

A Comparison of Digital Elevation Models to Accurately Predict Stream Locations

A Thesis

Presented to the

Department of Geography/Geology

and the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment
of the Requirements for the Degree

Master of Arts

University of Nebraska at Omaha

by

Spencer Trowbridge

May 2014

Supervisory Committee:

Dr. Rex Cammack, Chair

Dr. Michael Peterson

Dr. John McCarty

A Comparison of Digital Elevation Models to Accurately Predict Stream Locations

Spencer Trowbridge, MA

University of Nebraska, 2014

Advisor: Dr. Rex Cammack

Three separate digital elevation models (DEMs) were compared in their ability to predict stream locations. The first DEM from the Shuttle Radar Topography Mission had a resolution of 90 meters, the second DEM from the National Elevation Dataset had a resolution of 30 meters, and the third DEM was created from Light Detection and Ranging (LiDAR) data and had a resolution of 4.34 meters. Ultimately, stream locations were created from these DEMs and compared to the National Hydrography Dataset (NHD) and stream channels traced from aerial photographs.

Each bank of the named streams of the Papillion Creek Watershed were traced and samples were obtained that represent error in the placement of the derived stream locations. Measurements were taken from the centerline of the traced stream channels to where orthogonal transects intersected with the derived stream channel of the DEMs and the streams of the NHD.

This study found that DEMs with differing resolutions will delineate stream channels differently and that without human assistance in processing elevation data, the finest resolution DEM was not the best at reproducing stream locations.

ACKNOWLEDGEMENTS

While volunteering at Midland Lutheran College (MLC) in Fremont, NE, where I earned a Bachelor of Science degree, I was asked by Dr. Allyson Backstrom if I had ever thought about GIS. While Dr. Backstrom was at Midland during my undergraduate years, she was not one of my instructors. However, every single meeting I had with her pushed me forward academically. I thank her for pointing me in this direction, and I am thankful for ultimately finding myself at the University of Nebraska at Omaha (UNO) studying Geography, GIS, and Cartography with the most honest, skillful, and inclusive group of professionals I have known since leaving Midland.

I am particularly grateful to Dr. Cammack, my thesis advisor, for convincing me to pursue a Master's degree. He has guided me through this thesis, and he has helped to provide me with opportunities for travel, as well as personal and professional growth, that I never thought were possible for me. He always had ideas for projects that helped me develop professionally, and he helped me become an independent learner by allowing me to explore my own ideas and capabilities. I will always be appreciative of your honesty and openness to talk at a moment's notice. I will always strive to emulate that during my career. I know you have prepared me well.

Dr. Peterson has significantly helped me during my progress in graduate school. He allowed me to explore new mapping techniques and share them with his classes. I find it interesting how intimidated and lost I felt during the first week of his class, but now find myself wishing to make the things I was learning a large part of my career. I wish to thank you for continuing to make learning enjoyable, as well as challenging and relevant for all of your students. Our talks are always entertaining, as well as enlightening for me.

Dr. Dando gave me the writing advice I needed and the chance to put my cartographic skills to use. I am thankful for the opportunity she gave me to develop as a researcher and to be a colleague of true professionals. Thank you for taking a chance on me.

I truly admire Brenda Todd. With her talents of being inclusive, and her willingness to lend an ear, she is amazing. I could not have succeeded without you being there.

I wish to recognize and thank my fellow Teaching Assistants. Their abilities to challenge themselves, remain humble, and still have fun, even when the pressure was on, was mind-blowing. I was constantly in awe of their talents. Gentlemen, to me, you are all good guys, and where I come from, that is the biggest compliment one can give. I know you will all find success. Thanks for sharing the challenges, and for sharing your knowledge and skills with me.

I would also like to recognize and thank the Cartography and Geographic Information Society (CaGIS) and the North American Cartographic Information Society (NACIS). Those organizations provided funding for me to attend professional conferences which served to inspire me and validate my decision to study Geography, GIS, and Cartography at an advanced level.

Dr. Carlson provided me with my desire for a lifetime of learning. He cultivated my interest in Earth Science while at MLC. He also provided me with projects that have facilitated my professional development on an international level. He also set the example for me to follow while educating students at UNO through my assistantship. Thank you for all you continue to do for me.

I wish to thank Chad Fox. Without his help I would not have been in graduate school. He is not only an inspiration to his students and athletes, but to me as well. I have tried very hard to not let you down.

I would genuinely be remiss if I did not thank Holly Lusk. She was a friend to me when I needed one most. Her checking in on me from time to time made all the difference. I always smiled when sending her reports of my progress, hoping she was proud of me. Thank you, Holly. I love you.

TABLE OF CONTENTS

Title Page.....	i
Abstract.....	ii
Acknowledgments.....	iii
Table of Contents.....	vi
List of Figures.....	viii
Chapter	Page
I. INTRODUCTION	
1.1 Introduction.....	1
1.2 Data Terms.....	1
1.3 Software and Hardware.....	1
1.4 Problem Statement.....	2
1.5 Research Objective and Hypothesis.....	3
1.6 Study Area.....	3
1.7 Significance of Research.....	11
1.8 Organization of Thesis.....	11
II. LITERATURE REVIEW	
2.1 Introduction.....	12
2.2 Early Research and Pioneering Methods.....	12
2.3 D8 Method.....	13
2.4 D9 Method.....	15
2.5 Other Methods of Stream Extraction.....	17
2.6 Problems Observed with D8 Method.....	20
2.7 Rho 8 Method.....	21
2.8 Methods for Evaluating Stream Delineation.....	23
2.9 DEM Resolution Comparison Studies.....	24
2.10 Conflicting Results in Different Terrain.....	25
2.11 Lack of Evidence Supporting DEM Use for Feature Extraction.....	25
2.12 Summary.....	26
III. METHODOLOGY	
3.1 Introduction.....	27
3.2 Datasets Used.....	29
3.3 Naming Conventions and Geodatabase Creation.....	29
3.4 Obtaining NED and SRTM Data.....	30
3.5 Raster Creation from LiDAR Data: ArcGIS Tools Used.....	31
3.6 LiDAR Storage Location and Settings.....	31
3.7 Average Point Spacing Determination.....	31
3.8 Initial Surface Creation.....	32
3.9 Raster Creation.....	32

3.10 Derivation of Stream Channels.....	32
3.11 ArcGIS Hydrology Tools and Feature Extraction.....	32
3.12 Individual Stream Geodatabase Creation.....	37
3.13 Stream Tracing.....	38
3.14 Stream Tracing Quality Control Using NHD.....	39
3.15 Stream Tracing Generalization.....	40
3.16 Baselines.....	43
3.17 Casting Transects.....	44
3.18 Fine Tuning Traced Stream Channels.....	46
3.19 Distance Data Collection Feature Requirements.....	47
3.20 DSAS Extension.....	47
3.21 Data Collection Quality Control.....	51
3.22 Initial Distance Calculations and Data Manipulation.....	52
3.23 Spreadsheet Setup.....	53
3.24 Traced Stream Width Calculation.....	53
3.25 Summary.....	55
IV. RESULTS	
4.1 Introduction.....	58
4.2 Rationale.....	58
4.3 Distance Data storage and Naming Conventions.....	59
4.4 Determination of Overall Accuracy of Predicted Streams.....	59
4.5 Rationale for Determining Stream Width.....	60
4.6 Determination of the Number of Correctly Placed Samples.....	61
4.7 Raster Cell Size Considerations.....	62
4.8 Distances of Derived Stream Channel Locations.....	64
4.9 ANOVA.....	64
4.10 Box-Cox Transformation.....	65
4.11 Initial Analysis.....	66
4.12 Rationale for regrouping datasets.....	69
V. SUMMARY	
5.1 Introduction.....	72
5.2 Datasets Examined.....	72
5.3 Procedure Review.....	73
5.4 Study Area Considerations.....	74
5.5 Resolution Comments.....	76
5.6 Preprocessing Comments.....	77
5.7 Overall Cost Considerations.....	77
5.8 Conclusions.....	79
REFERENCES.....	81

LIST OF FIGURES AND TABLES

Figures	Page
1.1 Location of the Papillion Creek Watershed in eastern Nebraska.....	4
1.2 Named stream locations.....	6
1.3 Average precipitation amounts and temperatures.....	7
1.4 Calculated placement of flood control lakes.....	8
1.5 Physiographic provinces of the United States.....	9
1.6 Relief map of the watershed and the surrounding region.....	10
2.1 D_{∞} Method of flow distribution.....	16
2.2 DEM created using D8 method showing parallel flow lines.....	18
2.3 DEM created using Rho8 method.....	19
2.4 Parallel extracted stream locations in floodplain areas using D8 method.....	21
3.1 Stream Extraction Flowchart.....	28
3.2 Flow Accumulation diagram of D8 method.....	34
3.3 Flow Accumulation – LiDAR dataset.....	36
3.4 Flow Accumulation followed by Stream to Feature tool.....	37
3.5 Example of precise stream tracing.....	41
3.6 Stream channel tracing.....	42
3.7 Single baseline.....	44
3.8 Multiple baselines.....	45
3.9 Cast transects setting tab.....	48
3.10 Shoreline calculation settings tab.....	49
3.11 Calculation Change Statistics tab.....	50
3.12 Distance calculation method.....	55
4.1 Number of correctly placed samples.....	62
4.2 Number of correctly placed samples within respective cell sizes.....	63
4.3 Bimodal distribution of LiDAR sample data.....	66
4.4 Mean distances of samples to traced stream banks.....	67
5.1 Computer processed stream locations near the mouth of Papillion Creek.....	75
Tables	Page
4.1 ANOVA summary of all datasets.....	68
4.2 SNK of all datasets.....	68
4.3 ANOVA summary of computer processed and human assisted datasets.....	70
4.4 SNK of computer processed and human assisted datasets.....	71

INTRODUCTION

1.1 Introduction

This research examines the abilities of three digital elevation models (DEMs) from various sources and different resolutions to predict stream locations. All were used to produce the stream channel locations of the Papillion Creek Watershed in eastern Nebraska. The purpose is to compare how accurately each locates the streams so that better decisions can be made with regard to issues pertaining to drainage networks, and in order to produce a method to evaluate light detection and ranging (LiDAR) data, the newest method for creating DEMs. This study also inspects the differences between elevation data processed solely by computer and data processed with human intervention to aid in feature extraction.

1.2 Data Terms

Three DEMs were compared in this study: 1) Shuttle Radar Topography Mission (SRTM) 2) National Elevation Dataset (NED) based on United States Geological Survey (USGS) topographic maps and 3) a DEM created from Light Detection and Ranging (LiDAR) data obtained from the Nebraska-Iowa Regional Orthophotography Consortium (N-IROC). A fourth dataset from the National Hydrography Dataset (NHD) provided additional centerlines for the stream network. The distances of extracted stream channels to the traced stream channel locations from aerial photographs were used to evaluate the DEMs in this study.

1.3 Software and Hardware

The Environmental Sciences Research Institute (ESRI) ArcGIS software programs, ArcMap and ArcCatalog were used extensively during this project. The ArcGIS extension, Digital Shoreline Analysis System (DSAS) was used to obtain

distance, or error, data. The DSAS extension is available as a free download from the U.S. Geological Survey (USGS) at <http://woodshole.er.usgs.gov/project-pages/dsas/version4/index.html>.

Microsoft Excel was used to manipulate the distance sample values into a usable format. Statistical Analysis Software (SAS) was used for inspecting the data and performing analysis.

The majority of computational work for this project was done using a desktop home computer with a 3.00 GHz processor and 8 GB of RAM. Several external storage devices were employed for storage of photographs and LiDAR data, and for backup in the event of a catastrophic failure

1.4 Problem Statement

There is a need to know if there is a difference between DEMs with varying resolutions to accurately represent or reproduce stream channel locations that are automatically extracted with a GIS. Specifically, there was a need to know the differences between LIDAR, SRTM, and NED DEMs to accurately determine stream channel locations. This LiDAR dataset has not been compared to the others for this application. Furthermore, while LiDAR is often used to represent various geomorphological features such as stream channels, there has been little research to truly evaluate its effectiveness to do so (Rayburg 2009).

Logic would tell us that increased resolution of a DEM would allow for more precise mapping. However, this increase in resolution may be unnecessary when mapping a stream channel. There must be an upper limit to the resolution needed in order to obtain adequate feature locations. Quinn et al. (1991) and Zhang and Montgomery (1994) point out that the scale of elevation change, and resolution needed to extract stream channels in

different types of terrain, are directly related. In flatter terrain, it becomes difficult to apply the same algorithms to the DEM to obtain the proper flow directions that would allow the software to differentiate a stream from flat ground.

While not as mountainous as the terrain used in Mouton's (2005) and McMaster's (2002) DEM stream extraction studies, the Papillion Creek Watershed was evaluated here to determine how appropriate each of the datasets are for producing stream networks in flatter areas. The Papillion Creek watershed has also been mostly channelized to protect against flooding, which is different than similar comparison studies. This project will add knowledge to the fields of remote sensing, mapping, and fluvial geomorphology by providing a way for researchers to evaluate the mapping of streams directly from DEMs.

1.5 Research Objectives, Hypothesis

Objectives:

Determine the resolution of a digital elevation model that best predicts stream channel locations.

Develop a method to determine a DEM's ability to predict a stream network.

Hypotheses:

A finer resolution DEM does not affect its ability to accurately predict a stream channel's location.

Human and computer processed DEMs produce stream networks in the same way.

1.6 Study Area

The Papillion Creek Watershed in eastern Nebraska was selected for this study due to the availability of datasets with varying resolutions. The watershed lies in portions of Washington, Douglas, and Sarpy counties. According to the U.S. Census Bureau (2010), the populations of the three counties total 696,184, with 20,234 people living in

Washington, 158,840 in Sarpy, and 517,110 in Douglas. This population constitutes nearly one-third of the total population of the entire state of Nebraska.

While Nebraska is a large state having a total area of over 77,000 square miles, the study area is relatively small at only 402 square miles (<http://www.papillioncreek.org/overview.html>). It occupies an area between the much larger watersheds of the Elkhorn, Platte, and Missouri Rivers. The location of the Papillion Creek Watershed can be seen in Figure 1.1.



Figure 1.1. Location of the Papillion Creek Watershed in eastern Nebraska.

The drainage system generally trends toward the southeast in a dendritic pattern from the city of Blair in Washington County through the cities of Omaha, in Douglas County, and Papillion and Bellevue, in Sarpy County. In the lower portion of the

watershed, the stream channels have nearly all been channelized to protect against flooding.

Large tributaries of the of the Big Papillion Creek are the South Papillion Creek, West Papillion Creek, Little Papillion Creek, Papillion Creek, Thomas Creek, Cole Creek, and the Northwest Branch of West Papillion Creek. The other named streams used in this study to evaluate the various DEMs are Walnut Creek, Southwest Branch, Richter Branch, North Branch, Mud Creek, Leach Branch, Hell Creek, Falls Branch, East Fork, Copper Creek, Butter Flat Creek, Boxelder Creek, Boston Branch, and Big Elk Creek. The streams used in this research project are shown in Figure 1.2.

The study area has received an average of 30.62 inches of rainfall annually from 1981 to 2010 (<http://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data>). The area receives the majority of this precipitation between the months of May, June, July, and August. Precipitation and temperatures observed from 1981 through 2010 are summarized in Figure 1.3 from the National Oceanographic and Atmospheric Association (NOAA) at <http://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data>. The study area is prone to periods of drought as well as thunderstorms and heavy rainfall that can lead to flooding.

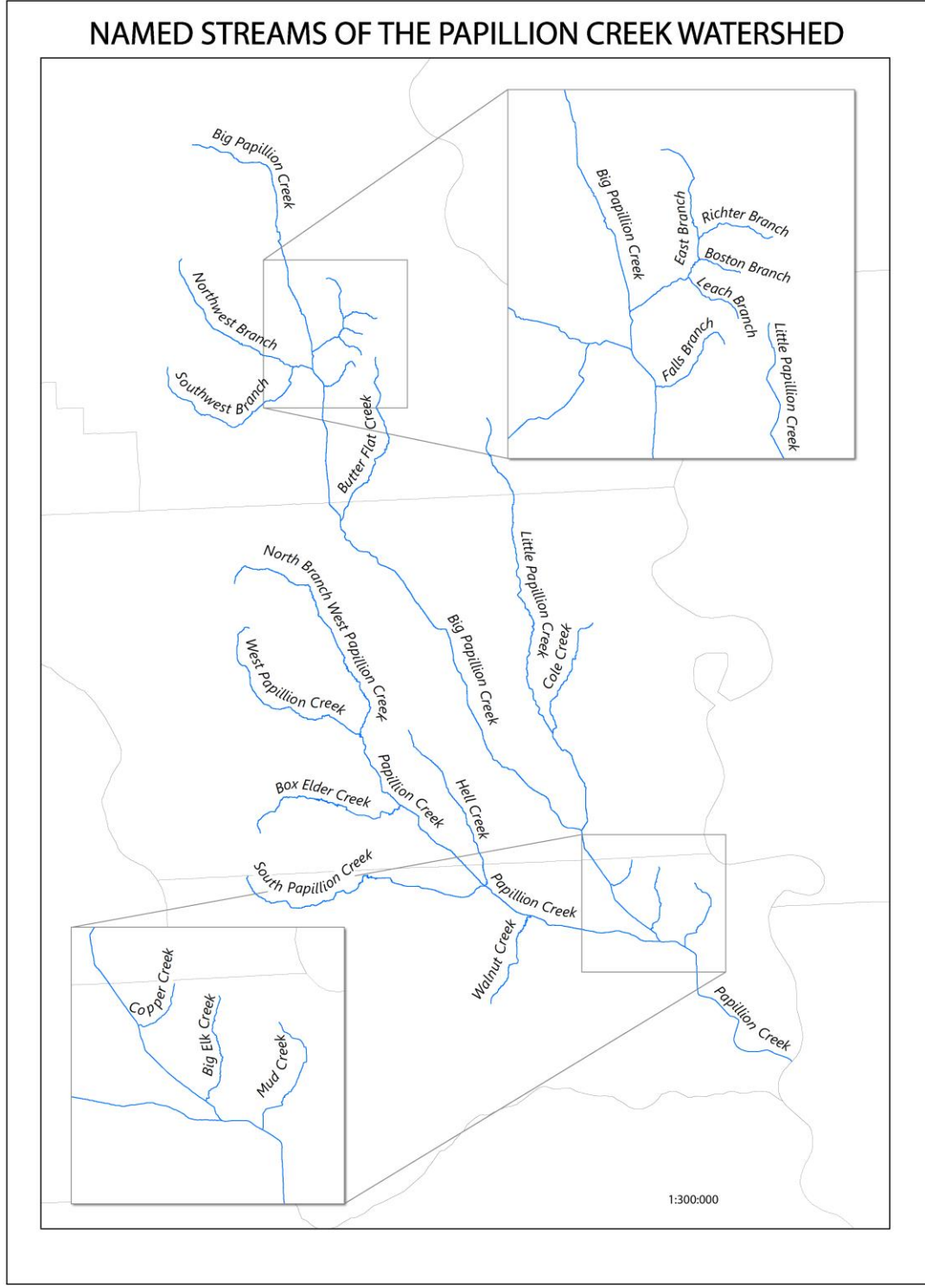


Figure 1.2 Named streams of the Papillion Creek Watershed analyzed in this study. DEMs with different resolutions were evaluated in their ability to reproduce stream locations.

