ASSESSMENT OF TERRAIN ATTRIBUTE MODELS FOR THE IDENTIFICATION OF EROSION PRONE AREAS SUITABLE FOR THE ESTABLISHMENT OF GRASSED WATERWAYS IN AN AGRICULTURAL FIELD SETTING IN THE OUT BLUEGRASS REGION OF KENTUCKY

Adam Clellon Pike
University of Kentucky, scpike2@uky.edu

Recommended Citation
https://uknowledge.uky.edu/gradschool_theses/544
The speed and accuracy of conservation planning could be improved if maps indicating areas where grassed waterways should be placed to reduce erosion could be easily created. For five central Kentucky fields, elevation data were obtained with real time kinematic (RTK) global positioning system (GPS) and from US Geological Survey (USGS) digital elevation models (DEMs). Terrain attributes were calculated from these datasets which were used as predictor variables for neural network and logistic regression analyses. Grassed waterway prediction models were developed with these analyses. The type of activation function, type of standardization procedure, number of neurons, number of preliminary runs, and number of hidden layers had little impact on the results of the neural network analysis. Logistic regression and neural network analyses produced similar erosion prediction maps. The type of flow direction algorithm used to calculate terrain attributes did not change prediction maps substantially. Grassed waterways could be predicted in most cases with the RTK data but only in some cases with the USGS data. This modeling approach was robust and could aid conservation planners in identifying suitable areas for waterways more efficiently if accurate elevation data can be acquired.

KEYWORDS: RTK, Terrain Attributes, Neural Networks, Logistic Regression, Erosion
ASSESSMENT OF TERRAIN ATTRIBUTE MODELS FOR THE IDENTIFICATION OF EROSION PRONE AREAS SUITABLE FOR THE ESTABLISHMENT OF GRASSED WATERWAYS IN AN AGRICULTURAL FIELD SETTING IN THE OUT BLUEGRASS REGION OF KENTUCKY

By

Adam Clellon Pike

Tom G. Mueller

Charles T. Dougherty

7/24/08
RULES FOR THE USE OF THESES

Unpublished theses submitted for the Master’s degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the thesis in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this thesis for use by its patrons is expected to secure the signature of each user.

Name

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

__________________________________________

Date

__________________________________________
THESIS

Adam Clellon Pike

The Graduate School

University of Kentucky

2008
ASSESSMENT OF TERRAIN ATTRIBUTE MODELS FOR THE
IDENTIFICATION OF EROSION PRONE AREAS SUITABLE FOR THE
ESTABLISHMENT OF GRASSED WATERWAYS IN AN
AGRICULTURAL FIELD SETTING IN THE OUT BLUEGRASS
REGION OF KENTUCKY

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
College of Agriculture
at the University of Kentucky

By

Adam Pike

Lexington, Kentucky

Director: Dr. T. G. Mueller, Associate Professor of Plant and Soil Sciences

Lexington, Kentucky

2008
To my parents,

Gary and Robin Pike,

my brother,

Andy Pike,

my grandmother,

Gladys Buckler,

my grandfather,

Earl Pike,

for the unconditional love and support which they have so graciously given me.
I am extremely grateful to Dr. Tom Mueller for his guidance and friendship throughout my time at the University of Kentucky. Appreciation is also expressed to my graduate committee: Dr. Scott Shearer for helping me to understand the applied aspects of my research, Dr. Tasos Karathanasis for strengthening my understanding of soils both in the classroom and in the field, and Dr. Thomas Nieman for his insight throughout my thesis. I am also grateful to Blazan Mijatovic for the countless hours spent in the field. I would also like to thank Mike Ellis for providing access to his farm to conduct this research. Randall Rock for going out to the fields and evaluating many of the grassed waterways. Jack Kuhn and Danny Hughes for their information and discussion on grassed waterways. I finally thank Dr. Srinivasan and Dr. Viswanathan for their help with many neural network questions.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................................................ iii

LIST OF TABLES ................................................................................................................................................ vii

LIST OF FIGURES ........................................................................................................................................... x

CHAPTER ONE: INTRODUCTION ........................................................................................................................ 1

Objectives .......................................................................................................................................................... 3

Hypotheses ....................................................................................................................................................... 3

Organization of Thesis .................................................................................................................................... 4

CHAPTER TWO: LITERATURE REVIEW ........................................................................................................... 5

Real Time Kinematic (RTK) Global Positioning Systems (GPS) ................................................................. 5

Digital Terrain Models .................................................................................................................................... 6

Digital Terrain Analysis .................................................................................................................................. 8

Grassed Waterways ........................................................................................................................................ 11

Conservation Reserve Program (CRP) ........................................................................................................... 12

Continuous CRP ........................................................................................................................................... 13

General CRP ................................................................................................................................................ 14

Artificial Neural Network Models ................................................................................................................ 14

Logistic Regression Models .......................................................................................................................... 18
LIST OF TABLES

Table 1. Soil Survey information for each study area. ..................................................... 26

Table 2. Misclassification results for neural network (NN) models created using RTK data with different numbers of neurons and with the following variables held constant: Hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model). .................................................................................................................. 37

Table 3. Misclassification results for neural network (NN) models created using RTK data with different activation functions and with the following variables held constant: 20 neurons, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model). ................................. 41

Table 4. Misclassification results for neural network (NN) models created using RTK data with different normalization procedures and with the following variables held constant: 20 neurons, hyperbolic activation function, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model). ................. 42

Table 5. Misclassification results for neural network (NN) models created using RTK data with different number of preliminary runs and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model). ................................. 43
algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model). ................................................................. 44

Table 6. Misclassification results for neural network (NN) models created using RTK data with different number of hidden layers and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model). ................................................................. 45

Table 7. Misclassification results for neural network (NN) models created using RTK data with different models consisting of various combinations of terrain attributes and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model). ........................................ 47

Table 8. Misclassification results for neural network (NN) and logistic regression (REG) models created using RTK data with different flow direction algorithms used and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer (Fields 1 and 2 were used to validate and train the model and Field 3, 4, and 5 was used to test the model). ................................................................. 51

Table 9. Misclassification results for neural network (NN) and logistic regression (REG) models created using RTK and USGS data for the different number of fields used to train and validate the model and with the following variables held constant: 20
neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm).
LIST OF FIGURES

Figure 1. Layout of an artificial neural network model comprising of an input layer with 8 variables, one hidden layer with 6 neurons and an output layer.......................... 16

Figure 2. Layout of overall analysis represented in this thesis........................................... 19

Figure 3. Boundary of Field 1 (57 hectares) from 2007 overlain on a 2004 aerial image
(Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from

Figure 4. Boundary of Field 2 (36 hectares) from 2007 overlain on a 2004 aerial image
(Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from

Figure 5. Boundary of Field 3 (23 hectares) from 2007 overlain on a 2004 aerial image
(Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from

Figure 6. Boundary of Field 4 (33 hectares) from 2007 overlain on a 2004 aerial image
(Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from

Figure 7. Boundary of Field 5 (11 hectares) from 2007 overlain on a 2004 aerial image
(Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from

Figure 8. Terrain attributes for the test dataset (Field 3) calculated from the RTK GPS and
USGS DEMs. (Upslope Contributing Area, m2; Plan Curvature, radians/100 m). ................................................................. 34
Figure 9. Terrain attributes for the test dataset (Field 3) calculated from the RTK GPS and USGS DEMs. (Profile Curvature, radians/100 m)................................. 34

Figure 10. Comparison maps of grass waterways created with different number of neurons for the Field 3 test data set. The FD8 algorithm was used to calculate flow direction and the data were normalized with the standard deviation. For the neural network analyses, 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 2.......... 39

Figure 11. Comparison maps of grass waterways created with different normalization methods used for the Field 3 test data set. The FD8 algorithm was used to calculate flow direction. For the neural network analyses 20 neurons, hyperbolic activation function, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 4. ........................................ 43

Figure 12. Comparison maps of grass waterways created with different variable scenarios for the Field 3 test data set. The FD8 algorithm was used to calculate flow direction. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 7. ................................................................................................................. 49

Figure 13. Upslope contributing area maps for the RTK and USGS data created with different flow direction algorithms (i.e., D8, FD8, and Stream-Tube Demon). ... 53

Figure 14. Comparison maps of grass waterways created with different flow direction algorithms for the Field 3 test data set (RTK Data). For the neural network
analyses 20 neurons, hyperbolic activation function, normalization by the
standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps
correspond to the comparisons made in Table 8 (for test dataset 3)................. 54

Figure 15. Comparison maps of grass waterways created with different flow direction
algorithms for the Field 3 test data set (9.1-m USGS DEMs). For the neural
network analyses 20 neurons, hyperbolic activation function, normalization by the
standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps
correspond to the comparisons made in Table 8 (for test data set 3)................. 55

Figure 16. Comparison maps of grass waterways created with the RTK and USGS
datasets for Field 1. For the neural network analyses 20 neurons, hyperbolic
activation function, normalization by the standard deviation, 5 preliminary runs, 1
hidden layer, and the FD8 flow direction algorithm were used. These maps
correspond to the comparisons made in Table 9.................................................. 57

Figure 17. Comparison maps of grass waterways created with the RTK and USGS
datasets for Field 2. For the neural network analyses 20 neurons, hyperbolic
activation function, normalization by the standard deviation, 5 preliminary runs, 1
hidden layer, and the FD8 flow direction algorithm were used. These maps
correspond to the comparisons made in Table 9.................................................. 58

Figure 18. Comparison maps of grass waterways created with the RTK and USGS
datasets for Field 3. For the neural network analyses 20 neurons, hyperbolic
activation function, normalization by the standard deviation, 5 preliminary runs, 1
hidden layer, and the FD8 flow direction algorithm were used. These maps
correspond to the comparisons made in Table 9.................................................. 59
Figure 19. Comparison maps of grass waterways created with the RTK and USGS datasets for Field 4. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm were used. These maps correspond to the comparisons made in Table 10.

Figure 20. Comparison maps of grass waterways created with the RTK and USGS datasets for Field 5. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm were used. These maps correspond to the comparisons made in Table 9.
CHAPTER ONE: INTRODUCTION

Accelerated soil erosion is a major concern in areas of the world where crop production occurs on rolling and steep terrain. Erosion is the process of displacing solids usually by means of water, wind, or ice. As erosion occurs, topsoil is swept away leaving a thin A-horizon behind. This negatively affects crop productivity. Further, less biomass is produced and incorporated into the soil which means that there will be less organic material protecting the soil from the erosive forces of water and eroded particles. Therefore, past erosion makes the soil even more susceptible to future erosion. This all translates into an economic loss that may not pay for the cost of planting crops in the short term and definitely is not sustainable in the long term. It would be more beneficial to protect areas prone to erosion or leave them out of production.

