Basin-scale spatiotemporal analysis of hydrologic floodplain connectivity

David Michael McCann

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Biological Systems Engineering

Durelle T. Scott, Chair
Zachary M. Easton
Kevin J. McGuire

April 16, 2014
Blacksburg, Virginia

Keywords: floodplain connectivity, Geographic Information Systems (GIS), inundation area, flooding, regional regression

Basin-scale spatiotemporal analysis of hydrologic floodplain connectivity

David Michael McCann

ABSTRACT

Floodplain inundation often provides water quality benefits by trapping sediment and biogeochemically transforming other pollutants. Hydrologic floodplain connectivity is a measure of water exchanges and interactions between the main channel and the floodplain via surface (inundation) and subsurface (groundwater) connections. Using an automated model combining GIS and numerical analysis software, this study examined floodplain inundation patterns and measured floodplain connectivity for the Mahantango Creek watershed (Pennsylvania, USA). Connectivity was quantified by developing a metric that included inundation area and duration. Long-term hydrographs at each reach in the watershed were developed via QPPQ (Flow-Percentile-Percentile-Flow) methodology using regional regression analysis to calculate the ungauged flow duration curves (FDC). Inundation area (normalized to stream length) was found to increase with drainage area, suggesting larger streams have more area available for biogeochemical activity. Annual connectivity increased with drainage area, suggesting larger streams, having higher connectivity, should be the focus of individual reach restoration projects due to higher potential for water quality benefits. Across the watershed as a whole, however, the total annual connectivity across first order streams was greater than higher order streams, suggesting the collection of small streams in a watershed may have a stronger effect on outlet water quality. Connectivity was consistently higher during the non-growing season, which was attributed to higher flows. Despite higher connectivity during the non-growing season, increased floodplain biological activity may be negated by low temperatures, reducing microbial activity. Correlations between land use and connectivity were also found, emphasizing dynamics between flow, channel morphology, and floodplain inundation.

Acknowledgements

The completion of this thesis and Master's project would not have been possible without the help and assistance of several individuals. First, I thank my committee chair and advisor, **Dr. Durelle Scott**. You encouraged and advised me throughout my undergraduate career and through an undergraduate research project, and because of that, I knew I wanted to work with you for my graduate work. You have pushed me to research well and to strive for a better understanding of science. To my other committee members, **Dr. Zachary Easton** and **Dr. Kevin McGuire**, I thank you for your continual technical guidance and advice. Your knowledge on the subject matter was extremely helpful in completing the project in a way that will push forward the science of flood hydrology.

I also must thank my fellow research teammate, **Nate Jones**. You have been more than a teammate, but a friend throughout my entire project. Your passion for the subject led to many conversations and discussions that shaped my project throughout. You also were always there to encourage me and tell me that I would, indeed, finish. You also helped me keep a balance to life and graduate school.

And lastly, I thank my parents, **Ed & Debbie McCann**. You were the ones who originally pushed me to pursue a Master's degree. Thank you for pushing me to stick with it and finish. Thank you for instilling in me a tough work ethic from an early age, which has helped on numerous occasions throughout my life, and especially in finishing this degree.

Table of Contents

List of Figures

List of Tables

1. Introduction

Flooding is one of the natural processes affecting the landscape at a wide variety of scales and perspectives. Across the globe, communities strive to protect the human infrastructure from flood damage, often without examining the impacts of flooding beyond anthropogenic-related concerns. From the human perspective, floods are a nuisance and cause significant expenses to rebuild damaged infrastructure. When examining flooding from an ecosystem perspective, however, the processes involved become significantly important to the function of the river-floodplain ecosystem.

Junk et al. (1989) present the flood pulse concept in which periodic flooding supports biological activity by providing a dynamic environment of metabolic needs (nutrients, etc.). Flooding occurs when the flow is at such a level to breech the primary banks of the river channel, and thus exceeds "bankfull" flow. Once bankfull flow is exceeded, the additional water begins to inundate the floodplain. The spatial extent and duration of the inundation is dependent on many factors, one of which is the size of the system at the point in question. Larger drainage area systems can be characterized by fewer flood occurrences (Tabacchi et al., 1998), but the floods that do happen have a longer duration than in smaller streams (Swanson et al., 1998; Spink et al., 1998; Junk et al., 1989). Small streams, on the other hand, have opposite flooding characteristics including a short duration (Swanson et al., 1998) and occur more times throughout the year, which could be attributed to the relationship between rainfall and in these smaller systems (Junk et al., 1989). These differences in flooding patterns can impact the net ecological effect of floodplain inundation on water quality and other ecosystem services. Understanding the role of system size (as measured by drainage area, stream order, etc.) on flooding characteristics and floodplain inundation is critical to realizing the full ecosystem benefits associated with river-floodplain networks.

Once inundated, multitudes of processes occur in floodplains, which interact to provide ecosystem services as well as to stimulate biogeochemical activity. Inundation in the floodplain can have a wide variety of effects ranging from sedimentation (Hupp, 2000; Tabacchi et al., 1998) to water-air gas exchange (Bustillo et al., 2011).

Another key benefit of floodplains is the potential for nutrient retention. Due to the impacts of excess nitrogen leading to eutrophication and other ecological problems (Mosier et

al., 2001; Diaz and Rosenberg, 2008; Rabalais et al., 2002; Downing et al., 1999), removing nitrogen prior to confluence with coastal waterways has become of significant importance. One common form of such nutrient retention is denitrification. Denitrification is the conversion of nitrate to nitrous oxide and nitrogen gas, essentially removing the nitrogen from the aquatic system. Denitrification, as well as other forms of nutrient retention have been observed in several floodplain/riparian zone studies (Forshay and Stanley, 2005; Roley et al., 2012). Junk et al. (1989) stress, however, that biogeochemical activity in the floodplain is not a direct result of solely the main channel water composition, but is a compilation of multiple water inputs acting under different circumstances.

Despite all the potential benefits associated with floodplains and riparian zones, many of these benefits may not be realized if water does not inundate the floodplain, triggering many of the process-related benefits. River-floodplain hydrological connectivity is a concept used to describe the transport and exchange of water between the main channel and the floodplain via surface or subsurface flow. This study focused on measuring floodplain surface inundation patterns, assuming groundwater flow, and thus subsurface connectivity, occurs with surface connectivity (floodplain inundated area connected to the main channel). Given this, connectivity was able to be examined via remote spatial analysis techniques.

Floodplains with high surface connectivity have the potential for wide expanses of inundation combined with long durations of inundation. Low connectivity, on the other hand, results in infrequent floodplain inundation due to channel morphology that raises the river stage associated with a bankfull flow event. The National Research Council (2002) described how low connectivity corresponds to reduced benefits associated with riparian zones. The degree of hydrological connectivity in a river-floodplain system can impact the ecosystem in a wide variety of ways. Low connectivity can have negative ecosystem effects including the reduction of available in-stream metabolic resources since over-bankfull flooding is infrequent, representing fewer opportunities for floodplain-to-main channel transport (Tockner et al., 1999a). Several studies have shown that connectivity can affect biogeochemical activity (Noe et al., 2013) including denitrification (Racchetti et al., 2011) as well as nitrogen transport (Richardson et al., 2004). Thus, understanding the factors, both controllable and natural, affecting river-floodplain connectivity is critical to maintaining floodplain ecosystem services and sustaining water quality for future generations.

The objectives in this study involved quantifying and analyzing river-floodplain connectivity from both a spatial and temporal perspective. GIS analysis, regional regression equations, and other concepts of hydrology were used to develop both the spatial and temporal components of the river-floodplain connectivity metric. My hypothesis was that both inundation area and connectivity patterns would have significant relationships with watershed variables. The specific objectives of this study were to:

1. Examine spatial inundation patterns and relationships across the watershed;
2. Quantify river-floodplain connectivity using both area of inundation as well as inundation duration at a watershed scale; and,
3. Examine connectivity relationships with various watershed and reach characteristics.

2. Literature Review

2.1 Flooding & Flood Theory

Starting with climate as the catalyst, flooding is a series of overlapping processes dependent on water source (rainfall vs. snow, etc.), land cover, and topography (Swanson et al., 1998). The timing and intensity of flooding events also depend on the specific sub-factors related to climate, as well as the subsurface soil properties of the catchment, as well as the channel itself. Depending on physiographic region, natural precipitation patterns of frequency, duration, and intensity all affect the size and duration of floods. More rainfall at a higher intensity has a higher likelihood of resulting in floods, by concept. When this precipitation occurs, depending on degree of intensity, flooding will generally only occur once the catchment becomes fairly saturated. Thus, the antecedent conditions of the soils will determine how quickly surface flow will occur, leading to flooding. On the topic of soils, soil hydraulic conductivity, among other soil characteristics, plays a key role in infiltration. It is when infiltration slows and begins to stop that surface flow occurs, once again, leading to flooding.

Among all ecosystem disturbances that could occur to a river-floodplain system, Swanson et al. (1998) claim floods to be "...the most frequent and intense natural physical disturbances". Impacts from floods can span large areas on a varying temporal basis (Gregory et al., 1991). Flooding can be part of the inherent fluvial processes for a river system, with floods merely causing a jump in the evolutionary process (Michener and Haeuber, 1998). It helps sustain the water feature and terrestrial heterogeneity (Ward et al., 1999b). The general flooding process consists of increased river stage leading to flow onto the floodplain, depositing material, causing "natural levees" to form along the channel (Hupp, 2000).

River-floodplain systems may contain a variety of specific water features including the river, side arms (eupotamal), tributaries, dead arms connected at the downstream end (parapotamal), former braids (plesiopotamal), abandoned meanders (palaeopotmal), alluvial spring brooks, and wetland areas (Ward et al., 1999b). Floodplains themselves, common to nearly all river systems (Stanford and Ward, 1993) can develop on varying levels over time on a river (Gregory et al., 1991).

2.1.1 Flood Theory

Because floods are such common and occasionally catastrophic events, it is important to understand their roles in ecosystem function. Several theories have been developed to serve as a framework for examining floods including the flood pulse concept (Junk et al., 1989), the River Continuum Concept (Vannote et al., 1980), the Riverine Productivity Model (Thorp and Delong, 1994), the Serial Discontinuity Concept (Ward and Stanford, 1983), and more.

First, the flood pulse concept stresses the importance of flood events to the ecological activity and biogeochemical cycling between the floodplain and river. Junk et al. (1989) present the aquatic/terrestrial transition zone (ATTZ) and its role in both sporadic headwater floods and higher order lowland floods. Also, horizontal transfer between channel and floodplain is a key component of the flood pulse concept presented by Junk et al. (1989). This area of interaction between the channel and floodplain can begin to serve as a base definition for a hydrologic connection. Although complete floodplain inundation is a critical component of stream function (Junk et al., 1989), Tockner et al. (2000) express the equal importance of below-bankfull events. This expansion and contraction concept focuses on applying similar principles from the flood pulse concept on a much broader spectrum of systems, not only high order rivers. They emphasize both large flooding events, as well as "expansion-contraction" and smaller "flow pulses" and their varying effects on connectivity in a river-floodplain system.

One base theory to studying river systems is the River Continuum Concept (RCC) (Vannote et al., 1980). Developed based on "natural, unperturbed ecosystems...", the RCC suggests that river processes follow gradients in physical property transitions (such as channel dimensions, flow properties, etc.) moving from upstream low order streams, to downstream large order rivers. For example, they hypothesize that ecological structure correlates to the expenditure patterns of the river's kinetic energy; thus, a disturbance in kinetic energy would result in a change in the ecological community. Another key tenet of the RCC is that of energy transfer from upstream to downstream. Vannote et al. (1980) suggest that as coarse particulate organic matter (CPOM) originating from the riparian zone is processed upstream, any inefficiencies in metabolism would be transferred downstream in the form of an energy source for high order river biological processes. They emphasize the energy continuum from upstream to downstream and suggest that floodplain energy sources in high order rivers are "insignificant..." and attribute additional organic matter sources downstream to be from low

order tributaries rather than floodplain materials. Support for this comes from the gradient of metabolic energy source, ranging from primarily allochthonous in the headwaters to more autochthonous production near the confluence (Vannote et al., 1980).

Many studies, however, do not support the river continuum concept (RCC) proposed by Vannote et al. (1980), as their results and analysis have supported the integral nature of floodplains and hydrologic connectivity (Heiler et al., 1995; Junk et al., 1989; Sedell et al. 1989). Sedell et al. (1989) notes limitations in the RCC, including the concept of river-forest interactions, something not taken into account in the RCC. They discount the RCC by describing it as being "of limited value for predicting large river ecosystem function," and expressing the complexity of river systems as "not a simple continuum."

Yet another base theory for large river ecosystem function is the Riverine Productivity Model (Thorp and Delong, 1994). This model, fit for large, constricted systems, emphasizes that the carbon source for in-stream metabolism is derived from direct riparian allochthonous sources in combination with in situ production (Thorp and Delong, 1994). They claim the RCC and flood pulse models "downplayed or virtually ignored the role of local instream primary production and riparian litterfall" (Thorp and Delong, 1994).

Because most large river ecosystem theories were developed based on relatively pristine conditions, Ward and Stanford (1983) propose the Serial Discontinuity Concept to begin understanding the increasingly common regulated river network. This idea emphasizes changes in water quality and quantity parameters from dams placed at various locations along the length of a river system. They propose a model to begin quantifying these effects by introducing two factors: discontinuity distance (DD) and parameter intensity (PI). Model analysis would include determining how far parameter values shift up or downstream (DD) and also the change in magnitude of each parameter (PI).

None of the theories are entirely the "best" way to examine river-floodplain systems, but in most situations, come combination of theories may explain hydrologic and biogeochemical phenomena better. This combination of theories was shown by Spink et al. (1998) when in their floodplain vegetation and nutrient study, they found that both longitudinal and lateral position within the river-floodplain system were important.

2.1.2 Factors Affecting Flooding

Spink et al. (1998) found in their floodplain vegetation and nutrient study that flooding, inundation, and residence time characteristics depend on several factors, including riparian zone and stream order. With flooding depending on so many factors, it is reasonable to see a high degree of variation in the effects of flooding (Swanson et al., 1998). Several studies have found flooding characteristics to vary with stream order. Lower order (headwater) type systems usually have flooding patterns characterized by short duration and unpredictability (Sedell et al., 1989; Junk et al., 1989; Swanson et al., 1998). Wissmar and Swanson (1990) hypothesize that the repeated and irregular disturbances of low order streams lead to decreased stability. Swanson et al. (1998), describe flooding in a mountainous setting with, "Water, soil, sediment, and woody debris move down hillslopes and stream channels by a cascade of processes, following the gravitational flow path." High order, lowland river, on the other hand, have flooding patterns that are often individually longer and more predictable (Junk et al., 1989; Swanson et al., 1998; Spink et al., 1998). Large rivers have a higher chance of flooding with relatively long floodplain residence times (Burt et al., 2002).

With respect to flooding patterns in general, high order systems experience events less frequently while steep, mountainous streams have flooding more often (Tabacchi et al., 1998). Meade et al. (1985) describe the Amazon River (large, high order system) flooding pattern as, "fairly simple" due to its high seasonality, with generally a single main annual flood. When discussing how stream systems are formed, however, Tabacchi et al. (1998) suggest that high order streams are formed by events happening more often, while low order streams are formed during rare, high-return period events.

Besides stream order, other factors affecting flooding include the topography, floodplain connectivity, erosive tendencies, and other soil characteristics. Combinations of stream power and bed material size help determine the geomorphologic structure of the river (Tabacchi et al., 1998). These two factors most likely relate to river slope, which Hupp and Osterkamp (1996) identify as a primary contributor to geomorphological characteristics. Particularly, the shape of the valley floor can impact whether a flood will have limited effects due to confined channels (narrow valley floor) or having mixed effects due to a wider valley floor with areas of fast and slow moving water (Swanson et al., 1998). In areas of high sediment transport, floods may have a higher influence due to low hydraulic conductivity of the soils (Brunke and Gosner, 1997).

Geographic region may also affect flooding characteristics. For example, Hupp and Osterkamp (1996) characterize coastal-plain systems as having generally lower slopes with clear periods of high and low flows, resulting in flood inundation cycles throughout the year. Floods occurring in areas that have connected floodplains may experience a lower degree of morphology change after extreme disturbances due to less destructive and lower energy flow and sediment transport associated with active floodplains (Michener et al., 1998).

2.1.3 Ecological Impact of Floods

Several factors affect the biological and chemical activity in river-floodplain systems including connectivity (Heiler et al., 1995; Sedell et al., 1989), flood pulses (Heiler et al., 1995), inundation area (Sedell et al., 1989), stream order (Montreuil et al., 2011; Vannote et al., 1980), etc. When floods occur, although situational, some of the characteristics affecting the system's reaction include flood size, habitat access (Michener and Haeuber, 1998), and area history (Michener and Haeuber, 1998; Swanson et al., 1998). How an ecosystem recovers from a flood depends on the types of organisms in the system, their interactions, as well as the shift in energy source, which leads to a shift in the entire food web (Swanson et al., 1998). Also, Swanson et al. (1998) note that, in response to a flood event, ecosystem reestablishment rates differ between main sections of the river-floodplain system, with the riverine areas redeveloping quicker than the terrestrial sections.

By nature, water is an integral part of the flooding process. Water in the river-floodplain system plays critical roles in biogeochemical cycling, ecological processes, and erosional activities (Hupp, 2000). Grimm et al. (2003) suggest that biogeochemical activity, and thus the resulting water composition, is dependent on both residence time as well as sediment properties.

Because headwater streams have relatively short residence times on the floodplain, direct allocthonous organic matter input from the riparian zone becomes more important than organic matter transfer between the floodplain and the channel (Sedell et al., 1989). This concept regarding low order streams is supported by Vannote et al. (1980) in discussion about the importance of floodplain vegetation as a source of detritus, and therefore energy input.

Floodplain soils are often highly productive because of the water availability and organic material introduced during flooding (Gregory et al., 1991). Molles et al. (1998) saw ecological

activity in the form of bacterial and fungal growth in presence of floodwater in a managed flood experimental setup. On the other hand, however, increased levels of suspended solids in floodwaters can reduce productivity in the stream (Heiler et al., 1995).

Besides aiding microbial activity, water is also critical to river-floodplain systems by determining redox characteristics for the channel and floodplain (Grimm et al., 2003) and also acting in biogeochemical activities (McClain et al., 2003). Anaerobic conditions can develop from a combination of long residence times and microbial metabolism (Grimm et al., 2003). Also related to anaerobic conditions and processes are electron acceptor composition/concentration and carbon supply (Grimm et al., 2003).

These are just some of the ecological factors and processes associated with flooding in a river-floodplain system. More detailed syntheses of flooding and floodplains in their relationships with wildlife, riparian zones, and nutrient cycling are provided below.

2.2 Hydrology

In order to gain a full understanding of natural systems, it is important to understand the intricacies of water movement between floodplain and channel. When discussing water movement between the floodplain and channel, this water transfer may occur as groundwater movement or as surface flow (Mertes, 1997). This is especially important in braided and meandering type systems (Ward and Stanford, 1995). The National Research Council (2002) suggests this as a key distinguishing factor from other landscape areas since riparian areas, in particular, are affected by hydrologic activity from the channel as well as the terrestrial environment. Ward and Stanford (1995) describe this lateral movement between channel and floodplain as one of the three interactive spatial dimensions of rivers.

For a base equation for floodplain hydrology, Tabacchi et al. (1998) define the water flux as the difference between the sum of losses (evapotranspiration and deep percolation) and the sum of inputs (precipitation, runoff, and subsurface flows). Given the many components of the equation, each individual piece can both affect and be affected by other hydrologic components in addition to various biotic and abiotic factors. Mertes (1997) suggests several factors that affect the way floodplains are inundated including channel structure, substrate content, floodplain topography, and water sources. Also, Lowrance et al (1984) suggest that above and

below-ground flow distribution towards the channel can be affected by climate, land use, and soil characteristics (Lowrance et al., 1984). During drier periods and when floodplains have lower slopes, water in the channel can flow back into the floodplain, versus the normal groundwater flow direction into the stream (Vidon and Hill, 2004b).

Flooding on the floodplain may be caused by surface overflow from the river itself, or it may occur from local inundation from groundwater, precipitation, basin runoff, tributaries, etc. (Mertes, 1997). Mertes (1997) describes the area where these two flood water sources mix as the perirheic zone, which can have significant floodplain impacts.

In addition to field-based studies, Bates et al. (2000) used numerical modeling to examine floodplain hydrology. In their study of a larger river-floodplain system, they suggest the importance of horizontal water movement between the uplands and channel.

2.2.1 Inundation

The time during which the floodplain is inundated and the processes leading to inundation are critical for many biogeochemical and fluvial processes, as well as for understanding the flooding process. The study conducted by Mertes (1997) of several large global river systems showed that local flooding of the floodplain could develop prior to surface transfer from the channel. Thus, the flow rate in the channel is not the only important factor to consider when analyzing flooding patterns and characteristics. Catchment surface runoff and topographic depressions, tied with soil moisture properties and conditions, can also play a role in flooding, and therefore biogeochemical cycling in a river-floodplain system.

As for the purposes of flooding, Hupp (2000) suggested that sedimentation benefits of floodplains might be realized most when the floodplain is inundated. Inundation of the floodplain is characterized by flow shallower than that in the channel, therefore with lower velocities, allowing for more settling of suspended loads and particulate matter. In the Amazon River study conducted by Bustillo et al. (2011), they found that the amount of inundation area during a flood has an effect on carbon dioxide movement between the water and atmosphere of the river-floodplain system. Besides sedimentation, floodplains that have been inundated for a length of time can support denitrification and other biogeochemical processes (Baldwin and Mitchell, 2000).

2.2.2 Connectivity

In a very general sense, connectivity with respect to the landscape refers to the ability to facilitate transfer between spatial locations (Taylor et al., 1993). They also suggest that connectivity between different "patches" in the environment is more important and can have a higher influence than physical distance of separation. Ward et al. (1999b) define connectivity "...as the ease with which organisms, matter or energy traverse the ecotones between adjacent ecological units." These definitions hold true to the hydrological connectivity of interest between the channel and associated floodplain, which is often a very complex subject, involving many components and processes (Brunke and Gosner, 1997).

Hydrologic connectivity, in a broad sense, can be related to connections starting from a point in the channel interacting with another point in the channel downstream, groundwater (downward), and lastly with other groundwater connections (e.g. hillslope aquifers) (Amoros and Bornette, 2002). This correlates with Amoros and Bornette's definition for riverscape, regarding it as the collection of bodies of water connected via groundwater or surface flow to the river itself, upland area, or alluvial aquifer (Amorose and Bornette, 2002). In the context of this study, hydrologic connectivity is primarily focused on those surface and subsurface interactions between the floodplain and channel.

Specifically measuring connectivity can be accomplished in several different ways. One particular study on the Danube River of Austria analyzed three different connection levels: disconnection, seepage inflow, and upstream/downstream surface connection (Tockner et al., 1999a). Each of these conditions corresponded to an increasing level of flood inundation (low flow, floodplain connected via groundwater, and overflow on surface back into main channel, respectively), measured by water level relative to the mean water level. Another study characterized connectivity in terms of water quality measures such as turbidity, nutrient levels, and productivity limitations (Knowlton and Jones, 1997). Described as "conductivity", Hughes (1980) expresses the importance of a hydrologic floodplain metric (enabling comparisons across landscape and hydrologic schemes), supporting the further analysis of floodplain inundation and storage. Jones et al. (2008) begin analyzing hydrologic connectivity by first instituting the concepts of hydrologic facets and impedance. Hydrologic facets are depressions in the landscape or otherwise small areas that are connected hydrologically. Once these facets are initially established, they can be aggregated to make a coarser map of facets, depending on the

hydrologic impedance. Hydrologic impedance is defined as, "the maximum change in river stage necessary to inundate the divide", with the divide being the boundary between two adjoining facets. With this concept model, low impedance correlates to high connectivity (Jones et al., 2008).

2.2.2.1 Factors Affecting Connectivity

The degree to which the river channel is connected to the floodplain can be influenced by varying factors at different levels of impact. In general, though, connectivity depends on the hydrologic nature and position of the section within the river system, as well as floodplain properties and degree of establishment (Tabacchi et al., 1998). Natural and anthropogenic erosion and sedimentation patterns of the river system can affect connectivity, with excessive scouring or deposition yielding different geomorphological features, thereby impacting connectivity (Amorose and Bornette, 2002). A study by Montreuil et al. (2011) had riparian wetlands near a 2nd and 5th order stream, which had higher connectivities with the upland versus channel, respectively. Thus, there may be a correlation or relationship between stream order and connectivity.

Floodplain topographic characteristics also play a major role in hydrologic connectivity. Varying levels of hydrologic connectivity exist between the channel and surrounding lentic bodies depending on the elevation at which the disconnection can be breached (Amorose and Bornette, 2002). Even if low-lying floodplain areas are no longer connected to the main channel via surface flow, there is the potential for filling (although slower than surface connection) via connection from groundwater flow (Ward et al., 1999b). Vidon and Hill (2004b) found that both slope and depth to a confining layer affected connectivity between the upland and the riparian zone. Topographic obstructions in the floodplain, such as levees, can decrease the exchange of water, and therefore nutrients and organic matter, between the channel and backwater areas (Johnston et al., 2001). More discussion about anthropogenic impacts on connectivity can be found in the section regarding human impacts to river-floodplain systems.

2.2.2.2 Impacts & Effects of Connectivity

Just as connectivity is impacted by different factors, the impacts of varying connectivity have been examined extensively. First, floodplain connectivity has important effects on aquatic and terrestrial habitat and species diversity. The benefits of riparian zones may become undermined if low connectivity between the channel and floodplain develops (National Research Council, 2002). Tockner et al. (1999b) used mollusks, macrophytes, odonate, and amphibians to determine a relationship between hydrologic connectivity and species diversity. Their study resulted in finding that a river with intermediate connectivity promoted the highest diversity given all four faunal metrics. They explained this by a balance of both hydrological (flow & sediment transport) and ecological processes, seen in highly and lowly connected areas, respectively. The intermediate level of connectivity is also supported by Knowlton and Jones (1997) for an ideal connectivity for autotrophic production. Tabacchi et al. (1998) also suggested that the highest level of species richness can be found where the most habitat develops, which is generally where intermediate flooding patterns exist. Similarly, Ward et al. (1999b) suggest that maximum biodiversity occurs with not a single flood scheme, but a wide distribution. The combination of small, medium, and large events would facilitate those organisms thriving under each respective condition (Ward et al., 1999b).

Vegetative development often follows flooding gradients, and therefore connectivity gradients moving horizontally from the river (Gregory et al., 1991). Flooding in connected areas of river-floodplain systems can result in turbulence and erosion characteristics that are detrimental to plant growth (Amorose and Bornette, 2002). Low connectivity can also result in ecological impairment of the stream due to reduced bioavailable organic matter. When low connectivity persists, nutrients and organic matter necessary for aquatic life become rare since they only enter the channel during high flow events (Tockner et al, 1999a). Hupp et al. (2009) note that excessively high connectivity, however, may lead to ecologically problematic sedimentation rates on the floodplain. This excessive sedimentation, potentially caused by high suspended sediment loads, can be detrimental to riparian ecology (Hupp, 2000). When floods occur, the associated connections with the floodplain provide opportunities for sediment and nutrient transfer, enabling the floodplain to be a resource sink or source (Taylor et al., 1993; Ward and Stanford, 1995).

Floodplain connectivity can also impact microbial function. Subsurface biological activity may be induced by the mixing of groundwater and surface water due to floodplain connections to groundwater (Brunke and Gosner, 1997). Noe et al. (2013) saw a clear effect of varying connectivity on biogeochemical cycling. In a related manner, Tabacchi et al. (1998) found that riparian nitrogen uptake is based primarily on "the length of the [subsurface] hydrologic contact zone". Racchetti et al. (2011) found significant differences between connected versus isolated wetlands and their denitrification performance. The connected wetlands consistently had higher denitrification rates than the isolated areas (Racchetti et al., 2011). In a similar manner, Richardson et al. (2004) on the Upper Mississippi River identified hydrologic connectivity as being an important factor in nitrogen cycling, with emphasis on connectivity's role in transporting nitrate. With regards to denitrification specifically, Wolf et al. (2013) did not see an effect of connectivity with the channel on denitrification potential in floodplain wetlands.

Surface and subsurface water chemistry may differ based on the degree of connectivity between the floodplain and river channel. Johnston et al. (2001) suggests that ecological processes in backwater locations impact surface water characteristics more so than the geological characteristics of the area. On the Danube River in Austria, Preiner et al. (2008) studied different levels of connectivity between side-arms and the main channel, and the impacts on carbon cycling in the aquatic ecosystem. Their results suggest that the channel energy web may be impacted by restoring, and therefore higher, connectivity due to a pulse of autochthonous organic matter.

2.3 Vegetation & Riparian Zones

Simply put, Gregory et al. (1991) define the riparian zones as, "the interfaces between terrestrial and aquatic ecosystems." For a more extensive definition, the National Research Council (2002) defines riparian areas as,

> "...transitional between terrestrial and aquatic ecosystems and are distinguished by gradients in biophysical conditions, ecological processes, and biota. They are areas through which surface and subsurface hydrology connect waterbodies with

> their adjacent uplands. They include those portions of terrestrial ecosystems that significantly influence exchanges of energy and matter with aquatic ecosystems (i.e., a zone of influence). Riparian areas are adjacent to perennial, intermittent, and ephemeral streams, lakes, and estuarine-marine shorelines."

Riparian zone development originates from the area's geomorphology, as well as the soil properties, erosion tendencies, and precipitation and flooding patterns (Tabacchi et al., 1998). Riparian zones can be affected by floods by creating, eliminating, and changing areas for plant growth and development (Wissmar and Swanson, 1990). Once hardier vegetation becomes more densely populated, riparian and streambank areas become more resistant to damage from regular flood events (Tabacchi et al., 1998). Channel organic matter characteristics may even be affected by this riparian vegetation (Gregory et al., 1991).

Several studies have shown floodplains and riparian zones to be effective filters of and sinks for sediment (Cooper et al., 1987; Dunne et al., 1998; Mertes, 1994). Vegetation in riparian zones and on streambanks promotes sediment deposition by providing flow resistance from the plants themselves, but also from litter buildup (Tabacchi et al., 1998). Heffernan et al. (2008) also showed that vegetation encouraged sedimentation of fine particles in their study of a desert stream system. Cooper et al. (1987) showed that riparian zones have the capability to capture sediment from upland agricultural sources. Brunet et al. (1994) note that inundated riparian area rather than floodplain surface was more important to river-floodplain system retention properties. Floodplain sedimentation can even help reduce the impact of increased erosion rates of anthropogenic origin (Hupp et al., 2009).

Vegetation also plays a role in soil characteristics that are important to flooding and associated biogeochemical processes. For example, vegetation type and age may affect soil moisture, as Hunter and Faulkner (2001) suggested that the older vegetation in their natural wetlands might have removed significant water from the soil. Another function of riparian vegetation is to serve as the primary carbon source for redox processes (Tabacchi et al., 1998). This function is discussed in more detail later in the section on nutrients and denitrification. In their study on vegetation in desert streams, Heffernan et al. (2008) often found anaerobic subsurface conditions associated with areas of vegetated cover.

In the concept model proposed by Vidon and Hill (2004a), the width of riparian zone needed for high nitrate removal depends on soil hydraulic conductivity, slope, and depth to a confining layer. For optimal riparian zone performance, Burt et al. (2002) suggest a riparian zone of adequate width to sustain appropriate water table depths, remains hydrologically connected, and is located along an intermediate-sized stream. Haycock and Burt (1993), however, suggest that soil properties are more important than site width in determining riparian zone nutrient removal effectiveness. Hill et al. (2000) support this by stating that riparian vegetation width is less of a factor in denitrification than soil carbon distribution and confining layer depth.

The development, distribution, and heterogeneity of floodplain vegetation can be affected by many factors including river current and wave action (Roberts and Ludwig, 1991), flood-caused erosion and deposition (Gregory et al., 1991), flooding and ecological disaster history (Gregory et al., 1991), topographic characteristics (Hupp and Osterkamp, 1996), vegetative competition (Hupp, 2000), plant stress tolerance (Hupp, 2000), and plant tolerances to flooding and erosion (Gregory et al., 1991). Hupp (2000) discussed distributional patterns based on slight elevation differences, most likely related to moisture/inundation regime. Floodplain vegetation utilize both surface water and groundwater for metabolic processes, but Stanford and Ward (1993) suggest that the groundwater and its components may have larger impact on plant growth. Spink et al. (1998) found that the limiting growth factor for riparian vegetation was not actually soil nutrient levels for their large river study. They hypothesize this occurred in their particular study because of excess nutrient supply or another more stringent growth limiter (e.g. water availability).

2.4 Wildlife & Habitat

Flooding and floodplains also affect habitat and play critical roles in the life cycles of many floral and faunal species. When examining habitat in floodplains, Amorose and Bornette (2002) discuss four factors affecting species diversity: suspended matter, nutrient composition, underlying geology, and water temperature. Connell (1978) suggests the intermediate disturbance hypothesis with the highest biodiversity coming from a moderate frequency of disturbances, maintaining a balance of competition (older communities with more time since last

disturbance) and pioneer development (recently disturbed). He explains that in recently disturbed and older communities, small pools of available pioneer species and competition, respectively, drive down species diversity, making an intermediate disturbance patterns more ideal (Connell, 1978). Besides disturbance patterns, different levels of connectivity caused by varying flow rates as described by Tocker et al. (1999a) can help increase ecological diversity by providing a means for species transport between lentic and lotic water bodies in the river-floodplain system (Amorose and Bornette, 2002). Knowlton and Jones (1997) found oxbow and scour lakes to have high potential for certain species (i.e. birds, land-animals, etc.), but due to their lack of connectivity with the Missouri River, would not serve as habitat for river-specific species. The connected sour lakes in their study, although valuable habitat for river-specific organisms at the time of the study did not show to have much future due to the already shallow and continuously decreasing depths.

Because of floods' diverse levels of impact across a landscape, various sections of habitat are demolished while others remain intact, providing "refuges" for organisms to reestablish communities after the event (Swanson et al., 1998). After Tropical Storm Alberto, Michener et al. (1998) found several aquatic fauna were relatively unaffected by the flooding associated with the event, which they attribute to the development of acceptable conditions via active floodplains.

Sedell et al. (1989) mentions the importance of river-floodplain interactions to fish populations, independent of system scale. When flooded, floodplains serve as breeding and feeding grounds for fish (Ward and Stanford, 1995; Ross and Baker, 1983). Depending on species and preferences, flooding can impact fish populations by providing nutrients from the floodplain, as well as adequate habitat and reproduction areas (Ross and Baker, 1983). Among fish species, cutthroat trout is one that has shown to be resilient through flooding (Swanson et al., 1998).

2.5 Nutrients

2.5.1 General Nutrient Processing

Biogeochemical cycling in a river-floodplain system is critical to understanding overall ecosystem nutrient dynamics and is highly dependent on many different factors. In general, nutrient processing is not independent, but varies with both time and space (McClain et al., 2003). The large concept of aquatic biogeochemical cycling can be broken down into considerations of nutrient transport/movement, in-stream processes, riparian zone/floodplain effects. Although many chemical compounds are involved in processes surrounding a river-floodplain system, special focus here is paid to nitrogen because of its multiple redox states, and therefore continuous transformation through various constituents.

Understanding nutrient transport mechanisms is important to understanding associated concepts of both in-stream and floodplain processing and retention. Before being transported, the source of the nutrient or organic matter is important to consider. Boyer and Groffman (1996) found that agricultural soils developed more water-soluble labile carbon than forest soils, hypothesizing the potential for stimulating microbial function (Boyer and Groffman, 1996). This source/transport phenomena can potentially affect riparian zone function since, in their study, the labile carbon difference was found in the top 0.5 m of soil, making it susceptible for runoff and movement towards a riparian zone. Nutrient transport is also affected by the hydrological form and particular compound in question. For example, dissolved organic carbon (DOC) transport and flux from watershed to stream can be impacted by precipitation (Royer and David, 2005). When looking at nitrogen and phosphorus, Peterjohn and Correll (1984) found that groundwater flow was more significant for nitrogen transport to riparian zones, while surface runoff was more important for phosphorus. When looking at losses from the riparian zone itself, similar trends were found. Groundwater flow was still important in nitrogen loss, but both surface and subsurface flow contributed to phosphorus losses (Peterjohn and Correll, 1984). These findings are most likely due to the nature of the respective nutrients, with nitrogen usually being found in the dissolved form and phosphorus usually being adsorbed to sediment. The nutrient retention results from Venterink et al. (2003) support this with lower nitrogen retention compared to phosphorus, most likely explained by soil adsorption characteristics of the two nutrients.

When taking a large view of the entire nutrient cycle, transport processes before reaching the riparian zone must also be taken into consideration. Where are the nutrients coming from and how are they getting to the riparian zone? The source of the nutrient can impact transport, depending on their form (dissolved, airborne, particulate, etc.). As aforementioned, different nutrients have different affinities for dissolved versus adsorbed to particulates, making transport processes different. Agricultural land uses are often targeted for study because of their high concentrations of nutrients and high potential for loss. Peterjohn and Correll (1984) found that the majority of the nitrogen and phosphorus was lost from the system via crop harvesting, eliminating nutrients before reaching the riparian zone.

When examining nutrient removal from source to final outlet, in-stream (Seitzinger et al., 2002) as well as riparian/floodplain (Forshay and Stanley, 2005) removal processes and characteristics must be taken into account. In-stream biogeochemical processing may be nutrient dependent, as Webster and Patten (1979) found that there was little effect from in-stream processing on calcium and potassium at a watershed scale. Royer and David (2005) looked at watershed effects, suggesting that stream higher concentrations of DOC in the landscape could translate to higher processing rates in-stream. After flow has reached the channel, Admiraal et al. (2002) found that phytoplankton have a large effect on in-stream organic matter dynamics in certain situations. This would suggest that land use and riparian areas have the main impact on surface runoff water composition up until the complex relationship with phytoplankton begins in the channel.

2.5.2 Nutrients in the Floodplain & Riparian Zone

Floodplains and riparian zones, among other ecotones, impact biogeochemical cycling in a river-floodplain system (Brunke and Gosner, 1997; Fabre et al., 1996; Forshay and Stanley, 2005; Gregory et al., 1991). Generally speaking, the larger the system and the more nutrients in the channel indicate higher levels on the floodplains (Spink et al., 1998). The results from Grubaugh and Anderson (1989) concluded that organic matter composition and movement can be affected by floodplains on large river systems (Grubaugh and Anderson, 1989). In a study of the Ill River in France, Termolieres et al. (1994) found that even highly polluted flood water can be sufficiently filtered via the floodplain and associated soil matrix, leading to clean groundwater

after infiltration. The Wisconsin River is another example where the floodplains have been shown to retain nitrate via denitrification (Forshay and Stanley, 2005). In reference to riparian zones in particular, they can have a much higher nitrogen and phosphorus retention as compared to contributing upland agricultural areas (Peterjohn and Correll, 1984). Another study on an agricultural watershed showed that riparian zones contribute to trapping several nutrients including nitrate, calcium, magnesium, potassium, and sulfate (Lowrance et al., 1984). On the other hand, Fabre et al. (1996) showed that riparian zones can be both a source and sink of phosphorus during flood events, depending on the form of phosphorus being questioned. They emphasize the times when the riparian zone is a phosphorus source because labile phosphorus was the form in flux. Similarly, Hein et al. (2005) found the floodplains of the Danube River to be sources of phosphorus during flooding events. Racchettie et al. (2011) showed that a floodplain's source or sink designation for nitrate might be dependent on both connectivity as well as season.

2.5.3 Nutrient Retention

Grimm et al. (2003) stress an integrated and holistic approach when analyzing watershed nutrient retention, including land, water, and riparian areas. Nutrient retention during flood events depends on several factors including the nutrient and its associated properties, connectivity, residence time, flow, etc. Venterink et al. (2003) compared different rivers with different levels of floodplain flow and found that the river with more water flowing through the floodplain retained more phosphorus. They suggested that more floodwater in the floodplain might help retain more phosphorus since retention depends on soil/water interaction and swiftness of flow (Venterink et al., 2003).

There is currently a divide in the literature in regards to how stream size and associated flow characteristics affect nutrient removal. An extensive study on the Mississippi River basin conducted by Alexandar et al. (2000) showed that nitrogen removal rates in the river network decrease as depth and water velocity increase. Also shown in this study was that proximity to watershed outlet may not have a direct correlation to nitrogen delivery as some watersheds farther from the outlet had higher nitrogen delivery than watersheds close to the outlet (Alexander et al., 2000). In the Gulf of Mexico nitrogen loading study conducted by Goolsby et

al. (2001), they did not factor denitrification or floodplain retention into their calculations, suggesting the processes are insignificant in large rivers. Peterson et al. (2001) express the importance of small streams due to their impact on biogeochemical cycles in the channel, especially in regards to nitrogen removal. Mulholland et al. (2008) on the other hand, suggest that the relation between nitrogen removal rate and stream size is dependent on nitrogen concentrations and sources. They found that large river removal is dependent on the upstream small river removal rate, the latter of which decreases as nitrate inputs increase. If nitrogen loading is very high, the stream size has little effect on removal since all efficiencies decline (Mulholland et al., 2008). Also, Mulholland et al. (2008) note that in circumstances where hydrologic shortcuts are created and there is a higher loading directly to large rivers, a shifting of removal effectiveness towards larger reaches may take place. In a model study (excluding riparian nutrient removal) conducted by Seitzinger et al. (2002), they found that low order streams have better nitrogen removal efficiencies (inputs versus outputs) although larger rivers remove more on a per length basis because of the large total amounts of nitrogen entering from the rest of the watershed. From this, they suggest that both large (small percentage of total system length) and small river systems are important for nitrogen removal on total removal and efficiency bases, respectively (Seitzinger et al., 2002). From this, one might conclude that restoration of larger rivers may be most effective given smaller restoration lengths would yield more removal of nutrients, nitrogen in particular. A potential issue, however, is whether it is feasible to restore in-stream processes in a system since most restoration projects focus on riparian areas.

Among other factors, Grimm et al. (2003) suggest that the combination of abiotic with biotic factors can create a synergistic effect on nutrient retention. Nutrient losses from the riparian zone may also depend on the vegetation age, as Emmet et al. (1993) found that nitrate loss into local streams increased after forest age surpasses 30 years.

2.5.4 Nitrogen Processing

Of the many pollutants occurring in waterways today, nitrogen has come to the forefront of research and study. The National Academy of Engineering (2012) claimed clean water access and management of nitrogen cycle to be two of the "grand challenges" facing society. The

effects of nitrate overload on the hydrologic cycle have been studied and researched to see what the actual impact has been, is currently, and will be in the future. Mosier et al. (2001) conducted a review of nitrogen cycling trends as impacted by human population growth and noted several negative human and environmental impacts of excess nitrate, including eutrophication, biodiversity issues, oxygen depletion, water quality decline etc. The "dead zone" in the Gulf of Mexico is specific area of intense study, with Rabalais et al. (2002) reviewing the effects of higher nutrient loads on the hypoxic area, including negative effects (speciation shift, food web alterations, etc.) on various aquatic species. Similarly, Diaz and Rosenberg (2008) relate increased anthropogenic nitrogen to eutrophication, often leading to hypoxic dead zones which cause aquatic habitat loss and organism death. Smith's (2003) review of eutrophication examined the intricate dynamics of nutrient surpluses and algae production in relation to eutrophication in both freshwater and saline aquatic ecosystems. This review also included a brief outline of numerous negative human and environmental consequences of eutrophication. Among several explanations for increased algal blooms, Hallegraeff (1993) suggests anthropogenic eutrophication as a potential cause, leading to human health problems and aquaculture yield declines. Vitousek et al. (1997) warn of human health concerns, fish kills, and seafood intoxication as a result of algal blooms. With these global-scale negative impacts of nitrogen excess, nitrate removal has been a key focus in aquatic ecosystem studies. Goolsby et al. (2001) found that 61% of the mean annual nitrogen flux into the Gulf of Mexico from 1980-1996 was in the form of nitrate: the constituent often of highest interest when examining loading as well as denitrification. The model study conducted by Scavia et al. (2003) identified a nitrogen load reduction of 30% could result in a 20-60% reduction of the hypoxic zone area in the Gulf of Mexico.

When examining the nitrogen cycle in a river-floodplain system, many components and factors are involved, encompassing the air, water, and soil. One of the many benefits associated with floodplains and riparian zones is the potential for nitrate removal (Haycock and Burt, 1993; Lowrance et al., 1984; Pinay et al., 1995; Burt et al., 2002). Although nitrate is the compound of highest concern, several authors have noted the involvement of other nitrogen forms. Mulholland et al. (2000) found nitrification as a key step in the nitrogen cycle in regards to nitrogen removal from in-stream processing. Additionally, Hill (1990) found that ammonium

processing was more evident than nitrate processing in riparian zones when focusing on various groundwater sources.

As for nitrate, levels may be impacted by the surrounding land use as well as nitrogen cycling processes taking place in the stream itself (Peterson et al., 2001). Plant uptake and denitrification are the two main pathways that riparian zones utilize to remove nitrate from the system (Lowrance et al., 1984; Pinay et al., 1995). Both these processes are important, as Pinay (1995) noted that flooded conditions facilitate higher denitrification, whereas dry conditions support plant uptake. This study focuses on denitrification, given its influence on floodplain biogeochemistry.

2.5.4.1 Denitrification

Denitrification has the potential for significant reductions in overall nitrogen export (Howarth et al., 1996). Mulholland et al. (2008) define denitrification as, "...the primary mechanism by which bioavailable nitrogen is permanently removed from ecosystems." From a chemical compound perspective, denitrification is essentially microbial metabolism without oxygen as an electron acceptor. In the absence of oxygen, nitrate is often the next best acceptor of electrons. When nitrate is involved in the process without oxygen, dinitrogen gas is produced. Because of the natural solubility properties of dinitrogen gas, it is released from the riverine system after being produced during denitrification (Baldwin and Mitchell, 2000). If conditions are not completely anoxic, and some oxygen is present, nitrous oxide has the potential to be released. In general, the process could take place when nitrate-laden groundwater comes in contact with organic matter and bacteria that are often found in the upper soil horizons (Gold et al., 2001). McClain et al. (2003) explain that both aerobic and anaerobic aquatic conditions are necessary for nitrate generation and then denitrification, respectively. In fact, Haycock and Burt (1993) suggest that denitrification may be most effective where soil conditions transition from oxygenated to anaerobic. Similarly, in their connectivity modeling study on the Danube River, Tritthart et al. (2011) located the highest rates of microbial respiration where disconnected areas of the floodplain were inundated during a flood event. In channels were turbulent conditions may occur, reaeration of previously anoxic sediments in the hyporheic zone may occur. This introduction to oxygen back into the system has the potential to stimulate biogeochemical

cycling. This concept of reaeration has been modeled and quantified by several studies, summarized in Bott (2006) as well as Chapra (1997). In an anaerobic environment, denitrification generally occurs after the sulphfide and ammonium ions have been oxidized and soil manganate has been reduced (Tabacchi et al., 1998). This agrees with Mulholland et al. (2000) who suggest that ammonium is utilized for nitrogen before nitrate.

Several factors affect denitrification rates including temperature (Pinay et al., 2002; Strauss et al., 2006), soils (Burt et al., 2002; Vidon and Hill, 2004a), topography/slope (Vidon and Hill, 2004a), vegetation (Predick and Stanley, 2010), nitrate levels (Forshay and Stanley, 2005; Johnston et al., 2001; Scaroni et al., 2001), carbon source/availability (Hill et al., 2000; Predick and Stanley, 2010), etc. First, the soil composition and moisture content of the floodplain affects subsurface denitrification processing (Gold et al., 2001). Several studies have found that where water tables are lower (naturally or anthropogenically) and well below the soil surface, little denitrification takes place, nullifying many riparian zone benefits (Bohlke and Denver, 1995; Hill et al., 2000). Burt et al. (2002) suggest that adjacent lotic systems can influence groundwater levels in riparian zones, possibly improving denitrification potential. Hunter and Faulkner's (2001) study found that hydrology and appropriate soil moisture conditions were one of the most important factors in determining the denitrification capacity. While moisture is important, if the soil profile remains completely saturated, riparian zone nitrate removal effectiveness may decline due to surface flow rather than subsurface flow (Vidon and Hill, 2004a). Ideal soils for denitrification allow enough water movement for sufficient nitrate flow, while being impermeable enough to maintain a high water table and anaerobic conditions (Burt et al., 2002). Hunter and Faulkner (2001) suggested that agricultural versus groundwater and precipitation sources of wetland moisture could have impacted denitrification ability in their restored wetland comparative study.

Predick and Stanley (2010) clearly showed that both flow rates and vegetation impacted denitrification capacity. In their study on sandbars in the Wisconsin River, they found that bars with no vegetation shifted between sources and sinks of nitrate given high and low flow, respectively. Vegetated bars, on the other hand, were found to be sinks of nitrate during higher flows, but had decreased nitrogen retention during low flows tied to several factors including low nitrate concentrations. They emphasized the differences in hydraulic conductivity between vegetated and unvegetated bars, since denitrification ability is tied to nitrate, carbon, and oxygen

supplies. Their results suggest that reduced hydraulic conductivity associated with the vegetated bars limited denitrification ability. In contrast, Deforet et al. (2008) saw little effect of vegetation on nitrogen retention in their study on parafluvial gravel bars. Similarly, Burt et al. (2002) saw little to no correlation between denitrification and vegetation cover. Vegetation may have a slight impact on denitrification, though, given their main interaction with the process is contributing subsurface organic carbon (Haycock and Burt, 1993). Heffernan et al. (2008) also suggest that instream vegetation serves as a subsurface organic matter source for anaerobic processes.

Gergel et al. (2005) showed that flood size and volume might affect overall nitrate processing in a river-floodplain system. Although larger floods process more nitrate due simply to higher fluxes of nitrate, they conclude that small floods may have a larger impact given a long period of time once recurrence intervals are taken into account. Richardson et al. (2004) also suggest that channel flow levels impact denitrification, primarily by affecting the quantity and movement of nitrate through the river-floodplain system.

Several studies found that denitrification may still take place, even effectively, in low temperatures associated with winter months (Haycock and Burt, 1993; Groffman and Hanson, 1997; Burt et al., 2002). Correll et al. (1997) saw seasonal differences in floodplain nitrate removal, with winter having higher removal. In contrast, Racchetti et al. (2011) found higher rates of denitrification in the summer versus the winter. Other studies have also seen effects of season on denitrification (Forshay and Stanley, 2005) and nitrogen retention (Deforet et al., 2008). In the results found by Vidon and Hill (2004a), slope seemed to be a significant factor affecting the overall sink size designation with the reasoning that higher slopes promote higher N fluxes. These high fluxes may not always be helpful for nitrate removal. Correll et al. (1997) found that steeper subsurface hydraulic gradients in the floodplain combined with very permeable soils led to more, faster moving nitrate, and an overloading of the system.

