by creating preferential pathways and scouring grains from the substratum, weakening the load-bearing strata (Figure 2) (USACE, 2000). When piping is allowed to continue unabated, preferential pathways may widen to form crevasses resulting in extreme levee failure.

3.1 Mechanics of Piping

Many studies, notably Fisk (1945), Turnbull and Mansur (1959), and Kolb (1975), have listed two geologic controls necessary for piping development and sand boil formation during flooding: a pervious substratum overlain by a semi- to impervious top stratum. When a levee segment meeting these “conducive environment” controls is subjected to water loading, the force exerted by the weight of water on the riverside of the levee can be transferred through the

pervious substratum to the landside, resulting in a hydrostatic imbalance between strata and the surface landside of the levee (Turnbull and Mansur, 1959; USACE, 2000).

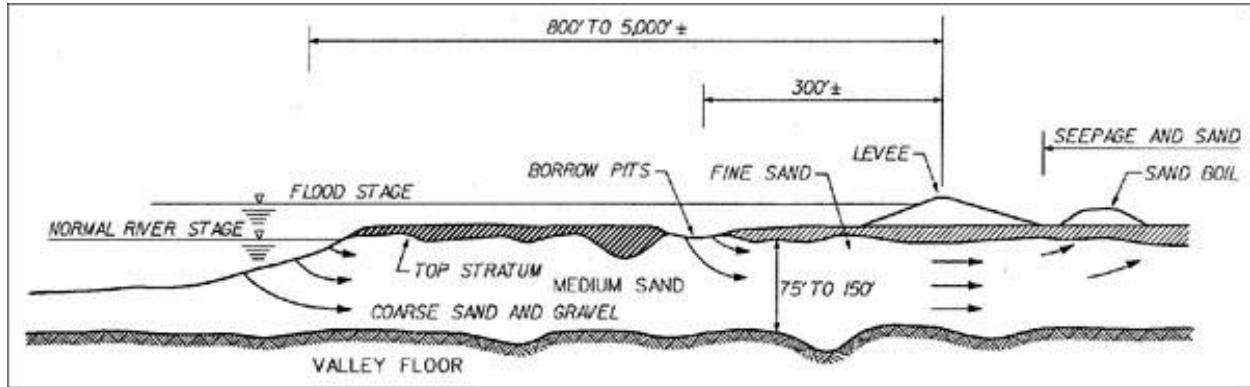


Figure 2. Depiction of underseepage, piping, and sand boil development through a cross-section of a levee. Underseepage and piping is represented by flow lines from the river and borrow pits, located on the left, under the levee, located on the right. Fisk (1945)

If large hydrostatic pressures, historically known as an artesian head, are allowed to develop in the pervious substratum, , rupture will eventually occur (Fisk, 1945). Rupturing of the top stratum may be spread through an entire section (e.g. non-localized underseepage), or channeled (e.g. localized underseepage or piping) (Fisk, 1945; USACE, 1956b; Kolb, 1975). The levee's hydrostatic gradient may only become "critical" at localized points while the average gradient remains well below critical at specific locations of the top stratum due to thin or weak spots (Mansur et al., 2000).

A critical gradient, i_c , is calculated to quantify the maximum level of hydrostatic imbalance allowable prior to rupture, by finding the ratio of the submerged or buoyant unit weight of soil, $\tilde{\alpha}'$, comprising the top stratum, to the unit weight of water, $\tilde{\alpha}_w$, where G_s is the specific gravity of soil solids and e is the void ratio (1) (USACE, 2000).

$$(1). i_c = \tilde{a}' / \tilde{a}_w = (G_s - 1) / (1 + e)$$

It is defined as “the gradient required to cause boils or heaving (flotation) of the landside top stratum” (USACE, 2000).

This value is surpassed when a flood’s head reaches or exceeds a height equating to a pressure force larger than the weight force of saturated soil landside of the levee (USACE, 2000; Mansur et al., 2000). If the force created by the submerged weight of soil is heavier than the force created by the hydrostatic imbalance, piping will not occur. This concept is very informative when discussing the mechanics of piping and can also be used when considering influential factors other than “conductive environment”. It will be included in the general databases for PDR and ECG; however, i_c , is currently not available for FTC. Several secondary factors, mentioned by Fisk (1945), Turnbull and Mansur (1959), Kolb (1975), and USACE (1956b), and the representative values used in this research will be discussed in Section 3.1.2.

3.1.1 Piping Preventative Measures

Installation of prevention measures designed for piping abatement began during the initial investigations in 1941 (Turnbull and Mansur, 1959). Starting in 1937, seepage berms were designed and installed along the Lower Mississippi River and in 1950, relief wells were installed at Trotters 54, Mississippi; the purposes of these installations was to study their effectiveness at reducing the occurrence of piping (Turnbull and Mansur, 1959).

Measures taken to prevent piping target the control of scouring and the minimization of excess hydrostatic pressure landside of the levee (USACE, 1956a). These measures are required

when values of h_0 , are expected to equal or exceed h_c (estimated at $0.75 \cdot z_t$), where h_0 is a factor of the raw head during a flood event and z_t equals the transformed confining layer thickness (USACE, 1956a). These values will be discussed more in subsections 3.1.2.2 and 3.1.2.3.

Techniques include the installation of cutoff trenches, pervious toe trenches, riverside impervious blankets, landside seepage berms, and pressure relief wells (USACE, 2000). Some of these measures may be expensive to install and are only temporarily effective. Local factors, such as levee foundation characteristics, cost of installation and maintenance, productive lifetime expectancy, levee constraints (e.g. length and width of the area landside of the levee), and dumping of seepage water, will determine the possible preventative measures to be taken in that location (USACE, 1956a).

3.1.1.1 Cutoff Trenches

USACE (2000) states a cutoff trench, also referred to as a “cutoff”, has the “most positive” results in eliminating seepage. In general, an excavated trench, below the location of a future or present levee susceptible to piping, is backfilled with slurry or compacted earth with a low permeability. The trench must be excavated through 95 percent or more of the pervious substratum and, in locations where the pervious substratum is extensively thick, (e.g. exceeding 12.2 m), cutoffs are not monetarily feasible (USACE, 2000).

If excavation reaches below the water table, dewatering of the levee foundation must be implemented (USACE, 2000). USACE (2000) suggests following dewatering system design guidelines described in a technical manual, TM 5-818-5, published by the Joint Departments of the Army: the Air Force, and the Navy (USACE, 2000; HDAAF-USA, 1983). Dewatering

system can be expensive and other preventative measures should be used if local conditions permit (USACE, 2000).

3.1.1.2 Pervious Toe Trenches

USACE (2000) suggests coupling a relief well system with pervious toe trenches when considering the case of an extensively thick pervious substratum. Shallow underseepage may be directed towards a “partially penetrating” trench excavated at or near the levee toe, which is designed to specifically protect the area around the toe trench (USACE, 2000). The release of hydrostatic pressure in shallow portions of the pervious substratum through toe trenches and the release of hydrostatic pressure in deep portions of the pervious substratum through the installation of relief well systems can be very effective if local conditions permit (USACE, 2000).

3.1.1.3 Riverside Impervious Blankets

For exposed portions of the substratum riverside of the levee, riverside impervious blankets may be installed to inhibit the development of hydrostatic pressure imbalances in the subsurface landside of the levee (USACE, 2000). Riverside impervious blankets, also referred to as “blankets”, may be placed in suspect areas to reduce the possibility of infiltration into the substratum (USACE, 1956a). This, in turn, will decrease seepage flow and prohibit the development of excess hydrostatic pressure landside of the levee (USACE, 1956a).

This type of technique is most useful for limited areas where weak or thin top stratum provides potential infiltration into the pervious substratum riverside of the levee (USACE, 2000). Factors, such as permeability, thickness, and length of the blanket, as well as its distance to the levee riverside toe, can control the overall performance of the blanket (USACE, 2000).

3.1.1.4 Landside Seepage Berms

Landside seepage berms, also known as “berms”, attempt to increase the thickness of the top stratum enough to withstand even the highest headwaters (USACE, 1956a). For this method to be successful, the weight of the impervious top stratum coupled with the weight of the berm must be large enough to overcome the uplift force (i.e. hydrostatic pressure imbalance) exerted by floodwaters in the substratum (USACE, 1956a). Also, the berm must extend lengthwise to a predefined point where the critical gradient no longer exists (USACE, 1956a; USACE, 2000).

Berms are easily recognizable in aerial photography and satellite imagery. They are frequently maintained and specific guidelines accompany the development of the various types of berms (USACE, 2000). Four unique types of berms, impervious berms, semipervious berms, sand berms, and free-draining berms, are chosen for construction dependent upon availability of space and fill material landside of the levee, as well as local economic constraints (USACE, 2000). For more information regarding construction guidelines, please reference USACE (2000).

3.1.1.5 Pressure Relief Wells

Pressure relief wells, usually referred to as “relief wells”, may be installed landside of the levee in areas where seepage has proven to be problematic, or where piping has occurred (USACE, 2000). The intent is not to prohibit the development of excess hydrostatic pressure, but instead, to control and direct seepage flow to an exposed surface landside of the levee (USACE, 1956a). This alleviates pressure buildup and reduces the possibility of piping and the formation of sand boils (USACE, 1956a). The wells must sufficiently penetrate the substratum and be spaced closely enough to fully reduce hydrostatic pressures between the wells (USACE, 2000).

Construction of relief wells is indicated where the pervious substratum is too thick for cutoffs or toe drains or where space landside of the levee is limited and berms are ineffective (USACE, 2000). Well screens can be cumbersome and maintenance is frequent. Loss in efficiency will occur with time due to clogging, bacteria growth, or carbonate incrustation (USACE, 2000). Maintenance of discharge disposal is also necessary for successful prevention (USACE, 2000).

3.1.2 Secondary Factors of Piping

Numerous secondary factors relevant to piping development have been suggested and are still considered when conducting studies on piping and sand boil development (Fisk, 1945; Turnbull and Mansur, 1959; Mansur et al., 2000). For example, subsurface erosion will not commence until turbulent flow is reached; however, the rate of flow through the substratum depends upon height of the floodwaters and soil characteristics of the substratum (Mansur et al., 2000). Other examples of secondary factors include characteristics of the riverside top stratum; source, velocity, and measure of seepage concentration; “seepage carrying capacity” of the

substratum; natural cavities such as shrinkage cracks, decay of roots, uprooting of trees, animal burrows, crayfish holes, etc.; or man-made holes such as drainage ditches, post holes, and seismic shot holes (Turnbull and Mansur, 1959; Mansur et al., 2000).

Some secondary factors have been suggested to play a more influential role in piping occurrence than others. Research, beginning with the investigations initially conducted by Fisk (1941), has shown a strong correlation between a secondary geologic characteristic, termed “unfavorable geologic conditions”, and the development of piping and location of sand boils (Fisk, 1945; Kolb, 1975). The types of “unfavorable geologic conditions” and their variables are further discussed in the following section. Other influential factors that will be considered in this study are top stratum thickness, perviousness of substratum, severity of flood, piping control measures, and previous piping events. These variables will be defined, discussed, and quantified in the following sections.

3.1.2.1 Unfavorable Geologic Condition

Unfavorable geologic condition refers to impermeable formations that impede or restrict flow pathways landside of the levee, which can lead to localized underseepage and piping. Kolb (1975) considered the influence of point bar deposits, natural levee deposits, backswamp deposits, and channel-fill deposits on piping and sand boil formation. He found that, generally, point bar deposits are the only regularly occurring formations thin or permeable enough to affect piping (Kolb, 1975). In rare cases, natural levee deposits can result in piping; however, they must be directly overlain by the constructed levee and their ancient crevasse channels must be oriented with the lateral flow of water through the subsurface (Kolb, 1975). Backswamp deposits and

channel-fill deposits are normally too thick or impermeable for lateral flow but can influence piping under special circumstances (Kolb, 1975).

Given Kolb's findings, this research will use data on the presence and orientation of swales, \bar{u} , to determine the influence of unfavorable geologic condition on piping. Swale and ridge complexes typically form from point bars deposited on the convex side of bends along meandering river systems (Fisk, 1945; Kolb 1976). Approximately 60% of the Mississippi River overlies point bars or other accretion deposits (Kolb, 1975). Ridges generally comprise silty sand or sand and are relatively permeable, whereas swales comprise silt and clay and are relatively impermeable (Kolb, 1975).

Fisk (1945) studied the influence of swale and ridge complexes on subsurface flow by comparing water levels in piezometers installed near swales to water levels in piezometers installed near ridges. He found that pressures formed by elevated headwaters near ridges translated through the substratum more rapidly than elevated headwaters near swales (Fisk, 1945). Ridge formations subjected to hydrostatic imbalances allow for non-localized, non-problematic seepage (USACE, 1956a).

Swale formations subjected to hydrostatic imbalances may not influence piping at all, in which case seepage will be similar to that caused by ridge formations under hydrostatic imbalance (Kolb, 1975). Many times, however, swale formations subjected to hydrostatic imbalances result in non-localized, problematic seepage and/or localized seepage, piping and sand boil formation (Fisk, 1945; Turnbull and Mansur, 1959; Kolb, 1975). In these cases, piping will concentrate along the swale's adjacent ridge (Kolb, 1975).

The orientation of swales, measured in relation to the direction of river flow along a levee, correlate with the number of available flow pathways landside of the levee. Kolb (1975) found that sand boil formation is least developed along obtusely oriented swales, where flow pathways in the direction of river flow are relatively unbound, and most developed along acutely oriented swales, where flow pathways in the direction of river flow are significantly reduced. This is because the availability of flow pathways directly correlates with substratum pressure dispersion and the reduction of hydrostatic imbalances in the subsurface. Piping and sand boil formation can develop in ridges adjacent to normally oriented swales but distribution is random (Kolb, 1975). The general database for all levee districts in this research has a categorical binary code representing the influence of swale orientation, \bar{u} , where acutely oriented swales are assigned a value of 1, “highly influential” and obtusely oriented swales are assigned a value of 0, “not influential”. Initially, Wilson (2003) used a continuous scale from 0 to 1 which supplied a more specific quantification of swale orientation; however, this scale was replaced by the binary code

3.1.2.2 Effectiveness of Top stratum

A top stratum with certain characteristics (e.g. specific values of thickness, variation, perviousness) can effectively prevent piping (Fisk, 1945; Mansur et al., 1956; Kolb, 1975). Thin or weak spots in the top stratum have already been identified in Section 3.1 as determining features in piping and sand boil formation that occurs when i_c is met or exceeded. A top stratum may be characterized based on either its horizontal or vertical extent oriented with the levee. For example, USACE (2000) suggests using the length of top stratum landside of the levee, L_3 , to

determine several factors including hydraulic heads defined in underseepage analysis. However, horizontal extent was not available for analysis in this study.

Mansur, Kaufman, and Schultz (1956) attempted to provide categories for the influence of a top stratum's vertical extent or thickness, z , on piping. While the classifications are not widely used due to quantifiable limitations, listing them does provide some valuable insight on the effectiveness of a top stratum on piping: (a) no significant topstratm; (b) top stratum of insufficient z to withstand the hydrostatic pressures that tend to develop; (c) top stratum of sufficient z to withstand any hydrostatic pressure that may develop during the maximum design flood.

The top stratum along the Mississippi River from Dupu, Illinois, to Gale, Illinois, was identified by Mansur et al. (2000) as category (b), the most potentially dangerous situation for the development of piping (USACE, 1956b; Mansur, Kaufman, and Schultz, 1956; Kolb, 1975). Category (b) has a z value high enough to outweigh relatively large hydrostatic pressures developed during “moderate” flooding, but does not reach the z of category (c), leading to rupture from hydrostatic pressures developed by more “considerable” headwaters (Mansur et al., 2000). The outcome is an excessive build-up of substratum pressures, resulting in sudden rupture of the top stratum in localized points (USACE, 1956b). Category (a) prohibits any development of hydrostatic pressures by permitting intermittent flow of non-localized underseepage to the surface; category (c) prohibits rupture of the top stratum under any conditions, eliminating the possibility of piping (Kolb, 1975; Mansur, Kaufman, and Schultz, 1956). “Moderate” and “considerable” are not quantified and are used only on a relative basis.

USACE (2000) suggests using a transformed confining layer thickness, z_t , to determine quantified influential thickness values. This is because vertical permeability of the top stratum,

k_b , is rarely uniform and is normally a composite of n number of soil types with a distinct vertical permeability, k_n , and thickness, z_n (USACE, 2000). Soils types can be reasonably weighted according to clay content in lieu of k_n measurements if z_n is known (e.g., z_t of clayey silt is greater than z_t of silty sand) (USACE, 2000). Applicable layers for z_t are all strata above the base of the least pervious stratum and underlying more pervious top strata (USACE, 2000).

3.1.2.3 Susceptibility of the Substratum

Underseepage cannot develop unless some portion of the substratum is exposed, to flood water riverside of the levee (Mansur et al., 2000; Kolb, 1975). Infiltration and flow through the exposed portion is limited by the extent of exposure and perviousness of the substratum.

Perviousness is a relative term used to distinguish soil types that allow water to flow relatively easily through their matrix from soil types that hinder (e.g., silty sand) or resist flow (e.g., silty clay) (Ranjan, 2005).

Currently, the extent of exposure of the substratum riverside of the levees along the MMR is not well documented and data are largely unavailable. However, riverside borrow pits, \check{R} , and landside ditches, \check{L} , are a known cause of substratum exposure riverside and landside of the levee, and can be easily identified through aerial photography and LiDAR data made available by USGS (Mansur et al., 2000). The variable \check{R} will be used to determine the influence of riverside borrow pits on infiltration by a binary code of present where $\check{R} = 1$, and not present, where $\check{R} = 0$. The variable \check{L} will be used to determine the influence of landside ditches on infiltration by a similar binary code.

The perviousness of a stratum has three general categories: pervious, semi- and impervious. These categories are ordinal and not quantified (Ranjan, 2005). Permeability describes the ease with which water flows through a soil and is calculated on a continuous scale for all soils through laboratory analysis (Shepherd, 1989). Increasing permeability values in the substratum directly correlate with increasing rates of flow into the subsurface and increase the probability of turbulent flow and piping (Wilson, 2003; Mansur et al., 2000). The most accurate values for intrinsic permeability, k , are analytically determined by either field pumping/injection test or the use of a permeameter on samples in the laboratory (Shepherd, 1989).