Water erosion losses are 2.4 tons of soil per acre per year on average in Kentucky which is relatively low considering that 54% of the crops are planted on highly erodible land (HEL) (NRCS, 2007). This is because no-till adoption is fairly high in Kentucky (e.g., 51% in 2002) (Core4, 1996). Nevertheless, erosion is still a problem in Kentucky. For example, nearly 19% of Kentucky cropland lost soil in 2003 at rates greater than the "tolerable" levels (NRCS, 2007), indicating that erosion is occurring faster than replacement rates in these areas.

One way to prevent soil losses on both tilled and untilled agricultural fields is to install grassed waterways. A grassed waterway is a “a natural or constructed channel that is shaped or graded to required dimensions and established with suitable vegetation” in order 1) “to convey runoff from terraces, diversions, or other water concentrations without causing erosion or flooding”, 2) to reduce gully or ephemeral gully erosion, or 3)
“to protect/improve water quality” (NRCS, 2003). Studies have shown that waterways can substantially reduce soil erosion (Briggs et al., 1999; Chow et al., 1999; Fiener and Auerswald, 2003) as described more fully in the literature review.

There are too few waterways in Kentucky but the exact number is difficult to quantify. Relatively few land-owners receive payments from the Federal Government for installing and maintaining grassed waterways through the Conservation Reserve Program (CRP). According to the Farm Service Agency, there are only 4,100 acres of grassed waterways in the conservation reserve program (FSA, 2008). This is equivalent to only about 0.08% of the cultivated land in Kentucky. However, many landowners that do not receive payments still use waterways but these numbers are not recorded. There are many fields in Kentucky with rolling topography that do not have grassed waterways protecting them from channel erosion.

To determine where waterways should be placed, an NRCS conservationist must make an on-farm site assessment which involves walking across each field examining potential areas for eroded channels. To make assessments on all fields in Kentucky and across the United States, the NRCS would be overwhelmed. It would be helpful to the land-owners, farmers, the NRCS, and the FSA if there were tools (i.e., maps) to help identify where channel erosion might occur.

Some of the tools that can help an NRCS planner identify areas potentially prone to erosion such as topo-maps, aerial imagery, and soil surveys. However, older aerial images may not show the erosion patterns that exist today. Erosion patterns may be difficult to discern in newer (i.e., less than 20 years old) images if there is substantial no-till residue on the surface. Soil surveys lack enough detail to identify all of the erosion
patterns in the field. Contour maps can describe some but not all drainage ways within a field and can be difficult to interpret for making accurate estimates.

One of the best ways to estimate channel flow is with terrain attributes as described in the literature review (Tomer et al., 2003; Srivastava and Moore, 1989; Thorne et al., 1986; Daba et al., 2003; Kheir et al., 2007). The speed and accuracy of conservation planning could be improved if digital terrain analysis data and site-specific sensors were used to aid with the identification areas that are most susceptible to erosion. Then NRCS conservationists could specifically target those areas. Farmers that do not wish to enroll ground into the CRP program may also use these tools to help them identify areas that are prone to erosion within a field. Statistical methods such as logistic regression and neural networks may help with the analyses of this information to produce simple maps such that planners and farmers can easily determine the appropriate location for grassed waterways.

**Objectives**

The objective of this study is to create a set of tools to help conservation planners identify potential locations for grassed waterways. The tools will include training artificial neural networks and logistic regression models to locate the ideal placement of waterways from digital terrain attributes.

**Hypotheses**

In this thesis, the following hypotheses were tested.

\[ H_{a1} \text{ Terrain attributes can be used to identify erosion channels eligible for enrollment in the CRP program.} \]
Hₐ₂ Models used to identify grassed water ways produce similar results regardless of whether RTK or USGS data are used as input.

Hₐ₃. Neural network models are superior to logistic regression models for identifying suitable grassed waterway locations.

**Organization of Thesis**

This thesis will first discuss the literature associated with the various technologies, conservation practices, and analytical techniques required for this thesis (Chapter 2). The Materials and Methods section (Chapter 3) will detail the procedures, collect and obtain data, calculate digital terrain attributes, and develop the neural network and logistic regression models. The Results and Discussion section (Chapter 4) will describe the comparison of terrain attributes calculated with USGS and RTK elevation data, sensitivity analysis of neural network modeling, model variable selection, differences between neural network and logistic regression, impact of flow direction algorithms, comparison of models created using USGS and RTK data, and the analysis of using four fields to train and validate the model while using the fifth field as an independent test data set. The summary and conclusions will be provided in Chapter 5 followed by future research questions in Chapter 6.
CHAPTER TWO: LITERATURE REVIEW

This literature review will discuss the major technologies and tools used to conduct the research presented in this thesis. These include Real Time Kinematic (RTK) Global Positioning Systems (GPS), digital elevation models (DEMs), and terrain analysis. In addition, various aspects of the CRP will also be discussed along with statistical procedures used in the prediction of grassed waterways.

Real Time Kinematic (RTK) Global Positioning Systems (GPS)

With technological advances in agriculture, more precise measurements can be acquired with RTK GPS receivers. Standard GPS uses coded time information from satellite signals to calculate position. RTK GPS uses the carrier phase measurements to eliminate some of the errors associated with carrier measurements. In other words, it observes the attributes (i.e., phase) of the actual waves that are used to transmit the GPS signal (Tyler, 1993).

RTK receivers are frequently utilized in everyday farming practices such as planting, spraying, and harvesting crops. RTK provides positioning information that has less than 1 cm horizontal and 2 cm vertical relative accuracy (Personal communications, T.S. Stombaugh, 2008). These elevation measurements can be collected on-the-go while traversing a field resulting in an intensive grid of points.

Most base stations require a rover within 10 kilometers to ensure centimeter level accuracy. When outside that range, the base station should be moved closer to the rover. The base station broadcasts the phase of the carrier it measures to the rover unit. Then the Rover uses this information to make accurate estimates of position. This was the
approach used to make the elevation measurements used for this thesis. The mathematics used for these calculations are complex and are described by Strang and Borre (1997).

Now, however, a virtual reference station (VRS) network can be used (in lieu of a base station) to provide RTK correction. Information from several reference stations are processed to generate corrections that are valid over a much larger region (Wanninger, 2003). Establishing this kind of network involves placing RTK base stations at different locations throughout the area of interest. Satellite data received at the base stations are transmitted to a central location. Then, software is used to model corrections for regions between base stations within the network. This approach allows coverage over wider areas with high levels of accuracy and a relatively low density of network base stations. These systems can be very expensive but this could actually lower overall costs if the corrections are used by a large number of users for many different applications. One benefit of using a VRS network is increased reliability and availability (Fotopoulos and Cannon, 2001) with fewer base stations. If one base station goes down, then the remaining stations continue to provide acceptable corrections.

**Digital Terrain Models**

A digital elevation model (DEM) is a mathematical representation of ground surface topography. A DEM is a raster (regular grid) of elevation points. Alternative representations of elevation surfaces include triangular irregular networks (TINs) and contours.

There are currently varying resolutions of freely available data that may be obtained from the internet. The U.S. Geological Survey (USGS) currently has the contiguous United States, Hawaii, Puerto Rico, and Alaska mapped out at the UTM-
based 7.5-minute (30m) scale. The 7.5-minute DEM data are produced in 7.5- by 7.5-minute blocks either from digitized cartographic map contour overlays or from scanned National Aerial Photography Program (NAPP) photographs (USGS, 2006). These contours were obtained from scanning and digitizing 1:24,000 USGS topographic maps. The vertical accuracy of these 30-m DEMs is equal to or better than 15 meters. Most of the 7.5-minute DEMs produced thus far are categorized as Level-1 DEMs which have a RMSE accuracy standard between 7 and 15 meters (USGS, 2006).

Some USGS DEMs have been created at finer scales. In Kentucky, 10-m DEMs were developed for the entire commonwealth by the USGS under contract for Kentucky Division of Geographic Information (KDGI). These were re-interpolated by the KDGI on 9.1-m (30-ft) grids. The source data used by the USGS to create these DEMs were the same digitized cartographic map contour overlays created from the topographic maps (Demetrio Zourarakis, personal communications, 2007) that were used to create the 30-m DEMs. These 9.1-m DEMs are of interest because they were one of the data sources used in this thesis.

While RTK elevation measurements are considered to be very accurate, few if any studies have been conducted to quantify the absolute errors of mobile RTK elevation. One study did however compare RTK data with DEMs created from a photogrammetric survey, a DGPS RTK (virtual base station), DGPS unit (beacon base), single survey-grade GPS (beacon base), 10 ft. contour lines (USGS), and 30-m USGS DEMs (Renschler and Flanagan, 2008). They found that the more precise measurements made with the RTK, photogrammetric survey (TIN), and DGPS yielded more precise on-site
soil loss predictions. At the smaller watershed scale, the USGS 10 ft. contour lines are just as good as the most accurate data (Renschler and Flanagan, 2008).

With elevation measurements obtained from RTK GPS, highly accurate DEMs can be created. RTK provides a more accurate and precise elevation model versus the USGS DEMs. One study compared the RTK data against the 9.1-m and 30-m USGS DEMs at 2 locations within central Kentucky (Pike et al., 2006). They found that visually the USGS DEMs matched up very well with the RTK data but since several of the depressional areas were somewhat shifted that many of the correlations between the two were poor.

**Digital Terrain Analysis**

Terrain analysis has many applications in hydrology (topographic influences on soil moisture and runoff behavior), geomorphology (pedological and geomorphologic), and biology (topographic influences on vegetative cover) (Wilson and Gallant, 2000). Digital terrain attributes are landscape parameters that describe the relief of the landscape and can be used to understand the influence of topography on environmental processes such as erosion. They are mathematically calculated generally from DEMs but can also be derived from TINs.

Terrain attributes are classified as being either primary (slope, aspect, plan and profile curvature, specific catchment area [Eq. (1)], and upslope contributing area) or secondary (topographic wetness index [Eq. (2)], stream power index [Eq. (3)], length-slope factor [Eq. (4)], and channel initiation threshold [Eq. (5)]). Secondary terrain attributes are simply computed from two or more primary attributes and offer a way to describe landscape patterns as a function of process (Wilson and Gallant, 2000).
The topographic wetness indices are useful for predicting zones of saturation usually found along drainage paths and where water tends to concentrate on the terrain (Wilson and Gallant, 2000). The Revised Universal Soil Loss Equation has been heavily used in predicting erosion. One component of this equation is the Length-Slope factor which is used to measure the erosion potential for a specific slope. Based on the assumption that discharge is proportional to specific catchment area, stream power index is the measure of erosive power of flowing water (Wilson and Gallant, 2000). The final secondary terrain attribute is the channel initiation threshold. “This is a variation of the stream power index used to predict the location of headwaters of first-order streams” (Wilson and Gallant, 2000).