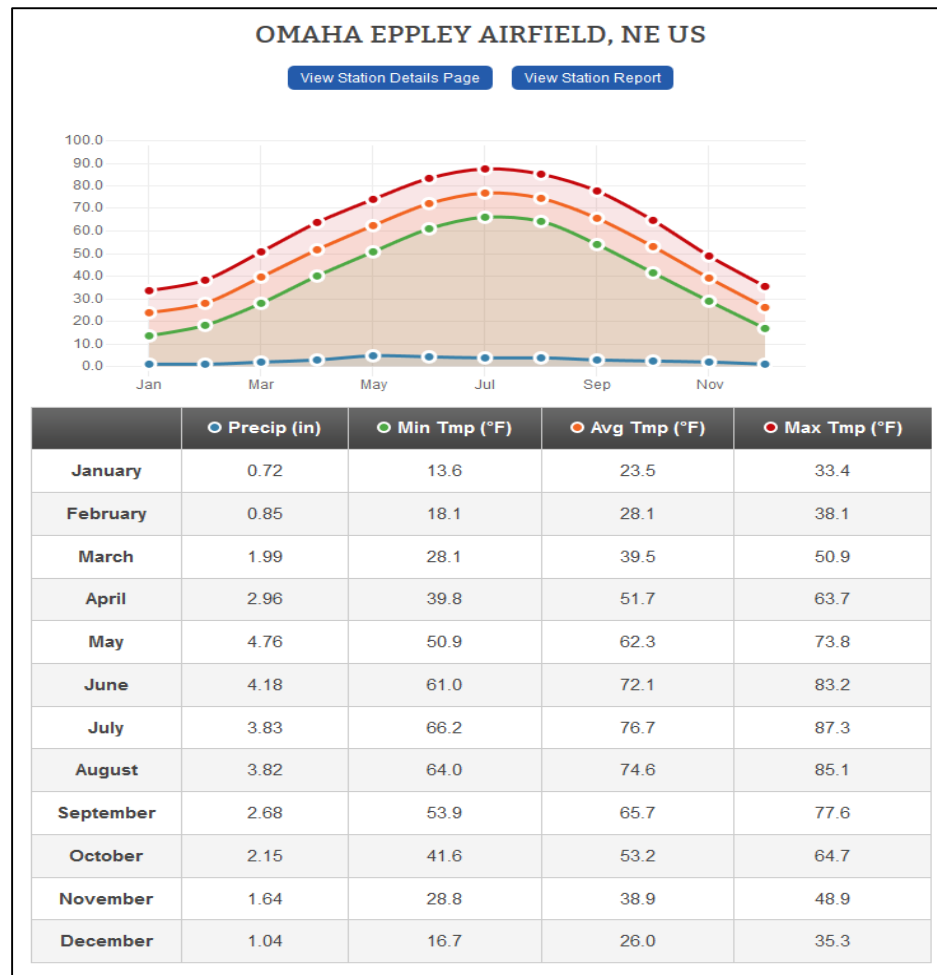


Figure 1.3. Average precipitation amounts and temperatures observed from 1981-2010.
<http://www.ncdc.noaa.gov/land-based-station-data/climate-normals/1981-2010-normals-data>

Accurate locations of stream channels and the calculation of flow accumulation data can affect many people. It is important to understand the true nature of the area so that better decisions can be made during flood events. Several flood control lakes are spread throughout the watershed to protect the more populated areas of Douglas and Sarpy Counties. Some of the larger flood prevention lakes are Zorinsky Lake on Boxelder Creek, a tributary of the West Papillion Creek, Standing Bear Lake, a tributary of the Big Papillion Creek, and Glenn Cunningham Lake on the Little Papillion Creek, and Wehrspann Lake on Wehrspann Creek, a tributary of the South Papillion Creek

(<http://www.papillioncreek.org/overview.html>). One can observe the calculated placement of these man-made lakes in Figure 1.4.

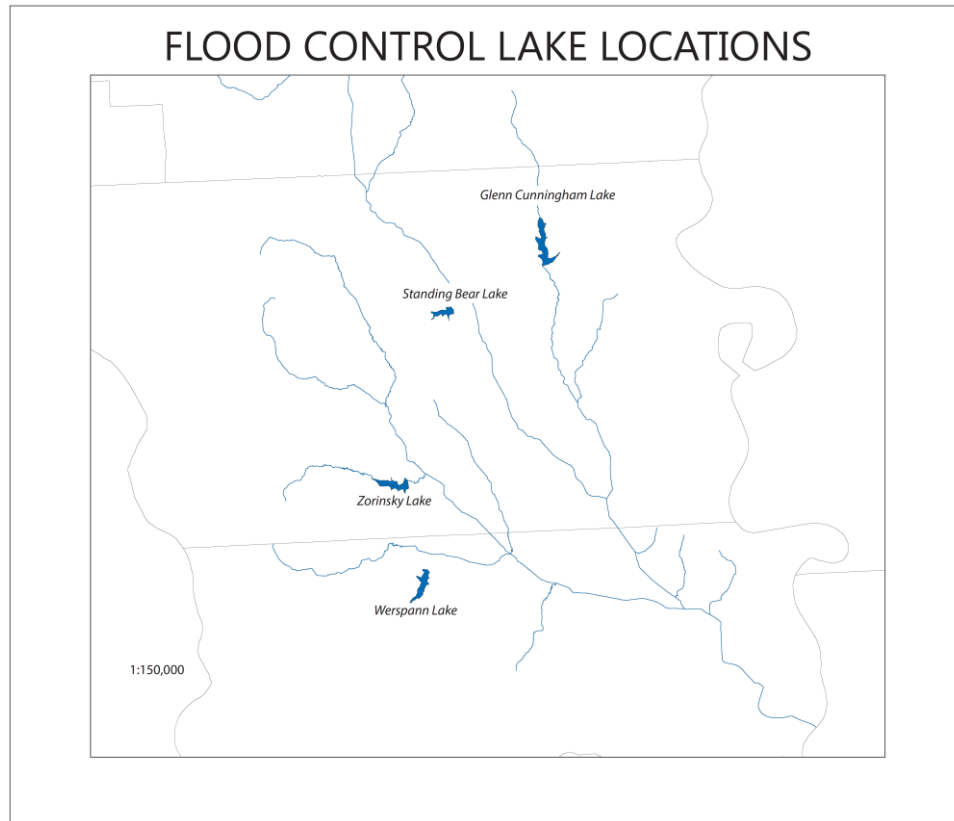


Figure 1.4. Calculated placement of man-made flood control lakes protecting the metropolitan area of Omaha, NE.

The study area lies at the western edge of a large geographic area within the North American continent known as the Central Lowlands (Atwood, 1938). The study area lies in a transition zone between the Central Lowlands and the Great Plains region that steadily rises in elevation by approximately 10 feet per every mile traveled west toward the Rocky Mountain region of the western United States. These provinces are shown in Figure 1.5.

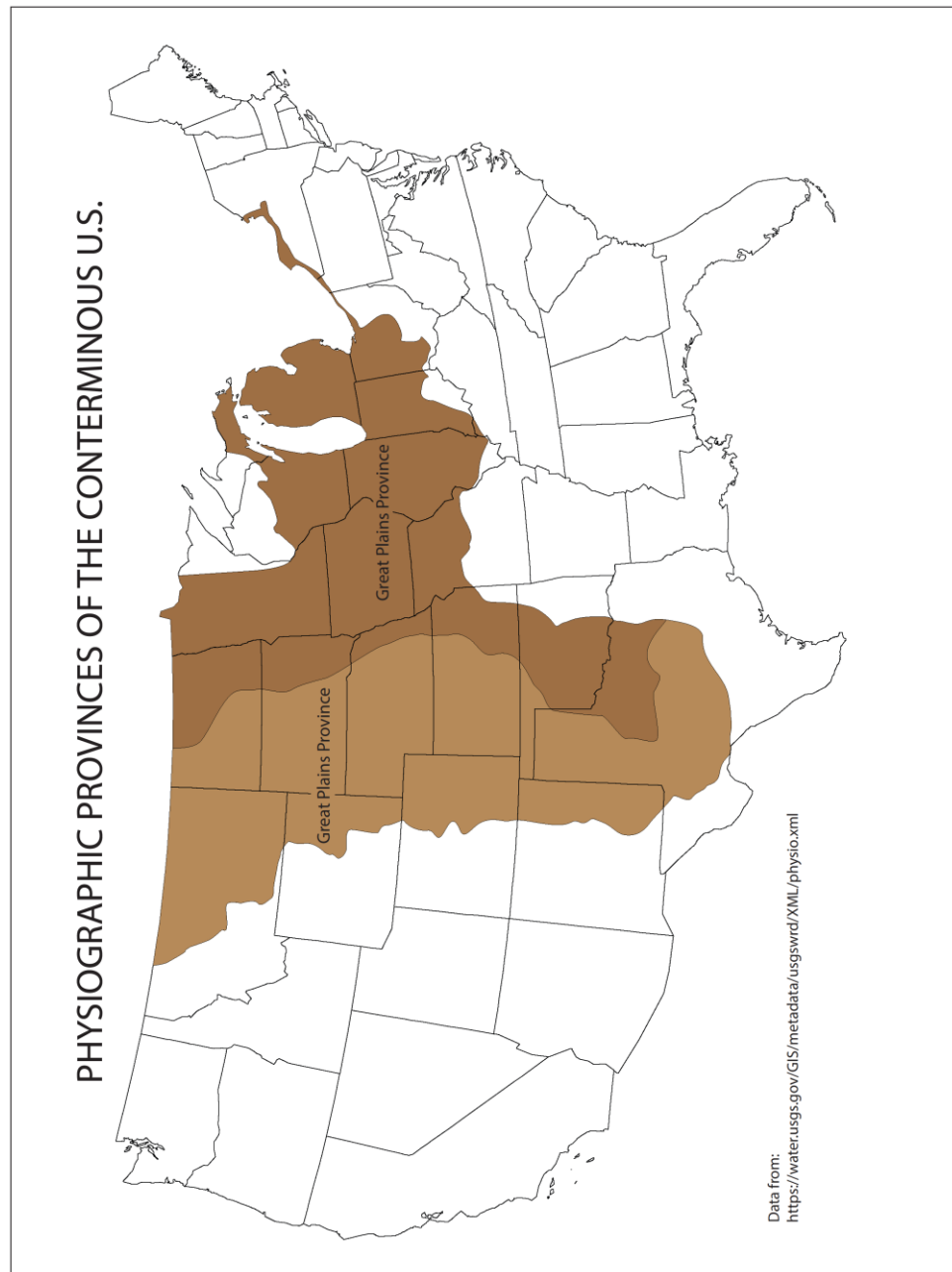


Figure 1.5. The Great Plains and Central Lowlands physiographic provinces of the United States. Data Source: <http://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml>.

The Central Lowlands, and the study area in particular, are characterized by their rolling hills as bluffs next to the Missouri, Platte, and Elkhorn River valleys. In the study area, these rolling hills may be remnants of the last glacial maximum where the debris transported by the ice sheets were left in place as till following the glacial retreat. While

the watershed has a low elevation minimum of 286 meters above sea level at the mouth of the Missouri River, the maximum elevation is over 400 meters at the north and west edges of the watershed. The streams of the watershed have incised the glacial till and helped to create the hilly topography of the area.



Figure 1.6. Relief of the watershed and the surrounding region.

1.7 Significance of Research

This project offers a new way to evaluate DEMs and their abilities to produce fluvial channels in a landscape. Often, DEMs are evaluated based solely on visual inspection rather than assigning any quantitative assessment, particularly its ability to reproduce a stream network. LiDAR, in particular, has been growing in use. This is partly due to the speed in which a highly detailed, and vertically accurate, surface can be created. Yet, there is very little evidence to support its use in determining the locations of fluvial channels. The method used in this study can provide a way to validate its use. Various features derived from a DEM could be evaluated in the same way using the method presented here.

The hope is that this study will allow researchers to advance flow routing algorithms and surface modeling by giving them a tool to evaluate their work in a mathematical way. It should allow comparisons to other methods currently in use.

1.8 Organization of Thesis

This thesis consists of chapters organized in the following way. Chapter 2 focuses on literature used to determine the current research trends in the field. Much of the chapter discusses various algorithms currently in use to extract stream channel locations from DEMs. Other reviews will lead one to understand the need for continued study of the abilities of DEMs to reproduce surface features in a variety of terrains. Chapter 3 outlines the method used in this research project to create a DEM from LiDAR data. Chapter 3 also demonstrates how distance sample measurements were collected for comparison of the datasets. Chapter 4 shows the results of the distance measurements obtained. Chapter 5 is a summary of the research and provides some conclusions of the results.

LITERATURE REVIEW

2.1 Introduction

While there have been numerous studies comparing the abilities of digital elevation models (DEM) to reproduce stream networks, there is a lack of any quantifiable way of doing so to determine the horizontal accuracy of the reproduced streams. Once stream locations are extracted from a DEM using various computer algorithms, a visual inspection method is often employed for comparison. Overall stream length and sinuosity characteristics have been computed, and this tends to serve as a way of validating newer algorithms and DEMs, but the actual position of a reproduced stream channel can differ from reality even though stream length, sinuosity, and other measurable characteristics of a stream can be exact. This literature review focuses on the history of the delineation of stream channels from DEMs and offers examinations of several studies that compare DEMs and their abilities to reproduce stream channels through a visual comparison. The review also establishes that horizontal accuracy measurements are typically not provided in a quantifiable way.

2.2 Early Research and Pioneering Methods

Some of the first efforts to determine physical features from elevation data was performed by Peucker and Douglas (1975). Using codes for elevation data for individual cells with regard to their surrounding cells, they were classified as either pits, passes, ridges, ravines, or breaks. Moore, Grayson, and Ladson (1991) give a detailed review of early surface studies using DEMs, as does Florinsky (1998) where he explains many of the concepts in the field of terrain analysis and fluvial morphology feature extraction. More recently, Wilson et al. (2008) evaluated flow routing algorithms and described the current research being done in this field.

Kiss (2004) discussed how the Peucker and Douglas (1975) method was limited by the minimal computing power of the time. O'Callaghan and Mark (1984), with their much used method for stream extraction, were aware of the limitations due to the lack of computer memory at the time, which was ten years after Peucker and Douglas' research. In the years since, computing power has advanced. In fact, the stream extractions for this thesis were performed at home using a desktop computer.

2.3 D8 Method

Much of the current research to evaluate the extraction of stream channels builds on a few simple terrain analysis techniques. The most commonly used algorithms for this purpose are known as D8 and D ∞ .

The most widely accepted method for extracting stream channels from a DEM is O'Callaghan and Mark's (1984) D8 method (Jones, 2002). This method has been named D8 for the eight directions in which water can flow from the center of one square cell of a DEM raster to one of the eight neighboring cells.

The D8 method is known as a single flow direction algorithm (SFD) where all of the water is assigned a single flow direction based on the elevation of cells in a 3 X 3 grid. Each cell of the gridded raster has an elevation value assigned to it. When determining a cell's flow direction, the algorithm compares the values of all the surrounding cells and assigns the direction of flow to the cell that has the lowest elevation value. In the natural environment water follows the gradient, or the steepest descent, and so, by assigning the cell with the lowest elevation value surrounding the cell in question the stream flow direction, it is reasoned that this direction will most closely match the reality of the stream following the gradient.

O'Callaghan and Mark (1984) do discuss a channel and how it differs from just a depression in the land whereby fluvial processes are more pronounced than slope processes such as mass wasting. This means there would be more total flow from the water going in the direction of steepest descent, and transporting material as the water flows, rather than any kind of slope failure, another material transport process that creates surface features and can also produce channel-like features. In the D8 algorithm these differences would still be shown as flow. It is the overcoming of slope processes by fluvial processes that the D8 algorithm, and subsequent stream channel delineation, is based upon (O'Callaghan and Mark, 1984).

To assist in the determination of this tipping point, the D8 method calculates how many upstream cells have contributed to each subsequent cell considered to be downstream within the surface model. Using this data, the user can specify a threshold value so as to not delineate every single flow path visually from one cell to another on a map. This could be considered a form of generalization where some of the natural deviations in the surface are simply not shown, although the algorithm still takes these cells into consideration when it produces the final stream channel locations. The adding of cells within the algorithm has been termed Flow Accumulation, and it helps to delineate the stream channels in the event of a pit, or depression, in the raster representation of the surface where water would not continue to flow.

O'Callaghan and Mark (1984) reference Speight's (1968) method in which he calculates how many contour lines a 100 foot long line drawn following the gradient would cross if a user were to draw one on a contour map. In essence, he was accumulating how many drops in elevation occurred along those 100 foot lines. A location that had the most number of drops, or accumulations, was assigned the stream

location. O'Callaghan and Mark's (1984) D8 method accumulated how many drops occurred from cell to cell regardless of the overall change in elevation. Their method did not bother with specific contour intervals, or the exact amount of drop in elevation that occurred, but instead focused on the overall number of cells that contributed to each downstream cell.

Once a line of cells representing stream locations has been calculated and represented, vector lines can be drawn from the center of each cell to the center of each downstream cell. These vector lines are then displayed to users of the D8 method for stream delineation. Flow Accumulation is explained further in Chapter 3.

2.4 D_{∞} Method

Another popular algorithm for stream delineation from elevation data is known as the D_{∞} method. This method, introduced in 1997 by David Tarboton, attempted to represent flow directions and upstream areas in a new way. Rather than assigning an entire cell's worth of stream flow to the lowest of the eight cells surrounding it, Tarboton's method (1997) distributes the flow between two cells. The distribution is based on how close the calculated direction of flow is to the center of that cell that is lowest in elevation of the eight surrounding ones. O'Callaghan and Mark's method differs in that it assigns all of the flow from one cell to the next lowest cell, regardless of any calculated exact direction of flow.

In the D_{∞} method, the edges of a cell are divided by placing nodes at each midpoint, and at the corners of that cell. Water can be thought to flow freely from one cell to the two lower cells that are adjacent to it, or to only the single lowest cell, depending on the direction of the gradient from the initial cell. Again, the amount of water that is assigned by the algorithm to flow from one cell to another, or possibly two,

The D8 and D_{∞} methods for determining stream flow are similar in that they compare elevations in a gridded format. They also assign flow to cells lower in elevation and accumulate this flow over an area. Thus, one is able to extract stream locations. However, they are different in that the accumulation from one cell to another might diverge slightly at fine scales in the D_{∞} method presented by Tarboton.

Both the D8 and D_{∞} methods suffer from an elevation raster containing pits, or areas of the surface where water will not flow out (O'Callaghan and Mark, 1984; Tarboton, 1997). In O'Callaghan and Mark's method, a filling of these pits, or flooding them as Tarboton (1997) refers to the process, to the lowest available cell elevation corresponding to the surrounding cell with the lowest elevation, results in that pit "spilling out" and continuing on to that cell as a stream channel. Tarboton's D_{∞} method makes use of O'Callaghan and Mark's (1984) method for pit filling to solve this problem (Tarboton, 1997).

Mark (1988) does make note of these pits being rare in the natural world at scales larger than 10 meters, besides in unique geomorphologic features. These pits in DEMs typically correspond to the overly precise measurements of elevation data, and based on Mark's (1988) study, it may be interpreted that when high resolution data is used to automatically map something on the scale of a stream channel, excessive resolution may not be necessary. Again, D8, and the pit filling method put forth by O'Callaghan and Mark (1984) is the most common way to process a DEM for hydrological analysis (Jones 2002).

2.5 Other Methods of Stream Extraction

While the D8 and D_{∞} methods for stream channel extraction are well-known, there are other methods in use. Wilson and Gallant (2000) point out that the Rho8

method, developed by Fairfield and Leymarie (1991), is a single flow direction algorithm that attempts to eliminate a problem of D8 methods where parallel flow lines occur in representing stream networks due to the limited eight directional choices for flow. The images for Fairfield and Leymarie's Rho8 method compared to D8 are striking, and are shown in Figures 2.2 and 2.3.

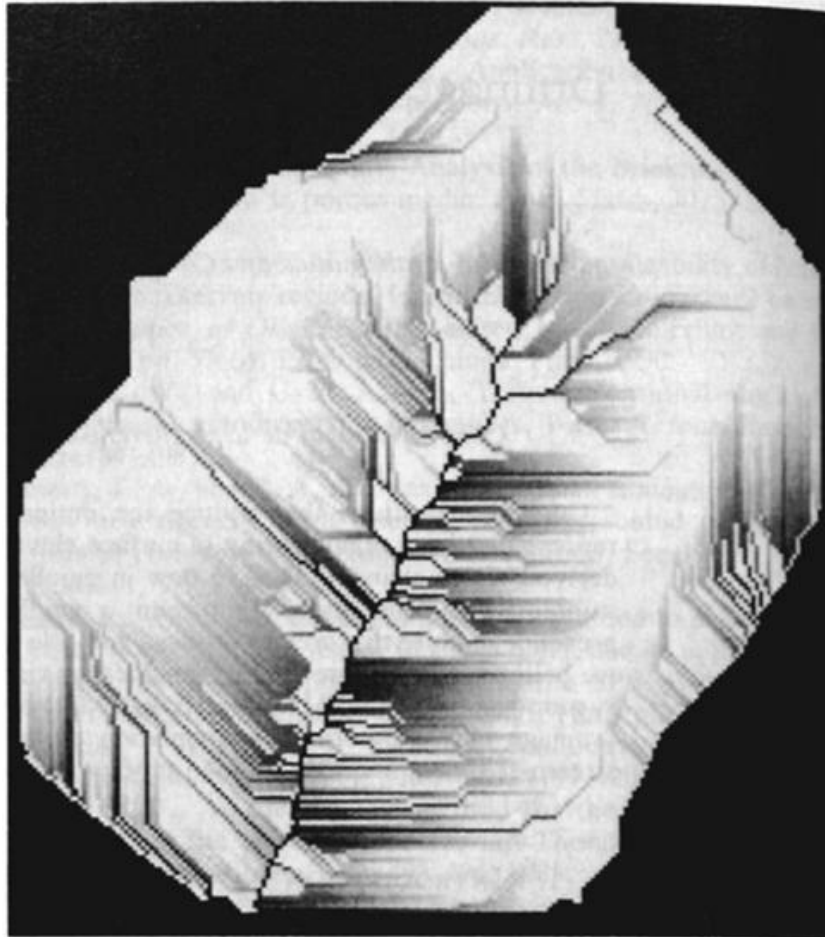


Figure 2.2. DEM created using D8 method with parallel flow lines clearly visible due to the limited number of choices for flow direction from cell to cell of a raster. From Fairfield and Leymarie (1991).