The limiting factor of denitrification varies across the river-floodplain lateral transect, with carbon being limited further away from the channel and nitrate being more limited closer to the channel (Hill et al., 2000). Predick and Stanley (2010) found that both nitrate and carbon can be limiting factors to denitrification, depending on flow and vegetation schemes. Racchetti et al. (2011) also found that denitrification was closely related to nitrate supply, and thus connectivity with the main channel. In their habitat study on the Atchafalaya River, Scaroni et al. (2011)

found denitrification to be limited by available nitrate rather than carbon. Richardson et al. (2004) found denitrification in the Upper Mississippi River to be primarily limited by nitrate levels, with only a minimal affect from additions of bioavailable carbon. Many other studies note nitrate as the limiting nutrient for denitrification (Forshay and Stanley, 2005; Johnston et al., 2001; Strauss et al., 2006).

Even with all these factors being taken into account, a combination of them may lead to optimal denitrification. In contrast, Groffman and Hanson (1997) found that neither vegetation type, soil drainage class, nor water table level were correlated to denitrification levels in a wetland.

2.6 Human Impact

A common trend in river systems today is to add hydrologic and hydraulic features to reduce risk of severe flood damage to farmland, homes, infrastructure, and other investments in the floodplain. These changes to the system including dams, levees, dredging, etc. have caused changes in the natural system (Swanson et al., 1998; Ward and Stanford, 1983; Gregory et al., 1991; Pinay et al., 2002; Junk et al., 1989; Vitousek et al., 1997). Human impact can be found across the span of river-floodplain systems in the United States, with Benke (1990) finding only 42 large (>200 km) rivers in the continental US that have not been introduced to significant regulation. Even with the uncertainty associated with parameters and the study methods, this still outlines the need for action.

Ward et al. (1999a) outline four general outcomes from river regulation: simplification of channel structure, repression of existing processes, hydrologic disconnection, and overall discontinuity. Changes in flow characteristics (channel and groundwater), channel bed modifications, and floodplain modifications can impact the fluvial characteristics of the river system (Amorose and Bornette, 2002). Friedman et al. (1998) found that over a wide spatial scale, most rivers generally narrowed after the implementation of upstream dams. Braided channels shift towards a narrower meandering channel and meandering channels become less dynamic in their movement (Friedman et al., 1998). In contrast, McClain et al. (1997) describe the Amazon River floodplain, generally considered non-impacted by human influence, as, "a

prominent feature on the landscape," containing a complex array of lentic and lotic features other than the main channel.

Pringle (2003) addresses the importance of understanding the relationship between human impact and connectivity by stating that "...we have grossly underestimated the power of the hydrologic cycle to transport human-generated wastes throughout the biosphere-- just as we have often undervalued the positive aspects of hydrologic connectivity in providing essential ecosystem services and transporting essential elements." Anthropogenic impact in the form of dams, hydroelectric infrastructure, weirs, levees, etc. has been seen and/or noted in several river connectivity studies and analyses (Heiler et al., 1995; Junk et al., 1989; Amoros and Bornette, 2002; Knowlton and Jones, 1997; Ward et al., 1999b). Sometime termed, "river regulation", river modifications often decrease the amount of water movement between the channel and floodplain, thereby decreasing connectivity (Ward and Stanford, 1995). Using a simulated "neutral-terrain" model, Gergel analyzed the effects of dams and levees on inundation area in the floodplain. After modeling various landscapes, anthropogenic scenarios, and flood levels, she found that levee and dam effects are highly dependent on flood magnitude. The impact of levees are seen by reduced inundation area up until the levees are breached. Dams were simulated using a lower peak flow and decreased temporary pond area for small floods, but increased temporary pond area during larger events. For the more frequent events, therefore, she concluded that connectivity decreased given an upstream dam situation. The 100-year flood event was found to be a certain threshold where the effects of dams and levees on temporary pond area were maximized due to channelization from the levees and lower peak stage caused by the dam.

Tabacchi et al. (1998) support a holistic approach to analyzing impacts of human water usage on the stream system since connectivity can lead to effects in the river and on the floodplain. Human impacts on the Missouri River have changed the natural geomorphological processes of the river, thereby changing the deposition patterns and reducing connectivity (Knowlton and Jones, 1997). Because of this, lateral water movement between the channel and floodplain has been severely attenuated, resulting in lower flood impact on the flood plain and impeded surface runoff no longer reaching the river except for during extreme flow events (Knowlton and Jones, 1997). Pringle (2003) illustrates how changes in hydrologic connectivity

can be related to several contemporary water quality issues related to the storage and transport of pollutants including methylmercury and endocrine disrupters.

Another problem of anthropogenic impact on river systems in regards to decreased connectivity is the restricted transport of organic matter, thus lowering the ability for aquatic production (Ward and Stanford, 1995). McClain et al. (2003) suggest that biogeochemical processing abilities in certain areas may be impacted by forest clearing and wetland draining.

Retention properties of the river system may also be impacted by river regulation (Gregory et al., 1991). Bruet et al. (1994) suggest a decrease in river retention because of smaller floodplains caused by human activity. This could be critical, as Dunne et al. (1998) found that Amazonian floodplains retain more sediment than that produced via bank erosion, thus making them important factors in overall sediment transport. Since the Amazon is still generally considered "natural" and free from anthropogenic impact, the study by Dunne et al. (1998) shows insight into the potential implications of reduced inundation area caused by anthropogenic influences on the watershed.

In the inundation model study conducted by Gergel et al. (2005), they found that simulated dams and levees significantly reduced nitrate processing in the floodplain simply because less nitrate was exposed on the floodplain. They even differentiated the effects of dams above the reach in question, as well as floodplain levees, with dams having a larger impact. Levees may shrink floodplain area (Hupp, 2000), thus potentially reducing retention benefits associated with floodplains. On a global scale, Syvitski et al. (2005) found that overall sediment transport to the oceans has actually been reduced by 1.4 BT/year since before major human impact on the hydrological system. This net change was realized despite increased anthropogenic erosion and was primarily attributed to reservoir retention (Syvitski et al., 2005).

If climate change proceeds in future years, significant hydrologic changes could occur (Grimm et al., 1997; Kundzewicz et al., 2007), which may have effects on nitrogen cycling due to changes in nitrogen transport and water distribution (Pinay et al., 2002). Vitousek et al. (1997) outline agricultural fertilizer, fossil fuel use, and planted legume fixation as major anthropogenic influences on the nitrogen cycle. Agricultural development has caused more nutrients to be introduced to waterways, such as increased groundwater nitrate concentrations (Bohlke and Denver, 1995), causing environmental issues and has led to, "...a monetary and energy loss to society" (Peterjohn and Correll, 1984). In a large-scale study conducted by

Howarth et al. (1996), fertilizer was found to make up approximately two-thirds of the total anthropogenic nitrogen influence on the North Atlantic Ocean. Nitrate removal potential is often decreased or removed when anthropogenic changes to the landscape cause a drop in the water table away from the soil surface (Gold et al., 2001). Land drainage may lower the water table enough to remove the benefits of riparian zones (Burt et al., 2002). Haycock and Burt (1993) also suggest that intact floodplains (not drained) have high potential for nitrate removal, and therefore should be a part of restoration efforts.

The creation of reservoirs can significantly affect hydrologic, and therefore geomorphological processes leading to increased channel uniformity and decreased channel movement (Shields et al., 2000). Grimm et al. (1997) also note significant effects of reservoirs on both the hydrology of the system, as well as water quality. The model developed by Seitzinger et al. (2002), however, found that reservoirs had little, yet variable, impact on in-stream nitrogen removal.

Vegetation and wildlife in the riparian zone and floodplain can have interactions with anthropogenic hydrologic change. Because of human interaction, non-native, or exotic vegetation species can develop and begin to flourish within river-floodplain systems (Tabacchi et al., 1998). Ward et al. (1999b) warn of the potential extermination of the current, "islands of biodiversity" that exist as the remaining river systems not affected by anthropogenic activities. On the other hand, because of the aforementioned geomorphology change that can be associated with dams on braided channels, Friedman et al. (1998) saw increased vegetation growth into the channel. On meandering channels, however, they saw a decrease in pioneer forest growth. Similar geomorphologic change and vegetative growth trends from anthropogenic alterations for braided and meandering channels were found by Johnson (1998) on the Platte and Missouri Rivers, respectively. Thus, the effects of anthropogenic hydrologic change on riparian vegetation may be based on the geomorphologic nature of the river-floodplain system. An earlier study on the Platte River by Johnson (1994) also suggested that anthropogenic water abstraction from the river system might have resulted in low flows that supported forest growth further into the channel. The Rio Grande is another river that has experienced significant effects from anthropogenic impact, including riparian zone changes and flood pattern adjustment (Molles et al., 1998). Hupp et al. (2009) note potential negative impacts on biodiversity and riparian habitat loss as a result of anthropogenic flow changes. In their study on the Roanoke

River (North Carolina, USA), they noted higher rates of bank erosion downstream from dams, causing high, steep banks and overall soil loss, leading to the riparian zone area loss. Based on their findings of wildlife flood tolerance, Michener et al. (1998) warn against flood management infrastructure because of its impact on the river-floodplain system.

While most studies have expressed the negative effects of anthropogenic impact on a river system (Stanford and Ward, 1993; Ward and Stanford, 1995; Knowlton and Jones, 1997; Ward et al., 1999b; Molles et al., 1998; Grimm et al., 1997; Gergel et al., 2005), Swanson et al. (1998) suggest mixed effects of both benefits and damage, depending on the type and location of the infrastructure. Howarth et al. (1996) suggest that not all nutrient problems found in aquatic ecosystems may be directly attributed to anthropogenic causes. For example, they found the Amazon basin to be the largest exporter of phosphorus (total and per area bases) out of several large basins emptying to the North Atlantic Ocean (Howarth et al., 1996). In their study, nitrogen outputs from the Amazon were also significantly high, despite low human influence (Howarth et al., 1996).

Effects of river regulation infrastructure placed at different locations along the river may vary, but independent of the location, some degree of downstream ramifications will take place (Ward and Stanford, 1983). Decrease in connectivity has varying affects among types of river systems, but has the potential for the highest effect on meandering reaches, which have many different lateral connections (Ward and Stanford, 1995). In a different realm, Molles et al. (1998) note that the effects of river regulation are most notably seen in the dry areas of the world, where water is a rare commodity.

2.7 Restoration

Restoration efforts are often not one-way, single attempt efforts, but are a part of a larger process, proceeding from poorly connected floodplains to a fully functional system with many processes in between (Molles et al., 1998). Restoring a river-floodplain system involves many parts, which experience high temporal heterogeneity during redevelopment (Molles et al., 1998). In their study on the "Fiume Tagliamento", one of the few remaining large-scale un-impacted river systems in Europe, Ward et al. (1999a) emphasize the importance of comprehending the natural system and how much prior impact has occurred before undertaking restoration practices.

Understanding the hydrologic and chemical interactions between lotic and the surrounding lentic environments during flooding is important if restoration of connectivity is to be considered (Heiler et al., 1995). Among many benefits, Bayley (1995) suggests economic benefits to restoration projects in addition to the environmental benefits, primarily from increased fish production during flooding. Hein et al. (2005) noted an increase in biogeochemical cycling on section of the Danube River with restored connectivity.

Different studies recommend different practical suggestions regarding restoration of river-floodplain systems, especially in regards to overall improvement and reestablishing denitrification potential. Some studies recommend small streams for restoration since their riparian zones have been shown to have more impact (Brunet et al., 1994; Peterson et al., 2001). The National Research Council (2002) recommends restoring low-order streams for purposes of pollutant removal because more of their flow originates from runoff, whereas in large river systems, less of the flow is introduced to the channel via floodplain runoff at a particular location. In contrast, Cooper et al., 1987 suggest that riparian zones are most effective in areas where the receiving water body is a larger order river where sediment settling is less likely (Cooper et al., 1987). Forshay and Stanley (2005) also support large river restoration as a way to increase nitrogen retention in the floodplain. For actual restoration methodology and design, a study conducted by Molles et al. (1998) on the Rio Grande showed that managed floods have the potential to help restore riparian zones. Riparian zone width along the channel should increase as erosive potential and the soil's clay content increase (Cooper et al., 1987). Location is also a factor, as restoration and protection may serve the best in the southeastern United States, since this is the area having the highest concentration of unregulated rivers with little protective legislation (Benke, 1990). Because denitrification is an important function of riparian zones, the water table level of a restored site should be raised back to appropriate levels if previously lowered due to human impact (Gold et al., 2001). If water table management is used as a method to reduce subsurface nitrate levels, Jacinthe et al. (2000) found that a high water table for extended period of time maximizes removal while minimizing N_2O emissions. Even still, higher water tables, although more efficient at nitrate removal, have the potential for elevated N_2O emissions (Jacinthe et al., 2000). Hunter and Faulkner (2001) studied the biogeochemical differences between natural and restored wetland areas, finding that although natural wetlands still provided the highest denitrification potential, restored wetlands could support denitrification

if the hydrology is correct. Similarly, Klocker et al. (2009) examined restored and natural stream systems for denitrification, concluding that if connectivity and residence time are restored, denitrification capacity may be affected.

Molles et al. (1998) comment on the widespread impact of decreasing floodplain connectivity, emphasizing the need for more research and concept development to formulate future implications. McClain et al. (2003) even note connectivity as an important consideration when choosing riparian zone restoration projects. Similarly, because of the many floodplain benefits, Hupp et al. (2009) recommend restoring river-floodplain systems to their "natural" connectivity regimes. Racchetti et al. (2011) also support floodplain reconnection to the channel for nitrate removal and other subsequent beneficial purposes.

Because of the changes occurring to connectivity, it is important to consider floodplain connectivity for restoration projects, depending on given circumstances and the restoration focus for each situation. If the focus is to reduce downstream nitrogen loads, Wolf et al. (2013) suggest that higher connectivity between floodplain wetlands and the channel can support more nitrogen retention. On the other hand, in the Danube Restoration study by Tockner et al. (1999b), the focus was on faunal species diversity. For this situation, the final recommendations involved a holistic approach involving the improvement of connectivity as well as hydrologic and geomorphologic functions. Amorose and Bornette (2002) call for a very broad approach to restoration, involving examining not only the species in question and its optimal level of connectivity, but considering the fully diverse nature of the system. They emphasize finding the cause to the problem at hand and ensuring that both temporal and spatial scales are taken into account (Amorose and Bornette, 2002). One practical method being considered for restoring connectivity is installing diversions onto major river channels. These diversions allow floodwater to enter the floodplain, essentially reconnecting previously utilized wetland areas (Mitsch et al., 2001). Mitsch et al. (2001) describe and suggest diversions in the Mississippi River as one way to help reduce nitrogen loads to the Gulf of Mexico in addition to the potential for land aggradation. These diversions could reduce nitrate in flood flow by upwards of 10% to 15% (Mitsch et al., 2001). Specifically for the Caernarvon diversion in Louisiana, Lane et al. (1999) found it effective for nitrogen and sediment retention. Depending on overall vision for the area, they even pose the possibility of increasing flow through the diversion into the wetland due to the benefits already realized (Lane et al., 1999).

On the other hand, improving connectivity may not be best suited in all circumstances. Van den Brink et al. (1993) suggest the main focus of restoration for the Lower Rhine and Meuse (heavily polluted rivers) be reducing pollutant loadings rather than floodplain connectivity. The reasoning was that if water quality is extremely poor in the area, increasing connectivity may result in decreased water quality of lentic waters near the river system. Strauss et al. (2006) note the potential denitrification benefits of increasing connectivity in backwater areas on the Mississippi River, however warn that decreased residence time and other geomorphic alterations may simultaneously occur, actually limiting nitrogen processing.

3. Methods

3.1 Study Site

This study was focused within the Mahantango Creek watershed located in eastern central Pennsylvania in the Valley and Ridge physiographic province Figure 3.1. This creek drains directly into the Susquehanna River and then into the Chesapeake Bay. The Mahantango Creek watershed (starting at the outlet into the Susquehanna River) is approximately 425 km^2, with stream orders ranging from 1-4. The watershed is primarily agricultural (40%) and agricultural (52%), with only 7% developed area. This watershed was chosen because of long-term flow record (~74 years) and other data availability, including high resolution (~3 m) DEM data. According to the USGS gauge near the outlet, the growing season (May-October) accounts for approximately 32% of the total annual flow, whereas the non-growing season represents approximately 68% of the total flow.

The Chesapeake Bay has been the focus of numerous water quality and eutrophication studies, with scientists seeking solutions to improve the Bay ecosystem. Improvement of water quality and reduction in nutrient delivery to downstream areas are the tenets in reducing eutrophication, which is the "essential core" for the Bay's improvement. Also, Mahantango Creek is one of many tributaries to the Susquehanna, the Bay's largest input (Chesapeake Bay Program, 1997). Thus, this watershed is an example where significant pollutant sources located, identified, and estimated for their potential impact on downstream water quality based on the connections between the main channel and floodplain. More improvement in Susquehanna water quality could be had on top of the improvements in nitrogen, phosphorus, and suspended solids seen by Zhang et al. (2013) over the past few decades.

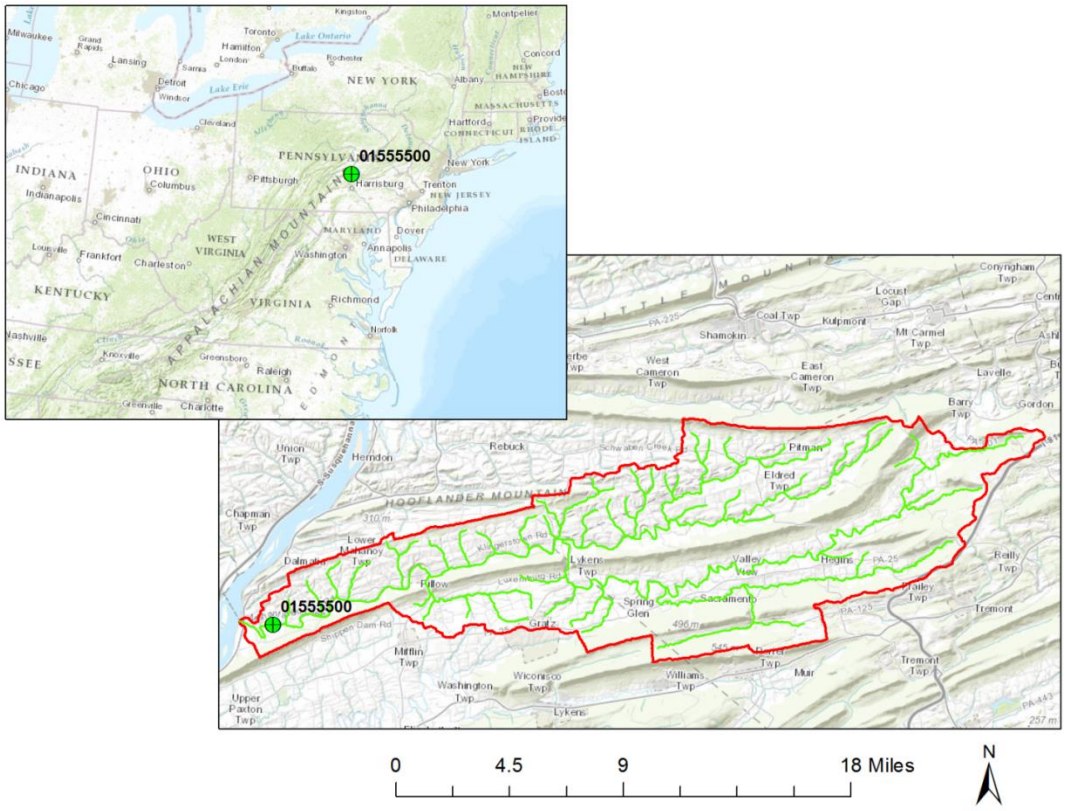

Figure 3.1. Mahantango Creek study site located in east central Pennsylvania (Topographic map from ArcGIS, 2012).

3.2 Approach

For the purposes of this study, connectivity was defined in terms of both spatial inundated area and time of inundation duration. For each time step when the flow was greater than bankfull, the associated inundated area was multiplied by the time step, which for this study was 1 day, yielding units of [m^2*days]. This was summed across the entire flow record, and properly normalized to an annual basis.

In order to calculate this connectivity metric, however, flow (a hydrograph), as well as relationships between inundated area and flow (A_i vs. Q) were needed at each location (Figure 3.2). In order to develop the hydrograph, a combination of regional regression techniques, as well as QPPQ methodology were used. Regional regression was used to develop a flow duration curve (FDC) at each ungauged location. Then, that FDC, in combination with a nearby gauge's long-term flow record and calculated FDC were used in the QPPQ method to predict the flow at

each ungauged location in the watershed. The A_i vs. Q relationship came from relationships between inundated area and stage (A_i vs. h) as well as flow and stage (Q vs. h). The Ai vs. h relationships came from GIS flood simulations, while a Manipulated form of Manning's equation was used to develop the Q vs. h relationships. Each of the primary operations in Figure 3.2 is described in detail in the following sections. A comprehensive detailed description of the analysis procedures, including GIS models and MATLAB scripts used can be found in Appendix A.

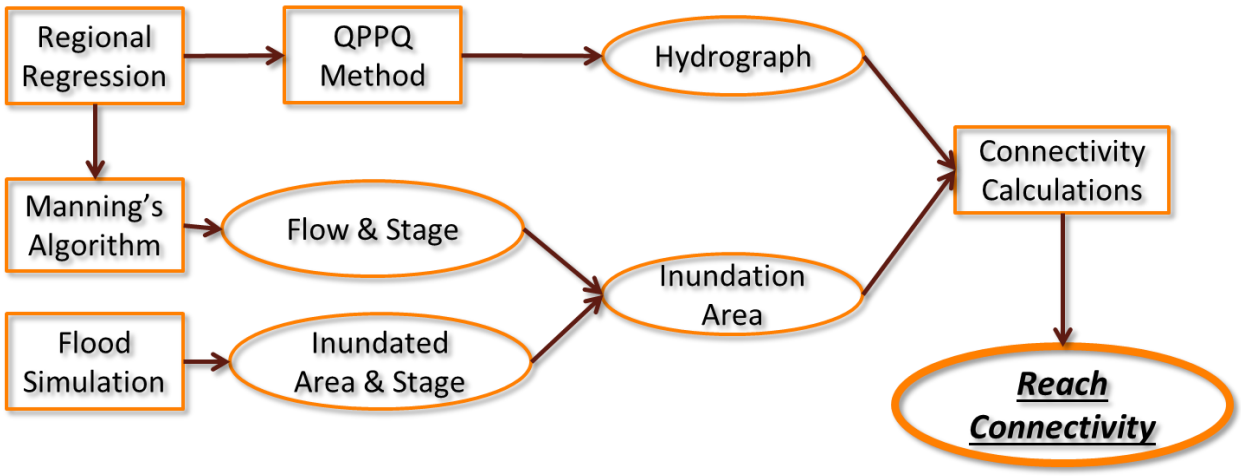

Figure 3.2. General algorithm used to develop the quantitative connectivity metric values.

3.3 Regional Regression

The primary purpose of regional regression in this study was to use watershed characteristics to predict 14 flow percentiles (99, 90, 70, 50, 40, 30, 25, 20, 10, 5, 2.5, 1, 0.1 and 0.01 %) to form an FDC at each ungauged location. Regional regression equations were developed for each of the 14 flow percentiles using watershed attributes from USGS gages in the surrounding area. Gages for the regression were selected based on a 100 mile search radius from the centroid of the Mahantango Creek watershed, a minimum 20-year record length, and a location within the same physiographic region as the study watershed. These criteria yielded 72 available gages, six of which were used for model validation, including the Mahantango Creek

outlet gage (USGS 01555500), resulting in 66 gauges for regression analysis (Figure 3.3). To make the range of drainage areas smaller and more homogeneous to the drainage area of Mahantango Creek, a factor of 2.5 was used to eliminate gages from the regression with a drainage area 2.5x above or below that of Mahantango Creek, yielding a final 28 gauges that were used in the regression.

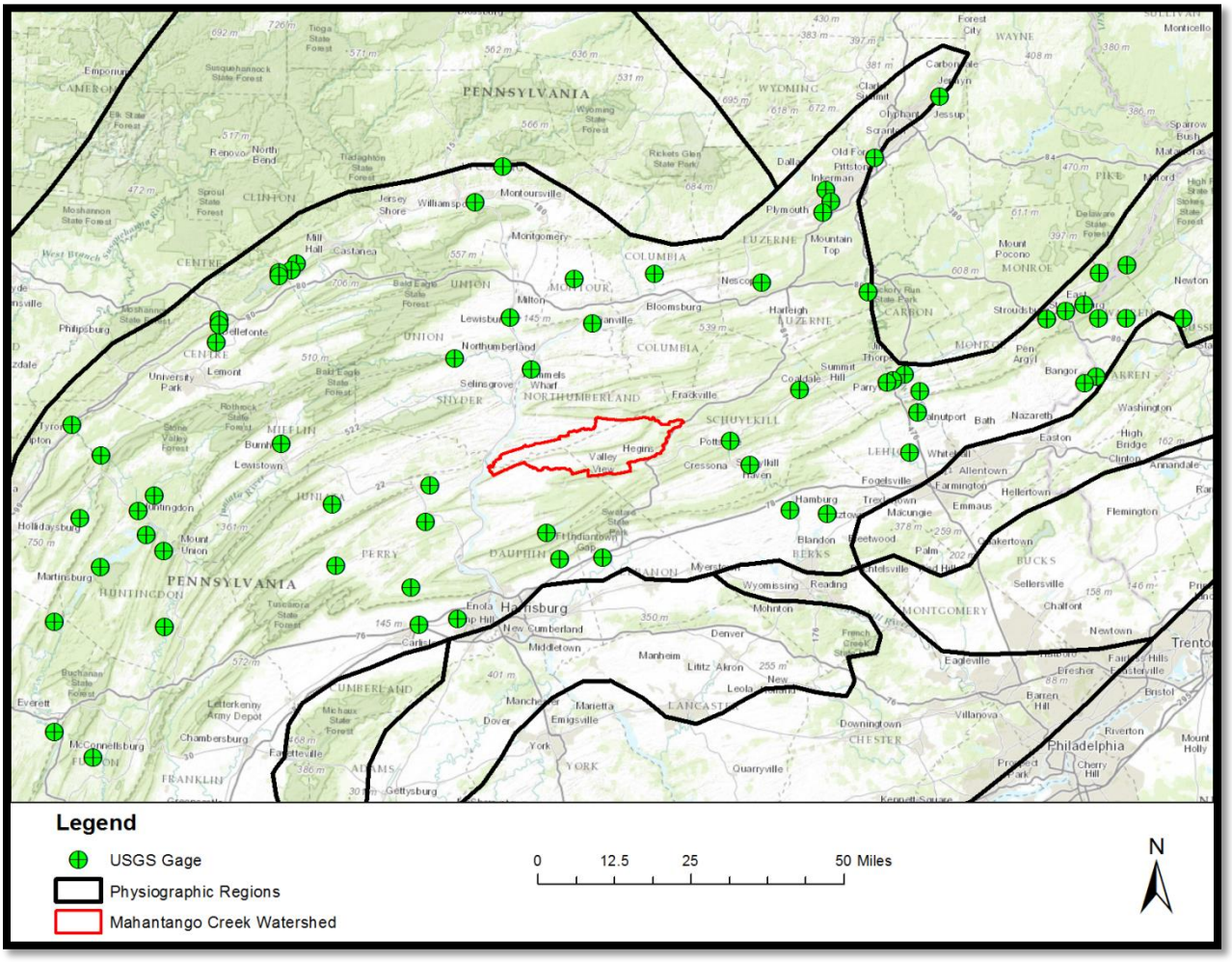

Figure 3.3. USGS stream gages used in the regression analysis prior to screening by drainage area.

GIS models were used to collect the watershed attributes for both the known gages (to build the regression equations) as well as the unknown reach locations (to which the regression equations would be applied). Attributes collected and used in the regression analysis included drainage area, mean hydrologic soil group (HSG), percent agricultural cover, percent developed area, percent agricultural area, minimum elevation, elevation range, mean slope, 2-yr, 24-hr maximum precipitation, 2-yr, 24-hr mean precipitation, 100-yr, 24-hr maximum precipitation, and 100-yr, and 24-hr mean precipitation.

In a similar fashion to the regression equations built by Ssegane et al. (2013), a three-variable power-function model was selected to predict the 14 flow percentiles across all the ungauged reaches. A uniform set of three variables was selected from the 16 independent variables using a stepwise algorithm. A three-variable model was chosen over a two or four variable model because it was found to have the best prediction of the Mahantango Creek outlet gauge FDC while maintaining model simplicity. A uniform set of variables was used to predict flows across all percentiles in order to maintain proper scaling to yield a strictly monotonically increasing flow duration curve: a problem discussed by Archfield et al. (2010). Other studies (Roland and Stuckey, 2008; Capesius and Stephens, 2009) have used a uniform set of variables to predict various magnitude peak flows. The three variables chosen for the model were based on the number of times those variables were found significant across all flow percentiles (significance frequency).

3.4 QPPQ Method

The long-term hydrograph for each reach was developed via the QPPQ method (as seen in Hughes and Smakhtin (1996) and developed by Fennessey (1994)) using the predicted FDC, as well as a nearby gage's long-term hydrograph and calculated FDC. The concept of the QPPQ method begins by determining the flow percentile (P) for every daily flow measurement (Q) on the known hydrograph (Q→ P). Then, the primary assumption associated with the method is that the flow percentile at the nearby gage occurs at the same time (same day) as at the ungauged location (P→ P). Finally, flow at the ungauged location (P→ Q) is found via the flow percentile and the ungauged FDC, as predicted from the aforementioned regression analysis (Figure A.10).

Thus, at the end of the QPPQ method, a long-term hydrograph had been developed for each reach, and could provide the basis for the connectivity metric calculations, as described later.

Among many methods of determining the "nearby" gage from which the timing of flows were developed, this study utilized the nearest gage to the unknown location on a watershed centroid-to-centroid basis. This was used when validating and corroborating the model. For actual application on the individual reaches within the Mahantango Creek watershed, the outlet gage was selected for use in the QPPQ method. This was done due to the relationship between the overall Mahantango Creek watershed flow characteristics and the subwatersheds located within it.

3.5 Flood Simulations

As aforementioned, artificial flood simulations in GIS were used to develop the relationships between inundated area and stage, as well as cross-sectional area and stage (A_i vs. h and A_x vs. h, respectively). This study focused on floodplain inundation once the river stage exceeded bankfull. Because of this, the flood simulations started from the approximate location of bankfull. Thus, one of the first tasks was to develop a water surface polygon for the entire flow network, which represented the horizontal area occupied by the stream system immediately prior to floodplain inundation.

The surface water polygon was developed via a combination of filtered DEMs. The five different DEM filters included the ArcGIS curvature tool, SAGA GIS (SAGA, 2007) Laplace with majority filter, high pass with curvature filter, catchment area mass flux module, and DTM slope-based filter. The DEM for the Mahantango Creek watershed was run through each of the five filter/filter combinations, and then the output grids were examined and combined to compile the most representative surface water area. Once converted to a polygon, aerial imagery and manual editing was used to correct the water surface as necessary. This polygon, although representative of the surface area, was not representative of the approximate elevation on the DEM to which artificial flood stages could be added in order to develop inundated area relationships. One challenge with associating an elevation to the bankfull water surface was that in a natural system, the surface is sloped, but for modeling purposes, a flat surface was needed. To account for this, the water surface was split into ~20 m sections and assigned the elevation of

the maximum DEM cell within each section. Thus, a "stair-step" pattern was created of short, flat surfaces, increasing in elevation from the outlet to the headwaters. The water surface was split via cross-sectional lines created using tools from Geospatial Modelling Environment (GME) (Beyer, 2012) and ArcGIS.

Once the water surface was properly prepared, artificial floods were simulated in order to develop the relationship between inundated area and stage for each reach. Artificial stages ranging from 0 to 8 m at 0.5 m increments were added to the bankfull water surface elevation to create a "flood surface". Then, the area inundated according to the DEM's elevation relative to the flood surface (DEM < flood surface, → inundated) was calculated and summarized based on the contributing area to each reach. Additionally, at the cross-section most representative of each reach as a whole, the cross-sectional area was calculated by multiplying the flood depth (flood surface-DEM elevation) by the analysis grid cell width (3 m). Thus, at the end of the flood simulations, relationships were developed for A_i vs. h as well as A_x vs. h, both necessary for future analysis.

3.6 Manning's Algorithm

The next relationship needed to calculate connectivity was that between flow and stage (Q vs. h) at each reach. Given the large scale of this study, little detailed topographical information was known for each reach, thus eliminating the potential use of a hydraulic model to develop the stage-discharge relationships. Manning's equation was manipulated and used to relate river stage to flow, thus finding a stage for each flow percentile, as calculated in the regional regression steps. First, it was assumed that the W:D ratio of flood waters was high enough such that the hydraulic radius (R, m) could be estimated by the water depth (stage, h, m) (Equation 3.1). Next, the relationship between cross-sectional area and stage (A_x vs. h) could replace the A_x in Manning's equation. Then, the equation was rearranged to have all functions of stage on one side, and flow (Q, m^3/s), Manning's n, and slope (S, fraction) on the other (Equation 3.1). A root finding method was then used to calculate a stage for each given flow value on the predicted FDC for each reach, thus yielding the stage-discharge (Q vs. h) relationship for every reach in the Mahantango Creek watershed.

Equation 3.1. Manning's manipulation for stage-discharge relationship calculations.

$$Q = A_x V = A_x R^{\frac{2}{3}} S^{\frac{1}{2}} \left(\frac{1}{n}\right) = A_x h^{\frac{2}{3}} S^{\frac{1}{2}} \left(\frac{1}{n}\right) = f(h) * h^{\frac{2}{3}} S^{\frac{1}{2}} \left(\frac{1}{n}\right) \rightarrow \frac{Qn}{S^{\frac{1}{2}}} = f(h) h^{\frac{2}{3}}$$

3.7 Connectivity Calculations

After the Manning's algorithm and flood simulations were performed, the relationship between inundated area and flow (A_i vs. Q) was ready for incorporation within the connectivity calculations. Using the long-term hydrograph for each reach (as calculated in the QPPQ method) and the relationship between inundated area and flow for each reach, every flow value greater than bankfull was converted to an inundation area. A combination of literature values (Table A.4, Appendix A), was used to assume that bankfull flow was equivalent to the 2.5% daily mean flow exceedance percentile. The calculated inundation areas (one for each day in the long-term flow record where the mean daily flow was above bankfull) were then multiplied by the time step (1 day), summed across the entire flow record, and normalized to an annual basis, resulting in the annual connectivity metric for each reach. Seasonal connectivity metrics were developed by summing the inundation area values across each season and normalizing accordingly, with the growing-season defined by the months of May through October and the non-growing season defined as November through April. The seasonal metrics were then examined for their relationships with drainage area, while the annual metric was analyzed for relationships with various watershed attributes including drainage area, reach slope, developed land use, agricultural land use, forested land use, and the 100-yr, 24-hr precipitation depth.

4. Results

The results from the spatiotemporal floodplain connectivity model are presented under three main categories: regional regression, inundation area, and connectivity. The regional regression results include the regression statistics as well as the development of the final model to predict the 14 percentile flow values on the FDCs throughout the stream network. The inundation area results show the relationships between normalized inundation area and various other factors including drainage area, reach slope, etc. Finally, the frequency of connection across the network is presented. Any coefficient of determination (R^2) values presented in this study is reflective of the fit between the linearized data with associated linear model (rather than the actual data with the exponential, logarithmic, or power function model). Unless otherwise noted, the p-values presented are the probability values associated with the model significance and the F-statistic.

4.1 Regional Regression: Obtain hydrographs across stream network

The temporal GIS and temporal MATLAB processes both produced output valuable for examination in this study. First, the watershed attributes as developed in the temporal GIS model (precursor to the regional regression analysis) are summarized and presented. Second, the variable selection results for the FDC prediction regression equations are presented. Lastly, the other regression results including the independent variables used in the regression as well as the regression equation coefficients are examined. Corroboration of the ungauged hydrograph development (including ungauged FDC prediction) was also performed.

4.1.1 Watershed Attributes

The first main step of the regression script was to prepare the gauge watershed attributes into an independent variable matrix. The gauges' attributes were summarized as described in the method section prior to performing the regression analysis. The full matrix of watershed attributes can be found in Table E.1 (Appendix E). There were a total of 66 gauges that were analyzed, but after eliminating those outside the 2.5 times the drainage area threshold, 28 gauges remained in the regression analysis. The gauges used in the regression had drainage areas

ranging from 169 km^2 to 928 km^2, which was within the drainage area factor (2.5) of the Mahantango Creek USGS gauge watershed area (419.8 km^2). The upstream watershed land use at each gauge site used in the regression was primarily forested, with percent forest land cover ranging from 37% to 85%, with a mean of 66%. There was a wide range of agricultural influence in the regression gauge watersheds, ranging from nearly no agricultural input (0.04%) to approximately half the watershed area (50.2%). The goal of dealing with relatively low human-impacted watersheds was met, with developed area ranging from 3-26% with an average of 11% developed area. In terms of precipitation on the gauge watersheds, the depths had a fairly even distribution, ranging from 5.8-8.4 in, with a mean of 7.0 in.

The output from the summarization GIS model included the watershed attributes and characteristics associated with each reach in the study area (Mahantango Creek). These attributes were summarized and processed in the temporal MATLAB model (regression script). The full tabular dataset of the individual reach watershed attributes can be found in Table E.2 (Appendix E). Reach drainage areas ranged from 0.5 km^2 to 425 km^2 with a mean of 54.4 km^2. Most of the reaches in the watershed remained unaffected by human impact, with developed area ranging from 1.5% to 21.2% (mean of 7.3%). Both forested and agricultural area distributions had similarly wide ranges of values. Forested area ranged from 0.7% to 96.3% with a mean of 47.9%. Agricultural area, on the other hand, also had a wide range from 0% to 88.0% and an average of 44.2%. As far as the 100-year, 24-hr storm precipitation depth, the values ranged from 6.9 to 7.9 in (mean of 7.4 in).

4.1.2 FDC Prediction Equation Regression Model Variable Selection

After performing the stepwise regression in MATLAB, the significance frequency (number of times the variable was found to be significant across the 14 flow percentile equations) was calculated for each of the 16 independent variables tested (Table 4.1). The full binary matrix (shown as the transpose of the actual output for display purposes) of which independent variables were found to be significant in which regression equations can be found in Table E.3 (Appendix E).

Table 4.1. Significance frequency of independent regression variables.

Variable	Significance Frequency (#)
DA	14
HSG	0
PCT_FOR	2
PCT_DEV	8
PCT_AG	3
ELEV_MIN	0
ELEV_RNGE	0
SLOPE_MEAN	1
2yr24MAX	0
2yr24hrMEAN	0
100yr24MAX	0
100yr24MEAN	6
HSG*100max	0
HSG*100mean	1
HSG*2yrmax	0
HSG*2yrmean	5

This output was analyzed before finalizing the rest of the regression script. Based on Table 4.1, drainage area was clearly an important variable in predicting all the flow percentiles. The next most important variable was percent developed land area ("PCT_DEV") found in 8 out of 14 regression equations. Finally, as per the three-variable model, the last variable included in the model was the mean 100-year, 24-hr precipitation depth (in*1000) for the reach watershed ("100yr24MEAN"), having been significant in 6 out of the 14 equations. The model variables were selected solely on their significance frequency. This variable selection suggests that these variables have the highest flow prediction ability across all flow magnitudes for this study area. The final regression equations for each of the 14 flow percentiles were based on drainage area, percent developed area, and the mean 100-year, 24-hr precipitation depth. The final linearized regression model with the variable transformations shown can be found in Equation 4.1. The regression coefficients changed for each flow percentile equation.

Equation 4.1. Linearized FDC prediction model

$$\log(Q_{pp} + 1) = a_{pp} + b_{pp} * \log(DA) + c_{pp} * \log(PCTDEV + 1) + d_{pp} * \log(100yr24MEAN)$$

Where:

Q= Flow value (ft^3/s);

pp= Flow percentile (%);

DA= Drainage area (m^2);

PCTDEV= Percent developed area (%);

100yr24MEAN= Mean 100-year, 24-hr precipitation (in*1000); and,

a, b, c, d= regional regression coefficients.

Equation 4.1 can be transformed to a non-linearized form to show the underlying power-relationship of the variables (Equation 4.2).

Equation 4.2. Overall regional regression model

$$Q_{pp} = 10^{a_{pp}} DA^{b_{pp}} (PCTDEV + 1)^{c_{pp}} 100yr24MEAN^{d_{pp}} - 1$$

4.1.3 Final Regression Results

The independent variable matrix (as log-transformed) was narrowed to only the variables to be kept in the final regression (best three previously described variables). These independent variable values used to build the regression equations can be found in Table E.4. (Appendix E). Note the first column containing values equal to one, which allows for the calculation of a constant within the regression. After the final regression took place, the models' statistics were

summarized in Table 4.2. These model statistics were based on the 14 equations built from the known regression gauge data.

Table 4.2. Regression results for the 14 FDC prediction models.

Exceedance Probability (%)	R^2	p-Value
99	0.65	1.24E-05
90	0.72	9.35E-07
70	0.77	9.48E-08
50	0.86	2.94E-10
40	0.89	9.94E-12
30	0.92	2.83E-13
25	0.93	5.51E-14
20	0.94	6.03E-15
10	0.96	1.80E-16
5	0.95	5.14E-16
2.5	0.95	1.65E-15
1	0.95	2.17E-15
0.1	0.90	2.13E-12
0.01	0.78	3.95E-08

Based on Table 4.2, all 14 regression equations had p-values less than 0.05, thus indicating the models (regression equations at each flow percentile) were statistically significant. Also, all of the R^2 values were greater than or equal to $R^2 = 0.65$, indicating adequate model fits. The actual output coefficient values (Table 4.3) (from the MATLAB "regress" function) follow the form presented in Equation 4.2, with "a" representing initial constant coefficient, "b" representing the coefficient for drainage area, "c" for the developed area, and "d" for the 100-year, 24-hr mean precipitation depth.

Table 4.3. Regression coefficients for FDC prediction equations.

Exceedance Probability (%)	a (Constant)	b (Drainage Area)	c (Developed Area)	d (100-yr, 24hr precip.)
99	-6.96	0.89	0.98	-0.09
90	-9.97	0.90	0.78	0.80
70	-12.21	0.93	0.46	1.48
50	-11.66	0.96	0.22	1.39
40	-11.37	0.98	0.14	1.32
30	-11.22	1.02	0.07	1.25
25	-11.15	1.03	0.04	1.23
20	-11.19	1.05	0.00	1.22
10	-11.20	1.10	-0.08	1.19
5	-11.13	1.13	-0.13	1.16
2.5	-11.02	1.13	-0.17	1.16
1	-11.32	1.16	-0.24	1.24
0.1	-11.54	1.23	-0.38	1.27
0.01	-11.08	1.35	-0.33	0.94

Using the coefficients in Table 4.3, Equation 4.1, and the transformed watershed attributes for the ungauged reaches in the study watershed, the FDCs for every reach in the watershed were calculated (Table E.5, Appendix E). These FDCs were later used to develop the full ungauged hydrographs and the final connectivity metric.

4.1.4 Hydrograph & FDC Corroboration

As described in the methods, six gauges were treated as "unknown" locations for purposes of model and overall methods corroboration. Thus, their attributes were also summarized (Table E.6, Appendix E) and applied to the previously developed regression equations. Then, the actual hydrographs from the gauges were compared to the predicted hydrographs (developed via predicted FDC and nearby gauge hydrograph). The Nash-Sustcliffe Efficiency (NSE) of the full hydrograph, NSE of the 14 points on the flow duration curve, and

the NSE of the upper 50 percentile flows were examined to evaluate the validity of the ungauged hydrograph prediction (Table 4.4).

Table 4.4. Hydrograph and FDC prediction validation statistics.

Site Number	DA (km^2)	Site Number (Nearest Gauge)	DA (Nearest Gauge, km^2)	DA_{gauge} / DA_{near}	Dist. (km)	NSE	RMSE	NSE_{FDC}	NSE_{HIGH}
1470779	181.6	1573000	872.8	0.2	20.1	0.64	68.7	0.65	0.66
1554500	141.2	1468500	344.5	0.4	26.8	0.60	34.2	0.88	0.65
1555500	419.8	1568500	58.3	7.2	18.0	0.43	361.5	0.72	0.46
1563000	22.6	1567000	8686.8	2.6E-03	6.0	-0.42	601.7	-0.50	-0.41
1570500	62339.5	1552000	1126.6	55.3	30.7	-8.30	34427.3	-69.94	-7.94
1614500	1300.0	1564500	530.9	2.4	32.2	0.45	430.0	0.59	0.47

There was a wide range of validation statistics based on the values found in Table 4.4. Sites 156300 and 1570500 had extremely poor hydrograph prediction based on the NSE values. Site 1470779 had the best hydrograph prediction (NSE=0.64), but two other sites had better FDC prediction. The study watershed outlet gauge (USGS Site 1555500) had a full hydrograph NSE of 0.43, and an NSE of the FDC prediction of 0.72. On all the sites, there was a slight increase in NSE from the full hydrograph to flows above median flow (50% on the FDC).

To begin qualitatively examining the hydrograph and FDC prediction results, the FDCs of two validation gauges were examined to determine whether high or low flows were predicted better. The first validation gauge which had the FDC qualitatively examined was the outlet gauge (USGS 01555500) (Figure 4.1).

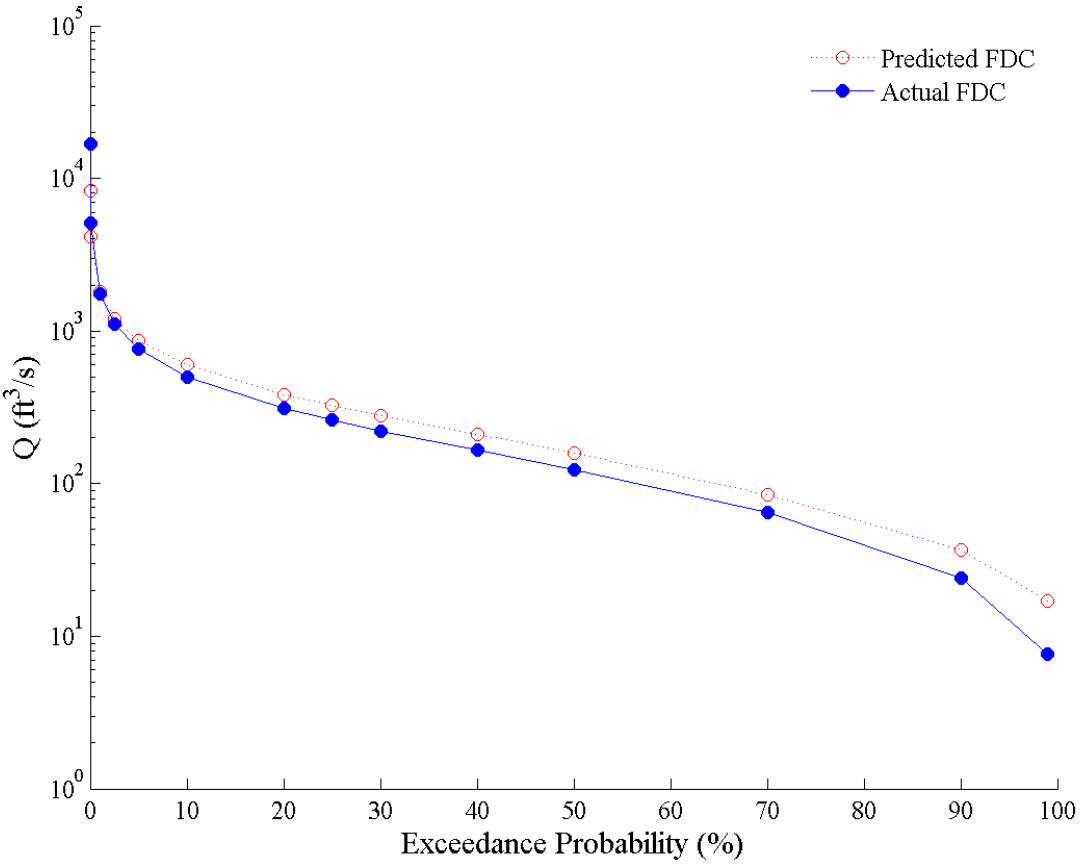

Figure 4.1. Predicted and actual FDC for the Mahantango Creek outlet gauge, USGS 01555500

It appears that the predicted FDC for the outlet gauge followed relatively close to the actual FDC for most of the high flow range. Prediction of these higher flows was most important for this study since these are the flows which could cause flooding. For low flows, the predicted FDC tended higher than the actual FDC. As a comparable example of extremely poor prediction (NSE_{FDC}= -69.9), the actual and predicted FDC for another gauge, USGS 01570500, was also examined (Figure 4.2).

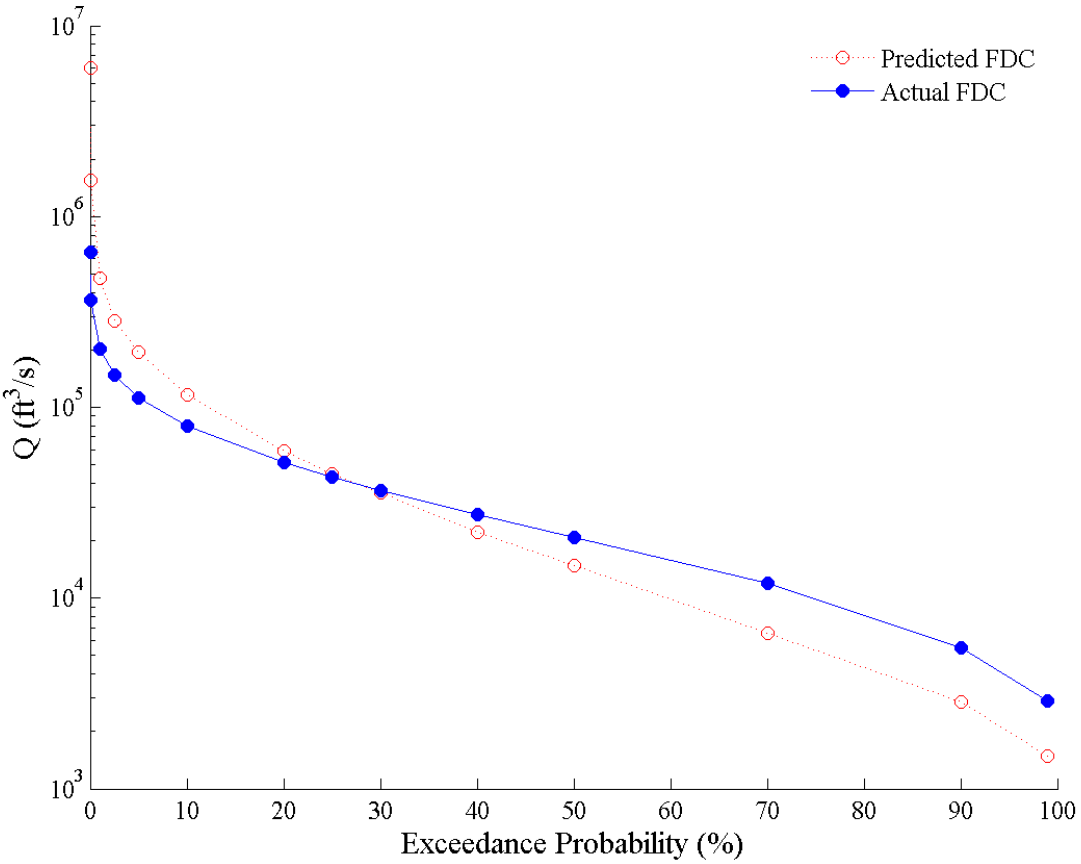

Figure 4.2. Predicted and actual FDC for USGS 01570500 gauge

For this gauge, the high flows were severely over predicted, while the low flows were under predicted. Only those flows around the 25% exceedance were relatively accurate. Because the final hydrograph prediction was used in the development of the connectivity metric, it was important to examine the prediction accuracy of the final full hydrographs in addition to the FDCs as predicted via the regression. This allows for analysis of the methodology used to develop the predicted hydrograph from the nearby gauge hydrograph, nearby gauge FDC, and predicted location FDC. Portions of the predicted and actual hydrographs for the outlet gauge (USGS 0155500) are shown in Figure 4.3. Generally, the predicted hydrograph follows the actual hydrograph with the same trends. The peaks in December and January were relatively close in following the actual flow pattern. The peak in November, however, was severely under

predicted. In general, the peaks were under predicted and baseflow was over predicted. This correlates to trends in the FDC prediction (Figure 4.1). As previously mentioned, the NSE for this hydrograph was 0.43, and the upper 50 percentile flows had an NSE of 0.46.