Several studies have used primary and secondary terrain attributes to either help predict or assess the potential for erosion to occur. One study used wetness index along with upslope contributing area to develop maps that would help plan the placement of riparian buffers and constructed wetlands within a watershed (Tomer et al., 2003). Another study used specific catchment area, slope, aspect, upslope contributing area, and profile and plan curvature as a way to predict hydrologically sensitive zones (Srivastava...
and Moore, 1989). Some have used terrain attributes to predict the possibility of ephemeral gullies forming (Thorne et al., 1986; Daba et al., 2003; and Kheir et al., 2007).

Flow direction algorithms can have a substantial impact on the calculation of the secondary terrain attributes. The Tapes program uses five different flow routing algorithms although only three have been examined in this thesis (i.e., D8, FD8, and DEMON stream tube). These flow routing algorithms are based on either the single or multiple flow direction grids. For the D8 approach, flow is routed in only one of eight directions (i.e., the one with the steepest descent). There are eight directions because each cell has only eight neighbors. The multiple flow direction approaches (i.e., FD8 and the Stream-Tube method) partitions the flow out to multiple neighboring cells. The FD8 approach distributes flow to multiple cells only until the cross-grading area threshold is exceeded (i.e., 100,000 m²) and then the single direction (D8) algorithm is used. The DEMON method routes flow down the stream tubes, which can expand and contract, until the edge of a DEM or a pit is encountered (Wilson and Gallant, 2000). One study compared six routing algorithms based on single and multiple flow direction algorithms to see how they impacted the prediction of ephemeral gullies (Desmet and Govers, 1996). They found that the multiple flow algorithms predicted wider ephemeral gullies whereas the single flow algorithm predicts them to be one cell thick. They also noted that single flow algorithms predict a higher starting point for gullies. This approach was more sensitive to errors in topography than multiple flow algorithms which were more robust and better represented the distribution of ephemeral gullies.
Grassed Waterways

A grassed waterway, by definition, is a natural or constructed waterway, typically broad and shallow, seeded to grass that is used to protect against erosion and to conduct surface water away from cropland (Soil Survey Staff, 1980). A grassed waterway is designed to reduce the speed of flow because of the retardant effect of the vegetation (Morgan, 2005). One study used terraces and grassed waterways in a potato field and found that soil losses were reduced from 20 t/ha/yr to 1 t/ha/yr (Chow et al., 1999).

There are several grasses recommended for use in a waterway management plan (i.e., Kentucky bluegrass, tall fescue, brome, reed canary, and Bermuda). These species are selected based on several criteria including: soil and climatic conditions, duration, quantity, and velocity of runoff, time required to develop a good cover and ease of establishment, availability of seed or plant materials, suitability for utilization as a seed or hay crop, and spreading of vegetation to adjoining areas (Schwab et al., 1996; and Troeh et al., 2004). Once a grassed waterway has been established, it is important to maintain it. Many times a waterway can be sprayed out from applying herbicides or it may be damaged by tilling through it in a non no-till management system. It is very important to repair any area that becomes damaged. Failure from erosion may also occur when a storm of much higher magnitude occurs than that for which the waterway was designed but, the most common cause of failure is inadequate maintenance (Morgan, 2005).

There are several benefits to implementing grassed waterways. Not only do they control erosion but they can also help reduce other environmental contaminants (i.e., fertilizers, pesticides, and herbicides). One study found that grassed waterways reduced runoff by 39%, sediment delivery by 82% and mineral nitrogen content by 84% (Fiener
and Auerswald, 2003). Another study found that grassed waterways reduced runoff volume by 47% and herbicide runoff by 56% compared to non-grassed waterways (Briggs et al., 1999).

According to the NRCS grassed waterways should be designed in such a way that they will convey the peak runoff expected from a storm of 10-year frequency (NRCS 2003). According to this guide they can be either parabolic or trapezoid in shape and shall not exceed 100 feet in width with a minimum width being whatever is necessary to carry the designed capacity.

Conservation Reserve Program (CRP)

Over the last several decades, much has been done to prevent soil erosion from occurring. Several environmental laws and practices have taken effect that helps govern how erosion is controlled. Erosion control provides many benefits to the general public and is likely to show economic efficiency gains (Reichelderfer and Boggess, 1988). The Farm Service Agency (FSA) sponsors the Conservation Reserve Program to reduce soil erosion, minimize the transport of sediment into streams, improve water quality, and create a wildlife sanctuary for many animal species. This program provides economic incentives to producers who implement conservation practices in areas that are susceptible to soil degradation. The FSA enlists the help of the Natural Resources Conservation Service (NRCS) to determine the eligibility, planning, and implementation of the CRP. By offering farmers incentives such as rental payments and cost-sharing to establish the vegetative cover, many producers are willing to adopt conservation practices.
There are two kinds of CRP: general and continuous. Which both are discussed in this section, the research in this thesis pertains only to continuous CRP.

**Continuous CRP**

Continuous CRP was developed to protect agricultural fields that are susceptible to soil erosion and to act as sediment filters around water sources. These areas can be protected for 10 to 15 years by installing grass waterways, filter strips, riparian buffers, field windbreaks, contour grass strips and shallow water areas for wildlife. As an incentive to implementing continuous CRP, landowners receive annual rental payments for land taken out of production and for repairs of the conservation practices. Landowners can receive up to 50 percent cost share for the installation and an additional 40 percent of eligible installation costs as an incentive payment or receive a one-time incentive payment of $100 per acre. As of April 2008 there were 1.64 million hectares enrolled in continuous CRP (USDA-Farm Service Agency, 2008).

To determine if a strip of land would be eligible for CRP payments for installing grassed waterways, an NRCS conservationist must make a site visit to determine if 1) acreage is suitable for the offered practice, 2) practice is needed and feasible, and 3) the practice to establish a vegetative cover on eligible cropland will enhance environmental benefits (FSA 2008). In Kentucky, grassed waterways are usually placed in areas of concentrated flow such as an ephemeral gully or old drainage ditch (Jack Kuhn, 2008, Personal Communications; Randall Rock, 2008, Personal Communications). However, in other states grassed waterways are often used in terrace systems to address sheet and rill erosion (Jack Kuhn, 2008, Personal Communications).
General CRP

The General CRP program provides farm owners and operators with an annual per-acre rental payment and half the costs of establishing a permanent land cover in exchange for retiring highly erodible land from production for 10-15 years (USDA-Farm Service Agency, 2008). As of April 2008 there were 12.4 million hectares enrolled in general CRP (USDA-Farm Service Agency, 2008). An environmental benefits indicator (EBI) ranks offers from land owners based on environmental indices and cost (USDA-Farm Service Agency, 2008). This indicator was adapted in 1990 by the Food Agriculture Conservation and Reform Act (FACTA).

Artificial Neural Network Models

Artificial neural networks are used to estimate, classify, and make predictions often in ways that may be better than traditional regression modeling (Matignon, 2005). By definition artificial neural networks are non-linear predictive models that learn through training and resemble biological neural networks in structure. Neural networks usually consist of one input layer representing all the variables applied to the model. These variables are then connected to a hidden layer consisting of user-definable number of neurons. Nonlinear transformations are applied to the input layer and hidden layer to generate predictions (Matignon, 2005). The exact number of neurons in the hidden layer can vary greatly. Each input variable is connected with different weights to each neuron within the hidden layer. Every hidden layer neuron is then connected to the output layer which produces the final model used in prediction. An example of the neural network structure is given in Figure 1.
In this thesis a multi-layer perception (MLP), feed-forward neural network architecture will be used. The MLP uses various linear combination functions and nonlinear sigmoidal activation functions to form a nonlinear regression model (Matignon, 2005). There are important assumptions that may need to be considered for the neural network models. The first is the independence assumption which requires the observations to be independent of each other. Matignon (2005) says that neural networks require errors to be uncorrelated, unrelated, or
Figure 1. Layout of an artificial neural network model comprising of an input layer with 8 variables, one hidden layer with 6 neurons and an output layer.
independent of each other. However, Lagazio and Russett (2004) state that neural network analysis does not require independent observations. Matignon (personal communications, 2007) supports the premise that independence issues may be ignored to some extent when using multilayer perception architecture and the Levenberg-Marquadt (default in SAS Enterprise Miner 4.3) convergence technique which was used in this thesis. The normality assumption is very subjective but requires error terms to be normally distributed (Matignon, 2005). The various standardization procedures were tested for this study in order to normalize the data. Further, neural networks have long been used successfully in situations where data are spatially dependent such as pattern recognition (Bishop, 1995).

Artificial neural networks have been used numerous times throughout the literature in some form to help predict landslide susceptibility. One study used artificial neural networks to predict the susceptibility of shallow landslides in Venezuela with 90% overall accuracy (Gomez and Kavzoglu, 2005). Their input variables include remote sensed imagery, documentary data, and terrain attributes (aspect, elevation, slope angle, and slope length). Another study used artificial neural networks in the Riomaggiore catchment, a sub watershed of the Reno River basin located in Northern Apennines (Italy), to forecast landslide susceptibility with 73% of the mapped landslides being correctly identified (Ermini et al., 2005). The input variables used were lithology, slope angle, profile curvature, land cover, and upslope contributing area. An additional study used neural networks along with logistic regression to predict landslide susceptibility in the Hendek region in Turkey (Yesilnacar and Topal, 2005). They used a total of 19 variables as input into the model with the most relevant to this thesis being plan and
profile curvature, slope length, stream power, and topographic wetness index. Neural networks did outperform the logistic regression models but both methods did very well with the neural network model producing 82% overall accuracy with the test dataset.

**Logistic Regression Models**

Logistic regression is used in many instances to predict the probability of an event occurring by fitting data to a logistic curve. An equation illustrating the logistic function is given below.

\[
f(x) = \frac{1}{1 + e^{-x}}
\]

Logistic regression is an easier way to predict and understand the occurrence of events than using neural networks. In some cases it may not yield as good of results (Yesilnacar and Topal, 2005) as an artificial neural network model. The same study also found that the forward stepwise logistic regression model had a tendency to remove several more variables that neural network models termed were important parameters. Another study used logistic regression to map landslide susceptibility in forested watersheds (Gorsevski et al., 2006). They used several input variables along with the compound topographic index (CTI) (which in this study is called the wetness index) and found very promising results. Logistic regression was also used to map the probability that substantial soil erosion had occurred in the past (Mueller et al., 2005). They found that by using various precision agriculture tools along with logistic regression that second order soil surveys could be improved which would give better erosion probability maps.
CHAPTER THREE: MATERIAL AND METHODS

The general layout of the analysis in this thesis is illustrated in Figure 2 below. Each of these steps will be discussed in more detail throughout this chapter.