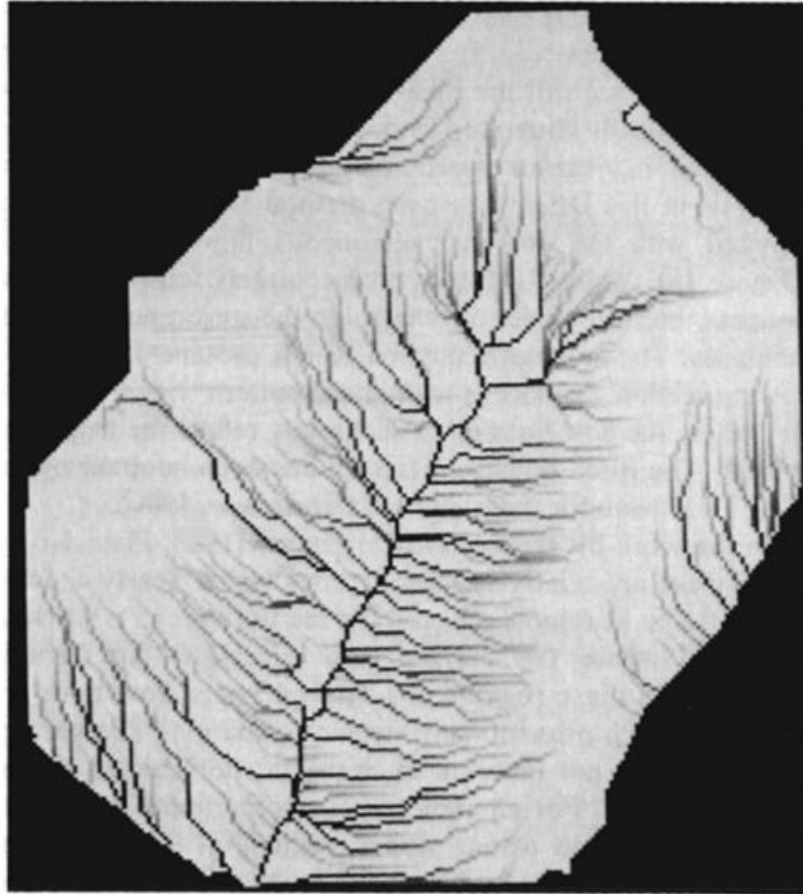


Figure 2.3. DEM created using Rho8 method. Parallel flow lines are not as evident when the flow can be routed to more than eight cardinal directions. From Fairfield and Leymarie (1991).

Comparing Figures 2.2 and 2.3 it would appear that the Rho8 method is desirable, and O'Callaghan and Mark (1984) do note that these parallel channels created with their D8 method would not be wanted in certain circumstances such as in cartographic representations. However, the D8 method is repeatable. A researcher will get the same stream channels extracted in the same place every time. This characteristic is desirable when trying to scientifically evaluate a new dataset's quality, but Moore, O'Loughlin, and Burch (1988) argue that the D8 method represents flow only in a rough sense, and that the discrete directions for which the method is known are too coarse to be used for any detailed modeling of hydrologic surfaces. In the Moore, O'Loughlin, and Burch (1988)

paper they use this notion as a means to promote the use of contour-based flow routing algorithms. The research for this thesis compares only the named, higher order streams of the Papillion Creek Watershed. Most of the parallelism was generalized away prior to analysis due to the scale at which these named streams are mapped by the algorithm. Thus, the D8 method for stream extraction is employed by ArcGIS.

2.6 Problems Observed With D8 Method

Fairfield and Leymarie (1991) did note that the creation of parallel stream channels was not scale dependent, and in this thesis parallelism was evident for all datasets representing a variety of resolutions. Even after the removal of some smaller streams, Figure 2.4 still shows this phenomenon. However, this thesis was more concerned in mapping the named streams, which, again, typically correspond to higher order streams. So, upon generalization of the stream network further, the parallel flow lines disappeared.

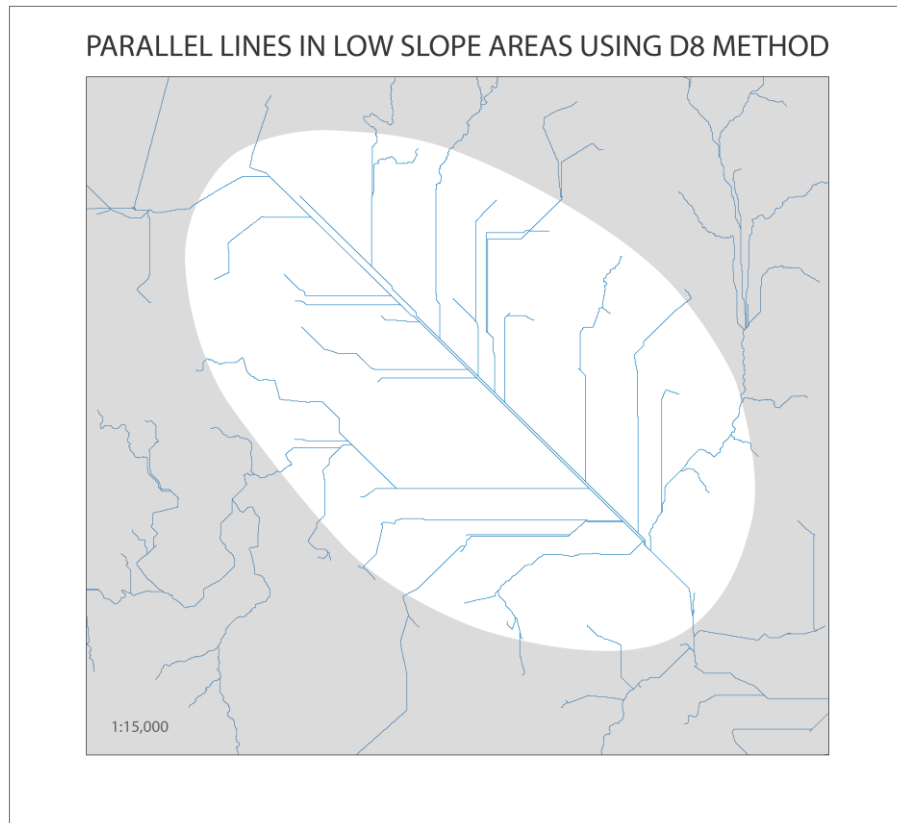


Figure 2.4. Parallel extracted stream locations in floodplain areas using the D8 method. NED DEM (30 meter cell size).

2.7 Rho 8 Method

The Rho8 method employs a bit of randomness to the flow direction from cell to cell. The actual direction is assigned based on a multiplier taken from a random value obtained from a cumulative distribution function (cdf) (Moore, O'Loughlin, and Burch, 1988). Those authors do address an assumption by many that this introduces noise into the data of a method that already contains some amount of noise in the way of imprecision in the data.

Moore, O'Loughlin, and Burch (1988) argue that in areas of low relief, such as gently rolling plains or tidewater regions, and at the end of an incorrect run of flow cells, the position could be off by a great deal. This phenomenon is independent of the resolution of the DEM in deterministic models such as D8 where there is a limited

number of choices for the flow out of any particular cell. They offer that in stochastic models which are designed to introduce probabilistic, or random noise into a DEM to smooth the surface, flow direction can be made to follow correct positions more accurately regardless of the resolution of the DEM. However, it is not clear how these new positions were evaluated when compared to other algorithms.

In O'Callaghan and Mark's (1984) paper they offer that unrealistic patterns appeared in areas where the gradient was very low such as the parallel flow lines, for instance. They also make note that in one of their study areas, which corresponded to a floodplain, there were a large number of pits and depressions in the surface, and they even referred to the DEM as "noisy" while in steeper terrain pits are seen as "dams" in the flow path, yet they are easily overcome by the large drop in elevation.

Further, O'Callaghan and Mark (1984) do imply that any errors in the data in fluvial landscapes are more likely to produce pits in the digital surface where the slope is slight; in other words, where the "signal," as they refer to it, is weak. There have been many studies to determine the appropriate signal to noise ratio for stream channel reproduction. Signal being the actual drop in elevation from cell to cell over a given distance, or slope, and noise being the variation in the accuracy of the elevation data for the surface prior to stream channel extraction.

There have been many iterations of algorithms to create digital surfaces. Many of these algorithms employ a variety of weighting factors to various surface features including slope, aspect, rainfall received, soil and vegetation types, and extensive mathematic smoothing of elevation data to aid in better reproduction of a digital surface. Researchers in this area include Passalacqua et al. (2010).

Anderson and Ames (2011) have been researching the extraction of stream channel locations directly from LiDAR point cloud data. This practice would eliminate any post processing of that data which could save time and resources.

Zhou and Liu (2004) have looked at different algorithms that determine the creation of a DEM and analyzed the amount of error in them. They summarize by stating that the choice of algorithm made when used to determine physical characteristics is important when the data has a high amount of precision. Clearly, as detail increases the interpretation of that data must be done with a great amount of care and manipulation by an algorithm.

2.8 Methods for Evaluating Stream Delineation

Measurements of stream length and sinuosity are the most common way of determining a DEM's stream delineation ability. Several researchers have utilized these types of methods including Callow, Van Niel, and Boggs (2006) where they weighed the impact that DEM manipulations had on overall stream length and catchment area. Desmet and Govers (1996) evaluated algorithms by locating ephemeral streams and calculating overall stream length to determine upslope area. Pryde et al. (2007) investigated the abilities of SRTM and ASTER datasets to delineate overall watershed boundaries. In their study, watersheds were created using spatial analysis tools in ArcGIS. These tools included the hydrology toolset used for this thesis of "Fill", "Flow Direction", and "Flow Accumulation" followed by the "Watershed" tool. They utilized a Cartesian coordinate system and compared x values for explicit y values for the boundaries of the datasets to see how alike they were.

2.9 DEM Resolution Comparison Studies

There have been many studies of comparing DEMs with different resolutions. These have been performed in a variety of ways. However, many of the studies do not investigate the horizontal locations of stream channels in a quantifiable way. Many researchers simply compare accuracy of stream locations by visual inspection. Even Pryde et al. (2007) evaluated their watershed comparisons by visual inspection, but they did support their interpretations with measured values. The Callow, Van Niel, and Boggs (2006) study references known hydrology to evaluate how stream burning, or manufacturing channels in the DEM, would have on terrain analysis. It is not clear how the known hydrology for their study is determined.

The literature for comparison of DEMs with differing resolutions is fairly extensive with a great variety of source data being used. However, the methods used to determine how well a DEM performs in general, is also numerous. McMaster (2002) did say how very little had been done to establish a method for determining a threshold for the grid size of a DEM where it would no longer be appropriate for stream extraction. Since that statement was written, research investigating appropriate cell sizes for DEMs has been a current trend. Colson et al. (2006) felt that the use of LiDAR in the creation of DEMs, and the subsequent increase in point density, would result in higher resolution, as well as higher accuracy of stream networks. But, as Mark (1988) points out, for certain applications, this may be unnecessary.

Jancso and Melykuti (2011) studied DTMs that had been interpolated from contour lines with grid spacings of 5 and 50 meters, although they were not investigating stream extraction techniques. Rather, they were investigating the overall height differences of the models.

2.10 Conflicting Results in Different Terrain

There are others that have focused their hydrological DEM research on grid resolution and have come to differing conclusions about whether or not a decrease in cell size produces better stream networks. In these comparison studies that have shown differing results, the datasets, as well as the study areas and topography in particular, are different. Mouton (2005) comes to the conclusion that as DEM resolution is increased, the ability to accurately detect stream channels and the horizontal positions of stream heads, or starting positions, is increased. He makes the connection that this information then decreases the need for more sophisticated models obtained from some of the advanced algorithms previously mentioned.

McMaster's (2002) study showed that after a threshold value of approximately 180 meters was obtained for grid spacing, the accuracy of stream prediction declined. The study area for McMaster's project was in mountainous terrain and the average drop from cell to cell was high which should help to minimize the impact of a DEM's pits on stream channel prediction, as Mark (1988) mentioned.

2.11 Lack of Evidence Supporting DEM use for Feature Extraction

While many researchers have utilized DEMs created from high point density LiDAR data to study a variety of surface feature extraction, including stream networks, there has been little done to quantifiably prove that it is appropriate to be using these DEMs in this way (Rayburg, 2009). Hosseinzadeh (2011) agrees by stating that focusing on the accuracy of the data may let someone come to an incorrect conclusion that the highest resolution data would be the most appropriate to use in certain situations.

Recently, Passalacqua et al. (2010) challenged the results of O'Callaghan and Mark (1984), Band (1986), Mark (1988), Tarboton et al. (1989, 1991), and McMaster

(2002) stating that there was no empirical evidence from the field to support their findings regarding stream delineation. However, in McMaster's study (2002) there were site visits mentioned and at those locations it was likely that a stream existed where predicted, which helped to validate the findings.

There has been research performed to find the locational positions of stream heads, or concave areas in DEMs by Tribe (1991, 1992), and Colson et al. (2006). Heine, Lant and Sengupta (2004) also supported their findings with field work. Much of the current effort in stream location research focused on determining the exact locations where a first order stream begins. Lacking in the literature is the attempt to determine the horizontal position of channels downstream from those initiation points.

2.12 Summary

The literature regarding Digital Elevation Models, terrain analysis and hydrological research is vast. This thesis is not concerned with mathematical algorithms and the following DEM creation, but focuses on a DEM's ability to extract a stream channel's location accurately. This thesis offers a way for researchers to quantifiably state the performance of a raster surface or DEM algorithm to accurately reproduce a stream channel network. This type of work is missing from hydrological research and other surface feature extraction studies.

METHODOLOGY

3.1 Introduction

This chapter relates the procedure used to delineate stream channels from the datasets. The procedure for sample collection will also be explained so that the study can be repeated. To avoid any bias, the DEMs were not preprocessed in any way by the user. However, the NED is constantly being manipulated and updated by people prior to being made available to the public. Ultimately a comparison was performed of DEMs that have received some human assistance in their creation were combined and compared to datasets that have had no such influence. Once in ArcGIS, all DEMs were processed in the same way using the ArcGIS hydrology tools. Figure 3.1 is a summary of the steps used.

Distances from the center of the traced reference stream to the derived stream channels of the various datasets were calculated using transect lines that were cast from manually drawn baselines. The baselines followed the general flow direction of all of the derived streams using the ArcGIS extension, DSAS. Distances were obtained by collecting the points where a transect intersected a derived stream channel line. The following sections in Chapter 3 detail the method used during this study to obtain these distance measurements.

The aerial photographs used for tracing of the stream channels had a resolution of six centimeters per cell. These effectively served as ground truth for this study and all datasets were compared relative to the traced streams obtained from them.

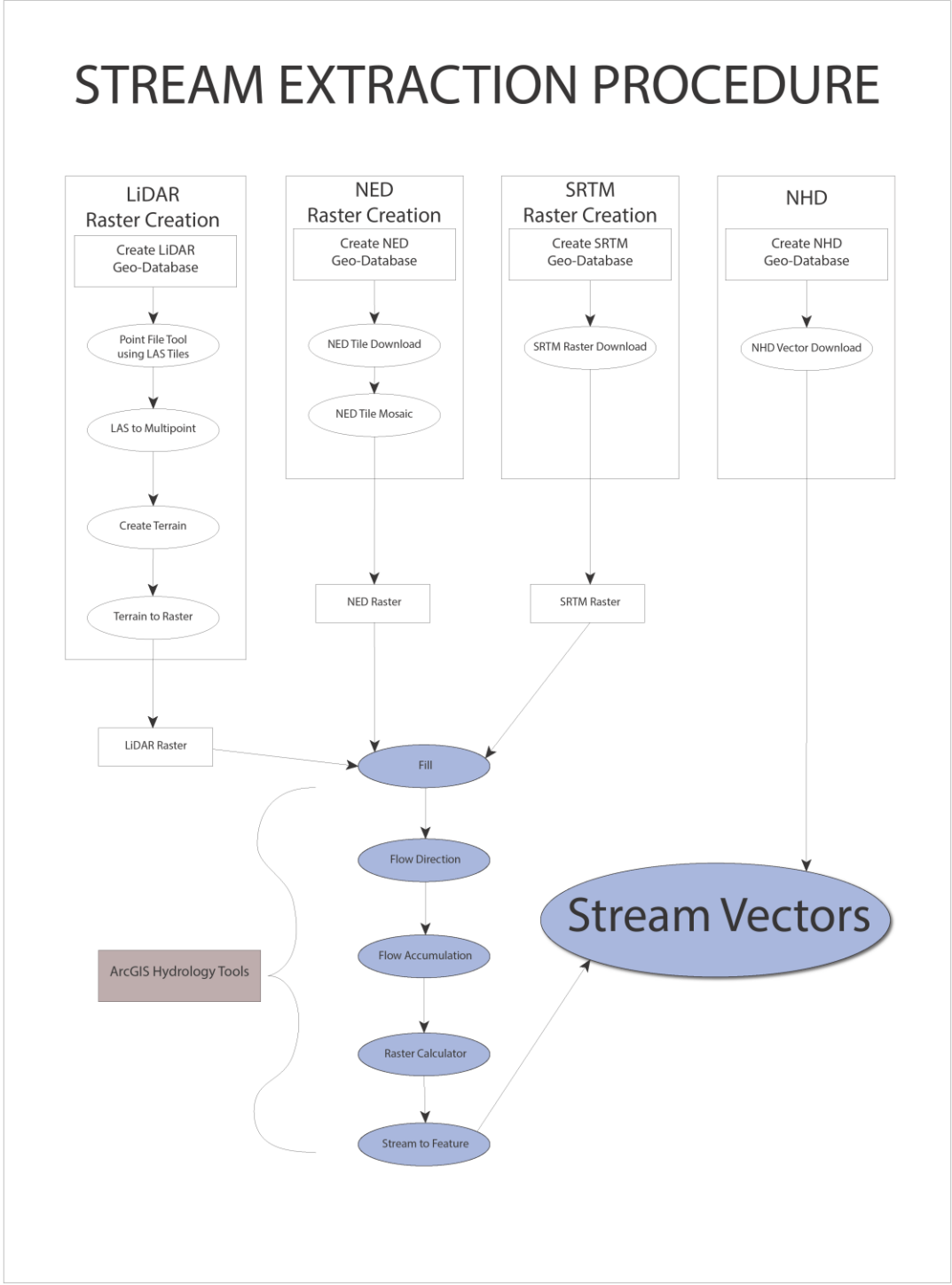


Figure 3.1. Flowchart showing the method used to extract stream channels from the various datasets.

3.2 Datasets Used

The National Elevation Dataset (NED) and the Shuttle Radar Topography Mission (SRTM) have resolutions of 30 meters along each side of a raster cell and 90 meters per cell, respectively. The DEM created from the raw LAS files was assigned a cell size of 4.34 meters in keeping with the average point spacing obtained from the Point File tool used to evaluate the raw data within ArcGIS. A fourth dataset, the National Hydrography Dataset, containing vector stream data, was also compared. The orthophotographs used for stream tracing were collected at the same time as the LiDAR data.