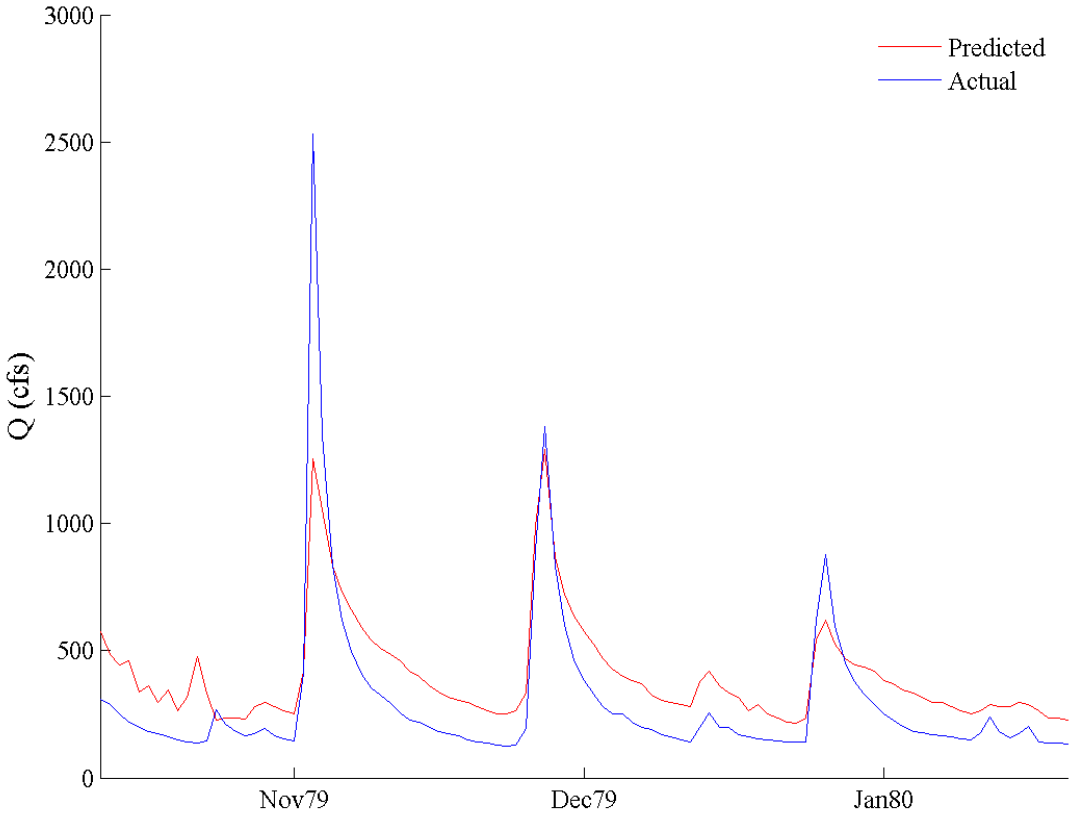

Figure 4.3. Predicted and actual hydrograph portions for the outlet gauge of the study watershed (Mahantango Creek, USGS 01555500).

To compare, the predicted and actual hydrographs of site USGS 01570500, which had poor validation results, are shown in Figure 4.4. The NSE of this gauge's hydrograph prediction was -8.3, with an upper 50 percentile flows having an NSE of -7.9. The peaks were highly over predicted and baseflow was under predicted. Once again, this correlates to the same trends as

shown in the FDC for this gauge (Figure 4.2). Underlying these predictions, however, was the FDC prediction, which was also extremely poor at an NSE of -0.5.

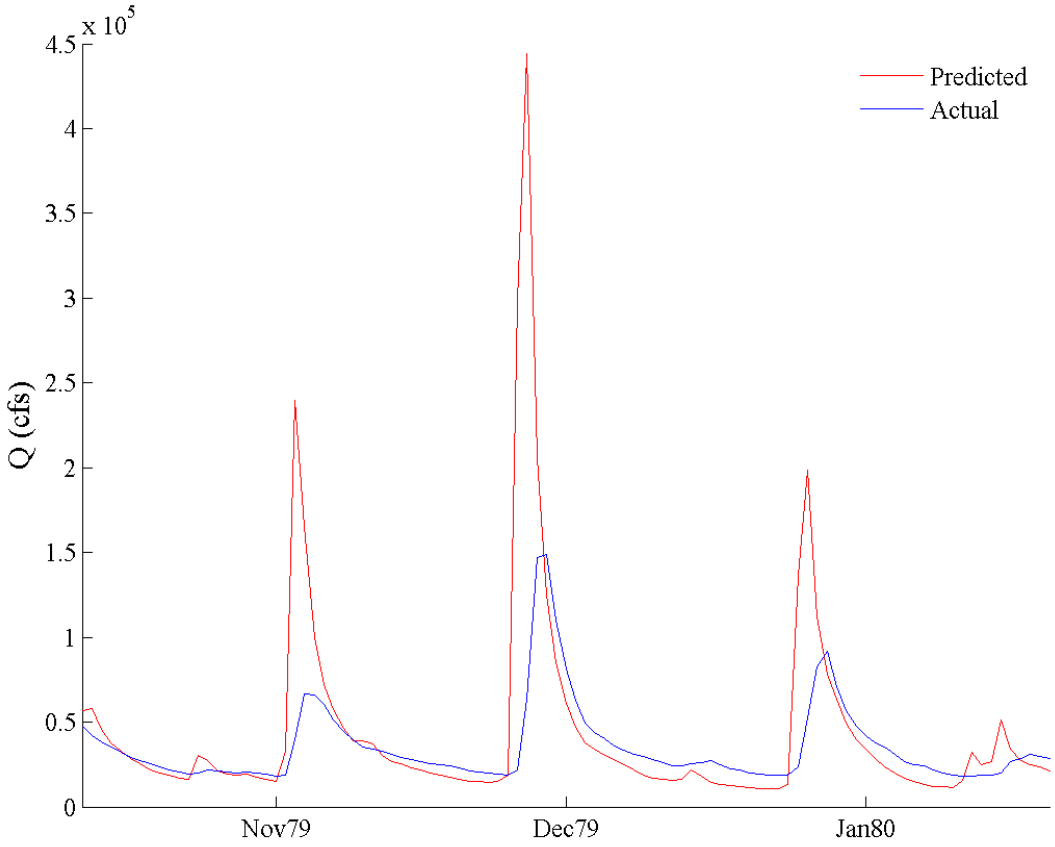

Figure 4.4. Predicted and actual hydrograph portions for USGS 01570500, which had notably poor validation statistics.

The FDC and hydrograph portion qualitative analysis was important for relating the validation statistics to what was actually happening in a physical sense. The observations confirmed the calculated validation statistics.

4.2 Inundation Area

The spatial component of the overall model yielded results regarding inundation patterns for each reach. These patterns were analyzed for relationships with drainage area as well as differences among storm sizes. First, the inundation area values were normalized to the length of the reach. This resulted in inundation area for a given storm percentile per meter of stream length. This allowed for analogous comparison between reaches of different lengths. Figure 4.5 shows the normalized inundation area versus drainage area to examine the effects on flooding patterns.

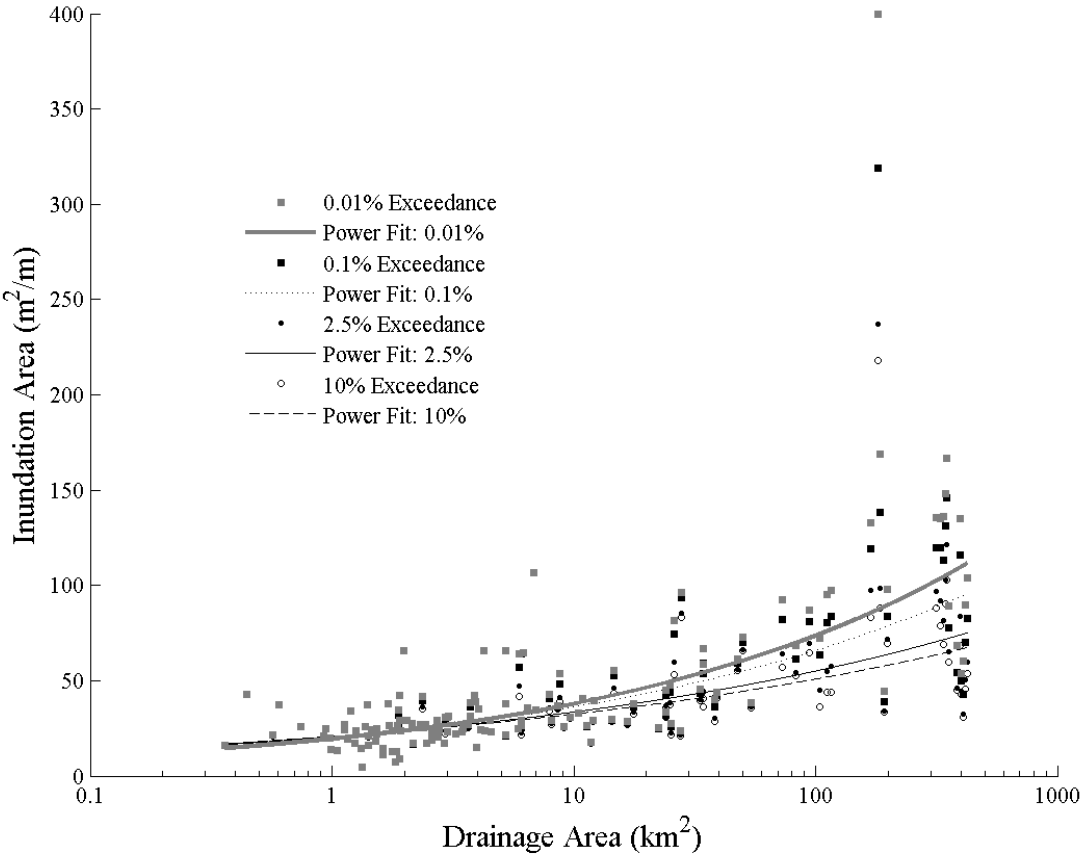

Figure 4.5. Inundation area per unit of stream length versus drainage area.

Each of the trendlines in Figure 4.5 represents a fit to a power function of the form: $y = ax^b$. As would be expected, the largest storm event (0.01% exceedance) had the highest inundation areas. Meanwhile, the smallest storm displayed (10% exceedance) had the lowest inundation areas. The R^2 of the trendlines ranged from 0.60 for the largest storm (0.01%), and declined with 0.55, 0.47, and 0.43 for the 0.1%, 2.5%, and 10% storms, respectively. For all storms, the inundation area increased as drainage area increased. Thus, larger river systems have larger flooded area per length of stream than smaller streams, given the same storm percentile.

Besides normalizing the inundation area to the stream length, it was also normalized to the total floodplain size. The maximum floodplain size was assumed to be the inundation area extent of the largest modelled flood event, which for this study was the 0.01% storm. Thus, each inundation area for each reach was divided by the inundation area for the 0.01% storm event. This resulted in fractions of floodplain coverage. This floodplain coverage fraction was then related to reach drainage area (Figure 4.6).

As with the inundation area normalized to stream length (Figure 4.5), the largest storms (lowest percentile values) have the highest floodplain coverage. The power fit trendlines clearly show the decreasing nature of floodplain coverage with drainage area. As drainage area increases, the percentage of the floodplain covered by the same storm (same flow percentile) decreases. Also, the effect of storm size changes with drainage area as well. For reaches with large drainage areas, there are large differences in the floodplain coverage for different storm events (flow percentiles). For smaller streams, however, there is little to no difference in floodplain coverage across storm sizes. For small streams in Figure 4.6, nearly the entire floodplain (coverage≈1) is inundated for all the storms displayed. Once again, the data were fit to power trendlines which had relatively constant levels of variance explanation, with each model explaining approximately 70% of the data variance. The R^2 values were 0.68 for the 0.1% storm, and 0.69, 0.68, and 0.67 for the 1%, 2.5%, and 5% storms, respectively. Thus, the power relationship for the floodplain coverage in relation to drainage area explained an adequate amount of the variance for the model.

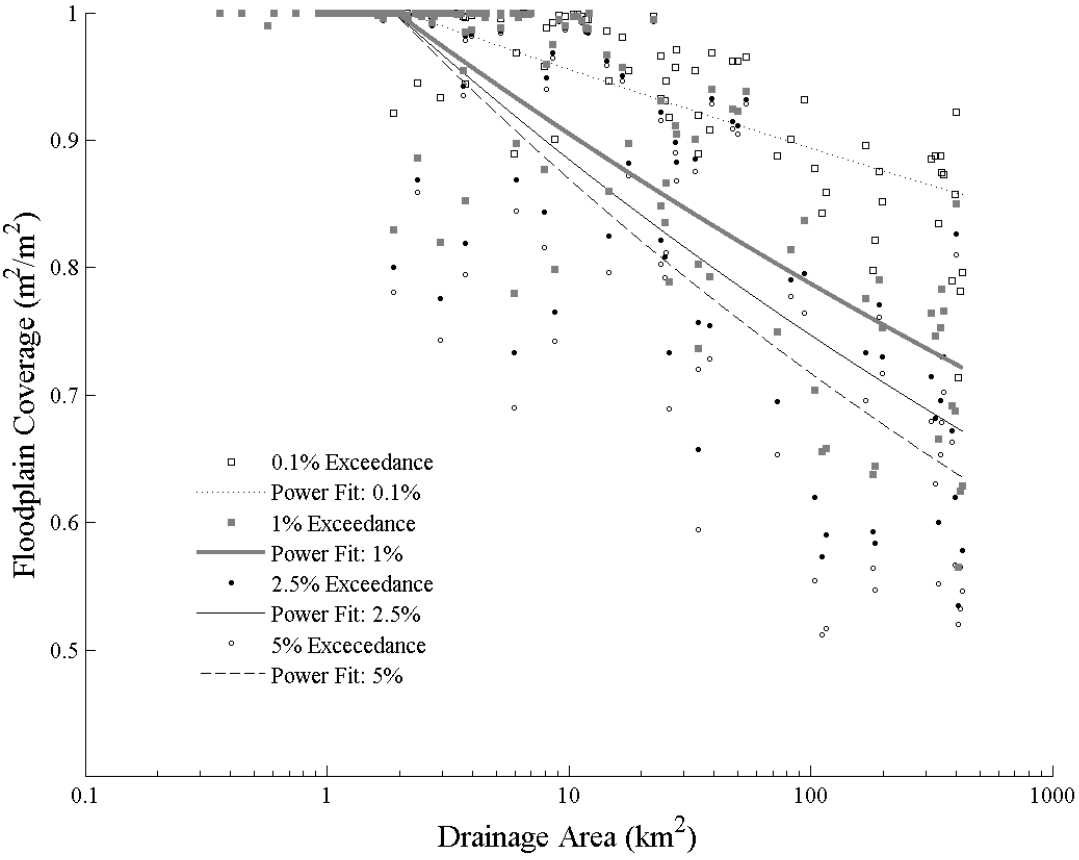

Figure 4.6. Inundation area normalized to largest inundation area, resulting in fraction of floodplain covered versus drainage area by varying storm sizes.

To examine the validity of the inundation area patterns and results, two case study reaches were analyzed and qualitatively interpreted for their inundation area patterns. First, a first order headwater reach (COMID 4518000) was presented (Figure 4.7) to examine the inundation patterns of a reach with a small drainage area (1.8 km^2). The inundation stage shown represents the inundation area for the given artificial flood simulation stage. Thus, the lower stages (darker blue) represent the area covered by the 0.5 m or 1.0 m stage flood simulations. The lighter colors (higher stages) represent the extent of the higher stage flood events. Thus, the 8.0 m stage event had the largest area, and is resembled by the light blue color.

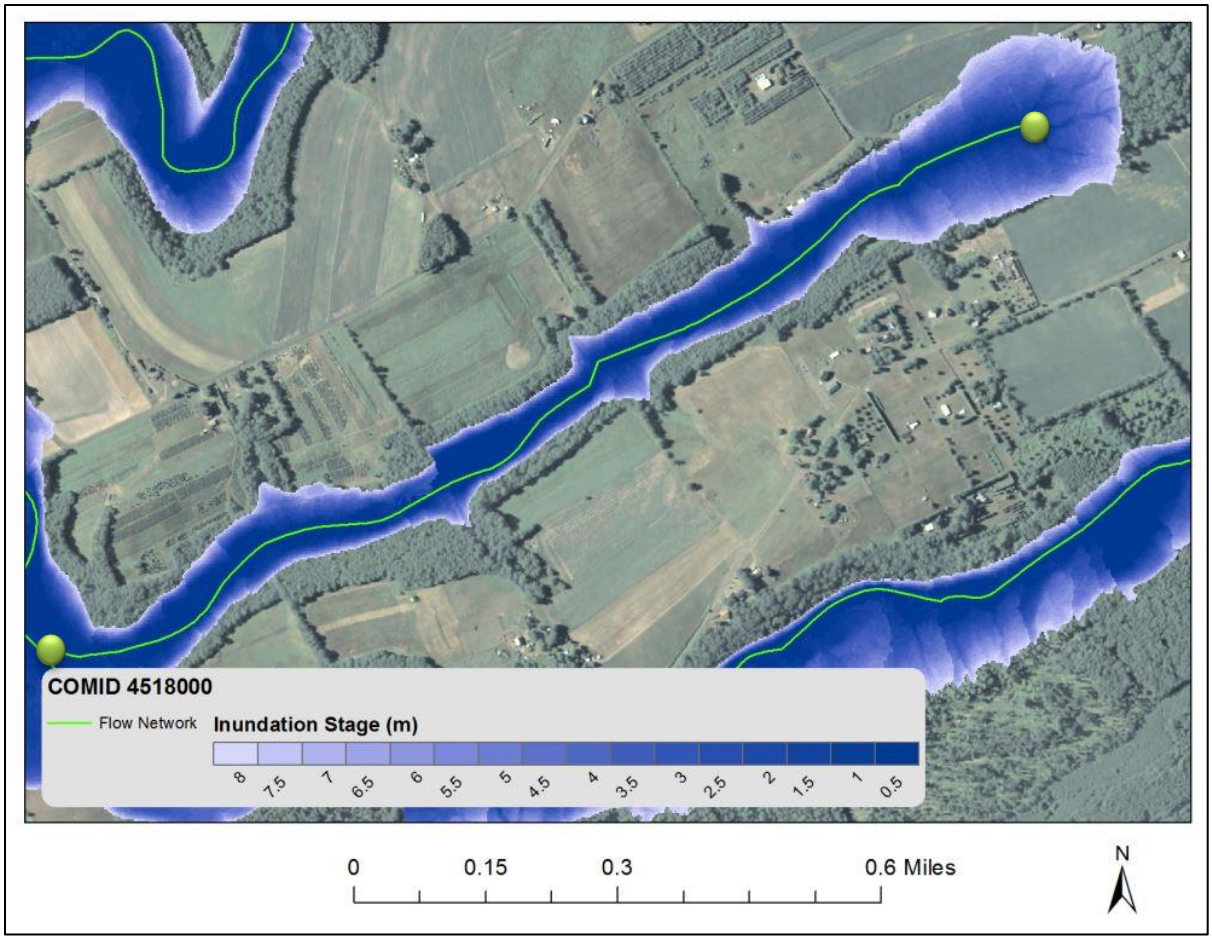

Figure 4.7. Inundation mapping for reach 4518000, exemplifying a small drainage area (Imagery from NAIP, 2010).

To compare, a reach with a large drainage area was also examined (Figure 4.8). For the large drainage area reach example, a fourth order reach (COMID 4518306) was chosen which had a drainage area of approximately 425 km^2. As with the smaller drainage area, the simulated flood inundation layers were shown to observe the spatial extent of inundation as stage increased.

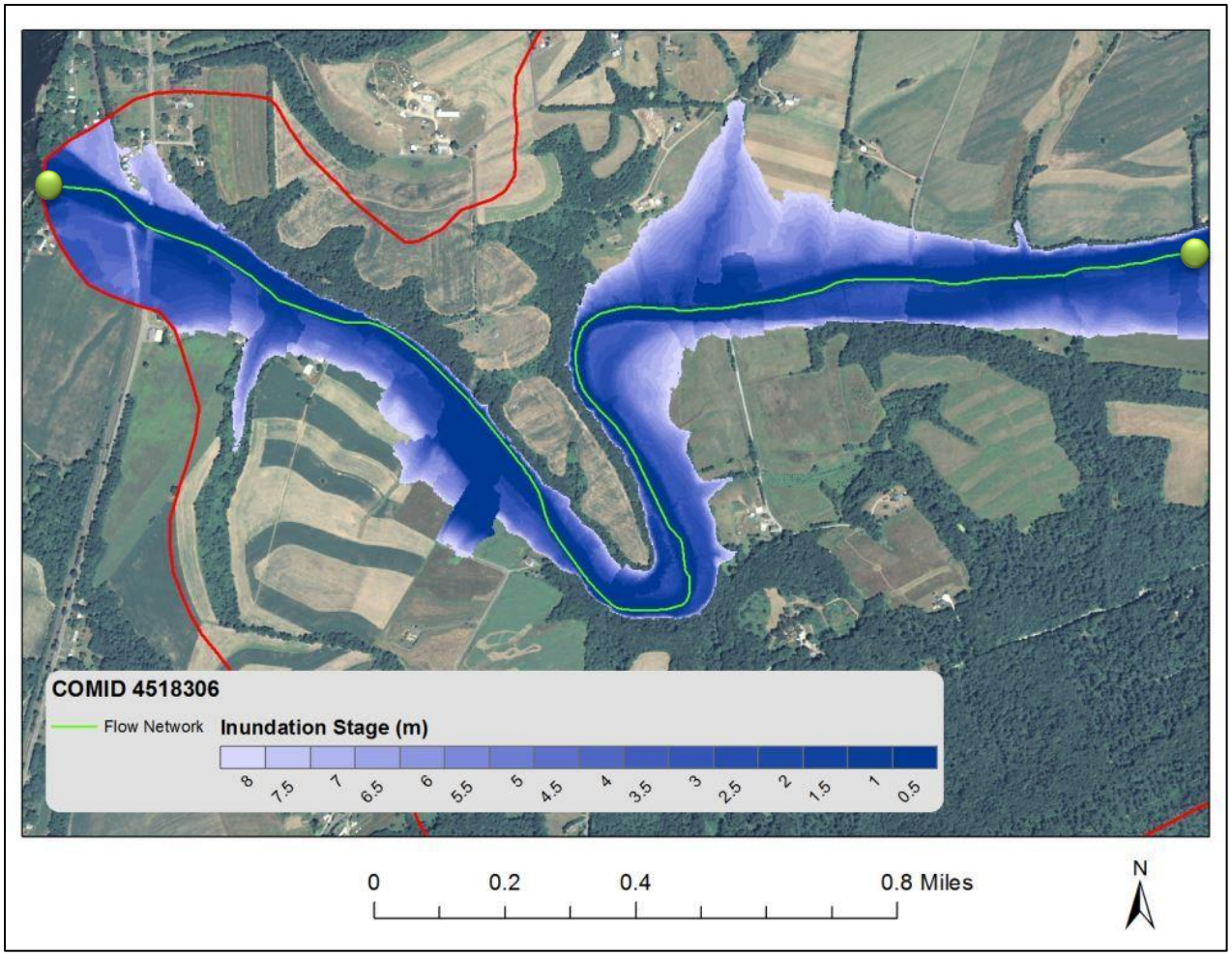

Figure 4.8. Inundation mapping for reach 4518306, exemplifying a large drainage area (Imagery from NAIP, 2010).

4.3 Connectivity

As described in the methodology, there were several connectivity-related output metrics from the MATLAB connectivity model. The primary outputs from the model included the total annual connectivity metric (m^2*days/m), the metric as summed for the growing season (m^2*days/m), and the metric as summed for the non-growing season (m^2*days/m). The annual metric was normalized by summing the connectivity across the entire flow record, dividing by the record length, and then multiplying by 365 days/year. The seasonal metrics were normalized by summing the connectivity across the days in each respective season, dividing by the total days

in each season across the whole record, and then multiplying by 182.5 (half of 365) to approximate that each season was half of a year. Also, these metrics were calculated based on inundation areas of flows greater than bankfull, which was defined as 2.5% exceedance for daily flow. In order to gain a better understanding of river-floodplain systems, these connectivity metrics were related to several watershed characteristics including watershed area, reach slope, developed area, agricultural area, forested area, and the 100-year, 24-hr precipitation depth.

4.3.1 Drainage Area

Drainage area was the primary watershed characteristic that was examined for its relationship with connectivity. First, the annual connectivity metric was related to watershed area with all reaches being considered (Figure 4.9).

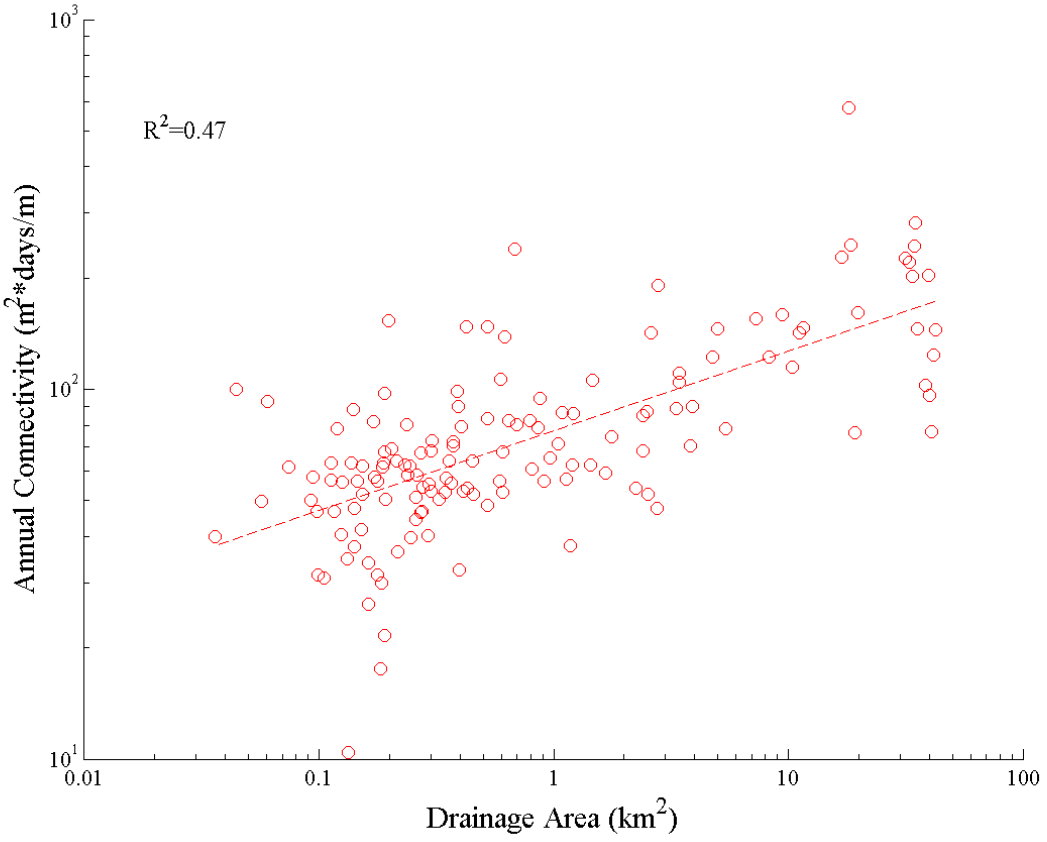

Figure 4.9. Annual connectivity as normalized to stream length related to cumulative watershed area.

The trend line shown in Figure 4.9 was the result of a power function fit (Form: $y = ax^b$), hence the linearity on the log-log plot. The p-value associated with the model was <0.001, thus signifying a significant relationship between the two parameters with α=0.05. The R^2 value of 0.47 indicates that the power fit model explains nearly half of the data variance.

Next, the same data was plotted (annual connectivity versus drainage area), but the data were separated into four series: one for each stream order represented in the study watershed (Figure 4.10).

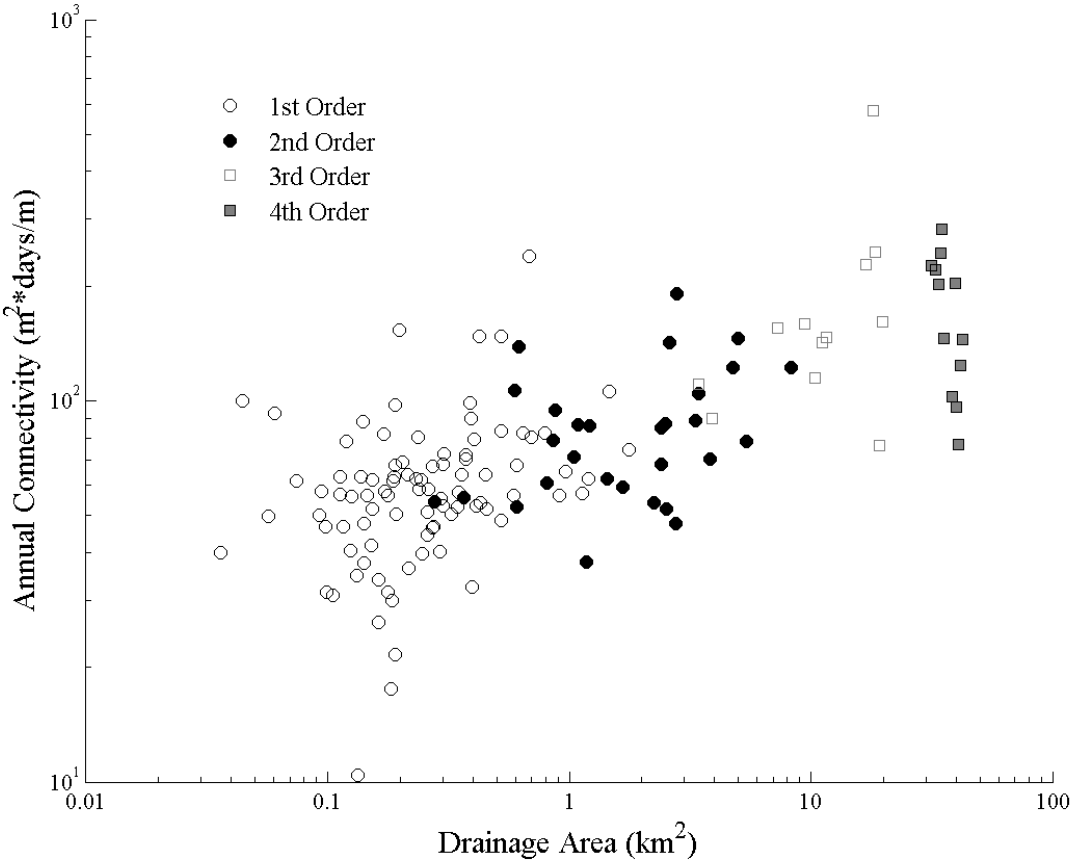

Figure 4.10. Annual connectivity vs. watershed area separated by stream order.

Figure 4.10 shows the relatively dendritic nature of the hydrologic system for the Mahantango Creek watershed. There were a large number of first order streams, with a relatively large range of drainage areas. There were significantly fewer second, third, and finally fourth order streams. The fourth order streams appear to have relatively similar drainage areas. This could possibly be due to the way the NHDplus data were segmented. Within the fourth order streams, there was a relatively large range of connectivity for a small range in drainage area. To address the overall impact of smaller streams versus larger rivers, the total connectivity metrics were summed based on stream order (Table 4.5). Since the metric output was originally in units of [m^2*days/m], the metric values were first multiplied back by the stream length to result in units of [m^2*days]. Thus, the results in Table 4.5 reflect the generally dendritic nature of a hydrologic network, with more first order streams than higher order streams. There was a clear downward trend in annual connectivity as stream order increased. The first order streams had the highest total annual connectivity, whereas the fourth order streams had the lowest total annual connectivity.

Table 4.5. Annual connectivity summed by stream order

Stream Order	Total Annual Connectivity (m^2*days/10^6)
1	8.8
2	7.4
3	6.0
4	4.6

Finally, the annual connectivity metric was examined for differences between the growing season and non-growing season in relation to drainage area (Figure 4.11). As aforementioned, the growing season was defined as May – October, and thus the non-growing season was represented by November-April.

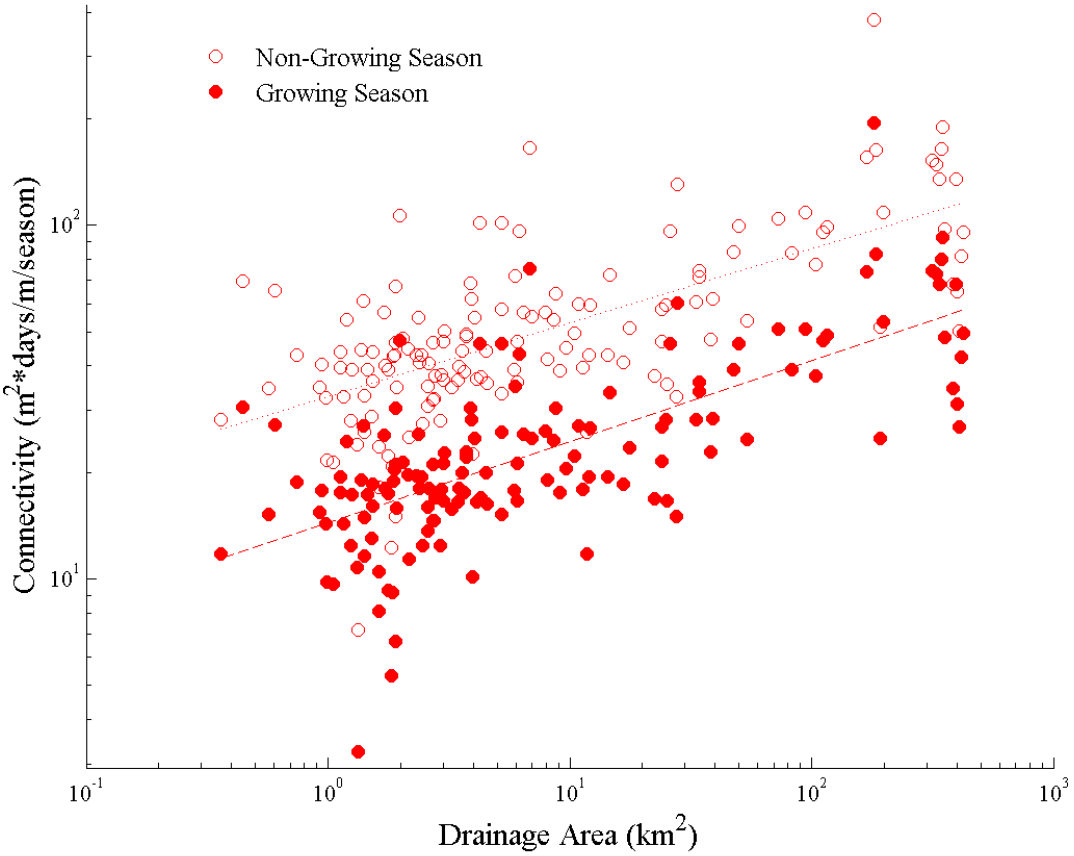

Figure 4.11. Seasonal connectivity related to drainage area with Non-Growing season defined as November-April and the Growing season defined as May-October.

There was a clear difference between the floodplain connectivity in the growing season versus the non-growing season. Based on Figure 4.11, connectivity in the non-growing season was consistently higher across all drainage areas. Both power-fit models expressed were significant, having $p<<0.001$. The power-fit models both explained nearly half of the data variance, with R^2 values of 0.45 and 0.50 for the non-growing season and growing season, respectively.

To examine the sensitivity and effects of the Manning's n selection, the model was re-run with Manning's n values a factor of 10x (1000%) lower than those previously used. The 10x factor was selected to mimic similar seasonal changes in Manning's n found by De Doncker et al. (2009). This was done to simulate a "Non-growing season Manning's n" value to see the

effects of Manning's n selection on seasonal connectivity differences. Figure 4.12 represents the seasonal connectivity with the adjustment to the Manning's n values. The non-growing season connectivity values had the adjusted ("low") roughness coefficients, while the growing-season connectivity values maintained the original ("regular") Manning's values. This should simulate a more accurate representation of the natural system with seasonal changes in channel roughness taken into account.

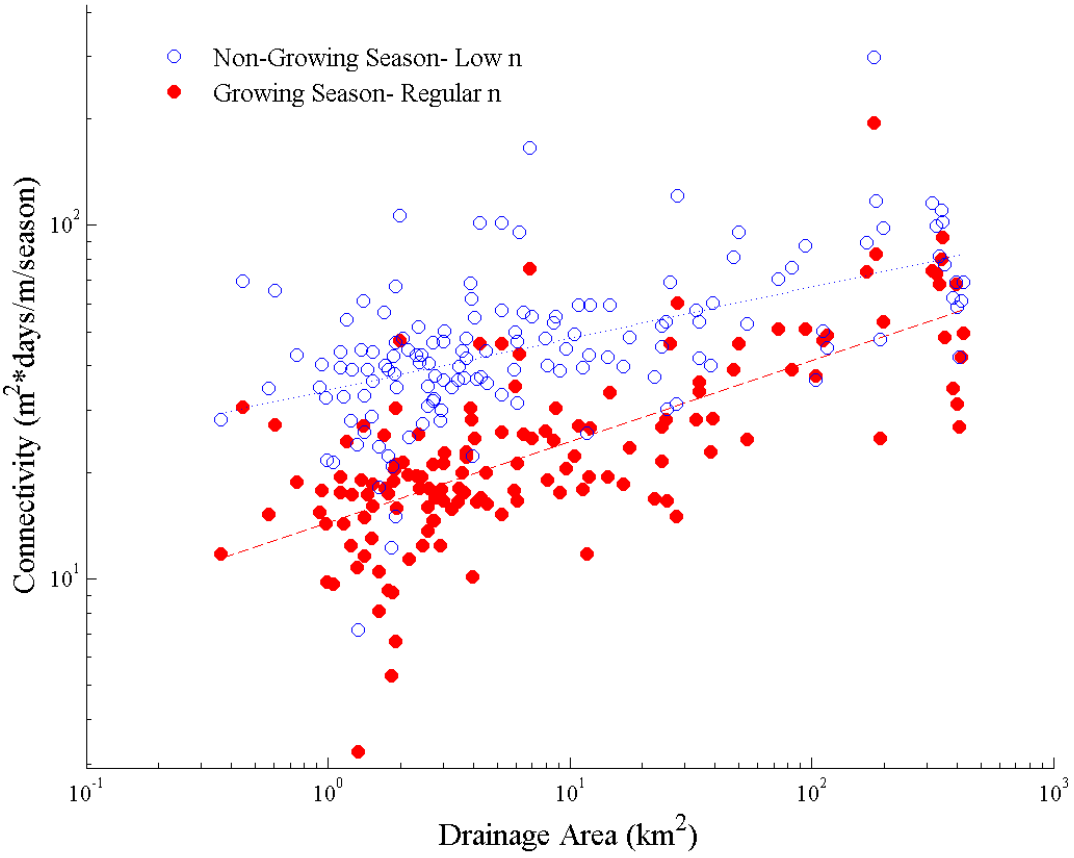

Figure 4.12. Seasonal connectivity with adjusted Manning's n for the non-growing season

There was little change when incorporating the lowered/adjusted Manning's n into the connectivity model. There was an average decrease in non-growing season values of 7.2% when

comparing the results from using the original Manning's n and the lowered value. To examine the effect of changing Manning's n for the non-growing season, the non-growing season connectivity values for both high and low Manning's n values were plotted against drainage area (Figure 4.13).

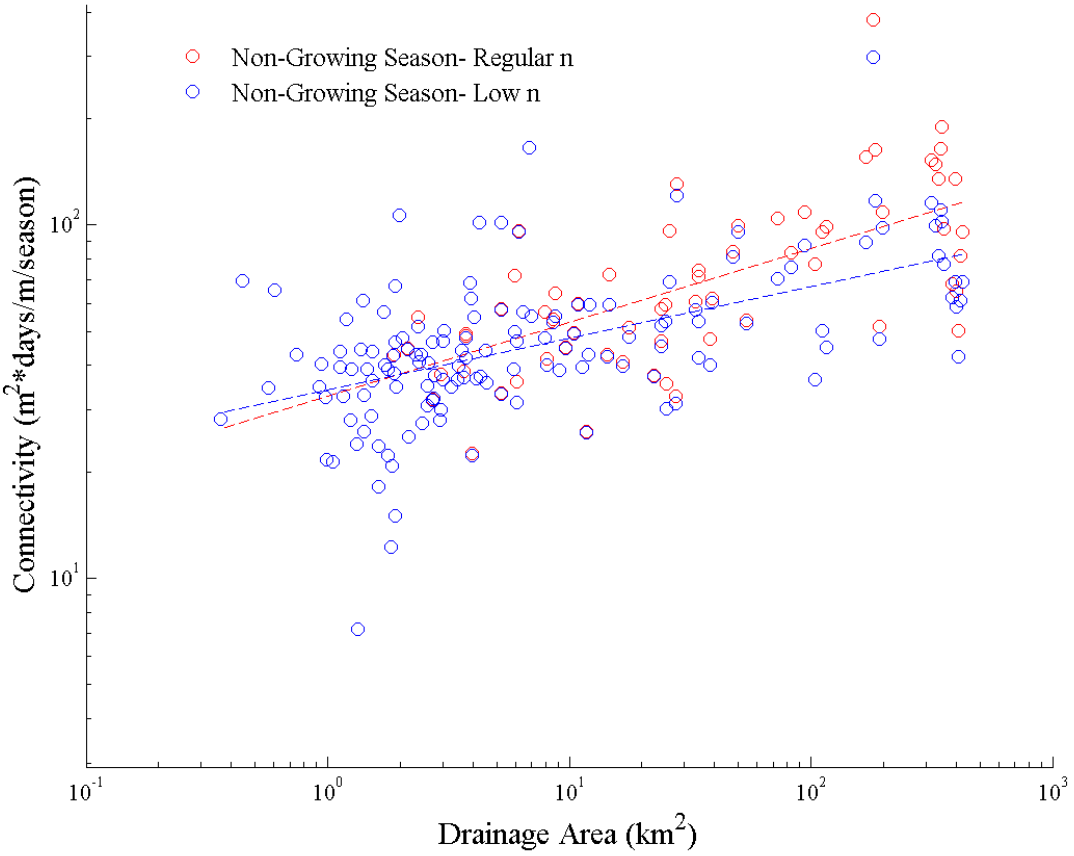

Figure 4.13. Effect of "high" (regular) and low Manning's n values on non-growing season connectivity across drainage areas.

As can be seen in Figure 4.13, there was a slight decrease in the non-growing season connectivity values when reducing the Manning's n by a factor of ten. The effect of the change was most clearly seen in the largest drainage areas. Additionally, the non-growing season

connectivity values were plotted against each other to examine the general data trend between the two series, without respect to drainage area.

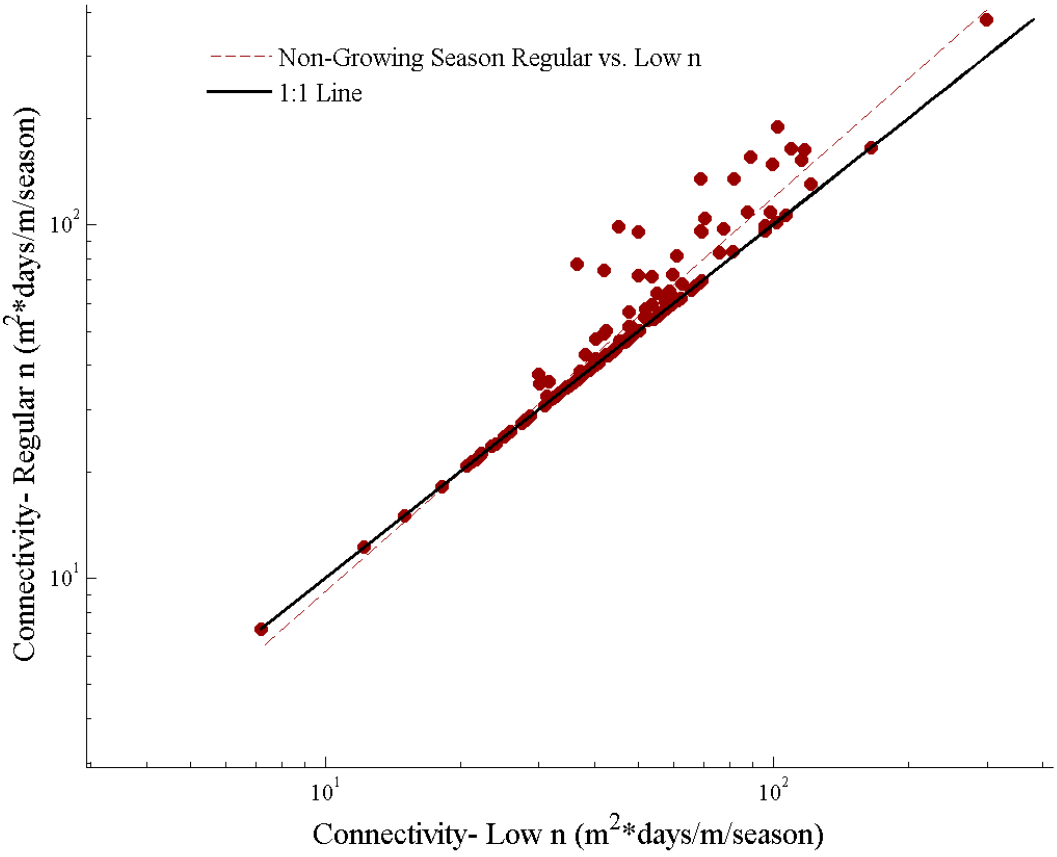

Figure 4.14. Comparison of connectivity values from the high versus low Manning's n values for the Non-Growing season.

The 1:1 relationship between the non-growing season connectivity for the regular Manning's n value versus the low Manning's n value showed that the data are generally higher with the regular Manning's n (Figure 4.14). The Manning's value appeared to have the largest effect at the highest connectivity values.

4.3.2 Slope

Local reach slope was also examined for its relationship with annual connectivity. Figure 4.15 represents the entire dataset of reaches (without the previously mentioned outlier) and their relationship between annual connectivity and reach slope.

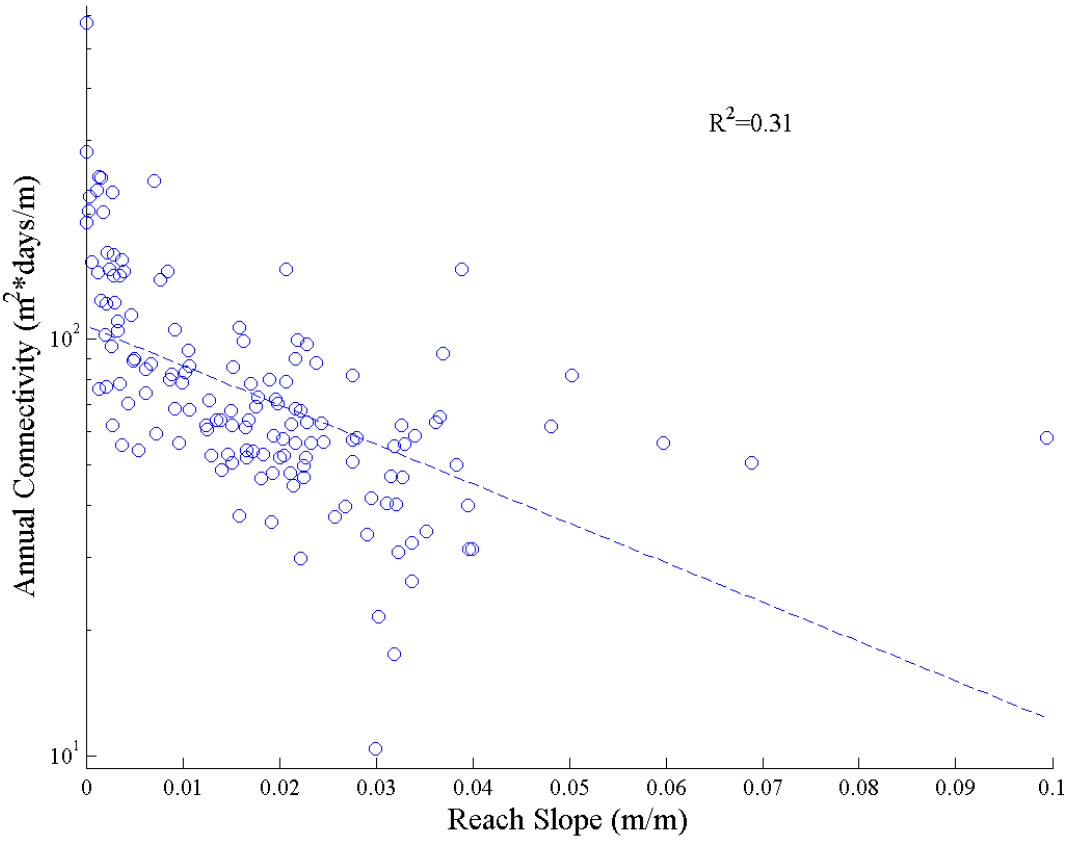

Figure 4.15. Annual connectivity as normalized to stream length related to reach slope.