The USACE, St. Louis District, geologic investigations of the MMR resulted in boring log data for all USACE monitored levee districts along the MMR, with the exception of Kaskaskia Levee District (due to its location on the western side of the river). Permeability values were not measured in these investigations but effective grain size, d_{10} , a justified proxy measure of permeability, was determined. Dunn (1980) defines d_{10} for a given soil sample as the determined particle size for which 10% of the sample by weight is smaller than that size. Studies have shown a general relationship between permeability values and grain size of the strata (Shepherd, 1988). Other relationships pertain to textural maturity of grains, depositional environments of grains (such as dune, beach, and river), and maximum grain size of strata as opposed to average grain size (Shepherd, 1989), but these relationships were not used for this paper.

3.1.2.4 Severity of Flood

The severity of a flood event controls the extent of piping and sand boil formation (Fisk, 1945). Severity can be analytically represented in several ways (e.g., velocity (m s^{-1}), discharge ($\text{m}^3 \text{s}^{-1}$), flux ($\text{kg s}^{-1} \text{m}^{-2}$)). With increasing severity comes the greater probability of elevated substratum pressures larger than the critical gradient, leading to top stratum rupture and the possible development of piping (USACE, 2000; Fisk, 1945). An analytical measure of severity that controls piping extent is the mass of flood water (kg), and the resulting downward force (N) at levee segments that meet the “conductive environment” (Fisk, 1945).

Flood water directly atop exposed pervious strata will result in infiltration if its weight (force) is greater than the strata’s resistance to flow. Section 3.12.3 explained that the perviousness of the exposed strata at a given point (e.g. measured effective grain size), can represent the strata’s resistance to infiltration. Under a similar concept, the weight (force) may be estimated by considering the observed net head elevation, the height of water on the riverside measured from the natural ground surface on the landside, for a specific location along the levee (Wilson, 2003; USACE, 2000).

Net head, **H**, is a controlling factor of excess hydrostatic head beneath the top stratum (USACE, 2000). Net head is directly measured by USACE at predefined waypoints along maintained levee districts during flood events (USACE, 2000). USACE (2000) identifies two helpful quantified variables related to **H** on a levee: (a) excess hydrostatic head; (b) head beneath the top stratum at a distance x .

Excess hydrostatic head, **h₀**, is related to **H**, the dimensions of the levee and foundation, the permeability of the foundation, and top stratum conditions (i.e. perviousness, length) on the both sides of the levee. The head beneath the top stratum at a distance x , **h_x**, is related to **H** and the distance x but is most commonly related to **h₀** (USACE, 2000). This is because **h_x** as a

function of h_0 depends only on the type and thickness of the top stratum and pervious foundation landward of the levee (USACE, 2000). This research will include the independent variables H and h_0 into the general databases for PDR and ECG. The FTC flood reports did not include H , eliminating the possibility of calculating h_0 . To reduce redundancy, the variable h_x will not be included because of its dependence on h_0 .

3.1.2.5 Effectiveness of Piping Prevention Measures

The installation of piping prevention measures should reduce a levee's potential for piping under "conductive environment" (Turnbull and Mansur, 1959; Kolb, 1975; USACE, 2000; Mansur et al., 2000). USACE (2000) lists several measures available for piping prevention: cutoff trenches, riverside impervious blankets, landside seepage berms, pervious toe trenches, and pressure relief wells. Installation is dependent upon established need, available funding, and type of geologic conditions in the area (USACE, 2000).

Wilson (2003) quantified prevention measures in the levee districts by assigning a binary code for the presence or absence of the variable, where 1 equals the presence and 0 equals absence. Landside seepage berms, B , and relief wells, R , were located for PDR and FTC districts using aerial photography and LiDAR data supplied by the USGS.

3.1.2.6 Role of Previous Piping Events

Preferential channels created during piping will remain intact following the flood event unless some process (e.g., levee failure or maintenance) disturbs the soil. Piping through these

previously established channels may require a less severe flood for development in subsequent events. Previous piping events, **P**, were shown to have a highly significant relationship to piping potential, $P - \text{value} < 0.0001$, in the research conducted by Wilson (2003).

3.2 Wilson's Efforts

Wilson (2003) employed several techniques during model building and selection. PDR and FTC were divided into equal segments for which independent variables, based off suggestions by previous studies, were obtained and interpolated, if applicable, to estimate the most representative value for that segment. He used piping observations from the Great Flood of 1993 and a lesser known Mississippi River flood of 1995 as dependent variables in the regression analyses. Eight total models were developed in XLSTAT, a statistical software suite for Microsoft Excel. Four were created using linear regression, two for PDR and two for FTC, and four were created using logistic regression, two for PDR and two for FTC. His methods for data acquisition and interpolation, and model building and selection, as well as his final results and conclusions are discussed in the following sections.

3.2.1 Compilation of Databases

Wilson divided PDR into 349 levee reaches and FTC into 278 levee reaches. Each reach is 250 feet long and is associated with one representative value for each variable. Variables were chosen based off of analyses by Fisk (1945), Turnbull (1959), Kolb (1975), and USACE (2000). Boring data were limited in some areas and interpolation techniques were used to correct for this limitation. Interpolated or analyzed variables₇ are discussed in Section 3.2.1.1₂ whereas variables

that required no interpolation or analysis are discussed in 3.2.1.2. Table 1 lists all variables used in the model building. The last three variables listed in Table 1 are only available for PDR. Available data on dependent and independent variables included boring log data, aerial photography, LiDAR, and underseepage analysis conducted by USACE St. Louis District.

Table 1. Description of independent variables used for regression analysis.

Independent Variables: Wilson (2003) Database		
#	Variable	Quantitative Description for 250 ft Levee Segment
1	Presence and orientation of Swales (\tilde{u})	Presence and orientation of swales; binary code
2	Transformed Confining Layer Thickness (z_t)	Minimum interpolated thickness of confining layer; continuous scale
3	Riverside Borrow Pits (\check{R})	Presence of borrow pit located on riverside of the levee; binary code
4	Effective Grain Size (d_{10})	Average interpolated effective grain size of the pervious substratum unit; continuous scale
5	Relief Wells (\acute{R})	Presence of relief wells; binary code
6	Landside Ditches (\acute{L})	Presence of borrow pit or ditches on the landside of the levee; binary data
7	Berms (β)	Presence of landside seepage berms; binary code
8	Net Head (H)	Elevation difference between flood head and surface elevation; calculated by USACE, St. Louis; continuous scale
9	Excess Hydrostatic Head (h_0)	Function of levee dimensions, dimensions and permeability of foundation, and topstratum conditions; calculated by USACE, St. Louis; continuous scale
10	Critical Gradient (i_c)	Ratio of submerged or buoyant unit weight of soil comprising the topstratum unit to the weight of water; calculated by USACE, St. Louis; continuous variable

Fort Chartres data is restricted to variables 1-8. All variables listed are available for Prairie du Rocher.

3.2.1.1 Analyzed Variables

Analyzed variables required either quantification through interpolation or interpretation of geologic conditions conducted by Wilson or by the original data compiler (i.e. USACE, St. Louis District). Transformed confining layer thickness, z_t , effective aquifer grain size, d_{10} , and

presence and orientation of swales, \bar{u} , were interpolated or analyzed by Wilson (2003). Net head, H , excess hydrostatic head, h_0 , and critical gradient, i_c , values were transferred from USACE, St. Louis District, fact sheets on levee performance east of the Middle Mississippi River following flood events in 1973 and 1993. However, these variables are still considered analyzed variables and the methods of determining such variables are discussed below.

Transformed Confining Layer Thickness

Transformed confining layer thickness, z_t , values of each boring log were calculated using empirical criteria established by Turnbull and Mansur (1959). The transformed confining layer thickness value gives a more accurate representation of the substrata's resistance to flow for strata of varying clay and silt content. Once z_t was determined from each boring log, Wilson used ordinary kriging to determine the minimum value of confining layer thickness along each 250-foot reach. Kriging analysis takes a regional variable and estimates the value for the variable at a specific location using a semivariogram or covariogram (Davis, 2002).

A regional variable is neither completely deterministic nor truly random, which is characteristic of many geological variables (Davis, 2002). It possesses spatial structure. A semivariogram is a graph of the semivariance of a variable, which finds a rate of change for the regionalized variable at a specific orientation. Covariograms are a plot of the covariances of all data points a specified distance apart (Davis, 2002).

Wilson determined the spatial structure for z_t by creating a variogram using Variowin 2.2 software created by Pannatier (1996). Wilson then used geostatistical interpolation software available in ArcGIS 8.1 to input the transformed confining layer thickness data from each of the 218 boring sample locations. Wilson used a circular neighborhood shape with 10 neighbors for the

spatial structure. He determined this shape and size based on the orientation and spacing of given sample locations.

Permeability of the Substratum

Wilson used the correlation between d_{10} and permeability to describe the aquifer or pervious substratum in which piping occurred. Only 78 sample locations were available for interpolation and variography was unsuccessful in determining spatial structure. Inverse distance weighting was used instead. Inverse distance weighting is a geostatistical interpolation method that does not require spatial structure (Davis, 2002). Once again, Wilson chose a circular neighborhood shape with 10 neighbors and a maximum search radius of 16,530 feet because of the sparseness of sample locations.

Presence and Orientation of Swales

Initially, Wilson used a range of values from 0 to 1 to describe \bar{u} . A continuous scale was developed based upon the orientation of the swales in a section. For example, a value of 0.5 corresponded with the intersection of a swell and levee at an angle less than 90° . A value of 0.7 corresponded with a swale that perpendicularly intersected the levee. However, this procedure was replaced by a simpler binary description of the variable with 0 being no presence of swales and 1 being presence of swales. Wilson chose to use any value equal to 0.7 and above as the presence of swales and any value below 0.7 as no presence of swales.

Net Head, Excess Hydrostatic Head, Critical Gradient

Net head, H , excess hydrostatic head, h_0 , and critical gradient, i_c , were measured by USACE, St. Louis District, during the 1973 and 1993 flood events. The variable's definitions and

their influence on piping occurrence have been previously discussed in Section 3.1.2. Wilson used linear interpolation to estimate values between data points.

3.2.1.2 Direct Variables

Direct variables are variables that were directly observable without the use of formulae or interpretation of geologic features. The presence of riverside borrow pits, presence of landside ditches, presence of relief wells, presence of landside seepage berms, and previous piping occurrences from 1993 and 1995 are all considered direct variables. These features are quantified as a binary code and do not depend upon orientation. Each observation was assigned a 1 if there was any presence of the feature (e.g., presence of relief wells, along the 250-foot reach) and a 0 if there was no presence. Riverside borrow pits, landside ditches, relief wells, and landside seepage berms, were identified using aerial photography, LiDAR, and on-site visits to both levee districts.

3.2.2 Model Building

Once each data point and its associated variables were determined, Wilson conducted a linear regression analysis and logistic regression analysis for piping in 1993 and piping in 1995 at the PDR and FTC. Models using 1993 piping as the dependent variable can indicate where piping could occur without any knowledge of piping in the area. Models using 1995 piping as the dependent variable include previous piping, **P**, as an independent variable database. A total of eight models was created. Tables 2 through 5 show the final models for the PDR database and both dependent variables. A more detailed description of the models and the methods used are discussed below.

Table 2. Description of PDR-93 Linear model.

Model Name	Model Test	Model Significance	Independent Variables	Coefficient Value	P-value
PDR-93 Linear	R^2	0.082	Intercept	-0.2107	0.0096
			Transformed Confining Layer Thickness, Z_b	-0.0031	0.2276
	F-significance	1.84×10^{-6}	Effective Aquifer Grain Size, D_{10}	1.6872	0.0003
			Presence and Orientation of Swales, \tilde{u}	0.1067	0.0003

Coefficient value corresponds with the value of β in general form of equation.

Table 3. Description of PDR-93 Logit model.

Model Name	Model Test	Model Significance	Independent Variables	Coefficient Value	P-value
PDR-93 Logit	Mc Fadden psuedo- R^2	0.165	Intercept	-7.7432	<0.0001
			Transformed Confining Layer Thickness, Z_b	-0.0526	0.4630
			Effective Aquifer Grain Size, D_{10}	26.9941	0.0010
			Presence and Orientation of Swales, \tilde{u}	1.9084	0.0010

Coefficient value corresponds with the value of β in general form of equation.

Table 4. Description of PDR-95 Linear model.

Model Name	Model Test	Model Significance	Independent Variables	Coefficient Value	P-value
PDR-95 Linear	R^2	0.147	Intercept	0.0841	0.0004
	F-significance	9.45×10^{-13}	Piping Potential Values from PDR-93 Linear	0.3253	0.1921
			Piping Locations in 1993	0.4938	1.86×10^{-11}

Coefficient value corresponds with the value of β in general form of equation.

Table 5. Description of PDR-95 Logit model.

Model Name	Model Test	Model Significance	Independent Variables	Coefficient Value	P-value
PDR-95 Logit	Mc Fadden psuedo-R ²	0.128	Intercept	-2.3532	< 0.0001
			Piping Potential Values from PDR-93 Logit	2.9881	0.0880
			Piping Locations in 1993	2.3851	< 0.0001

Coefficient value corresponds with the value of β in general form of equation.

Wilson created four models using 1993 piping as the dependent variable: PDR-93 Logit, PDR-93 Linear, FTC-93 Logit, and FTC-93 Linear. Logit refers to the models built using logistic regression. Linear refers to the models built using linear regression. A stepwise regression backward-elimination procedure, defined by Le (1998), was used for both linear and logistic models. All independent variables from each database were input into the respective regression model (PDR or FTC; linear or logistic) and systematically eliminated based on the P-values of the variables in the model.

If the variable was not significant at a 95% confidence level, equivalent to a P-value < 0.05, or had a non-sensible sign, either positive when logically the sign should be negative or vice versa, on the coefficient associated with it, it was removed from further variations of the model. For example, the presence of relief wells was found to have a P-value of 0.644 and a positive sign on the coefficient, meaning the null hypothesis that relief wells are independent of piping can not be rejected without a 34.6% chance of it actually being true, or 34.6% significant. Logically, the presence of relief wells should decrease the possibility of sand boil formation by redirecting water flow, giving this variable an inverse relationship with piping, which directly conflicts with a

positive coefficient in the model. These values do not conform to the model building process and the variable was eliminated.

Transformed confining layer thickness is one exception to Wilson's model building process. While the variable had a sensible sign, it was not significant within a P-value < 0.05 . Most literature lists confining layer thickness as an influential variable to piping (Fisk, 1945; Turnbull, 1959; Kolb, 1975; etc.), therefore Wilson chose to retain the variable despite the loss in validity of the model.

Four models using 1995 piping as the dependent variable were created under the same categorical concept: PDR-95 Logit, PDR-95 Linear, FTC-95 Logit, and FTC-95 Linear. However, models used the original independent variables in each dataset plus 1993 piping as the independent variable \mathbf{P} . Instead of using backward-elimination as in step 1, Wilson chose to use just two independent variables for these models: (1) known piping occurrences in 1993, (2) $P(\pi=1)$ from the respective indicative model. For example, PDR-95 Logit uses 1993 piping in PDR and $P(\pi=1)$ values from PDR-93 Logit.

3.2.3 Model Thresholds

Wilson determined thresholds of high-, medium, and low-potential for piping by analyzing the distribution of Y, for linear regression models, or $P(\pi=1)$, for logistic regression models. Different thresholds were set for 1993 logistic models, 1993 linear models, 1995 logistic models, and 1995 linear models and were applied to both PDR and FTC datasets. Table 6 lists the performance of all final models applied to their respective thresholds.

Table 6. Performance of Wilson (2003) models under his applied thresholds.

Regression Type	Database Type	Threshold Model Values	Model Name	Applied Threshold	% Piped for Applied Thresholds	
Linear	Pure	High ≥ 0.1825	PDR-93 Linear	High	21.40%	
		Medium $\in (-1.0000, 0.1825)$		Medium	31.10%	
				Low	8.10%	
		Low ≤ -1.0000		FTC-93 Linear	High	15.30%
					Medium	23.00%
					Low	6.90%
	Previous	High ≥ 0.1600	PDR-95 Linear	High	57.69%	
		Medium $\in (-1.0000, 0.1600)$		Medium	16.07%	
				Low	7.58%	
		Low ≤ -1.0000		FTC-95 Linear	High	57.69%
					Medium	12.35%
					Low	6.14%
Logistic	Pure	High ≥ 0.1950	PDR-93 Logit	High	25.00%	
		Medium $\in (0.0000, 0.1950)$		Medium	28.20%	
				Low	7.90%	
		Low ≤ 0.0000		FTC-93 Logit	High	14.10%
					Medium	23.10%
					Low	6.10%
	Previous	High ≥ 0.2200	PDR-95 Logit	High	57.69%	
		Medium $\in (0.0000, 0.2200)$		Medium	15.45%	
				Low	7.50%	
		Low ≤ 0.0000		FTC-95 Logit	High	55.56%
					Medium	11.54%
					Low	6.12%

His most accurate model, PDR-95 Linear, accurately predicted 57.69% of piping along levee segments in the “High Potential for Piping” category. Levee segments in the “Low Potential for Piping” category piped just 7.58% and segments in the “Medium Potential for Piping” category

pipied 16.07%. While the ability to predict any piping potential is useful, these models can be improved upon by applying the concepts previously stated in this study's hypothesis (e.g., a different stepwise regression method, the use of interaction terms).

3.2.4 Conclusions on Wilson's Efforts

Wilson used McFadden- R^2 value for linearly regressed models and the pseudo-McFadden R^2 logistically regressed models, a goodness-of-fit parameter, to determine how well the final independent variables could describe the dependent variable. This parameter is questionable as a method for determining significance (Davis, 2002). However, Wilson used it for comparison and discussion of his models. For a more detailed explanation of the McFadden- R^2 value and other goodness-of-fit measures, see Appendix A. Tables 2-5 show the McFadden- R^2 (or McFadden-pseudo R^2) for each model.

Three models, PDR-95 Linear, PDR-95 Logit and FTC-95 Linear, have the highest accuracy in predicting reaches with a high-potential for piping. However, PDR-95 Linear also has the highest inaccuracy in predicting a low-potential for piping. FTC-93 Linear has the lowest accuracy of predicting reaches with a high-potential for piping.

Wilson determined the most accurate models were derived using 1995 piping as the dependent variable and incorporating 1993 piping as an additional independent variable. Notably, the significance of adding of 1993 piping as an independent variable superseded the significance linear versus logistic regression techniques. His most accurate model was a linearly-regressed model with 1993 piping as an additional dependent variable. Wilson's research was

eventually employed for a published paper in a USACE-ERDC report by M.E. Glynn and J. Kuszmaul (2004).