ArcGIS hydrology geoprocessing tools required a raster surface in order to extract the stream locations. The LiDAR data was obtained in raw LAS format, so a raster had to be created within ArcGIS using extra geoprocessing tools.

3.3 Naming Conventions and Geodatabase Creation

To avoid confusion in while performing the various conversion tools in ArcGIS, folders were created for each dataset examined in this study. This was also done to help prevent the accidental overwriting of data, and quick referencing if needed. Checklists were made for each dataset and each stream length to avoid repetition.

Larger File geodatabases proved to be necessary for processing the LiDAR data and to contain the large amount of data created from each geoprocessing tool. Personal geodatabases were utilized for the tracing of individual streams and the creation of distance measurements. The personal geodatabase is also a prerequisite of DSAS that was used to obtain the sample data distance measurements.

3.4 Obtaining NED and SRTM Data

The NED DEM of the study area was obtained from <http://viewer.nationalmap.gov/viewer/>. The National Elevation Dataset has resolutions of 1/9 arc-second (~3 meters per cell), 1/3 arc-second (~10 meters per cell), 1 arc-second (~30 meters), and 2 arc-seconds (~3 meters). Since only the 1 arc-second data covers the entire study area, it was chosen.

The data was supplied via email attachment as a compressed folder. These files were moved to the appropriate NED folders and uncompressed. Using ArcMap, the files were inspected in the map view. Depending on the bounding box one draws, it is possible to receive extra raster tiles that are outside the area of interest. The two tiles that were received had names of grdn42w096_1 and 1grdn42w097_1. These were subsequently merged together using a mosaic operation in ArcGIS.

The SRTM data for the study area was obtained from <http://gdex.cr.usgs.gov/gdex/>. The SRTM 90m radio button was selected to view the coverage and availability on the website's map. The general outline of the study area was then enclosed by a manually drawn polygon. The polygon selection method acts as a clipping tool for the data and eliminates the unnecessary areas outside of the study area. This step greatly speeds processing time because it eliminates unneeded data.

One must be logged in with a Reverb/ECHO account in order to download data. Once logged in, the download button can be selected followed by the selection of the dataset one is interested in using. GeoTIFF files were selected for download in a lat/lon format rather than a projection of either UTM zone 15 or 16. The file was then stored in the appropriately named folder for ease of identification when running the GIS hydrology tools.

3.5 Raster Creation from LiDAR Data: ArcGIS Tools Used

The LiDAR data used for this study was provided in raw LAS format. In order to extract stream locations from the raw LiDAR LAS files, they had to be converted to a raster for the hydrology tools in ArcGIS. The ArcGIS tools used for this method were: Point File, LAS to Multipoint, Create Terrain, and Terrain to Raster.

3.6 LiDAR Storage Location and Settings

To make sure all of the LiDAR processing data was contained in a single location, a feature dataset was created within a file geodatabase to hold this data. The feature dataset was assigned a coordinate system of NAD 1983 State Plane Nebraska FIPS 2600 (Meters) through the “New Feature Dataset” creation dialog window. No vertical coordinate system was chosen for the z coordinates in the feature dataset. The default values of 0.001 were applied for the XY Tolerance, Z Tolerance, and M Tolerance. The name LiDAR_Feature_Dataset was given to it in order to aid in identification.

3.7 Average Point Spacing Determination

To determine the point spacing for the LAS to Multipoint tool, it was necessary to run the Point File tool in ArcGIS. The output feature class name was set as Point_File within the LiDAR feature dataset. This tool calculated the average point spacing to be 4.34 meters.

All of the LAS files for the study area were then added using the LAS to Multipoint 3D analyst tool. The LiDAR_Feature_Dataset was selected as a location with LAsToMultipoint assigned as a name, along with the 4.34 point spacing. Only the bare earth returns, code 2, were chosen for this tool.

3.8 Initial Surface Creation

A new Terrain was created within the LiDAR_Feature_Dataset by right-clicking on it and selecting new - Terrain. The feature class, LAStoMultipoint, was chosen to participate in the terrain with 4.34 set as the approximate point spacing. Window size was chosen as the pyramid type, and pyramids were calculated for display purposes. The values obtained were 8.5, 17, 34, 68, 136, and 272. No secondary thinning was applied.

3.9 Raster Creation

The Terrain to Raster tool created the raster surface from the LiDAR data. The Terrain to Raster tool was ran with FLOAT selected as the output data type along with a LINEAR method. The sampling distance was set to CELLSIZE, and the value was amended to 4.34 meters. Pyramid Level Resolution was kept as zero.

3.10 Derivation of Stream Channels

The ArcGIS hydrology tools used in this study for the initial extraction of stream channels were Fill, Flow Direction, and Flow Accumulation. The Raster Calculator tool was then ran to limit the number of derived stream channels output by the Flow Accumulation tool for display. The Stream to Feature tool was ran to create the final vector lines that represented the centers of the derived streams.

3.11 ArcGIS Hydrology Tools and Feature Extraction

3.11.1 Fill Tool

The Fill tool inspects all of the cells surrounding each cell and evaluates their elevations. If a cell is surrounded by elevation values that are all higher than that cell, a pit, any flow would cease at that cell. By altering the elevation of that cell to the next highest value of the surrounding cells, any water that would accumulate in that cell would then be allowed to flow out.

3.11.2 Flow Direction Tool

The Flow Direction assigns a direction based on the eight cardinal directions from which water could flow out of a cell. The Flow Direction tool makes use of the D8 method according to the ESRI ArcGIS online help pages (<http://resources.arcgis.com>). Using the heights of one cell and comparing that value to the eight cells that surround it, water is assigned to flow in the direction of the cell with the lowest value. This creates stream channels in raster format from the DEM. They are drawn as filled in cells, and are connected either by their sides or at the corners of the cells. The cell side dimensions are the same as the originally created raster: NED 30 meters, SRTM 90 meters, and LiDAR 4.34 meters.

3.11.3 Flow Accumulation Tool

While the Flow Direction tool assigns flow of water from any one cell to another, the Flow Accumulation tool back calculates the number of upstream cells that have merged to form the stream flow lines. By addition, a rough network of flow can be achieved by piecing these cells together. Thus, two cells flowing into the same cell would give that cell a value of two. Figure 3.2 shows the addition mechanism for flow accumulation.

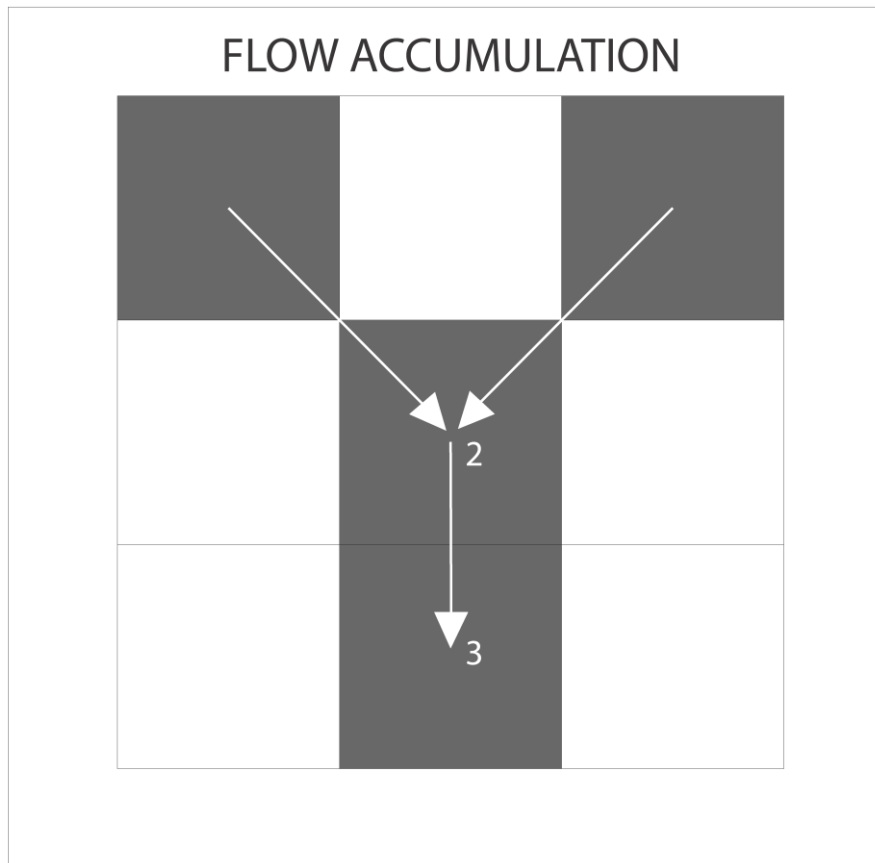


Figure 3.2. Flow accumulation diagram showing the addition of upstream cells flowing into subsequent cells.

While the Flow Accumulation tool ultimately creates a stream network, it can be difficult to locate and pick out the exact location of the derived streams. There may simply be too many flow lines on the map to pick the appropriate stream.

3.11.4 Raster Calculator Tool

To remedy the problem of too many flow lines created from the Flow Accumulation tool, the Raster Calculator tool was utilized in order to limit the number of streams that are visible. This tool inspects the information from the Flow Accumulation tool. Through the use of a conditional statement, the user can set how many cells flowing together constitute a stream for display as a map. An example of the equation used to create the extracted streams as a series of cells is given here:

```
rastercalcLiDAR=Con("Accum_Dir_FILL_mosaic_LiDAR">3000,1)
```

This conditional statement results in a nominal set of data where the cells are either assigned a stream flow, or not, based on a minimum accumulation count of 3000.

The Flow Accumulation raster is used as the input raster with all of the cells that have greater than 3000 cells flowing into them assigned a value of 1. This signifies that those cells should be kept while the remaining cells are not added to the stream network.

Some trial and error were needed to determine an appropriate value to assign to the conditional statement in order for the derived stream network to be easily recognizable, yet not too generalized which would lead to shorter stream lengths.

The NED and SRTM cell sizes are much larger than the LiDAR dataset used, and a value of 250 was used to limit the number of derived streams for them. Figure 3.2 shows the result of the Raster Calculator tool used on the LiDAR dataset with the conditional value of 3000 applied to it. The stream channels are defined in this map, but, upon zooming, the stream displays as a string of cells rather than a single line.

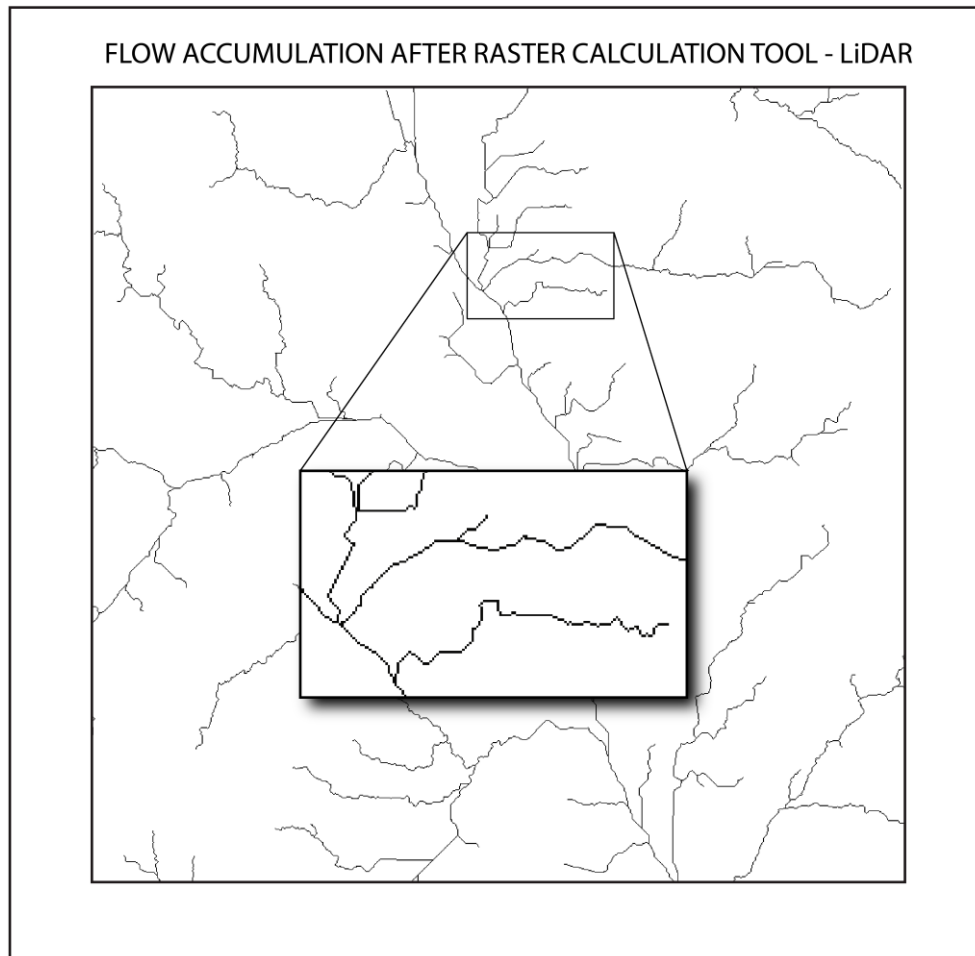


Figure 3.3. Representative sample of the LiDAR dataset after a conditional statement using a value of 3000 in the Raster Calculator tool. The stream network is represented by individual cells rather than vector lines at this point in the stream extraction process.

3.11.5 Stream to Feature Tool

The final hydrology tool used for stream channel extraction was the Stream to Feature tool. Rather than creating another raster with individual cells, this tool creates a vector line joining the centers of each connected cell as water would flow into them. Figure 3.4 shows the same portion of the watershed as Figure 3.3 after running the Stream to Feature generalization tool.

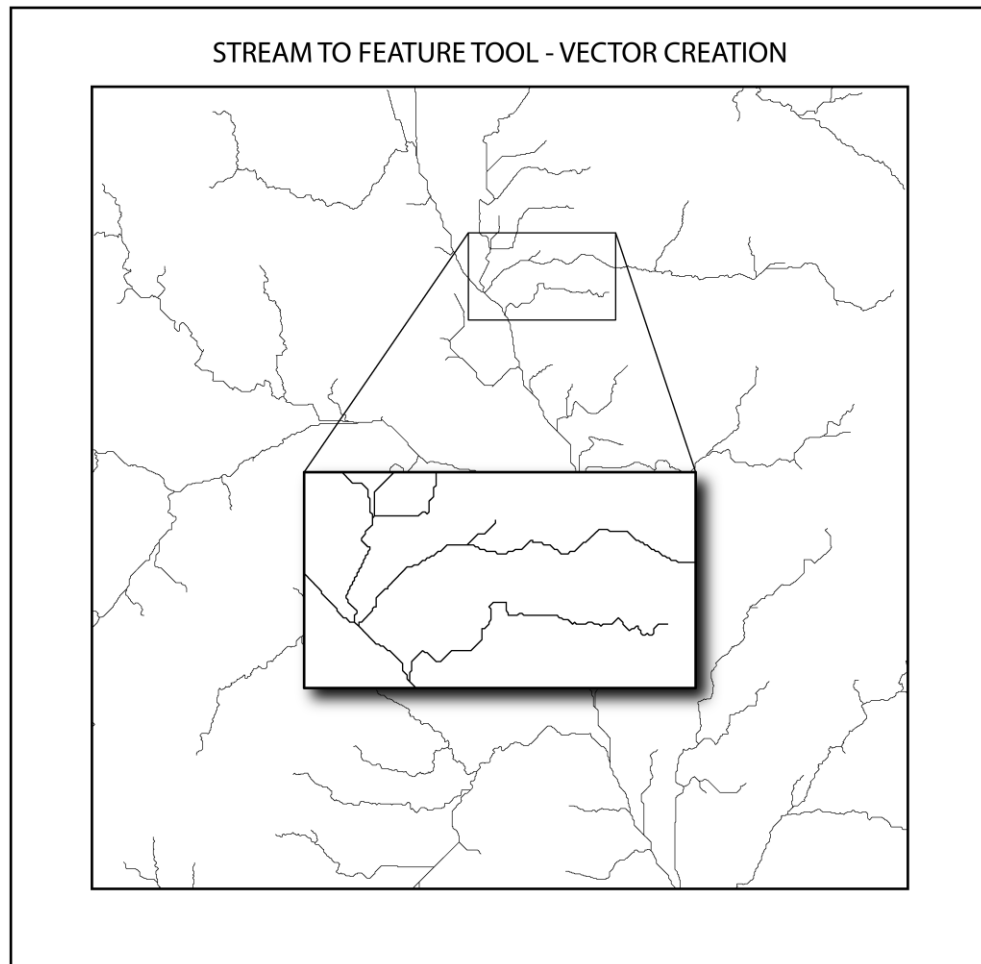


Figure 3.4. Representative sample of the LiDAR dataset after a conditional statement using a value of 3000 in the Raster Calculator tool followed by using the Stream to Feature tool to create individual vector lines from the cells of the raster.

This vector Feature Class of vector lines representing the delineated stream location was stored in the Feature Dataset of the appropriately named geodatabase. It was named fittingly such as LiDAR_Stream_to_Feature to aid in identification. These vectors are what were ultimately compared to the traced stream channels in this study.

3.12 Individual Stream Geodatabase Creation

Working in ArcMap, personal geodatabases were created for each named stream in the watershed. Even though personal geodatabases in ArcGIS are limited to only 2 GB

of data, they were used rather than file geodatabases. This is due to requirements of the DSAS extension when it is used to compile the measurement data.

The geodatabases were named according to what stream's data would be housed in it. A feature dataset was created of the traced streams and the vectors created by the Stream to Feature tool for each dataset compared in this study. The feature dataset was georeferenced to the North American Datum of 1983, Nebraska State Plane coordinate system.

Once the initial geodatabases and feature datasets were created, they were simply copied and then renamed with the appropriate stream names. i.e., "Copper Creek Geodatabase," followed by "Copper Creek Feature Dataset." This naming convention was useful throughout the project for finding and editing the appropriate files as needed for work to progress.

No vertical coordinate system was chosen for the z coordinates in these Feature Datasets. The default values of 0.001, 0.001, and 0.001 were applied to the Feature Datasets for the XY Tolerance, Z Tolerance, and M Tolerance, respectively.

3.13 Stream Tracing

New feature classes (polyline) were created with a similar naming convention used for the personal geodatabases and feature datasets. Only "North" or "South" was added to the names to differentiate them such as Copper_Creek_North, indicating that particular feature class was representative of the north stream bank. The terms north or south were arbitrary, and often the north and south sides of the stream banks are switched in the map view. This depended on the sinuosity of the stream. The calculation that ultimately determined the center of the actual stream channel along the transect is an average of the two distances from the baseline, so the order or the names did not matter.

The creation of the feature classes automatically adds them to the table of contents in the ArcGIS map window. This action (the first item added to the data frame) sets the properties for any additional items added to the map, such as the orthophotographs used for tracing, even though they are not stored within the feature dataset or the geodatabase.

3.14 Stream Tracing Quality Control Using NHD

To ensure the appropriately named stream was being manually traced for each stream length, a shapefile was downloaded and imported from the National Hydrography Dataset (NHD). It is available at <http://viewer.nationalmap.gov/viewer/>. The shapefile, named NHDFlowline.shp, is one of a suite of shapefiles available after uncompressing a selected area's downloaded file. The original NHD file was stored separately for use for each stream length in the following way.

The original NHDFlowline shapefile was added as a feature class to the feature dataset being traced. One must be in an edit session to perform this action. The shapefile was then dragged into the map view for display. The attribute table was opened and sorted according to the column GNIS_ID.

The appropriate rows, or stream names, were highlighted corresponding to the stream being worked on. The "Switch Selection" button was then clicked which resulted in all other streams being highlighted. These other records were simply deleted from that particular feature class leaving only the stream desired. The edit session was closed and the edits were saved. The original shapefile containing all of the streams was left intact for use in the other stream geodatabases and the process was duplicated for those streams.

The NHD Feature Class in each stream's geodatabase was named NHD. Since each NHD feature class was stored within the appropriately named personal geodatabase

and feature dataset, it was clear which stream it represented. The closing of the edit session and quick renaming was done to avoid confusion in the event of any interruptions while tracing the streams.

To create the north or south stream line feature classes, edit sessions were started by selecting the “Create Features” button on the editor toolbar in ArcGIS then selecting the appropriate polyline followed by “Line” in the construction tools menu. For this study, “Line” was used extensively. Once all of the geodatabases were created with the appropriately named feature datasets, NHD shapefiles, and north and south line empty feature classes housed in them, tracing of the streams could begin.