There was a significant ($p<<0.001$) exponential decreasing relationship between annual connectivity and reach slope, although the model only explained nearly one third of the data variance. As with drainage area, the relationship between reach slope and annual connectivity

was also examined in terms of the different stream orders within the study watershed (Figure 4.16).

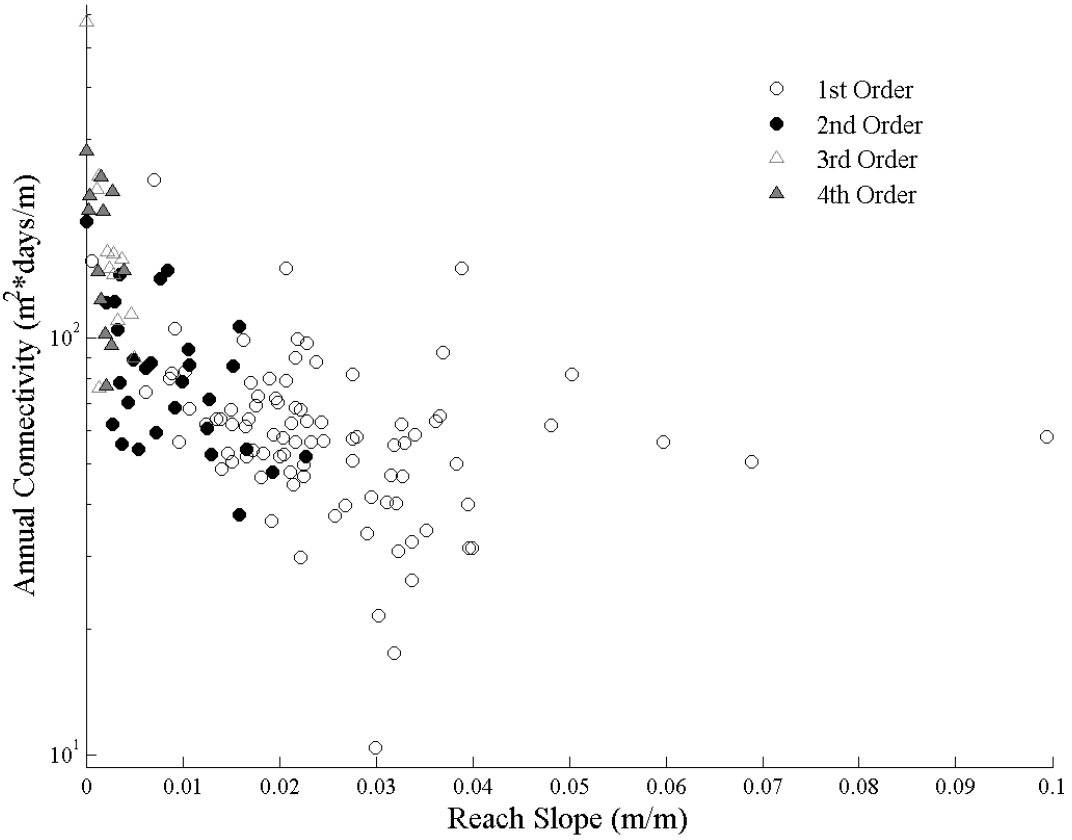

Figure 4.16. Annual connectivity vs. reach slope separated by stream order.

Based on Figure 4.16, there were a wide range of reach slopes for fist order streams. This wide range of slopes only somewhat corresponded with the wide range of annual connectivity that was also seen. Nearly all of the second through fourth order streams had slopes of 0.02 or less and exhibited a large range of connectivity.

To help explain the inverse trends between connectivity and drainage area/reach slope, the relationship between drainage area and reach slope was also plotted (Figure 4.17). There was

a clear decreasing power relationship ($p<<0.001$) between drainage area and reach slope. As slope increased, drainage area decreased.

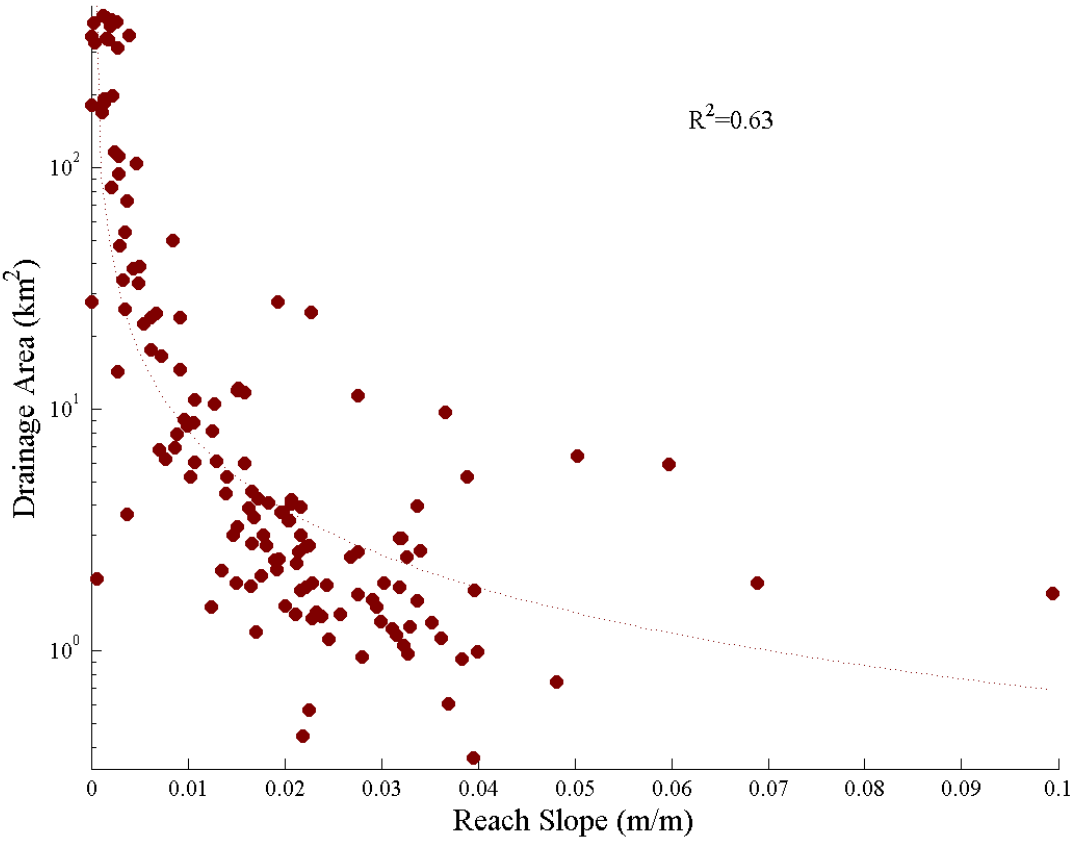

Figure 4.17. Power relationship between reach slope and drainage area within the Mahantango Creek watershed.

4.3.3 Watershed Characteristics

Even though drainage area and slope were both the primary watershed characteristics examined for their impact on connectivity, several other characteristics were also examined. First, three land use categories were related to annual connectivity. The percent of developed

area was examined due to its inclusion in the regional regression equations for FDC prediction Figure 4.18).

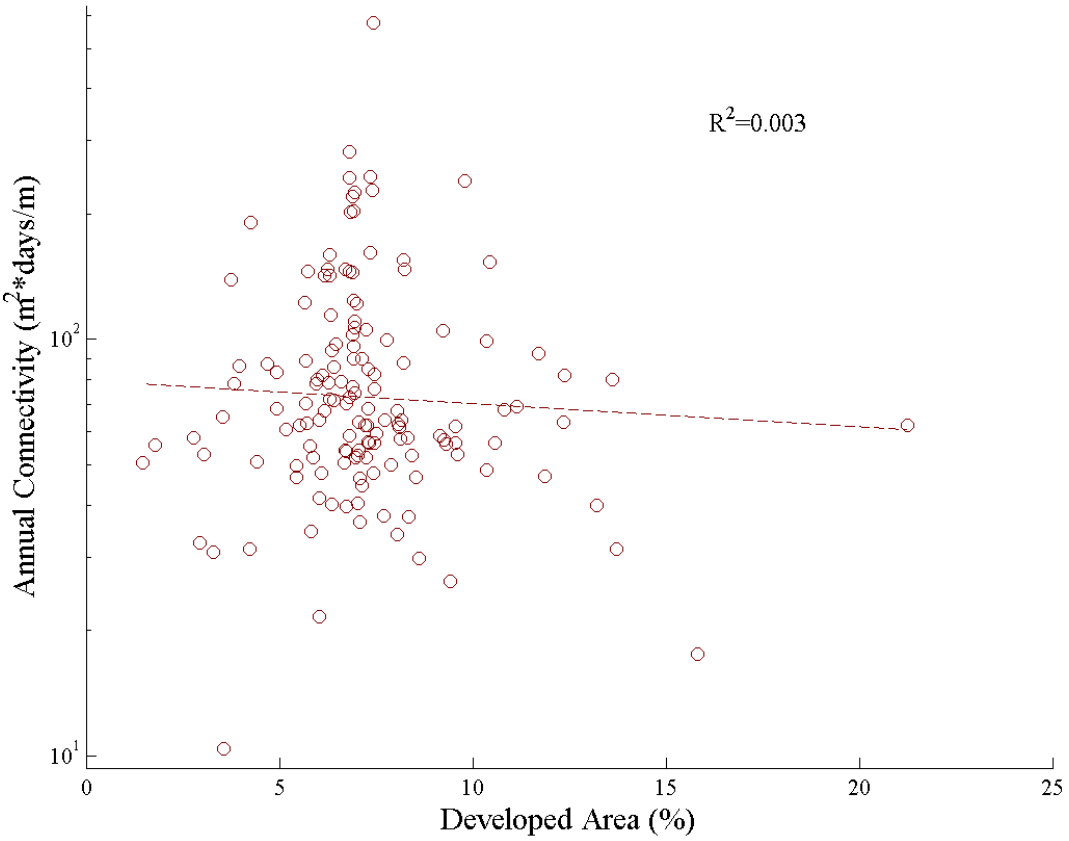

Figure 4.18. Relationship between annual connectivity and percent developed area.

Despite the significance of developed area in the FDC prediction regression model, there was not a significant relationship between annual connectivity and developed area (exponential fit, p=0.51). The exponential model fit did not explain the large amount of variance in the relationship.

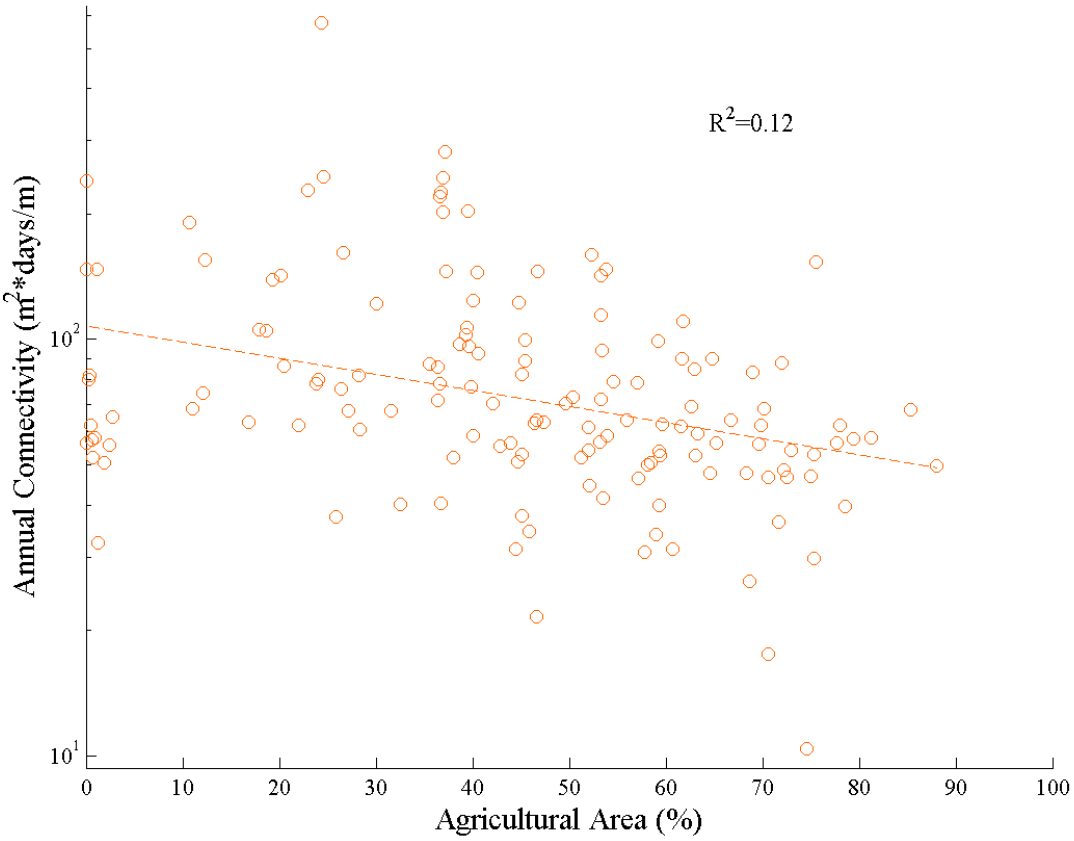

Figure 4.19. Relationship between annual connectivity and percent agricultural area.

Next, the percentage of agricultural area in the watershed was related to annual connectivity (Figure 4.19). Although a poor-fitting model, the exponential fit did have a significant decreasing relationship with agricultural area ($p_{model}<<0.001$). Thus, connectivity increased as the percent of agricultural land decreased in the watershed.

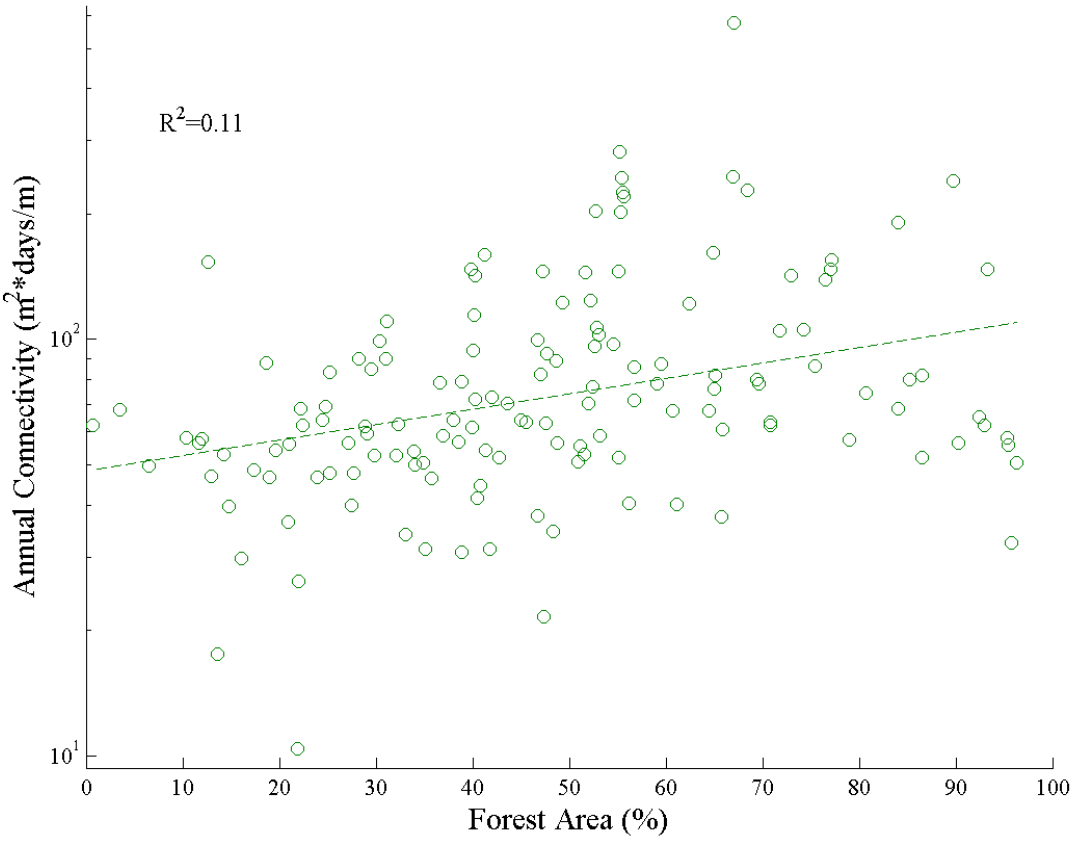

Figure 4.20. Relationship between annual connectivity and percent forested area.

Since agricultural area correlates to other land-use factors, the relationship between forested area and annual connectivity was also examined (Figure 4.20). The relationship between forested area and annual connectivity (Figure 4.20) had an opposite (increasing) trend and similar R^2 with an exponential fit. The model p-value for the exponential fit was $p<<0.001$, thus making it a significant relationship at $\alpha=0.05$. As forested area increased, there was a general increase in annual connectivity.

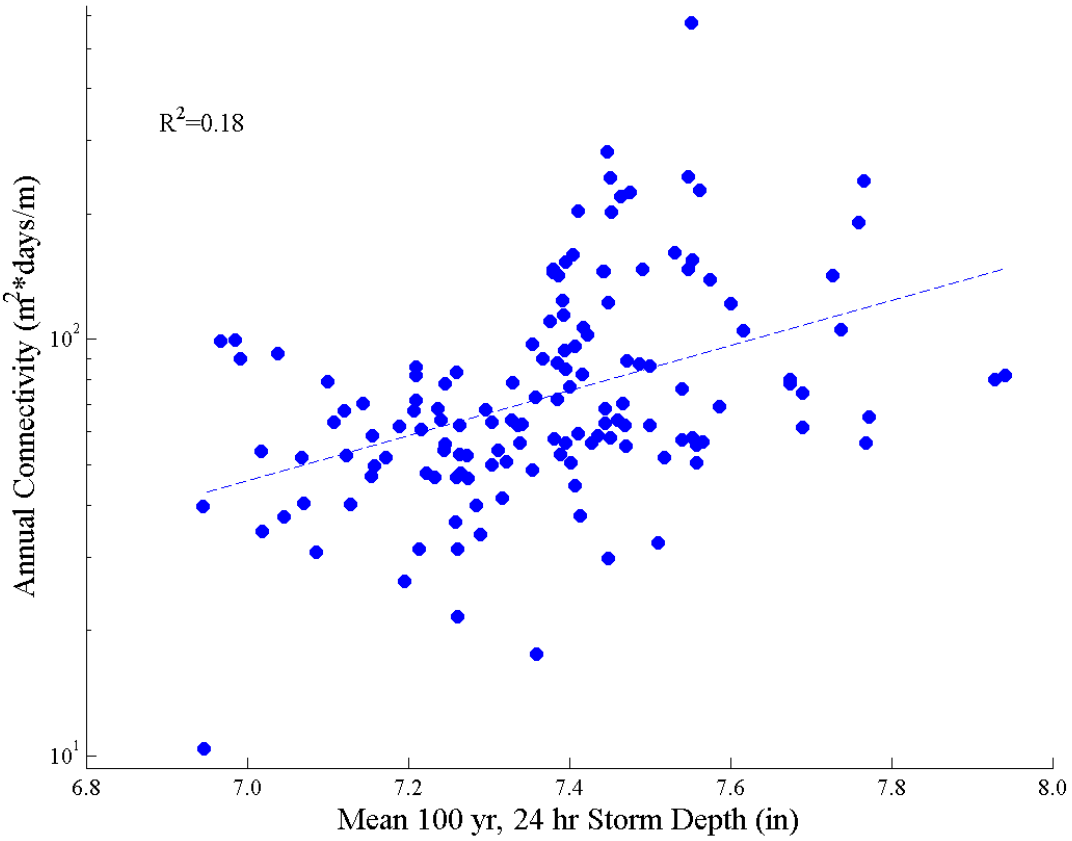

Figure 4.21. Relationship between annual connectivity and the 100-year, 24-hr precipitation depth.

The last watershed characteristic examined for its relationship with annual connectivity was the 100-year, 24-hr precipitation depth (Figure 4.21). This analysis was completed since this watershed attribute was included in the FDC prediction regional regression equations. There was an increasing exponential trend between the 100-year, 24-hr precipitation depth and annual connectivity. The exponential fit model was significant ($p<<0.001$) and had relatively low explanation for the data variance ($R^2= 0.18$).

5. Discussion

5.1 Regional Regression

5.1.1 Watershed Attributes

5.1.1.1 Drainage Area

The first major watershed attribute taken into consideration was the drainage area of the gauges chosen to be used in the regression analysis. Out of the 66 gauges, only 28 remained after eliminating those that were too small or too large based on drainage area. Ssegane et al. (2013) performed a slightly larger scale regression analysis, examining 51 gauges with drainage areas ranging from 24.8 km^2 to 7226.1 km^2, but performed the regression based on 47 of those gauges (four were used for validation). Our study utilized a much narrower range of drainage areas for the regression gauges (169 to 928 km^2) due to the application on the Mahantango Creek watershed (and the individual reaches within). The goal was to avoid regressing by gauges outside of the physiographic region of the study watershed and also to avoid gauges with drainage areas drastically different than the study watershed. If multiple application watersheds were used of varying drainage areas, the drainage area elimination factor (2.5 for this study) may have to be increased to include enough gauges for a representative analysis.

The reaches within the Mahantango Creek watershed (those locations upon which the regression equations were applied) had a wide range of drainage areas, from 0.5 km^2 to 425 km^2. There were approximately 125 reaches out of the 144 (87%) analyzed that were smaller than 169 km^2: the minimum drainage area for which the regression equations were built. Thus, many of the reaches that the final equations were applied to were outside the drainage area range that the regression equations were built with. This difference in drainage areas (reach drainage areas smaller than the minimum regression gauge drainage area) could have had impacts on the quality of FDC prediction, as discussed later.

5.1.1.2 Land Use

The watersheds used to develop the regression equations were primarily forested, with an average forested area of 66%. The remaining land uses were highly varied, including both agricultural area as well as developed areas. On average, there was more agricultural land (19%) than developed land (11%) in the gauged watersheds used in the regression analysis. Because of the presence of these two land use categories, there was the potential for anthropogenic influence on the flow regime. These alterations to the flow regime from land use could support these factors being included in the flow percentile regression equations, as discussed later.

The reaches within the Mahantango Creek watershed had relatively similar land use characteristics, supporting the regression methodology used in this study. These reaches had an average percent developed area in the watershed of 7%, slightly less than the 11% average from the regression gauges, yet still similar enough for application of the regression equations. The agricultural and forested land had wider ranges of influence on these reach watersheds than the regression gauge watersheds. The range of forest coverage in the gauged watersheds was 37-85%, as compared with the 1-96% in the reach watersheds. The range of agricultural land use also followed a similar comparison, with 0-50% in the regression gauge watersheds as compared to the 0-88% range for the reach watersheds. These wider ranges in the reach watersheds are most likely a result of the high number of headwater watersheds. These first order stream watersheds have small drainage areas, thus have the higher likelihood that they are all a single type of land use (0% of one land use, or nearly 100% or another). Since percent developed area, and not percent forested or agricultural land was included in the regression equations, the equations' applicability remains. As aforementioned, the regression gauge watersheds as well as the reach watersheds within the Mahantango Creek watershed had similar degrees of developed area, supporting the usability of the regression equations including an input for developed area.

5.1.1.3 Precipitation

The last watershed characteristic of importance that was examined was the 100-year, 24-hr precipitation depth due to its inclusion in the FDC prediction regression equations. The range of precipitation depths for the regression gauged watersheds (5.8-8.4 in) was wider, and thus inclusive of the range of the reach watersheds (6.9-7.9 in). This inclusion supported the

applicability and usage of the regression equations using the 100-yr, 24-hr precipitation depth as an input. Although acceptable as is, a narrower range of precipitation depths in the regression gauged watersheds closer to the range observed within the Mahantango Creek watershed may have led to better FDC prediction and validation results.

These discussed watershed attribute results (drainage area, land use, and precipitation) indicate that the gauges used in the regression analysis were acceptable for application of the equations to the reaches within the Mahantango Creek watershed. Besides these characteristics, the gauge selection factors promoted acceptability of the regression practices used in this study. As mentioned in the methods, only gauges in the same physiographic region were used in the regression analysis. Additionally, gauges were selected from only those within 100 mi of the Mahantango Creek watershed and at least 20 years of data. These selection criteria suggest that the regression gauge watersheds should experience relatively similar hydroclimatic influences as the Mahantango Creek watershed. This should aid in the similarity of flow regimes and thus prediction ability of the FDC regression equations.

5.1.2 FDC Regional Regression

5.1.2.1 Regression Model Variable Selection

The binary matrix output from the initial stepwise regression was examined for independent variable significance across the flow percentiles (Table E.3, Appendix E). For the following discussion, "significance" was determined by the "stepwisefit" MATLAB function with an entrance p-value of 0.1 and an exit p-value of 0.15, which determined each variable's inclusion in the model, and thus significance as discussed here. When summing the frequency of significance by column (by flow percentile), each regression only had 2-3 variables that were significant at each flow percentile. Literature's clearly established relationship between drainage area and flow supports the significance and model inclusion of drainage area for all flow percentiles. The inclusion of developed area in the regression model may have relation to the effect of urbanization on flow regimes, affecting hydrology during times of drought as well as flooding. The review conducted by Hollis (1975) clearly shows the common result of increased development in a watershed: higher flood flows; while Simmons and Reynolds (1982) noted that

development in the Long Island, NY area led to decreases in baseflow (in terms of baseflow as a percentage of total stream flow). As far as percent agricultural area, it was not found to be significant enough across all flow percentiles to be included in the final regression model. In the annual peak flow modeling and inundation analysis study conducted by Sarhadi et al. (2012), however, they found that agricultural land in Iran was important in predicting the 10 year storm, while range land was important in the 10, 20, 50, 100, and 200-year storm events.

Once the significance frequencies were summed for each variable, the final model included drainage area, percent developed area, and the 100-year, 24-hr mean precipitation depth. These parameters for regional regression prediction of FDCs were similar to those found in other related studies. For example, the final regression models developed by Ssegane et al. (2013) included drainage area, hydrologic soil index (HSI), and the maximum 100-yr, 24-hr precipitation depth (area-weighted average, Feaster et al., 2009). Out of 19 flow percentiles regressed, they found drainage area significant in all 19 equations, HSI significant in 10 equations, and the 100-yr, 24-hr precipitation depth significant in 7 equations. Thus, their significance frequencies across the flow percentiles were similar to those found in this study. In the regional regression study conducted by Roland and Stuckey (2008) in Pennsylvania, they found that both drainage area and percent urban area were important factors for predicting peak flows. As far a drainage area specifically, our study's findings correspond to other regional regression studies showing drainage area as an important factor of consideration. In their study in the northeast United States, Vogel and Kroll (1992) used drainage area as a variable in their final regression models for prediction low-flows. Bisese (1995) found that drainage area was "the most significant explanatory variable" in his peak flow regionalization study. In the large scale regression study on mean annual streamflow conducted by Vogel et al. (1999), drainage area was used as a base independent variable for the models, but the regression performance was improved by the addition of climate-related variables. Although not specifically the 100-yr, 24-hr precipitation as examined in our study, Mohamoud (2008) found that the mean August and February monthly precipitation for the period of record were significant at the 0.1% and 0.5% exceedance values, respectively. In regards to the inclusion of developed area in our FDC regression equation models, Hejazi and Moglen (2007) included the level of imperviousness in their low-flow prediction regression model. They did find, however, that the climatic variables in their study (temperature and precipitation) outweighed the degree of imperviousness in terms

of model importance. Our study found that developed area was more important than the 100-year, 24-hr precipitation depth, although the purposes of the studies in question were different: daily flow percentile prediction versus prediction of a single low flow statistic. Including the developed area in the regression may have been advantageous for this study, as Booker and Snelder (2012) noted that a source of error in their FDC prediction modeling could have been from development within their analysis watersheds.

The final regression model variables in this study were important statistically, but also corresponded with the physical relationships between the watershed attributes and flow. Drainage area has significant impacts on flow due to the fact that with more area, there is more influence from precipitation, and thus more water available to contribute to streamflow. As previously discussed, developed area can have impacts on both low flows as well as peak flows. With the increase in impervious area and overall higher runoff potential that corresponds with development in a watershed, flows will inevitably be affected. The small range of impervious area (maximum of 21% in the Mahantango Creek) coupled with the variable's significance in the regression equations suggest that even minor land development in the watershed could influence flow characteristics. And lastly, the 100-yr, 24-hr precipitation is important physically to predicting various flow percentiles because with higher levels of extreme precipitation, there will be more runoff, and thus streamflow.

5.1.2.2 Regression Equations & Performance

The resulting R^2 values in Table 4.2 show that the regression equations for predicting each flow percentile had a good fit ($R^2 \geq 0.9$) for 8 out of the 14 flow percentiles. The poorest model fits, as expected, occurred at the low flows (99-40% exceedance) and at the highest percentile (0.01%). This is most likely due to the difficulty and high degree of variation in extremely low and high flows. The R^2 values for the FDC prediction models in this study were comparable and even slightly improved to the peak flow regression models used by Sarhadi et al. (2012). In their study, the R^2 of the models ranged from 0.743 at the most extreme event (1000 year storm) to 0.881 at the 5 year storm event. The most extreme event modeled in our study was the 0.01% mean daily flow exceedance percentile, which, although not equivalent in storm magnitude to the 1000 year event, had an R^2 of 0.78. The regression model for the 2 year storm

event in their study had an R^2 of 0.824, whereas the most comparable model in our study (2.5% mean daily flow exceedance) had a model R^2 of 0.95. Thus, the R^2 of our daily flow percentile regression equation models were adequate for use in predicting FDCs at the ungauged locations within the Mahantango Creek watershed. Besides the R^2 values, the p-values associated with the F-statistic in the regression deemed all models statistically significant at α=0.05. In fact, each model had a regression $p<0.0001$.

The coefficients presented in Table 4.3 reveal the varying importance of each model variable to the prediction of the varying flow percentiles. Drainage area (Column "b") and percent developed area (Column "c") had opposite trends of importance going from low flows to high flows. Drainage area was least important at low flows and most important at the highest flows. The opposite was true for developed area, having the highest importance at the lowest flow (99% exceedance) and then dropping sharply to having a negative coefficient value from 10-0.01% exceedance. As far as the coefficient values for the mean 100-yr, 24-hr precipitation depth (Column "d"), most of the coefficient values remained constant from 70-0.1% exceedance. The 99% exceedance had the most drastic change in coefficient value at -0.09. This could be related to the relatively poor model fit for this 99% exceedance regression model (R^2=0.65, Table 4.2). The relatively poor prediction model for the 99% exceedance percentile was not of concern for this study since low flows (anything below bankfull) would not have affected connectivity results.

5.1.2.3 FDC Development

Using the developed regional FDC regression equations, the FDC for each reach in the Mahantango Creek watershed was predicted for all 14 flow percentiles. When examining Table E.5, it was clear that there was significant error in the FDC prediction, given the negative flow values at the low flow percentiles for many of the reaches. Many of the reaches predicted a negative value for the 99% flow, whereas some reaches had negative flows for all 14 percentiles on the FDC. These negative flows were most likely a result of the differences in parameter values used in the regression equations. As aforementioned, the regression equations were built on drainage areas ranging from 169-928 km^2, but the equations were applied to watersheds primarily smaller than the lower threshold. Over half of the reaches had drainage areas less than

10 km^2 and nearly 75% of the reaches had drainage areas less than 25 km^2. The importance of drainage area in the regression could be seen in the highest coefficient values (excluding consideration of the constant, Table 4.3) as well as the highest magnitudes of the three independent variables. The relative magnitudes of the independent variables can be estimated from the independent variable data used to build the equations, found in Table E.4 (Appendix E). Because of drainage area's high importance in the regression equations, the decrease in drainage areas from those used to build the equations (gauged watersheds) to those on which the equations were applied (unknown locations) was most likely the reason for negative flow prediction at low flow percentiles. Ssegane et al. (2013) also realized poor FDC prediction for small drainage areas (< 5.0 km^2). They, for similar reasons as in this study, attributed the poor predictions to applying the regression equations on locations where the independent variables (watershed characteristics) were outside the range used to build the regression equations.

The two highest flow percentiles on the predicted FDCs also experienced some issues related to the reach drainage areas. In theory, FDCs should be monotonically increasing from high percentiles (i.e. 99% exceedance) to low percentiles (0.01% exceedance). When examining the predicted FDCs, 25 out of the 144 reaches had the $Q_{0.01}$ prediction as nearly equal to or slightly less than the $Q_{0.1}$ prediction. This was most likely due to the small drainage areas associated with these reaches, as discussed with regards to negative flow predictions. Each of these "problem" reaches had drainage areas less than 3 km^2. The $Q_{0.01}$ regression equation (0.01% exceedance) had the highest weight (1.35) for drainage area out of any of the 14 regression equations (Table 4.3) in addition to having a relatively poor fit (R^2=0.78) as compared to the $Q_{0.1}$ equation . Thus, all these considerations combined most likely resulted in some of the reaches having the highest flow percentile (0.01%) be equal to or less than the immediately previous flow percentile (0.1%). Mohamoud (2008) discussed this phenomenon, suggesting similar causes as just mentioned for this study, including poor model fit and regression gauge watersheds with a wide range of attributes.

5.1.3 Hydrograph & FDC Validation

The statistics in Table 4.4 represent the accuracy of the FDC prediction regression models as well as the overall methodology of developing the ungauged hydrograph. The three

primary statistics examined were the NSE of the full hydrograph (represented in the table as "NSE"), the NSE of the 14 points on the FDC ("NSE_{FDC}"), and the NSE of the upper 50 percentile of flows (NSE_{HIGH}). The validation statistics found in this study were comparable to other similar ungauged hydrograph prediction studies. Overall, our results were slightly worse than those found by Ssegane et al. (2013). They found NSE values for FDC predictions to be greater than or equal to 0.86. When excluding the two poorly predicted gauges (as discussed earlier), this study had NSE prediction values ranging from 0.59 to 0.88. They had full hydrograph predictions with NSE values ranging from 0.09 to 0.84, whereas our study had values across a narrower range from 0.43 to 0.64. Mohamoud (2008) also developed ungauged hydrographs using a similar FDC prediction regional regression technique, but used another algorithm for developing the final predicted hydrograph. That study had similar daily hydrograph prediction quality as Ssegane et al. (2013), with NSE values of the predicted hydrograph ranging from 0.46 to 0.85 for the similar FDC method to that used in this study.

There are several reasons why some of the validation statistics (Table 4.4) were poor. First of all, for site 1570500, the NSE values for the full hydrograph, the upper 50% of flows on the hydrograph, as well as for the FDC were all negative, indicating that the predicted hydrograph was a poorer estimation than the mean. One possible explanation to this is the difference in drainage area sizes between the timing gauge (nearest gauge as measured by linear distance between watershed centroids which was the source for the hydrograph timing) and the "unknown" location, or validation gauge. The two poorest validation locations (1563000 and 1570500) both had drainage area ratios (DA_{gauge}/DA_{near}) that were significantly large, indicating that the timing gauge and validation gauge locations had drainage areas that were vastly different. For the poorest prediction, the validation location drainage area was over 50x the size of the timing gauge. This difference in sheer magnitude between watershed sizes could have affected the hydrograph prediction. When one location is tremendously larger than the other, the assumption that the same hydroclimatic events would affect the watersheds evenly may not remain true. The larger watershed may have been affected by significant storm events that the smaller watershed may not have experienced, thus breaking one of the primary assumptions for the ungauged hydrograph development algorithm. Also, to recall, the regression equations were built based on drainage areas ranging from 169-928 km^2. Thus, the two poorest validation predictions were for locations where the drainage area was outside that range.

Another issue that could have related to the poor validation statistics was the distance between the timing gauge and the prediction location. Mohamoud (2008) showed that with increased distance between the timing and prediction locations, the prediction ability of the modeling procedures decreases, leading to poorer hydrograph prediction at the ungauged location. In this study, however, the difference in drainage area sizes appeared to matter more. The distance values (watershed centroid to watershed centroid) for each of the validation gauges were all relatively similar and did not seem to correlate with model performance. Gauge 1614500 had the largest distance between timing and validation gauge watersheds, yet had fair NSE statistics that were reasonable with the other gauges. Gauge 1563000 had some of the poorest validation statistics, yet had the smallest distance between watersheds. Thus, for this study, the drainage area size and comparable size to the timing gauge may be more important in hydrograph prediction than distance between the watersheds. This differs from the findings of Ssegane et al. (2013) who, using similar methodology, suggested that distance between the watersheds was more important than difference in watershed sizes between the timing and validation gauges.

The accuracy of the validation gauge's predicted hydrograph was based on two primary components: the regression-based FDC prediction and the timing sequence taken from the nearest gauge, referenced as the "timing gauge". Thus, any error in the FDC prediction should carry through as error in the final hydrograph prediction. This can be seen by the NSE_{FDC} values being consistently higher than the NSE values for the full hydrograph prediction. The high flow (upper 50 percentile) NSE values were often higher than the NSE_{FDC} values, however. Also, the high flow NSE values were consistently slightly higher than the full hydrograph NSE values. This suggests that the overall hydrograph prediction method performed better at the higher flows. This is conducive to this study given our focus on flooding, and thus high flows.

As previously described, the validation methodology used the "nearest gauge" method to find the timing sequence for the unknown locations. When determining the "nearest gauge" for the ungauged reaches within the Mahantango Creek watershed, the "nearest gauge" was manually set to the outlet gauge for the watershed (USGS Site number 01555500). This was done because of the assumption that all reach watersheds within the Mahantango Creek watershed would be affected by similar hydroclimatic events, and would thus have a similar timing of flows. Because of this, it was important that the outlet gauge had satisfactory

validation statistics. The NSE of the full hydrograph was 0.43, which was not the best predicted hydrograph, but was acceptable. More importantly, the NSE_{FDC} was 0.72. Since the regression technique (drainage area factor, model variable selection, etc.) was the area which could be manipulated most, the NSE_{FDC} value was more important because it was more able to be controlled. The full hydrograph NSE value was more a result of the overall ungauged hydrograph development algorithm rather than any intermediate controllable parameters or factors. Thus, the validation statistics for the Mahantango Creek gauge deemed the overall ungauged hydrograph development methodology (QPPQ method) acceptable for application to the individual reaches within the Mahantango Creek watershed.

As far as the qualitative analysis of the validation gauges (01570500 and 01555500), the trends in the FDCs correlated to the trends in the hydrographs. The Mahantango Creek FDC (01555500) had fair correlation for the upper flows, but over predicted the low flows. This trend showed on the hydrograph portion as well (Figure 4.3). The NSE versus NSE_{HIGH} values also corresponded to the qualitative analysis. In the Mahantango Creek hydrograph section, the peaks were more accurately predicted than the lower, more baseflow-like, conditions. Similarly, the validation gauge exemplifying poor prediction (USGS 01570500) had matching qualitative trends between the FDC and the hydrograph. The over-prediction of high flows and under-prediction of low flows as seen on the FDC (Figure 4.2) correlated to the over-predicted peaks flows and slightly under predicted low flows in the hydrograph (Figure 4.4). This qualitative analysis helped confirm and add validity to the quantitative statistical analysis for the ungauged hydrograph prediction model.

The over-prediction of high-flows for the USGS 01570500 location and under-prediction of high flows for the Mahantango Creek gauge could have implications in the connectivity analysis. Since the "nearest gauge" was designated to be the Mahantango Creek gauge for the reaches within the watershed (as was the focus for the connectivity analysis section of this study), particular attention was paid to the hydrograph prediction characteristics of this gauge. As aforementioned, the peaks were under-predicted for the Mahantango Creek outlet gauge (USGS 01555500). This suggests that connectivity values, as discussed later, could be under-predicted. This same trend may not hold, however, when predicting hydrographs of significantly different drainage areas (i.e. the headwater streams within the Mahantango Creek watershed).

5.2 Inundation Area

5.2.1 Floodplain Inundation Area

The first measure of flood inundation area was the inundation area for each reach as normalized to the length of the reach, thus resulting in a [m^2/m] metric of inundation. As seen in Figure 4.5, inundation had an increasing power relationship with drainage area. This suggests more riparian inundation for large rivers versus small streams. This could be due to a multitude of factors, but is most likely a result of both increased flow at larger drainage areas, as well as floodplain topography that supports wide expanses of floodplain leading to a larger inundation area. Because of this, floodplains associated with larger river systems have the potential for a higher degree of water quality improvement since more water is exposed to the floodplain surface. Although this does not take temporal duration in the floodplain into account, the degree of spatial inundation is still important when examining the importance of floodplains for ecosystem services.

When examining the inundation area versus drainage area relationship, there appeared to be a higher degree of inundation area variation at larger drainage areas. Inundation areas at smaller drainage areas appeared to fit tighter to the trendlines, whereas the data spread out more at larger drainage areas. Additionally, based on the trendlines, there were only small differences in inundation areas between the different storm events at small drainage areas. This could be explained by the geomorphology and hill slope terrain only supporting a certain size floodplain width. Once that floodplain is fully inundated, the water rises vertically, therefore not increasing the spatial area of inundation. Thus, no matter the flow after a certain point, the inundation area would not increase. For larger drainage areas, however, there was more variation in the data and a much larger difference between the different storm sizes. This suggests that the floodplains of large drainage area systems are wide enough to carry the flow of infrequent and extreme flood events.

5.2.2 Floodplain Coverage

Besides normalizing the inundation area results to the length of stream within the respective reach segment, the inundation area values were also normalized to the total floodplain

area. This total floodplain area was assumed to be equal to the largest modelled inundation area, or the inundation area for the 0.01% exceedance flood. Thus the values shown in Figure 4.6 represent the inundation area divided by the inundation area for the 0.01% flood, or: $AI_{pp}/AI_{0.01}$. Theiling and Burant (2012) suggest a similar normalization method for its ability to allow examination of floodplain activity during flooding. Since one of the objectives of this study was to examine flooding characteristics across the landscape, floodplain coverage (fraction of total floodplain inundated) was compared across drainage areas as well as across different storm events. Based on Figure 4.6, floodplain coverage decreases as drainage area increases in a power relationship. Therefore, for larger river systems, a smaller percentage of the floodplain will be covered as compared to small streams at the same flow exceedance percentile (Same point on the FDC). The opposite holds true for small streams: nearly the entire floodplain is inundated (floodplain coverage ≈ 1) for all modelled storm events. This is most likely due to the total size of the floodplain in small streams. The small size of headwater floodplains would result in complete inundation any time the stream breeches bankfull stage. Some reaches had floodplain coverage values of greater than 1, which can be tied back to poor FDC prediction. As aforementioned, some of the $Q_{0.01}$ values were slightly less than the $Q_{0.1}$ values. This translated into having $AI_{0.01}$values slightly less than the $AI_{0.1}$ values, leading to a floodplain coverage of greater than one when dividing by the $AI_{0.01}$ area. The larger total size of floodplains in larger, higher order streams most likely explains the larger effect of storm event size on floodplain coverage. For smaller streams, the storm size had little to no effect on floodplain coverage. Since there was more available floodplain in larger drainage area systems, different size storm events would have the capacity to flood more area, with more pronounced differences between the different stages for the varying events.

The inundation maps (Figure 4.7 & Figure 4.9) provided a way to qualitatively analyze and validate the inundation area relationships previously discussed. The two case study reaches' inundation patterns exemplified the conclusions brought forth by the inundation area and drainage area relationships. In the small reach (Figure 4.7), the low stage events extend to most of the full floodplain width. There is very little light coloration along the edges of the reach, meaning that the higher stage events did not lead to significantly more inundation area. Thus, the lower stage flood simulations "filled" the floodplain so that higher stage events did not significantly extend the spatial area of inundation. The large reach (Figure 4.8), on the other

hand, showed that higher stages had a significant impact on extending the inundation area. The larger areas of lighter blue show that the higher stages did extend the area of inundation beyond that which was covered by the lower flood simulations. Thus, as with the varying stages, the same conclusion as previously discussed could be made: that larger drainage areas have floodplain topography that is conducive to larger areas of inundation as compared to reaches with a smaller drainage area. Smaller reaches have a narrower floodplain that once inundated at a relatively low stage (or a storm event just above bankfull), the additional volume from more flow is stored vertically (deeper water in the floodplain rather than expanding the area of inundation horizontally.

Most flood inundation studies have been conducted on large river systems due to the potential damage to human infrastructure. Because of this, there is a gap in the literature in regards to flooding in smaller systems or across system sizes, as modelled in this study. This helps support the novelty and importance of this study for contributions and better understanding of a range of system sizes and their biogeochemical impacts while hydrologically connected.

On the lower Mississippi River, the floodplain inundation modeling study conducted by Hudson et al. (2013) showed that floodplains could be inundated at relatively low magnitude flow percentiles. They found 9% floodplain inundation at a 50% flow while the 10% exceedance flow resulted in 87% of the floodplain being inundated. The floodplain coverage results from our study are only somewhat comparable given the vast difference in system drainage areas from the lower Mississippi to the Mahantango Creek watersheds. Nonetheless, the largest drainage areas from this study experienced approximately 65% floodplain inundation for the 5% exceedance event. A similar percentile event in the Hudson et al. (2013) study would correspond to approximately 90% floodplain inundation. This difference is most likely to differences in the definition of the full floodplain area as well as system type (large lowland system versus smaller, Piedmont style system). Theiling and Burant (2012) showed that on a section of the Upper Mississippi River, the floodplain was nearly entirely covered (floodplain coverage ≈ 0.9) at a storm percentile of approximately 5-10%. Once again, our study showed that for larger systems, full floodplain coverage would not occur except during extreme events of greater than 0.1% exceedance. As with the Hudson et al. (2013) study, however, this difference is most likely due to differences in defining the full (100% coverage) floodplain. Differences could also be due to

the different channel and floodplain geomorphologies associated with such a large system of the Mississippi River as compared to the Mahantango Creek watershed.

These differences in floodplain coverage can have implications with regards to floodplain management. One aspect of floodplain management includes determining where restoration projects should take place. These results show that a restoration project may be smaller and thus include a lower capital cost for small streams, but restoring the floodplain riparian zone of a larger river could result in more water quality and ecosystem benefits given the additional inundated area during flood events. A floodplain manager may decide, however, to restore the floodplain riparian zone of a smaller stream, realizing that the entire project area could be flooded during all events greater thank bankfull, meaning that the entire restoration area would be utilized for water quality improvement. Larger projects may have sections that are inundated only during extreme and infrequent events, thus lessening the support for restoring such an area. To gain a full understanding of the potential ecosystem services rendered by floodplains during flood events, both duration and areal extent of inundation, and thus connectivity, must be taken into account.

5.3 Connectivity

5.3.1 Drainage Area

Connectivity had a significant increasing power relationship with drainage area. As drainage area increased, there was a significant increase in annual connectivity (m^2*days/m). Thus, a larger area of the floodplain is inundated for a longer period of time across a year's time span in larger systems as compared to smaller streams. This most likely relates back to the geomorphology associated with floodplains of small versus large systems. Since local topography is a driving factor of when inundation of the floodplain occurs, channel geomorphology differences associated with different sized systems make up the driving factor most likely responsible for differences in connectivity across the landscape. Headwater systems often have smaller floodplains available for inundation and have fairly steep hill slopes leading to the channel. In large, lowland rivers, on the other hand, the floodplain is often much wider and more expansive, leading to a large inundation area once bankfull flow is exceeded.

Some components from the development of the model may have led to a falsely low representation of annual connectivity in small systems. As previously described, the daily flow data at the gauges and used throughout the model were daily mean flows at a particular location. This most likely underrepresented floodplain inundation in small stream systems. Using daily maximum data would have given a better representation of flooding characteristics, especially in smaller streams which often have flashier hydrograph patterns. Some inundation events in small streams occur over a time span of less than a day's time. This could lead to a mean daily flow value of less than bankfull, even though inundation did occur at some point over the course of that storm event. Using daily peak flow instead of daily mean flow would have a less prominent effect on larger streams since their hydrographs generally capture more of the full flood wave and are not quick to rise and fall.

The same increasing relationship between annual connectivity and drainage area held true when examining the data across the different stream orders. The first order streams within the Mahantango Creek watershed had a wide range of drainage areas and a corresponding wide range of annual connectivity values (Figure 4.10). Despite the wide ranges, there was still an increasing trend visible, even across only the first order streams. The fourth order streams had a very narrow range of drainage areas yet a wide range of annual connectivity values. As aforementioned, this could be due to the segmentation of the NHD data, but it could also be a result of model variation and error.

As aforementioned, it appeared that connectivity was higher in larger stream systems (per unit length of stream). To gain a more holistic understanding of connectivity across the entire watershed, the connectivity metrics were summed with length of stream taken into account. Thus, the "total annual connectivity", as shown in Table 4.5, reflects the impact of stream order on river-floodplain connectivity from a full watershed perspective. The clear downward trend of total annual connectivity with stream order shows that on average, there is more total area inundated for a longer period of time across a given year for the entirety of small headwater streams rather than the high order reach segments. With more floodplain area inundated over a longer period of time, there is the higher potential for increased water quality improvement (sedimentation, denitrification, etc.). Thus from a drainage area (system size) perspective, large rivers have the potential for more connectivity on a "per unit reach length" basis, and therefore more long-term water quality impacts over a shorter length of stream (higher "efficiency").

When accounting for the dendritic nature of the system, however, connectivity in headwater streams as a whole may be more significant in terms of impact on watershed outlet water quality.

As far as implications, these results suggest that headwater streams as a whole may have more effects on water quality since they have a higher total annual connectivity. From a restoration perspective, however, our results support restoring larger systems since the floodplains of a larger river are more connected. Individual floodplains that are more connected will inundate a larger area of the riparian zone more often, thus reaping more benefits from the restoration. From floodplain connectivity perspective, several first order streams would have to be restored to have the equivalent effect of restoring a single reach in of a higher order.

There was a clear difference between connectivity in the growing season and non-growing season. There was consistently higher connectivity during the winter months (November-April) across all drainage areas (Figure 4.11). This difference could be due to several factors. When examining the makeup of the connectivity metric, there were two components: time and area. Thus, to have higher connectivity, either water has to be in the floodplain for a longer period of time or across a wider span of the floodplain. For this study, the most likely explanation for the higher connectivity during the non-growing season was the fact that there was overall more flow during the non-growing season. As aforementioned in the study site description, 68% of annual flows occur during the non-growing season according to the data for the outlet gauge (USGS 01555500). More flow leads to the potential for more flooding, leading to more inundated area for a longer period of time, thus leading to the higher connectivity. Despite this observation, there are several other factors that could have led to the higher connectivity values during the non-growing season.