CHAPTER FOUR

Methods: Model Building, Selection, and Application

Using the Wilson (2003) database, also used in Glynn and Kuszmaul (2004), for PDR and FTC, several datasets were defined for regression analysis during model building. Single-variable regression analysis was used to determine the significance of each independent variable. Multi-variable regression analysis was to determine the final models for the datasets. Models were created for each dataset using a modified forward stepwise regression procedure, also called a stepwise regression procedure, suggested by Le (2010). Model selection is based on the χ^2 -statistic value and each model's performance under thresholds is discussed in subsection 4.2.1. Two types of model were created using this procedure: "Pure" and "Previous". A third type of model, "Raw Previous", considered the efficacy of predicting future piping based solely upon previous piping where no geologic or flood specific variables were used, a concept not explored during the forward stepwise regression procedure. The significance of these models will be discussed in section 4.1. Maps of the best performing models in PDR and FTC were made using ArcGIS.

Blind testing on the models was conducted on a database compiled for a randomly selected levee district. Boring logs and flood event documents provided data for almost all necessary variables. No interpolation was necessary for application to the chosen field areas due to the proximity of boring data for each of the candidate levee segments. Analysis of LiDAR

imagery provided the data necessary for determining $\tilde{\mathbf{u}}$, presence and orientation of swales.

Models from each dataset were tested on this database. A more complete discussion is provided in Section 4.4.

4.1 The Engineering Applications of Developed Models

To accomplish a broader spectrum of engineering applications two necessary modeling tasks are performed: 1) ranking large stretches of levee with no data on prior piping events, and 2) ranking large stretches of levee in areas with data on prior piping events. A model of type 1 is a regression analysis in the truest form. Type 1, titled “Pure” models, strictly uses the geologic variables and flood specific variables recorded for any given environment. These models are not influenced by knowledge of prior piping events and show the direct relationship between the independent variables and the potential for piping. A model of type 2 is not considered to be a true regression analysis because a possible dependent variable, previous piping, is used as an independent variable. Type 2 models may incorporate both geologic data and previous piping, titled “Previous” models, or strictly use previous piping as the sole predictor of future piping, titled “Raw Previous”. While relationships between the original geologic and flood specific environments can be made for type 2, they are skewed due to the highly significant relationship between past and present piping events.

The difficulty in recording piping or problematic seepage due to the intensity of the event was discussed in Chapter 3. Despite this challenge, efforts are continually made by USACE and others to systematically record piping events along maintained levees (e.g., 2011 Ohio River

flood event). The significance of previous piping events to future piping events was suggested and confirmed by Wilson (2003).

4.2 Model Building

Model building refers to the process by which independent variables are defined, added, and eliminated from a logistic regression analysis. Model building was conducted for “Pure” and “Previous” models only. “Raw Previous” is exempt from model building as it uses only previous piping to predict a future piping event and will not be revisited until model utility is discussed in section 4.3.2.

A global database encompassing all known independent and dependent variables and their interaction terms divided into four general datasets were used for model building in this research. Interaction terms will be discussed in the following section on dataset definition.

Model building methods are the same for each dataset. In general, the significance of each independent variable to the corresponding dependent variable was determined by single-variable univariate logistic regression. Variables found to be significant were then added to a multi-variable univariate logistic regression model by the stepwise regression procedure method suggested by Le (2010). Le (2010) did not consider interaction terms when discussing stepwise regression procedure and adjustment was made to the method to include interaction terms. This method will be described more fully in section 4.2.2.2.

4.2.1 Defining Datasets

The global database used in model building and selection encompasses all independent and dependent variables collected and calculated by Wilson (2003), for PDR and FTC.

Interaction terms were added to the database to determine the added significance, if any, two variables combined would have on piping potential. For example, the presence of a riverside borrow pit and a pervious substratum with a large grain size together might have a greater effect on the occurrence of piping than the presence of just one of the variables. Interaction terms either take the form $(x_i x_j)$ where i and j designate each independent variable available. They describe effect modification in which one variable controls or modifies the effect of another variable and have been suggested to help create a better fit model in regression analysis (Le, 2010). Historically, interaction terms have been used in regression analyses in the health and biological sciences (e.g. the effect modification of smoking cigarettes with increasing age should accelerate health problems more quickly than just increasing age).

Interaction terms are considered as independent variables in the regression models (Le, 2010). The direct relationship between an interaction term, its factors, or a different interaction term containing one of its factors, will result in a false positive goodness-of-fit value (e.g., χ^2 statistic or McFadden pseudo- R^2) if they are included in the model simultaneously. This is due to collinearity, or multicollinearity; the process in which two independent variables, x_1 and x_2 , are highly correlated and the contribution of x_1 mirrors the contribution of x_2 , resulting in an overlap of data and an unreasonably large goodness-of-fit value (Mela and Kopalle, 2002). Careful consideration must be taken during multi-variable regression analysis to reduce this possibility. The process of reducing collinearity between independent variables is described in Section 4.2.2.2.

As mentioned in Chapter 3, a few independent variables (e.g. h_0 , H , i_c) are calculated from variables pertaining to a specific flood and are more difficult to ascertain. Independent variables that must be calculated from a set of other variables are less reliable than variables taken from raw data (e.g., grain size of pervious substratum). To account for the questionability of some variables, two datasets were defined: Unlimited Dataset and Limited Dataset, where the Unlimited Dataset contains all types of variables, and the Limited Dataset contains only directly observed (e.g., d_{10}) or measured variables (e.g., \bar{u}), omitting variables that are the result of analysis and not strictly interpretation.

Table 7. Allocation of independent variables for each defined dataset.

Independent Variable Datasets				
UNLIMITED PREVIOUS		Previous Piping (\mathbf{P})		LIMITED PREVIOUS
	UNLIMITED PURE	Presence and Orientation of Swales ($\hat{\mathbf{u}}$)	LIMITED PURE	
		Transformed Confining Layer Thickness (\mathbf{z}_t)		
		Riverside Borrow Pits ($\check{\mathbf{R}}$)		
		Effective Grain Size (\mathbf{d}_{10})		
		Relief Wells ($\dot{\mathbf{R}}$)		
		Landside Ditches ($\dot{\mathbf{L}}$)		
		Berms (\mathbf{B})		
		Net Head (\mathbf{H})		
	Excess Hydrostatic Head (\mathbf{h}_0)			
	Critical Gradient (\mathbf{i}_c)			

The Unlimited and Limited datasets will be further subdivided to accomplish these two tasks. “Pure” and “Previous” datasets correspond to the dependent variable, 1993 piping or 1995 piping, respectively, used in model building (Table 7). “Previous” datasets contain the variable \mathbf{P} , previous piping events from 1993, as an available independent variable. Interaction terms for the four resulting dataset vary for each dataset. For example, the “Unlimited Previous” dataset

results in the largest number of interaction terms due to the larger amount of variables in the dataset and the addition of 1993 piping as an independent variable.

Finally, the types of dataset were applied to values strictly pertaining to Prairie du Rocher Levee District (PDR) and values pertaining to both Prairie du Rocher and Fort Chartres Levee District (PDR_FTC). They were not applied to values strictly pertaining to Fort Chartres Levee District (FTC). Therefore, eight total datasets were used for model building: PDR Unlimited Pure and Previous; PDR Limited Pure and Previous; PDR_FTC Unlimited Pure and Previous; PDR_FTC Limited Pure and Previous. The model building process was applied to these eight different datasets. Please note that some variables were not available for FTC so, for example, PDR Unlimited has more variables available for model building than PDR_FTC Unlimited.

4.2.2 Stepwise Regression Procedure

Stepwise regression is a form of regression analysis that seeks to identify significant variables to the outcome in an attempt to reduce the possibility of a Type I (false positive) error (Le, 2010; Davis, 2002). This is useful for regression analyses, such as these, where the significance of some or all variables is unknown or questionable, which could result in an ill-fitted model (Davis, 2002). The method consists of defining criteria for selecting a model and specifying a strategy for applying the criteria (Le, 2010).

Criteria for selecting a model are determined using single-variable logistic regression for every independent variable and their corresponding dependent variable in the dataset. A “cut off standard” related to significance is applied to all variables, and those not meeting the standard

are removed from further analysis (Le, 2010). This process will be further discussed in section 4.2.2.2.

Strategy (i.e. forward or backward) defines how variables are selected for the multi-variable model and their order of addition (Le, 2010). A forward procedure involves the addition of significant variables to a multi-variable model whereas a backward elimination procedure involves the elimination of insignificant variables from a multi-variable model that initially includes all variables (Le, 2010; Davis, 2002). A modified strategy of the forward procedure, called a stepwise regression procedure, uses a re-examination method at each addition to the model (Le, 2010). If a re-examined variable has lost its significance to the model at any step, it is removed and the forward procedure continues (Le, 2010).

4.2.2.1 Stepwise Regression Procedure: Defining Criteria

Single-variable regression analyses, performed in XLSTAT statistical analysis software created by Addinsoft and run through Microsoft Excel, were conducted on the independent variables in each dataset and the corresponding dependent variable to determine their significance to the dependent variables. The “cut-off standard” used for this step is set at a $P - \text{value} \leq 0.20$, so that any variable which meets this standard or has a larger P-value will be eliminated from the model. This cutoff eliminated a majority of independent variables in each dataset and effectively reduced time spent during multi-variable regression analysis.

Some variables were found to have a $P\text{-value} \leq 0.0001$, the minimum P-value XLSTAT can calculate. These variables are “highly” significant and indistinguishable from one another

when considering their significance to the dependent variable. The approach to this problem is presented in the next section.

4.2.2.2 Stepwise Regression Procedure: Strategy

The order in which independent variables from a dataset are added to the multi-variable model is determined by the variable's P-value, so that the most significant variable is added first and the least significant is added last (Le, 2010). During reexamination, if any variable has lost significance to a P-value > 0.15 , it is removed from the model and the addition of variables is resumed.

The possibility of collinearity, or multicollinearity, a result of two highly correlated variables within a regression model, was introduced in the discussion on interaction terms from Section 4.2.1. This can be avoided by first, calculating the correlation between the independent variables in each dataset and second, omitting any variable from addition to the model if a significantly correlated variable, P-value < 0.15 , is already used in the model. For example, if a model contains the interaction term ($\mathbf{d_{10}} * \mathbf{P}$), then any factor of that term (e.g., $\mathbf{d_{10}}$, \mathbf{P} , or an interaction term with either of those variables) cannot be added to the model.

This process is accepted for datasets where the significances of the independent variables are distinct (i.e., each independent variable has a unique P-value). However, this process can result in the omission of important independent variables if the dataset contains highly significant, independent variables (i.e., any variable with a single-variable regression P-value ≤ 0.001). Highly significant, independent variables were mentioned in Section 4.2.1. These indistinguishably significant variables are all eligible for first addition to the model, which will

invariably affect the addition of any correlated variable regardless of its significance to the model. Furthermore, the omission of correlated variables can change the significance of the remaining possible variables to the model.

All possible permutations of the highly significant, independent variables for a given dataset must be considered to completely represent the dataset during model building. This resulted in several model variations for each dataset. For example, a dataset with four highly

Table 8. Number of models created during model building.

Type	Dataset	Levee District	No. of Models
Unlimited	Pure	PDR	2
		PDR_FTC	3
	Previous	PDR	133
		PDR_FTC	6
Limited	Pure	PDR	2
		PDR_FTC	3
	Previous	PDR	5
		PDR_FTC	6

significant, indistinguishable, variables would have 16 possible combinations that must be applied. The largest number of permutations was a result of 16 highly significant variables. However, some permutations of those variables were redundant and did not provide new information. Table 8 shows the number of models created in each dataset, the top model in each dataset (determined through goodness-of-fit values) and the variables associated with that model.

4.3 Model Selection

For this research, model selection depended upon the chi-square (χ^2) statistic and the application of thresholds to all possible models. Although model building was conducted on eight datasets, model selection is restricted to the four general types of dataset and not dependent

upon specific levee district data. This equates to four selected models from each type of dataset: Unlimited Pure, Unlimited Previous, Limited Pure, and Limited Previous. Therefore, four separate thresholds were chosen to apply to models in each dataset. Because each dataset is distinctly defined, unique thresholds may be set for each type without the risk of inconsistency.

Thresholds refer to the range of model values that define areas of high, medium, and low potential for piping. They are determined from the model with the highest χ^2 statistic value in each dataset, “threshold definers” listed in Table 9. The methods used to determine thresholds are defined in section 4.3.1. Once the thresholds are applied to all models, the percentage of segments that piped in each category can be determined and the first assessment of model utility can be performed. Initial model selection is possible after the application of thresholds to all datasets. Once selected, the models are compared with “Raw Previous” models and tested on a new field area, a procedure known as blind testing.

Table 9. Top two performing models from each dataset.

Dataset	Type	Data	Model	Chi ² Statistic	Variables
Unlimited	Pure	PDR	A	29.23732	$\tilde{u} * d_{10}$
					$z_t * h_0$
		PDR_FTC	A	24.739	$\tilde{u} * d_{10}$
					\tilde{R}
	Previous	PDR	DC	141.3113388	$H * i_c$
					$h_0 * P$
					$z_t * z_t$
					$\tilde{R} * \tilde{R}$
		PDR_FTC	D	121.5149	$d_{10} * P$
					$z_t * z_t$
					$\tilde{R} * \tilde{R}$
					$d_{10} * \tilde{L}$
Limited	Pure	PDR	B	28.91403	$\tilde{R} * \tilde{R}$
					d_{10}
		PDR_FTC	A	24.739	$\tilde{u} * d_{10}$
					\tilde{R}
	Previous	PDR	E	82.25785	$\tilde{u} * d_{10}$
					P
					z_t
					$\tilde{R} * \tilde{R}$
		PDR_FTC	D	121.5149	$d_{10} * P$
					$z_t * z_t$
					$\tilde{R} * \tilde{R}$
					$d_{10} * \tilde{L}$

4.3.1 Thresholds

The model with the highest χ^2 statistic in each general dataset was chosen as the “threshold definer” for that given dataset. The model values determined for raw piping observations were compared to the model values for all observations for each “threshold definer”. A MATLAB function, written by the researcher, produced two figures: (a) stacked histograms of the distributions (Figure 3 and Figure 5), and (b) a graph defined by the division of piping observation model values over all observation model values, defined as the “piping ratio” (Figure 4 and Figure 6). This can be better thought of using an equation (1) to calculate the

“piping ratio”, where $Y_1=1$ or piping observations, Y_2 = all observations, and \hat{y}_i refers to the associated model values.

$$F(Y_1, Y_2) = \frac{\hat{y}_i(Y_1)}{\hat{y}_i(Y_2)} \quad (1)$$

Plotting $(F(Y_1, Y_2), x)$, where x ranges from the maximum model value to the minimum model value, shows the relationship between model values that piped and all model values. Ideally, relatively high model values (e.g., $\hat{y}_i = 0.7$) are associated with piping observations only and will display a one-to-one ratio on the graph. While relatively low model values (e.g., $\hat{y}_i = 0.1$) are strictly associated with non-piping observations and will display no ratio on the graph. Previously unobserved, natural sills in the distribution of model values as they relate to piping observations can be easily located on this graph and various limits for threshold values can be assigned.

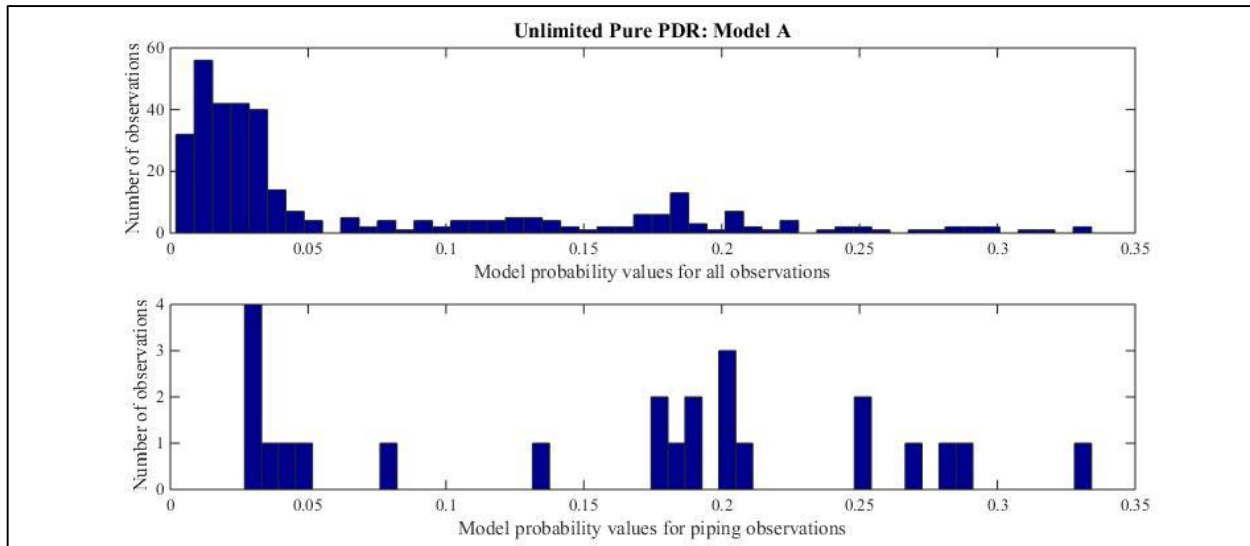


Figure 3. Best Fit top Unlimited Pure model. The top histogram depicts the distribution of model values for all observations. The bottom histogram depicts the distribution of model values for observations associated with piping events.