**Analysis**

![Diagram of analysis process]

**Figure 2.** Layout of overall analysis represented in this thesis.

**Site Description**

This study was conducted on five fields located on a farm in the Outer Bluegrass region of Kentucky. Field 1 (38° 20.191' N, 85° 10.727' W), 2 (38° 17.901' N, 85° 11.006' W), 3 (38° 16.913' N, 85° 9.275' W), 4 (38° 20.280' N, 85° 14.090' W), and 5 (38° 20.193' N, 85° 11.978' W) are 57, 36, 23, 33, and 11 hectares in size, respectively. Aerial images of these fields are produced in Figures 3 through 7. This farm was chosen
because detailed grassed waterways had already been established. These fields had been in a no-till, corn (Zea mays L.), wheat (Triticum aestivum L.), and double-crop soybean (Glycine max [L.] Merr.) two year rotation for more than 15 years. Soils in this region developed primarily from limestone residuum overlain with pedisediment from limestone weathered materials and loess (Table 1). According to the soil survey manual for Shelby County, all soils but the Elk and Shelbyville soils are easily eroded when implementing a grassed waterway.

**Grassed Waterways**

Numerous grassed waterways had been established by the farmer in the five study fields at various times over the past 10 to 20 years with the interest of controlling and preventing channel erosion as described in the literature review above. The farm owner was informally trained by the NRCS conservationist in his county to identify areas within fields that required grassed waterways. In 2007 the landowner developed additional waterways in these fields.
Figure 3. Boundary of Field 1 (57 hectares) from 2007 overlain on a 2004 aerial image (Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from ftp://ftp.kymartian.ky.gov/fsa/).
Figure 4. Boundary of Field 2 (36 hectares) from 2007 overlain on a 2004 aerial image (Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from ftp://ftp.kymartian.ky.gov/fsa/).
Figure 5. Boundary of Field 3 (23 hectares) from 2007 overlain on a 2004 aerial image (Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from ftp://ftp.kymartian.ky.gov/fsa/).
Figure 6. Boundary of Field 4 (33 hectares) from 2007 overlain on a 2004 aerial image (Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from ftp://ftp.kymartian.ky.gov/fsa/).
Figure 7. Boundary of Field 5 (11 hectares) from 2007 overlain on a 2004 aerial image (Source: FSA NAIP Digital Ortho Photo Imagery obtained on-line from ftp://ftp.kymartian.ky.gov/fsa/).
Table 1. Soil Survey information for each study area.

<table>
<thead>
<tr>
<th>Location</th>
<th>Soil Symbol</th>
<th>Soil Name</th>
<th>Slope Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>LoB</td>
<td>Lowell sil 2 - 6 %</td>
<td>Fine, mixed, active, mesic Typic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LoC</td>
<td>Lowell sil 6 - 12 %</td>
<td>Fine, mixed, active, mesic Typic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NhB</td>
<td>Nicholson sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NhC</td>
<td>Nicholson sil 6 - 12 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Nolin sil 0 - 2 %</td>
<td>Fine-silty, mixed, active, mesic Dystric Fluventic Eutrudepts</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>ElB</td>
<td>Elk sil 2 - 6 %</td>
<td>Fine-Silty, Mixed, Active, Mesic Ultic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LoB</td>
<td>Lowell sil 2 - 6 %</td>
<td>Fine, mixed, active, mesic Typic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LoC</td>
<td>Lowell sil 6 - 12 %</td>
<td>Fine, mixed, active, mesic Typic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NhB</td>
<td>Nicholson sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Nolin sil 0 - 2 %</td>
<td>Fine-silty, mixed, active, mesic Dystric Fluventic Eutrudepts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ShB</td>
<td>Shelbyville sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Mollic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td>Site 3</td>
<td>LoC</td>
<td>Lowell sil 6 - 12 %</td>
<td>Fine, mixed, active, mesic Typic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NhB</td>
<td>Nicholson sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Nolin sil 0 - 2 %</td>
<td>Fine-silty, mixed, active, mesic Dystric Fluventic Eutrudepts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OtB</td>
<td>Otwell sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shb</td>
<td>Shelbyville sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Mollic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td>Site 4</td>
<td>ElB</td>
<td>Elk sil 2 - 6 %</td>
<td>Fine-Silty, Mixed, Active, Mesic Ultic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LoC</td>
<td>Lowell sil 6 - 12 %</td>
<td>Fine, mixed, active, mesic Typic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NhB</td>
<td>Nicholson sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Nolin sil 0 - 2 %</td>
<td>Fine-silty, mixed, active, mesic Dystric Fluventic Eutrudepts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OtB</td>
<td>Otwell sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shb</td>
<td>Shelbyville sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Mollic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td>Site 5</td>
<td>ElB</td>
<td>Elk sil 2 - 6 %</td>
<td>Fine-Silty, Mixed, Active, Mesic Ultic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LoC</td>
<td>Lowell sil 6 - 12 %</td>
<td>Fine, mixed, active, mesic Typic Hapludalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NhB</td>
<td>Nicholson sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Nolin sil 0 - 2 %</td>
<td>Fine-silty, mixed, active, mesic Dystric Fluventic Eutrudepts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OtB</td>
<td>Otwell sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Oxyaquic Fragiudalfs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shb</td>
<td>Shelbyville sil 2 - 6 %</td>
<td>Fine-silty, mixed, active, mesic Mollic Hapludalfs</td>
<td></td>
</tr>
</tbody>
</table>
In February of 2008, Randall Rock, the NRCS District Conservationist for this region examined approximately 40% of these grassed waterways by walking into each field with preliminary model results and the boundaries created by the land-owners. The district conservationist found that that every grassed waterway that was examined would have been eligible for continuous CRP payments. However, if precision spraying and planting equipment had not been utilized, the district conservationist would not have recommended that farmers enroll the small waterways (< 1/4 acre) in the program because they would have otherwise been logistically too difficult for most farmers to manage. Terraces are evident for Field 3 (Figure 5). These were put many years ago to help control erosion but were removed when no-till practices were adopted on this farm. The farmer takes great measures to keep waterways intact. Many times waterways will be sprayed out within a crop year leaving them more susceptible for erosion. After each crop year the farmer normally goes into the field and either repairs the grassed waterways by seeding it back into tall fescue or by drilling rye grass into the sprayed out areas (Mike Ellis, personal communications 2008).

Figures 3, 4, 5, 6, and 7 indicate the boundaries of the current grassed waterways as mapped in 2007. These boundaries were overlain on the aerial ortho-images that were taken in 2004. Notice that the 2007 grassed waterways do not match the waterways that were apparent in the aerial photograph. This was the case because more waterways were added and existing waterways expanded in 2007 because map-based section control for spray and planting equipment were adopted by the land-owner in this same year. This allowed these complex waterway designs to be managed more efficiently.
Elevation Data

To create the RTK GPS surveys, a Trimble (Trimble, Ltd., Sunnyvale, CA) AgGPS 214 receiver was used as a base station, and a Trimble 5800 receiver was used as a rover. Relative elevation measurements were logged each second along parallel passes with approximately 4 m between consecutive measurements and 12 m between passes. A 9.1-m DEM was obtained from the Kentucky Division of Geographic Information (KDGI; http://ogi.ky.gov/) website. Elevation measurements were removed from each field where RTK data was taken along with outside edges of the field to create a smoother transition. To help line the RTK GPS data up with the 9.1-m USGS DEMs, estimates of elevation from the 9.1-m USGS data sets were obtained with bilinear interpolation. The RTK measurements were then corrected using the average distance between the two datasets. The RTK GPS data was taken in 2000 for Field 1 (Mueller et al., 2003), 2004 for Fields 2 and 4 (Pike et al., 2006) and in 2007 for Fields 3 and 5.

DEM Creation

Two datasets were created for each field. The first included only the USGS data for the field and surrounding area. The second included the RTK data for the field and USGS data for areas outside the field border. These will be referred to as the RTK datasets (even though they include USGS data for areas outside the fields).

The TOPOTORASTER ArcGIS 3D analysts command (ESRI, Redlands, CA) was used to create each 4m USGS and RTK DEM as outlined by Sears et al. (2005) and Pike et al. (2006). Drainage enforcement was not used. There can be noise present around the borders of the fields due to the matching of the USGS and RTK data sets. This was removed by creating a 1m contour map that helped smooth the data along the
borders. The contour maps were then used to create a new 4m DEM with the TOPOTORASTER command using no drainage enforcement for each field.

**Terrain Modeling**

TAPESG for Windows 7.1 (University of Southern California, Los Angeles, CA) was used to calculate the different terrain attributes (i.e., upslope contributing area, length-slope, plan and profile curvature, stream power index, channel initiation threshold, wetness index, and delta) described earlier for each flow direction algorithm (D8, FD8, and Stream-Tube Demon). The output slope method used in this study was the finite difference algorithm which computes in the direction of the maximum slope of a curved surface fitted to the point and its 8 neighbors (Pike et al., 2006).

A point file representing each 4-m grid cell in each field was then created. Each terrain attribute along with the rasters representing the grassed waterways and croppable ground were then sampled to get the value at each grid cell location. This was used as input into the neural network and logistic regression models.

**Logistic Regression and Neural Network Modeling**

Fields 1 and 2 (Figures 3 and 4) were quantitatively combined in this study for use in SAS Enterprise Miner 4.3. The data were partitioned using a stratified method so that an unbiased representation of each field could be included in each of the training (50% of data) and validation (50% of data) datasets. The target variable was determined by the presence (1) or absence (0) of existing grassed waterways.

Multilayer perception neural network and logistic regression analyses were conducted with SAS Enterprise Miner 4.3. For neural network analysis, one hidden layer was used. Additionally the Levenberg-Marquardt convergence technique was used.
because it is the default in SAS Enterprise Miner 4.3 when less than 100 weights are used and it almost always converges to the correct weight estimates (Matignon, 2005).

For initial testing, Field 3 (Figure 5) was the primary test dataset but Fields 4 (Figure 6) and 5 (Figure 7) were also included in some cases as indicated in the results section. Fields 1 and 2 were used to train and validate the model but the test fields were not.

Preliminary sensitivity analysis involved evaluating the effects of different numbers of neurons (4, 8, 12, 16, 20, and 24), activation function (hyperbolic, logistic, gaussian, and elliot), standardization procedures (none, mid-range, range, and standard deviation), numbers of preliminary runs (0, 1, 3, 5, 10, and 40), and the number of hidden layers (1, 2, and 3). The predictor variables used in the sensitivity analysis of the neural network models included all of the following variables: upslope contributing area, length-slope, plan curvature, wetness index, stream power index, channel initiation threshold, delta, and profile curvature. The results of the sensitivity analyses were used to select analytical procedures for the remainder of the thesis results.

Next, different models were evaluated with different numbers and combinations of these variables to determine which variables should be used throughout the remainder of the study. Two of the terrain attributes were retained in the candidate model (i.e. plan curvature and length-slope factor). Plan curvature measures the topographic convergence and divergence of an area (Wilson and Gallant, 2000). It has been used numerous times in the literature to help predict landslide susceptibility and hydrologically sensitive zones respectively (Yesilnacar and Topal, 2005; Srivastava and Moore, 1989). Plan curvature also matched up extremely well visually with waterway boundaries.
Length-slope is one component of the Revised Universal Soil Loss Equation and has appeared in the literature as well for predicting the susceptibility of shallow landslides (Gomez and Kavzoglu, 2005). It too remained in this study even though it did not match perfectly with waterway boundaries.