3.15 Stream Tracing Generalization

During the course of the study, it was realized that the initially traced stream bank positions could be generalized then edited later to be more precise at the exact locations where the transects were intersecting them. Rather than tracing the stream channels precisely, as in Figure 3.5, the stream banks were traced widely and quickly on either side of the stream channel while generally following the route of the NHD dataset. Selecting the appropriate stream length from the original NHD shapefile and housing it in the appropriate geodatabase served as a guide to ensure that the appropriate channel was being traced.

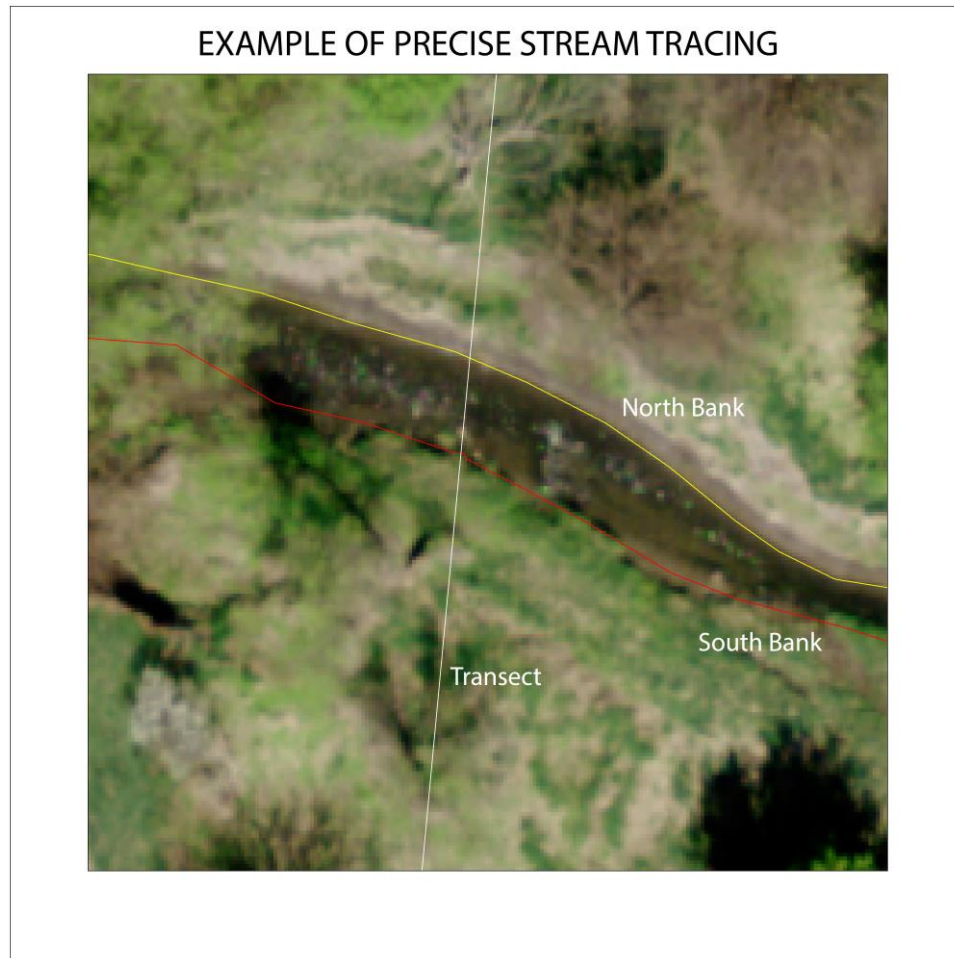


Figure 3.5. Example of tracing the north (yellow line) and south (red line) stream banks in a precise manner. Note the area at the left of the image where vegetation obscured the stream channel. In that area the stream was traced in a generalized manner.

After the north and south stream banks were traced in this generalized manner, the transects were cast using the DSAS extension. The final adjustment of the north and south traced stream banks required knowing where the transects would intersect, so the stream banks could be quickly traced in a generalized way, followed by fine tuning the positions once the transects were cast from the baseline. In other words, the traced stream was drawn in a wide fashion and then later “pinched” to their precise locations at the intersection points. Figure 3.6 shows the generalized tracing technique and the pinching

of the traced banks where the transect crosses the stream. Figure 3.6 also shows how a vegetation obscured stream channel was not edited where the transect crossed it.

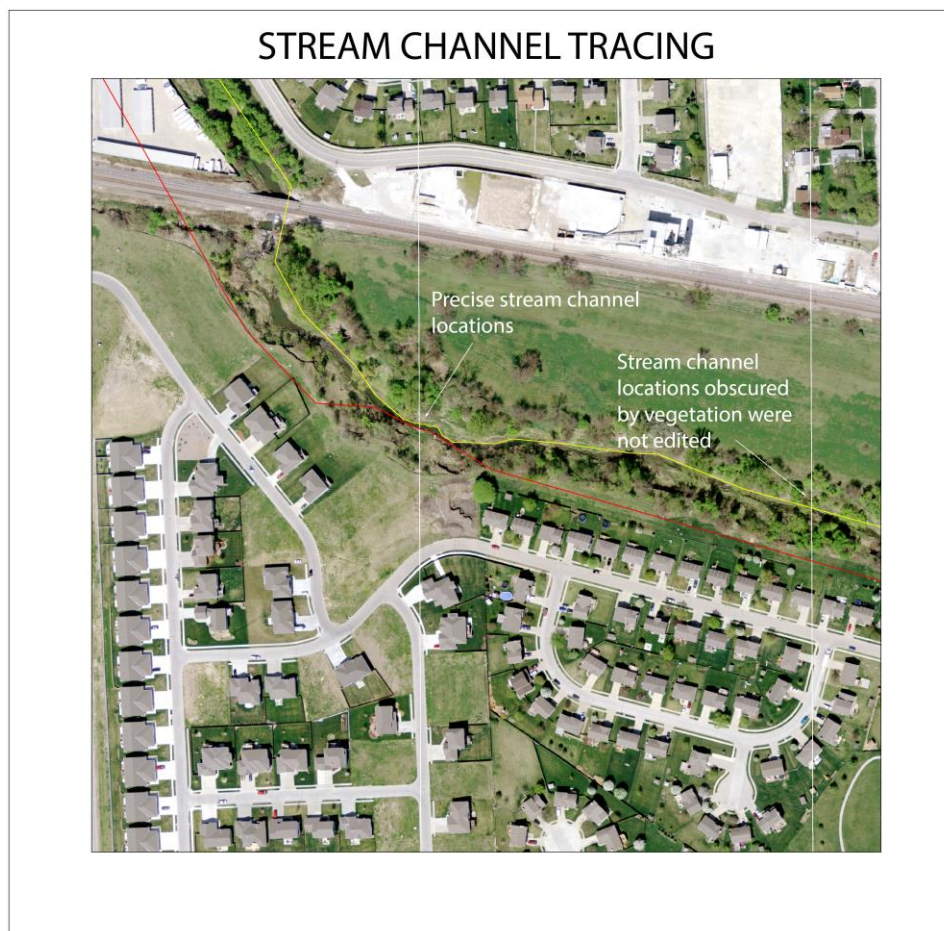


Figure 3.6. Generalized stream tracing followed by editing the north (yellow line) and south (red line) stream banks to a more precise final location prior to collecting the intersection sample distances. Note how the obscured stream channel locations were not edited in order save time.

This fine tuning of the position of the traced stream bank was necessary after the transects were cast. At that point in the creation of traced stream locations there was no way to know exactly where a transect line would intersect the stream, so generalizing was beneficial. This realization greatly sped up the tracing process. While fine tuning the position of the traced stream banks, certain transects that crossed the traced stream locations that were obscured due to vegetation, roadways, or shadow were documented for later removal from the distance data. They were not included in the final results.

It is possible to simply delete individual transects prior to recording the distance measurements. However, it was felt that if newer photographs were to become available, then the original baselines and transects created for the DSAS tool to run could be used for further analysis after tracing new stream locations.

3.16 Baselines

DSAS requires that baselines be created. From these baselines the transects are cast at 90 degree angles at user-defined spacing in user-defined lengths. The user must display all of the layers in the stream geodatabase and create a baseline that tends to follow the general direction of all of the streams traced, as well as all of the derived stream channels.

The baselines were drawn in such a way as to hug the nearest stream as close as possible. One must realize that any major changes in direction of the stream due to sinuosity would need another baseline drawn, or several, in order to better represent the general direction of the stream channel. A mature meandering stream would require many baselines to be drawn. In the case of the Papillion Creek Watershed, many of the stream channels have been urbanized so their general direction is rather straight, but on a few occasions several baselines were needed in order to follow the streams better. Figure 3.7 shows a fairly straight stream length with all layers displayed and a typical baseline that was drawn for them.

The baseline feature class was created within the feature dataset, and is drawn using the line option for this study. Again, these feature classes are named appropriately for each stream, such as, West_Papillion_Baseline. This naming step helped during the selection of the appropriate lines when the distance measurements were ultimately calculated using the DSAS extension.

3.17 Casting Transects

The casting of transects requires some manual adjustment on the part of the user to ensure all of the traced streams and derived streams will be intersected by the largest number of transects, and that the transects cross the streams as close to perpendicular as possible.

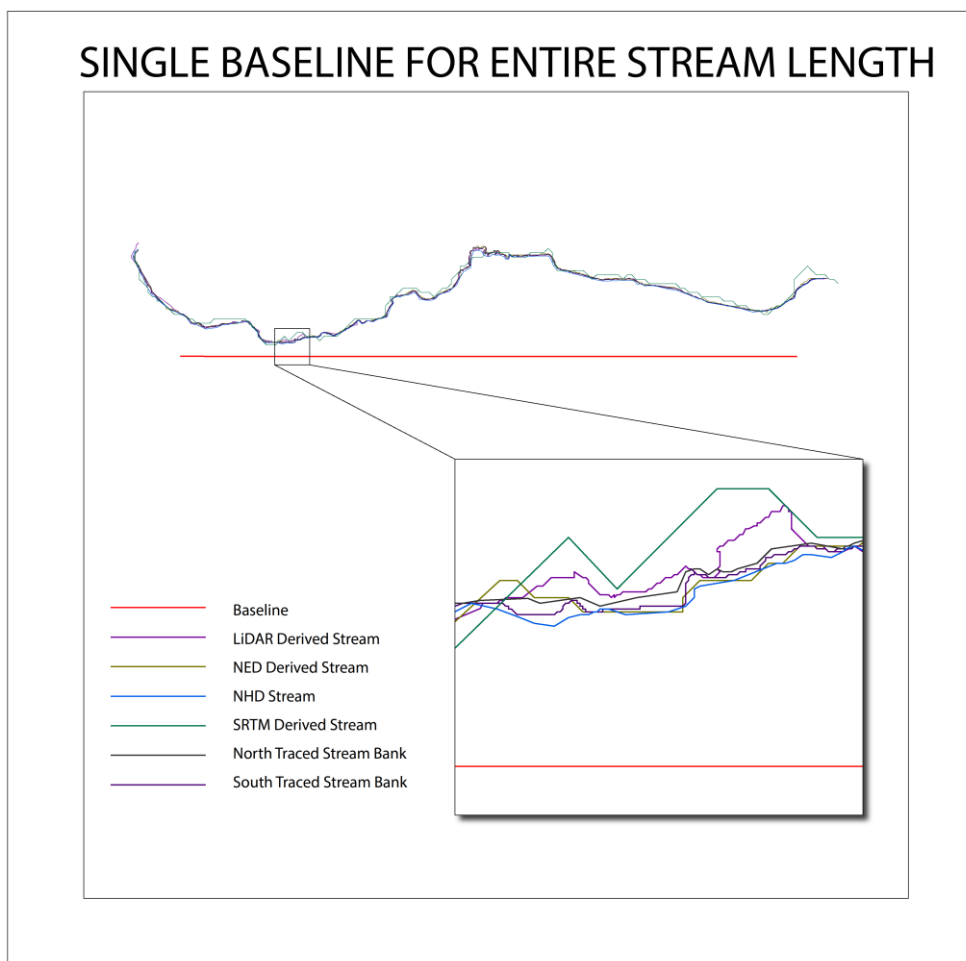


Figure 3.7 Single baseline used for the entire stream length following the general direction of the stream.

Transects are cast perpendicular from the baseline by the DSAS extension, so curved baselines, although possible, were not feasible for this study. Instead, several smaller straight baselines were drawn in order to follow the more sinuous sections of the streams of the study area. Figure 3.8 illustrates a section of stream length that is relatively

sinuous with several baselines and the orthogonal transects cast from them to intersect with the streams.

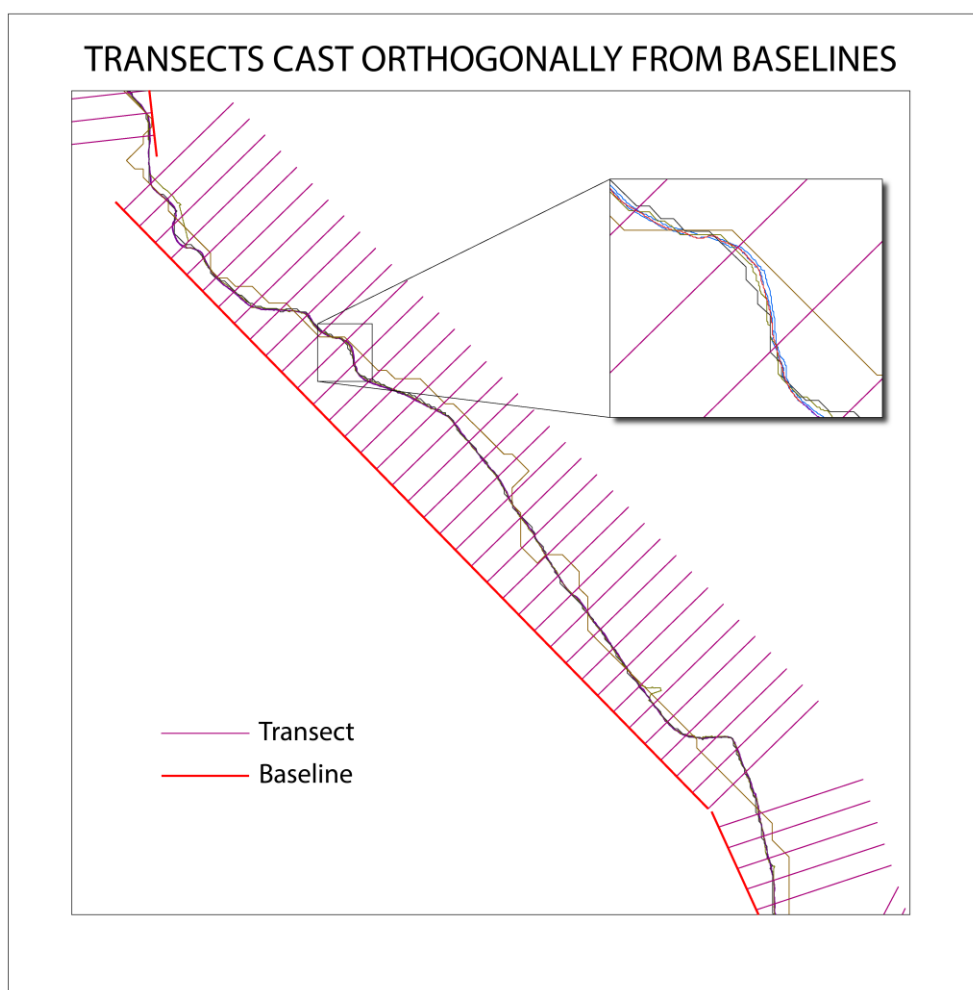


Figure 3.8. Transects cast orthogonally from a baseline intersecting with all traced and derived stream channels.

Transect lengths were determined by measurement and trial and error so that all of the derived stream centerlines, and both the north and south stream banks, would be intersected by the transects for the full length of each stream being worked on. The transects for each named stream section were evaluated independently for appropriate length. Some stream sections required a large transect length to encompass all of the derived stream centerlines, such as near the mouth of the Missouri River, while other

stream lengths required smaller transect lengths, such as along the middle of the Big Papillion Creek. The transect spacing distance for this study was set at 250 meters.

The DSAS extension automatically adds the appropriately named transect shapefile to the personal geodatabase. This shapefile is eventually used by the DSAS extension when calculating distances from the baseline to intersection points.

If the transect shapefile is edited in any way, such as extending transects in order to intersect a line that is far away, or adjusting the angle of the transect to obtain a better sample, the attributes are automatically updated. Subsequent calculations would be obtained based on those new transect positions. During this study, no transects were altered once they were cast. It was ensured that all stream lines were intersected upon initially casting them. Also, working on one stream geodatabase at a time prevented accidentally collecting an incorrect intersection distance. Great care was taken to insure the measurements ultimately collected by the DSAS extension accurately reflect distances to the appropriate streams.

Any distance data obtained from a transect where the stream channel in the photograph was obstructed and could not be traced, were eliminated from the final spreadsheet used to manipulate the data for analysis.

3.18 Fine Tuning Traced Stream Channels

Following the casting of transects, the traced stream bank positions, previously traced in a generalized way, were made more precise. This was accomplished by opening an editing session and simply moving the traced line into a more precise position on either side of the intersecting transect by adding a vertex to the line and dragging it to a more accurate position in ArcMap, or “pinching” the traced stream at the transect intersections. The derived stream locations were not altered in any way. In fact, these

layers were “turned off” for this pinching procedure. Again, Figure 3.7 shows a section of the watershed’s traced stream channel that was adjusted where a transect intersected with them as well as a portion of the traced stream channel that was not adjusted due to the intersection points being obstructed due to vegetation.

Again, this was done since the traced stream bank would need to be verified and made more precise following the initial trace. The original tracings were done in a generalized way to simply save time. Also, since the actual intersections of the transects and the traced stream locations would need to be inspected for obstructions, and possibly eliminated from the final results, this practice allowed for close inspection, verification, and documentation of the obstructed transects.

If, while making the final positions of the traced stream channels more precise, a transect was found to be obstructed, the traced stream channels were not adjusted, and the transect’s ID at that location was documented for later removal from the final set of data. These same transects, and the distances that were calculated by DSAS were not included in the final results.

3.19 Distance Data Collection Feature Requirements

Working on one stream at a time, the following feature classes were added to the table of contents within ArcMap: baseline, transects, NED derived centerline, SRTM derived centerline, LiDAR derived centerline, north traced line, and south traced line. These were the features needed in order for DSAS to calculate the various distances and record them properly.

3.20 DSAS Extension Toolset

In the DSAS toolbar in ArcMap, the Set/Edit default parameters button must be selected followed by the selection of the appropriate feature classes, or stream layers,

from the drop down menus. The baseline layer for the stream section one is working on must be selected from the “Cast Transect Settings” tab as shown in Figure 3.9.

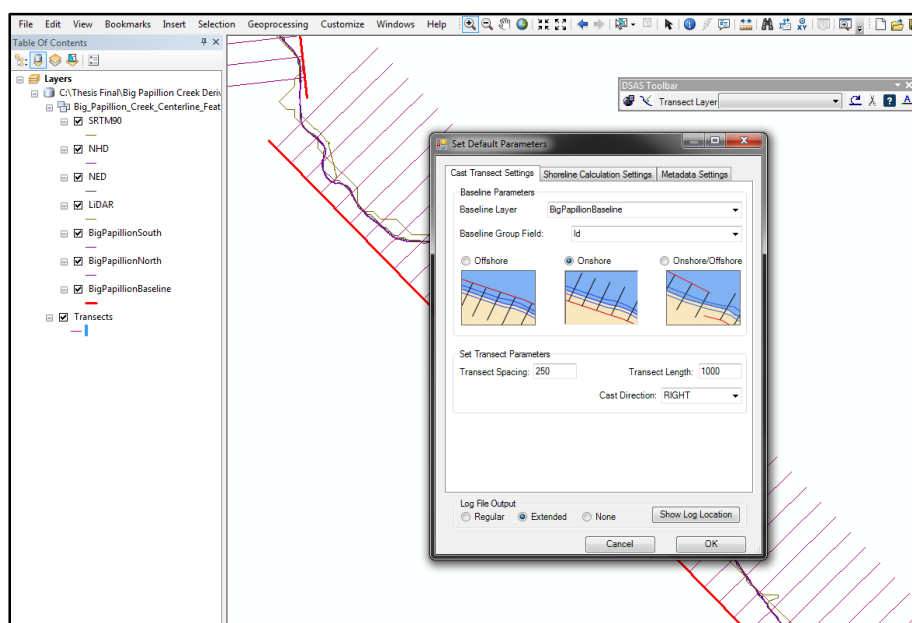


Figure 3.9. Cast Transect Settings tab of DSAS extension. Transect distances and spacing are set in this window along with the selection of the baseline.