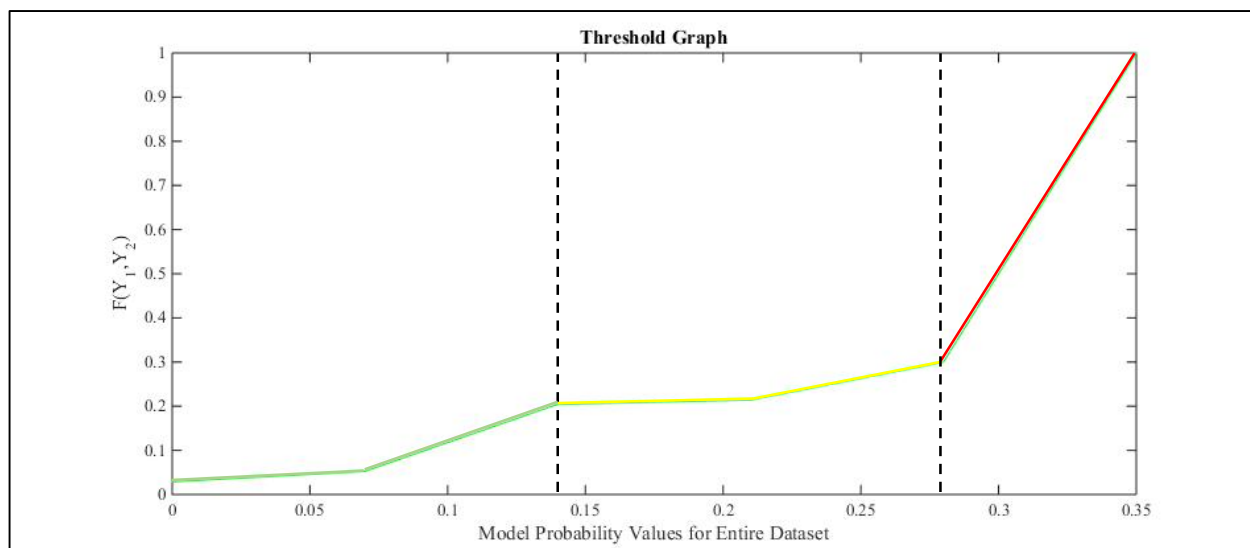


Figure 4. Best Fit top Unlimited Pure Model. Comparison of “piping ratio” to the distribution of model probability values for the entire dataset.

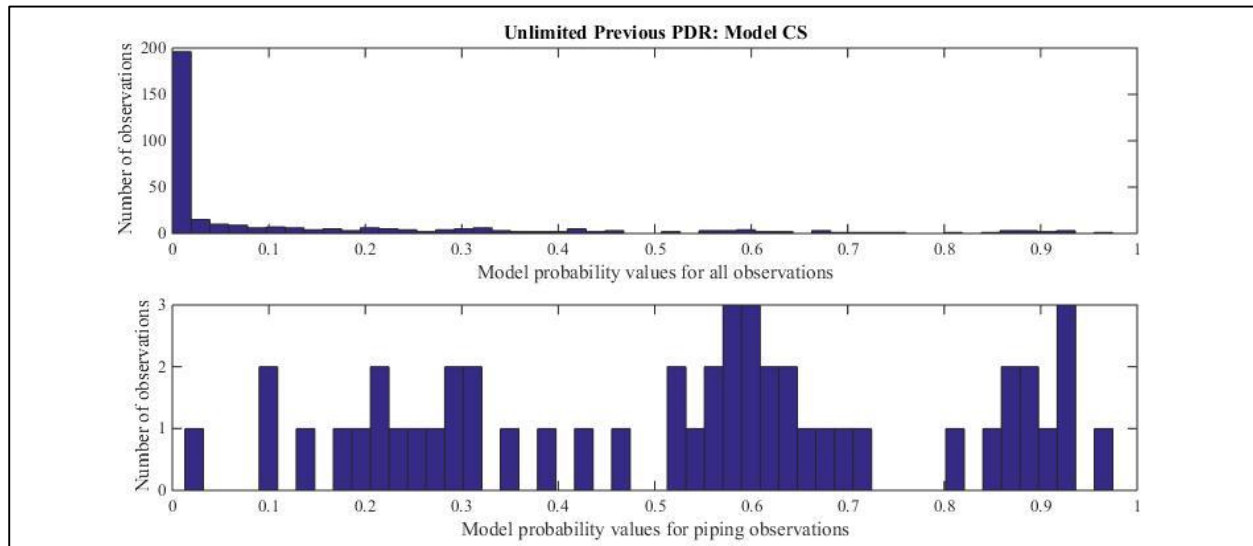


Figure 5. Best Fit top Unlimited Previous model. The top histogram depicts the distribution of model values for all observations. The bottom histogram depicts the distribution of model values for observations associated with piping events.

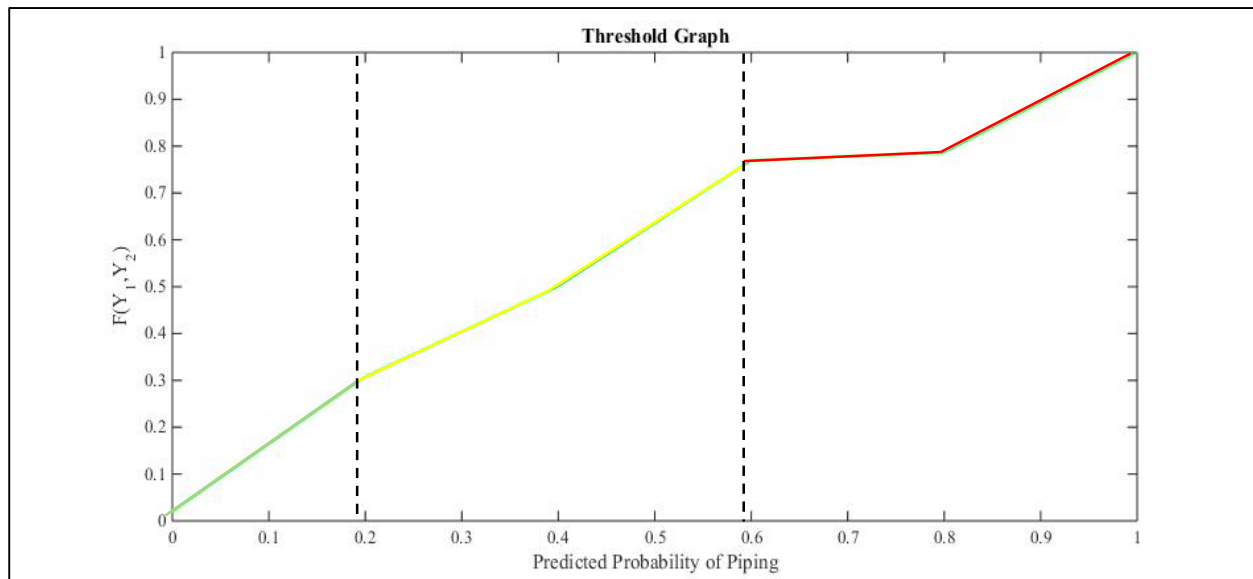


Figure 6. Best Fit top Unlimited Previous Model. Comparison of “piping ratio” to the distribution of model probability values for the entire dataset.

Threshold limits were applied to the top four model variates in each specific dataset, or all model variates in the dataset if less than four were created, this includes PDR and PDR_FTC datasets. To perform this action, a MATLAB function was written for each general dataset (i.e., Unlimited Pure and Previous and Limited Pure and Previous). The functions were used to determine how many total observations fell into in each category and how many piped.

4.3.2 Determining Model Utility

Model utility may be determined after application of the thresholds. Two forms of accuracy were used in model selection. The first form uses “percent piped” in the categories of high, medium, and low potential for piping for the four models built using the stepwise forward selection procedure. The “percent piped” value for all three categories is found by the division of piped reaches in the respective category by non-piped reaches in the same category.

The second form uses 2x3 contingency tables to determine overall accuracy and individual accuracy of the four selected models, and 2x2 contingency tables to determine overall accuracy of the “Raw Previous” models and of each category in the four selected models. For the 2x3 contingency tables, overall accuracy is found by dividing the row total of the correctly predicted observations by the sum total. Correctly predicted observations are represented by the non-piped segments for the Low and Medium potential for piping categories and the piped segments for the High potential for piping category. Individual accuracy is found by dividing the row total by the column total of the respective category. For the 2x2 contingency tables, overall accuracy is found by dividing the sum of the correctly predicted observations by the sum total.

Model utility is then assessed qualitatively by creating maps of each levee district and the determined piping potential of each segment using ArcGIS software. Functionality of the model can be easily shown by identifying actual piping locations along the levee and comparing each category to their locations. These maps are presented in Chapter 5.

4.4 Blind Tests

Three levee districts along the MMR were available for blind testing of the models. Each was assigned a value: Clear Creek Levee District (CCL) (1), Columbia Levee District (CL) (2), and East Cape Girardeau Levee District (ECG) (3). No other levee districts within the MMR were available for blind testing due to the unavailability of necessary variables. A random number generator from MATLAB was used to determine which levee district would be used in the blind tests. This eliminates any bias the researcher might have in choosing the field area in which to perform blind tests.

Using the same method, a small sample of segments, maximum of 50 segments, were randomly selected within ECG. However, only segments with a complete, or nearly complete, set of needed model variables are included as candidates for random selection. In that sense, there may be some bias in the selection of levee segments. The most limited variable sets are transformed confining layer thickness, z_t , and d_{10} size. D_{10} is only available in borings chosen for soil sampling and laboratory analysis. The percentage of segments that piped in the random sample is compared to the percentage of segments that piped for the district to confirm the selected segments are a justified representation of the levee district.

Data were acquired through the available materials described in previous sections. Interpolation was not necessary because data points were selected among available borings so that d_{10} is a direct measurement and z_t can be directly calculated from the semi- to impervious strata thicknesses. Some analysis of LiDAR was required to interpret \bar{u} , presence and orientation of swales. All other variables were either located in flood event documentation or aerial photographs provided by USGS. Only interaction terms used in the selected models were defined for this database. Again, the database was defined into four subsets: Unlimited Pure, Unlimited Previous, Limited Pure, and Limited Previous. The selected models for each dataset were run and the resulting model values were then applied to their respective thresholds. Model utility was then assessed for a second time by determining the percentage of segments that piped in each category.

CHAPTER FIVE

Results

Model results are categorized by Unlimited versus Limited; Unlimited Pure versus Unlimited Previous; and Unlimited Pure versus Limited Pure. To succinctly list model results, the models will be compared based on Unlimited versus Limited datasets and Pure versus Previous datasets. These comparisons will be shown in the following sections. Also, model utility describes both the accuracy of the model after applying thresholds and the accuracy of the model after blind testing so these two types of utility will be listed for each model. Significant independent variables for each model will also be assessed.

5.1 Significant Independent Variable Analysis

Interaction terms account for 19 out of 24 total variables used or 79.17% of total model variables. This value includes repeated interaction terms. The most repeated interaction terms in both datasets are $(\check{\mathbf{R}} * \acute{\mathbf{R}})$ with a frequency of five, where $\check{\mathbf{R}}$ represents riverside borrow pits and $\acute{\mathbf{R}}$ represents relief wells, $(\tilde{\mathbf{u}} * \mathbf{d}_{10})$ with a frequency of four, where $\tilde{\mathbf{u}}$ represents presence and orientation of swales and \mathbf{d}_{10} represents effective aquifer grain size. Five terms remained highly significant, P-value < 0.0001, even after additional terms are incorporated into the overall model (e.g., $(\mathbf{d}_{10} * \mathbf{P})$, $(\mathbf{z}_t * \mathbf{H})$, $(\check{\mathbf{R}} * \acute{\mathbf{R}})$ for the Unlimited Previous dataset, $(\tilde{\mathbf{u}} * \check{\mathbf{R}})$ for the Limited Pure

dataset, and **P** for the Limited Previous dataset). Table 10 shows the final independent variables and their significance to the Pure models. Only Pure model variables are shown because of their directly measured relationship with previous piping.

Table 10. Significance of final independent variables to piping locations.

Variable Significance to Piping for Pure Models		
MODEL	VARIABLE	SIGNIFICANCE
PDR Limited Pure	Riverside Borrow Pits * Unfavorable Geologic Condition, ($\check{R} * \check{u}$)	<0.0001
	Aquifer Grain Size, d_{10}	0.001
FTC_PDR Limited Pure	Unfavorable Geologic Condition * Aquifer Grain Size ($\check{u} * d_{10}$)	<0.0001
	Riverside Borrow Pits, \check{R}	0.059

The two continuous variables, d_{10} and z_t , were compared to the original piping observations and calculated model values associated with a “high potential for piping” using histograms similar to those presented in Chapter 4 (Figure 7 through Figure 14).

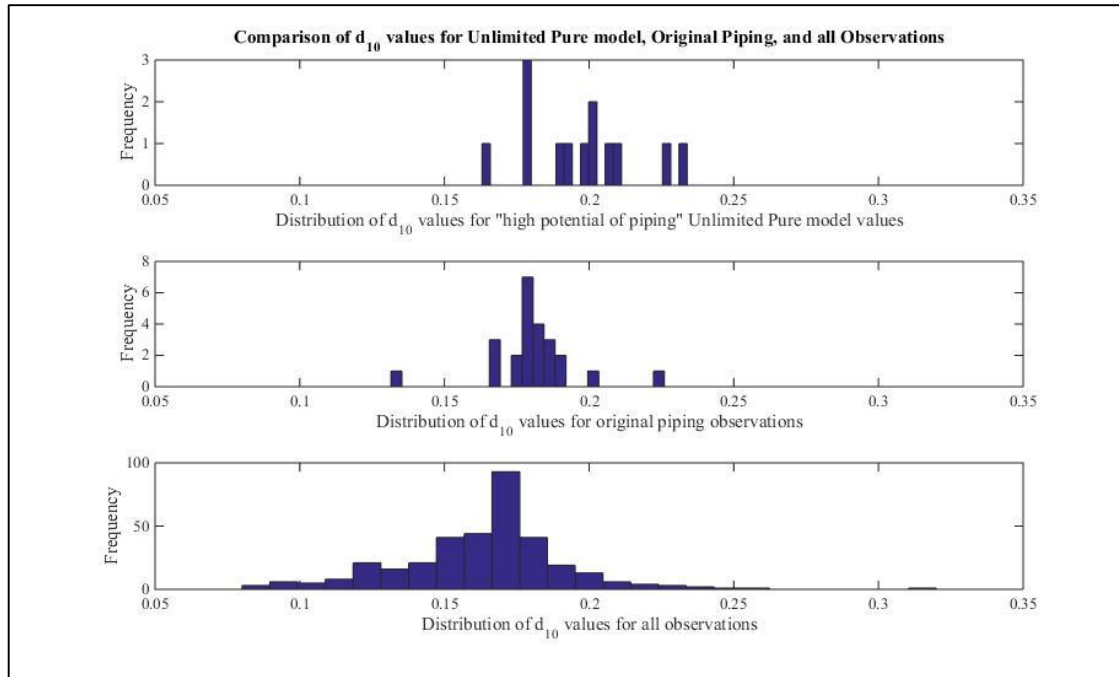


Figure 7. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Unlimited Pure model, original piping observations, and all observations.

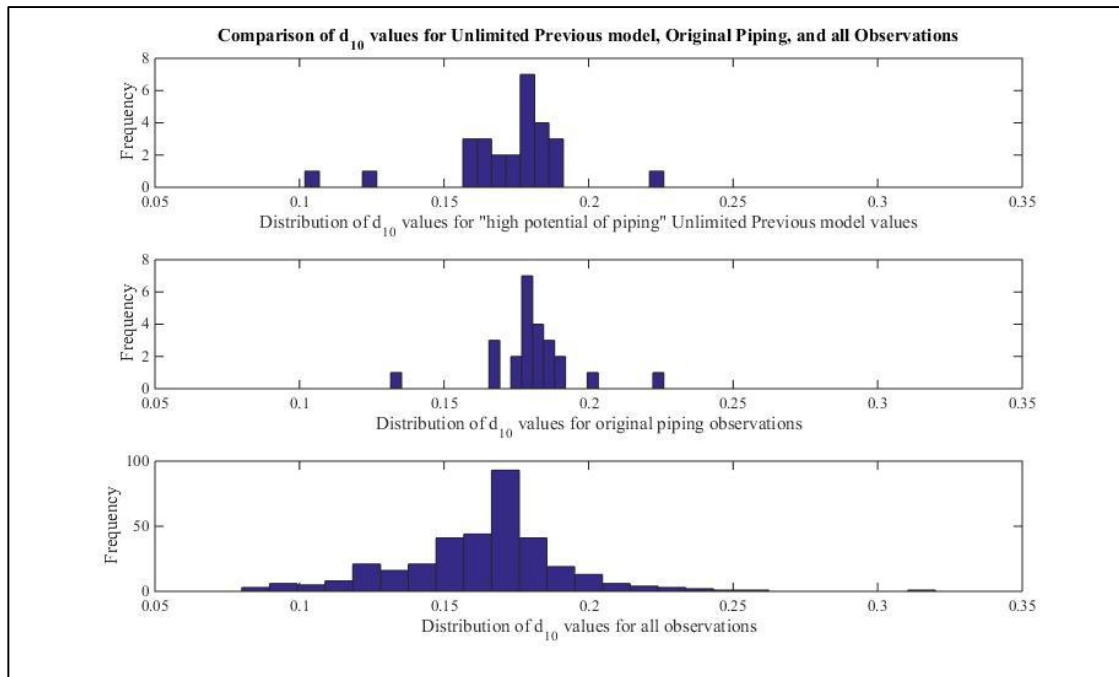


Figure 8. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Unlimited Previous model, original piping observations, and all observations.

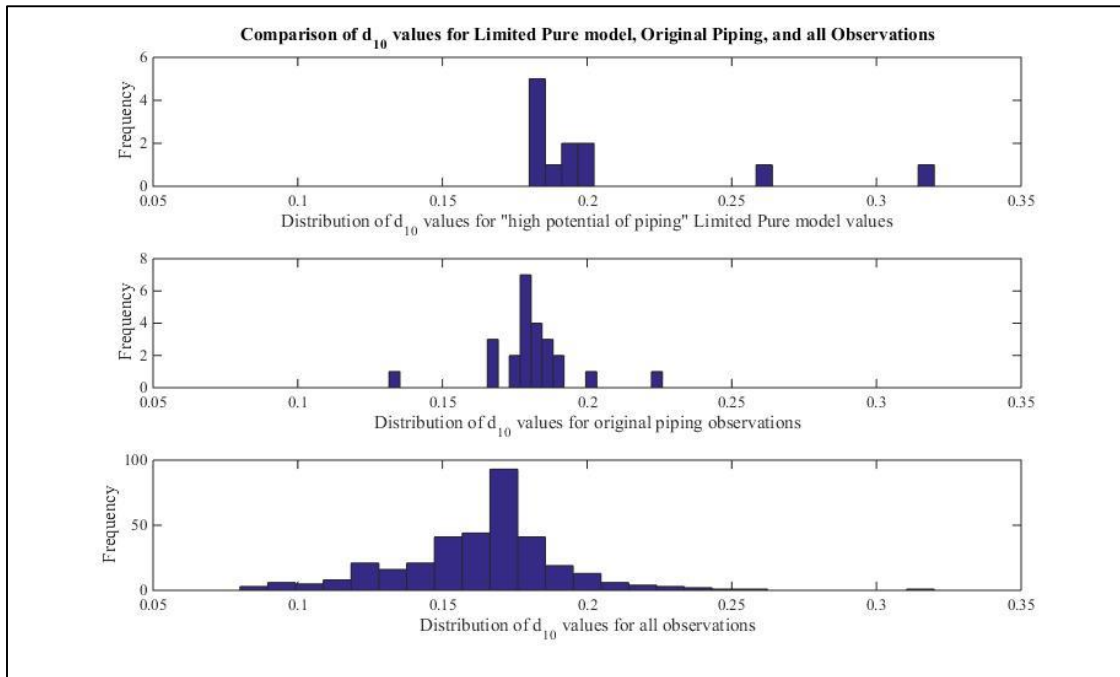


Figure 9. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Limited Pure model, original piping observations, and all observations.