Another issue to consider regarding seasonal connectivity was with regards to the model algorithm and selection of a Manning's n value. For the purposes of this study, a single Manning's n value was chosen for each reach representative cross-section. It was assumed that this was representative for the reach across all temporal scales, not taking into account changes in roughness over various seasons. As seasons change, however, bank and bed vegetation changes, leading to changes in roughness, and therefore the Manning's n coefficients. Manning's n coefficients have been found to differ as much as tenfold between seasons, with the highest coefficient values being found in the summer months (De Doncker et al., 2009). In their study in a subtropical marsh, Shih and Rahi (1982) found a three-fold seasonal difference in

Manning's roughness. During the non-growing season, therefore, Manning's n coefficients are lowest, leading to the highest flow velocities. Given the same flow value and higher velocity, the cross-sectional area of flow would decrease, leading to lower stage. Thus, according to this Manning's theory, flow in the winter months should be at a lower stage. The Manning's n coefficients chosen for this study may have been more representative of the roughness of the growing season. This would lead to a Manning's n coefficient higher than in reality, leading to higher stages. This could be the explanation for the higher connectivity during the non-growing season: a non-representative Manning's n value.

When the non-growing season connectivity values incorporated the lowered Manning's n value (Figure 4.12), there was little change in the final connectivity metric. The overall model appeared to have a low sensitivity to changes in Manning's n value, given a 10x decrease (1000%) in Manning's n led to a 7.2% decrease in non-growing seasonal connectivity averaged across all reaches in the study watershed. Based on a qualitative interpretation of Figure 4.13, the change in Manning's n coefficient appears most prevalent at high drainage areas. This indicates that the change in roughness was most reflected in lowering the connectivity values for watersheds with a relatively large drainage area. Similarly, when examining Figure 4.14, only slight differences between the non-growing season connectivity for the regular Manning's n versus the low Manning's n were visible. The lower Manning's n value led to lower connectivity values during the non-growing and was most apparent for the highest connectivity values. This was apparent by the data trendline lying above the 1:1 trendline, indicating that the data trended higher towards the higher (regular) Manning's n connectivity values.

Overall, the potential changes in channel roughness associated with the varying seasonal vegetation did not have a major effect on the seasonal connectivity values. It appeared that the difference in flow volume between seasons was the driving factor explaining why connectivity was consistently higher in the non-growing season than during the growing season.

In theory, higher connectivity during any particular season would suggest a higher potential for ecosystem services from the floodplains in the form of biogeochemical nutrient transformation and consequent improvement of water quality. This higher potential, however, may not be realized in actuality since the season during which connectivity is higher is during the non-growing season. The non-growing season (November-April) can be generally characterized by low temperatures and therefore less biological activity. Since microorganisms are the

responsible parties for much of the biogeochemical processes, their inactivity will result in lower nutrient transformations and thus fewer water quality benefits from inundated floodplains. Even still, water quality benefits of floodplains not associated with microbial activity, such as sediment removal, may still increase during the non-growing season with the increased connectivity. During the non-growing season, when much of the vegetation has died off, there is the potential for higher erosion rates, introducing more sediment into the channel, supporting the importance of higher connectivity during that season.

5.3.2 Slope

Reach slope had a significant decreasing exponential trend with annual connectivity. As reach slope increased (realizing that higher reach slopes are generally characteristic of smaller, headwater streams), connectivity decreased. Fourth order streams had a large range of connectivity given small range in reach slope. This overall decreasing relationship was inverse to the relationship shown in regards to drainage area in Figure 4.9. The inverse relationship in connectivity patterns can be tied back to the inverse relationship between drainage area and reach slope (Figure 4.17). As general landscape hydrology suggests, and as exemplified in this study, drainage area decreases as reach slope increases. This corresponds with steeper slopes in low order, headwater-type streams and lower slopes for high order, lowland rivers. Thus, due to the inverse relationship between drainage area and reach slope, both relationships suggest higher connectivity in larger rivers and lower connectivity in smaller streams.

5.3.3 Watershed Characteristics

There was no significant relationship between the proportion of developed area in the watershed with annual connectivity. This watershed attribute, however, was found to be significant in the regression analysis (out of the 16 characteristics analyzed) and was included in the FDC prediction regression equation models. Despite this, there was no clear relationship between developed land and connectivity. White and Greer (2006) found higher flows associated with an increase in developed land use from 9-37%, although they also noted that the effect of development decreased as storm size increased, up to a 10-year storm event. Hollis

(1975) also noted that development had a decreasing effect as storm size increases. Additionally, Hollis (1975) concluded that 5% imperviousness would not significantly affect the 1-year or greater flood. This concept could have come into effect in our study given 17 out of 144 reaches had less than 5% developed area and nearly 90% of the reaches (129) had less than 10% developed area. The findings of these other studies pose possible explanations to the poor relationship between annual floodplain connectivity and the percentage of developed area in a watershed that was found in our study.

The level of watershed development, although potentially related to flow, could be related to channel geomorphology, and thus connectivity. In a review of the effects of overall "urban transformation" on stream systems, Chin (2006) showed "clear evidence for larger channels in urbanizing rivers" by summarizing changes in channel geomorphology (capacity, width, depth) from relevant studies. Hession et al. (2003), however, found that the composition of the riparian zone area has more impact on channel geomorphology than upstream developed area. Therefore, despite development in the watershed, the local topography (and therefore connectivity) in a particular reach may be affected most by the riparian zone attributes. Despite this, the study in southeast Pennsylvania by Cianfrani et al. (2006) suggested that the effects of urbanization on streams are variable and geographically dependent. These findings may help explain the lack of relationship between annual floodplain connectivity and developed area.

Because the proportion of developed area in a watershed correlates to the proportions of other land uses in the watershed, it was important to examine the connectivity relationships with other watershed characteristics as well. Annual connectivity had inverse and opposite relationships with the percentage of agricultural and forested land uses. When the proportion agricultural land increased, connectivity decreased, whereas when the proportion of forested area increased, connectivity also increased. These differences in connectivity could be related to the differences in channel geomorphology related to different land uses. Since local topography plays such a critical role in when streams inundate the floodplain, it is plausible that channel geomorphology could be a root difference between the landscapes, and thus the annual floodplain connectivity, as our results suggest. Sweeney et al. (2004) found that there were significant differences in channel geomorphology between streams with riparian zones that were forested versus non-forested in southeast Pennsylvania and northern Maryland. They were clear to clarify that their non-forested reaches had no external agricultural or development influence,

so any effects seen would be from solely the presence or absence of forest area. They found non-forested channels to be narrower than reaches with forested riparian areas. Additionally, they found that forested streams had more bed roughness and slower water movement through the channel, both of which could have impacts on water stage, and thus connectivity. Hession et al. (2003) also noted changes in channel width between forested and non-forested riparian zone streams, but noted that there were no significant differences between the two conditions with regards to bankfull depth. Thus, the width of the channel may change with different riparian land use, but the depth from the bottom of the stream to the stage where floodplain inundation would occur may not change. Davies-Colley (1997) suggested that the higher channel widths in forested areas may be due to erosion following bank vegetation loss from tree canopy shading associated with forested riparian areas.

These findings supporting higher connectivity in forested areas relate back to the importance of maintaining riparian buffers in river floodplain systems. Since there is higher connectivity in highly forested watershed, water is on the floodplain for a longer time covering more area. This can, in turn, lead to more ecosystem benefits in terms of water quality improvements. As aforementioned, riparian buffers offer many benefits including sediment removal (Cooper et al., 1987), among others as discussed in the literature review. Riparian zones are known to have significant benefits to the surrounding ecosystem (Gilliam, 1994; Pusey and Arthington, 2003), but these benefits could be even more drastic when realizing that forested areas have higher levels of floodplain connectivity, thus introducing more opportunities for these preexisting benefits to be realized.

The last watershed characteristic to be examined for its relationship with annual floodplain connectivity was the watersheds' mean 100-year, 24-hr precipitation depths. There was a significant increase in connectivity as the 100-year, 24-hr precipitation depth increased. This could potentially be due to the shear amount of water in systems where there is a larger mean 100-year, 24-hr precipitation depth. More precipitation generally correlates to more flow, and thus the potential for more above-bankfull events, leading to a higher annual connectivity metric value. Currently, large, infrequent precipitation events are suggested to increase with climate change (Mendizabal, et al., 2014), thus potentially having effects on river-floodplain connectivity, given the results from our study. Trenberth (2011) outlined the detailed climate-related effects of climate change, including an increase in both drought and precipitation

intensity as temperatures increase. In a study relating to the effects of climate change, Muzick (2002) found that if rainfall mean and standard deviation increased by 25%, the 2 and 100 year flood events would increase significantly (80% and 41%, respectively).

6. Conclusions

6.1 Model Improvements & Future Research Opportunities

This spatiotemporal floodplain connectivity model could be improved in several ways throughout all the major model components. First, the flood simulations and the spatial GIS model could have modifications and improvements to further advance the accuracy of the model. In the development of the flood surface ("surface water polygon"), various DEM filters were used to estimate the location of the bankfull water surface. After "slicing" this polygon of the water surface into ~20 m sections, the maximum elevation in that zone was assigned to the zone and served as the base to which the artificial stages were added for flooding purposes. This is a relatively crude methodology and could be improved significantly. Besides not being entirely accurate, the initial DEM filtering and proceeding polygon development steps were extremely manually intensive, thus potentially introducing error. Automating the development of the flood surface would greatly improve the efficiency as well as accuracy of the model, reducing human bias and influence. The "stair-stepping" method of providing a flat surface upon which to add the simulated flood stages is also another part of the model that could be improved or changed. The stair-stepping pattern showed up as artifacts in some of the inundation area maps during the analysis process. Another alternative to this flood surface is to flatten, or "detrend" the entire surface, as seen in the methods used by Jones et al. (2008). Besides the actual flood surface, some potential error was introduced during the statistical summary of inundation areas during the flood simulations. As mentioned in the methods, some cross-sections were manually moved upstream (contrary to the automated placement) due to effects of the confluence (Figure). Even after moving the cross-sections, the question remains as to whether the cross-section location was measuring a representative cross-sectional inundated area for the reach and whether inundation at the confluence still had effects at the new cross-section location. Additional error could have come from the DEM data source, relating to the 3 m horizontal resolution of the data. Given the small size of headwater streams, this elevation resolution may not have been fine enough to fully capture topography and floodplain connectivity characteristics of each reach. Higher resolution data, however, would limit the available areas where this spatiotemporal hydrologic floodplain connectivity model could be used.

Next, there are areas for improvement and elimination of error in the temporal MATLAB model, which was focused on developing the regression equations for FDC prediction as well as on the ungauged hydrograph development. The primary way that the model could be improved is in the development of the ungauged hydrograph. Improvement could come in the area of the FDC regression or the methodology used to create the full ungauged hydrograph. For the regression, an automated procedure could be developed to estimate the validation statistics for all possible combinations of independent variables. This could reveal another combination of watershed characteristics that better explain the data variation to predict various points on the FDC. Also, more points along the FDC could be developed to reduce interpolation error when utilizing the FDC. Besides adding flow percentiles to the FDC, the existing percentiles could be adjusted and changed to more accurately capture the nuances of the actual FDC. As described in the methodology, a "drainage area factor" was used to eliminate gauges from the regression if they were outside the factor-determined range of drainage areas as compared to the study watershed's drainage area. The study watershed's area was actually the maximum area the regression equations would be applied to. Many of the reaches and sub-watersheds within the full Mahantango Creek watershed had drainage areas much below the lower threshold set with the drainage area factor. Thus, in the future, the drainage area factor may be used to set a maximum drainage area limit, but there should be no lower threshold to the drainage area size allowable to include that particular gauge in the regression analysis. Besides improvement in the regression itself, there are other options for determining the gauge from which the ungauged flow timing would be based. In the validation portion of this study, the nearest gauge by distance between watershed centroids was used. For the application portion of this study, all the reaches within the Mahantango Creek watershed utilized the Mahantango outlet gauge (USGS 1555500) to determine the flow timing. Other studies have utilized more complex methods of determining the gauge by which the timing is best estimated. Future research in this area could further examine ways to select the best gauge to use for flow timing at an ungauged location while utilizing the QPPQ method.

Lastly, the connectivity model could be improved by adjusting the way in which the metric was normalized. As built, the model normalized the annual connectivity metric by summing the connectivity across the entire record and then dividing by the length of record to get a connectivity value per day. This was then multiplied by 365 to estimate the annual metric.

This may not be the best way to summarize the connectivity data. A similar normalization technique was used with the seasonal connectivity metrics, both of which could be adjusted to more accurately represent the connectivity over the course of that season. The FDCs created in this study were for the full record, and thus "period of record" FDCs, as described by Vogel and Fennessey (1994). Using a combination of annual FDCs may lead to better hydrograph prediction, and thus more accurate connectivity estimates (Vogel and Fennessey, 1994).

Future work on this spatiotemporal model for analyzing floodplain hydrologic connectivity should include both sensitivity analysis and model validation. Although some adjustment of the Manning's n value was completed in the seasonal connectivity analysis, a more specific and detailed sensitivity analysis should take place on the Manning's n variable as well as the bankfull flow percentile (assumed 2.5% in this study). Additionally, due to the uniqueness of this study, there is currently not enough accessible data to perform an intensive validation of the model results. A long term time series of inundation area field measurements and durations would be needed to validate the connectivity results.

6.2 Spatiotemporal Floodplain Connectivity

According to the results presented in this study, QPPQ method of developing an ungauged hydrograph may be acceptable for some applications, but might not be a fully accurate way of predicting flow. The hydrographs predicted for validation purposes in this study had varying results, with some being moderately accurate (NSE= 0.64), while others were extremely poor (NSE< 0). Although there were several areas for improvement in the specific QPPQ methodology used in this study, another alternative for ungauged hydrograph prediction may be better for future studies.

The spatial component results from this study suggest that smaller streams have less inundation area, but are more frequently completely inundated than larger streams. The patterns of area of inundation in terms of both strict inundation as well as floodplain coverage show that larger river systems have large floodplains that are rarely completely inundated, but have different inundation areas for different storm events. Storm size has little effect on inundation area in small streams less than 10 km^2.

There were significant relationships found between connectivity and drainage area, slope, and land use characteristics of the watersheds studied. Drainage area and slope had inverse relationships with connectivity due to the inverse relationship between reach slope and drainage area. Larger stream systems (large drainage area, lower reach slope) generally had a higher annual connectivity, indicating that there was more inundated area for a longer period of time when comparing to a small stream (small drainage area, high reach slope) system. These results suggest that there is higher potential for pollutant removal and water quality improvement from the floodplains of larger rivers rather than small streams. Given this, maximum water quality benefits may be realized by restoring the floodplain of a higher order reach since this study showed the floodplain was more connected to the channel on an annual basis. When examining the total annual connectivity (m^2*days), which accounted for reach length and number of reaches in each stream order, however, headwater streams showed more potential for overall impact on watershed outlet water quality. Thus, from a restoration perspective, many small streams would need restoration to counterbalance potential benefits from a single large-river restoration.

This study furthered the understanding of relationships between flooding patterns, land use, flow regime, and channel morphology. There was higher connectivity during the non-growing season; however, water quality benefits may be limited due to decreased microbial activity during that time period. The inverse effects of agricultural versus forested land use on connectivity support the concept model relating land use and flooding patterns. First, land use change occurs, leading to changes in the local flow regime. This change in flow regime creates new erosion patterns, leading to a change in channel geomorphology. When channel geomorphology changes, flooding may occur at a different flood stage, thus affecting river-floodplain connectivity. The findings of this study showed that increased agricultural area, related to decreased forested area, resulted in decreased connectivity, and therefore decreased potential for floodplain benefits.

References

Admiraal, W., D. Jacobs, P. Breugem, and E. D. D. Vansteveninck. 1992. Effects of phytoplankton on the elemental composition (C, N, P) of suspended particulate material in the lower River Rhine. Hydrobiologia 235:479-489.

Alexander, R. B., R. A. Smith, and G. E. Schwarz. 2000. Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico. Nature 403(6771):758-761.

Amoros, C., and G. Bornette. 2002. Connectivity and biocomplexity in waterbodies of riverine floodplains. Freshwater Biology 47(4):761-776.

ArcGIS. 2012. World Topographic Map. Available at: http://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f. Accessed 23 May, 2012.

Archfield, S.A., R.M. Vogel, P.A. Steeves, S.L. Brandt, P.K. Weiskel, and S.P. Garabedian. 2010. The Massachusetts Sustainable-Yield Estimator: a decision-support tool to assess water availability at ungauged stream locations in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2009-5227.

Baker, M. E., M. J. Wiley, and P. W. Seelbach. 2001. GIS-based hydrologic modeling of riparian areas: Implications for stream water quality. Journal of the American Water Resources Association 37(6):1615-1628.

Baldwin, D. S., and A. M. Mitchell. 2000. The effects of drying and re-flooding on the sediment and soil nutrient dynamics of lowland river-floodplain systems: A synthesis. Regulated Rivers-Research & Management 16(5):457-467.

Bates, P. D., M. D. Stewart, A. Desitter, M. G. Anderson, J. P. Renaud, and J. A. Smith. 2000. Numerical simulation of floodplain hydrology. Water Resources Research 36(9):2517-2529.

Bayley, P. B. 1995. Understanding Large River-Floodplain Ecosystems. BioScience 45(3):153-158.

Benke, A. C. 1990. A perspective on America vanishing streams. Journal of the North American Benthological Society 9(1):77-88.

Beyer, H. L. 2012. *Geospatial Modelling Environment*. Ver. 0.7.2.0. Spatial Ecology LLC.

Bisese, J. A. 1995. Methods for estimating the magnitude and frequency of peak discharges for rural, unregulated streams in Virginia. U.S. Geological Survey.

Boesch, D. F., R. B. Brinsfield, and R. E. Magnien. 2001. Chesapeake Bay eutrophication: Scientific understanding, ecosystem restoration, and challenges for agriculture. Journal of Environmental Quality 30(2):303-320.

Bohlke, J. K., and J. M. Denver. 1995. Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in 2 agricultural watersheds, Atlantic Coastal-Plain, Maryland. Water Resources Research 31(9):2319-2339.

Booker, D. J., and T. H. Snelder. 2012. Comparing methods for estimating flow duration curves at ungauged sites. Journal of Hydrology 434:78-94.

Bott, T. L. 2006. Primary productivity and community respiration. In Methods in stream ecology. R. F. Hauer, and G. A. Lamberti, eds. Amsterdam: Academic Press.

Boyer, J. N., and P. M. Groffman. 1996. Bioavailability of water extractable organic carbon fractions in forest and agricultural soil profiles. Soil Biology & Biochemistry 28(6):783-790.

Brinson, M. M., A. E. Lugo, and S. Brown. 1981. Primary productivity, decomposition and consumer activity in fresh-water wetlands. Annual Review of Ecology and Systematics 12:123-161.

Brunet, R. C., G. Pinay, F. Gazelle, and L. Roques. 1994. Role of the floodplain and riparian zone in suspended matter and nitrogen-retention in the Adour River, South-West France. Regulated Rivers-Research & Management 9(1):55-63.

Brunke, M., and T. Gonser. 1997. The ecological significance of exchange processes between rivers and groundwater. Freshwater Biology 37(1):1-33.

Bunya, S., J. C. Dietrich, J. J. Westerink, B. A. Ebersole, J. M. Smith, J. H. Atkinson, R. Jensen, D. T. Resio, R. A. Luettich, C. Dawson, V. J. Cardone, A. T. Cox, M. D. Powell, H. J. Westerink, and H. J. Roberts. 2010. A High-Resolution Coupled Riverine Flow, Tide, Wind, Wind Wave, and Storm Surge Model for Southern Louisiana and Mississippi. Part I: Model Development and Validation. Monthly Weather Review 138(2):345-377.

Burt, T. P., G. Pinay, F. E. Matheson, N. E. Haycock, A. Butturini, J. C. Clement, S. Danielescu, D. J. Dowrick, M. M. Hefting, A. Hillbricht-Ilkowska, and V. Maitre. 2002. Water table fluctuations in the riparian zone: comparative results from a pan-European experiment. Journal of Hydrology 265(1-4):129-148.

Bustillo, V., R. L. Victoria, J. M. S. de Moura, D. D. Victoria, A. M. A. Toledo, and E. Colicchio. 2011. Biogeochemistry of Carbon in the Amazonian Floodplains over a 2000-km Reach: Insights from a Process-Based Model. Earth Interactions 15:29.

Capesius, J. P. and V. C. Stephens. 2009. Regional regression equations for estimation of natural streamflow statistics in Colorado: U.S. Geological Survey Scientific Investigations Report 2009-5136. U. S. Geological Survey. Reston, VA.

Chapra, S. C. 1997. Gas transfer and oxygen reaeration. In Surface water-quality modeling. Long Grove, Illinois: Waveland Press, Inc.

Chesapeake Bay Program. 1997. Chesapeake Bay Nutrient Reduction Progress & Future Directions: Nutrient Reduction Reevaluation Summary Report. Chesapeake Bay Program.

Chin, A. 2006. Urban transformation of river landscapes in a global context. Geomorphology 79(3-4):460-487.

Cianfrani, C. M., W. C. Hession, and D. M. Rizzo. 2006. Watershed imperviousness impacts on stream channel condition in southeastern Pennsylvania. Journal of the American Water Resources Association 42(4):941-956.

Connell, J. H. 1978. Diversity in tropical rain forests and coral reefs- high diversity of trees and coral is maintained only in a non-equilibrium state. Science 199(4335):1302-1310.

Cooper, J. R., J. W. Gilliam, R. B. Daniels, and W. P. Robarge. 1987. Riparian areas as filters for agricultural sediment. Soil Science Society of America Journal 51(2):416-420.

Correll, D. 1997. Failure of agricultural riparian buffers to protect surface waters from groundwater nitrate contamination. 162-165. T. Jordan, ed. Cambridge: Cambridge University Press.

Davies-Colley, R. J. 1997. Stream channels are narrower in pasture than in forest. New Zealand Journal of Marine and Freshwater Research 31(5):599-608.

De Doncker, L., P. Troch, R. Verhoeven, K. Bal, P. Meire, and J. Quintelier. 2009. Determination of the Manning roughness coefficient influenced by vegetation in the river Aa and Biebrza river. Environmental Fluid Mechanics 9(5):549-567.

Deforet, T., P. Marmonier, D. Rieffel, N. Crini, C. Fritsch, P. Giraudoux, and D. Gilbert. 2008. The influence of size, hydrological characteristics and vegetation cover on nitrogen, phosphorus and organic carbon cycling in lowland river gravel bars (Doubs River, France. Fundamental and Applied Limnology 171(2):161-173.

Diaz, R. J., and R. Rosenberg. 2008. Spreading dead zones and consequences for marine ecosystems. Science 321(5891):926-929.

Downing, J. A., J. L. Baker, R. J. Diaz, T. Prato, N. N. Rabalais, and R. J. Zimmerman. 1999. Gulf of Mexico Hypoxia: Land and Sea Interactions. Council for Agricultural Science and Technology Task Force Report No. 134.

Dunne, T., L. A. K. Mertes, R. H. Meade, J. E. Richey, and B. R. Forsberg. 1998. Exchanges of sediment between the flood plain and channel of the Amazon River in Brazil. Geological Society of America Bulletin 110(4):450-467.

Emmett, B. A., B. Reynolds, P. A. Stevens, D. A. Norris, S. Hughes, J. Gorres, and I. Lubrecht. 1993. Nitrate leaching from afforested Welsh catchments- interactions between stand age and nitrogen deposition. Ambio 22(6):386-394.

Engineering, N. A. o. 2012. NAE Grand Challenges for Engineering. Available at: http://www.engineeringchallenges.org/cms/8996/9132.aspx.

Fabre, A., G. Pinay, and C. Ruffinoni. 1996. Seasonal changes in inorganic and organic phosphorus in the soil of a riparian forest. Biogeochemistry 35(3):419-432.

Feaster, T. D., A. J. Gotvald, and J. C. Weaver. 2009. Magnitude and frequency of rural floods in the Southeastern United States, 2006- Volume 1, Georgia: U. S. Geological Survey Scientific Investigations Report 2009-5043. U. S. Geological Survey.

Fennessey, N., and R. M. Vogel. 1990. Regional flow-duration curves for ungauged sites in Massachusetts. Journal of Water Resources Planning and Management-Asce 116(4):530-549.

Fennessey, N. M. 1994. A hydro-climatological model of daily streamflow for the northeast United States. Tufts University, Medford, MA

Forshay, K. J., and E. H. Stanley. 2005. Rapid nitrate loss and denitrification in a temperate river floodplain. Biogeochemistry 75(1):43-64.

Friedman, J. M., W. R. Osterkamp, M. L. Scott, and G. T. Auble. 1998. Downstream effects of dams on channel geometry and bottomland vegetation: Regional patterns in the Great Plains. Wetlands 18(4):619-633.

Gergel, S. E. 2002. Assessing cumulative impacts of levees and dams on floodplain ponds: A neutral-terrain model approach. Ecological Applications 12(6):1740-1754.

Gergel, S. E., S. R. Carpenter, and E. H. Stanley. 2005. Do dams and levees impact nitrogen cycling? Simulating the effects of flood alterations on floodplain denitrification. Global Change Biology 11(8):1352-1367.

Gilliam, J. W. 1994. Riparian wetlands and water quality. Journal of Environmental Quality 23(5):896-900.

Gold, A. J., P. M. Groffman, K. Addy, D. Q. Kellogg, M. Stolt, and A. E. Rosenblatt. 2001. Landscape attributes as controls on ground water nitrate removal capacity of riparian zones. Journal of the American Water Resources Association 37(6):1457-1464.

Goolsby, D. A., W. A. Battaglin, B. T. Aulenbach, and R. P. Hooper. 2001. Nitrogen input to the Gulf of Mexico. Journal of Environmental Quality 30(2):329-336.

Gregory, S. V., F. J. Swanson, W. A. McKee, and K. W. Cummins. 1991. An ecosystem perspective of riparian zones. BioScience 41(8):540-551.

Grimm, N. B., A. Chacon, C. N. Dahm, S. W. Hostetler, O. T. Lind, P. L. Starkweather, and W. W. Wurtsbaugh. 1997. Sensitivity of aquatic ecosystems to climatic and anthropogenic changes: The Basin and Range, American Southwest and Mexico. Hydrological Processes 11(8):1023-1041.

Grimm, N. B., S. E. Gergel, W. H. McDowell, E. W. Boyer, C. L. Dent, P. Groffman, S. C. Hart, J. Harvey, C. Johnston, E. Mayorga, M. E. McClain, and G. Pinay. 2003. Merging aquatic and terrestrial perspectives of nutrient biogeochemistry. Oecologia 137(4):485-501.

Groffman, P. M., and G. C. Hanson. 1997. Wetland denitrification: Influence of site quality and relationships with wetland delineation protocols. Soil Science Society of America Journal 61(1):323-329.

Grubaugh, J. W., and R. V. Anderson. 1989. Upper Mississippi River- seasonal and floodplain forest influences on organic-matter transport. Hydrobiologia 174(3):235-244.

Hallegraeff, G. M. 1993. A review of harmful algal blooms and their apparent global increase. Phycologia 32(2):79-99.

Haycock, N. E., and T. P. Burt. 1993. Role of floodplain sediments in reducing the nitrate concentration of subsurface run-off- a case-study in the Cotswolds, UK. Hydrological Processes 7(3):287-295.

Heffernan, J. B., R. A. Sponseller, and S. G. Fisher. 2008. Consequences of a biogeomorphic regime shift for the hyporheic zone of a Sonoran Desert stream. Freshwater Biology 53(10):1954-1968.

Heiazi, M. I., and G. E. Moglen. 2007. Regression-based approach to low flow prediction in the Maryland Piedmont region under joint climate and land use change. Hydrological Processes 21(14):1793-1801.

Heiler, G., T. Hein, F. Schiemer, and G. Bornette. 1995. Hydrological connectivity and flood pulses as the central aspects for the integrity of a river-floodplain system. Regulated Rivers-Research & Management 11(3-4):351-361.

Hein, T., W. Reckendorfer, J. H. Thorp, and F. Schiemer. 2005. The role of slackwater areas for biogeochemical processes in rehabilitated river corridors: examples from the Danube. Archiv fuer Hydrobiologie Supplement 155(1-4):425-442.

Hession, W. C., J. E. Pizzuto, T. E. Johnson, and R. J. Horwitz. 2003. Influence of bank vegetation on channel morphology in rural and urban watersheds. Geology 31(2):147-150.

Hill, A. R. 1990. Ground-water flow paths in relation to nitrogen chemistry in the near-stream zone. Hydrobiologia 206(1):39-52.

Hill, A. R., K. J. Devito, S. Campagnolo, and K. Sanmugadas. 2000. Subsurface denitrification in a forest riparian zone: Interactions between hydrology and supplies of nitrate and organic carbon. Biogeochemistry 51(2):193-223.

Hisdal, H., and O. E. Tveito. 1992. Generation of runoff series at ungauged locations using empirical orthogonal functions in combination with kriging. Stochastic Hydrology and Hydraulics 6(4):255-269.

Hollis, G. E. 1975. The effect of urbanization on floods of different recurrence interval. Water Resources Research 11(3):431-435.

Howarth, R. W., G. Billen, D. Swaney, A. Townsend, N. Jaworski, K. Lajtha, J. A. Downing, R. Elmgren, N. Caraco, T. Jordan, F. Berendse, J. Freney, V. Kudeyarov, P. Murdoch, and Z. L. Zhu. 1996. Regional nitrogen budgets and riverine N&P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences. Biogeochemistry 35(1):75-139.

Hudson, P. F., M. A. Sounny-Slittine, and M. LaFevor. 2013. A new longitudinal approach to assess hydrologic connectivity: Embanked floodplain inundation along the lower Mississippi River. Hydrological Processes 27(15):2187-2196.

Hughes, D. A. 1980. Floodplain inundation- processes and relationships with channel discharge. Earth Surface Processes and Landforms 5(3):297-304.

Hughes, D. A., and V. Smakhtin. 1996. Daily flow time series patching or extension: A spatial interpolation approach based on flow duration curves. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 41(6):851-871.

Hunter, R. G., and S. P. Faulkner. 2001. Denitrification potentials in restored and natural bottomland hardwood wetlands. Soil Science Society of America Journal 65(6):1865-1872.

Hupp, C. R. 2000. Hydrology, geomorphology and vegetation of Coastal Plain rivers in the south-eastern USA. Hydrological Processes 14(16-17):2991-3010.

Hupp, C. R., and W. R. Osterkamp. 1996. Riparian vegetation and fluvial geomorphic processes. Geomorphology 14(4):277-295.

Hupp, C. R., A. R. Pierce, and G. B. Noe. 2009. Floodplain geomorphic processes and environmental impacts of human alteration along coastal plain rivers, USA. Wetlands 29(2):413-429.

Jacinthe, P. A., W. A. Dick, and L. C. Brown. 2000. Bioremediation of nitrate-contaminated shallow soils and waters via water table management techniques: evolution and release of nitrous oxide. Soil Biology & Biochemistry 32(3):371-382.

Jenson, S. K., and J. O. Domingue. 1988. Extracting topographic structure from digital elevation data for geographic information-system analysis. Photogrammetric Engineering and Remote Sensing 54(11):1593-1600.

Johnson, W. C. 1994. Woodland expansion in the Platte River, Nebraska- patterns and causes. Ecological Monographs 64(1):45-84.

Johnson, W. C. 1998. Adjustment of riparian vegetation to river regulation in the great plains, USA. Wetlands 18(4):608-618.

Johnston, C. A., S. D. Bridgham, and J. P. Schubauer-Berigan. 2001. Nutrient dynamics in relation to geomorphology of riverine wetlands. Soil Science Society of America Journal 65(2):557-577.

Jones, K. L., G. C. Poole, S. J. O'Daniel, L. A. K. Mertes, and J. A. Stanford. 2008. Surface hydrology of low-relief landscapes: Assessing surface water flow impedance using LIDAR-derived digital elevation models. Remote Sensing of Environment 112(11):4148-4158.

Junk, W. J., P. B. Bayley, and R. E. Sparks. 1989. The flood pulse concept in river-floodplain systems. Canadian Special Publication of Fisheries and Aquatic Sciences 106:110-127.

Kalyanapu, A. J., S. J. Burian, and T. N. McPherson. 2009. Effect of land use-based surface roughness on hydrologic model output. Journal of Spatial Hydrology 9(2):51-70.

Klocker, C. A., S. S. Kaushal, P. M. Groffman, P. M. Mayer, and R. P. Morgan. 2009. Nitrogen uptake and denitrification in restored and unrestored streams in urban Maryland, USA. Aquatic Sciences 71(4):411-424.

Knowlton, M. F., and J. R. Jones. 1997. Trophic status of Missouri River floodplain lakes in relation to basin type and connectivity. Wetlands 17(4):468-475.

Kundzewicz, Z. W., L. J. Mata, N. W. Arnell, P. Doll, P. Kabat, B. Jimenez, K. A. Miller, T. Oki, Z. Sen, and I. A. Shiklomanov. 2007. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to theFourth Assessment Report of the Intergovernmental Panel on Climate Change. 173-210. Cambridge, UK: Cambridge University Press.

Lane, R. R., J. W. Day, and B. Thibodeaux. 1999. Water quality analysis of a freshwater diversion at Caernarvon, Louisiana. Estuaries 22(2A):327-336.

Li, M., Q. X. Shao, L. Zhang, and F. H. S. Chiew. 2010. A new regionalization approach and its application to predict flow duration curve in ungauged basins. Journal of Hydrology 389(1-2):137-145.

Lowrance, R. R., R. L. Todd, and L. E. Asmussen. 1984. Nutrient cycling in an agricultural watershed 1. Phreatic movement. Journal of Environmental Quality 13(1):22-27.

McClain, M. E., E. W. Boyer, C. L. Dent, S. E. Gergel, N. B. Grimm, P. M. Groffman, S. C. Hart, J. W. Harvey, C. A. Johnston, E. Mayorga, W. H. McDowell, and G. Pinay. 2003. Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. Ecosystems 6(4):301-312.

McClain, M. E., J. E. Richey, J. A. Brandes, and T. P. Pimentel. 1997. Dissolved organic matter and terrestrial-lotic linkages in the central Amazon basin of Brazil. Global Biogeochemical Cycles 11(3):295-311.

Meade, R. H., T. Dunne, J. E. Richey, U. D. Santos, and E. Salati. 1985. Storage and remobilization of suspended sediment in the lower Amazon River of Brazil. Science 228(4698):488-490.

Mendizabal, M., J. Sepulveda, and P. Torp. 2014. Climate Change Impacts on Flood Events and Its Consequences on Human in Deba River. International Journal of Environmental Research 8(1):221-230.

Mertes, L. A. K. 1994. Rates of floodplain sedimentation on the central Amazon River. Geology 22(2):171-174.

Mertes, L. A. K. 1997. Documentation and significance of the perirheic zone on inundated floodplains. Water Resources Research 33(7):1749-1762.

Michener, W. K., E. R. Blood, J. B. Box, C. A. Couch, S. W. Golladay, D. J. Hippe, R. J. Mitchell, and B. J. Palik. 1998. Tropical storm flooding of a coastal main landscape - Extensive foodplains ameliorated potential adverse effects on water quality, fishes, and molluskan communities. BioScience 48(9):696-705.

Michener, W. K., and R. A. Haeuber. 1998. Flooding: Natural and Managed Disturbances - A special issue of BioScience devoted to flooding as a disturbance. BioScience 48(9):677-680.

Mitsch, W. J., J. W. Day, J. W. Gilliam, P. M. Groffman, D. L. Hey, G. W. Randall, and N. M. Wang. 2001. Reducing nitrogen loading to the Gulf of Mexico from the Mississippi River Basin: Strategies to counter a persistent ecological problem. BioScience 51(5):373-388.

Mohamoud, Y. M. 2008. Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 53(4):706-724.

Molles, M. C., C. S. Crawford, L. M. Ellis, H. M. Valett, and C. N. Dahm. 1998. Managed flooding for riparian ecosystem restoration - Managed flooding reorganizes riparian forest ecosystems along the middle Rio Grande in New Mexico. BioScience 48(9):749-756.

Montreuil, O., C. Cudennec, and P. Merot. 2011. Contrasting behaviour of two riparian wetlands in relation to their location in the hydrographic network. Journal of Hydrology 406(1-2):39-53.

Moore, M. R. 2011. Development of a high-resolution 1D/2D coupled flood simulation of Charles City, Iowa. University of Iowa, Iowa Research Online

Mosier, A. R., M. A. Bleken, P. Chaiwanakupt, E. C. Ellis, J. R. Freney, R. B. Howarth, P. A. Matson, K. Minami, R. Naylor, K. N. Weeks, and Z. L. Zhu. 2001. Policy implications of human-accelerated nitrogen cycling. Biogeochemistry 52(3):281-320.

Mulholland, P. J., A. M. Helton, G. C. Poole, R. O. Hall, S. K. Hamilton, B. J. Peterson, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, W. K. Dodds, S. E. G. Findlay, S. V. Gregory, N. B. Grimm, S. L. Johnson, W. H. McDowell, J. L. Meyer, H. M. Valett, J. R. Webster, C. P. Arango, J. J. Beaulieu, M. J. Bernot, A. J. Burgin, C. L. Crenshaw, L. T. Johnson, B. R. Niederlehner, J. M. O'Brien, J. D. Potter, R. W. Sheibley, D. J. Sobota, and S. M. Thomas. 2008. Stream denitrification across biomes and its response to anthropogenic nitrate loading. Nature 452(7184):202-U246.

Mulholland, P. J., J. L. Tank, D. M. Sanzone, W. M. Wollheim, B. J. Peterson, J. R. Webster, and J. L. Meyer. 2000. Nitrogen cycling in a forest stream determined by a N-15 tracer addition. Ecological Monographs 70(3):471-493.

Muzik, I. 2002. A first-order analysis of the climate change effect on flood frequencies in a subalpine watershed by means of a hydrological rainfall-runoff model. Journal of Hydrology 267(1-2):65-73.

NAIP. 2010. 10:1 Compressed NAIP. National Agricultural Imagery Program. USDA FSA. Available at: http://viewer.nationalmap.gov/viewer/. Accessed 17 January, 2014.

NASS, USDA. 1997. Usual Planting and Harvesting Dates. National Agricultural Statistics Service Agricultural Handbook Number 628. Available at: usda.mannlib.cornell.edu/usda/nass/planting/uph97.pdf. Accessed 10 February 2014.

National Research Council. 2002. Riparian Areas: Functions and Strategies for Management. National Academy Press, Washington, DC.

NHDPlus. National Hydrography Dataset Plus. Version 1. Horizon Systems Corporation. Available at: http://www.horizon-systems.com/nhdplus/nhdplusv1_home.php.

NOAA. 2006. Precipitation Frequency for Ohio River Basin, USA- NOAA Atlas 14 Volume 2. Silver Spring, MD: NOAA/NWS/Office of Hydrologic Development, Hydrometeorological Design Studies Center. Available at: hdsc.nws.noaa.gov/hdsc/pfds/pfds_gis.html. Accessed 16 December 2013.

Noe, G. B., C. R. Hupp, and N. B. Rybicki. 2013. Hydrogeomorphology Influences Soil Nitrogen and Phosphorus Mineralization in Floodplain Wetlands. Ecosystems 16(1):75-94.

Peterjohn, W. T., and D. L. Correll. 1984. Nutrient dynamics in an agricultural watershed-observations on the role of a riparian forest. Ecology 65(5):1466-1475.

Peterson, B. J., W. M. Wollheim, P. J. Mulholland, J. R. Webster, J. L. Meyer, J. L. Tank, E. Marti, W. B. Bowden, H. M. Valett, A. E. Hershey, W. H. McDowell, W. K. Dodds, S. K. Hamilton, S. Gregory, and D. D. Morrall. 2001. Control of nitrogen export from watersheds by headwater streams. Science 292(5514):86-90.

Phillips, C. B., and F. N. Scatena. 2013. Reduced channel morphological response to urbanization in a flood-dominated humid tropical environment. Earth Surface Processes and Landforms 38(9):970-982.

Pinay, G., J. C. Clement, and R. J. Naiman. 2002. Basic principles and ecological consequences of changing water regimes on nitrogen cycling in fluvial systems. Environmental Management 30(4):481-491.

Pinay, G., C. Ruffinoni, and A. Fabre. 1995. Nitrogen cycling in 2 riparian forest soils under different geomorphic conditions. Biogeochemistry 30(1):9-29.

Poole, G. C., J. A. Stanford, C. A. Frissell, and S. W. Running. 2002. Three-dimensional mapping of geomorphic controls on flood-plain hydrology and connectivity from aerial photos. Geomorphology 48(4):329-347.

Predick, K. I., and E. H. Stanley. 2010. Influence of vegetation and seasonal flow patterns on parafluvial nitrogen retention in a 7(th)-order river. Journal of the North American Benthological Society 29(3):904-917.

Preiner, S., I. Drozdowski, M. Schagerl, F. Schiemer, and T. Hein. 2008. The significance of side-arm connectivity for carbon dynamics of the River Danube, Austria. Freshwater Biology 53(2):238-252.

Pringle, C. 2003. The need for a more predictive understanding of hydrologic connectivity. Aquatic Conservation-Marine and Freshwater Ecosystems 13(6):467-471.

Pusey, B. J., and A. H. Arthington. 2003. Importance of the riparian zone to the conservation and management of freshwater fish: a review. Marine and Freshwater Research 54(1):1-16.

Rabalais, N. N., R. E. Turner, and W. J. Wiseman. 2002. Gulf of Mexico hypoxia, aka "The dead zone". Annual Review of Ecology and Systematics 33:235-263.

Racchetti, E., M. Bartoli, E. Soana, D. Longhi, R. R. Christian, M. Pinardi, and P. Viaroli. 2011. Influence of hydrological connectivity of riverine wetlands on nitrogen removal via denitrification. Biogeochemistry 103(1-3):335-354.

Richardson, W. B., E. A. Strauss, L. A. Bartsch, E. M. Monroe, J. C. Cavanaugh, L. Vingum, and D. M. Soballe. 2004. Denitrification in the Upper Mississippi River: rates, controls, and contribution to nitrate flux. Canadian Journal of Fisheries and Aquatic Sciences 61(7):1102-1112.

Roberts, J., and J. A. Ludwig. 1991. Riparian vegetation along current-exposure gradients in floodplain wetlands of the River Murray, Australia. Journal of Ecology 79(1):117-127.

Roland, M. A., and M. H. Stuckey. 2008. Regression equations for estimating flood flows at selected recurrence intervals for ungaged streams in Pennsylvania: U. S. Geological Survey Scientific Investigations Report 2008-5102.

Roley, S. S., J. L. Tank, and M. A. Williams. 2012. Hydrologic connectivity increases denitrification in the hyporheic zone and restored floodplains of an agricultural stream. Journal of Geophysical Research-Biogeosciences 117:16.

Ross, S. T., and J. A. Baker. 1983. The response of fishes to periodic spring floods in a southeastern stream. American Midland Naturalist 109(1):1-14.

Royer, T. V., and M. B. David. 2005. Export of dissolved organic carbon from agricultural streams in Illinois, USA. Aquatic Sciences 67(4):465-471.

SAGA. 2007. *System for Automated Geoscientific Analyses*. Ver. 2.0.8.

Sanders, B. F. 2007. Evaluation of on-line DEMs for flood inundation modeling. Advances in Water Resources 30(8):1831-1843.

Sarhadi, A., S. Soltani, and R. Modarres. 2012. Probabilistic flood inundation mapping of ungauged rivers: Linking GIS techniques and frequency analysis. Journal of Hydrology 458:68-86.

Scaroni, A. E., J. A. Nyman, and C. W. Lindau. 2011. Comparison of denitrification characteristics among three habitat types of a large river floodplain: Atchafalaya River Basin, Louisiana. Hydrobiologia 658(1):17-25.

Scavia, D., N. N. Rabalais, R. E. Turner, D. Justic, and W. J. Wiseman. 2003. Predicting the response of Gulf of Mexico hypoxia to variations in Mississippi River nitrogen load. Limnology and Oceanography 48(3):951-956.

SDV. *Soil Data Viewer*. Ver. 6.1. USDA NRCS.

Sedell, J. R., J. E. Richey, and F. J. Swanson. 1989. The river continuum concept: a basis for the expected ecosystem behavior of a very large rivers? Canadian Special Publication of Fisheries and Aquatic Sciences 106:49-55.

Seitzinger, S. P., R. V. Styles, E. W. Boyer, R. B. Alexander, G. Billen, R. W. Howarth, B. Mayer, and N. Van Breemen. 2002. Nitrogen retention in rivers: model development and application to watersheds in the northeastern USA. Biogeochemistry 57(1):199-237.

Shields, F. D., A. Simon, and L. J. Steffen. 2000. Reservoir effects on downstream river channel migration. Environmental Conservation 27(1):54-66.

Shih, S. F., and G. S. Rahi. 1982. Seasonal variations of Manning's roughness coefficient in a subtropical marsh. Transactions of the ASAE 25(1):116-119.

Simmons, D. L., and R. J. Reynolds. 1982. Effects of urbanization on base flow of selected south-shore streams, Long Island, New York. Journal of the American Water Resources Association 18(5):797-805.

Smith, V. H. 2003. Eutrophication of freshwater and coastal marine ecosystems - A global problem. Environmental Science and Pollution Research 10(2):126-139.

Spink, A., R. E. Sparks, M. Van Oorschot, and J. T. A. Verhoeven. 1998. Nutrient dynamics of large river floodplains. Regulated Rivers-Research & Management 14(2):203-216.

Ssegane, H., D. M. Amatya, E. W. Tollner, Z. H. Dai, and J. E. Nettles. 2013. Estimation of daily streamflow of southeastern coastal plain watersheds by combining estimated magnitude and sequence. Journal of the American Water Resources Association 49(5):1150-1166.

Stanford, J. A., and J. V. Ward. 1993. An ecosystem perspective of alluvial rivers- connectivity and the hyporheic corridor. Journal of the North American Benthological Society 12(1):48-60.

Strahler, A.N. 1957. Quantitative analysis of watershed geomorphology. Transactions of the American Geophysical Union 38(6):913-920.

Strauss, E. A., W. B. Richardson, J. C. Cavanaugh, L. A. Bartsch, R. M. Kreiling, and A. J. Standorf. 2006. Variability and regulation of denitrification in an Upper Mississippi River backwater. Journal of the North American Benthological Society 25(3):596-606.

Stuckey, M. H., E. H. Koerkle, and J. E. Ulrich. 2012. Estimation of baseline daily mean streamflows for ungaged locations on Pennsylvania streams, water years 1960-2008: U. S. Geological Survey Scientific Investigations Report 2012-5142.

Swanson, F. J., S. L. Johnson, S. V. Gregory, and S. A. Acker. 1998. Flood disturbance in a forested mountain landscape - Interactions of land use and floods. BioScience 48(9):681-689.

Sweeney, B. W., T. L. Bott, J. K. Jackson, L. A. Kaplan, J. D. Newbold, L. J. Standley, W. C. Hession, and R. J. Horwitz. 2004. Riparian deforestation, stream narrowing, and loss of stream ecosystem services. Proceedings of the National Academy of Sciences of the United States of America 101(39):14132-14137.

Syvitski, J. P. M., C. J. Vorosmarty, A. J. Kettner, and P. Green. 2005. Impact of humans on the flux of terrestrial sediment to the global coastal ocean. Science 308(5720):376-380.

Tabacchi, E., D. L. Correll, R. Hauer, G. Pinay, A. M. Planty-Tabacchi, and R. C. Wissmar. 1998. Development, maintenance and role of riparian vegetation in the river landscape. Freshwater Biology 40(3):497-516.

Taylor, P. D., L. Fahrig, K. Henein, and G. Merriam. 1993. Connectivity is a vital element of landscape structure. Oikos 68(3):571-573.

Theiling, C. H., and J. T. Burant. 2013. Flood inundation mapping for integrated floodplain management: Upper Mississippi River system. River Research and Applications 29(8):961-978.

Thorp, J. H., and M. D. Delong. 1994. The riverine productivity model- an heuristic view of carbon-sources and organic-processing in large river ecosystems. Oikos 70(2):305-308.

Tockner, K., F. Malard, and J. V. Ward. 2000. An extension of the flood pulse concept. Hydrological Processes 14(16-17):2861-2883.

Tockner, K., D. Pennetzdorfer, N. Reiner, F. Schiemer, and J. V. Ward. 1999a. Hydrological connectivity, and the exchange of organic matter and nutrients in a dynamic river-floodplain system (Danube, Austria). Freshwater Biology 41(3):521-535.

Tockner, K., F. Schiemer, C. Baumgartner, G. Kum, E. Weigand, I. Zweimuller, and J. V. Ward. 1999b. The Danube restoration project: Species diversity patterns across connectivity gradients in the floodplain system. Regulated Rivers-Research & Management 15(1-3):245-258.

Tremolieres, M., U. Roeck, J. P. Klein, and R. Carbiener. 1994. The exchange process between river and groundwater on the central Alsace floodplain (Eastern France) 2. The case of a river with functional floodplain. Hydrobiologia 273(1):19-36.

Trenberth, K. E. 2011. Changes in precipitation with climate change. Climate Research 47(1):123-138.

Tritthart, M., N. Welti, E. Bondar-Kunze, G. Pinay, T. Hein, and H. Habersack. 2011. Modelling highly variable environmental factors to assess potential microbial respiration in complex floodplain landscapes. Environmental Modelling & Software 26(9):1097-1111.

USDA/NRCS. 2006. Digital General Soil Map of U.S. USDA/NRCS- National Geospatial Center of Excellence.

U. S. Geological Survey (USGS). 1/9 Arc Second National Elevation Dataset. Sioux Falls, SD. Available at: http://viewer.nationalmap.gov.

U.S. Geological Survey. 2001. National Water Information System data available on the World Wide Web (USGS Water Data for the Nation). Available at: http://waterdata.usgs.gov/nwis/dv/?referred_module=sw.

U. S. Geological Survey. 2011. National Land Cover Database 2006. Sioux Falls, SD. Available at: http://datagateway.nrcs.usda.gov.

Vandenbrink, F. W. B., J. Deleeuw, G. Vandervelde, and G. M. Verheggen. 1993. The impact of hydrology on the chemistry and phytoplankton development in floodplain lakes along the lower Rhine and Meuse. Biogeochemistry 19(2):103-128.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. River continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37(1):130-137.

Venterink, H. O., F. Wiegman, G. E. M. Van der Lee, and J. E. Vermaat. 2003. Role of active floodplains for nutrient retention in the river Rhine. Journal of Environmental Quality 32(4):1430-1435.

Vidon, P. G. F., and A. R. Hill. 2004a. Landscape controls on nitrate removal in stream riparian zones. Water Resources Research 40(3):14.

Vidon, P. G. F., and A. R. Hill. 2004b. Landscape controls on the hydrology of stream riparian zones. Journal of Hydrology 292(1-4):210-228.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human domination of Earth's ecosystems. Science 277(5325):494-499.

Vogel, R. M., and N. Fennessey. 1994. Flow-duration curves I: new interpretation and confidence intervals. Journal of Water Resources Planning and Management 120(4):485-504.

Vogel, R. M., and N. M. Fennessey. 1995. Flow duration curves 2. A review of applications in water-resources planning. Water Resources Bulletin 31(6):1029-1039.