The appropriate stream centerlines must be selected individually for the shoreline layer of the “Shoreline Calculation Settings” tab. The closest intersection radio button was selected at this point in the shoreline calculation settings tab, but other settings such as distance and spacing in the cast transect settings tab do not need to be changed at this point since the transects have already been cast. Figure 3.10 shows the options under the Shoreline Calculation Settings tab.

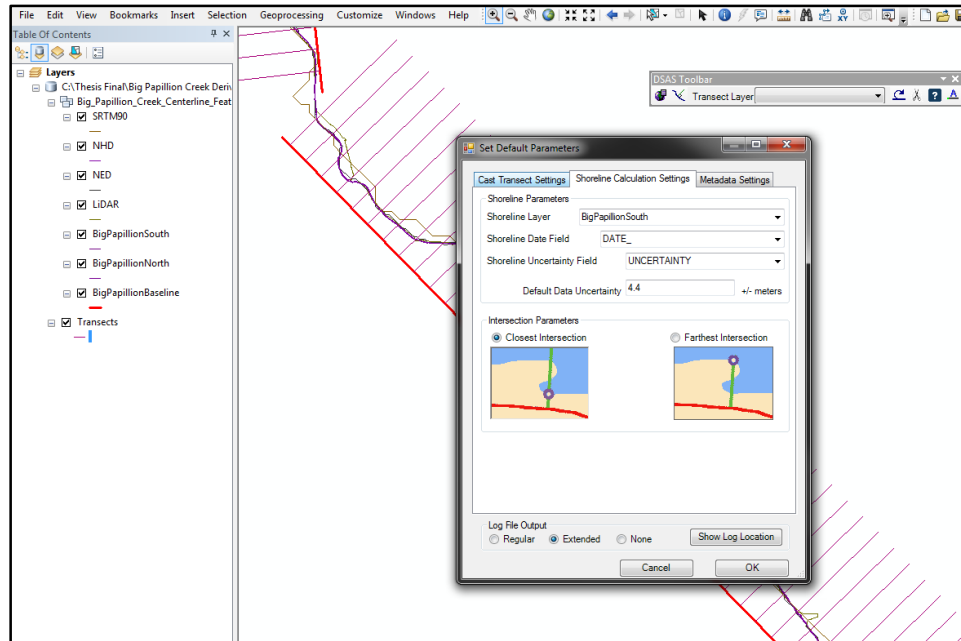


Figure 3.10. Shoreline Calculation Settings tab of DSAS extension. Individual shoreline, or stream layers are selected along with the type of intersection point, nearest or farthest.

Clicking the calculate statistics button of the DSAS toolbar opens a window which allows for the selection of the net shoreline movement checkbox. When the “Calculate” button is clicked, the extension begins calculating distances along each transect from the baseline to the intersection of the shoreline, or stream layer, that the user selected. Figure 3.11 shows the DSAS “Calculate Statistics” dialog box.

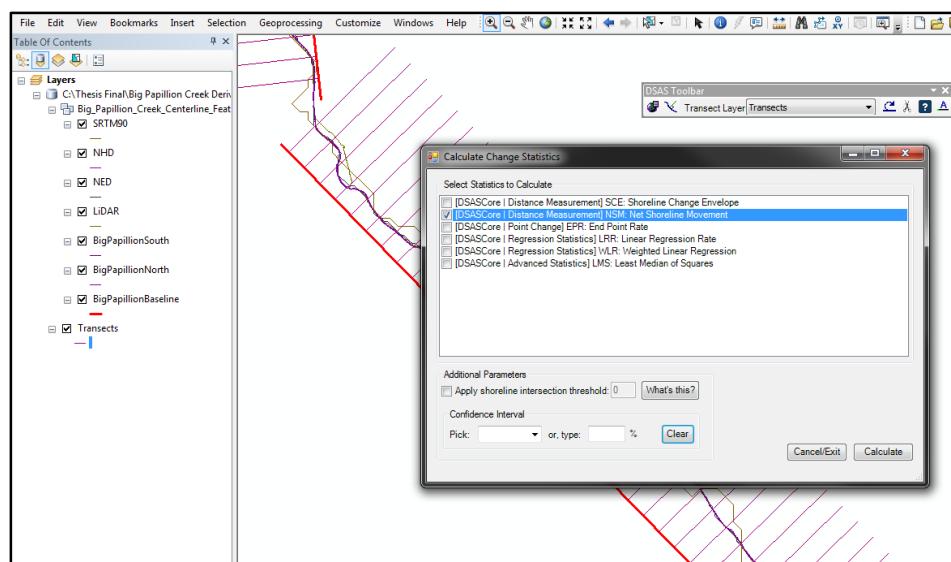


Figure 3.11. Options of the Calculate Change Statistics dialog window of the DSAS extension to ArcMap used to calculate distances of traced and derived stream channels from baselines.

Two tables are created within ArcGIS using this extension. These are named as a “Transects_rates” table and a “Transects_intersect” table. For this study, only the “Transects_intersect” table is needed. Future studies could utilize the rates table and determine how fast certain areas of the watershed are moving.

One must remember to change the shoreline layer from the DSAS Set/Edit default parameters button and the shoreline calculation settings tab after calculating distances. The removal of the layer following distance calculations, and the manual creation of a list of the stream layers with checkboxes, aided in preventing accidentally recalculating the same intersection distances repeatedly.

Another way to avoid accidentally recalculating distance measurements would be to only have one stream layer in the ArcMap table of contents at a time, but that practice would be time consuming, and may cause more opportunities for error by the user accidentally selecting an inappropriate layer when adding it back into the table of contents.

For this study, after each layer's distances were calculated they were removed from the map so they would not be accidentally recalculated. The workflow for distance data collection included creation of a list of the stream layers in the geodatabase. Once a stream layer had been selected in the "Set Default Parameters" dialog window, it was manually crossed off the list. Then, once the distance data had been produced by the DSAS extension, it was manually crossed off the list for that step. This further helped to prevent the accidental recalculation of a stream's distance data by allowing one to always know where they are at in the distance collection process in case of interruption.

The personal geodatabase in which the distance data was stored was refreshed following the calculation. This refresh step allowed for the table containing the sample distances to be renamed from a generic title to a more appropriate one. Since these tables were stored in the individual stream geodatabases it was found that this naming convention helped to clarify which stream the distance data corresponded to.

3.21 Data Collection Quality Control

Each dataset's distance measurements were calculated in this manner for each stream of the study area. These distance tables were opened and their intersection distances at several transects were compared to measurements using the ArcMap measuring tool. The first and last transects were measured from the baseline to the intersection point and verified to the table for quality control purposes.

These tables contained transect identification numbers in rows and intersection distances for them in columns. These were the distances were used for the data analysis.

Due to the process of manual creation of the baselines used to cast transects, the transect identification numbers are often not in perfect numerical order. Transect number one is located where the baseline originates and the identification number progress

numerically along it. For most of the stream lengths multiple baselines were used. Subsequently, the numerical progression is altered when a second or third baseline was placed on the opposite side of all of the streams and drawn backwards in relation to the others because of the limitations of the DSAS extension when casting transects.

Throughout the analysis, the original transect identification numbers were kept intact, along with their original distance measurements, in the event that a distance needed to be verified, or to recheck for any obstructions. Additionally, this was done to be able to ultimately line up the appropriate transect measurements for each of the four datasets being compared in the final spreadsheet. This made editing easier when eliminating transect measurements due to obstructions of the stream banks in the photographs. An obstructed transect measurement could be eliminated for all datasets at once by simply deleting that corresponding row in the final data manipulation spreadsheet.

3.22 Final Distance Calculations and Data Manipulation

The stream centerlines extracted for the three Digital Elevation Models (DEM) and the National Hydrography Dataset were tested for horizontal accuracy. Ultimately, the derived stream channel lines of the DEMs, and the available vector lines of the NHD, were compared to the location of a calculated center of the traced stream channel along a transect that crossed the traced stream banks of all of the named streams within the study area.

For ease of distance data manipulation, the attribute table data for all of the stream distance measurements for all of the streams calculated by DSAS were copied into a spreadsheet and kept for backup in the event of data loss. A copy of this spreadsheet was made and was used for manipulating the data into a format usable by SAS. The original

file containing all of the original distance values, including those transect distances that would eventually be deleted due to obstructions, was always kept as a backup, and for verification purposes.

3.23 Spreadsheet Setup

Using this data manipulation spreadsheet, the extra columns created by the DSAS extension were eliminated. These included the columns titled “object identifier”, “BaselineId”, “ShorelineId”, “IntersectX”, and “IntersectY.” The only columns remaining in the data manipulation workbook were “TransectId” and “Distance”, but there were 6 sets of this which represented the distances to the traced north and south stream banks, as well as the distances to the streams derived from the three DEMs and the NHD.

This action condensed the final number of samples to 535. The spreadsheet was further condensed by deleting the headers for each of the copied streams. This left only one row that served as a header and two columns named “TransectId” and “Distance.”

3.24 Traced Stream Width Calculation

The center of the traced stream channel was calculated using a spreadsheet. The distances measured by the DSAS extension along the transect line from the baseline to the intersection points of the north and south traced stream banks, D_N and D_S respectively, were added. This value was then divided two. The distance from the baselines to the centerlines of the traced stream channels, D_C , is given in the equation below.

$$D_C = (D_N + D_S) / 2$$

This stream width calculation was needed to determine the distance of the derived streams to the traced stream, and it was eventually used to determine if a derived stream fell within the banks at that location.

Again, the distances from the manually drawn baseline to the derived stream centerlines for each of the datasets were originally obtained from the DSAS extension. These measurements are represented here as D_{NEDbase} , D_{NHDbase} , $D_{\text{LiDARbase}}$, and D_{SRTMbase} (measurement from the baseline). The distance from the calculated center, D_C , of the traced stream location to the centerlines of the derived streams is given for the datasets NED, LiDAR, and SRTM90 as D_{NED} , D_{LiDAR} , and D_{SRTM} , respectively. These measurements were calculated as follows within the spreadsheet data manipulation file:

$$D_{\text{NED}} = \text{ABS}(D_{\text{NEDbase}} - D_C)$$

$$D_{\text{NHD}} = \text{ABS}(D_{\text{NHDbase}} - D_C)$$

$$D_{\text{LiDAR}} = \text{ABS}(D_{\text{LiDARbase}} - D_C)$$

$$D_{\text{SRTM}} = \text{ABS}(D_{\text{SRTMbase}} - D_C)$$

The center point determination of the traced stream channel is represented in Figure 3.12.

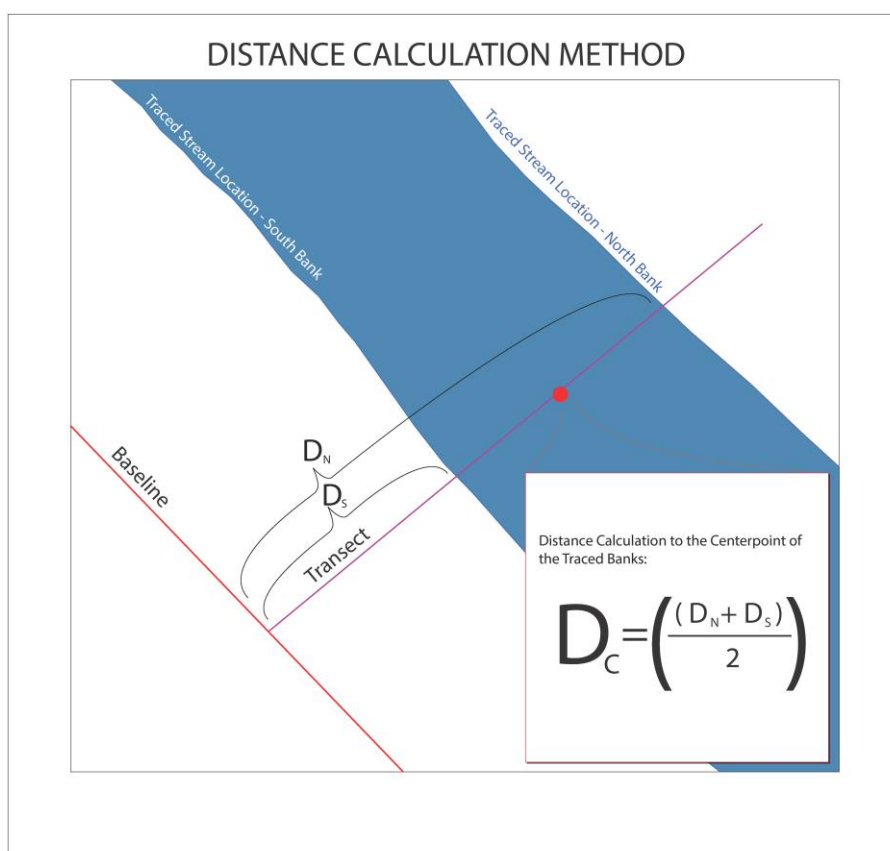


Figure 3.12. Traced stream channel center determination. Ultimately, if a derived stream intersection distance is smaller than the distance to the center, plus half of the stream's width, then it was deemed to be within the traced stream and recorded as an accurately predicted sample. The process is explained in Chapter 4.

3.25 Summary

The procedures for obtaining NED and SRTM elevation data, as well as the stream locations for the NHD, are straightforward. While LiDAR data is not readily available for all portions of the United States as is the case for the other datasets compared in this study, its use is growing. There were more steps needed within ArcGIS to convert the raw LAS files to a raster, and despite the large amount of data that required a large amount of processing time, using the outline available from ESRI to create a LiDAR DEM from LAS files did proceed smoothly. The creation of the NED and SRTM DEMs proceeded quickly and the process to extract the stream locations using the

hydrology tools in ArcMap was direct. Only the LiDAR dataset required large amounts of time to perform the stream extraction.

Tracing the stream channels was tedious, and consumed the most amount of time during this study. This was part of the reason the NHD was included in the analysis. Once the baselines for the DSAS tool were drawn and transects were cast, the precise work of drawing the stream at the locations of each viable transect where the stream was not obscured, was laborious.

The only real drawback during the tracing process, besides the anticipated time needed to actually trace the channels, was the necessary creation of the individual geodatabases for each named stream length. DSAS uses personal geodatabases within ArcMap in order to store the distance values, but file size of a personal geodatabase is limited to 2 GB, so each stream length was traced individually rather than doing the entire watershed at once.

It would be difficult to know which transect to remove from the data collection in the event of an obscured stream channel if the entire watershed was done at the same time. It could be done, but the numeric progression of the transects did prove to be useful at times for quality control purposes when the manual use of the ArcMap measuring tool was employed to insure that the DSAS tool was measuring to the correct extracted stream channel.

The DSAS tool's ability to capture data automatically did prove to be extremely useful. Ultimately, when combined with the use of the DSAS tool during data collection, the ability of the GIS to export distance tables in a spreadsheet was valuable and saved time. The work proceeded well although large amounts of storage were needed for the photographs and LiDAR data. In order for work to continue, that data was stored in

multiple locations and new, temporary ArcMap .mxd files were created while editing. Throughout the study, the entire work was copied many times with version numbers to aid in recovery in the event of catastrophic failure of storage devices.

Ultimately, the method used to produce distance values proceeded smoothly and can be easily repeated for other datasets or stream extraction algorithms.

RESULTS

4.1 Introduction

Using the locations of each DEM's derived stream locations and comparing them to the nearest bank of the traced stream channel, sample distances were produced. Using these distances, this method was able to determine how often a derived stream location was accurately predicted. If a sample was not accurately predicted, the amount of discrepancy for each sample and each dataset could be compared. The goal was to determine if the amount of error for each dataset could be used to determine if they produced stream channel locations in the same way.

4.2 Rationale

It was felt that if a derived stream channel was sampled to be within the traced stream's banks it would be considered to be an accurate placement due to the continuous migration and dynamics of the stream. The total number of correct samples could be calculated and a higher number of accurately placed stream locations would be an indicator of a dataset's ability to produce the overall stream pattern.

If a sample did not fall within the traced stream channel banks, then further calculations would be performed for that sample to quantify the distance from either of the banks. It is this distance from the traced stream channel banks that was considered error rather than the distance from the center of the traced stream to the center of the derived stream. However, the center point of the traced stream did need to be determined in order to conclude if a stream was inside or outside of the traced banks.

Using the raw distances obtained, and ultimately filtering out the samples that fell within the stream banks, a Box-Cox transformation of the data was performed in order to create a normal distribution of the sample values. A normally distributed sample

population is necessary for ANOVA to be used to evaluate data and compare central tendencies (Myoung Park, 2008). Following this transformation, ANOVA was performed for the four datasets in this study and while investigating the combination of computer-processed vs human-assisted datasets. The calculations used to determine the raw distances and the summaries of the Box-Cox transformation results and ANOVA are outlined below.

4.3 Distance Data Storage and Naming Conventions

The usage of the DSAS transect casting method for collecting distance data produced 535 samples. These samples were collected from where both the north and south sides of the stream banks were visible along a transect. All other transects that were obscured by vegetation, shadow, or roadways were eliminated.

The streams' distance data were kept in spreadsheet files with appropriate names such as "Big Papillion Creek all transects", indicating all of the transects cast are still intact. The name "Big Papillion Creek transects deleted" indicates that the transects, where the intersection points were obscured, have been deleted from the table.

Each of the edited spreadsheet files were kept in a folder titled "All transects combined after deletion." The title of the file housing the stream data with the deleted transects was "All streams all columns after transect deletion." Within this same folder, the unnecessary columns for the distance data were eliminated. The resulting file was titled "Distance data after deletions."

4.4 Determination of Overall Accuracy of Predicted Stream Channels

To determine the overall accuracy of a derived stream, more calculations were needed within the spreadsheet to be able to determine if the derived streams were

produced accurately. First, the overall traced stream width at each transect would need to be calculated. Second, these width measurements would need to be divided in half. Doing this allowed for a center point to be determined. Third, the distances from the baseline to each derived stream channel location would need to be compared to a value representing either the north or south stream channel to see if the predicted channel was within the traced banks, thus a correctly placed sample.

If a derived stream channel's distance from the center of a traced stream was determined to be closer than the traced bank was to that center point, it was deduced that it fell within the stream channel, and was included as an accurately predicted stream channel location. If a derived stream channel location was farther from the center of a traced stream than either of the traced banks, it was regarded as an inaccurate prediction. The amount of discrepancy between the predicted stream channel location's center and the nearest traced stream bank was then calculated. To clarify, if a derived stream channel location fell outside of the traced stream channel's banks, the distance from the derived centerline to the closer of the two traced stream banks was kept and considered error in the placement.

Ultimately, these correctly placed samples were included and another comparison was made to determine if they would change the results significantly.

4.5 Rationale for Determining Stream Width

Calculating the stream width along the transects by applying the absolute value function to a subtraction of the south distance from the north distance, as outlined in Chapter 3, allows for some interesting examinations.

The formula =ABS(AF3-AE3) was first used to obtain the stream width, V3, which represented a cell in the data manipulation spreadsheet where the original

calculations of distance measurements were kept. The variable in this equation, AF3, corresponded to the north bank. AE3 represented the south bank. The Absolute Value function was used in this equation since some of the transects were cast from opposite directions in order to parallel the general direction of the streams better, This procedure was outlined in Chapter 3 due to limitations of the DSAS extension.

The stream's width, V3, was then divided by two using the simple calculation $= (V3/2)$ in order to obtain the distance from the banks to the center of the stream channel along each transect. This stream width value was ultimately used to determine if a dataset was able to place a stream channel within the actual traced stream.

The distance to the center of the traced stream channel from the baseline was calculated with the equation $= (AE3 + AF3) / 2$. This calculation was performed for each transect in the spreadsheet.