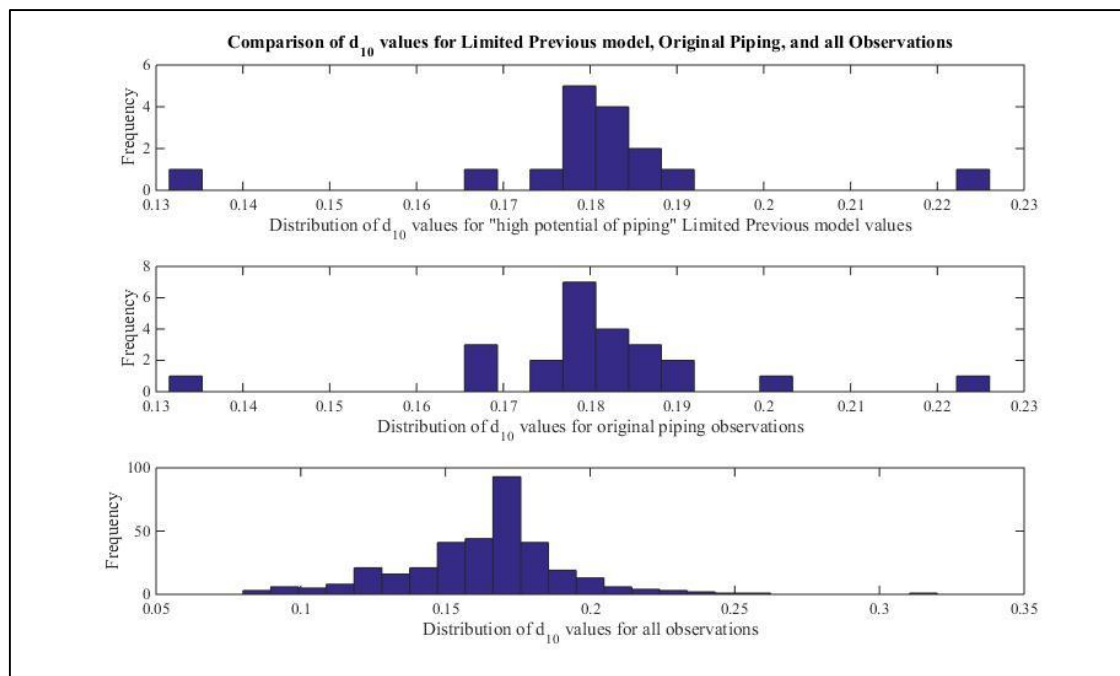


Figure 10. Comparison of Prairie du Rocher d_{10} values associated with a “high potential of piping” from the top Limited Previous model, original piping observations, and all observations.

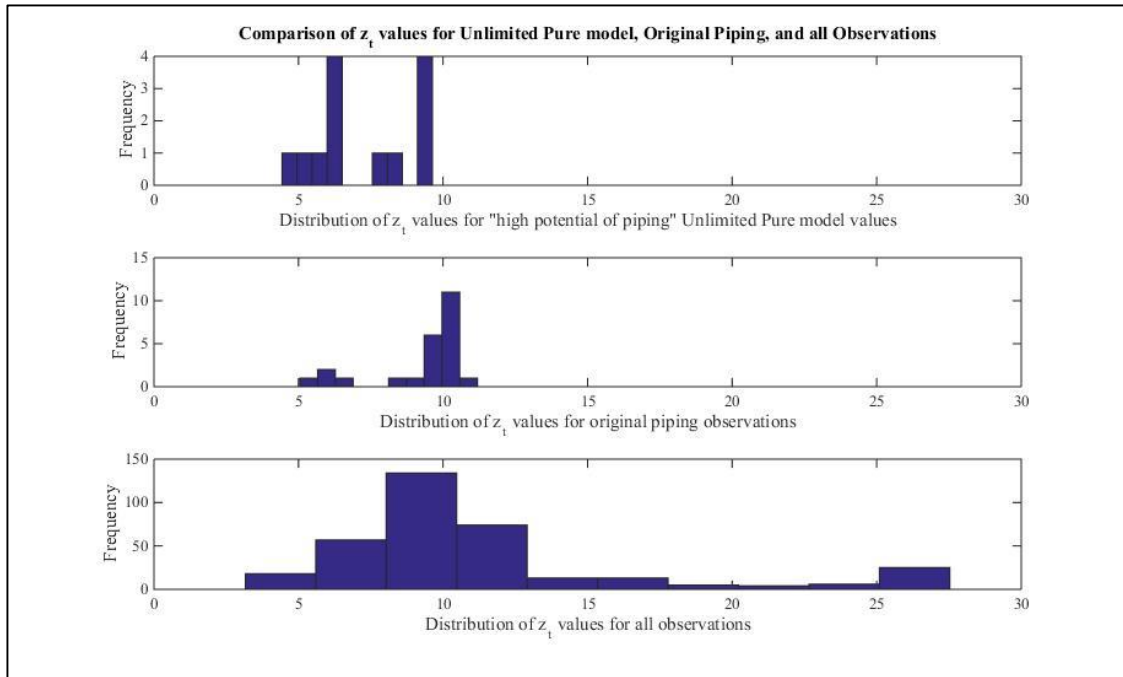


Figure 11. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Unlimited Pure model, original piping observations, and all observations.

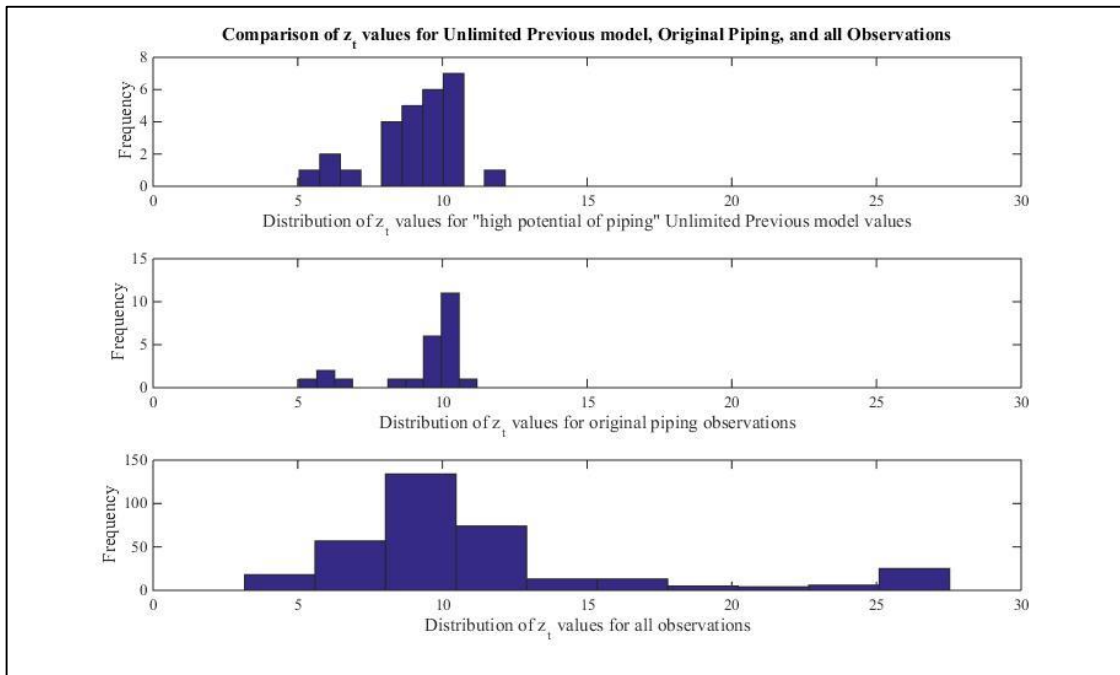


Figure 12. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Unlimited Previous model, original piping observations, and all observations.

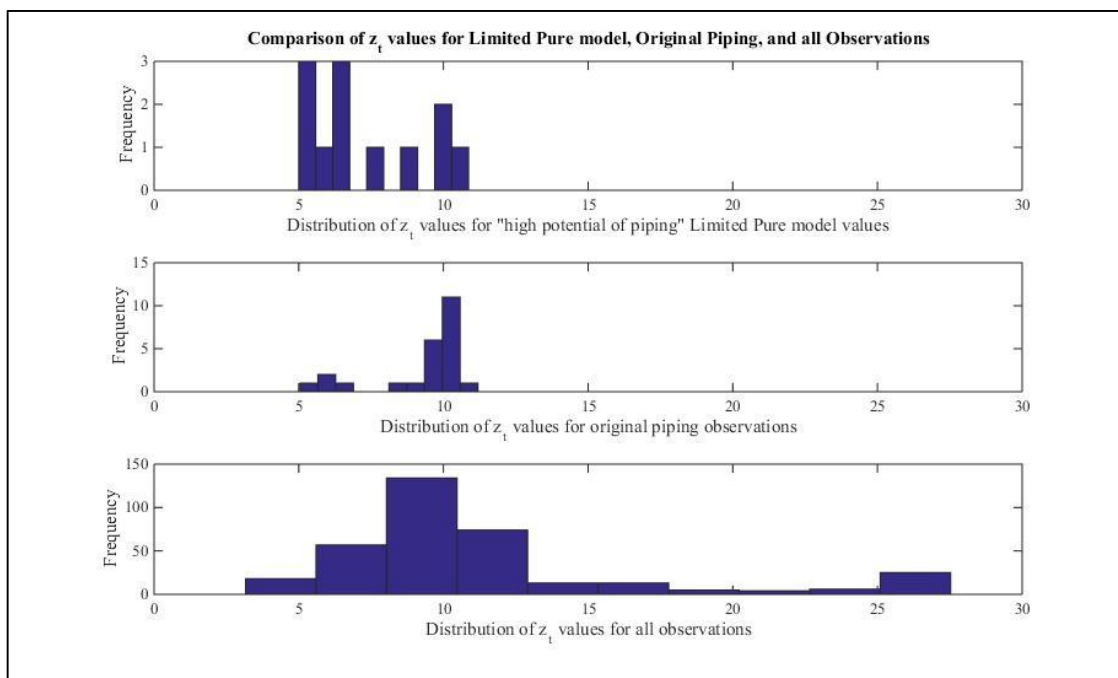


Figure 13. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Limited Pure model, original piping observations, and all observations.

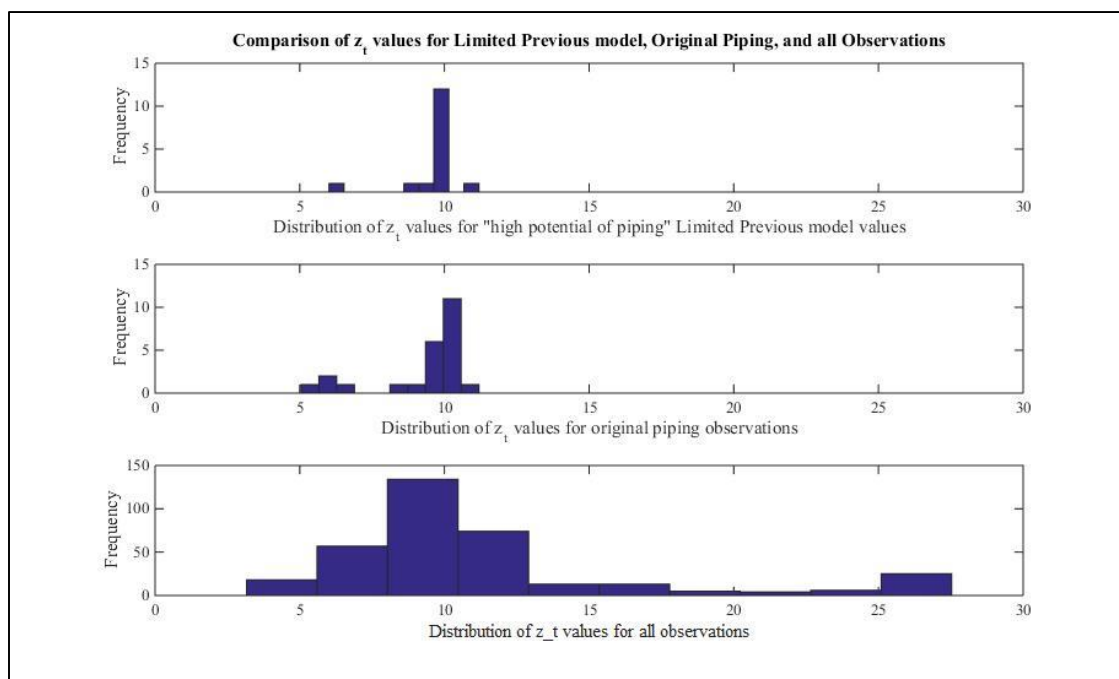


Figure 14. Comparison of Prairie du Rocher z_t values associated with a “high potential of piping” from the top Limited Previous model, original piping observations, and all observations.

The distribution of \mathbf{d}_{10} and \mathbf{z}_t values associated with a “high potential for piping” model values most resembled the distribution of values associated with actual piping observations for the Previous models. Pure models were unable to clearly show the relationship between these values and piping observations.

5.2 Model Utility: Selected Models

The availability of variables in the “Unlimited” dataset resulted in many more possible models than the “Limited” dataset (i.e., 144 models versus 16 models). The same is true for the availability of variables in the “Previous” dataset versus the “Pure” dataset (i.e., 150 models versus 10 models) (see Table 9.). Best fit is determined by the χ^2 statistic. The top model for the Unlimited dataset was slightly better fit to its parameters than the top model for the Limited dataset (e.g., Unlimited Previous PDR, $\chi_{LR}^2(344, N=349) = 144.31$, Limited Previous PDR_FTC, $\chi_{LR}^2(344, N=349) = 121.51$) where χ_{LR}^2 is the log-likelihood ratio chi-square test and the form (A,N) refers to degrees of freedom, A, and sample size, N (see Table 9). The PDR_FTC specific models are identical in Unlimited and Limited due to the unavailability of analyzed variables (e.g., \mathbf{h}_0). The minimum standard deviation for the dataset is 14.00, when the top Unlimited model and the top Limited model are considered. The maximum standard deviation for the dataset is 82.43, when the top Unlimited and bottom Limited models are considered.

5.2.1 Comparison of Limited and Unlimited Datasets

The top Limited model, Previous dataset PDR_FTC Model A, with $\chi_{LR}^2(345, N=349) = 112.56$, outperformed the top Unlimited model, Previous dataset PDR Model CS, with $\chi_{LR}^2(344, N=349) = 138.62$ in predicting piped levee segments, where 83.87% of segments piped in the

“high potential for piping” category and 9.51% of segments piped in the “low potential for piping category” (Figure 15 and Figure 16). However, the top Unlimited model outperformed the top Limited model in predicting where piping would *not* occur, in which 77.78% of segments piped in the “high potential for piping” category and 1.92% of segments piped in the “low potential for piping” category. The interaction term ($\check{\mathbf{R}} * \hat{\mathbf{R}}$) appears in both models. Also, \mathbf{P} and \mathbf{z}_t appear in both models but in different forms, i.e. ($\mathbf{d}_{10} * \mathbf{P}$) for Unlimited and \mathbf{P} for limited.

5.2.2 Comparison of Pure and Previous Datasets

“Previous” selected models outperformed “Pure” selected models in predicting a “high potential for piping” by a maximum of 63.3%, for the Previous Unlimited Model DC and the Pure Limited Model B, and a minimum of 46.4%, for the Previous Limited and Pure Limited (Figures 15, 16, 17, and 18). “Previous” models predicted 77.8% of piped levee segments and 83.87% of piped levee segments, categorized as “high potential for piping”, for Unlimited and Limited datasets, respectively. “Pure” models predicted 41.67% of piped levee segments and 30.77% of piped levee segments, categorized as “high potential for piping”, for Unlimited and Limited datasets, respectively.

The top Previous model, Limited Model A with $\chi^2_{LR}(345, N=349) = 112.56$, outperformed the top Pure model, Limited Model B with $\chi^2_{LR}(347, N=349) = 28.91$, in predicting high, medium, and low “potential for piping” along levee segments, where 83.87% piped versus 41.67% for “high potential for piping”, 42.86% piped versus 12.94% for “medium potential for piping”, and 9.51% versus 3.17% “low potential for piping”.