Vogel, R. M., and C. N. Kroll. 1992. Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics. Water Resources Research 28(9):2451-2458.

Vogel, R. M., I. Wilson, and C. Daly. 1999. Regional regression models of annual streamflow for the United States. Journal of Irrigation and Drainage Engineering-Asce 125(3):148-157.

Ward, J. V., and A. S. Jack. 1983. The serial discontinuity concept of lotic ecosystems. In *Dynamics of Lotic Ecosystems,* 29-42. T.D. Fontaine and S. M. Bartell, eds. Ann Arbor Science Publishers.

Ward, J. V., and J. A. Stanford. 1995. The serial discontinuity concept- Extending the model to floodplain rivers. Regulated Rivers-Research & Management 10(2-4):159-168.

Ward, J. V., K. Tockner, P. J. Edwards, J. Kollmann, G. Bretschko, A. M. Gurnell, G. E. Petts, and B. Rossaro. 1999a. A reference river system for the Alps: The 'Fiume Tagliamento'. Regulated Rivers-Research & Management 15(1-3):63-75.

Ward, J. V., K. Tockner, and F. Schiemer. 1999b. Biodiversity of floodplain river ecosystems: Ecotones and connectivity. Regulated Rivers-Research & Management 15(1-3):125-139.

Webster, J. R., and B. C. Patten. 1979. Effects of watershed perturbation on stream potassium and calcium dynamics. Ecological Monographs 49(1):51-72.

White, M. D., and K. A. Greer. 2006. The effects of watershed urbanization on the stream hydrology and riparian vegetation of Los Penasquitos Creek, California. Landscape and Urban Planning 74(2):125-138.

Wissmar, R. C., and R. J. Swanson. 1990. Landscape disturbances and lotic ecotones Ecology and Management of Aquatic-Terrestrial Ecotones 4:65-89.

Wolf, K. L., G. B. Noe, and C. Ahn. 2013. Hydrologic Connectivity to Streams Increases Nitrogen and Phosphorus Inputs and Cycling in Soils of Created and Natural Floodplain Wetlands. Journal of Environmental Quality 42(4):1245-1255.

Yokoo, Y., and M. Sivapalan. 2011. Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis. Hydrology and Earth System Sciences 15(9):2805-2819.

Zhang, Q., D. C. Brady, and W. P. Ball. 2013. Long-term seasonal trends of nitrogen, phosphorus, and suspended sediment load from the non-tidal Susquehanna River Basin to Chesapeake Bay. Science of the Total Environment 452:208-221.

Appendix A Detailed GIS & MATLAB Methodologies

A.1 General Algorithm

The entire goal of the spatiotemporal model was to develop a metric to compare the degree of connectivity between various reaches and their respective floodplains. The model that was developed can be broken down into five main processes using two primary programs: MATLAB and ArcGIS (Figure A.1).

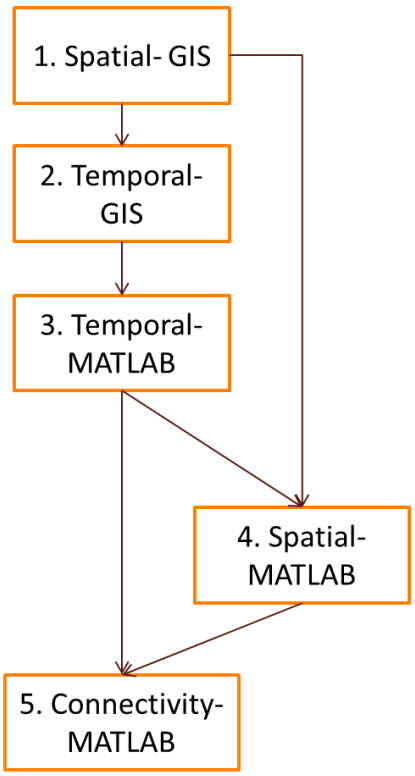

Figure A.1. Overall model diagram

The first main process (Step 1 in Figure A.1), Spatial-GIS, performed a surface water analysis and output simulated flood data establishing relationships between stage and cross-sectional area as well as stage and inundation area. This process was responsible for the bulk of

the actual flood simulations and initial data manipulation. After the raw data was processed and prepared for analysis, the first main task was to develop a polygon representing the water surface of the river network. This polygon was critical as it was the area to which artificial flood stages are added, as described in detail later. Next, inundation area was simulated at 0.5 m stage intervals for each reach, with horizontal spatial area and cross-sectional area data being summarized and tabulated. This raw data formed the base for developing key numerical relationships between both stage and inundation area as well as between stage and cross-sectional flooded area. Other processes in this first GIS spatial process include establishing a Manning's n coefficient for the floodplain as well as final data table manipulation and preparation for analysis in MATLAB.

To begin the temporal GIS process (Step 2, Figure A.1), some data is input from the first GIS model (spatial), in addition to USGS gauges and multiple datasets of raw data used to describe watershed characteristics (soils, land cover, etc.). Watershed characteristics were developed for each reach and each selected gauge and are used to develop regression equations in the final portion of the model: MATLAB temporal.

The temporal MATLAB process (Step 3, Figure A.1) predicted the long-term hydrograph at each ungauged reach based on regional regression equations for each of the 14 exceedance percentiles. The watershed attribute data (From Step 2, Figure A.1) were used to build a set of 14 regional regression equations to predict flow at each of the 14 flow percentiles. The ungauged flow duration curves (FDCs) as predicted by the regional regression equations, in addition to the hydrograph of the gauge nearest to the ungauged location, were then used to develop the ungauged hydrographs.

The flood data was then passed on to the fourth main process of the model: the MATLAB spatial process, which had the main goal of establishing the duration curve for the area of inundation. Thus, the output from this process included inundation areas for each reach for each of the 14 duration percentiles calculated earlier in the regional regression analysis (Step 3-Temporal MATLAB). First, the simulated flood data (cross-sectional areas at 0.5 m stage increments) and Manning's equation were used to relate the flows on the FDCs to a particular stage, thus first establishing a stage duration curve (SDC, stages at each of the 14 exceedance percentiles). Then, the relationship between inundation area (horizontal) and 0.5 m increment stage data was quantified. Lastly, using the newly created inundation area-stage relationship, the

previously created SDC was then be transformed to create the area of inundation duration curve (AIDC).

The last main process (Step 5, Figure A.1) to the overall model involved taking in all the data calculated in previous processes and calculating the connectivity metric values for each reach. First, the predicted hydrograph (From Step 3, Figure A.1) was converted from flow to inundation area using the AIDCs and FDCs for each reach. Then, flow at the point where flooding would occur was determined based on the 2.5% exceedance flow. Then, the long-term inundation area series, long-term hydrograph, and bankfull flow threshold were used to establish the connectivity metric (inundation area*time) by summing inundation areas when the flow for a particular day was above bankfull stage.

The detailed processes of each of the five primary processes are described in the rest of the methods section. All GIS ModelBuilder files and MATLAB scripts are available upon request from the author. For reader ease, a table of common acronyms has been provided to aid in understanding of the methodology (Table A.1). Additionally, the GIS model names used throughout the methods are provided in a summarized table in Appendix B.

Table A.1. Common acronyms and abbreviations

Acronym/Abbreviation	Meaning
AIDC	Area of inundation duration curve
ArcGIS	Program used for spatial GIS analysis
COMID	NHD reach ID number
DEM	Digital Elevation model
FDC	Flow duration curve
GIS	Geographic Information System
MATLAB	Numerical analysis program used in analysis
NHD	National Hydrography Dataset
SDC	Stage duration curve

A.1.1 Spatial GIS Process (Step 1)

The spatial GIS model consists of several components and processes that yield data necessary for the development of relationships between stage, discharge, inundation area, and

cross-sectional area. Inundation area was defined as the spatial area (length^2) of water in the floodplain when flow in a channel exceeds bankfull stage. Cross-sectional area (length^2) is defined as the area of water in 2-D cross-sectional space at specified representative cross-sections near the bottom of each defined reach segment. The GIS spatial algorithm outputs outlet nodes (points) and cross-sectional lines for each reach in addition to reach attribute and flood inundation area and cross-sectional area data (Figure A.2).

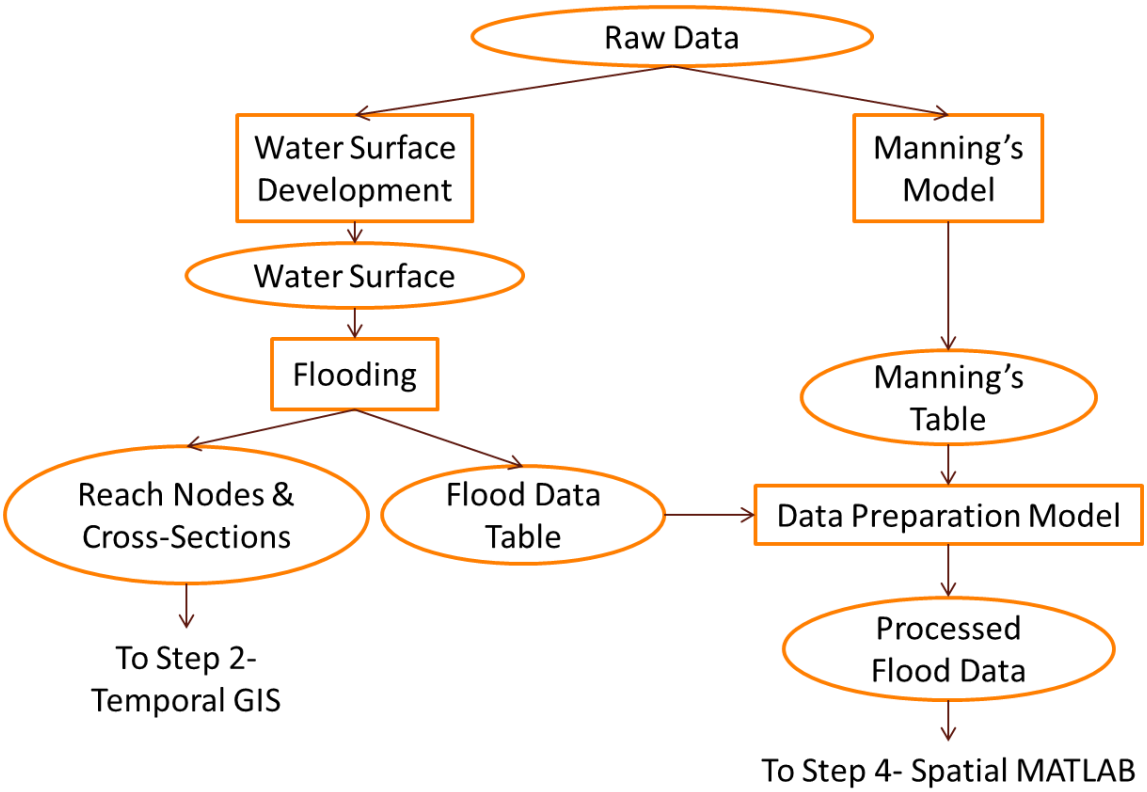

Figure A.2. GIS Spatial Process (Step 1)

A.1.1.1 Data Collection

The first step in the modeling process was to collect the necessary raw data, including elevation, land cover, and hydrography. Digital Elevation Models (DEMs) were downloaded from the USGS National Map Viewer (http://viewer.nationalmap.gov) at a 3 m horizontal resolution for the area of interest (USGS). The 1/9 arc second (~3 m) horizontal resolution was selected to maximize the areas where data was available in addition to providing enough detail to

capture the intricacies of floodplain topography. There was a data gap in the downloaded DEMs for the study watershed. This gap was approximately 1100 m x 100 m in size, and was filled by resampling and mosaicking a 10 m DEM grid. The data gap was located in an area (>0.2 km away from and ~20 m above nearby stream) such that it was assumed not to have significant effects on the artificial flood simulations and resulting inundation area relationships.

The National Land Cover Database (NLCD-2006) was downloaded from the USDA Geospatial Data Gateway at a 30 m resolution (USGS, 2011). This dataset was used to determine the Manning's n coefficient in the floodplain and to estimate land cover characteristics of the entire basin, thus it was deemed that at 30 m resolution would be adequate.

The last piece of data required to run the first main process was the NHDPlus data (NHDPlus). NHDPlus data includes not only the normal flowlines and catchment information, but has compiled many attributes regarding stream slope, flow information, etc. for each reach. The watershed boundary dataset (WBD) from the USDA Geospatial Data Gateway was also an important piece of the hydrography data that was not included in the NHDPlus dataset. The main advantage to this dataset is that it includes the Hydrologic Unit Code (HUC) boundaries (delineated watersheds) at the desired resolution (HUC8, HUC10, HUC12, etc.). These HUC boundaries were used later in the data processing to narrow down base datasets to the area of interest (the HUC boundary).

A.1.1.2 Water Surface & Flow Network Development

After data collection, the spatial GIS processes could be utilized to establish the relationships associated with different flow values and the flooded area. Thus, the end product from the spatial GIS process was flooded inundation area and cross-sectional flooded area for a range of artificial stages (Inundated Area, Cross-sectional area,$A_i, A_x = f(0.5\ m\ stages)$) in order to create relationships between these variables. The specific workflow of the spatial GIS model can be seen in Figure A.3.

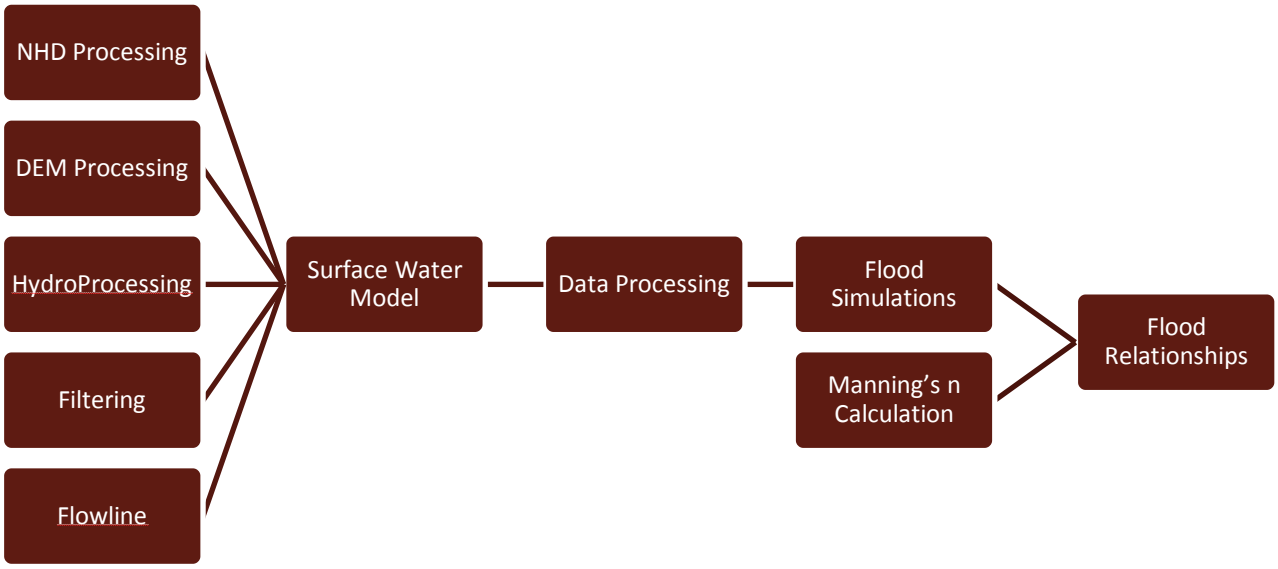

Figure A.3. Spatial GIS process specific workflow

As can be seen in Figure A.3, the first main step to the spatial GIS process was establishing what represents the surface water of the stream. This "surface water" was used to represent the location of the streams and rivers as represented by a polygon rather than a polyline to better represent the aerial nature of river surfaces. This data would later be used in the Flood model to determine inundation areas associated with various stages above the elevation of the surface water polygon.

Included in the surface water model (Second main process in Figure A.3, the set of tools used to develop the polygon representing the water surface of the rivers and streams in the watershed, diagrammed as the "Surface Water 2" GIS model in Figure A.4) were submodels and steps to initially process the downloaded data and begin developing the surface water polygon dataset. These include the NHD Processing, DEM Processing, Flowline, HydroProcessing, and Filtering submodels.

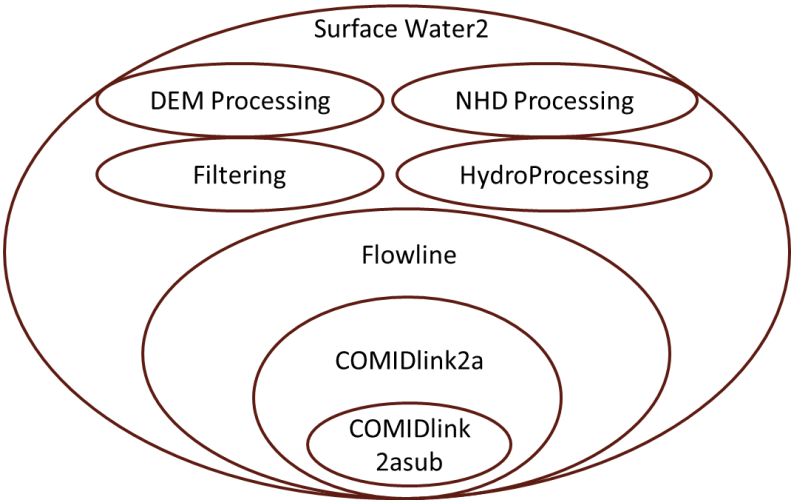

Figure A.4. Surface water model and submodel hierarchy.

First in the surface water model was the NHD processing submodel. The primary purpose of this submodel was to process the NHD data and develop useful datasets for later processing. The inputs to this submodel were the NHD flowline polyline shapefile from the NHDplus datasets, the HUC boundaries dataset, and the Stream Gauge point shapefile, also from the NHDplus datasests. The first steps in this submodel included projecting each of the three vector datasets into an appropriate coordinate system for the area, which for this study's watershed was NAD1983, UTM Zone 18N. Next, the desired HUC12 boundaries were selected with the ArcGIS "Select" tool. The modeler could change the select SQL expression based on the desired HUC12 number, or if there were multiple HUC12s that made up the study watershed. In this study, four HUC12s composed the Mahantango Creek watershed, and thus, the selection SQL expression was, ""HUC12" LIKE '020503010802' OR "HUC12" LIKE '020503010801' OR "HUC12" LIKE '020503010702' OR "HUC12" LIKE '020503010701"'. This expression utilized the "HUC12" attribute of the WBD dataset and selected the four HUC12s making up the Mahantango Creek watershed: 020503010802, 02050301081, 020503010707, and 020503010702. Next, because there were duplicate polygons with the same WBD data, the dissolve tool was used to create a single polygon from those selected previously, making the full watershed boundary as a polygon shapefile. Then, this watershed boundary was buffered by 1 km to use to clip later datasets for purposes of data file size and processing time management.

The last step of the NHD processing submodel was clipping the projected NHD flowline dataset to the buffered watershed boundary.

The next submodel in the surface water model was the DEM Processing submodel. First, this model mosaics all the raw DEMs into a single TIFF file. The TIFF file type was chosen in order to accommodate processing rasters in other open source GIS software, as described later. Next, this full DEM was extracted to the buffered watershed boundary. The last main portion of this submodel was to convert the floating point grid into an integer. In order to maintain the accuracy contained in the decimal places, the floating point grid was multiplied by 100, and then 0.5 was added to simulate rounding. This kept the accuracy contained in the decimal places, but allowed for integer conversion, which was necessary to keep file size low and to accommodate certain geoprocessing tools.

The clipped and buffered DEM was then sent to the HydroProcessing submodel. This submodel executed the geoprocessing tasks associated with delineating watersheds. First, the DEM was filled (using the ArcGIS Fill tool) to remove any data error pits that could result in flow direction or accumulation errors. This filled DEM was then sent to the flow direction tool, which determined the D8 flow direction according to the algorithm in Jenson and Domingue (1988). Next, the flow accumulation grid was developed from the flow direction grid. A temporary flow network grid was established by keeping the cells in the flow accumulation grid above a 200 cell threshold value. Lastly, the ArcGIS stream order tool was executed with option for the Strahler methodology (1957).

Next, the Flowline GIS model (Figure A.4) was used to develop a flow network that could then be processed manually to most accurately represent the system. The submodels, as described in detail later, were meant to properly assign the NHD COMID to each reach in the derived flow network. The NHD flowline data could not be directly used for modeling purposes, since due to differences in source elevation data, the datasets would not exactly match up. If the flow network does not match directly with the topography of the DEM, flood inundation characteristics developed later in the model would be inaccurate. The first step in the flowline model was to remove extremely high order (small streams) that were in the stream order dataset from the HydroProcessing submodel. A threshold of stream order 3 was set to eliminate stream segments designated as stream order 3 or less. These segments were most likely remnants of low lying areas on the DEM, and not necessarily perennial streams, as determined by comparisons

with the NHD data. The goal of the flowline model was to replicate the NHD dataset as closely as possible, but using this study's source elevation data. Thus, the next steps in the flowline model were data processing in order to most accurately mimic the NHD flowline dataset. The raster version of the flow network was then converted to a polyline with a minimum dangle length of 0.005. Then, the Simplify Line tool was used with the "BEND_SIMPLIFY" algorithm, a 100 m tolerance, and all other options left unchecked. The Smooth Line tool was then used with the PAEK algorithm option, a 75 m smoothing tolerance, and all other options unchecked. This output was then projected to the NAD1983 UTM Zone 18N coordinate system in order to yield a linear measurement unit. This polyline was then converted back to a raster of cell size 10 m. The 10 m cell size used here was chosen to help eliminate some of the artifact lines that could be created in the vector to raster conversion process. The Thin tool was then used to further process the flow network with a background of zero, round shape, and thin distance of 10 m (same as the most recent polyline to raster conversion). At this point, the flow network was processing all pieces of the flow network within the buffered boundary. The next steps (Make Feature Layer and Select Layer by Location) were used to select only the flow network pieces inside the actual watershed boundary. All the tolerance values were determined based on trial and error in order to get the cleanest possible outcome flow network.

This "clean" flow network was then sent to another submodel ("COMIDlink2a", Figure A.4) to assign the reach identification number associated with the closest reach in the NHD dataset to each feature in the derived flow network. Additionally, the "COMIDlink2a" model created a dataset of cross-sections to be used later in the modeling process. Each reach in the NHD dataset was assigned an identification number called a "COMID". Since the NHDplus reach attributes were used later in the modeling process, it was important to match the COMID reaches in the NHD dataset as closely as possible to the reaches in the developed flow network in this study's model algorithm.

The "COMIDlink2a" submodel (as a submodel to the Flowline model, Figure A.4) "linked" a COMID (reach identification number) from the NHD dataset to the nearest, most representative feature in the derived vector flow network. The first steps (Make Feature Layer, Select Layer By Location, Feature Class to Feature Class) included narrowing down the NHD dataset to only those reaches within the watershed boundary. Then, a copy of the COMID field was added (to COMID2 field) and calculated to have a numeric version of the COMID since the

original COMID was in a string format. This clipped and processed NHD dataset was then sent along with the previously derived "clean" derived flow network to a final submodel ("COMIDlink2asub").

The next model ("COMIDlink2asub", Figure A.4) actually processed and modified the flow network to achieve the most accurate flow network with NHD COMID as an attribute. The "COMIDlink2asub" GIS model was an automated process to assign a COMID; however, manual editing had to be done afterwards since the automated process was unable to recognize some of the intricate differences between the NHD network and the derived flow. The first step in "COMIDlink2asub" was to copy the "clean" derived flow network so further edits and modifications would not affect that version and it could be referenced if needed in the future. Next, the clipped NHD was converted to points. This essentially created a point for each NHD reach segment. The purpose of this was related to the selection of the best COMID for each derived flow network reach. Next, the flow network copy was buffered by 100 m and used in a spatial join with the previously developed NHD reach points. The spatial join match option of "CONTAINS" was used to join only the NHD points that were within the 100 m buffer to the buffered flow network reaches. At this point, the buffer dataset had polygon features for each reach, with spatially joined attributes related to the NHD points that fell within each respective reach's flow network buffer. Then, a second spatial join was performed in order to renames the "Join_Count" field that resulted from the first spatial join (number of NHD points within the 100 m buffer) to "NumNHDPtsNear." The buffer "FID" field and the newly developed "NumNHDPtsNear" field were both joined spatially from the buffer back to the derived flow network. Now, the flow network had an attribute, "NumNHDPtsNear," representing the number of relatively close NHD reaches. For those reaches with only one close NHD reach, the assignment of a COMID was relatively simple, and the nearest COMID could be utilized. For those with more than one close NHD reach, visual observations yielded that the closest NHD point did not always represent the COMID reach corresponding to the same derived flow network reach. For example, a flow network reach may have two close NHD reaches if the flow network is one feature spatially representing both NHD reaches. Essentially, the NHD reaches were subdivided smaller than the derived flow network reaches. Thus, the most effective automated process to match each NHD reach with a spatially equivalent derived flow network reach was to subdivide the derived flow network.

The next steps in the "COMIDlink2asub" GIS model (Figure A.4) were related to subdividing the derived flow network where there was more than one nearby NHD reach (Figure A.5). The Make Feature Layer and Select Layer By Attribute tools were used to isolate the derived flow network reaches with more than one nearby NHD reach in order to subdivide them. Next, these reaches were converted to points, using the Feature to Point tool, in a similar fashion to previously, when the NHD reaches were converted to points. The Split Line at Point tool was then used to essentially divide the features in half, approximately where the reach's point (from converting the derived flow network to points, previous step) was located. At this point, the derived flow network pieces that needed to be subdivided were split, and then they had to be merged back with the original "clean" flow network shapefile. In order to do this, the whole (un-split) lines in the flow network had to be removed, and then the newly split lines had to be merged in with the rest of the flow network that did not need to be subdivided. To accomplish this, the Symmetrical Difference tool was used to remove the whole lines (that were later split) from the "clean" flow network to make room for the newly split versions of those lines to be merged in. Then, the merge tool was used to merge: the split lines, and the original flow network minus the lines that needed to be split. Now that the flow network was subdivided where necessary, the nearest NHD reach's COMID still had to be joined to the derived flow network (as per the goal of this submodel). To do this, the new flow network was converted to points again, and a spatial join was performed to join the nearest (Match Option: CLOSEST) NHD reach's COMID to the spatially correspondent flow network reach point. Lastly, the nearest NHD reach's COMID (now an attribute on the flow network reach points shapefile) was joined to the polyline flow network. Thus, at the end of the COMIDlink2asub submodel, the flow network has been derived from elevation data and processed to most closely represent the NHD flow network.

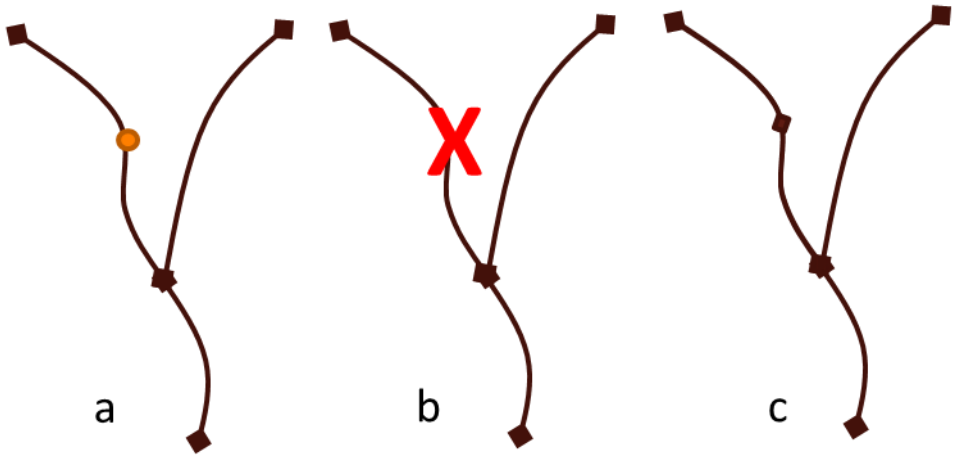

Figure A.5. The derived flow network reaches were identified where there were two nearby NHD reaches (different COMIDs). These reaches were (a) split by points and then (c) reinserted (b) where the non-split reach was located in the flow network shapefile.

Some manual editing was completed on the flow network shapefile immediately after the Surface Water model (See Figure A.3) was done, while other editing was done later in the methodology. After the full Surface Water model was completed, just the Flowline submodel (See Figure A.3) was repeated several times with slightly adjusted settings to determine which combination of smoothing and processing settings (as previously described) would output the most accurate representation of the actual flow network based up on the DEM. Some artifact lines (extremely short polylines surrounding what was once a raster cell) were still left in the output shapefile, which were eliminated based on their length. Any noticeably short line features (as determined by calculating the geometry of each feature) were examined and, if needed, manually edited or eliminated if they did not represent the actual flow network, but were remnants of poor raster to vector conversion. Also, the linking of the most representative COMID from the NHD flow network was not an ideal automated process. Some of the COMID associations on the derived flow network were found to be incorrect during the visual inspection process, and were thusly corrected based on the analogous NHD reach's COMID. Also, during the visual inspection process, it was found that the flow network lines were very choppy and had sharp corners, unlike common natural channel geomorphology. Because of this, the Smooth Line tool was used again (first used inside the Flowline submodel to the "Surface Water 2" GIS model, Figure A.4) with the PAEK smoothing algorithm, a 75 m smoothing tolerance, and the

"Preserve endpoint for closed lines" was left unchecked. The output from this tool created a more realistic flow network that had less sharp corners that would not actually exist in the natural environment. More manual editing to this flow network was completed as errors in the flow network were found while the next steps of the model were undertaken. Overall, the derived flow network matched fairly closely to the NHD flow network, with the exception of delineating some of the first order headwater streams. Because of the small size of these locations, some error and differences between the derived flow network and the NHD network were deemed acceptable.

After the flow network was developed with a properly assigned COMID, this flow network was used as a base river network file to create the cross-sectional lines used later in the overall model (End of the "COMIDlink2a" submodel to the Flowline model, Figure A.4). The last portion of the "COMIDlink2a" model was an ArcGIS script tool (Appendix C) ("GME Transect Tool") created to run the Geospatial Modelling Environment (GME) (Beyer, 2012) tool, "sampleperppointsalonglines" (Sample Perpendicular Points Along Lines). GME is a GIS software package that includes tools that ESRI's ArcGIS does not contain. The specific tool used in this study ("sampleperppointsalonglines") creates pairs of points perpendicular to a line segment at specified distance intervals along the flowline and specified widths (distances from the flowline). The flow network was fed into the GME Transect tool with the transect width set to 200 m (for a total cross-sectional width of 400 m) and the transect interval distance (along the flowline) was set to 20 m. The output from the tool, the sampled points perpendicular to the flow network, was then converted to lines with the ArcGIS Points to Line tool, thus creating a network of cross-section lines perpendicular to the flow network.

The last submodel in the surface water model was the filtering submodel (Figure A.4). The filtering submodel was built to utilize the terrain features from the DEM and develop the polygon representing the water surface of the flow network. The filtering model contained five different filtering procedures yielding five outputs. The goal was that after filtering, a combination of these outputs would be able to yield a decent representation of the water surface. Multiple filters were used in order to optimize the surface water polygon development. One filter passed the DEM through the ArcGIS curvature tool. The other four filters were ArcGIS script tools (Appendix C) that utilized modules built into the open source SAGA GIS program (SAGA, 2007). Each of the filtering script tools passed the DEM output from the DEM

Processing submodel (clipped, integer version of the DEM) through a single or combination of filters. The four filter/filter combinations were: 1) Laplace and Majority filters, 2) High Pass with Curvature filters, 3) Catchment Area Mass Flux module, and 4) DTM Filter (slope-based). Within each of the script tools, in addition to the filter commands, the DEM grid (.tif) had to be imported into the SAGA grid format (.sgrd), and then after the filtering process was finished, the grid had to be exported back to a (.tif) format to be utilized in ArcGIS again. Previous work showed that this filtering technique was appropriate for determining the location of the channel water surface (Daniel Chuquín Broad, personal communication).

The next major step in the spatial GIS component of the spatiotemporal floodplain connectivity model was the surface water (manual) processing steps (Figure A.3). The first main step to this was processing the DEM filtering outputs. The five filtered DEM outputs from the "Surface Water 2" GIS model (Figure A.4) were examined and observed to detect any patterns that could isolate the water surface of the active channel. After a brief observational analysis, the "Filter PostProc2" (Filter Post Processing Version 2) GIS model was built and used to condense the filtering outputs into the best possible raster representation of the water's surface. This processing model selected certain ideal values from each raster that best represented the water surface and then combined them to maximize the accuracy of the output surface water. Each filtered DEM was converted to a binary grid with 1's representing the filtered output's best representation of surface water, and 0's being everything else. First, the maximum and minimum values along the flow network of the SAGA Catchment Area Mass Flux filtered output were found. Then, any values in the full raster that were outside of this range were kept as representing surface water (Raster values >min & <Max were excluded, output=0). This same process was used to process the DTM filtered output, but keeping any values between the maximum and minimum values found along the flow network. For the High Pass-Curvature combination filtering output, surface water values were kept based on filtered DEM values within ¼ of a standard deviation from the mean of the filtered raster. Thus, surface water was best represented by near-mean values of the filtered High Pass-Curvature filtered DEM. The Laplace-Majority filter DEM combination utilized a specific set of values to represent the active channel surface water that were found via trial and error with the symbology classifications. It was found that a range between -2 and 2 best represented the surface water, and were therefore kept (value=1) as surface water. Finally, for the ArcGIS Curvature filtered DEM, values were

kept as surface water that were within ½ standard deviation from the mean value of the filtered DEM (similar to the High Pass-Curvature filter combination processing step). These five binary grids were then summed to find areas of high value which best represented the water surface. Based on a qualitative, observational inspection, values in the summed grid that were less than three (less than three filtered and processed grids had that cell as representative of surface water) were eliminated and assumed to be unrepresentative of the true surface water. Thus, only where three or more filtered and processed DEMs included the cell as surface water were actually kept as surface water cells.

It was observed that none of the filtered and now processed DEMs accurately represented the active channel water surface of extremely small reaches (first order streams). This is due to data resolution limitations and the ability of topography to accurately represent channels that have a width less than the DEM resolution (~3 m). In these cases, it was assumed that the raster version of the derived flow network was an acceptable approximation of what the water surface would be in reality. Thus, in addition to the five now binary filtered DEMs, the derived flow network grid was added as another binary grid to ensure that the small reaches still had an active channel water surface defined.

At this point, the present grid had cell values that best represented where the active channel water surface was located. Due to the DEM filtering techniques, there were still many locations were small artifacts of area were included as the surface water, but were disconnected from where the main channel actually was. To eliminate these artifact areas, Raster Calculator was used to remove any groups (as found via the Region Group and Lookup tools, in series) of "surface water" cells not along the flow network and were smaller than five cells (four cells and smaller) were eliminated. These areas generally represented highlighted by the filtering process, but were not along the main flow network and were not part of the active channel water surface area. This final grid at the end of the "FilterPostProc2" model still needed to be processed further to get most accurate results of where the active channel water surface was in reality.

Despite efforts to automate the process of delineating the active channel water surface, manual edits were still necessary. To begin, the raster version of the water surface, as previously developed from the "Filter PostProc2" GIS model, was converted to a polygon, with the option to simplify polygons checked. Next, because of relatively poor conversion from raster to polygon, the Aggregate Polygons tool was used with a 25 m tolerance to begin "cleaning up" the

surface water polygons. At this point, there were still many holes, gaps, and multipart features in the surface water polygon. To fix this, a combination of the Union tool (with no gaps allowed) and the Dissolve tool (not allowing multipart features) were used. The union tool filled in any gaps, which then created some concentric features (polygon features within other polygon features). The dissolve tool removed the features to create a "clean" surface water polygon. At this point, there were still many artifact polygons that needed to be deleted, so the Select by Location tool was used to select only the polygons that were intersecting the flow network. The Select tool was then used to export these polygons intersecting the flow network to a new shapefile. This final shapefile was then copied and renamed in order to preserve it for historical reference if needed.

This surface water polygon was then manually edited for more accuracy of its representation of the active channel in reality. This manual editing process consisted of comparing the developed surface water polygons to aerial imagery of the area. In many situations, the previously described delineation process included much low-lying floodplain in the surface water, which had to be manually adjusted. In some situations, (first order reaches), the aerial imagery was not sufficient to give a good estimation of what the water surface should be like, due to trees obstructing the view of the channel. In these situations, best engineering judgment was used to approximate the active channel water surface area, realizing that these channels would be very small with a very narrow active channel width along the flow network. During this extensive manual editing and examination process, the flow network was sometimes incorrectly delineated according to the aerial imagery. In these situations, both the flow network and the surface water polygon shapefiles were edited to mimic the NHD flow network and aerial imagery.

The purpose of developing the water's surface was to have an aerial representation of the channel and to have a starting point of where to begin adding stage heights to model the flooding processes. In reality, flood channels are not level, but have varying slopes going from upstream to downstream. To most easily represent this sloped channel in a modeling environment, the channel was essentially "sliced" at short (~20 m) intervals to create a stair-step pattern of the water surface going downstream. Instead of having a sloped surface to add flood stages to, there were many short, flat, water surface sections, which the flood stages could be added to. The concepts associated with this are described more fully in the flooding section of the

methodology. For this section, however, the water surface polygon, as described in the previous steps, needed to be sliced at small enough intervals to accurately depict a sloped surface with many short flat pieces at different elevations. To accomplish this, after the DEM filtering output was manually edited for maximum accuracy, the "SurfProc2", or Surface Water Processing (Version 2) model was used. This GIS ModelBuilder model was designed to "slice" the surface water into many sections that could be assigned a different base elevation later in the methodology. The beginning of the "SurfProc2" GIS model contained the Union and Dissolve tools to catch any remaining artifact features that may have been missed during the manual editing phase. Then, the cross-sectional lines, as developed in the COMIDlink2a submodel to the Flowline model (submodel to the "Surface Water 2" GIS model, Fig) were used to "cut" or "slice" the surface water polygon into approximately 20 m sections (the cross-sectional lines were developed with a 20 m distance between them along the flow network. To accomplish this, the cross-sectional lines were first clipped to the water surface. A 5 m buffer was then created around the cross-sectional lines and only those lines that were not close enough to another cross-sectional line were kept (Select Layer by Location tool, Relationship: Crossed by the Outline of); the rest were deleted. This step of eliminating some of the cross-sections was to improve the "slicing" process output in the next step. The next step utilized the "Feature to Polygon" tool to "slice" the surface water polygon with the spaced cross-sectional lines.

Because cross-sectional lines were removed, some of the water surface polygon features were not 20 m apart. Also, the water surface was not sliced in areas where there were large sections where no cross-sections were developed originally. To remedy this situation, and to still accurately represent a sloped surface, the "Cut Polygons Tool", as part of the Editing tools, was used to manually "slice" the larger water surface sections where observed to be necessary. Larger polygon features were "sliced" into lengths of approximately 20 m along the flow network. Also, despite the algorithm to remove cross-sectional lines that were close in proximity, some still remained that were intersecting, and thus created unnecessary slices and artifact polygons in the water surface that would cause problems later in the modeling process. Thus, the area of each polygon feature was calculated, and areas less than 30 m^2 were dissolved into neighboring polygons via the Eliminate tool. A last remaining visual inspection of the shapefile revealed a few more artifact polygons that had to be manually eliminated. At the end

of this process, the water surface was represented by a series of polygon segments, spaced approximately 20 m apart was ready for flood modeling.

After the "sliced" surface water was manually edited and prepared for flood simulations, the next step involved reexamining the flow network for COMID accuracy when compared to the representative/analogous NHD reach. Also, the flow network was examined for any further artifact remnant line features from the raster to vector processing procedures. The goal was for the derived flow network to match as closely as possible to the NHD network, but with the advantage of being derived from the DEM. In total, there were 149 NHDPlus reaches (defined individually by their identification number, or COMID). In theory, there would be a corresponding 149 derived flow network reaches. There were some NHD reaches, however, that did not have a corresponding flow network reach. In some situations, there were NHD reaches that were headwater streams that did not have a corresponding derived flow network reach, due to delineation practices. The nearest flow network reach was assigned the COMID of the NHD reach immediately downstream, thus the headwater COMID was not included, but the next reach downstream was included. This occurred for COMIDs: 4517908, 4517950, 4517998, and 4518256. The last NHD reach that was not included in the derived flow network was in an area where the NHD had delineated a braided section of stream (Figure A.6). Because of anticipated complications with the flooding algorithm, professional engineering judgment was used to leave the flow network as a single stream instead of the braided network, thus eliminating the COMID 4518206 from consideration.

Several flow network reaches were improperly assigned a COMID from the linking algorithm, and so they were manually reassigned a COMID based on visual inspection and comparison to the NHD flow network. After the flow network was edited, the next GIS model was used, the "COMIDlink2b" model.

Figure A.6. Braided section of NHD network (blue), along with the derived flow network with no braiding taken into consideration.

The main purpose of the "COMIDlink2b" GIS model was to delineate the contributing areas (non-cumulative) to each of the flow network reaches. Thus, the end product from this model was an allocation polygon shapefile with the contributing area to each reach. The goal was for the number of features in this dataset to match the number of COMIDs in the flow network, with one feature for each reach. To begin, the ArcGIS Cost Allocation tool was utilized with the flow network as the source feature data (COMID field as the source field, so the tool would delineate based on each flow network COMID). An inverse flow accumulation grid was used as the cost raster, as it would be similar and analogous in flow properties to the DEM. The output from this tool was then converted to a polygon with the "Simplify polygons" option checked. Since the flow accumulation grid was still in just a geographic coordinate system, the polygon allocation file was projected into the common projected coordinate system of the project (NAD_1983_UTM_Zone_18N). Next, a field was added and calculated to determine the area of each feature, in square meters. Because of the difficulties with raster to polygon conversion, as experienced on several occasions in the project, there were some artifact polygons that had to be dealt with. To do this, the Eliminate tool was used to dissolve artifact/small features less than a threshold of 150 m^2 with their nearest/surrounding polygon feature. The individual feature areas

were then recalculated (Calculate Field tool), and several NHDPlus flowline attributes were joined based on the COMID field.

The final allocation polygon shapefile, henceforth called "COMIDalloc" file, was then examined to ensure that the number of features in this file matched the number of flow network COMID reaches. An iterative process was then completed which included reexamining the flow network, correcting COMID assignment errors missed previously, and then re-running the COMIDlink2b model until both the flow network and the COMIDalloc file were correct, in terms of having the same COMIDs represented.

A.1.1.3 Outlet Delineation & Representative Cross-section Selection

After the flow network and COMIDalloc files were properly delineated, the next steps necessary prior to flood modeling were to delineate the outlet points for each reach (COMID nodes) and to identify representative cross-sections for each reach. The "LinesandNodes2" GIS model was developed to accomplish these two goals.

The theory to find the outlet for each reach was to find the location of the maximum flow accumulation in each COMIDalloc (contributing area polygon shapefile) area. Thus, the Zonal Statistics tool was used, masked by the project boundary, to find the maximum flow accumulation grid (after projecting into the projected coordinate system used for the project) value in each reach's contributing area. In theory, where the flow accumulation grid matched the zonal maximum grid, that would be where the reach output nodes would be placed. When performing a manual check of the output, however, a few of the reaches had wrongly identified the maximum flow accumulation value. In fact, the flow accumulation maximum value that the Zonal Statistics tool "found" was not a value that even existed on the original flow accumulation grid. Because of this, a work-around (alternative methodology) had to be developed to handle the incorrect maximum flow accumulation values found by the zonal statistics tool.

A.1.1.3.1 Alternative Methodology for Delineating Watershed Outlets

After the Zonal Statistics tool was used, the output grid had the maximum value for each zone as a grid, covering each zone with the maximum value. Next, the raster calculator was used

with additional conditional statements to find where the original flow accumulation grid was within two flow accumulation units of the maximum value, thus being the approximate outlet location. This output led to some COMIDalloc zones having a single cell (representative of the outlet), whereas others had several cells due to the allowance for cells within 2 flow accumulation units of the actual maximum value. These extra cells had to be eliminated to narrow it down to only the most representative output location. By the raster calculator expression, the output was the maximum value as found by the Zonal statistics tool. Thus, where there were multiple cells in one zone, they all had the same value. To seek the actual maximum flow accumulation value, raster calculator was used again to replace the maximum cell value of these individual (up to a few cells per zone) with the original flow accumulation value. These cells were then converted to a point shapefile. These points had the attribute of its flow accumulation value at that location.

Besides multiple cells containing the actual maximum value, other cells were present for some COMIDs which had flow accumulation values within two flow accumulation units of the maximum value. These points had to be eliminated based on which was closest to the "maximum" value. A python script within the calculate field tool was used to mark where there were duplicate flow accumulation values. The Spatial Join tool was next, which assigned the appropriate COMID to the points that were in the respective COMID zone. Next, the Summary Statistics tool was used to find the maximum flow accumulation value within each COMID. Thus, the "FaccVal" field was the statistics field, and the COMID was the Case field by which the statistics were found/summarized. This output table provided the actual maximum flow accumulation value based on the one or several flow accumulation points in each COMID zone. This table was joined back to the main points, or COMID outlets, shapefile based to match where the flow accumulation value in the points shapefile matched the maximum value found in the summary statistics tool. Thus, wherever there was a proper join (Flow Accumulation=Max Flow Accumulation), this signified the actual maximum flow accumulation point, and thus the outlet to the reach. After this join, the Select tool was used twice: once to remove those that were not equal to the maximum, and another to remove duplicates were there were two or more points within a COMID zone that had the same flow accumulation value (both or more were equal to the maximum value).

A.1.1.3.2 Selecting a Representative Cross-section

The second half of the "LinesandNodes2" GIS model (after defining the outlet points for each reach, as previously described) was designed to select a cross-section that was representative of the reach, and thus near the reach's outlet. The algorithm was designed to select the closest cross section within the corresponding COMID zone to the outlet location. The input transect/cross-section lines were from the output of the COMIDlink2a model, as previously described. The first step to this second half of the LinesandNodes2 model was a Spatial join to assign a COMID value to each of the cross-section lines that have their center in the COMID zone. Next, the Near tool was used with the cross-section lines (now with an associated COMID) as the input features and the COMID outlet nodes (as developed in the first half of this model) as the near features. This calculated the distance from each cross-section line to the nearest outlet location. Since it was ideal to have the COMID associated with the distance in the cross-sectional lines shapefile table, the Join Field tool was used to join the COMID associated with each "Near Distance" as calculated with the near tool. Thus, the COMID values were pulled from the outlets table based on the FID corresponding to the "NEAR_FID" in the cross-sectional lines table. Then, the select tool was used to narrow down the cross-sectional lines shapefile by selecting only those lines that the nearest outlet COMID was the same as the COMID zone that they were in. This fulfilled the algorithm requirement that the representative cross-sectional line was located within the COMID zone corresponding to the outlet to which it is nearest. Then, the Summary Statistics was used to generate a table with the minimum "Near Distance" summarized for each COMID. Thus, among all the cross-sectional lines located within a COMID zone that had their closest outlet being the node of the COMID zone they were located in, the minimum was found. This table was then joined back to the cross-sectional lines table to find where the "Near Distance" matched the minimum near distance, thus being the closest cross-sectional line to the outlet while remaining in the correct COMID zone. The select tool was then used to select only those lines where the near distance matched the minimum near distance. These output cross-sectional lines were then ready for manual editing.

Manual editing was required on the cross-sectional lines due to the nature of the algorithm, and that the most representative cross-sectional line may not actually be the one that is closest. In the development of the pilot model with a different watershed, it was noted that some cross-sections were falsely recording high inundation values. This was because their cross-

section was so close to a confluence, inundation area from the river to which the tributary was emptying was influencing the inundation area readings of the tributary cross-section. An illustration of this phenomenon is provided in Figure A.7. Thus, cross-sections located further up the tributary (pulled from an intermediate version of the cross-sectional lines file from within the LinesandNodes2 model run), away from potential influence from flooded area from the main channel were selected and appended to the main COMID-based cross-sectional lines file. Best engineering judgment was used to select these cross-sections so that the cross-section would be most representative of the reach. Some areas of the watershed not covered by the original cross-sectional lines file (from the COMIDlink2a model output) had to be drawn and placed manually.

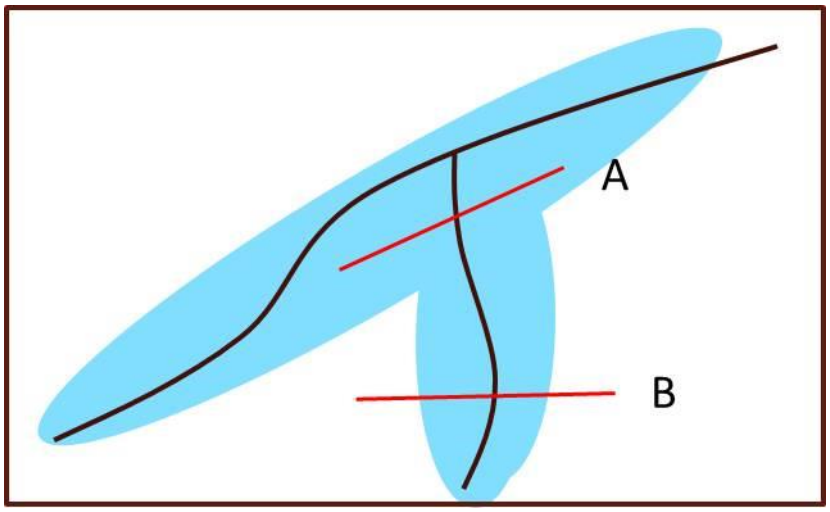

Figure A.7. Cross-section A represents the original automated selection of cross-sections, including a false representation of the reach's inundation area. Cross-section B represents the manually selected and more representative cross-section to be used in the flood modeling.

A.1.1.4 Flood Simulations

The artificial flood simulations provided the data foundational to calculating the desired spatiotemporal connectivity metric. The purpose of the flood simulations was to develop a relationship between flooded area and stage. Thus, the main function of this part of the methods was to simulate a certain water stage (flood), and then calculate both the inundation area and the cross-sectional inundation area. The inundation area was used later to analyze the level of

connectivity. The cross-sectional area was used in developing the stage-discharge relationships, based on the Manning's equation for flow.

Since the relationship flow (Q, m^3/s) and inundation area at each reach was needed to calculate connectivity, an intermediate relationship of artificial stage and flow was needed. The concept of the stage-discharge relationships ($Q = f(h)$) and stage-inundation area relationships ($A_i = f(h)$) are based on Manning's equation for velocity (V, m/s) (Equation A.1) where A_x represents cross-sectional area (m^2), R is the hydraulic radius (m), S is the channel slope (fraction), and n is the Manning's n coefficient (unitless).