4.6 Determination of the Number of Correctly Placed Samples

The distances to the center of the derived streams for each dataset was then calculated using the equation $= IF(A3 < (V3/2), 1, 0)$. In this equation the value in the spreadsheet cell A3 represents the dataset's extracted stream center distance from the baseline. If that value was less than half of the overall stream width, or between the center and either of the traced banks, then a value of 1 was assigned to the cell. This indicated that at that particular transect the stream was within the traced stream banks, or a correctly predicted stream location. Ultimately, all of the transects where a correct placement of the derived stream location occurred were summed.

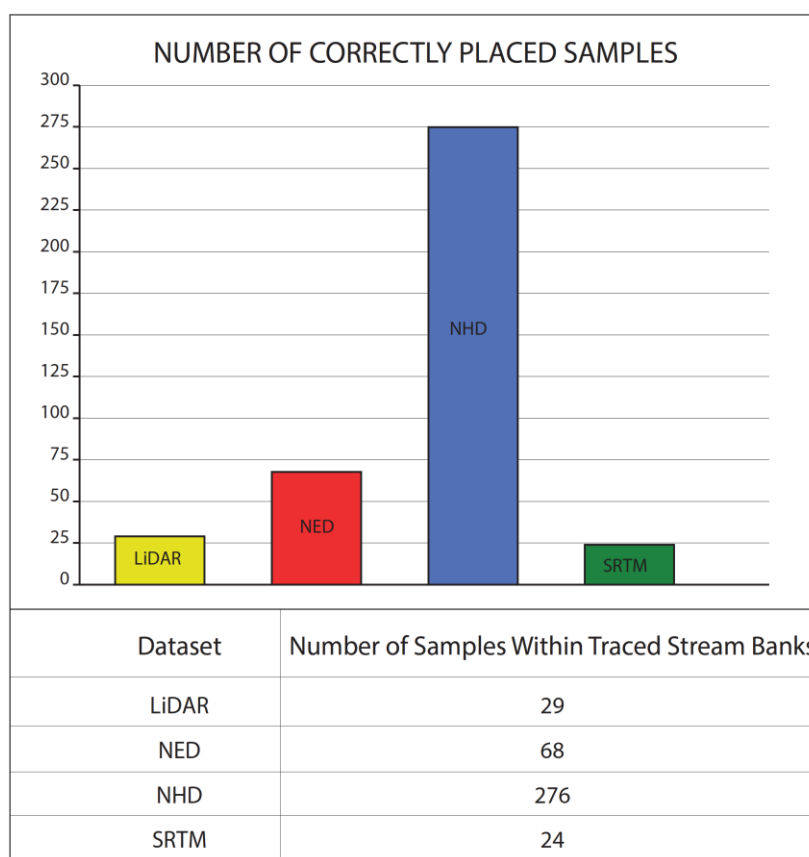


Figure 4.1. Number of samples found to lie within the traced stream banks for DEMS created from data obtained from LiDAR, the NED, and the SRTM along with stream locations obtained from the NHD.

4.7 Raster Cell Size Considerations

Within the data manipulation spreadsheet, a determination of how many samples were within the traced stream banks when cell size was taken into account was performed. If a derived stream location was found to be closer than that raster's cell size, then that sample value was deemed to be accurate. An assumed cell size value of just 0.167 meters was given to the NHD which corresponded to the use of aerial photographs with a resolution of 6 inches per pixel.

The number of samples where the derived centerline is within the respective limits of those cell sizes for each DEM show a large increase in the number accurately

placed samples for each dataset. This investigation produced 110 samples for LiDAR, 437 samples for the NED, 288 samples for the NHD, and 372 samples for the SRTM.

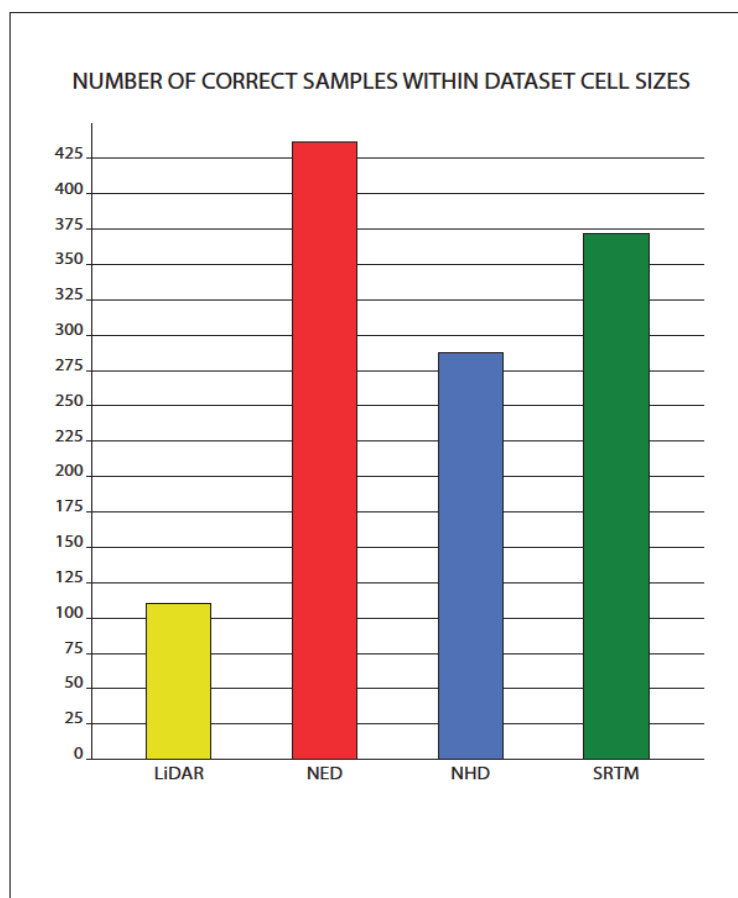


Figure 4.2. Number of sample distances that were within the respective dataset raster cell sizes from the traced streams.

The NED produced the most number of samples that were within the DEM's raster cell size when compared to the traced stream. This dataset was also the best performing DEM when examining the number of samples correctly placed within the traced stream banks with 68. At most locations of the Papillion Creek Watershed the stream channel width was far less than the 30 meter cell size of the NED DEM, so that dataset producing the most number of samples within the traced stream of the three DEMs investigated in this study was notable.

4.8 Distances of Derived Stream Channel Locations

The various datasets and the corresponding sample distances from their centerlines to the nearer traced stream bank were placed into a new spreadsheet for manipulation. The values that were ultimately compared using ANOVA were the distances from the traced stream banks to the extracted centerlines of the streams of the various datasets.

The spreadsheet utilized three different columns for eventual use in SAS, and for ANOVA. Column one was named “identifier” and was simply a repeating sequence of the numbers 1, 2, 3, or 4 which corresponded to the datasets under scrutiny. The “identifier” column’s values, corresponding to each of the various datasets, was used within SAS to combine the two computer processed datasets (LiDAR and SRTM) and the two human-assisted datasets (NED and NHD).

The second column in the data manipulation spreadsheet was labeled “transect id.” This column corresponded to the individual sample identification numbers taken along each transect to the various dataset distances from their centerlines to the nearer stream bank. In this second column one would see the same sample number four consecutive times followed by the next transect number four times, and so on.

The third column in this spreadsheet contained the actual distances from the derived stream channel centerlines to the nearer traced stream bank.

4.9 ANOVA

Results of ANOVA of the distances of the derived stream channels from the traced stream banks for each dataset are presented here. In order to perform ANOVA, the distributions are required to be normally distributed. If a set of data’s distribution is skewed left or right then the mean may not be a good indicator of its central tendency.

This study attempted to determine the central tendency of various datasets. Utilizing ANOVA, one can determine if the various datasets are the same. If they are not, then I would reject the null hypothesis that there is no difference between them and accept the alternative that there are differences in the various dataset's abilities to produce the stream channels accurately.

4.10 Box-Cox Transformation

After importing the data into SAS, with values of zero assigned to the correctly placed stream channels, the datasets were initially tested for normality using the following Distribution Analysis tests: Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling.

The four tests for normality failed for all four datasets. A Box-Cox Transformation was performed on was performed on all four datasets using the equation, $(y_i^\lambda - 1)/\lambda$ within the data manipulation spreadsheet. After importing this file into SAS, the Distribution Analysis for normality was again performed.

With the correctly placed samples excluded, passing normality values for all of the tests and all of the datasets, except for LiDAR, were obtained. Inspection of the LiDAR distribution shows a somewhat bimodal distribution, and can be seen in Figure 4.3. No amount of transformation was going to correct this. Since all other datasets did pass for normal distributions, an ANOVA of these newly transformed values was performed including the transformed LiDAR data. The correctly placed samples were both excluded and included in a series of ANOVA, and the results of these analyses follows.

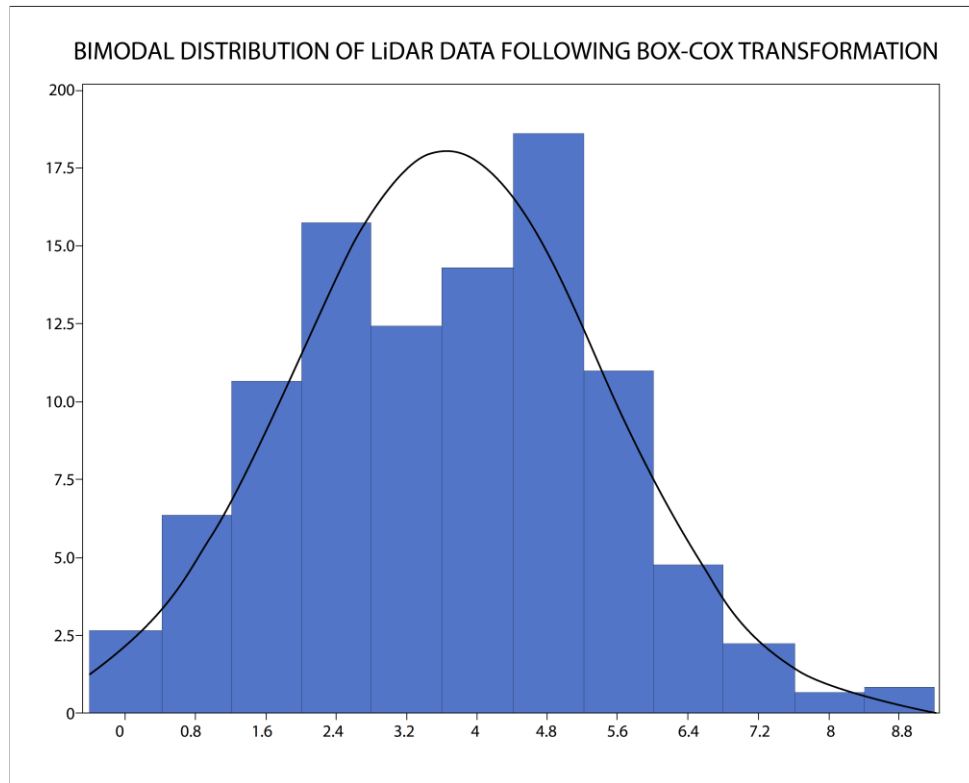


Figure 4.3. Distribution of LiDAR data following Box-Cox transformation.

4.11 Analysis

The average distances of the dataset's error with correctly placed samples included and assigned distance measurements of zero were: LiDAR, 93.78938 meters, 18.95377 meters for NED, 1.590288 meters for the NHD, and 93.92602 meters for SRTM. These average distances are non-transformed values.

The average distances of the dataset's error with correctly placed samples excluded from the calculation were: LiDAR, 99.16467 meters, 21.71363 meters for NED, 3.284959 meters for the NHD, and 98.33742 meters for SRTM. Again, these average distances are non-transformed values.

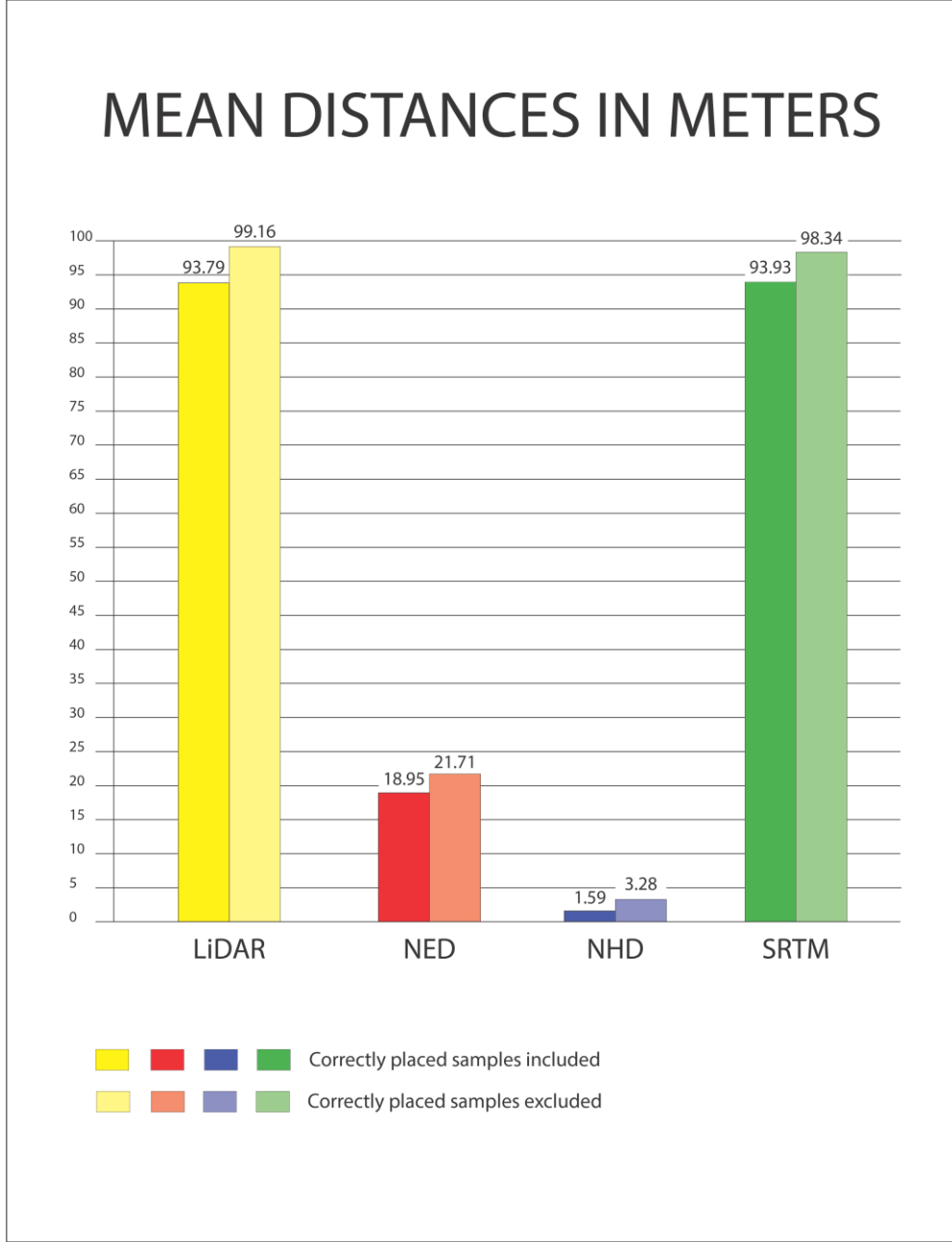


Figure 4.4. Means of sample distances to traced stream banks. Correctly placed samples found to be inside the traced stream banks were included and assigned distance measurements of zero. This graphic also shows the means obtained when those samples are excluded.

One-Way ANOVA					
Box-Cox Transformed Data - All Four Datasets					
	Number of Samples	Sum of Squares (Transformed Data)	Mean Square (Transformed Data)	F Value	Pr>F
Correctly placed samples Excluded	1640	1635.853	545.284	249.53	<0.0001
Correctly placed samples Included	2140	4142.561	1380.854	541.73	<0.0001

Table 4.1 Summary of ANOVA of Box-Cox transformed data for all four datasets with the correctly placed samples excluded and included.

Student-Newman-Keuls			
Box-Cox Transformed Data - All Four Datasets			
Means with the same letter are not significantly different			
SNK Grouping Correctly Placed Samples Excluded	Mean (Transformed Value)	Number of Samples	Dataset
A	4.3295	505	SRTM
B	3.8173	490	LiDAR
C	2.7912	450	NED
D	1.2007	196	NHD
SNK Grouping Correctly Placed Samples Included	Mean (Transformed Value)	Number of Samples	Dataset
A	4.08675	535	SRTM
B	3.49622	535	LiDAR
C	2.34777	535	NED
D	0.43989	535	NHD

Table 4.2 Student-Newman-Keuls test results of Box-Cox transformed data for all four datasets with the correctly placed samples excluded and included.

After inspecting Tables 4.1 and 4.2, there are definitely differences in the DEM's abilities to predict stream channels accurately. Using the extraction methods in the hydrology tools available in ESRI's ArcGIS software for the datasets with differing resolutions revealed that no dataset delineated stream network is the same.

However, further inspection of Tables 4.1 and 4.2 reveals that the finest resolution DEM (LiDAR) is not the best. In fact, it appears to be nearly identical to the coarsest DEM (SRTM). Grouping the DEMs according to whether or not there has been any human assistance in their creation reveals more convincing evidence that there are strong differences in their abilities to predict stream channels.

4.12 Rationale for Regrouping Datasets

For further analysis, the LiDAR data was grouped with the SRTM data and the NED data was grouped together with the manually-drawn NHD data. The LiDAR DEM was created following ESRI's recommendations, and during the creation of it the cell size could be manually set to whatever the user wished. Besides applying the 4.34 meter cell size, there was no pre-processing of the LAS files for this particular study. Also, the SRTM data was downloaded directly as one DEM and not manipulated prior to analysis. However, the NED and NHD datasets are pre-processed significantly by people. The NHD is manually drawn from aerial photographs and the NED is created using a variety of sources depending on the resolution of the DEMs available for download. According to the USGS, 30 meter NED DEMs were prepared from stereo photography, as well as derived from 1:24,000 topographic maps (<http://ned.usgs.gov/>), so a DEM created from those analog sources is open to human interpretation.

Tables 4.3 and 4.4 contain the results of the analysis of computer processed and human assisted datasets. Again, the computer processed datasets are nearly identical as are

the human assisted datasets. Also, it is clear to see that people are still needed to produce the best results.

<p style="text-align: center;">One-Way ANOVA Box-Cox Transformed Data – Computer Processed and Human Assisted Datasets</p>					
	Number of Samples	Sum of Squares (Transformed Data)	Mean Square (Transformed Data)	F Value	Pr>F
Correctly placed samples Excluded	1640	1225.205870	1225.205870	503.56	<0.0001
Correctly placed samples Included	2140	3075.571249	3075.571249	1009.82	<0.0001

Table 4.3 Summary of ANOVA of Box-Cox transformed data for computer processed datasets (SRTM and LiDAR) and human assisted datasets (NED and NHD). Correctly placed samples excluded and included.

Student-Newman-Keuls Box-Cox Transformed Data - Computer Processed and Human Assisted Datasets			
Means with the same letter are not significantly different			
SNK Grouping Correctly Placed Samples Excluded	Mean (Transformed Value)	Number of Samples	Dataset
A	4.07727	995	Computer Processed (LiDAR, SRTM)
B	2.30867	646	Human Assisted (NED, NHD)
SNK Grouping Correctly Placed Samples Included	Mean (Transformed Value)	Number of Samples	Dataset
A	3.79148	1070	Computer Processed (LiDAR, SRTM)
B	1.39383	1070	Human Assisted (NED, NHD)

Table 4.4 Student-Newman-Keuls test results of Box-Cox transformed data for computer processed datasets (SRTM and LiDAR) and human assisted datasets (NED and NHD). Correctly placed samples excluded and included.

SUMMARY

5.1 Introduction

This chapter will recap the rationale, methods, and the outcomes of this study. The purpose of this research was to determine if DEMs with different grid spacing automatically delineated stream channels in the same horizontal locations, determine if human assisted datasets were the same as computer processed datasets. The project also created a method to evaluate the abilities of DEMs to create stream networks.

This thesis used the hydrology tool set from ESRI's ArcGIS software to delineate stream locations for the Papillion Creek Watershed. The ArcGIS extension Digital Shoreline Analysis System (DSAS) was used to collect sample measurements from a calculated centerline of stream channels traced from aerial photographs to the extracted centerlines of the various datasets being examined. These measurements were ultimately compared using SAS and evaluated using ANOVA to determine if there were differences in their means.