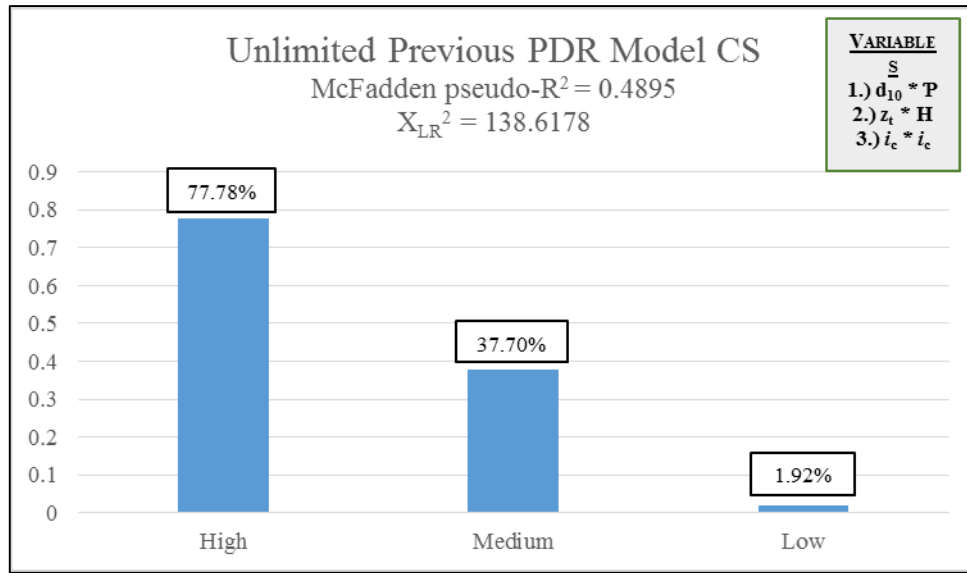


Figure 15. Bar chart for the Unlimited Previous model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

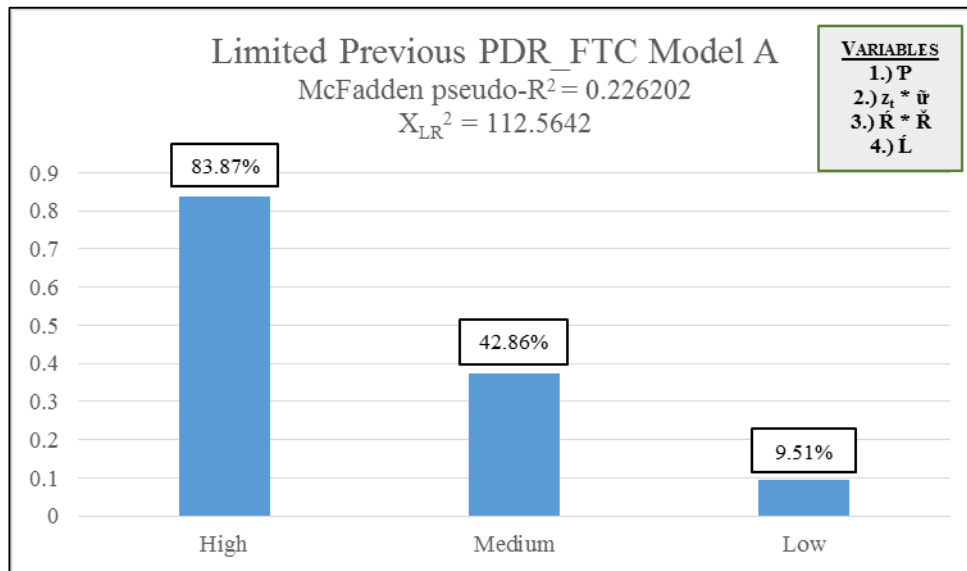


Figure 16. Bar chart for the Limited Previous model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

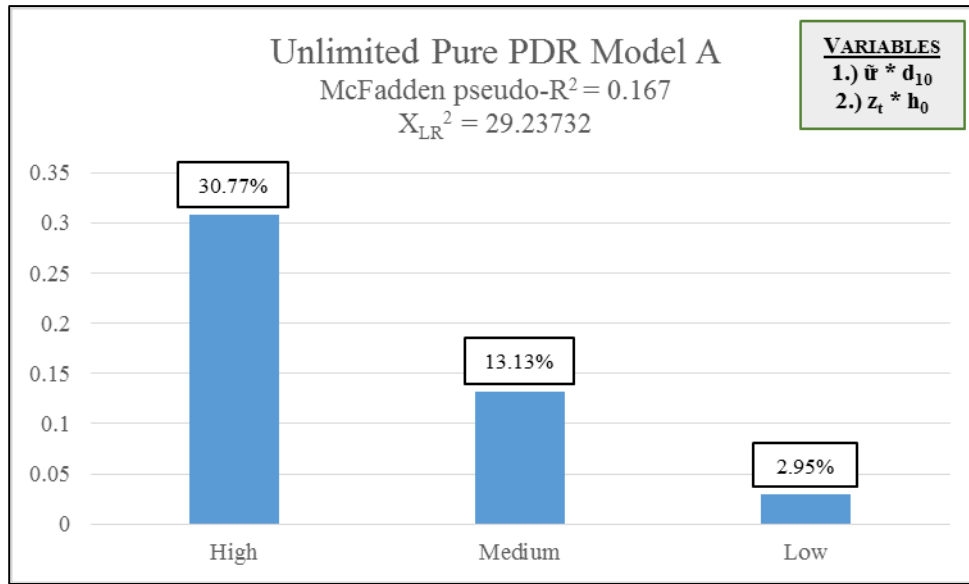


Figure 17. Bar chart for the Unlimited Pure model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

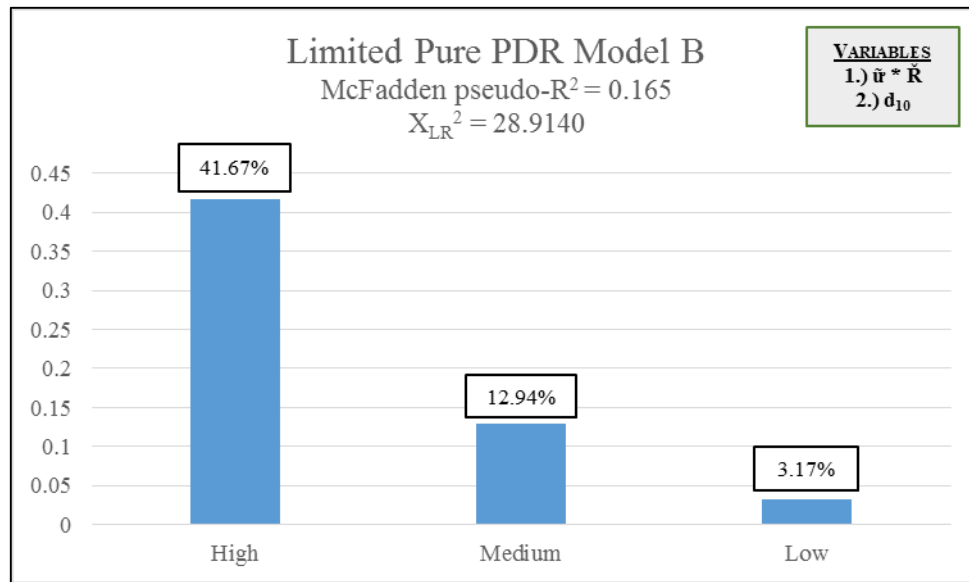


Figure 18. Bar chart for the Limited Pure model of the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

5.2.3 Contingency Table Comparisons

To consider the use of previous piping as the sole predictor of piping, contingency tables were used to evaluate 1993 piping versus 1995 piping and each of the selected models. Two by three tables are made to determine overall accuracy of the selected models and individual accuracy of the categories in each model. Two by two tables are made to determine overall accuracy of the Raw Previous model and the high, medium, and low categories for all four of the selected models. The division of the sum of the number of correctly classified observations for each category by the total number of observations for the dataset determines overall accuracy for the two by three tables (Table 11-14). Correctly classified observations refer to those observations that pipe in the high category and those observations that do not pipe in the medium and low category. The division of the correctly classified observations in a specific category by the total observations found in that category determines individual accuracy of the respective category for the two by three tables.

Table 11. PDR Pure Limited Model B.

PDR Pure Limited Model B		Model Predicts Piping			Total
		Low	Medium	High	
1993 Piping	Correct Prediction	244	74	7	325
	Incorrect Prediction	8	11	5	24
	Total	252	85	12	349
Low Potential Accuracy		96.83%			
Medium Potential Accuracy		87.06%			
High Potential Accuracy		58.33%			
Overall Correct Prediction Accuracy		93.12%			

Two by three contingency table showing overall accuracy of the model and individual accuracy of each category as it relates to its dependent variable, 1993 piping. Correct Prediction is the correctly predicted observations and incorrect prediction is the incorrectly predicted observations.

Table 12. PDR_FTC Pure Limited Model A.

PDR_FTC Pure Limited Model A		Model Predicts Piping			Total
		Low	Medium	High	
1993 Piping	Correct Prediction	417	171	1	589
	Incorrect Prediction	14	24	0	38
	Total	431	195	1	627
Low Potential Accuracy		96.75%			
Medium Potential Accuracy		87.69%			
High Potential Accuracy		100.00%			
Overall Correct Prediction Accuracy		93.94%			

Two by three contingency table showing overall accuracy of the model and individual accuracy of each category as it relates to its dependent variable, 1993 piping. Correct Prediction is the correctly predicted observations and incorrect prediction is the incorrectly predicted observations.

Table 13. PDR Previous Unlimited Model CS

PDR Previous Unlimited Model CS		Model Predicts Piping			
		Low	Medium	High	Total
1995 Piping	Correct Prediction	256	38	21	315
	Incorrect Prediction	5	23	6	34
	Total	261	61	27	349
Low Potential Accuracy		98.08%			
Medium Potential Accuracy		62.30%			
High Potential Accuracy		77.78%			
Overall Correct Prediction Accuracy		90.26%			

Two by three contingency table showing overall accuracy of the model and individual accuracy of each category as it relates to its dependent variable, 1995 piping. Correct Prediction is the correctly predicted observations and incorrect prediction is the incorrectly predicted observations.

Table 14. PDR Previous Limited Model E

PDR_FTC Previous limited Model E		Model Predicts Piping			
		Low	Medium	High	Total
1995 Piping	Correct Prediction	535	2	26	563
	Incorrect Prediction	56	3	5	64
	Total	591	5	31	627
Low Potential Accuracy		90.52%			
Medium Potential Accuracy		40.00%			
High Potential Accuracy		83.87%			
Overall Correct Prediction Accuracy		89.79%			

Two by three contingency table showing overall accuracy of the model and individual accuracy of each category as it relates to its dependent variable, 1995 piping. Correct Prediction is the correctly predicted observations and incorrect prediction is the incorrectly predicted observations.

The two by two tables analyze two types of model: each category as it relates to the entire dataset of observations for the selected models and the Raw Previous models (Table 15-28). For example, overall accuracy of the low category for the dataset PDR Pure Limited Model B is found by dividing the sum of the correctly classified observations; i.e. observations found in the low category which have not piped and observations found outside the low category which did pipe, i.e. the medium and high categories, which have piped; over the total number of observations for the dataset. Using the same methods, overall accuracy for the Raw Previous models is found by dividing the sum of the correctly classified observations; i.e. observations which during both 1993 and 1995 events and observations which did not pipe during both 1993 and 1995 events.

Table 15. PDR Pure Limited Model B

PDR Pure Limited Model B		Low		
		Predicted Piping	Predicted No Piping	Total
1993 Piping	Piping	16	8	24
	No Piping	81	244	325
	Total	97	252	349
Overall Correct Prediction Accuracy		74.50%		

Two by two contingency table showing the overall accuracy of the Low category.

Table 16. PDR Pure Limited Model B

PDR Pure Limited Model B		Medium		
		Predicted Piping	Predicted No Piping	Total
1993 Piping	Piping	11	13	24
	No Piping	251	74	325
	Total	262	87	349
Overall Correct Prediction Accuracy		24.36%		

Two by two contingency table showing the overall accuracy of the Medium category.

Table 17. PDR Pure Limited Model B

PDR Pure Limited Model B		High		
		Predicted Piping	Predicted No Piping	Total
1993 Piping	Piping	5	7	12
	No Piping	19	318	337
	Total	24	325	349
Overall Correct Prediction Accuracy		92.55%		

Two by two contingency table showing the overall accuracy of the High category.

Table 18. PDR_FTC Pure Limited Model A

PDR_FTC Pure Limited Model A		Low		
		Predicted Piping	Predicted No Piping	Total
1993 Piping	Piping	25	14	39
	No Piping	171	417	588
	Total	196	431	627
Overall Correct Prediction Accuracy		70.49%		

Two by two contingency table showing the overall accuracy of the Low category.

Table 19. PDR_FTC Pure Limited Model A

PDR_FTC Pure Limited Model A		Medium		
		Predicted Piping	Predicted No Piping	Total
1993 Piping	Piping	24	15	39
	No Piping	417	171	588
	Total	441	186	627
Overall Correct Prediction Accuracy		31.10%		

Two by two contingency table showing the overall accuracy of the Medium category.

Table 20. PDR_FTC Pure Limited Model A

PDR_FTC Pure Limited Model A		High		
		Predicted Piping	Predicted No Piping	Total
1993 Piping	Piping	1	38	39
	No Piping	0	588	588
	Total	1	626	627
Overall Correct Prediction Accuracy		93.94%		

Two by two contingency table showing the overall accuracy of the High category.

Table 21. PDR Previous Unlimited Model CS

PDR Previous Unlimited Model CS		Low		
		Predicted Piping	Predicted No Piping	Total
1995 Piping	Piping	44	5	49
	No Piping	44	256	300
	Total	88	261	349
Overall Correct Prediction Accuracy		85.96%		

Two by two contingency table showing the overall accuracy of the Low category.

Table 22. PDR Previous Unlimited Model CS

PDR Previous Unlimited Model CS		Medium		
		Predicted Piping	Predicted No Piping	Total
1995 Piping	Piping	23	26	49
	No Piping	262	38	300
	Total	285	64	349
Overall Correct Prediction Accuracy		17.48%		

Two by two contingency table showing the overall accuracy of the Medium category.

Table 23. PDR Previous Unlimited Model CS

PDR Previous Unlimited Model CS		High		
		Predicted Piping	Predicted No Piping	Total
1995 Piping	Piping	21	28	49
	No Piping	6	294	300
	Total	27	322	349
Overall Correct Prediction Accuracy		90.26%		

Two by two contingency table showing the overall accuracy of the High category.

Table 24. PDR_FTC Previous Limited Model E

PDR_FTC Previous limited Model E		Low		
		Predicted Piping	Predicted No Piping	Total
1995 Piping	Piping	10	75	85
	No Piping	13	529	542
	Total	23	604	627
Overall Correct Prediction Accuracy		85.96%		

Two by two contingency table showing the overall accuracy of the Low category.

Table 25. PDR_FTC Previous Limited Model E

PDR_FTC Previous limited Model E		Medium		
		Predicted Piping	Predicted No Piping	Total
1995 Piping	Piping	3	82	85
	No Piping	533	9	542
	Total	536	91	627
Overall Correct Prediction Accuracy		1.91%		

Two by two contingency table showing the overall accuracy of the Medium category.

Table 26. PDR_FTC Previous Limited Model E

PDR_FTC Previous limited Model E		High		
		Predicted Piping	Predicted No Piping	Total
1995 Piping	Piping	7	78	85
	No Piping	4	538	542
	Total	11	616	627
Overall Correct Prediction Accuracy		86.92%		

Two by two contingency table showing the overall accuracy of the High category.

Table 27. PDR Raw Previous

PDR Raw Previous		1995 Piping		
		No	Yes	Total
1993 Piping	No	291	34	325
	Yes	9	15	24
	Total	300	49	349
Correctly Classified		306		
Accuracy of Model		87.68%		

Two by two contingency table showing the overall accuracy.

Table 28. PDR_FTC Raw Previous

PDR_FTC Raw Previous		1995 Piping		
		No	Yes	Total
1993 Piping	No	532	56	588
	Yes	10	29	39
	Total	542	85	627
Correctly Classified		561		
Accuracy of Model		89.47%		

Two by two contingency table showing the overall accuracy.

5.2.4 Geographic Information System Applications

Maps were created for both levee districts to show the locations of actual piping versus the piping potential categories determined through application of thresholds. The top Pure model and the top Previous model are shown for PDR field area (Figures 19, 20, 21, and 22).

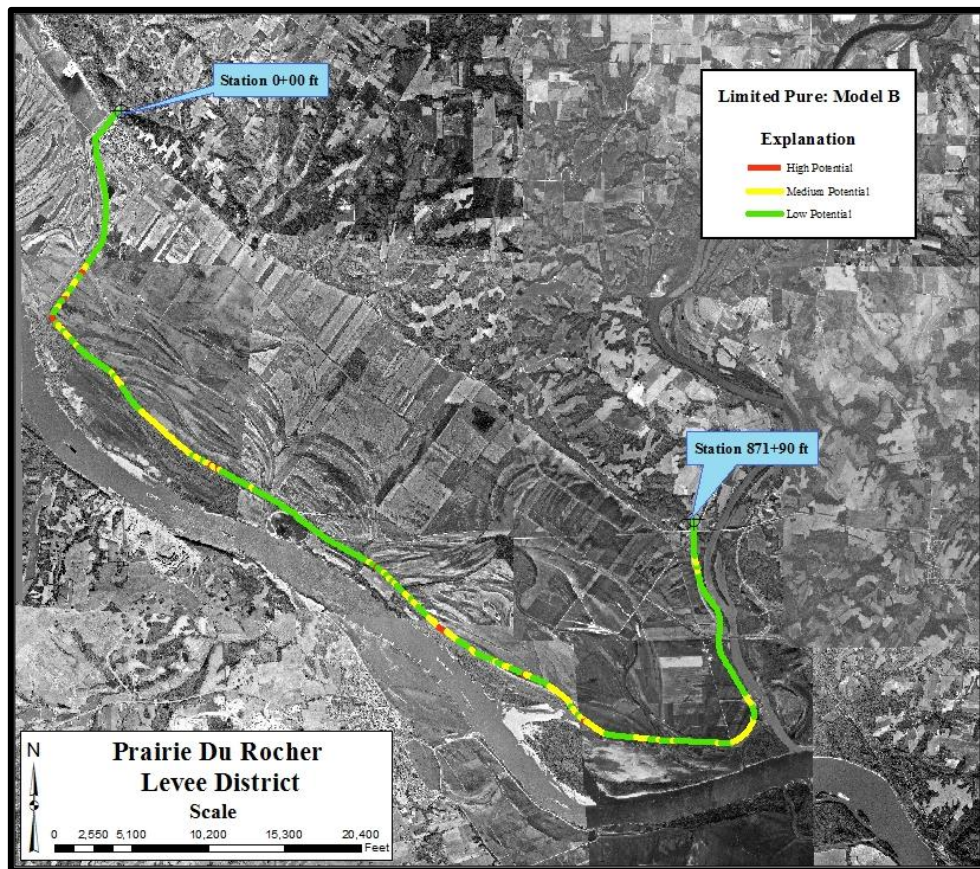


Figure 19. Map of piping potentials for Prairie du Rocher levee district determined from Limited Pure: Model B.