Equation A.1. Manning's Equation

$$Q = A_x V = A_x R^{\frac{2}{3}} S^{\frac{1}{2}} \left(\frac{1}{n}\right)$$

The cross-sectional area, as in Equation A.1, is the area measured as part of the flood model. Another key assumption in this model was the use of an infinite width assumption in regards to the hydraulic radius (Daniel Chuquín Broad, personal communication). Generally, the hydraulic radius is measured by the cross-sectional area divided by the perimeter (Equation A.2). On a regional scale, channel geometry is impractical to directly measure. Simplification of the channel geometry is necessary in order to utilize Manning's equation for purposes of developing a stage-discharge relationship. Once floodplains become connected to the main channel, the majority of the flow is largely wide with a small depth (Figure A.8). Thus, the majority of the flow is above the main channel and in the floodplain. At this point, the channel was assumed as rectangular in shape with a large width to depth ratio.

Figure A.8. High width to depth ratio assumption for the floodplain allowing the hydraulic radius to be assumed equal to the depth of water in the floodplain.

The cross-sectional area would be the width multiplied by the stage, and the wetted perimeter would be the width summed with twice the stage (depth). Because of the assumption that the width is infinitesimally larger than the depth, the equation yields just the depth (stage) as equivalent to the hydraulic radius (Equation A.2).

Equation A.2. Infinite Width Assumption

$$R = \frac{A}{P} = \frac{wh}{w + 2h} \approx \frac{wh}{w} = h$$

Once the infinite width assumption is made to substitute the stage for the hydraulic radius, the new equation for flow becomes as follows:

Equation A.3. Updated Manning's Equation

$$Q = A_x V = A_x h^{\frac{2}{3}} S^{\frac{1}{2}} \left(\frac{1}{n}\right)$$

This equation can then be solved to make stage a function of flow, cross-sectional area, slope, Manning's n.

Equation A.4. Stage as a function of other Manning's variables

$$h^{\frac{2}{3}} = \frac{Qn}{A_x S^{\frac{1}{2}}}$$

One primary goal of the flood simulations, as aforementioned, is to develop a relationship between stage and cross-sectional area. Thus, cross-sectional area could be written as a function of stage ($A_x = f(h)$). Substituting this relationship into Equation A.4 yields the following:

Equation A.5. Manning's flow as a function of stage

$$h^{\frac{2}{3}} f(h) = \frac{Qn}{S^{\frac{1}{2}}}$$

This equation was then used to develop the actual stage-discharge relationship. Moving all terms to one side yields an equation equal to zero. At this point, a root finding technique was used to solve the equation for a stage given a flow value. The flow values plugged into this equation were the 14 points on the flow duration curve for each reach, as calculated in the regional regression methodology, discussed later (Section A.1.4). Thus, for each reach for each flow percentile, the equivalent stage was calculated.

In the flood model, as previously described, artificial stages were run through the flood model in order to yield an inundation area. From this, a stage-inundation area (A_i) relationship could be established. The stage previously calculated using the Manning's equation and root finding methodology could then be plugged into the stage-inundation area relationship to find an inundation area for the given stage. Thus, using both relationships (stage-cross sectional area and stage-inundation area) and the manipulated Manning's equation, an inundation area could be calculated for a given flow value for a reach. Later in the modeling process, after the ungauged hydrograph was developed using regression techniques, the inundation area was calculated and summed for each point of flow on the hydrograph above bankfull.

The processes involved in developing the polygons representing the water surface were the first main steps (Steps from beginning through "Data Processing" in Figure A.3) in the spatial GIS main first step in the overall model (Figure A.1). The next main step of the spatial GIS process was performing simulated floods on the DEM terrain. Encompassed within the flood simulations were several processes, including the development of a Manning's coefficient

for each reach, the actual flood simulations, and then various table combining and manipulations to prepare the data for usage in MATLAB.

A.1.1.4.1 Manning's Coefficient Model

Next, the Manning's roughness coefficient was developed for the area designated as the floodplain for each cross-section. As previously mentioned, there was a representative cross-section for each reach segment, and the average Manning's was found in the area represented by each cross-sectional line. Since the Manning's flow equation was used to develop relationships only at the cross-sections, it was assumed that a cross-sectional representative Manning's roughness coefficient was sufficient rather than taking an average Manning's n value across the entire reach's floodplain. NLCD data were used as the base for this roughness coefficient assignment. The highest resolution data that could be attained was at a 30 m resolution. As described later, the cross-sectional lines were gridded at a resolution of 3 m. To maximize the ability of the model to "find" the NLCD values along the gridded cross-sectional lines, the coarse resolution NLCD data was resampled to a 3 m resolution. Using engineering judgment and literature values, a lookup table (Table A.2) was established to link the NLCD code values with appropriate Manning's coefficient values. The documentation and literature references for the development of the lookup table can be found in Appendix D.

As far as the specific GIS ModelBuilder model to prepare the cross-sectional Manning's n value, the model required the pre-processed (resampled) NLCD data, the NLCD to Manning's n lookup table, and the manually edited cross-sectional lines (one per reach). The first step in the model was converting temporarily to a GIS INFO table for use in the model. Next, the Join Field tool was used to assign a Manning's n value to each NLCD code. The Lookup tool was then used to create a Manning's n grid. Meanwhile in the model, the cross-sectional lines were converted to a grid at a 3 m resolution, as aforementioned. In order to use these cross-section line grids as the "zones" for zonal statistics, their cell values had to be in an integer format, thus the Raster Calculator was used to convert the existing cell values to an integer format. Then, the Zonal Statistics as Table tool was used with the now gridded cross-sectional lines as the zones and the Manning's n grid as the input raster data. To this table, a "MANNINGSN" field was added and populated with the calculated mean for each zone (each cross-section). The last steps in the model were managerial, and related to re-adding the COMID identification field and

removing unnecessary fields from the table. This table was used later when preparing the flood data for MATLAB analysis.

Table A.2. NLCD code to Manning's coefficient lookup table

NLCD Code	NLCD Description	Mannings n
11	Open water	0.025
12	Ice/snow	0.022
21	Developed, Open Space	0.028
22	Developed, Low Intensity	0.079
23	Developed, Medium Intensity	0.071
24	Developed, High Intensity	0.087
31	Barren Land	0.027
41	Deciduous Forest	0.213
42	Evergreen forest	0.207
43	Mixed Forest	0.230
51	Dwarf Scrub	0.130
52	Shrub/Scrub	0.130
71	Grassland/Herbaceous	0.144
72	ALASKA ONLY	0.000
73	ALASKA ONLY	0.000
74	ALASKA ONLY	0.000
81	Pasture/Hay	0.133
82	Cultivated Crops	0.038
90	Woody Wetland	0.109
95	Emergent Herbaceous Wetland	0.141

A.1.1.4.2 Artificial GIS Flood Simulations

To perform the flood simulations, a GIS "Flood" model was built to first perform some pre-processing of the data, and then to actually simulate the floods and tabulate/summarize some of the output results. To begin, the manually edited "sliced" surface water polygon was input to the model, with a generic integer ID field added to identify each analysis (~20 m) section. Next, the Zonal Statistics tool was used to find the maximum elevation in each of these small analysis

sections. This created a raster with each analysis section assigned an elevation. It was assumed that this single elevation was an adequate representation of the sloped surface, given the short segment length. It was assumed that the drop in elevation from the top of the analysis section (upstream) to the bottom (downstream) across a length of 20 m was insignificant. These "MaxElevation" groups of cells were then distributed across the contributing area to each analysis segment. This was accomplished by using the Cost Allocation tool with the "MaxElevation" grid as the input raster, and the inverse flow accumulation grid as the cost raster. Thus, any cells that drained into an individual 20 m section would have the value of the maximum river elevation for that segment. This essentially created a flat plane to which the artificial stages could be added. This maximum elevation allocation grid was then fed into the flood iteration models.

There were two flood iteration models used as submodels to the GIS "Flood" model. These flood iteration models essentially had the same functionality, but one was designed to accommodate an artificial stage of 0 m (no stage change), while the other model was used to simulate multiple artificial stages greater than 0 m. For the purposes of this study, stages in 0.5 m increments from 0.5 m to 8 m was found to be sufficient to develop the stage vs. inundation area and stage vs. cross-sectional area relationships analyzed later in the modeling process. Other inputs (besides the maximum elevation allocation grid) to the flood iterations submodel included the manually edited cross-sectional lines for each reach, the COMID allocation (contributing area to each reach) polygon shapefile, and the integer version (also scaled by 100x) DEM.

The primary functionality of the "Flood Iterations" model was in the GIS ModelBuilder iterator "Iterate Multivalue". The artificial stages (0 to 8 m on 0.5 m increments) were input to this Flood Iterations submodel as a Multiple Value data type, which drove this submodel. The Flood Iterations submodel would run completely for each stage. The first step in the submodel was the actual simulation of a flood depth with Raster Calculator. Essentially, the raster calculator executed a conditional statement that if the elevation of any given cell was below the flood depth, then the value of the artificial stage was placed in that cell, indicating it was inundated. If the terrain had a greater elevation than the flood depth, the cell got a value of 0, indicating it was not inundated. The "flood depth" was determined by adding the value from the maximum elevation allocation grid (flat plane used as base for flooding, as previously described)

and the current artificial stage value. Thus, the flood depth was the maximum elevation of the river for that analysis section plus the artificial stage. The artificial stages were scaled by 100x to match the DEM scaling. From this, the output grid had the artificial stage value wherever the terrain would be inundated, given the model algorithm and assumptions. Figure A.9 illustrates the artificial flood simulation concept.

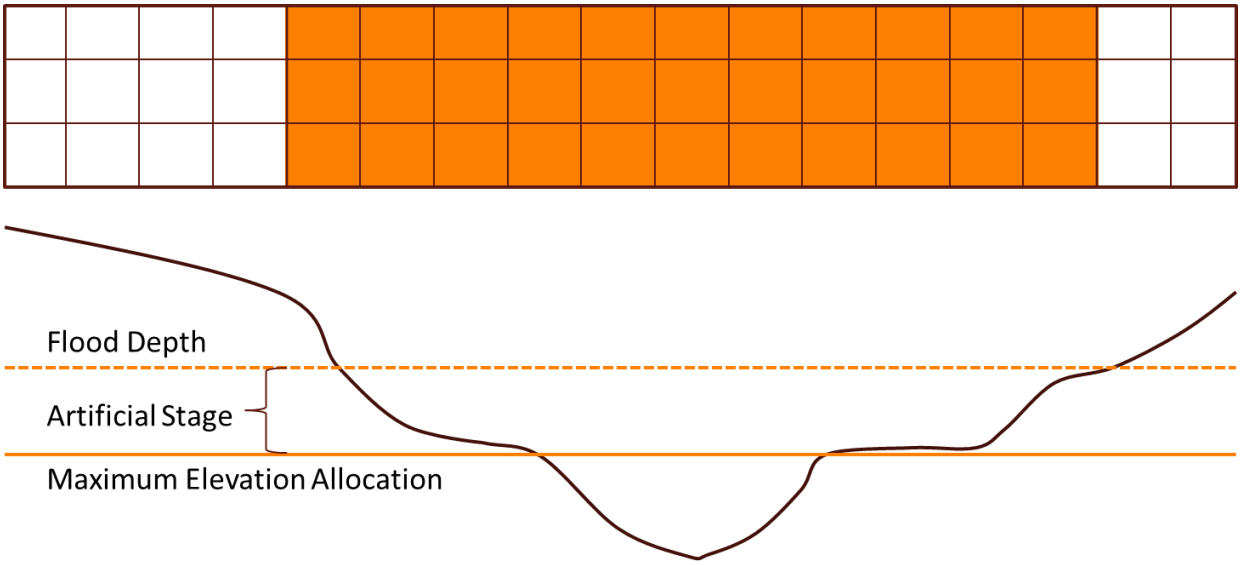

Figure A.9. Illustration diagram of the artificial flood simulation process. The bottom half represents a cross-sectional view of the DEM, while the top half shows what the inundation area output grid would look like for this small section.

As can be seen in Figure A.9, the artificial stage is added to the maximum elevation allocation grid to get a flood depth. Then, cells are designated as inundated if the elevation is below the flood depth, as seen in the top view of the inundation grid representation.

After the Raster Calculator flood simulation, the output inundation area grid was converted to an integer, and then the zero values were all converted to NoData. The Region Group tool was then used to group the cells that were actually connected to the main channel and separate the areas that were simulated as inundated, but disconnected from the main channel (i.e. depressions or low areas in the surrounding landscape). Using a combination of the Lookup, Get Raster Properties, and Raster Calculator tools, the disconnected areas were removed to end with

a grid containing only the connected inundated area. This inundation area grid was then sent to another submodel, the "Flood Summary" model to begin calculating the numeric inundation area statistics.

Besides the inundation area grid, the other primary input to the Flood Summary submodel was the COMID allocation (contributing area) polygon. The primary function of this submodel was in the Zonal Statistics as Table tool. This tool calculated the inundated area within each COMID allocation zone, thus being the inundated area for each reach. Before the statistics were calculated, however, the input flooded area raster was clipped to the project boundary (full watershed boundary) using the Extract by mask tool. This "area" was not actually the spatial area, however, since it was merely summarizing the cell values, which were equivalent to the current artificial stage. The actual area was calculated later in the Flood Summary submodel. Next, an output table was created by copying the COMID allocation polygon table. The semi-processed zonal statistics table was joined to this first output table. The food stage was added as a field, and calculated by retrieving the "MAX" value from the zonal statistics calculations. This maximum value was equivalent to the stage since the only values in the inundation raster were zeroes and the stage value. After re-adding a field for the identifier (COMID), the Calculate field tool was used to finally calculate the actual inundation area value. The inundation area was calculated by taking the "SUM" from the zonal statistics and dividing by the flood stage (zonal statistics "MAX" value). This yielded the number of inundated cells, which was then multiplied by the cell size squared to get the inundation area in square meters (Equation A.6).

Equation A.6. Inundation area calculation algorithm.

$$A = \frac{SUM}{MAX} * s^2$$

Where:

A= Inundation area (m^2);

SUM= Sum as found by the Zonal Statistics tool (m);

MAX= Maximum value as found by the Zonal statistics tool (m); and,

s= Cell size (m).

The output inundation area table of values was then sent back out to the Flood Iterations model. Once back in the Flood Iterations model, the table was then converted to a DBF and copied for combination with the cross-sectional area tabulated results.

A similar methodology was used within the Flood Iterations model to calculate the cross-sectional inundation area. Before the area could be calculated, the first step was to convert the cross-section polylines to a grid, in a similar fashion to the processes used when calculating the Manning's n coefficient. Next, Raster Calculator was used to create the actual simulated flood. An equivalent syntax was used in the tool as was used for the full inundation area flood simulation. The output form this tool, however, was the depth of flooding located only where there were cross-sectional lines. The previously developed grid of cross-sectional lines was then converted to an integer format and then used as the "zones" in the Zonal Statistics as Table tool. The purpose of this was to being summarizing the flooded area located at each cross-section, thus the need for the cross-sections as the zones. The polylines could not be used since lines do not necessarily have a spatial area. This created the need to convert the cross-sectional lines into grid format, thus introducing a spatial area. The "flooded depth at the cross-sections" raster (output from the Raster Calculator simulating the flood) was used as the input value raster to the Zonal Statistics as Table tool in order to summarize the flood characteristics by reach segment.

After joining fields from the integer-version of the gridded cross-section lines, the cross-sectional area was calculated by multiplying the "SUM" statistic by the cell width. The "SUM" statistic was output from the Zonal Statistics tool and essentially represented to total depth of water across the cross-section. Multiplying then by the cell width would yield a cross-sectional area of the water column that could be used later in the Manning's Equation for flow. This value had to be divided by 100 to accommodate for the 100x scaling of the DEM that took place earlier in the model. This tabular data (including the cross-sectional flooded area) was then copied to a DBF table, which was then joined to the inundation area results table, as previously discussed.

At this point, the cross-sectional area as well as inundation area had been calculated and joined together in a "combo" table for each artificial stage. These tables were then combined and manipulated for later modeling purposes. As previously mentioned, an analogous submodel to the "Flood Iterations" model was the "Flood Zero" model, designed to calculate the "base" inundation area. One key difference in this model was in regards to the way the disconnected inundation areas were removed. Due to small errors in the DEM and flood simulations, some

main sections of inundation area were disconnected from the rest of the inundation area. This created sections of flooded area that were correct (not truly disconnected), but were seen as disconnected in terms of the raster itself. To accommodate for this, the Zonal Statistics tool was used to essentially keep any areas that intersected the surface water polygon. This removed the truly disconnected areas and maintained the connected inundation areas that may have been disconnected by only a single raster cell.

A.1.1.4.3 Table Manipulation & MATLAB Preparation

After the flooding was completed, it was necessary to combine the tables created by the flood model into a single table that could be analyzed in MATLAB. The flood combination tables output by the Flood model consisted primarily of a row for each reach (COMID) and included the measured and summarized inundation area as well as cross-sectional area for that particular stage value. These tables were then combined with the "Table Summary" GIS model.

The "Table Summary" GIS model processed the flood combination tables, and then sent an output table for further manipulation out to a MATLAB script. First, the flood combination tables were combined via the Merge tool. Next, several NHDPlus attributes were joined based on the COMID field, both from the "flowlineattributesflow.dbf" file, as well as from the flowline attributes. The COMID contributing area (non-cumulative) attribute was then joined from the COMID allocation shapefile (polygon shapefile of contributing area delineations for each reach). Then, the Table to Table tool is used to convert the table from an INFO table to a DBF file. Lastly, a script tool was used (via the "tblformatrun_gis.py" script) to run a MATLAB script, per the general methodology outlined in Appendix C.

In the "tblformatrun_gis.py" python script, the first step included in the script converted the GIS-based DBF file to a CSV file that was more compatible with MATLAB. The rest of the python script ran the batch file, which ran the MATLAB script designed primarily to maintain the same number of rows as the number of stream reaches, while adding columns for each additional artificial flood stage. In the existing table, there were rows for each for each flood stage. This script converted the table to include one row per reach, and columns for each artificial flood stage.

The first main step in the "tblformatrun.m" MATLAB script was to load in the "FloodSummary.csv" file, which was output from earlier in the Table Summary GIS model. Throughout the spatiotemporal modeling process, a common range of stages (0-8 m on 0.5 m increments) was used, but maintaining a dynamic nature (accommodating other possible ranges of stages) was important. Thus, the next part of the table formatting script was finding the specific stages used based on the input data, number of rows, etc. Next, the common attributes (drainage area, length, etc.) were added to the base output table. A "FOR" loop in combination with matrix indexing was then used to shift the rows of flood data to the appropriate columns. Finally, field names were established and the output CSV file was formatted and then created. The final output table was the "RUN_flooddata.csv" and served as input to the final step in the spatial GIS model: the data preparation & manipulation ("Q_S2") GIS model.

A final GIS model ("Q_S2") was developed in order to prepare and combine the necessary data tables into a single CSV file that would be used in later MATLAB modeling procedures. Most of the specific tools and procedures within this model were data manipulation and conversion processes that may have had only changes to file type, etc. First, the Manning's table (cross-section Manning's n for each reach, output from the Manning's model) was copied to prevent alterations to the original file. Next, the input flood data CSV file (output from the Table Summary model) was converted to an ArcGIS INFO table. These two input tables were then joined together. Another join was performed in order to add the stream order attribute to the output table. This table had irrelevant fields removed and was converted to a DBF table which served as the base for the final output from the "Q_S2" GIS model. Meanwhile, the input flood data table was copied to avoid direct changes to the original file, and then excess fields were removed so only the flood related (cross-sectional area and inundation area) fields remained, in addition to the COMID. These flood data fields were then joined back to the output data table. This final "qsdata.dbf" table then had any remaining unnecessary fields deleted and was then converted to a CSV.

A.1.2 Temporal GIS Process (Step 2)

The entire purpose of the temporal GIS process relates back to the challenge of receiving a flow record at an ungauged location. Flow at an ungauged location is needed for the desired

annual connectivity metric. In order to determine the time above bankfull for each unknown outlet location, a long-term hydrograph was estimated at each reach. Developing hydrographs at ungauged locations is a complex area of hydrology that has been examined in many studies (Hisdal and Tveito, 1992; Mohamoud, 2008; Hughes and Smakhtin, 1996; Ssegane et al., 2013). Many studies have been completed examining peak flows at ungauged locations, but due to the nature of the problem, developing an entire ungauged hydrograph is extremely challenging. Fewer studies have been completed that estimated historic daily discharge at an ungauged location, as was desired in this particular study. The particular method utilized by this study to develop the ungauged hydrograph was based heavily on the methods used by Hughes and Smakhtin (1996) and Ssegane et al. (2013). Originally proposed by Fennessey (1994), the "QPPQ" (Flow-Pecentile-Percentile-Flow) method utilizes FDCs and a known hydrograph to develop a hydrograph at an unknown location. The general concept involves utilizing the flow time series from a nearby gauge, the flow duration curve (FDC) of the gauge, and a flow duration curve for the location of unknown flow (Q_g, FDC_g, and FDC_u) (Figure A.10).

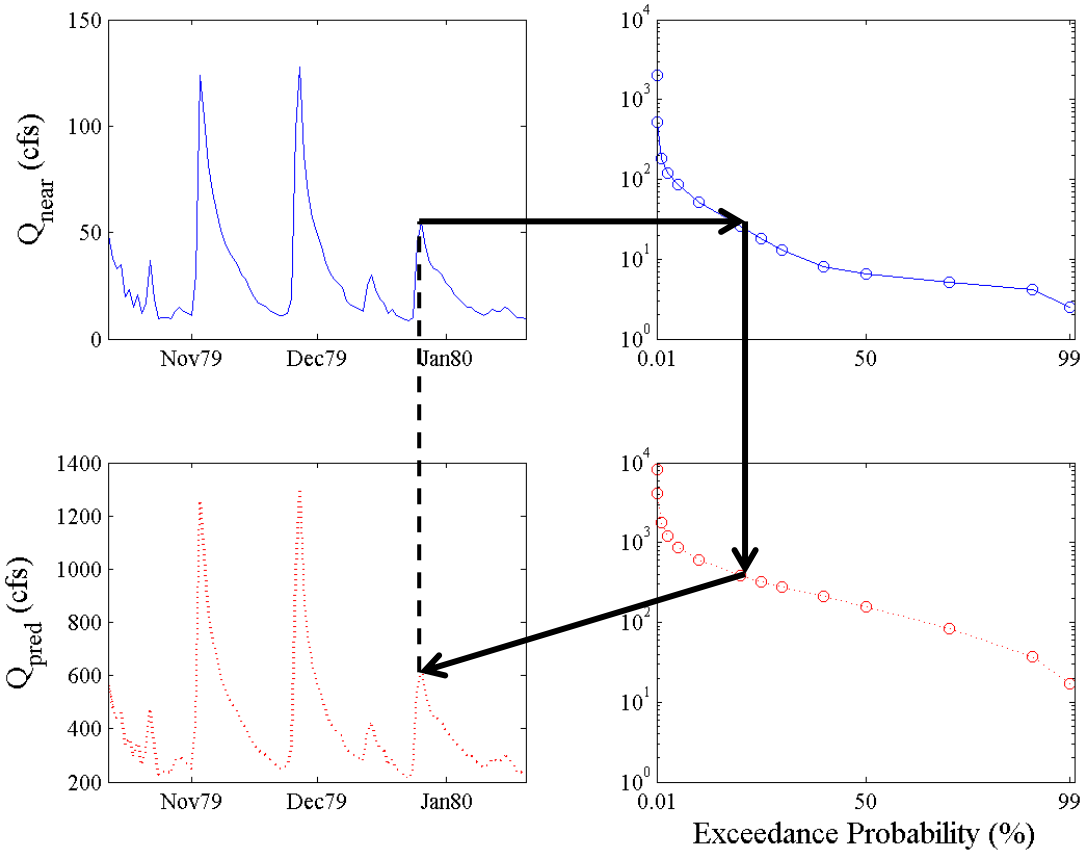

Figure A.10. QPPQ Theory (Concept adapted from Hughes & Smakhtin, 1996): The exceedance probability (top right) was found for any given flow on the source hydrograph (top left). Then, the same probability was found on the ungauged FDC (bottom right) and translated at the same time step to the ungauged hydrograph (bottom left). The two vertical lines indicate 1) the same daily value was estimated (left dotted line) and 2) the same exceedance percentile was assumed, per the QPPQ methodology.

To begin the process of developing an ungauged hydrograph, first, the gauge nearest to the ungauged location must be selected. This timing of flows from this gage will be used for the flows at the ungauged location. The assumption is that the gauge is near enough to the ungauged location to be influenced equally by the same storm events. For this study, the entire flow record was taken, and a flow duration curve was established with 14 points (99, 90, 70, 50, 40, 30, 25, 20, 10, 5, 2.5, 1, 0.1 and 0.01 % flow exceedance). Stationarity was assumed to justify using the entire flow record. This assumption was made in order to retrieve enough flow record to

accurately characterize the actual runoff properties of the watershed. For every point on the long-term hydrograph of the nearest gauge, the percent exceedance is found on the flow duration curve. It is then assumed that, due to the proximity of the gauge to the unknown location, that the percent exceedance at the gauge is the same as the flow percent exceedance at the ungauged location. From this, the flow is estimated from a flow duration curve at the ungauged location, and applied with the same timing to begin establishing a long-term hydrograph at the ungauged location. By definition, the ungauged hydrograph will have the same length of record as that of the nearest gauge. Besides the significant assumptions, the difficult part of this method is developing the ungauged flow duration curve (FDC).

Regression methodologies similar to those outlined by Ssegane et al. (2013) were used to develop the ungauged FDC. The goal of the regression was to be able to predict 14 different points on the flow duration curve (99, 90, 70, 50, 40, 30, 25, 20, 10, 5, 2.5, 1, 0.1 and 0.01 % flow exceedance). Thus, 14 different regression equations were developed in order to predict these 14 points on the flow duration curve. Since flow is known at the USGS gauges (dependent variables), the watershed characteristics of the USGS gages (independent variables) were used to develop the regression equations. The actual development of the regression equations was done in MATLAB, and thus is discussed later in the fourth part of the model: the temporal MATLAB portion. To develop the regression equations, factors and watershed attributes had to be developed that would be able to accurately predict the 14 different points on the flow duration curves. It was determined that a combination of watershed properties and characteristics would accurately predict the flow duration curve points due to the significant impact of the watershed on runoff characteristics. A GIS model was developed in order to collect watershed attributes given a point shapefile with the desired USGS gauges. Stuckey et al. (2012) also supported using GIS-based watershed characteristics for FDC prediction regression equations by stating that they "… [improve] the consistency, reproducibility, and ease-of-use of the resulting regression equations."

A.1.2.1 Gauge Selection

USGS gauges were selected from the surrounding area around the centroid of the Mahantango Creek watershed in east-central Pennsylvania. Gauges to be used in the regression

techniques were selected based on their proximity to the study watershed, physiographic region, and length of record. A 100 mi search radius was applied from the centroid of the Mahantago watershed. From these gauges, only those within the same physiographic region as the study watershed were selected. Finally, due to the regression on the flow data that would be completed, it was necessary to have a flow record of significant length. Various hydrologic regression studies have utilized gauges with varying minimum record lengths ranging from 20 years (Ssegane et al., 2013) to 5 years (Booker and Snelder, 2012). To encourage a comprehensive flow record, a minimum record length of 20 years was selected. Based on these criteria (same physiographic region, 100 mile radius, and 20 years of record), 72 USGS gauges were found. These gauges had to be split between those to use to build the regression equations and others to be used for model validation purposes. In the study conducted by Ssenege et al. (2013), they chose four gauges out of their possible 51 gauges to be used for validation, which is approximately 8% of the total available gauges. Thus, 8% of the 72 gauges found available in this study were used for validation, or six gauges. The outlet of the Mahantango Creek watershed was selected as a validation point since the degree to which this outlet is accurate will most likely correlate to the accuracy of the reaches' regressions within the watershed. Five other gauges were randomly selected to be used for validation.

A.1.2.2 Watershed Attribute Collection

The "GageAttrib" and "gage_iterator" GIS model and submodel, respectively, were created in order to develop a comma separated value (CSV) table with the USGS gauge site numbers along with their associated watershed characteristics/attributes. Analogous GIS models were built ("WSAttrib" and "ws_iterator") to the "GageAttrib" and "gage_iterator" models to calculate and summarize the watershed attributes for the ungauged reach locations.

The first main input to the model was a point shapefile of the gauges that were to be delineated and their watersheds summarized. There were a significant number of other inputs to the model due to the high number of watershed attributes that the model developed. Those other inputs included grids of flow direction, flow accumulation, hydrologic soil group (HSG), NLCD land use, elevation (DEM), slope, 2-year, 24-hour precipitation, and 100-year, 24-hour precipitation. The concept of the model was that all the input data was fed into the

"gage_iterator" model, which performed the analysis on each individual point (gauge) in the gauge shapefile. After all the gauges were analyzed, the data was summarized and put back together into a single output table.

The first step in the Gauge Attribute GIS model was to convert the site number field in the gauge shapefile to an integer, a necessary step for later processing within the GIS model. Next, the "gage_iterator" model was used to perform the summary analysis on a single gauge, looping through all the gauges in the input gauge shapefile. Once the individual gauge was singled out using the "Iterate Feature Selection" iterator within the ModelBuilder constructs, the Snap Pour Point tool was used to prevent poor watershed delineation due to discrepancies between the flow accumulation grid and the location of the gauge point. A snap distance of 180 m was used to ensure proper watershed delineation, which was equivalent to six times the input data resolution (~30 m). Next, the ArcGIS Watershed tool was used to delineate the watershed. Afterwards, it was then converted to a polygon and dissolved into a single feature (necessary due to artifacts from the raster to vector conversion). An ID field was created and then populated, and then the finalized watershed was then sent to each of several attribute summary steps, including attributes related to the DEM input, land use, soils, and precipitation, as described in detail below. Each of the following subsections details the purpose for attaining the attribute, the data source, as well as the specific ModelBuilder processes used to summarize the attribute.

A.1.2.2.1 DEM Attributes

Elevation-associated watershed characteristics were some of the many attributes examined to determine their effects in predicting flow percentiles. Other studies of regional regression to predict hydrologic characteristics have included DEM associated properties in their listing of variables. When first examining the possible watershed characteristics, Ssegane et al. (2013) examined the average, minimum, and maximum watershed elevation as independent variables in their regression. Booker and Snelder (2012) also incorporated DEM attributes by utilizing the average upstream watershed elevation in their regression study. Similarly, the slope grid (derivative from the elevation grid) was summarized for the gauge watersheds. Due to the nature of the data (integer version), the slope attributes represented were actually of the DEM*100 grid, and thus scaled by 100. Slope was incorporated into the list of attributes for its

obvious importance in flow characteristics, including its inclusion in the Manning's equation for flow.

A large coverage of elevation data was necessary for the regional regression to cover the extents of the USGS gauge watersheds. Because some of the gauges included in the regression were located on the Susquehanna River, this large watershed deemed an extensive DEM necessary. This large scale DEM was downloaded from the NHDPlus database (NHDPlus) and included the "a" and "b" elevation grid datasets for Region 2 (NHD Plus Version 1 Data). These datasets were brought into GIS, mosaicked into a single grid, projected into the study's common projected coordinate system (PCS), and then converted from integer to floating point. The units of the elevation grid were in centimeters. This was a result of the downloaded data being in an integer format while maintaining precision. The large extent flow direction and flow accumulation grids were also downloaded from the same site and processed (mosaicked and projected) in a similar fashion to the DEM data. The fully processed elevation data was used with the ArcGIS Slope tool to create a large-scale slope grid for use in the Gauge Attribute model. The percent slope option was used with a Z factor of 1.

Inside the gauge iterator model, the Zonal Statistics as Table tool was used on both the DEM and slope inputs, with the "zone" being the previously developed watershed polygon. All of the available statistics were calculated, and then the Collect Values tool was used to compile a full list of the output tables for later merging.

A.1.2.2.2 Land Use Attributes

Land use was another key dataset used and summarized to develop characteristics of the watershed that could be significant in determining flow percentiles on the FDC. As with the other attributes summarized, land use characteristics have been used in previous hydrologic regression studies (Ssegane et al., 2013; Mohamoud, 2008; Sarhadi et al., 2012). Different land uses will have significant impact on the runoff, and thus surface flow characteristics in the watershed. Because of this, we found it important to include a land use summary table as attributes to feed into the regression as independent variables.

The land use data was downloaded from the USDA Geospatial Data Gateway. The data was downloaded by state, pre-projected into the common PCS for the study in a 30 m resolution.

All the states downloaded that were in the UTM Zone 18N projection (NJ, PA, NY, MD) were mosaicked together in a single NLCD grid dataset that was input into the Gauge Attribute model.

Within the gauge iterator model, the Tabulate Area tool was used to compile the total area within each of the NLCD zones. This output only included headers of NLCD zones that had areas greater than zero. Since the tables for each watershed would be merged together later, it was critical to have a common table structure including all possible NLCD areas. Thus, after tabulating areas, this output table was then merged with a blank structure table, which included all possible NLCD codes as columns. Finally, Collect Values was used to compile a listing of the output tables to be merged later in the Gauge Attribute model.

A.1.2.2.3 Soils Attributes

Soils attributes were also considered when determining the watershed characteristics that could be related to flow within the watershed. Different soils have different infiltration and hydraulic conductivity characteristics, thus significantly affecting runoff. Because of this, we found it important to include the watershed average HSG as a possible independent variable for the future regression. Other studies have also included soils characteristics, with Ssegane et al. (2013) examining the hydrologic soil index specifically and Stuckey et al. (2012) examining several soils-based attributes.

The soils data was retrieved via the USDA Geospatial Data Gateway. The actual product retrieved was the STATSGO2 data (USDA/NRCS, 2006). This coarser soils data was utilized verses the finer resolution SSURGO data because it was assumed that the additional resolution and detail would have little affect given the size of the gauge watersheds and scale of the project.

The soils data had to have some processing completed before able to be used in the GageAttrib model, and thus the regression model development process. After the raw data was downloaded, the tabular soils data was imported to the database via the built-in macro inside the database. After this was completed, the soil layer shapefile from the downloaded data was brought into ArcMap. Then, the NRCS Soil Data Viewer (SDV) was used to create a layer of the soil delineations of hydrologic soil group (HSG). From this point, a "SoilsProcessor" GIS model was used to further process the data and ready it for utilization in the GageAttrib model.

The first step in the SoilsProcessor GIS model was to clip the data. This tool was used not to actually clip the data, but to serve as a way to make permanent the HSG data that was joined by the Soil Data Viewer (versus a temporary join that releases once other geoprocessing tools are used with that dataset). Next, the soils shapefile is projected in a projected coordinate system. At this point in the process, the HSG is still in a string format, using the common form of letters "A" through "D". A relatively simple field calculator python script was used to convert the HSG letters to numbers according to the table below (Table A.3). Lastly, the soils grid was converted to raster grid format with a cell size of 10 m to maintain adequate accuracy while managing a smaller file size than would be realized with a 3 m grid resolution.

Table A.3. Conversion methodology to go from HSG letter to numeric value.

HSG Letter	HSG Numeric Value
A	1
B	2
C	3
D	4
A/D	4
B/D	4
C/D	4

The concept of using numbers for HSG instead of the letters was vital for use in the quantitative modeling process. This process originated from the methods presented by Schwarz and Alexander (1995). Their algorithm for assigning a numerical value for the hydrologic soil group was adapted slightly for this study, but with the same primary relationships between character and numerical values (Table A.3).

As far as specific processes within the gauge iterator model, the Zonal Statistics as Table tool was used to summarize the HSG within the watersheds in the same fashion that the DEM and slope statistics were calculated previously. Instead of calculating all statistics on the HSG, only the mean HSG for the watershed was calculated. Finally, as with the rest of the attributes,

Collect Values was used to compile a listing of the output tables to be merged later in the Gauge Attribute model.

A.1.2.2.4 Precipitation Attributes

Precipitation also effects runoff characteristics, and thus was included in the GageAttrib model in order to collect precipitation characteristics of the watersheds. This data would be included in the future regression equations to help predict the various points on the flow duration curve. Other regression studies (Ssegane et al., 2013; Booker and Snelder, 2012; Mohamoud, 2008) also utilized precipitation characteristics in their hydrologic regression studies. Ssenege et al. (2013) found the 100 year, 24 hour precipitation depth to be one of the three most common and most significant watershed variables in their regression practices used to predict regionalized flow duration curves. Given this importance, and their initial use of the other precipitation exceedance depths as well, this study utilized the 100 year, 24 hour and 2 year, 24 hour precipitation depth characteristics of the watersheds.

The data utilized for precipitation were collected from the NOAA Hydrometeorological Design Studies Center (NOAA, 2006), specifically from the "Ohio River Basin and Surrounding States" sectional grid. This data, and therefore the actual regression variable came in units of inches*1000. This representation was most likely done to manage file size with integers, while retaining significant digits associated with precise precipitation depth estimates. According to the metadata associated with the precipitation grids, the precipitation estimates came from weather stations between 1893 to 2002, including only stations with 30 years of daily and/or 20 years of hourly data.

Within the "gage_iterator" GIS model, all the available statistics available with the Zonal Statistics as Table tool were collected for each gauge's watershed (iteratively) associated with the two precipitation grids. These output tables were consolidated with the Collect Values tool for further processing in the outer shell GIS model, "GageAttrib".

Once all of the individual attribute tables were created for each attribute for each gauge, the gauge iterator model was complete and the outer shell Gauge Attribute model continued. Thus, the output from the gauge iterator model included collections of tables (1 table per gauge) for each attribute (NLCD, HSG, DEM, slope, 2-yr 24-hr precipitation, and 100-yr, 24-hr

precipitation). A Python script tool was developed to merge all of the tables within each attribute collection (i.e. all the tables for NLCD statistics), thus creating a single output table containing the statistics for all the gauges that were analyzed. The Table to Table tool was then used to manage the fields in the consolidated tables. All the attribute tables (NLCD, DEM, HSG, etc.) were then joined together into a single table and output as a "RUN_attribtable.csv" file.

A.1.2.3 Gauge Analysis Preparation

Before completing the temporal GIS process (Step 2, Figure A.1, described in section A.1.2) and beginning the temporal MATLAB process (Step 3, Figure A.1, described in section A.1.4), significant data collection and gauge preparation had to be done. The "Gage3" GIS model was developed in order to perform this analysis and preparation. The main goal of the Gage3 model was to collect the online mean daily flow data for the gauges used in the regression as well as to determine the nearest gauge to the unknown outlet point. This "nearest gauge" would be the gauge used to determine the timing of flows when developing the ungauged hydrograph later in the modeling process (per the QPPQ methodology discussed in the overall temporal GIS process, Section A.1.2).

The first step in the "Gage3" GIS model was executing a Python script tool designed to retrieve the full record of daily mean flow data for each of the gauges to be used in the regression, saving the flow data in a text file. The data was retrieved from the available online database provided by the U.S. Geological Survey (2001). The actual Python script is available upon request from the author. One of the datasets that was downloaded via the Python script retrieval process had irregular syntax leading to errors in later processing, and so was modified appropriately. This one gauge's flow record contained notations for equipment malfunction (text) where the numeric flow value would have been. To avoid errors in the MATLAB data import process, the lines (daily records) with the text notation for equipment malfunction (~5 lines) were removed from the text file. After the flow data was collected for the regression gauges, the next step within the "Gage3" GIS model was to identify the nearest gauge to each reach which would later provide the timing of flows in the QPPQ method for forming the ungauged. This was done in the "NearestGage" submodel to the "Gage3" model.

The inputs to the "NearestGage" submodel are the ungauged outlet points, ungauged watershed attribute table (from the "WSAttrib" model), ungauged watershed centroids, regression gauge watershed centroids, and the full stream gauge shapefile. As aforementioned, the primary goal of the Nearest Gauge submodel is to identify and attribute the nearest gauge to each COMID reach. The nearest gauge was determined by the shortest distance on a watershed centroid-to-scentroid basis. Thus, the "nearest gauge" for any given reach would be the gauge whose watershed centroid is closest to the ungauged reach watershed's centroid.

The majority of specific GIS processes in the "NearestGage" model dealt with table manipulation and data management, rather than bulk actions/analysis. Several extra steps had to be added in order to accommodate running the model for an analysis run or a validation run (using reach nodes with a COMID identification number versus validation gauges with a SITENO identification number). The full validation methodology is described later.

On a tool-by-tool basis, the first step in the "NearestGage" model was to copy the ungauged outlet points to avoid altering the original file. Then, a generic "ID" field was added and populated with the original identification field (either the COMID or SOURCE_FEA field, depending on the ungauged point source). Meanwhile, the Generate Near Table tool was then used with the gauged and ungauged watershed centroids as inputs. Then, the Join Field tool was used to join the USGS site number of the nearest gauge to each ungauged FID. Then, Table to Table was used to clean up the fields and remove unnecessary fields before further processing. Join Field was then used again to join the ungauged ID to its respective FID, thus matching the ungauged ID with the nearest gauge site number. Join Field was then used again to match the generic ungauged ID field (previously developed) to the current ungauged ID in the table. Now, there is a single field with one name ("UKN_ID") containing the ID of the ungauged point. The near distance, nearest gauge's site number, as well as the UKN_ID was then all joined to the ungauged attribute table. This formed the basis for the final output table. The drainage area associated with the nearest gauge to each ungauged location was added to the table as an attribute for later analysis (from the raw stream gauge shapefile table). Lastly, the output table was converted to a CSV file and the daily mean flow data was collected (using the previously mentioned script tool) for the "nearest gauge" locations, in case any gauges were determined as "nearest" that were not included in the regression gauges.

A.1.3 Corroboration Methodology

Model corroboration was necessary to provide support for the methodology and assumptions made in the model for predicting hydrologic characteristics, flow, etc. at ungauged locations. Despite this, there was no feasible way to validate the inundation and connectivity results because of the site location and lack of data with which to validate. Validation was, however, possible for the FDC predictions from the developed regional regression equations. To validate the flow duration curve prediction, several gauges were selected and treated as "ungauged" locations. The attribute collection and regression procedures were then performed on these locations, assuming they had no flow record. The selection of the validation gauges was described previously, yielding six gauges to validate with, including the outlet of the Mahantango Creek watershed.

The regression-related GIS or MATLAB processes were applied to both the actual COMID reaches, as well as the validation gauges. This repetitive nature of procedures started with the attribute collection. Due to the construction of the models and their specificity to input field names, the "GageAttrib" model was used with the validation gauges as inputs (versus for the actual analytical run, the reach outlets were input to the "WSAttrib" GIS model). Next, the "Gage3" model was run, associating a "nearest gauge" to each of the "ungauged" locations (during the validation run of the Gage3 model, the "ungauged" locations were actually the USGS gauges selected for validation). The regression MATLAB script was performed on the validation gauges to develop a predicted FDC and hydrograph, just like would be created for the ungauged reach locations. Finally, the mean daily flow data collection Python script was used to collect the flow data for the validation gauges. Then, a validation MATLAB script was used to output validation statistics comparing both the FDC and full hydrograph of the predicted and actual data.

A.1.4 Temporal MATLAB Process (Step 3)

MATLAB was used for further data analysis and processing (regional regression, QPPQ methodology, area-stage/flow-stage relationships, connectivity metric calculations) following the GIS analysis for both the spatial and temporal components of the model. The main component of the temporal MATLAB process was a script designed to import tabular data from the temporal

GIS model and process it to develop the ungauged flow duration curves and ungauged hydrographs. The main steps for the temporal MATLAB process of the overall model were outlined in Figure A.11.

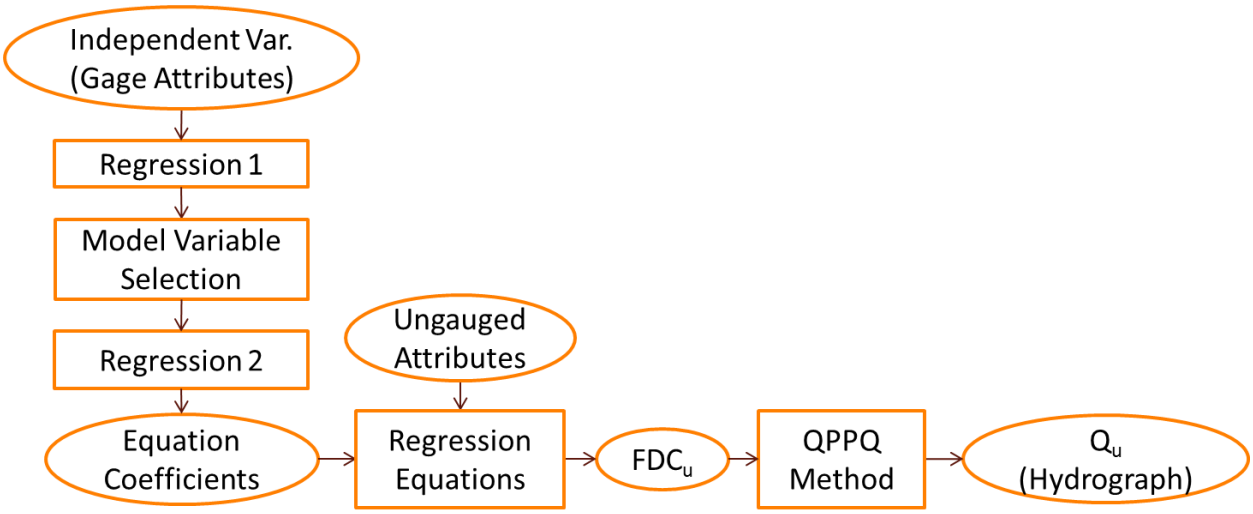

Figure A.11. General steps for the temporal MATLAB process (Step 3 of overall model)

Thus, a bulk of the script was developing the regional regression equations for 14 points on the flow duration curve based on the local gauge watershed characteristics and applying them to the study watershed's reaches and their respective watershed characteristics. The general form of the regression equations is shown in Equation A.7.

The first few sections of the regression analysis script (qregressrun) were designed to set up the file directories for files utilized throughout the script. The description of the specific processes and functions in the regression analysis script are provided below, primarily in sections A.1.4.1 to A.1.4.5). Also in the beginning of this script, the "defprobs" variable was set, which defines the exceedance probabilities at which the gauged and ungauged flow duration curves would be calculated. This variable could be set at the beginning of this one script, and any adjustments to the desired percent exceedance values would be reflected throughout the rest of the modeling processes.

Equation A.7. General regional regression equation

$$Q_{pp} = 10^{a} X_1^{b} X_2^{c} X_3^{d}$$

Where:

Q_{pp}= Flow at probability of exceedance, pp (ft^3/s);

a,b,c,d= Regression constants; and,

X_n= Independent watershed characteristic variables.

A.1.4.1 Regression Variable Development

The next part of the regression script imported the known USGS gauge site numbers that would be used during the rest of the modeling process. The site numbers were imported and processed into a proper MATLAB format from a "sitenumfile.csv" which was output as part of the "Gage3" model (Section A.1.2.3). After the site numbers were imported, a hydrologic frequency analysis was performed on the USGS gauges, finding the flows associated with the previously defined percentiles. A separate function, "gage_hfa" was written to perform the hydrologic frequency analysis given inputs of the site number, the file directory containing the flow data, and the desired flow percentiles ("defprobs" variable).

The "gage_hfa" function first read in the flow data that was previously downloaded online. Because the flow data was in a text file format, the first sections of the function were dedicated to accurately reading and processing the flow data. A message was displayed if the flow text file was empty (inaccurate download). If the flow text file is not empty, the function continues to process. The next part of the function was built to remove empty flow values and corresponding time stamps that may have been in the text file. For the actual hydrologic frequency analysis, the Weibull method was used. First, the flows were sorted in a descending order and assigned a rank. The probability of exceeding the given value was then calculated by Equation A.8.

Equation A.8. Weibull Probability of Exceedance

$$P = 100 * \frac{m}{n+1}$$

Where:

P= probability of equaling or exceeding the associated value;

m= rank from descending values; and,

n= number of data points.

After the probability of equaling or exceeding each flow value was calculated, the specific probabilities desired in the flow duration curve could then be calculated. For the purposes of this study, 14 points on the flow duration curve were calculated, but the specific "gage_hfa" function was designed to accommodate various desired probabilities. The actual probabilities for the flow duration curve were developed by performing a linear interpolation between the probabilities calculated via the previously mentioned Weibull method. These flow values at the desired percentiles are then sent out to the regression script as output from the "gage_hfa" function.

These flow duration curves provided the 14 different sets of dependent variables (1 flow percentile for all the regression gauges composed the dependent variable set for a single regression equation) used in the regression. To prevent errors in the regression, a value of 1 was added to each of the flow values prior to regression. This was done to avoid zero flow values at low flow percentiles, and thus errors when taking the logarithm of the flow value.

Before any regression could take place, the watershed characteristics had to be imported into the MATLAB ("Independent Variable (Gauge Attributes)", Figure A.11) workspace from the temporal GIS process (Step 2, Figure A.1). These variables included drainage area, mean hydrologic soil group (HSG), percent forest (calculated from NLCD codes 41, 42, and 43), percent developed area (calculated from NLCD codes 21-24), percent agriculture area (calculated from NLCD codes 81 and 82), minimum watershed elevation, range of elevation, mean slope, 2-yr, 24-hr precipitation depth maximum, 2-yr, 24-hr precipitation depth mean, 100-

yr, 24-hr precipitation depth maximum, and 100-yr, 24-hr precipitation depth mean. As with the flow values (dependent variables), a value of one was added to the NLCD percentage values to avoid values of zero and thus errors when taking the logarithm. The addition of a value of 1to the aforementioned attributes was consistent in both reason and application with the study conducted by Stuckey et al. (2012). These values were input from the CSV attribute table developed from the GIS "GageAttrib" model.

In addition to these 12 variables calculated directly from the temporal GIS model, four interaction terms were also calculated for their potential significance in the model regression. These four interaction terms were between the HSG value and each of the precipitation characteristics (100-yr max, 100-yr mean, 2-yr max, and 2-yr mean). Thus, in the end, 16 regression independent variables were used. The logarithm (base 10) was calculated for each of the independent variables prior to the multiple linear regression, similar to the methodology presented by Stuckey et al. (2012).

In order to improve the quality and applicability of the regression, some gauges were eliminated from the regression based on drainage area size. For this study, the regression gauges were going to be applied to the Mahantango Creek watershed and the reaches' watersheds within the full Mahantango Creek watershed. Thus, it was considered appropriate to ensure that the gauges used in the regression were within a reasonable tolerance of drainage area size as compared to the Mahantango Creek watershed area. It was suggested that excessively large or small drainage areas (in comparison to the study watershed) would skew the regression coefficients, and decrease model application efficiency. Thus, a drainage area "factor" was developed to quantify which gauges were within a certain tolerance of watershed size and which were not. A drainage area factor of 2.5 was set for this study, after examining the validation statistics for other possible factor values. This factor meant that any drainage areas above or below 2.5x the study watershed area would be eliminated from the regression. Using a factor of 2.5 resulted in the best prediction of the Mahantango Creek watershed's FDC. Since the regression equations were applied to those watersheds within the Mahantango Creek watershed, this watershed's validation statistics were of the highest priority and consideration. Once the factor value was assigned, it was used to develop a minimum and maximum watershed area threshold value based on the Mahantango Creek watershed area. Gauges that had watershed

areas outside the range developed by the drainage area factor were eliminated from consideration.