The selection criteria were as follows:

1. Parameter estimates for logistic regression should be biophysically meaningful. Specifically, logistic regression parameters should be positive for upslope contributing area, length-slope, wetness index, stream power index, channel initiation threshold, and deltas, but negative for plan and profile curvature.

2. Standard terrain attributes calculated by most software programs were generally preferred over those that were rarely used in the literature. For example, delta is not universal (e.g., it is not calculated by ArcGIS). So unless an attribute substantially improves the model, it should not be used unless it is “universal”.

3. Models should have low misclassification statistics for logistic regression and neural network analyses.

4. Models should only contain as many variables as necessary for accurate predictions.

Based on the results of these analyses, the following variables were included in the analyses for the remainder of the thesis: upslope contributing area, length-slope, plan curvature, and wetness index. Once the optimal parameters were found, the next step taken was to analyze different terrain attribute input scenarios for both neural network and logistic regression.
Next, different flow direction algorithms (D8, FD8, and Stream-Tube Demon) were compared using the optimal neural network and logistic regression parameters along with only using 4 terrain attributes to compute each model. The final analysis of this study involved using 4 fields to train and validate the models and using 1 field to test. This set of comparisons were conducted to determine how well the analytical procedures performed on all of the fields and to determine how adding more fields would impact the overall prediction.

The percentage of points that were misclassified for training, validation, and test data sets were evaluated and reported. The misclassification statistic for the test dataset was used to select the best procedures. Maps comparing the model predictions to the boundaries of the grass waterways were also used to compare the various analyses.
CHAPTER FOUR: RESULTS AND DISCUSSION

In the first section of the Results and Discussion, terrain attributes will be compared with the boundaries of the grassed waterways. Then there will be sections on Neural Network Sensitivity Analyses, Model Variable Selection, Comparison of Neural Networks and Logistic Regression, Selection of Flow Algorithm, Testing All Five Fields, Comparison of USGS and RTK data, Modeling Limitations, and Impact of Waterway Reshaping.

Comparison of Terrain Attributes with Grassed Waterway Boundaries

There were many similarities between terrain attributes created with the RTK and 9.1-m USGS DEMs although the maps were not identical (e.g., Field 3 in Figures 8 and 9). Pike et al. (2006) also made observations for two of the same fields considered in this thesis (i.e., Fields 2 and 4). The authors of that study indicated that many of the erosion features apparent in maps derived from the RTK data were also evident in those derived from the USGS data. However, specific erosion features that were visible in the terrain attribute data differed in intensity and in position in the landscape depending on the data sources. This was particularly true for compound topographic attributes that were based on flow accumulation (e.g., wetness index, stream power index, length-slope factor, and the channel initiation threshold) rather than the simple terrain attributes (e.g., plan and profile curvature, upslope contributing area, and delta). The observations the authors made for Fields 2 and 4 were also valid for Fields 1, 3, and 5 in this thesis.
Figure 8. Terrain attributes for the test dataset (Field 3) calculated from the RTK GPS and USGS DEMs. (Upslope Contributing Area, m²; Plan Curvature, radians/100 m).

Figure 9. Terrain attributes for the test dataset (Field 3) calculated from the RTK GPS and USGS DEMs. (Profile Curvature, radians/100 m).
Many of the terrain attributes calculated with the RTK GPS and USGS data corresponded well with the waterway boundaries (e.g., upslope contributing area, plan curvature, wetness index, stream power index, channel initiation threshold, and delta in Field 3 as shown in Figures 8 and 9). The length-slope data matched for some but not all waterways for most fields (e.g., Field 3 in Figure 8). This is consistent with a study that used the wetness index along with the length-slope factor to optimize the placement of riparian practices in a watershed (Tomer et al., 2003). They found that the length-slope factor highlighted many of the outside edges of stream meanders. Profile curvature did not match well in Fields 2, 3, 4, and 5 but did for Field 1 (data only shown for field 3 in Figure 9). Slope did not match well with the grassed waterways in any of the fields but slope was included in the calculations of several of the terrain attributes (length-slope, wetness index, stream-power index, and channel initiation threshold) according to Equations 2, 3, 4, and 5. One study used slope along with several other terrain variables within a regression tree model to map out gully erosion susceptibility, they found that most of the gullies occurred on steep slopes. The secondary topographic variables that incorporated slope and upslope contributing area in that study proved to be the most important variables (Kheir et al., 2007).

There were differences in the correspondence between grassed waterways and terrain attributes depending on whether RTK-GPS or USGS-DEM data was used to calculate the terrain attributes. The length-slope factor matched better for the RTK data than the USGS data. The example circled in Figure 8 demonstrates a more significant discrepancy where the RTK derived stream power index matched better with the grassed waterway boundaries. From a practical perspective, if these terrain attributes were used
as a guide for a NRCS personal to identify areas that may potentially qualify as waterways, a site visit would be required. Therefore, discrepancy in the exact location of erosion features as indicated by a terrain attribute would not be very important because the area for the planned waterway would be flagged by hand by the NRCS employee. What is more important is whether erosion features can be identified from the terrain information. It is also not very important that in some cases terrain attributes may incorrectly indicate that erosion features occur in an area when in fact they do not because it is better to over predict areas in the field verses under predicting, assuming there would be a site visit to examine and confirm features. The most serious case would be if an existing erosion procedure were not identified with a prediction model. In that case, an erosion feature could easily be missed during a site visit. However, the production of these kinds of maps help identify even 50% of erosion features that exist in agriculture fields, there would be a substantial reduction in erosion in Kentucky.

The visual analyses indicate that many secondary terrain attributes could be used to help conservationists better identify grassed waterways. After giving terrain attribute maps to Randall Rock, the NRCS conservation planner for the Shelby County area, he indicated that the maps were very busy and sometimes hard to understand but he saw that they had great value. It is be possible to train planners to better utilize these maps. However, it also may be possible to develop maps that integrate much of the data together with artificial neural networks or logistic regression. The advantages with models is that they may be sensitive to patterns that may otherwise be difficult for the human eye to discern and could be used to create a simple binary prediction map that are
easier for planners to interpret. The disadvantage is that they would need to be calibrated.

**Neural Network Sensitivity Analyses**

There are many options that can be used to conduct neural network analysis including (e.g., number of neurons used, activation function, standardization, number of preliminary runs, and the number of hidden layers). Generally, neural network options are selected based on preliminary experimentation and specifically observing how model errors and percent misclassification statistics are affected for validation and test data sets (C. Viswanathan, personal communications, 2008; Matignon, 2005). Some of the default options in SAS Enterprise Miner 4.3 depend on the data set size along with how many variables and other options within the neural network architecture are used (e.g., activation function).

**Table 2. Misclassification results for neural network (NN) models created using RTK data with different numbers of neurons and with the following variables held constant: Hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model).**

<table>
<thead>
<tr>
<th>Number of Neurons</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5.1</td>
<td>5.0</td>
<td>5.5</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>4.9</td>
<td>5.6</td>
</tr>
<tr>
<td>12</td>
<td>4.8</td>
<td>4.9</td>
<td>5.3</td>
</tr>
<tr>
<td>16</td>
<td>4.8</td>
<td>4.9</td>
<td>5.4</td>
</tr>
<tr>
<td>20</td>
<td>4.8</td>
<td>4.8</td>
<td>5.2</td>
</tr>
<tr>
<td>24</td>
<td>4.8</td>
<td>4.9</td>
<td>5.2</td>
</tr>
</tbody>
</table>
**Number of Neurons**

Increasing the number of neurons from 4 to 24 neurons generally reduced the misclassification rates for the training and validation data sets but not always for the test data set (Table 2). Since these models were developed from the data in Fields 1 and 2, the misclassification rates of greatest interest were those reported for the independent test field (i.e., Field 3) which was not used to develop the model. Adding more neurons slightly but not consistently reduced the test dataset misclassification statistic.

One rule of thumb is that the number of neurons should be no greater than three times the number of variables (C. Viswanathan, Personal Communications, 2008). Since 8 neurons were used, the maximum that would be recommended for this analysis (Table 2) would be 24. Increasing the number of neurons from 20 to 24 did not reduce the classification error for the test data set (Table 2) so this rule appears to be appropriate for this dataset. Matignon (2005) recommended testing with different numbers of neurons and suggested that it’s better to have too many rather than then too few neurons. The error rate appeared to be greatest for 4 and 8 neurons so this rule of thumb appears to also be appropriate for this analysis.
Figure 10. Comparison maps of grass waterways created with different number of neurons for the Field 3 test data set. The FD8 algorithm was used to calculate flow direction and the data were normalized with the standard deviation. For the neural network analyses, 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 2.

The maps comparing the model predictions for different number of neurons were not substantially different for the test dataset (Figure 10). The brown areas in this field indicate where the model predicted the croppable ground whereas the green areas represent the predicted grassed waterways. Brown areas within the boundary of the grassed waterway represent false negatives and the green areas outside the boundary represent the false positives.

Some false positives occurred because the model failed to accurately predict. In other cases, the false positives may have occurred where it would have been appropriate to have grassed waterways. A waterway prediction map similar to the one in Figure 10 was examined by Randall Rock, the NRCS planner. He determined that a waterway would have been justified in the circled waterway feature indicated on the west side of
the field but not the circled area in the center of the field. From a practical perspective, false positives would not be of great concern if an NRCS conservation planner were to field validate models before recommending that a grassed waterway be installed.

The number of neurons had a small effect on the maps (Figure 10). However, the waterways apparent in one map were still visible in the others. Clearly, differences in misclassification statistics do not necessarily indicate that one statistical modeling procedure is necessarily better than another for management.

This analysis suggested that relatively few neurons would be needed to generate high quality maps for this field. Adding more neurons to the model could substantially impact the computational time and strain on the system in some situations. However, this analysis used only used 8 input variables. However, in cases where hundreds of variables are used, fewer neurons might be appropriate if computational resources were a constraint. Based on this sensitivity analysis, subsequent analyses presented in this thesis use 20 neurons because of the lower misclassification rate observed and since 20 neurons did not require great computational resources.

**Activation Function**

The hyperbolic function is the default activation function in SAS Enterprise Miner when the linear-general combination function is used. This sigmoidal function is recommended for large sample sizes because it has a faster rate of convergence as compared to other methods including the logistic, gaussian, and elliot activation functions (Matignon, 2005).
Table 3. Misclassification results for neural network (NN) models created using RTK data with different activation functions and with the following variables held constant: 20 neurons, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model).