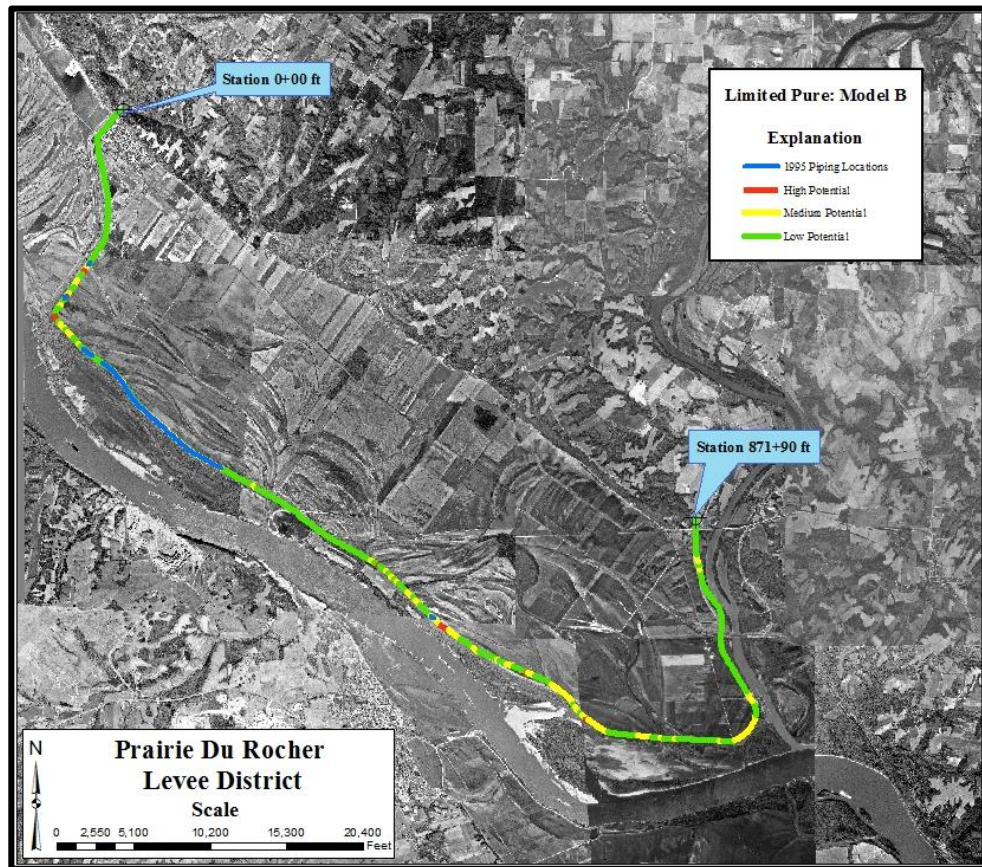


Figure 20. Map of actual piping locations atop piping potentials for Prairie du Rocher levee district determined from Limited Pure: Model B.

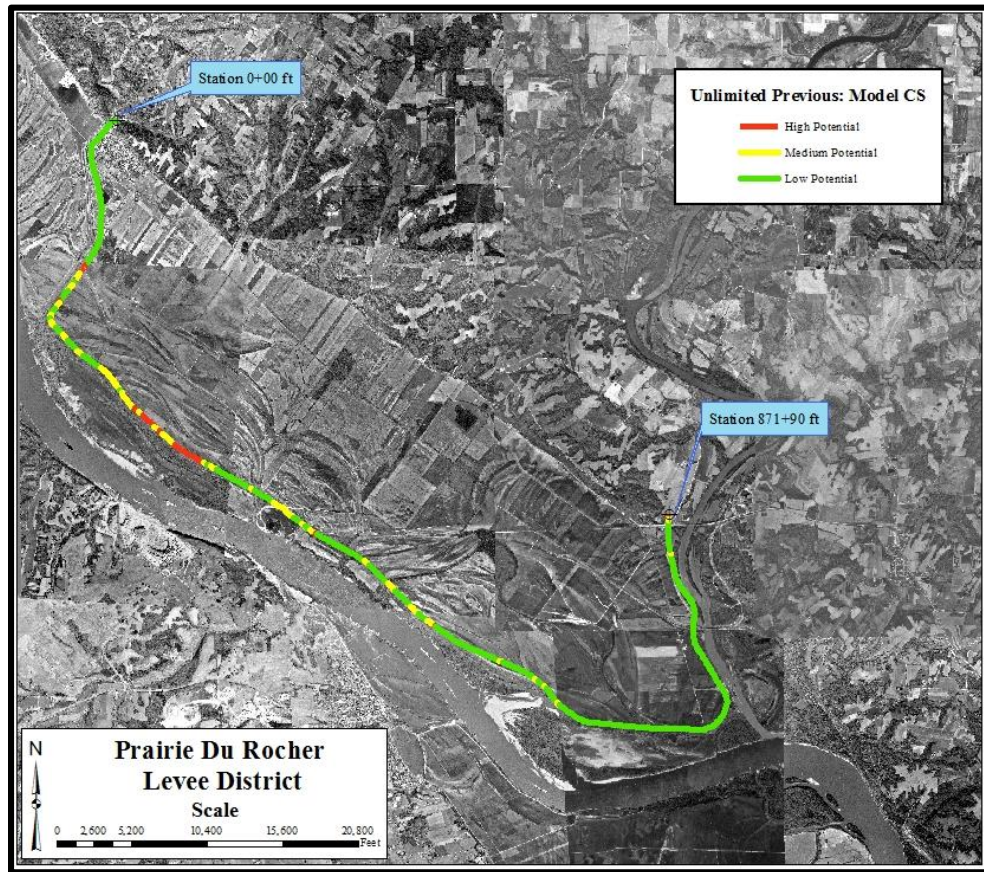


Figure 21. Map of piping potentials for Prairie du Rocher levee district determined from Unlimited Previous: Model CS.

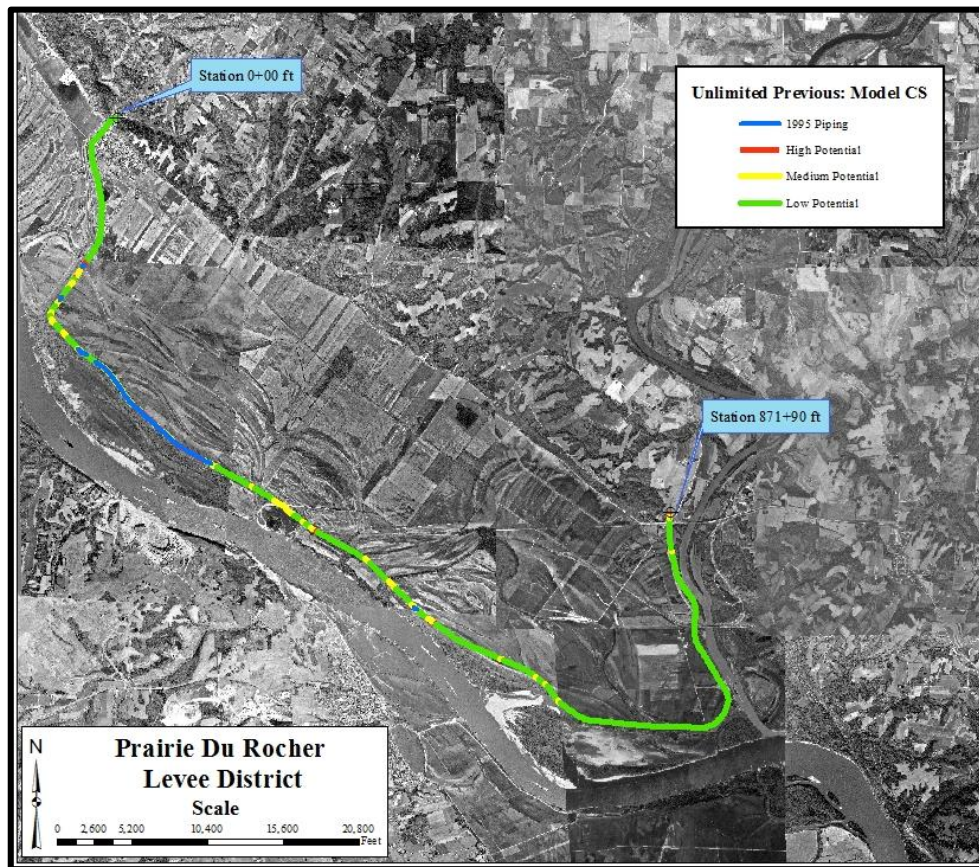


Figure 22. Map of actual piping locations atop piping potentials for Prairie du Rocher levee district determined from Unlimited Previous: Model CS.

5.3 Model Utility: Blind Testing

Blind testing of the four selected models, conducted on ECG Levee District, was restricted to 50 randomly chosen segments along the levee. However, that number was further limited by missing flood data from the 1993 event (i.e., h_0 and i_c were not calculated by USACE, St. Louis District). A smaller, but complete, dataset of 28 segments was applied to Unlimited models whereas the larger dataset of 50 segments was applied to the Limited models that do not need h_0 and i_c to run.

Figure 23 through Figure 26 display the performance of each model under their respective thresholds. A similar trend seen in the initial model utility tests can be partially said of the blind tests for model utility. The top Limited model, Previous PDR_FTC Model A, is the most accurate model in predicting locations where piping will occur (e.g., “high potential for piping”). However, during initial model utility tests, the top Limited model was outperformed by the top Unlimited model, Previous PDR Model CS, in predicting where piping will *not* occur (e.g., “low potential for piping”).

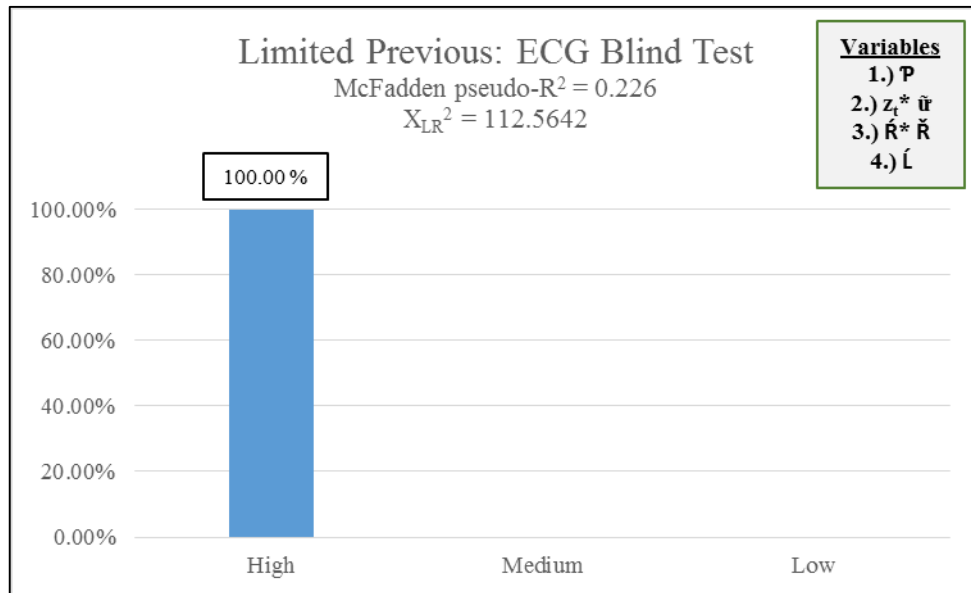


Figure 23. Bar chart for the application of the Limited Previous model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

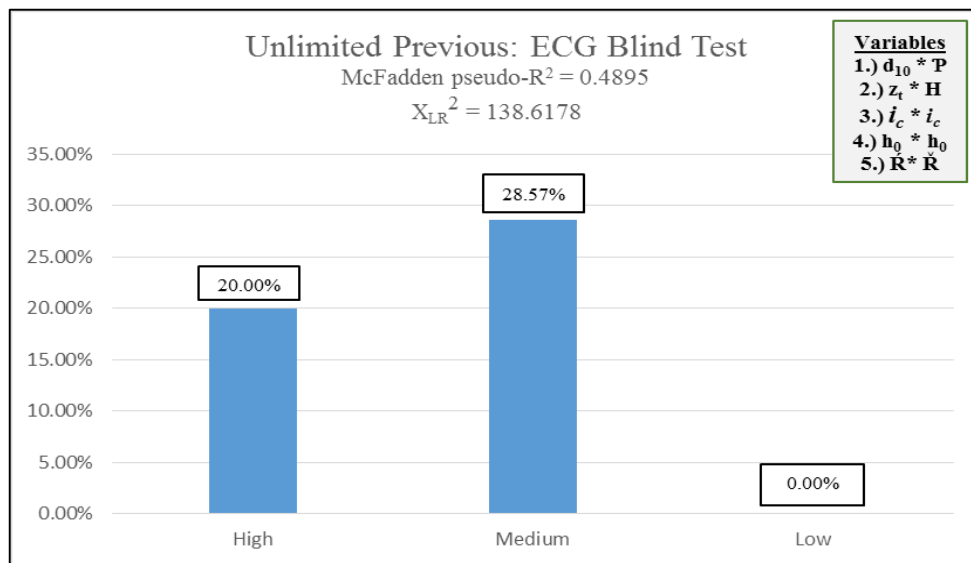


Figure 24. Bar chart for the application of the Unlimited Previous model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

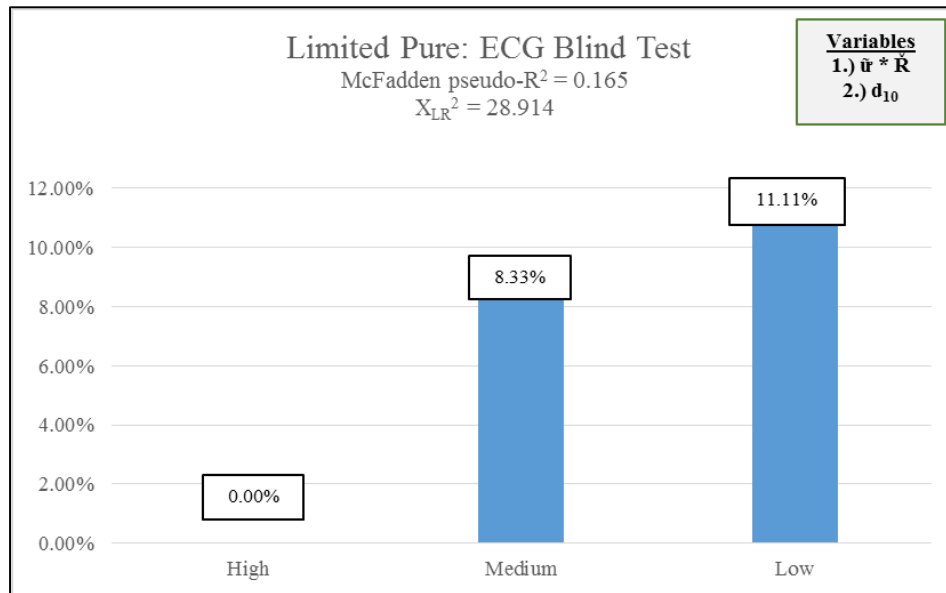


Figure 25. Bar chart for the application of the Limited Pure model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

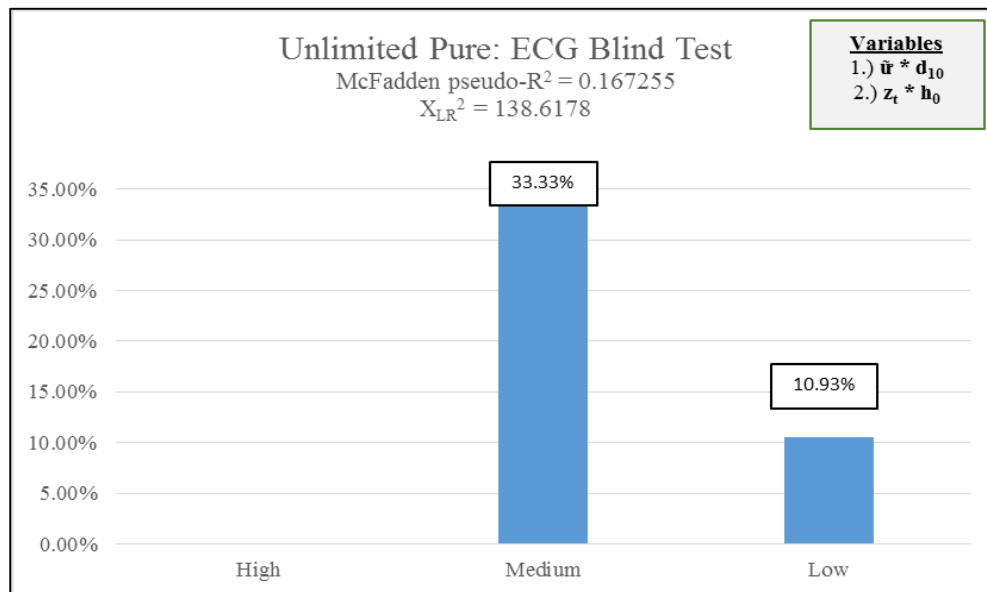


Figure 26. Bar chart for the application of the Unlimited Pure model to East Cape Girardeau Levee District for the percentage piped in each category of piping potential, i.e. high, medium and low potential for piping.

For the blind tests, both top models, Limited and Unlimited, accurately predicted segments with a “low potential for piping” where both models achieved a 0.00% piped. The top Unlimited model was unable to distinguish between areas of high potential and areas of medium, where 20.00% of segments piped and 28.57% of segments piped, respectively. The inconsistency in high versus medium potential can be attributed to the small amount of data points analyzed. The worst performing model, Limited Pure dataset PDR Model B, was unable to predict any category of piping potential: high=0.00%, medium=8.33%, and low=11.11%. Finally, the last model applied to ECG, Unlimited Pure dataset PDR Pure Model A, was also unable to predict areas of “high potential for piping”, but was more accurate in predicting medium and low potential: medium=33.33% and low=10.53%. The Unlimited Previous model was able to predict a “low potential for piping” (0.00%) but was unable to distinguish between high and medium potential, 20.00% piped and 28.57% piped, respectively. Neither Unlimited Pure or Limited Pure were able to predict areas of a “high potential for piping”

The Limited Previous model performed the best for categories of “high potential for piping” (100% piped) and “low potential for piping” (0.00%). No segments were assigned to the “medium potential for piping” so that value is 0.00% as well. This model outperformed the other models for every assessment of model utility and is selected as the top model developed during this study.

CHAPTER SIX

Conclusions

The model building process presented in this study proved to be a successful method for developing regression models meant to predict the potential for piping given the availability or lack of geologic and flood specific data. By determining the available data in a given levee district, one may use the processes presented to develop a model based off the distinctions between the core four datasets: Unlimited, Limited, Previous, and Pure. While the Raw Previous model was less successful than the models created using the forward regression procedure, it may still be considered an effective way of predicting piping when no other data are available.