This research can aid in the evaluation of new stream delineation algorithms and hydrologic mapping by providing a means to be certain of the quality of a raster surface and the accuracy of automatically extracted stream locations.

5.2 Datasets Examined

Three DEMs were compared in this study. The NED raster had grid spacing of 30 meters per cell. The SRTM data used for this study was 90 meters per cell. Using the Point File tool in ArcGIS as specified in the ESRI workflow for LAS files yielded an average spacing of 4.34 meters per cell. A fourth dataset was utilized during the study and was included in the analysis. The National Hydrography Dataset (NHD) is available

in vector form. No hydrology tools in ArcGIS were needed for the NHD. This dataset most closely matched the traced stream when compared to the three DEMs scrutinized.

5.3 Procedure

Both stream banks of the named streams of the watershed were traced from aerial photographs and have a resolution of six inches per cell. Baselines were drawn parallel to the general direction of flow of the traced streams and the ArcGIS extension DSAS was used to cast orthogonal transects from them. The transects intersected with the traced stream banks and the automatically delineated streams of the various datasets being examined. DSAS automatically captures distances of these intersection points from the baselines and they were exported to a Microsoft Excel spreadsheet for manipulation, further calculations, and comparison using SAS.

Averaging the intersection distances of the traced stream banks provided a centerline of the stream. The distance of the derived centerlines to the calculated traced stream centerline was then considered and it was determined whether or not a derived stream was within the traced stream banks and assigned an error distance value of zero. If any samples did not lie within the traced stream banks the distance from the derived stream to the nearer stream bank were kept. All viable transect measurements were then imported to SAS Enterprise for comparison.

An initial Box-Cox transformation of each of the datasets was needed to produce a normal distribution of the samples to perform ANOVA. Student Newman Keuls shows four separate groupings. One is then led to reject the null hypothesis that the datasets are the same and accept the alternative that they are not.

The datasets were regrouped according to whether or not there was human assistance in their creation. The LiDAR data and the SRTM DEM were combined as

computer-processed, and the NED and NHD were combined into a single human-assisted dataset.

5.4 Study Area Considerations

One thing to consider is the different study area being investigated in this and previous stream extraction studies. Mouton's (2005) study investigated mountainous terrain while the Papillion Creek Watershed is rolling hills and glacial till, while a large portion of the watershed lies within low flat lying terrain.

The flood plain area of the Papillion Creek, where it merges with the Missouri River, seemed to show the greatest distances of the derived LiDAR stream channel compared to other areas. In these flood plain areas the stream channels have been channelized. It is suspected that if the derived stream channel is placed outside of the traced stream location by the extraction tools it is difficult for the algorithm to place the stream back into the correct location. This may be due to the height of man-made levees which contain the streams. Another reason for this large difference may be that the LiDAR DEM was actually placing the derived stream into a former stream channel. However, the SRTM data also showed a large variation in this particular area.

Figure 5.1 shows the location of the floodplain area of the Papillion creek along with the correct stream channel that the NHD stream line follows well. In this location, the derived stream locations of the SRTM and LiDAR datasets are off by several thousand feet.

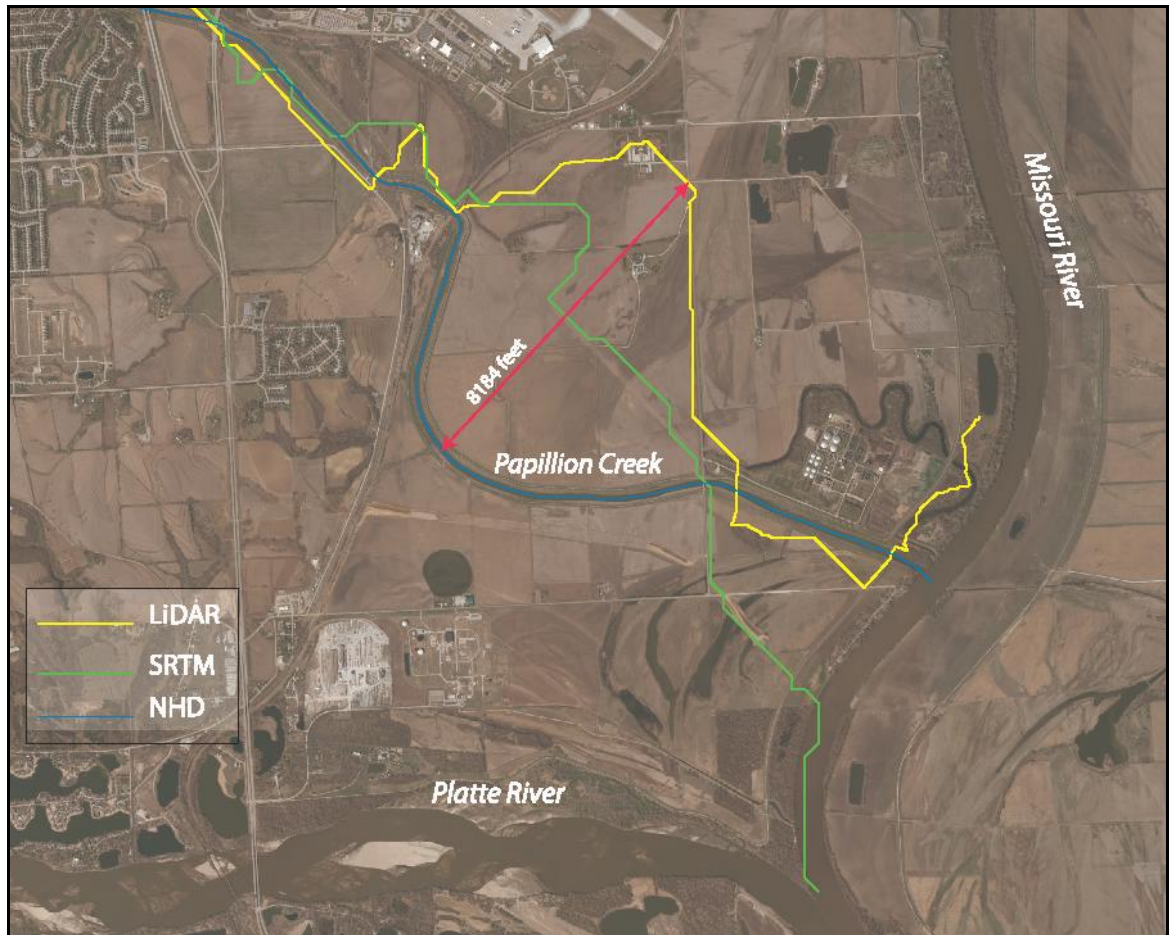


Figure 5.1. Computer processed DEM derived stream locations (SRTM and LiDAR) in the floodplain of the Papillion Creek near the confluence with the Missouri River. The NHD stream follows the stream channel in this area well.

However, previous research suggests that LiDAR performed best at picking out stream channels in flat areas and flood plains. This discrepancy may be caused by excessive preprocessing of the LiDAR data, or by researchers focusing on non-channelized areas.. It is not certain what other factors may contribute to the differences in LiDAR's ability to reproduce stream channels. There is very little research to validate LiDAR's use for such stream feature extraction purposes. More research is needed to eliminate these questions.

5.5 Resolution Comments

Based on the results of this study, while LiDAR does produce fine detail of a stream's sinuosity, photographs still seem to be the best option for correct positioning. Likewise, studies have shown that LiDAR data was better able to produce the look of a stream network once the extremely small streams are generalized away. This should come as no surprise, for there are simply more cells with which to delineate the flow paths. However, there was very little proof that these stream channels, albeit delineated at an extremely large scale, were accurate horizontally because many of these evaluations were performed using only visual inspection. Also, many of these studies evaluated DEMs and their ability to account for the total flow over a DEM. They were more interested in the sinuosity, or total distance of a stream and overall area of a watershed, and determining the total amount of water in the stream system, than with the quality of the positions of the streams themselves.

There were studies evaluating algorithms or datasets that showed how the streams were not placed in the proper locations when the vector was placed on the graphical representation of the DEM. Researchers commented that, visually, they were in the incorrect positions. Even in Mouton's (2005) study, the grid resolution of LiDAR data was altered. This change using the same data subsequently produced different routes for the stream channel. This leads one to believe that without extra processing of fine detail data to smooth out the digital surface in order to increase the signal to noise ratio, one should expect different results each time.

While there were differences in the datasets, it was surprising that the finest resolution was not the best at producing the positions of the streams. But, as Zhou and Liu (2004) point out, choosing the appropriate algorithm is important when fine detail

data is being used. However, Mouton (2005) claimed that the finer detail leads to increased accuracy. Clearly, after reviewing the results of this study, both of these researchers are correct in their respective conclusions. However, as was discussed previously, there must be an upper limit to the resolution needed in order to obtain adequate feature locations, and Quinn et al. (1991) and Zhang and Montgomery (1994) do point this out. Of the DEMs examined in this study, the NED performed the best. Its cell size resolution is 30 meters and represents a medium detail raster when compared to the other two. This result reflects the conclusions of Quinn et al. (1991) and Zhang and Montgomery (1994) by showing that the highest resolution DEM was not the best at producing stream locations.

5.6 Preprocessing Comments

There is an obvious association between a dataset's ability to reproduce a stream network and whether or not there has been human interference in manipulating the data prior to feature extraction.

5.7 Overall Cost Considerations

Another thing to consider after viewing the results of this study is the time involved in developing a stream network from the various datasets, and overall cost. LiDAR data is not cheap. While quick to produce a detailed surface, it requires time researching and calibrating the devices used to collect the information in addition to the cost of the flights themselves. The post processing needed to produce a final result must also be considered. SRTM and NED data are readily available in raster format for free, besides the time required for the ArcGIS tools to run. However, the time needed for the ArcGIS hydrology tools to run for SRTM and NED is minimal when compared to the

LiDAR raster. NHD data is already available in vector form, and, as determined in this study, produces the most accurate representation of reality.

The time involved to run the hydrology tools varied according to the dataset used, and the computer equipment. Of course, with finer resolution there would be an increase in the number of calculations involved, but the differences noticed in this study were staggering. While the SRTM and NED data required only a few minutes to complete a single tool, the same processes performed with the LiDAR data required several days to complete. Additionally, if a user wanted to manipulate the LiDAR data in some way in order to improve the performance of stream extraction, then there would be time and cost associated with that work as well.

An upgrade to ArcGIS from ESRI did allow for 64 bit background geoprocessing. This did provide noticeably quicker feature extraction, but during the time needed to run the tools, one would still have to hope there was not a power outage or other problems with the machine being used. Even with the upgrade, the LiDAR data required much more time to perform all the calculations of the extraction tools. Thus, eliminating the speed in which the data was collected to begin with. An aerial photograph would suffice for stream tracing, although tracing of the streams would consume time by the user. The NHD is a viable option for researchers when only higher order streams are of interest.

More investigation of the ability of LiDAR to reproduce stream networks is warranted after observing the results of this investigation. While the NED is updated using newer technology and the best elevation data available. As recently as April 2014, the NED was being updated continuously (<http://ned.usgs.gov/>). Even LiDAR is used to create DEMs for the NED. However, there remains some trepidation in using it. Even the USGS Center for LiDAR Information Coordination and Knowledge (CLICK) warns that,

while many people are interested in the use of LiDAR data for the creation of DEMs, there remains a great deal of research to be done to understand and use it appropriately (<http://lidar.cr.usgs.gov/>). The method presented here can assist in this work. It is repeatable and various changes can be made to the processing of the data prior to feature extraction in order to determine a best practice for using this type of data.

While investigating increased resolution of DEMs and stream extraction seems to be a current trend in research, the method outlined here may be employed to determine appropriate grid sizes for those investigations. Again, future testing of a dataset such as LiDAR where the resolution can be set by the user, or a dataset that can be resampled at different resolutions, can be compared easily in order to determine the best performing DEM grid size. Any result of new research can then be compared to reality, or the NHD, quite easily in order to determine the best match.

5.8 Conclusions

This study showed that once stream channels are traced and in place within ArcGIS, their locations can be sampled using the DSAS extension and compared to extracted streams from a DEM. Following the method applied here, DEMs and algorithms used to delineate streams can be evaluated in a scientific way.

It was found that grid size does impact a DEM's ability to extract stream channel locations, and that differences between datasets can be shown. It was also found that the finest detail DEM was not the best at reproducing a stream location. However, the reasons for the discrepancies in the DEM's abilities to extract stream locations still exist, and understanding the impact of human assistance in preprocessing elevation data prior to feature extraction is certainly needed after reviewing the results.

Ultimately, this study produced a method for researchers to be able to evaluate any DEM or stream extraction algorithm's ability to reproduce stream locations. It is clear after inspecting the results that there is still value in the collection and use of aerial photographs and human interpretation. The NED produced from USGS topographic maps and aerial photographs performed the best. It is also clear that more research is needed to increase LiDAR's ability to produce streams from DEMs created from it. The procedure presented here can aid in this research.

REFERENCES:

- Anderson, Danny L., Daniel P. Ames, *A method for extracting stream channel flow paths from LiDAR point cloud data*, Journal of Spatial Hydrology, Vol. 11, No. 1, Spring 2011.
- Atwood, Wallace W., *Physiographic provinces of North America*, Ginn and Company, 1940.
- Band, L. E., *Topographic partitioning of watersheds with digital elevation models*, Water Resources Research, 22, p. 15-24.
- Callow, John N., Kimberly P. Van Niel, Guy S. Boggs, *How does modifying a DEM to reflect known hydrology affect subsequent terrain analysis?*, Journal of Hydrology, 2007, 332, p. 30-39.
- Center for LiDAR Information Coordination and Knowledge (CLICK), <http://lidar.cr.usgs.gov/>.
- Colson, Thomas P., et al., *Comparison of stream extraction models using LiDAR DEMs*, Geographic Information Systems and Water Resources IV AWRA Spring Specialty Conference, Houston, Texas, May 8-10, 2006.
- Desmet, P. P. J. and Gerard Govers, *Algorithms to route flow over digital landscapes: A Comparison and Their Implications for Predicting Ephemeral Gullies*, Geomorphologie: Relief, Processus, Environnement, 1996, Vol. 2, Issue 2-3, p. 41-50.
- ESRI help pages, Last accessed 03-March-2014, <http://resources.arcgis.com>.
- Fairfield, John and Pierre Leymarie, *Drainage networks from grid digital elevation models*, Water Resources Research, May 1991, Vol. 27, No. 5, p. 709-717.
- Florinsky, Igor, *Combined analysis of digital terrain models and remotely sensed data in landscape investigations*, Progress in Physical Geography, 1998, 22, 1, p. 33-60.
- Heine, Reuben A., Christopher L. Lant and Raja R. Sengupta, *Development and comparison of approaches for automated mapping of stream channel networks*, Annals of the Association of American Geographers, 2004, 94(3), p. 477-490.
- Hosseinzadeh, Seyed Reza, *Drainage network analysis, comparison of digital elevation model (DEM) from ASTER with high resolution satellite image and aerial photographs*, International Journal of Environmental Science and Development, June 2011, Vol. 2, No. 3.

- Jancso Tamas and Gabor Melykuti, *Comparison of digital terrain models gained by different technologies*, 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), June 29, 2011-July 1, 2011, p. 324-329.
- Jones, Richard, *Algorithms for using a DEM for mapping catchment areas of stream sediment samples*, Computers and Geosciences, November 2002, Vol. 28, Issue 9, p. 1051-1060.
- Kiss, Richard, *Determination of drainage network in digital elevation models, utilities and limitations*, Journal of Hungarian Geomathematics, 2004, Vol. 2, p. 16-29.
- Mark, David M., *Network models in geomorphology*, Modeling Geomorphological Systems, 1988, John Wiley, New York.
- McMaster, Kevin J., *Effects of digital elevation model resolution on derived stream network positions*, Water Resources Research, 2002, Vol. 38, No. 4, p. 1-8.
- Moore, I D., E. M. O'Loughlin and G. J. Burch, *A contour-based topographic model for hydrological and ecological applications*, Earth Surface Processes Landforms, 1988, 13, p. 305-320.
- Moore, I. D., R. B. Grayson and A. R. Ladson, *Digital terrain modelling: A review of hydrological, geomorphological, and biological applications*, Hydrological Processes, 1991, Vol. 5, p. 3-30.
- Mouton, Adam, *Generating stream maps using LiDAR derived digital elevation models and 10m USGS DEM*, Master's Thesis, 2005, University of Washington.
- National Elevation Dataset, <http://ned.usgs.gov/>.
- National Hydrography Dataset, <http://viewer.nationalmap.gov/viewer/>.
- National Oceanographic and Atmospheric Association (NOAA), Last accessed 03-March-2014, <http://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data>.
- O'Callaghan, John F. and David M. Mark, *The extraction of drainage networks from digital elevation data*, Computer Vision, Graphics, and Image Processing, 1984, Vol. 28, p. 323-344.
- Papillion Creek Watershed Project, Last accessed 03-March-2014, <http://www.papillioncreek.org/overview.html>.
- Park, Hun Myong, *Univariate analysis and normality test using SAS, Stata and SPSS*, 2008, The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University.

- Passalacqua, Paola et al., *A geometric framework for channel network extraction from LiDAR: nonlinear diffusion and geodesic paths*, 7 January 2010, Journal of Geophysical Research, Vol., 115, F01002.
- Peucker, Thomas K. and David H. Douglas, *Detection of surface-specific points by local parallel processing of discrete terrain elevation data*, Computer Graphics and Image Processing, 1975, 4, p. 375-387.
- Pryde et al., *Comparison of watershed boundaries derived from SRTM and ASTER digital elevation datasets and from a digitized topographic map*, American Society of Agricultural and Biological Engineers, 2007, Paper Number: 072093.
- Quinn, P., et al., *The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models*, Hydrological Processes, 1991, 5, p. 59–79.
- Rayburg S., M. Thomas and M. Neave, *A comparison of digital elevation model generated from different data sources*, Geomorphology, 2009, 106, p. 261-270.
- Shuttle Radar Topography Mission, <http://gdex.cr.usgs.gov/gdex/>
- Speight, James G., *Parametric description of land form*, Land Evaluation, 1968, CSIRO/UNESCO Symposium, p. 239-250.
- Tarboton, D.G., R.L. Bras and I. Rodriguez-Iturbe, *Scaling and elevation in river networks*, Water Resources Research, 1989, Vol. 25, No. 9, p. 2037-2051.
- Tarboton, D.G., R.L. Bras and I. Rodriguez-Iturbe, *On the extraction of channel networks from digital elevation data*, Hydrological Processes, 1991 Vol. 5, p. 81–100.
- Tarboton, David G., *A new method for the determination of flow directions and upslope areas in grid digital elevation models*, Water Resources Research, 1997, 33(2): p. 309-319.
- Thieler, E.R., Himmelstoss, E.A., Zichichi, J.L., and Ergul, Ayhan, 2009, *Digital Shoreline Analysis System (DSAS) version 4.0— An ArcGIS extension for calculating shoreline change*: U.S. Geological Survey Open-File Report 2008-1278.
- Tribe, Andrea, *Automated recognition of valley heads from digital elevation models*. Earth Surface Processes and Landforms, 1991, Vol. 16, Issue 1, p. 33-49.
- Tribe, Andrea, *Problems in automated recognition of valley features from digital elevation models and a new method toward their resolution*, Earth Surface Processes and Landforms, 1992, Vol. 17, Issue 5, p. 437-454.

United States Census Bureau, Last accessed 03-Mar-2013, <https://www.census.gov>.

United States Geological Survey (USGS), Digital Shoreline Analysis System (DSAS) version 4.0, Last accessed 03-March-2014, <http://woodshole.er.usgs.gov/project-pages/dsas/version4/index.html>.

Wilson, John P., and J. P. Gallant, *Digital terrain analysis, Terrain Analysis: Principles and Applications*, New York: John Wiley and Sons: p. 1-27.

Wilson, John P., et al. *Water in the landscape: A review of contemporary flow routing algorithms*, *Advances in Digital Terrain Analysis: Lecture Notes in Geoinformation and Cartography*, 2008, p. 213-236.

Zhang, W., D. R. Montgomery, *Digital elevation model grid size, landscape representation, and hydrologic simulations*, *Water Resources Research*, 1994, 30, p. 1019– 1028.

Zhou, Qiming and Xue Jun Liu, *Analysis of errors of derived slope and aspect related to DEM data properties*, *Computer and Geosciences*, 2004, 30, p. 369-378.