<table>
<thead>
<tr>
<th>Activation function</th>
<th>Misclassification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>Hyperbolic</td>
<td>4.8</td>
</tr>
<tr>
<td>Logistic</td>
<td>4.9</td>
</tr>
<tr>
<td>Gaussian</td>
<td>4.8</td>
</tr>
<tr>
<td>Elliot</td>
<td>4.9</td>
</tr>
</tbody>
</table>

There were very small differences in misclassification rates for the different activation functions (Table 3). The differences between the model prediction maps for the various activation functions were virtually imperceptible (data not shown) but were comparable to the differences shown in Figure 10. The data suggested that the choice of activation function, in some cases, has little impact on neural network results. Therefore, the default hyperbolic activation function was used from this point forward in this thesis.

Standardization Procedure

Standardization is used to assure convergence (SAS, 2005). There are several standardization methods available in SAS Enterprise Miner (none, mid-range, range, and standard deviation). The Mid-range option subtracts the midrange and divides by half the range giving values that have a minimum of minus one and a maximum of plus one. The Range option subtracts the minimum and divides by the range with values having a minimum of zero and a maximum of plus one. The standard deviation option subtracts the mean and divides by the standard deviation resulting in values with a mean of zero and a standard deviation of one.
The default method is the standard deviation. However, SAS (2005) and Matignon (2005) recommend that none should be used if the variables are Gaussian. However, in this study, many of the variables deviated substantially from normal.

Table 4. Misclassification results for neural network (NN) models created using RTK data with different normalization procedures and with the following variables held constant: 20 neurons, hyperbolic activation function, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model).

<table>
<thead>
<tr>
<th>Standardization</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>5.3</td>
<td>5.2</td>
<td>6.1</td>
</tr>
<tr>
<td>Mid Range</td>
<td>5.0</td>
<td>5.0</td>
<td>5.2</td>
</tr>
<tr>
<td>Range</td>
<td>5.0</td>
<td>5.0</td>
<td>5.1</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4.8</td>
<td>4.8</td>
<td>5.2</td>
</tr>
</tbody>
</table>

The data that was not standardized produced the highest test data set misclassification statistics and the range procedure resulted in the lowest rates (Table 4). Figure 11 demonstrates that despite the quantitative differences, there were only very small visual differences between the two procedures. Since there were not great differences between the standardization procedures (i.e., Mid-Range, Range, and Standard Deviation) the default standard deviation option was used throughout the rest of this study.
Figure 11. Comparison maps of grass waterways created with different normalization methods used for the Field 3 test data set. The FD8 algorithm was used to calculate flow direction. For the neural network analyses 20 neurons, hyperbolic activation function, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 4.

Preliminary Runs

Preliminary runs are used to avoid local error minimums that can be sometimes found instead of the sought after global minimums while training the neural network model (Matignon, 2005). These preliminary runs are used to determine the more appropriate initial weight estimates. Given that preliminary runs can substantially increase the amount of time required for analyses, no more runs than necessary should be used.
Table 5. Misclassification results for neural network (NN) models created using RTK data with different number of preliminary runs and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model).

<table>
<thead>
<tr>
<th>Number of Preliminary Runs</th>
<th>Misclassification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>0</td>
<td>4.8</td>
</tr>
<tr>
<td>1</td>
<td>4.8</td>
</tr>
<tr>
<td>3</td>
<td>4.8</td>
</tr>
<tr>
<td>5</td>
<td>4.8</td>
</tr>
<tr>
<td>10</td>
<td>4.8</td>
</tr>
<tr>
<td>40</td>
<td>4.8</td>
</tr>
</tbody>
</table>

There were only very small differences in misclassification rates between the analyses using different numbers of preliminary runs (Table 5). Further, there were also only slight differences in prediction maps (data not shown). Therefore, only one or none preliminary runs may have been necessary here. However, because only one of the test data sets used in this study was analyzed here, it is possible that multiple preliminary runs may be helpful for some of the other fields. Therefore, 5 runs, the default setting for SAS Enterprise Miner, were used for all subsequent analyses presented in this thesis.

**Number of Hidden Layers**

Multiple hidden layers are rarely used because of the complexity introduced into the models often results in over-fitting; however, multiple layers can also potentially increase prediction accuracy in some cases (Matignon, 2005). In this study, 3 hidden layers produced slightly better results than 2 layers and very similar results with 1 layer (Table 6). There were, however, only slight differences in maps (data not shown).
Because of the substantial increase in computational time required and the only slight change in the misclassification statistic and maps, the default setting (i.e., 1 hidden layer) will be used for the rest of this study.

**Table 6. Misclassification results for neural network (NN) models created using RTK data with different number of hidden layers and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model).**

<table>
<thead>
<tr>
<th>Number of Hidden Layers</th>
<th>Misclassification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>1</td>
<td>4.8</td>
</tr>
<tr>
<td>2</td>
<td>4.9</td>
</tr>
<tr>
<td>3</td>
<td>4.7</td>
</tr>
</tbody>
</table>

**Summary of Sensitivity Analysis Results**

The sensitivity analysis indicated that the artificial neural networks procedure is robust to various changes in methodologies. The choice of standardization procedure had the largest impact on misclassification rate but prediction maps were very similar. Generally, varying the different parameters within the neural network architecture changed the misclassification rates but differences were generally miniscule and difficult to discern visually.

The sensitivity analysis allowed the selection of certain parameters to be used throughout the remainder of this study. These included the following: 20 neurons, 1 hidden layer along, and 5 preliminary runs. The hyperbolic activation function and the standard deviation standardization procedure were also used for the rest of the analyses presented in this thesis.
Model Variable Selection

Of the 10 models considered to be used throughout the remainder of the thesis (Table 7), numbers 1, 4, 6, 7, 8, 9, and 10 were rejected based on Criteria 1 described in the Materials and Methods section (see the Logistic Regression and Neural Network Modeling subsection). Model 3 produced small misclassification rates but were rejected because delta is uncommon (Criteria 2) and because delta contains both very high and very low values within waterways (e.g., Figure 9). Model 2 was selected over models 5 based on Criteria 3 (i.e., it resulted in lower misclassification rates). Therefore, these terrain attributes were used in the analyses that follow: upslope contributing area, length-slope, plan curvature, and the topographic wetness index.

From a practical point of view, the choice model and number of parameters had little impact on maps of prediction. For example, Model 7 (4-variable model), 8 (6 variable model), and 9 (7 variable model) produced similar maps (Figure 12) even though Model 7 had the smallest misclassification statistic.

Model 2 (Table 7) is intuitively meaningful. Upslope contributing area and length-slope are erosion index values and grassed waterways are used to prevent erosion in sensitive areas. They are therefore more positive as erosion index values increase. Waterways are also placed in areas where water accumulates and wetness index values become more positive in wetter areas. Therefore, it is not surprising that the parameter for upslope contributing area, length slope, and the wetness index are positive. Plan curvature is negative in areas that have convergent water flow and positive in areas where water diverges. Therefore, it is also not surprising that the coefficient is negative for plan curvature.
Table 7. Misclassification results for neural network (NN) models created using RTK data with different models consisting of various combinations of terrain attributes and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm (Fields 1 and 2 were used to validate and train the model and Field 3 was used to test the model).

<table>
<thead>
<tr>
<th>Model</th>
<th>Terrain Attributes Used</th>
<th>------Regression Analyses------</th>
<th>Neural Network Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T value</td>
<td>Parameter Estimate</td>
</tr>
<tr>
<td>1</td>
<td>Upslope Contributing Area</td>
<td>11.76</td>
<td>0.00205</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>7.06</td>
<td>0.664</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-27.14</td>
<td>-6.89</td>
</tr>
<tr>
<td></td>
<td>Stream Power Index</td>
<td>-2.74</td>
<td>-0.0554</td>
</tr>
<tr>
<td>2</td>
<td>Upslope Contributing Area</td>
<td>4.69</td>
<td>0.000411</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>15.19</td>
<td>0.751</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-19.47</td>
<td>-5.31</td>
</tr>
<tr>
<td></td>
<td>Wetness Index</td>
<td>14.73</td>
<td>0.848</td>
</tr>
<tr>
<td>3</td>
<td>Upslope Contributing Area</td>
<td>20.61</td>
<td>0.00138</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>10.68</td>
<td>0.507</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-25.36</td>
<td>-6.53</td>
</tr>
<tr>
<td></td>
<td>Deltas</td>
<td>7.61</td>
<td>0.0262</td>
</tr>
<tr>
<td>4</td>
<td>Upslope Contributing Area</td>
<td>22.60</td>
<td>0.00186</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>9.08</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-25.96</td>
<td>-6.66</td>
</tr>
<tr>
<td></td>
<td>Channel Initiation Threshold</td>
<td>-5.32</td>
<td>-0.988</td>
</tr>
<tr>
<td>5</td>
<td>Upslope Contributing Area</td>
<td>26.45</td>
<td>0.00192</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>10.25</td>
<td>0.533</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-13.45</td>
<td>-3.76</td>
</tr>
<tr>
<td></td>
<td>Profile Curvature</td>
<td>-37.44</td>
<td>-8.88</td>
</tr>
<tr>
<td>6</td>
<td>Channel Initiation Threshold</td>
<td>-5.09</td>
<td>-0.261</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>19.42</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-18.92</td>
<td>-5.04</td>
</tr>
<tr>
<td></td>
<td>Wetness Index</td>
<td>34.35</td>
<td>1.12</td>
</tr>
<tr>
<td>Model</td>
<td>Terrain Attributes Used</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T value</td>
<td>Parameter Estimate</td>
</tr>
<tr>
<td>7</td>
<td>Channel Initiation Threshold</td>
<td>14.70</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>-3.15</td>
<td>-0.274</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-32.98</td>
<td>-8.07</td>
</tr>
<tr>
<td></td>
<td>Deltas</td>
<td>13.65</td>
<td>0.0381</td>
</tr>
<tr>
<td>8</td>
<td>Upslope Contributing Area</td>
<td>5.47</td>
<td>5.47</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>13.24</td>
<td>13.2</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-17.62</td>
<td>-17.6</td>
</tr>
<tr>
<td></td>
<td>Wetness Index</td>
<td>14.82</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>Deltas</td>
<td>5.90</td>
<td>5.90</td>
</tr>
<tr>
<td></td>
<td>Channel Initiation Threshold</td>
<td>-5.76</td>
<td>-0.961</td>
</tr>
<tr>
<td>9</td>
<td>Upslope Contributing Area</td>
<td>-1.81</td>
<td>-0.000141</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>12.63</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-15.86</td>
<td>-4.45</td>
</tr>
<tr>
<td></td>
<td>Stream Power Index</td>
<td>6.12</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>Deltas</td>
<td>5.58</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Channel Initiation Threshold</td>
<td>-7.53</td>
<td>-2.17</td>
</tr>
<tr>
<td></td>
<td>Wetness Index</td>
<td>16.57</td>
<td>0.877</td>
</tr>
<tr>
<td>10</td>
<td>Upslope Contributing Area</td>
<td>-7.01</td>
<td>-0.000289</td>
</tr>
<tr>
<td></td>
<td>Length-Slope</td>
<td>9.17</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>Plan curvature</td>
<td>-4.41</td>
<td>-1.38</td>
</tr>
<tr>
<td></td>
<td>Stream Power Index</td>
<td>10.04</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>Deltas</td>
<td>2.62</td>
<td>0.00860</td>
</tr>
<tr>
<td></td>
<td>Channel Initiation Threshold</td>
<td>-8.01</td>
<td>-2.29</td>
</tr>
<tr>
<td></td>
<td>Wetness Index</td>
<td>14.47</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>Profile Curvature</td>
<td>-36.70</td>
<td>-8.95</td>
</tr>
</tbody>
</table>

Table 7. Continued.
Figure 12. Comparison maps of grass waterways created with different variable scenarios for the Field 3 test data set. The FD8 algorithm was used to calculate flow direction. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 7.