6.1 Model Selection: Conclusions

Out of the four selected models, the top Limited Previous had the greatest accuracies for both percentage piped and blind testing. The model contains only four independent variables, two original independent variables, and two interaction terms, a $\chi^2_{LR}(344, N=349) = 121.51$, and a McFadden psudeo- $R^2 = 0.226$. The goodness-of-fit values are considered unbiased due to the adherence of the step-wise procedure, the small number of final independent variables present, and the elimination of co-linearity during the model building process. This model can be used for general areas containing data on previous piping events. Because it is a Limited model, flood specific variables are not required and applicability of the model is much greater than Unlimited

dataset models. Unlike the other three models, this model does not contain \mathbf{d}_{10} , effective aquifer grain size, but does contain $\mathbf{\tilde{L}}$, the presence of landside ditches, which is not found in any other model. However, the significance of \mathbf{d}_{10} greatly outweighs the significance of $\mathbf{\tilde{L}}$ overall.

The performance of the Limited Previous model was very similar to the Unlimited Previous model where $\chi^2_{LR}(344, N=349) = 144.31$ and the McFadden pseudo- $R^2 = 0.490$. However, the Unlimited Previous model performed the best out of the two Previous models under contingency table analysis. Both models incorporated previous piping events, \mathbf{P} , and their interaction terms into their datasets. The Unlimited model was not chosen as the best performing model due to the datasets requirement of calculated independent variables (e.g., $\mathbf{h_0}$ in the dataset and more independent variables present in the final model, leading to a greater possibility of a false positive McFadden pseudo- R^2). However, the Unlimited Previous model's performance was very satisfactory, where only 1.92% of segments piped in the “low potential for piping” and 77.78% of segments piped in the “high potential for piping” category.

Pure models were the least accurate out of the selected models in reference to percentage piped and blind testing. However, they were still able to predict areas of high, medium, and low potential for piping and overall accuracies in contingency table analysis were higher than other models. Therefore, pure models may not be disregarded as unusable.

The Limited dataset is chosen for the best fit model, using just two easily obtainable independent variables to predict piping potential: $(\mathbf{\tilde{U}} * \mathbf{\tilde{R}})$ and \mathbf{d}_{10} . The Unlimited Pure model achieved roughly the same accuracies in the piping potential categories; however, because independent variables required for the model are from the Unlimited dataset, they are more difficult to obtain in real world situations.

6.2 Blind Testing Findings

Out of the four selected models, the Limited Previous model once again performed the best for the ECG Levee District, where 100% of piped segments were predicted in the “high potential for piping” category, no segments were predicted in the “medium potential for piping” category, and all segments predicted in the “low potential for piping” category were non-piped locations. The other models tested were unable to accurately predict the areas of high, medium, and low potential.

6.3 Recommendations

The final Limited Previous model outperformed all other models for all model utility assessments and is recommended for application to other levee districts (Table 11).

Table 29. Description of final selected model, Limited Previous: Model A.

Model Name	Model Test	Model Significance	Independent Variables	Coefficient Value	P-value
Limited Previous: Model A	Chi-square Statistic	112.5642	Intercept	-2.1970	<0.0001
			Previous Piping, P	3.7360	<0.0001
			Confining layer thickness * Unfavorable Geologic Condition, $(z_t * \tilde{u})$	0.0490	0.0920
	Mc Fadden psuedo-R2	0.165	Relief Wells * Riverside Borrow Pits, $(\hat{R} * \hat{R})$	-2.7450	0.0000
			Landside Ditches, \hat{L}	-1.4870	0.147

Previous piping has proven itself to be an important factor in the prediction of piping potential along levees. This finding is consistent with those made by Wilson (2003).

While this variable is difficult to ascertain due to the challenge of detecting and recording such data, its significance to future piping events is undeniable. During future flood fight efforts along levees of the region, levee inspectors and local parties are encouraged to record any and all observed piping or sand boil locations. This will provide the best possible prediction for any particular levee.

The undesirable performance of the Pure models during blind testing suggests further model building is needed for higher accuracies using those types of datasets. These models' value was proven during contingency table analysis and therefore, should not be disregarded. Variables not yet considered may improve these models, which are very valuable to levee districts where piping has never been documented.

Overall model performance for models created from the defined types of dataset can be improved by using the methods presented in this study on larger datasets which may or may not include new independent variables for analysis.

6.4 Recommendations on Future Research

While the models created during this research are capable of implementation immediately, further research could improve these models and their accuracy. A smaller sized levee reach would aid in a more detailed dataset and a closer fit along the regression line. Also,

other geologic and flood specific variables could be obtained to include in the forward regression procedure itself, e.g. levee dimensions, swale thickness, aquifer characteristics, continued observation of piping events, etc. Finally, several more levees should be analyzed and the models applied to the data for further confirmation.

LIST OF REFERENCES

- Anfinson, J. O., 2003, River of History: A historic resources study of the Mississippi National River and Recreation Area: USACE, St. Paul District, St. Paul, MN.
- Bhowmik, Nani G., Andrew G. Buck, Stanley A. Changnon, Robert H. Dalton, Ali Durgunoglu, Misganaw Demissie, Arlan R. Juhl, H. Vernon Knapp, Kenneth E. Kunkel, Sally A. McConkey, Robert W. Scott, Krishan P. Singh, Ta-Wei David Soong, et al. 1994, The 1993 Flood on the Mississippi River in Illinois: ISWS Miscellaneous Publication 151, Illinois State Water Survey, Champaign, IL
- Chrzastowski, M.J., Killey, M.M., Bauer, R.A., DuMontelle, P.B., Erdmann, A.L., Herzog, B.L., Masters, J.M., and L.R. Smith, 1994, The Great Flood of 1993: Geologic perspectives on the flooding along the Mississippi River and its tributaries in Illinois: Champaign, IL, Illinois State Geological Survey, Special Report 2, Digitized by: Internet Archive in 2012 with funding from University of Illinois Urbana-Champaign, <http://archive.org/details/greatfloodsof199302illi>
- Curtis, G. E., 2005, Military compensation background papers: Washington, DC, Dept. of Defense, Under Secretary of Defense for Personnel and Readiness, 6th ed., p. 944
- Davis, John C., 2002, Statistics and Data Analysis in Geology: 3rd ed., New York, New York, John Wiley & Sons, Inc.
- Dunn, I.S., Anderson, L.R., and Kiefer, F.W., 1980, Fundamentals of Geotechnical Analysis: New York, New York, John Wiley & Sons, Inc., p. 414.
- Fisk, H.N., 1941, Application of geological studies to underseepage problems in the Lower Mississippi Valley: Mississippi River Commission, Vicksburg, Mississippi.

- Fisk, H.N., 1944, Results of geological investigations of the Alluvial Valley of the Lower Mississippi River: Technical Lecture 3, United States Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Mississippi.
- Joint Departments of the Army, the Air Force, and the Navy, USA (HDAAF-USA), 1983, Technical Manual No. 5-818-5 Air Force Manual No. 88-5, Chapter 6, Navy Manual, No. P-418: Washington, DC, United States Government
- Kolb, C.R., 1975, Geologic control of sand boils along Mississippi River levees: S-75-22, United States Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Mississippi.
- Le, Chap T., *Applied Categorical Data Analysis*: New York, New York: John Wiley & Sons, 1998. Print.
- Le, Chap T., *Applied Categorical Data Analysis and Translational Research*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, 2011. Print.
- Lewis-Beck, M.S., 1995, *Data analysis: an introduction*. Thousand Oaks: Sage Publications, 1995. Print
- Mac, M. J., P. A. Opler, C. E. Puckett Haecker, and P. D. Doran. 1998. Status and trends of the Nation's biological resources. Vol. 1. U.S. Department of the Interior, U.S. Geological Survey, Reston, Va. 1-436 pp.
- Mela, C. F. and P. K. Kopalle, 2002, The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations: *Applied Economics*, Taylor & Francis Ltd, DOI: 10.1080/000368-40110058482

Mertler, C.A., and R. A. Vannatta. *Advanced and Multivariate Statistical Methods: Practical application and interpretation*. 2nd ed. Glendale, CA: Pyrczak, 2010. Print

National Committee on Levee Safety (NCLS), 2009, Draft: Recommendations for a National Levee Safety Program: A report to Congress from the National Committee on Levee Safety: An involved public and reliable levee systems: National Committee on Levee Safety, p. 13-16

Natural Hazards Research and Applications Information Center, University of Colorado at Boulder (NHRAIC-UCB), 1992, Floodplain management in the United States: an assessment report prepared for the Federal Interagency Floodplain Management Task Force: Washington, DC, Federal Emergency Management Agency (FEMA), v. 1, contract no. TV-72105A, p. 24-28, p. 57

Power, G., 1977, The Fox in the Chicken Coop: The regulatory program of the U.S. Army Corps of Engineers, 63 VA. L. REV. 503

Ozkan, S., 2003, Analytical Study of Flood Induced Seepage under River Levees, Ph.D. Dissertation, Louisiana State University, Department of Civil and Environmental Engineering

Ranjan, G. and A. S. R. Rao, 2005, *Basic and Applied Soil Mechanics*, New Age International (P) Ltd., New Delhi, India, 774 pp, pg 137

Turnbull, W. J., Krinitzsky, E. L., and S. J. Johnson, 1950, Sedimentary geology of the Lower Mississippi River and its influence on foundation problems: *Applied Sedimentation*, John Wiley & Sons, New York.

Turnbull, W.J. and Mansur C.I., 1959, Investigation of underseepage – Mississippi River levees.

Journal of the Soil Mechanics and Foundations Division, Proceedings of the American Society of Civil Engineers, vol. 8, p. 41-93.

U.S. Army Corps of Engineers (USACE), 2000, Design and construction of levees: EM-1111-1913, Department of the Army, Office of the Chief of Engineers, Washington, District of Columbia.

U.S. Army Corps of Engineers (USACE), 1956a, Investigation of underseepage, Mississippi River Levees, Alton to Gale, IL: TM-3-430, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Mississippi, 3 vols.

U.S. Army Corps of Engineers (USACE), 1956b, Investigation of underseepage and its control, Lower Mississippi River levees: TM-3-424, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Mississippi, 2 vols.

U.S. Army Corps of Engineers (USACE), 2010, Department of the Army, U.S. Army Corps of Engineers Civil Works Program, five-year development plan: Fiscal year 2011 to Fiscal Year 2015: U.S. Army Corps of Engineers, p. 34-54
<http://cdm16021.contentdm.oclc.org/cdm/ref/collection/p16021coll6/id/54>

Wilson, J.D., 2003, Middle Mississippi River levee flood performance: assessing the occurrence of piping through empirical modeling (Master's Thesis), University of Mississippi, University, Mississippi.

LIST OF APPENDICES

APPENDIX A: LOGISTIC REGRESSION

In general, regression analysis seeks a probabilistic relationship between a quantified dependent variable and one or multiple quantified independent variables, each bound to a specific location, whether in time or space (Davis, 2002). By doing so, the dependent variable may be estimated at any location for a given set of independent variable values. Several different types of regression analysis have been defined throughout the years: linear regression, curvilinear regression, orthogonal regression, logistic regression, etc. (Davis, 2002).

Regression analysis may be used in both univariate and multivariate statistics (Lewis-Beck, 1995). Univariate statistics are defined under two criteria: (1) represents types of single variable regression, which use data on one independent variable to estimate the value of a single dependent variable, and types of multiple variable regression, which use data on multiple independent variables to estimate the value of a single dependent variable (Davis, 2002). Multiple variable regression is often misappropriated under multivariate regression. Multivariate regression uses multiple independent variables to estimate the values of multiple dependent variables (MertlerVannatta, 2002).

Linear Regression

The simplest form of regression, univariate linear regression, uses a form of the general equation for a line to assign a value to the $y(x = 0)$ intercept, β_0 , and coefficients associated with j amount of independent variables, β_j for $j \in (1, k)$,

$$(1) \hat{y}_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji}$$

where \hat{y}_i equals the estimated value of the dependent variable, y_i for $i \in (1, n)$ where n represents the number of data points for the variables, and x_{ji} represents the values associated with the j^{th} independent variable for $i \in (1, n)$, (Davis, 2002; Le, 2010). The above equation is a representation of a multiple variable univariate regression analysis.

Coefficients are determined by backwards analysis on the general equation (1) used in linear regression analysis. The equation is multiplied by an additional x_i term, summed over all observations, and rearranged to solve for (2) β_j and (3) β_0 (Davis, 2002). This is easily accomplished by simple matrix algebra (Davis, 2002).

$$(2) \beta_j = \frac{\sum_{i=1}^n x_{ji} y_{ji} - \left(\left(\sum_{i=1}^n x_{ji} \right) \left(\sum_{i=1}^n y_{ji} \right) / 2 \right)}{\sum_{i=1}^n x_{ji}^2 - \left(\left(\sum_{i=1}^n x_{ji} \right)^2 / n \right)}$$

$$(3) \beta_0 = \frac{\sum_{i=1}^n y_{ji}}{n} - \beta_1 \frac{\sum_{i=1}^n x_{ji}}{n}$$

After finding the value of each coefficient, \hat{y}_i is easily interpolated between observations and goodness-of-fit can be determined (Davis, 2002). Plotting values of \hat{y}_i for $i \in (1, n)$ will ideally result in a normal distribution curve (Davis, 2002).

For example, values for \hat{y}_i in Wilson's (2003) linear regression models in the previous research conducted by Wilson (2000) is an estimated "piping value" ranging from 0 to 1, ideally follows normal distribution curve (Davis, 2002). The "piping value" for each location is an estimation determined by the y-intercept, β_0 , and the slope $\sum_{i=1}^n \beta_i$ as a function of $\sum_{i=1}^n x_i$.

If the model had a perfect goodness-of-fit, \hat{y}_i would equal either exactly 0 or exactly 1 with 100% accuracy. However, models are never perfect and \hat{y}_i values for piping are much more varied.

A very important distinction in linear regression analyses of piping occurrence is the value of \hat{y}_i does not equal the probability of piping for that section, e.g. $\hat{y}_i = 0.2$ does not mean there is a 20% chance of piping occurrence for that section. Rather, the value is a linearly scaled value in response to relationships between the independent data. When continuous variables are regressed in association with dichotomous, e.g. binary, variables, the functionality of linear regression models decreases and \hat{y}_i becomes less accurate and more difficult to interpret.

Logistic Regression

The benefit of using logistic regression instead of linear regression was touched upon in Section 1.3.1. Logistic regression is an attempt to determine the probability of the presence of a dependent dichotomous variable, π , by defining a relationship between multiple independent variables which can be either dichotomous or continuous (Davis, 2000; Le, 2000). The probability distribution for a random variable where the $Y=1$ has a probability of π and $Y=0$ has a probability of $(\pi-1)$ is expressed as

$$(4) \Pr(Y = y) = \pi^y(1 - \pi)^{1-y}$$

Using the concept of determining the probability of the presence of a variable instead of the estimation of the variable, the logistic function can be linearly expressed on the log scale (5),

$$(5) y_i = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{j=1}^k \beta_j x_{ji}$$

This form can be beneficial for dichotomous dependent variables because it mathematically transform a discontinuous variable into an S-shaped logarithmic curve which may be defined by continuous independent variables more easily (Le, 2010). After determining the y-intercept and weighted coefficients, the equation is rearranged into the form shown in equation (6),

$$(6) \pi_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_i)]}$$

This results in quantifiable probability values for the occurrence of a dichotomous dependent variable given the associated values of independent variables. Logistic regression is often utilized in biological and health sciences and has resulted in strong empirical support for its application (Le, 2010).

Testing Model Significance

Overall model significance for logistic regression is expressed by various forms of testing, e.g. likelihood ratio test, on the null hypothesis that” all k independent variables considered together do not explain the variation in the responses” of the dependent variable (Le, 2010).

$$(8) H_0: \beta_1 = \beta_0 = \dots = \beta_k = 0$$

This provides information on the significance the independent variables to the dependent variable.

Le (2010) lists three types of test for expressing model significance in logistic regression: an overall test, test for the value of a single factor, and test for contribution of a group of variables. The X^2 distribution is used for comparison in logistic regression model significance

tests (Le, 2010). A P-value is the probability that the same outcome could be obtained by using a non-unique or random variable (Davis, 2002). Therefore, a P-value < 0.05 corresponds with at least a 95% confidence level for the significance of the variable to the model.

Interaction Terms

These secondary factors are the basis for which the regression analyses are conducted. They will act as independent variables in an attempt to determine their correlation with piping as the dependent variables.

Interaction terms either take the form $x_i x_j$ or $x_i x_i$ where i and j designate each independent variable available. They describe effect modification in which one variable controls or modifies the effect of another variable. Because many geological variables are regionalized, interaction terms may be useful in regression analysis. Interaction terms are considered as an independent variable in the regression model and are found by the multiplication of one independent variable with another independent variable (Le, 1998). For example, if an interaction term between the presence of an unfavorable geologic unit and the thickness of the confining unit is found to be statistically significant to the dependent variable, then it may be useful in the regression model.

VITA

REBECCA SHIELDS

8 Beech Hills Tuscaloosa, AL 35404 (662)380-2162 rfshield@go.olemiss.edu

EDUCATION

M.S., Engineering Science, University of Mississippi, December 2015
 Emphasis: Geological Engineering
 Thesis: Empirical Modeling of Piping along Levees in Southwestern Illinois

B.S., Geological Sciences, University of Alabama, May 2011

EXPERIENCE

Research Internship with Univ of Mississippi: August 2014 to May 2015

Teaching Assistantship with Univ. of Mississippi: August 2013 to May 2014

Research Internship with Univ. of Alabama: August 2010 to May 2011

Research Internship with Univ. of Alabama: January 2009 to May 2009

Internship with TTL, Inc.: May 2008 to August 2008

Honors/Awards

2010 Outstanding Senior Award for Geological Sciences Department

 Advisory Board for Geological Sciences Dept.

 WM Ray Moore Phys Science Scholarship

2009 Chi Epsilon

 Geological Sciences Advisory Board Scholarship

Jelks Barksdale Memorial Scholarship

2008 Golden Key International Honor Society

2007 Engineering Supplement Scholarship

Phi Eta Sigma

Alpha Lambda Delta

Sigma Alpha Lambda

National Society of Collegiate Scholars

2006 President's Cabinet Scholarship

Alumni Scholar Award

Engineering Leadership Scholarship