Models 3 and 5 (Table 7) were also consistent with what is known about the way water behaves in landscapes. Specifically, delta is another erosion index so the parameter estimate is positive in Model 3. Plan curvature is the second derivative of elevation down a slope and values are negative for rapidly falling slopes. The coefficient is therefore negative in Model 5.

For these three models, the absolute values of the t-statistics were the greatest for plan curvature (Models 2 and 3) and profile curvature (Models 5) indicating that these
variables had the largest impact on the estimates. However, the combined impact of the complex terrain attributes (i.e., upslope contributing area, length-slope, wetness index, and delta) was substantially greater than the impact of the simple terrain attributes for Model 2 and 3 (the sum of the absolute values of the t-statistics). However the combined impact for simple terrain attributes was somewhat greater for Model 5.

**Comparison of Neural Network Analysis and Logistic Regression**

Differences between neural network and logistic regression models were not large. Logistic regression did produce lower misclassification statistics for the test datasets (Table 7). However, the prediction maps were very similar (e.g., compare the top row to the bottom row in Figure 12 for the different models). Despite the quantitative differences between the neural network and logistic regression, most of the same features are apparent for both methods. Clearly, the analytical method (i.e., neural networks or logistic regression) did not have a big impact for this field. However for a larger model over a larger geographic area, it is possible that neural networks would outperform logistic regression in some cases.

Results in other studies have been different. For example, Yesilnacar and Topal (2005) found that a neural network model did a better job of predicting landslides than logistic regression. Other studies have also found that neural network models perform better than logistic regression (Mahiny and Turner, 2003), although some have found no differences between these methods (Schumacher et al., 1996; Manel et al., 1999).

These other studies did not use independent datasets but rather set aside a random number of data points to test their models (i.e., they were from the same population as
their validation and training data sets). Sometimes complex modeling procedures can over fit the data used to generate the model. Therefore, while it may be that neural networks really did a better job than logistic regression, it is also possibly they did not have a good independent test to make this assessment. Had not an independent dataset been used in this study, one conclusion for this thesis would have been that neural networks out-performed logistic regression. Specifically, the misclassification statistics for the training and validation dataset indicated that the neural network model outperformed logistic regression which is contrary to what is indicated by the test dataset (Table 8).

**Table 8. Misclassification results for neural network (NN) and logistic regression (REG) models created using RTK data with different flow direction algorithms used and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer (Fields 1 and 2 were used to validate and train the model and Field 3, 4, and 5 was used to test the model).**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Statistical Procedure</th>
<th>Flow Direction Algorithm</th>
<th>Misclassification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>RTK</td>
<td>NN</td>
<td>D8</td>
<td>7.1</td>
</tr>
<tr>
<td>RTK</td>
<td>NN</td>
<td>FD8</td>
<td>5.9</td>
</tr>
<tr>
<td>RTK</td>
<td>NN</td>
<td>Stream Tube</td>
<td>6.4</td>
</tr>
<tr>
<td>RTK</td>
<td>REG</td>
<td>D8</td>
<td>7.2</td>
</tr>
<tr>
<td>RTK</td>
<td>REG</td>
<td>FD8</td>
<td>6.1</td>
</tr>
<tr>
<td>RTK</td>
<td>REG</td>
<td>Stream Tube</td>
<td>6.8</td>
</tr>
<tr>
<td>USGS</td>
<td>NN</td>
<td>D8</td>
<td>8.6</td>
</tr>
<tr>
<td>USGS</td>
<td>NN</td>
<td>FD8</td>
<td>8.6</td>
</tr>
<tr>
<td>USGS</td>
<td>NN</td>
<td>Stream Tube</td>
<td>8.2</td>
</tr>
<tr>
<td>USGS</td>
<td>REG</td>
<td>D8</td>
<td>8.9</td>
</tr>
<tr>
<td>USGS</td>
<td>REG</td>
<td>FD8</td>
<td>9.0</td>
</tr>
<tr>
<td>USGS</td>
<td>REG</td>
<td>Stream Tube</td>
<td>8.6</td>
</tr>
</tbody>
</table>
Logistic regression makes assumptions about multicollinearity, normality, and independence but neural networks do not make assumptions (C. Srinivasan, personal communications, 2008). Although not specifically tested, the logistic regression model errors were, without doubt, not spatially independent. Further, multicollinearity was likely problematic for the logistic regression Models number 1, 9, and 10 (Table 7) because the stream power index had a correlation value of 0.89 with upslope contributing area and a correlation value of 0.95 with channel initiation threshold. Multicollinarity is problematic for regression whenever correlation values are $\geq 0.9$ (P. Cornelius, personal communications, 2006). Because the results for neural networks and logistic regression were so similar in terms of misclassification statistics (Table 7) and the maps were accurate predictors of grass waterways (Figure 12), it is clear that violation of assumptions were not severe enough to have a substantial impact on logistic regression.

**Selection of Flow Algorithm**

The D8 procedure often produces narrow estimates of waterway widths because simulated water flow is channeled to only one neighboring cell. The effect of flow can best be observed in the terrain attribute maps for upslope contributing area (Figure 13) and is also apparent in the actual model output (Figure 14). The FD8 algorithm distributes flow to multiple cells until the cross-grading area threshold is exceeded and from then on, single direction (i.e., D8) flow is used. Therefore the predicted widths of the waterways are greater for the FD8 method in Figure 14. The DEMON stream tube method is an even more sophisticated approach for calculating flow that allows flow channels to expand and contract downhill but this was not very apparent in Figure 13.
The D8 approach can result in data defects (Wilson and Gallant, 2000) because single flow algorithms are (e.g., D8) also sensitive to small errors (Desmet and Govers, 1996). Examples of defects can be seen by the circled area in Figure 13. Here the drainage of the left-most center waterway has its own drainage out of the field whereas in field observations this area would most likely drain into the long center waterway and then drain out of the field as shown by the FD8.

Figure 13. Upslope contributing area maps for the RTK and USGS data created with different flow direction algorithms (i.e., D8, FD8, and Stream-Tube Demon).
method. The FD8 misclassification statistics were generally smaller for the Field 4 and 5 test data sets but not for Field 3 regardless of whether USGS or RTK data was used (Table 8). Therefore, the FD8 method was chosen to be used for all subsequent analyses presented in this thesis. However, for all flow direction three techniques and both the RTK and USGS data, the same waterways were identified in all maps (Figures 14 and 15).

Figure 14. Comparison maps of grass waterways created with different flow direction algorithms for the Field 3 test data set (RTK Data). For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 8 (for test dataset 3).
Figure 15. Comparison maps of grass waterways created with different flow direction algorithms for the Field 3 test data set (9.1-m USGS DEMs). For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, and 1 hidden layer were used. These maps correspond to the comparisons made in Table 8 (for test data set 3).

Testing All Five Fields

When models were created for all five fields (Table 9), errors were greatest for Fields 1 and 4 and smallest for Field 5. The corresponding maps are presented in Figures 16 through 20. Logistic regression produced similar results to neural networks for these
analyses as described earlier (in the subsection entitled “Comparison of Neural Network Analysis and Logistic Regression”).

Table 9. Misclassification results for neural network (NN) and logistic regression (REG) models created using RTK and USGS data for the different number of fields used to train and validate the model and with the following variables held constant: 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm.

<table>
<thead>
<tr>
<th>Test Data Set</th>
<th>Training and Validation Data Sets</th>
<th>Misclassification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RTK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>USGS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NN</td>
</tr>
<tr>
<td>1</td>
<td>2,3,4, and 5</td>
<td>7.5</td>
</tr>
<tr>
<td>2</td>
<td>1,3,4, and 5</td>
<td>4.8</td>
</tr>
<tr>
<td>3</td>
<td>1,2,4, and 5</td>
<td>4.3</td>
</tr>
<tr>
<td>4</td>
<td>1,2,3, and 5</td>
<td>6.6</td>
</tr>
<tr>
<td>5</td>
<td>1,2,3, and 4</td>
<td>2.3</td>
</tr>
</tbody>
</table>
Figure 16. Comparison maps of grass waterways created with the RTK and USGS datasets for Field 1. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm were used. These maps correspond to the comparisons made in Table 9.
Figure 17. Comparison maps of grass waterways created with the RTK and USGS datasets for Field 2. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm were used. These maps correspond to the comparisons made in Table 9.
Figure 18. Comparison maps of grass waterways created with the RTK and USGS datasets for Field 3. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm were used. These maps correspond to the comparisons made in Table 9.
Figure 19. Comparison maps of grass waterways created with the RTK and USGS datasets for Field 4. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm were used. These maps correspond to the comparisons made in Table 10.
Figure 20. Comparison maps of grass waterways created with the RTK and USGS datasets for Field 5. For the neural network analyses 20 neurons, hyperbolic activation function, normalization by the standard deviation, 5 preliminary runs, 1 hidden layer, and the FD8 flow direction algorithm were used. These maps correspond to the comparisons made in Table 9.

The impact of data support on model performance was considered for the test 3 dataset only. Specifically, the models developed with data from only two fields (Fields 1 and 2) had misclassification statistic of 4.9, 4.4, 5.1, and 4.0 % for the Field 3 test dataset (Table 8). In comparison, the model developed from four fields (i.e., Fields 1, 2, 4, and
5) had misclassification statistics of 4.3, 3.6, 4.4, and 3.6% also for the Field 3 test data set (Table 9). Clearly, adding additional data (i.e., from Fields 4 and 5) resulted in a reduction of the misclassification statistics. However, there were only small differences in the prediction maps. Specifically, observe that the FD8 neural network results in Figure 14 were very similar to the RTK neural network and logistic regression results presented in Figure 18. These results demonstrate that this modeling approach is robust. Specifically, only two fields were necessary to train and validate the model; however, they were two of the biggest fields in this study.

**Comparison of USGS and RTK data**

Many times highly precise elevation data is not an option because owning or renting an RTK GPS system makes is cost prohibitive. One alternative is to use freely available USGS DEMs that are available for the entire country. This study compared the use of the RTK data with that of the 9.1-m USGS DEMs. The misclassification statistics were similar but generally better for the RTK for the test datasets (Table 9). This was expected because the RTK data is generally considered to be much more precise.

When the prediction models were mapped, the RTK data generally better indicated the occurrence of waterways. This was particularly true for Field 1 and 2 (Figure 16 and 17). However, this was not always true. For example predictions for Field 3 (Figure 18) maps were very similar for both RTK and USGS data. The USGS does not always capture the extent of each waterway in the test dataset (e.g., Figures 16 and 20). For all fields there were waterways predicted with the RTK data that could not be with the USGS data.
Modeling Limitations

For Field 4 (Figure 19), many of the grass waterways were not identified by even the RTK datasets. This occurred because these channels did not convey surface runoff from large upslope areas but rather from springs discharging perched groundwater in upland areas (Randall Rock, personal communications, 2008). These kinds of waterways are specific to the geology of this area and can be difficult to predict with most occurring only during the wet months. Often water is not discharge at the bottom of the fields but it disappears at the toe slope (e.g., the waterways in the circled area in Field 4 shown in Figure 19).

The Field 4 waterways that were not predicted by the model (Figure 19) demonstrate that this modeling approach cannot predict all grassed waterways. However, since the majority of the waterways in the five fields studied can be identified, this modeling approach was still considered to be of value to the NRCS conservationist for this area, Randall Rock. However, with the price of RTK measurements, this may not be economic if there are fields where many eroded areas can not be identified.

If highly precise elevation data were free (e.g., LIDAR), this approach might be more feasible for these fields. However, more research will be required to determine whether LIDAR provides sufficiently accurate estimates of elevation to be adequate for predicting where grassed waterways should be located.

Impact of Waterway Reshaping

The reshaping of the landscape that occurs when installing CRP waterways likely had some impact on DEMs. This would have also affected the calculation of terrain attributes and impacted model results. However, the extent to which this occurred is not
known. However, it would only have affected those that were enrolled in continuous CRP.

Reshaping of waterways is been described by Morgan (2005) and summarized here. Reshaping is done several ways and the final shapes may be triangular, trapezoidal, or parabolic. Triangular sections are not recommended because of the risk of scour at the lowest point and most trapezoidal waterways become parabolic overtime. The NRCS design waterways by either a parabolic or trapezoidal method.

Many of the waterways in these fields were not enrolled in CRP. Fields 3 and 5 never had CRP waterways. Maps of the CRP waterways for the fields in this study are being requested from the NRCS and FSA. For those waterways that were not CRP, the farmers simply used a blade to smooth over the ephemeral gullies before planted grasses. This would not have had a substantial impact on the shapes of the waterways.

That waterway reshaping that occurred for the fields in this study did not have a substantial impact on the model results. For example, many of the new waterways installed in 2007 but shown as cropped in the 2004 aerial images (Figures 3, 4, 5, 6, and 7) were not enrolled in CRP and no major reshaping was performed. Nevertheless, the erosion models predicted well in those fields (Figures 16, 17, 18, 19, and 20). In some cases, waterways were established after the RTK DEMs were created. For example, the RTK DEM for Field 1 was created in 2000 prior to the addition of many new waterways added in 2007. Yet the RTK models predicted very well in the new waterway areas (e.g., Figure 16). The models also performed very well in Field 3 which never had CRP waterways (Figure 18). This analysis demonstrates that the modeling approaches presented in this thesis work very well.
CHAPTER FIVE: SUMMARY AND CONCLUSIONS

The first hypothesis examined whether terrain attributes could be used to identify erosion channels that may potentially be eligible for CRP. The analysis presented in this thesis clearly demonstrated that terrain attributes could be used to identify potential locations for grassed waterways. However, in some cases the models could not predict where waterways occurred when perched ground water exited from springs high in the landscapes where high flow would not be expected based on landscape position alone.

The second hypothesis investigated whether USGS data could be used to create high quality maps of potential waterway locations. In some but not all cases, the USGS data produced estimates similar in quality to those produced with the RTK data. This indicated that USGS data is likely not adequately reliable for the application of this modeling approach.

The third hypothesis considered whether more complex neural network rather than logistic regression models would do a better job of identifying suitable locations for grassed waterways. While there were differences in misclassification statistics, there were few fundamental differences in the maps created with the various models. Either modeling approach would be adequate for these datasets.

Of the 8 terrain attributes investigated in this study, many matched up well with existing grassed waterways boundaries (e.g., wetness index, stream power index, channel initiation threshold, plan curvature, upslope contributing area, and delta) whereas the length-slope factor matched up marginally well and profile curvature only matched with Field 1. Although slope matched poorly, slope was still important because it was used to
calculate many of the complex terrain attributes (e.g., wetness index, stream power index).

The sensitivity analyses demonstrated that the neural network procedures used were robust and that most default parameters (e.g., 1 hidden layer, 5 preliminary runs in the optimization step, hyperbolic activation function, and standard deviation standardization procedure) were appropriate for the datasets in this study. The rule of thumb that the number of neurons should be no larger than 3 times the number of variables (i.e., 24 neurons) also appears to be appropriate for this study.

The selection of the flow direction algorithm (e.g., D8, FD8, and DEMON stream-tube method) had a greater impact than some of the other procedures tested. While the D8 method had the lowest misclassification rates for Field 3, this method also produce errors because it is a simple single flow direction algorithm. The more sophisticated stream-tube method produced comparable results with the FD8 algorithm although it generally had higher misclassification rates.

In some cases, the models used in this study over-estimated the occurrence of waterways. However, this may not be very problematic since an NRCS conservationist would still be required to visit these fields in order to verify eligibility for CRP status. What is of greater concern is when the models under-predict areas that would benefit from waterways such as when unexpected flow in the landscape occurs (e.g., Field 4).

Grassed waterways can be predicted in most cases with the RTK data but only in some cases with the USGS data. This modeling approach was robust and could aid conservation planners in identifying suitable areas for waterways more efficiently if accurate elevation data can be acquired.
CHAPTER SIX: FUTURE RESEARCH DIRECTIONS

For many farms, RTK data may be too expensive and USGS data may be inadequate for predicting the locations of waterways. An alternative way to obtain high quality elevation data is with LIDAR. Obtaining LIDAR data for the entire Commonwealth of Kentucky would cost $3 to 4 million (M. Richie, Photo Science, Personal communications, 2008). This data may or may not be as accurate as RTK elevation data. Obtaining LIDAR for only agricultural applications would likely be prohibitively expensive. However, it may be feasible for a large consortium of industries in Kentucky to purchase LIDAR data considering the numerous useful applications of this information: transportation and engineering, soil mapping, disaster (flooding) simulation, quantitatively assessing forest biomass, locating invasive species, and a host of other applications. However, it is still not know if LIDAR is of sufficient accuracy for identifying waterways and this could be the focus of a future research investigation.

It may be possible to develop models that could predict areas that would benefit from waterways over sizable geographic areas such as physiographic regions. For example, it would be highly desirable if the same model parameters could be used to identify suitable locations for grassed waterways throughout many soils in the Outer Bluegrass Region. Potentially similar physiographic regions could be combined such as the Outer and Inner Bluegrass regions. Many of the soils in the Bluegrass and Pennyroyal regions are derived from limestone while many soils in the Purchase region are derived from thick layers of loess over coastal plain sediments. Silt is more easily erodible than clay and terrain varies differently in these regions. Therefore waterway
model parameters may be quite different between these geographic areas. These questions should be examined in future studies.

It is possible for the logistic and regression models to take into account the potential costs that would be incurred if important eroded areas in agricultural fields were not predicted. Because the consequences of missing these locations is more serious than over predicting waterways in other areas, the default threshold could be changed from 0.5 to 0.30 for example. So normally when the model predicts a value <= 0.5, the maps indicate no-waterways. When prediction values are > 0.5, then the maps indicate a waterway. The change of the threshold to 0.3 would mean that more waterways would be mapped in each field. This might produce a more useful map for management. An alternative would be to make maps of probability values rather than the binary logit (waterway or no waterway). This requires further investigation.

The modeling approach used in this thesis was to calibrate the terrain attribute data with site-specific observations (i.e., the presence or absence of waterways) in order to make useful interpretations with terrain attributes. If this approach were to be applied on a large scale (potentially with LIDAR derived terrain data) it may be possible to calibrate this data with existing information about waterways rather than collecting new data. Specifically, the NRCS keeps detailed records and maps of the locations of CRP enrolled grassed waterways across the country. One potential difficulty with using this data is that many different individuals (NRCS conservationists) have contributed to these dataset. Since identifying areas that would be eligible for CRP waterways depends on an NRCS conservationists individual philosophy (Danny Hughes, 2007, Personal communications), there may be considerable variability in these maps which could cause
difficulties with modeling. In addition, the conservationists also include what the farmer is willing and able to manage in their determination of where CRP waterways should be located. Therefore, more research will be required to determine whether it is actually possible to calibrate models with these existing NRCS datasets of CRP waterways.
REFERENCES


Kheir, Rania Bou, John Wilson, and Yongxin Deng. 2007. "Use of terrain variables for
mapping gully erosion susceptibility in Lebanon." Earth Surface Processes and
Landforms 32:1770-1782.

Lagazio, Monica, and Bruce Russett. 2004. "A Neural Network Analysis of Militarized
Disputes, 1885-1992: Temporal Stability and Casual Complexity." In The
Scourge of War: New Extensions on an Old Problem, by Paul Francis Diehl,

Through Remote Sensing and G.I.S: A Comparison of Neural Networks and
Logistic Regression Methods." Proceedings of the 7th International Conference
on GeoComputation. Southampton, United Kingdom: University of
Southampton.

discriminant analysis, neural networks and logistic regression for predicting
species distributions: a case study with a Himalayan river bird." Ecological
Modelling 120:337-347.

Author House.


NRCS. 2007. "National Resources Inventory 2003 Annual NRI (State Report)."


VITA

Adam Clellon Pike

Born November 15, 1982 in Campbellsville, Kentucky at Taylor County Hospital.

Education

- University of Kentucky College of Agriculture
- Bachelor of Science in Plant and Soil Science, emphasis in Crops and Soils

Publications

Research Journal:


Refereed Proceedings Papers:


Extension Publication:


Instructional Material:

Pike, A.C., T.G. Mueller. 2005. GIS Lab 1: Terrain Visualization. PLS468G Class.
Abstracts of Presentations


Pike, A.C., and T.G. Mueller. 2007. Using RTK to Identify Areas that Should be Enrolled in CRP. Presented at the ASA/CSA/SSSA annual meeting.


Other Presentations


In Preparation