For documentation and later analysis purposes, the next section of the regression script was designed to output a CSV table with the "X" values (independent variable watershed attributes) used in the regression. After this was completed, the data was processed and ready for the first multiple linear regression steps.

The dependent variables for use in the regression were the flow values associated with a single flow percentile from the flow duration curves of all the regression gauges. Since there were 14 points on the flow duration curve, there were, in the end, 14 different regression equations and respective sets of coefficients. The regression was performed independently for each flow percentile, and the process was looped to include each of the 14 flow percentiles.

A.1.4.2 Multiple Linear Regression

The MATLAB "stepwisefit" tool was used to perform the initial (stepwise) regression ("Regression 1", Figure A.11). This tool was used in order to automate the process of determining the best model for the regression data. The entrance and exit p-values used as input parameters for this tool were 0.1 and 0.15, respectively. This first regression (another regression was performed later to finalize the coefficients for the final model) was used to get a general idea of which variables were significant for which flow percentiles and which were not. One output from this tool was a binary vector which had a "1" for variables that were significant and a "0" when the variable was not found to be significant. When these vectors were summed by column, one could determine which variables were significant most frequently among the 14 different regression equations. The goal was to find the three variables that were significant in the most number of the equations.

A.1.4.3 FDC Regression Model Selection

One important consideration to the final model selection was the number of variables that would be included in the final model. Ssegane et al. (2013) utilized a uniform three variables for all flow percentiles (same regression model variables used to predict all flow percentiles). The

current study study also utilized a three-variable model for the prediction of the FDC values. The three-variable model was selected for multiple reasons. First, as just mentioned, another similar study utilized only three variables for predicting the various flow percentiles. Also, the regression and validation scripts were tested with the model choosing the best two, three, and four variables to include, respectively. This was done in order to justify how many variables should be included in the final FDC prediction regression models. After running these different tests, the validation statistics (Nash-Sutcliffe Efficiency (NSE) of the full hydrograph and NSE of the 14 FDC points) were examined to determine which model was best. The validation statistics revealed only slight differences in prediction accuracy between the two, three, and four-variable models. When focusing on the Mahantango Creek outlet gauge (since this study's focus was on reaches within that watershed), the NSE of the FDC did not change when going from a three to a four-variable model. There was a slight improvement when going from a two to a three-variable model. The other validation gauges showed similar trends, but overall had no or only slight differences between the different model validation results. Thus, for the study watershed's gauge, the three-variable model had the best FDC prediction while maintaining model simplicity. A uniform set of three variables was selected across all flow percentiles, as opposed to choosing the most significant variables at each flow percentile. This was done to ensure proper scaling and a final FDC that was monotonically increasing. Trial runs of the model that were conducted without utilizing a uniform set of variables across all flow magnitudes resulted in FDCs that were not monotonically increasing. Archfield et al. (2010) discuss issues related to predicting FDCs when not using a uniform set of variables.

The section of the regression script directly after the "stepwisefit" regression was made to find the best variables representing the system ("Model Variable Selection", Figure A.11). As previously described, the binary vectors of model inclusion were added together forming a vector that had the number of equations each variable was found to be significant in (a "Variable Column Sum" vector variable). From this vector, the variables with the highest values (significant in the highest number of regression models) were selected, with the "best variable" columns noted. In some situations, however, there would be ties between variables in their frequency of significance. Whenever there were ties leading to greater than three variables that could be selected for the model, the ties between variables were broken based on regression p-values. The variables included in the tie were then chosen based on the sum of the regression p-

values across all 14 flow percentiles. The variable with the lowest p-value total was selected for the model, while the others were not included in the model. This p-value sum was done in order to get a general sense of overall model performance when considering the future application of the regression equations to predict full hydrographs. Since it was important that all the regression equations had adequate predictability, it was important that the variable in question be as significant as possible. Any regression equations where the variable was found to be highly insignificant (i.e. p=1) had a high impact on its rejection from the model since a value of 1 would lead to a high p-value sum. Although a relatively crude technique to break ties between model significance frequency values, it was assumed to be adequate for the purposes of this study.

A.1.4.4 Final FDC Regression Model Development

With the best three regression variables identified, another regression was performed on just these three variables ("Regression 2", Figure A.11). The original independent variable matrix was narrowed down to only the variables selected for the model. This narrowed independent variable matrix with the addition of a constant column (all values equal to one) was the input into the MATLAB regress tool, with the same dependent variable looping process as used in the first regression. The regress tool was used for this second regression since it did not eliminate variables based on their significance in the regression model. The "stepwisefit" tool output coefficients were based only on variables that were found to be significant and included in the model (i.e. if two out of three variables were significant, the two significant variable coefficients were calculated based on only using those two variables in the model). Insignificant variables still had coefficient values, but they were based on if that variable had been included in the model. Thus, the regress tool was best suited for this second regression step so all three regression variables would be included in each of the final regression models for the 14 flow duration percentiles. The primary output from this second regression was a matrix with the coefficients for the best three variables (watershed characteristics) and a constant term for all 14 flow percentiles ("Equation Coefficients", Figure A.11). This matrix of coefficient values was used later to develop the FDC for the ungauged locations.

A.1.4.5 Ungauged FDC Development & QPPQ Methodology

The next step in the script was to import the watershed characteristics of the ungauged locations ("Ungauged Attributes", Figure A.11). This was accomplished with analogous procedures to those used previously for importing the attributes of the known gauge watersheds. In order to apply the developed regional regression equations, the logarithm (base 10) of the variables were taken since the developed equation was of linear form. The same NLCD summarization attributes (percent forest cover, percent developed, etc.) and interaction terms were also calculated. As with the gauged watershed attributes, the ungauged watershed attributes were also exported to a CSV table for later analysis, documentation, and interpretation. This full unknown watersheds attribute table was then narrowed based on which variables were used in the second and final regression. These processed attributes (independent variables) were then multiplied by the previously derived regression coefficients on a reach-by-reach basis ("Regression Equations", Figure A.11). The intercept of the multiple linear regression equation was then added to the sum of the coefficients multiplied by the processed independent variables (ungauged watershed attributes). Finally, this summed equation value was placed as the exponent to 10 (inverse of the base 10 logarithm) and a value of 1 was subtracted from it (to accommodate for the +1 addition to the flow values input as the dependent variables in the regression steps, See Equation 4.1 in Results). Once all 14 sets of regression equation coefficients were applied to the reach watershed's processed attributes, the final set of ungauged flow duration curves (FDC_u) were completed ("FDC_u", Figure) and output to a CSV file.

The next major step in the regression script was to utilize the ungauged FDCs to build the full ungauged hydrographs using the QPPQ concepts described in Section A.1.2 ("QPPQ Method", Figure A.11). The development of the ungauged hydrograph was performed on one location at a time, built into a loop in the script. The first main step to the developing the actual ungauged hydrograph was establishing the hydrograph of the nearest gauge. This was important because, as previously mentioned, the time sequence of the flows was based on the time sequence at the nearest gauge, as determined by distance between watershed centroids. This method of finding the appropriate gauge for the ungauged location's flow time sequence might be useful for a large application area, but this study involved application of the method on reaches within a single watershed. Because of this, the "nearest gauge" for all the reaches within the Mahantango Creek watershed was manually assigned to be the watershed outlet gauge

(USGS Site Number: 01555500). This was assumed appropriate since all the flows at the reaches within the watershed would be represented best by the flow at the outlet of the watershed. Thus, the timing of flows for all the study reaches was based on the Mahantango Creek watershed outlet gauge. Once the "nearest gauge" was established, the full hydrograph was attained from the mean daily flow data. Also, the FDC of the "nearest gauge" was pulled from the table of gauge FDCs previously established. With all the data in place, the main portion of ungauged hydrograph development could be developed, per the methodology presented by Hughes and Smakhtin (1996) (Refer to Figure A.10).

For each daily mean flow value in the nearest gauge's flow record, the percent exceedance according to the nearest gauge's FDC was calculated via linear interpolation (linear FDC flow values, logarithm base 10 of the probability values for more accurate interpolation) between the 14 FDC points. Using this percent exceedance value, the flow at the ungauged location was then calculated by linearly interpolating (linear FDC_u flow values, logarithm base 10 of the probability values for more accurate interpolation) from the ungauged FDC. Thus, a flow value for the same time as the nearest gauge was calculated. When this process was performed for each flow value in the record, an entire ungauged hydrograph was developed ("Q_u (Hydrograph)", Figure A.11). A log-linear interpolation technique was used for the two previously mentioned interpolations. The logarithm (base 10) of the exceedance probabilities was taken prior to usage in the interpolation in order to develop a more linear relationship between the variables in question. With a more linear relationship, the linear interpolation tool should give more accurate results. The rest of the regression script included the processes necessary to output each ungauged hydrograph to a CSV for later use.

A.1.5 Spatial MATLAB Process (Step 4)

The spatial MATLAB model inputs the processed and summarized flood data from the spatial GIS model and outputs inundation area for each flow point on a 14 point flow duration curve for each reach. This output can be considered a "water resource index duration curve", as discussed by Vogel and Fennessey (1995). In this study, the water resource being related to percent exceedance was floodplain inundation area. The relationship between flow and

inundation area was used later in the final MATLAB components of the overall model. The entire spatial MATLAB model consisted of a single MATLAB flood relationships script.

The "areastage" script was used to develop the necessary relationships between cross-sectional flooded area, inundation area, stage, etc. The first steps involved reading the stage and flood area data from the output CSV of the "Q_S2" GIS model (stage-discharge spatial GIS model, described in Section A.1.1.4.3). Next, the number of artificial stages used in the flood model was automatically calculated based on the number of columns in the data input table. After the flood and stage data was properly imported, the ungauged FDCs were imported as previously developed in the regression script (temporal MATLAB process). A data quality check was performed on the ungauged FDCs to ensure that they were all strictly increasing from high percentile flows (i.e. 99%, 70%, etc.) to low percentiles (high flows). Any flow duration curves found to not be strictly increasing caused an error in the model, forcing a manual check of the ungauged FDCs and fix the regression output problem causing the poor FDC development.

The next major portion of the flood relationships script involved calculating the stage for a given flow value. The flow values in the FDCs had to be converted to metric units from the English units that were residual from the downloaded USGS flow data. Two nested loops were used to calculate the various relationships, with the outer loop iterating through the 14 different flow percentiles while the inner loop iterated through each stream reach (COMID). Each iteration in this nested loop process calculated the stage associated with each flow in the FDC. The flow value (Q) was retrieved from the FDC data, while the reach slope and Manning's n values were attained from the previously imported flood data. Cross-sectional areas for each artificial stage (from the flood model) were also imported for each reach.

The first relationship to be developed (prior to finding a stage for any given flow) was the relationship between stage and cross-sectional flooded area, with all the data coming from the flood model output. Because there was no model relationship (linear, exponential, etc.) that fit the relationship between cross-sectional flooded area and artificial stage well, the relationship was determined by interpolation. Since the goal for the relationship was to find a cross-sectional area given an artificial stage value, it was assumed that interpolation between the known points would be most accurate rather than fitting a modelled relationship to the data.

Once the cross-sectional flooded area versus stage relationship (via interpolation) was established, cross-sectional flooded area could still be represented in equation form as a function

of stage ($f(h)$). Equation was then set to zero (Equation A.9) and solved using numeric root finding methods to find a value of stage (h) given flow (Q), Manning's n, and the reach slope (S).

Equation A.9. Root finding equation for stage-flow relationship

$$0 = \frac{Qn}{S^{\frac{1}{2}}} - h^{\frac{2}{3}} f(h)$$

An upper and lower guess of cross-sectional area was developed form interpolating from guesses at the resulting stages being between 0 and 8 m, respectively. When plugging these bounding cross-sectional flooded areas into Equation A.9, the goal was to achieve values with opposing signs (positive/negative). This would indicate that the stage solution was somewhere in between 0 and 8 m. If the bounding results were indeed of opposite signs, the MATLAB "fzero" function was used to find the actual stage value corresponding to the flow for that reach for each FDC percentile iteration. If the bounding resulted in values both of the same sign, it was an indication that a root could not be found, and a value of zero was assigned for the stage. The end of these nested loops resulted in a table with the stages associated with each reach for each flow percentile, thus creating a Stage Duration Curve (SDC, analogous to a flow duration curve).

When examining the stage prediction intermediate results from the spatial MATLAB model, there were many 0 m stage predictions for the low flow percentiles. These zero stage values appeared to correlate with the negative flow predictions in the predicted FDC (FDC_u). Using the Manning's-based algorithm, no root was found when a negative flow value was used from the predicted FDC, thus yielding a 0 m stage prediction. These negative flows at low flow percentiles (from the predicted FDCs in the temporal MATLAB process, overall model Step 3, section A.1.4.5) were deemed allowable since the focus of this study was on flows greater than bankfull that would cause floodplain inundation.

Before the SDC could be utilized for the development of an area of inundation duration curve (AIDC), the relationship between stage and inundation area had to be established. Another

nested loop structure similar to the one previously mentioned was used to iterate through each flow percentile and through each reach (COMID). For each reach, the polynomial functional relationship between stage and inundation area was established. Various relationships were examined with the raw data and it was determined that the best relationship between stage and inundation area was a 4^{th} order polynomial. Thus, for each reach, a vector of polynomial coefficients was established to be able to calculate an inundation area for any given stage. Then, using this relationship, the stages in the SDC table (water stage at each flow percentile for each reach) were converted to an area of inundation duration curve (AIDC). At this point in the modeling process, a table of values had been developed that contained the flood inundation area for 14 flow percentiles for each reach in the study watershed. Thus, since the percentiles were all the same, one could attain an inundation area value for any given flow (using proper interpolation techniques) since the FDC and AIDC were both known. The rest of the flood relationships script (spatial MATLAB process) was primarily dedicated to deriving a formal output AIDC for use in the connectivity model.

A.1.6 Connectivity (Step 5)

Thus far in the modeling process, the goal was to develop both an ungauged hydrograph as well as a relationship between flow and inundation area (the area of inundation duration curve, AIDC). These two primary datasets for each reach were then used in the MATLAB connectivity model to calculate the level of connectivity on both annual and seasonal bases for each reach. The main concept of the connectivity script was to calculate an inundation area for every point on the hydrograph, and then sum those inundation area values when the flow record was above bankfull. These values were summed and multiplied by the time step, which in the case of this study was 1 day. Thus, in the end, the value retained was a total area*time metric for the flow record. This value was then normalized by dividing by the length of the flow record and multiplying by 365 to attain an annual connectivity metric.

The first few steps in the MATLAB connectivity model (the "conn_metric" script) involved setting up the file directories used to import and export data from the model. The output AIDC data from the spatial MATLAB process ("areastage" script, Section A.1.5) was imported into the connectivity model. The FDC as well as the AIDC were pulled from this input

dataset and placed into respective variables for use in the model. To maintain consistency in modeling the duration curve, exceedance values were read from the input CSV file as well.

Next, a loop was set up to iterate through each ungauged location, calculating the connectivity for each reach. Given the ID of the ungauged location, the CSV file of the predicted hydrograph (as developed at the conclusion of the temporal MATLAB model) was read in, and essentially "converted" from flow values to inundation area values. This conversion was accomplished by looping through each flow value in the long-term hydrograph and linearly interpolating an inundation area using the FDC and AIDC. Prior to interpolation, the AIDC was normalized to the reach length. At the end of this process, a long-term record of inundation area values (per length of stream reach) had been calculated for the given reach that the model was currently iterating.

Before calculating the connectivity metric, the flows occurring above bankfull (therefore actually introducing flooding) had to be isolated from the rest of the flows. Thus, a bankfull flow for each reach was needed. Determining bankfull flow is a complex science by itself. For the purposes of this study, it was deemed adequate to utilize a particular flow percentile from literature that had been found equivalent to bankfull. Although not the best way of calculating bankfull flow, it was the best way given the constraints associated with the study and little information known at each ungauged location. The literature was examined and synthesized (Table A.4) for other studies that calculated the daily flow percentile associated with bankfull, nearly bankfull, or "effective discharge" events.

Table A.4. Bankfull event daily flow percentile exceedance

Name/Location	Drainage Area (km2)	Bankfull Discharge (cms)	Bankfull Exceedance (%)
North Fork Massie, OH	74.9	--	2.3
Loramie River, OH	394	--	3.3
Portage River, OH	1109	--	1.3
Stillwater River, OH	1303	--	2.3
Upper Great Miami, OH	1401	--	0.8
Grand River, OH	1774	--	0.4
Sandusky River, OH	3240	--	2.9
Scioto River, OH	13290	--	1.8
Maumee River, OH	16395	--	2.6
Muskingum River, OH	19223	--	6.2
Little Snake River near Lily, CO/WY	9,660	133	1.1
Little Snake River near Baggs, CO/WY	7,820	167	1.1
Little Snake River near Dixon, CO/WY	2,560	114	1.4
Savery Creek near Savery, CO/WY	855	28.9	0.7
Slater Fork near Slater, CO/WY	417	16.1	2.8
Little Snake River near Slater, CO/WY	738	72.2	1.2
Yampa River near Maybell, CO/WY	8,820	255	2.5
Wilson Creek near Axial, CO/WY	51.8	0.538	0.65
Goodsprings Creek near Axial, CO/WY	90.5	0.312	0.4
Williams Fork at mouth near Hamilton, CO/WY	1,200	46.7	2.8
Elkhead Creek near Elkhead, CO/WY	54.5	15.6	1.3
Yampa River below diversion near Hayden, CO/WY	3,700	167	3
Grassy Creek near Mt. Harris, CO/WY	67.4	1.4	0.7
Elk River near Trull, CO/WY	1,080	101	2.1
Elk River at Clark, CO/WY	535	60.9	2.1
Little Missouri, Marmath, ND	--	68.79	3.0
Little Missouri, Medora, ND	--	90.2	3.3
Sevier, Hatch, UT	--	20.82	1.0
East Form San Juan, Pagosa Springs, CO	--	10.51	6.0
White, Buckley, WA (upstream of Mud Mountain Dam)	--	80.1	6.5
White, Buckley, WA (downstream Mud Mountain Dam)	--	114.7	2.3
Lower Mississippi River	--	--	13.0
Little Grizzly Creek above Hebron, CO	135	6.71	4.6

Table A.4, ctd.

Name/Location	Drainage Area (km2)	Bankfull Discharge (cms)	Bankfull Exceedance (%)
Michigan River near Cameron Pass, CO	3.96	0.697	1.1
North Platte River near North Gage, CO	3706	85.2	1.1
Halfmoon Creek near Malta, CO	61.1	7.08	0.35
Chalk Creek near Nathrop, CO	251	9.77	1.5
East Inlet Creek near Grand Lake, CO	70.5	12.2	0.8
Stillwater Creek above Granby Reservoir, CO	45.3	2.21	1.0
Frazer River near Winter Park, CO	27.2	2.69	3.0
Little Muddy Creek near Parshall, CO	16.9	1.87	0.12
South Fork Williams Fork near Leal, CO	70.5	8.36	0.22
Williams Fork near Leal, CO	231	22.6	0.95
Castle Creek near Aspen, CO	104	4.45	6.0
North Fork Fryingpan River above Cunningham, CO	31.1	3.17	4.6
Cunningham Creek near Nome, CO	18.4	2.52	1.8
Crystal River above Avalanche, CO	433	49	2.0
East River at Almont, CO	749	37.5	6.5
Tomichi Creek at Sargents, CO	386	7.08	4.6
Lake Fork at Gateview, CO	865	42	1.7
Elk River near Trull, CO	1075	101	1.7
Yampa River near Hayden, CO	3700	167	3.0
Williams Fork at mouth near Hamilton, CO	883	46.7	2.9
Yampa River near Maybell, CO	8832	255	2.5
Little Snake River near Slater, CO	738	72.2	0.22
Little Snake River near Dixon, CO	2559	114	1.5

The average of the percentiles for daily mean flows in Table A.4 was approximately 2.4%. This suggests that for the study locations in the reviewed literature, flows equal to or exceeding bankfull occurred for approximately 2.4% of the year, or 9 days. A bankfull exceedance percentage value of 2.5% was chosen for this study to approximate bankfull flows. Once established, the bankfull flow value, predicted hydrograph, and calculated long-term inundation area record were fed into a separate function to actually calculate the connectivity metric value.

The connectivity calculator function (separate MATLAB function file that was used within the "conn_metric" script file) had inputs of the timing (t) and flow (Q) components of the hydrograph, the inundation area time series record (ai), and the bankfull flow value. The first part of the MATLAB function ensured that the length of record for the flow series and inundation area series matched. If they did not match, it was a sign that the flow and inundation area may not have come from the same time series, and thus, an error was produced. Next, the loop counter variables were initialized before being utilized in the rest of the function.

The main portion of the function looped through every flow record in the dataset. Each flow value in the flow record was compared to the bankfull value using an "if" statement. If the flow was above bankfull, the inundation area corresponding to that particular flow would be multiplied by the time interval (1 day) to derive the area*time metric. This value would then be added to the cumulative counter of connectivity. Thus, the connectivity metric consisted of the inundation areas (normalized to reach length) corresponding to flows greater than bankfull, all multiplied by a time increment of 1 day. A counter was established to count the number of flow records (the number of days, given daily flow data) that were found to be above bankfull.

Besides an annual connectivity value, seasonal connectivity metrics were also desired in order to examine differences in floodplain connectivity between both reach characteristics as well as the temporal season. Two seasons were examined in order to compare their respective connectivity values: the growing season and non-growing season. The growing season was assumed to be from May through October. This assumption was partially based on the planting and harvesting dates for corn in PA (NASS, USDA, 1997). Thus the non-growing season was November through April. The seasonal metrics were calculated in an analogous method to the full record/annual metric. An "if" statement was inserted into the code which triggered the summation of the area*time value if the date of the flow was within the months designated growing season or non-growing season. Also, counters were set up to count the number of days total within each season, regardless of the flow's relative magnitude to bankfull, so that normalization could take place later in the function.

After each record in the long-term hydrograph was compared to the bankfull value, and the cumulative connectivity value was calculated, the connectivity metrics were normalized to facilitate comparison between flow records of different lengths. Normalization was necessary because locations with longer flow records would have a higher chance of having a larger

connectivity as summed in the loop. Thus, to normalize the full (annual) connectivity sums, these values were divided by the length of the flow record and then multiplied by 365 to simulate an annual connectivity value. The seasonal connectivity sums were also normalized. These sums were divided by the total number of days in that season (per the previously described counters). Then, they were multiplied by 182.5 days (half of 365), since each season was approximately half of a calendar year. Thus, in the end, the annual metric represented connectivity across a year's time, and the seasonal connectivity metrics represented the connectivity across an average season.

Once all of the connectivity metrics were properly normalized, they were sent back to the full connectivity MATLAB script. After these calculations, the rest of the connectivity script was primarily focused on outputting a results table (CSV) of the connectivity values.

Appendix B GIS Model Listing

Table B.1. GIS ModelBuilder model names and general description of function and purpose

GIS Model Name	Reference Section	Brief Purpose/Description
Surface Water 2	A.1.1.2	Creates polygon flood surface
NHD Processing	A.1.1.2	Processes raw NHD data and selects desired HUC boundary
DEM Processing	A.1.1.2	Processes and clips raw DEM data
HydroProcessing	A.1.1.2	Performs hydrology ArcGIS tools (Flow direction, flow accumulation, etc.)
Filtering	A.1.1.2	Filters DEMs to approximate surface water
Flowline	A.1.1.2	Develops flow network
COMIDlink2a	A.1.1.2	Processes flow network and creates full cross-section dataset
COMIDlink2asub	A.1.1.2	Assigns NHD COMID to the derived flow network
Filter PostProc2	A.1.1.2	Combines filtered DEMs for surface water estimation/development
SurfProc2	A.1.1.2	Prepare surface water polygon for flood simulations
COMIDlink2b	A.1.1.2	Delineate contributing land area to each reach segment
LinesandNodes2	A.1.1.3	Create outlet nodes and select representative cross-section
Mannings	A.1.1.4.1	Estimate cross-sectional Manning's n coefficient from NLCD data
Flood	A.1.1.4.2	Calculates and summarizes inundation areas at range of 0.5 m stage increments
Flood Iterations	A.1.1.4.2	Looped model to simulate each flood stage simulation
FloodZero	A.1.1.4.2	Simulate inundation area for the 0 m stage increment
Table Summary	A.1.1.4.3	Table manipulation of "Flood" model outputs
Q_S2	A.1.1.4.3	Combine flood, Manning's, and stream order data into single table
GageAttrib	A.1.2.2	Collects USGS gauge watershed attributes
gage_iterator	A.1.2.2	Delineates individual gauge watersheds and calculates attributes
WSAttrib	A.1.2.2	Collects watershed attributes of individual reach watersheds
ws_iterator	A.1.2.2	Delineates individual reach watersheds and calculates attributes
Gage3	A.1.2.3	Gauge data collection and preparation
NearestGage	A.1.2.3	Determine nearest gauge to each ungauged location for QPPQ methodology

Appendix C ArcGIS Script Tool Development

ESRI's ArcGIS has become the industry standard for spatial analysis and has a wide variety of useful applications and valuable toolsets. Despite the ever-improving ArcGIS program, there is sometimes the need to utilize other programs to analyze data in a way that ArcGIS either cannot or is extremely cumbersome to do so. Some other GIS programs may have a tool that ArcGIS does not. A numerical processing program (such as MATLAB) may be more efficient at calculations and tabular manipulation. Because of this, it is important to be able to incorporate these outside programs into ArcGIS workflow. This process can be done by simply using ArcGIS, opening the outside program, doing the work in that program, then opening ArcGIS and continuing work. This workflow is not efficient, and in today's workplace, automation is critical.

There are a few options when pursuing automation within ArcGIS. One option is using ESRI's ModelBuilder: an automation component within the constructs of ArcGIS. The benefits to ModelBuilder include an easily-used graphical interface and no complicated coding to learn. The difficulties with ModelBuilder are that an analyst may become restricted to the default settings that ArcGIS has set as well as having to deal with bugs in the program. Another option for automation is programmatic coding. Python is the scripting language behind most all of the ArcGIS tools. The benefit to straight coding is the freedom associated with the language, allowing more functionality and flexibility to GIS models and workflows. The largest disadvantage to coding is the steep learning curve for many analysts. This learning curve may discount the benefits gained in flexibility. Throughout the spatiotemporal floodplain connectivity model, ESRI's ArcGIS ModelBuilder was used to automate various processes. For some model processes, this study utilized a combination of both ModelBuilder and coding to maximize the benefits of each option in order to incorporate outside programs (specifically MATLAB) into the automated ArcGIS workflow.

The general workflow going from the most nested procedure to the outermost "shell" procedure goes as follows: Outside Program → DOS Batch File → Python Script → ArcGIS Script tool → ArcGIS ModelBuilder model (Figure C.1) First, in order to run an outside program from within ArcGIS ModelBuilder, the program must have the capability to be run from the DOS command line. Most programs and other models have command-line capabilities, as this is the most base-level way of running a program. MATLAB is a good example of those

programs that can be extremely useful in spatial analysis and can be run from the command line. Once the desired program has been chosen, the next step is to write a DOS batch (.bat) file that will run the program. Generally, this batch file is relatively simple, consisting of less than 5 lines of code. For the most simplistic batch file, the outside program should be located on the system path variable. This will prevent the necessity to locate the executable file. For example, on most computers, MATLAB is located on the system path variable as a default setting during installation. Most likely, one line of code will reset the current directory to where the MATLAB script is located. The other primary line of code is that which actually runs the program. For MATLAB in particular, there are several options to establish the particular way the computer will run the program. One such command line run option is "-wait". Adding this syntax before the particular m-file to run prevents the batch file from closing until the MATLAB program has completed and is closed. Without the "-wait" syntax, the batch file will close while MATLAB is still processing, possibly causing problems later in the workflow. In order for the batch file to receive variables from a higher level/shell (the python script), place "%1", %2", etc. for the 1st, 2nd, etc. variables within the batch file. MATLAB function files work the best for automated flow; however, script files may be run in this batch file process as well. The benefit to using function files is the ability to pass variables into the function and allow for a dynamic work environment rather than a static m-file.

The next step in the process is to write the python script (.py). This python script will call & run the batch file using the "os.system" syntax. In the python script, the "arcpy" (ArcGIS python tools) and "os" (operating system) libraries will need to be imported at the beginning of the python script. Variables may also be passed into and out of the python script. To pass variables from the python script out to the batch file, one may simply place the variables inside the "os.system" syntax. The arcpy library is utilized to bring in variables from the ArcGIS interface. One of the most popular arcpy commands is "arcpy.GetParameterAsText(0)". This syntax will essentially "request" a variable from the ArcGIS interface. The first variable will have the numeric "0" in parentheses, whereas subsequent variables will have subsequent integer values. To send variables back to ArcGIS, the "arcpy.SetParameter()" or "arcpy.SetParameterAsText()" syntaxes may be used. Inside the parentheses would go the variable of interest, as well as the variable number (same subsequent variable numbering system as for receiving variables from ArcGIS). These are the pieces required to simply run a batch file,

but more processes and tools can be run from the python script if so desired. For example, one may desire to go online to retrieve data that will be processed in the outside program and then sent back to ArcGIS. This can be done in the python script before the syntax regarding running the batch file.

The python script is the essential part to the entire workflow process; however, to promote the easy-to-use graphical interface, a script tool should be created within ArcGIS from the python script. A script too may be created by adding one to an existing toolbox or toolset. When creating a script tool, the main steps include choosing the actual python script file and setting the tool's parameters. When setting the script tool parameters (variable specifications), the variables required by the python script should be given a name (name seen in the script tool's graphical interface) and a data type. Also, the programmer may desire to set a default value to the variable/parameter. It is important to set each parameter in the order the variables are requested in the Python script. For example, the variable number "0" should be the parameter specified first in the script tool. The highest variable number should be the last variable listed and specified. After these steps have been completed, the script tool is complete.
At this point, the created tool is now equivalent to other tools in the various ArcGIS toolboxes. In order to utilize the tool in an existing workflow within ArcGIS, the tool should be placed within a ModelBuilder model. Utilizing these processes in this fashion allows outside programs to be utilized within ArcGIS ModelBuilder models, maximizing automation. One of the benefits with creating tools is the increased flexibility in data input types to the tool. Many ArcGIS tools are very restrictive as to what data types may be used as input. When a tool is created via a python script tool, the programmer may design and select parameter specifications for optimum flexibility and efficiency. Once the script tool has been added into the ModelBuilder model, the incorporation process is complete. The outside program can now be utilized within the same automated process as the ModelBuilder model.

This standard operating procedure was utilized in several instances throughout the SPATFCON model. In addition to MATLAB usage and incorporation just mentioned, similar coding procedures were used to incorporate the SAGA and GME GIS programs into the model workflow, as noted in the Methods section.

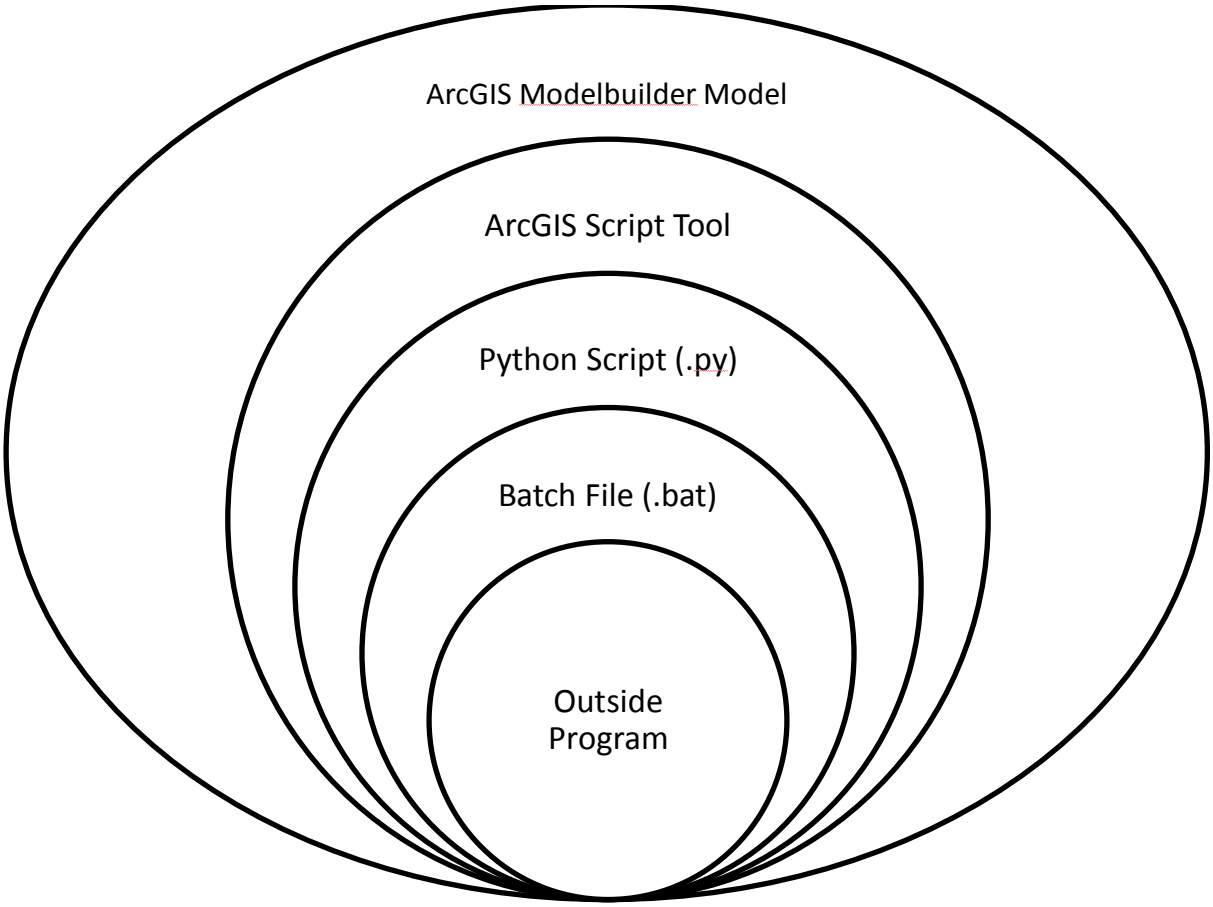

Figure C.1. Nested procedural schematic for incorporating outside programs within ArcGIS.

Appendix D Manning's Lookup Table Documentation

Where multiple sources are listed, the average was taken between the two assumed nearly equivalent land use categories.

Table D.1. NLCD code to Manning's n translation source documentation.

NLCD Code	NLCD Description	Mannings n	Source
11	Open water	0.025	1,2
12	Ice/snow	0.022	2
21	Developed, Open Space	0.028	1 ("Recreational Grass"), 2, 3
22	Developed, Low Intensity	0.079	1 ("Low Residential"), 2, 3
23	Developed, Medium Intensity	0.071	2, 3
24	Developed, High Intensity	0.087	1 ("High residential"), 2, 3
31	Barren Land	0.027	1, 2, 3
41	Deciduous Forest	0.213	1, 3, 3
42	Evergreen forest	0.207	1, 2, 3
43	Mixed Forest	0.230	1, 2, 3
51	Dwarf Scrub	0.130	Same as NLCD 52
52	Shrub/Scrub	0.130	1, 2, 3
71	Grassland/Herbaceous	0.144	1, 2, 3
72	ALASKA ONLY	0.000	No source
73	ALASKA ONLY	0.000	No source
74	ALASKA ONLY	0.000	No source
81	Pasture/Hay	0.133	1, 2, 3
82	Cultivated Crops	0.038	1, 2
90	Woody Wetland	0.109	1, 2, 3
95	Emergent Herbaceous Wetland	0.141	2, 3

[1]Bunya et al., Table 5

[2]Moore, 2011, Table 3.1

[3]Kalyanapu et al., 2009, Table 2

Appendix E Regression Results

Table E.1. Gauge watershed attributes from the temporal GIS model.

SITENO	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
1439500	303.0	2.9	81.3	4.6	0.0	126.7	495.6	7.7	3200	3016.7	7871	7425.7	22510.5	21236.9	9151.8	8627.4
1440000	168.6	2.6	78.5	3.2	6.2	105.3	394.5	11.5	3136	3049.5	7568	7376.5	19747.7	19247.9	8183.0	7957.2
1442500	677.2	2.8	70.1	19.4	4.0	91.4	574.4	9.9	3330	3204.4	8297	7935.3	22913.7	21914.7	9196.4	8849.5
1443500	329.7	2.6	49.3	12.1	23.7	101.7	386.6	9.1	3406	3186.6	7884	7529.3	20777.7	19843.0	8976.3	8398.0
1448000	829.8	2.9	69.2	9.9	0.8	315.1	376.9	5.7	3328	3181.6	8324	7984.9	23975.5	22998.9	9585.6	9163.8
1449800	267.0	2.6	66.9	15.1	14.2	149.8	478.6	9.8	3334	3267.8	8358	8127.6	21365.5	20776.6	8522.7	8353.3
1450000	283.8	2.6	65.9	15.3	15.3	135.4	493.0	10.0	3334	3263.8	8358	8116.6	21590.3	20966.8	8612.4	8431.1
1450500	198.2	2.6	69.1	9.8	19.8	118.8	396.4	15.5	3288	3191.8	7921	7735.7	20866.1	20378.0	8661.5	8408.0
1468500	373.5	3.1	73.6	16.0	5.3	141.9	405.2	14.4	3511	3379.7	8766	8328.8	27113.5	25761.3	10859.6	10453.5
1470500	928.3	2.8	71.2	12.6	12.4	96.5	539.4	14.6	3522	3390.9	8824	8356.0	25084.3	23753.8	10012.1	9639.5
1470756	413.8	2.7	39.4	8.9	50.2	89.7	414.4	12.0	3367	3270.3	8258	7868.6	22437.3	21379.4	9148.3	8885.6
1534500	281.1	3.3	70.0	9.0	12.3	264.6	542.3	10.5	3022	2921.7	7309	7004.2	24085.6	23081.1	9958.5	9628.1
1536000	861.0	3.3	63.5	20.5	6.7	178.7	627.5	10.9	3194	2867.9	8046	7052.9	26464.2	23197.8	10505.4	9433.0
1539000	704.7	2.9	65.9	4.4	27.1	167.9	620.7	15.3	3199	2918.7	7282	6926.3	21011.0	19984.7	9230.2	8421.4
1546500	223.9	2.0	37.1	26.1	36.3	243.3	496.6	9.3	2726	2658.5	6177	5964.6	12585.7	12153.0	5554.3	5416.6

Table E.1, ctd.

SITENO	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
1547100	377.7	2.0	42.4	21.4	35.5	215.1	524.8	10.7	2726	2652.6	6178	5936.9	12496.1	12008.5	5513.8	5365.3
1547200	694.8	2.3	61.6	15.0	22.9	208.0	555.8	15.0	2726	2636.4	6177	5871.6	14190.9	13489.2	6262.6	6056.7
1547500	878.9	2.3	62.5	13.7	22.5	177.0	586.8	15.3	2729	2640.9	6178	5874.7	14442.1	13733.0	6379.5	6173.6
1555000	793.0	2.1	71.6	5.4	22.5	153.5	582.6	16.6	3123	2909.9	7514	6855.6	16044.7	14638.9	6668.6	6213.6
1556000	755.0	2.3	65.6	14.2	19.6	249.6	704.7	16.8	2795	2669.0	6203	5948.3	14300.3	13713.1	6443.6	6153.0
1558000	579.8	2.3	74.3	11.1	14.0	230.2	579.4	16.4	2736	2602.0	6057	5763.8	13983.7	13306.8	6316.6	6007.1
1559500	335.2	2.4	84.7	4.6	10.4	187.9	551.6	17.3	2749	2670.9	6242	6012.5	15222.4	14662.7	6704.0	6513.6
1562500	220.7	2.3	79.5	5.5	13.5	238.5	505.5	11.4	2777	2718.6	6252	6121.9	14547.8	14245.0	6461.8	6325.8
1564500	528.3	2.6	77.4	6.3	16.0	187.8	559.8	17.3	2932	2742.8	6640	6189.3	17328.1	16152.0	7651.5	7157.8
1565000	422.1	2.1	62.2	7.3	30.3	166.6	549.3	17.2	3050	2796.1	7293	6430.5	15554.4	13714.9	6505.0	5963.5
1566000	544.8	2.7	75.8	4.1	19.6	122.8	571.0	18.3	2924	2815.5	7037	6605.2	18728.4	17579.1	7782.0	7493.3
1568000	534.9	2.4	69.9	5.6	23.9	129.8	556.5	16.4	2944	2863.3	7086	6969.9	16915.6	16638.4	7027.9	6835.1
1573000	871.4	2.8	51.0	10.9	36.4	108.7	414.0	10.7	3305	3161.4	8071	7723.1	22431.1	21464.2	9185.3	8786.2

Table E.2. Study reach (COMID) watershed attributes from the temporal GIS model.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4517826	5.0	2.0	86.5	12.4	0.3	295.6	239.0	16.3	3266	3232.7	8040	7941.3	16085.9	15888.4	6534.4	6467.8
4517836	2.8	2.3	50.9	4.4	44.7	235.8	232.4	17.2	3000	2992.3	7335	7322.5	16757.6	16729.1	6853.8	6836.2
4517838	2.4	2.5	70.8	7.2	22.0	285.8	207.4	19.5	3085	3074.8	7496	7469.6	18844.7	18778.5	7755.6	7730.0
4517840	2.0	2.2	54.6	6.5	38.7	244.7	260.8	17.8	3014	3007.1	7365	7354.0	16566.2	16541.4	6779.5	6763.9
4517842	3.8	2.4	95.7	2.9	1.3	285.8	248.6	19.3	3104	3091.4	7542	7510.8	18406.6	18330.4	7575.5	7544.8
4517848	1.7	2.2	40.8	7.1	52.1	248.0	223.3	16.4	3043	3035.6	7419	7406.5	16368.8	16341.3	6713.9	6697.6
4517850	1.4	2.4	48.8	7.3	43.9	227.4	274.9	16.4	3005	3000.1	7346	7338.8	17794.7	17777.1	7279.2	7267.2
4517856	1.4	2.1	95.4	2.8	1.0	258.9	278.4	20.2	3122	3106.2	7593	7552.5	15736.1	15652.2	6470.2	6437.4
4517858	3.5	2.1	95.5	1.8	2.4	255.9	281.4	21.0	3127	3109.1	7607	7558.6	16012.6	15910.7	6582.3	6544.7
4517862	3.3	2.3	34.9	6.7	58.4	232.4	238.9	15.8	3044	3033.1	7419	7402.1	17265.3	17225.9	7083.9	7058.5
4517868	2.2	2.6	32.3	8.1	59.6	219.2	282.9	14.0	3011	3002.2	7355	7341.1	19404.9	19368.2	7944.0	7920.8
4517870	6.8	2.6	38.0	6.0	56.0	219.1	283.2	15.0	3011	2996.2	7355	7328.4	19026.6	18957.9	7789.2	7750.9
4517872	16.6	2.5	29.0	7.5	63.3	221.1	261.0	14.7	3069	3038.5	7465	7410.4	18719.3	18582.3	7695.9	7619.4
4517874	3.1	2.5	42.0	6.8	50.4	222.0	283.5	15.2	3020	3009.8	7373	7357.5	18575.2	18536.1	7608.4	7582.6
4517876	3.5	2.1	96.3	1.5	1.9	258.1	279.2	21.1	3127	3109.0	7607	7558.1	15935.6	15833.2	6550.6	6512.9
4517878	9.1	2.6	27.2	7.5	65.2	231.9	250.2	13.9	3069	3049.0	7464	7427.9	19547.8	19453.3	8037.5	7985.2
4517882	6.5	2.3	76.5	3.7	19.3	250.1	301.9	17.8	3142	3114.6	7651	7574.8	17837.1	17659.4	7325.1	7261.2
4517890	20.2	2.4	74.2	7.2	17.9	249.7	302.3	16.8	3265	3168.7	8039	7738.0	19279.8	18557.8	7830.4	7599.5
4517900	3.5	2.7	24.5	8.1	66.8	257.1	227.2	16.0	3079	3068.3	7484	7460.1	20306.3	20241.4	8354.3	8325.2

Table E.2, ctd.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4517904	10.9	2.5	75.4	4.0	20.5	257.6	276.7	18.4	3106	3087.1	7545	7500.2	18947.4	18834.9	7799.9	7752.6
4517914	0.5	3.0	27.5	13.2	59.3	244.3	58.6	13.8	2982	2978.1	7294	7284.0	21882.0	21852.0	8946.0	8934.4
4517920	1.3	3.0	10.4	8.3	81.3	246.5	63.9	8.5	3069	3061.1	7464	7451.0	22392.0	22353.0	9207.0	9183.4
4517930	23.8	2.6	29.6	7.3	62.9	205.1	300.4	14.6	3069	3031.2	7465	7395.9	19402.0	19222.3	7976.5	7878.3
4517932	7.7	2.6	36.6	6.3	57.1	205.2	297.1	15.3	3011	2997.2	7355	7329.6	19407.0	19340.0	7944.9	7908.5
4517936	1.2	3.0	47.4	6.0	46.6	224.6	96.0	17.8	2977	2969.1	7281	7261.3	21843.0	21783.8	8931.0	8907.4
4517940	0.6	3.0	41.8	13.7	44.5	222.7	86.9	20.2	2955	2948.9	7226	7213.0	21678.0	21639.1	8865.0	8846.8
4517944	0.8	3.0	34.0	7.9	58.1	224.4	77.7	16.4	2992	2987.5	7311	7303.6	21933.0	21910.8	8976.0	8962.4
4517956	1.8	3.0	16.1	8.6	75.3	228.1	82.3	10.0	3069	3060.0	7464	7448.8	22392.0	22346.3	9207.0	9180.0
4517958	2.1	3.0	45.0	7.7	46.6	211.8	108.8	15.9	2969	2960.4	7260	7240.3	21780.0	21720.9	8907.0	8881.2
4517960	2.0	3.0	60.7	6.2	31.6	212.1	130.2	16.6	2955	2946.8	7230	7207.4	21690.0	21622.3	8865.0	8840.3
4517964	1.6	3.0	33.0	8.1	58.9	209.0	94.8	16.1	2986	2981.1	7299	7289.1	21897.0	21867.3	8958.0	8943.3
4517966	4.5	3.0	35.7	7.1	57.2	208.6	112.1	17.0	2986	2974.8	7299	7273.9	21897.0	21821.7	8958.0	8924.3
4517968	1.5	3.0	40.5	6.0	53.5	203.9	97.4	18.3	3001	2994.1	7329	7316.1	21987.0	21948.4	9003.0	8982.2
4517970	1.1	3.0	45.5	7.1	47.4	202.0	100.1	18.1	2994	2988.3	7314	7304.5	21942.0	21913.4	8982.0	8964.8
4517982	34.2	2.6	31.1	6.9	61.8	193.3	312.2	15.0	3069	3021.3	7465	7375.8	19698.9	19463.5	8098.6	7972.6
4517984	2.7	3.0	41.3	6.7	52.0	194.6	107.5	18.2	3001	2991.7	7329	7311.2	21987.0	21933.7	9003.0	8975.1
4517990	1.4	3.0	18.7	8.2	72.0	219.8	90.9	9.7	3035	3027.5	7400	7385.4	22200.0	22156.2	9105.0	9082.6
4518000	1.8	3.0	39.9	8.1	51.9	240.0	254.7	12.5	3178	3161.4	7737	7690.2	23036.8	22897.4	9462.4	9412.9

Table E.2, ctd.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4518002	36.9	2.5	73.0	6.2	20.1	238.2	313.8	16.7	3266	3169.1	8040	7726.6	20324.4	19532.2	8256.2	8011.1
4518004	11.4	2.6	84.0	4.3	10.7	239.1	293.7	18.4	3235	3186.6	7901	7758.7	20878.2	20502.2	8548.4	8420.4
4518010	9.6	2.6	92.5	3.5	2.8	239.9	292.8	19.5	3235	3191.5	7900	7772.2	20343.0	20013.9	8330.3	8218.3
4518018	0.9	3.0	35.1	4.2	60.6	203.6	94.2	16.7	2974	2970.0	7269	7261.2	21807.0	21783.7	8922.0	8909.9
4518026	0.9	3.0	21.1	9.3	69.6	212.7	75.8	13.5	2967	2963.2	7256	7246.1	21768.0	21738.3	8901.0	8889.6
4518032	7.6	3.0	29.8	7.0	63.1	184.2	136.4	16.2	2986	2975.0	7299	7273.5	21897.0	21820.6	8958.0	8925.1
4518034	1.5	3.0	25.2	6.1	68.3	185.9	111.9	15.4	2979	2971.8	7281	7265.1	21843.0	21795.4	8937.0	8915.3
4518036	1.0	3.0	23.9	5.4	70.6	203.5	90.2	17.2	2964	2957.9	7248	7233.0	21744.0	21699.0	8892.0	8873.6
4518038	49.4	2.7	47.6	5.7	46.4	186.9	347.4	16.3	3107	3058.5	7546	7444.5	20566.2	20289.7	8468.0	8335.8
4518040	3.4	3.0	12.0	8.1	79.4	203.5	107.3	9.7	3036	3025.9	7404	7381.4	22212.0	22144.3	9108.0	9077.8
4518044	33.2	2.7	48.7	5.7	45.4	208.8	325.6	16.0	3106	3072.7	7546	7472.1	20102.2	19905.2	8274.2	8185.6
4518046	36.2	2.6	51.2	5.8	42.8	208.7	325.7	16.7	3107	3071.4	7545	7469.7	19992.8	19793.3	8233.0	8138.7
4518050	38.2	2.7	52.0	5.7	42.1	201.7	332.7	16.9	3107	3069.9	7545	7466.9	20026.2	19818.8	8246.7	8148.2
4518052	49.8	2.7	47.2	5.7	46.8	183.8	350.6	16.2	3106	3057.8	7545	7443.0	20606.5	20327.8	8482.9	8351.1
4518054	39.1	2.7	31.0	7.1	61.7	183.8	321.7	15.3	3069	3017.6	7465	7367.6	20037.1	19775.8	8237.6	8099.6
4518062	47.4	2.7	49.3	5.6	44.8	187.1	347.3	16.5	3107	3060.6	7545	7448.8	20479.8	20218.8	8433.5	8307.7
4518064	7.9	3.0	65.8	5.2	28.3	167.2	183.9	18.9	2969	2950.6	7261	7216.2	21783.0	21648.6	8907.0	8851.7
4518066	9.6	3.0	65.1	6.1	28.2	167.0	184.0	19.1	2969	2948.0	7260	7209.8	21780.0	21629.3	8907.0	8843.9
4518068	0.9	3.0	28.9	9.6	61.5	171.9	158.0	12.7	2949	2939.9	7212	7189.4	21636.0	21568.2	8847.0	8819.6

Table E.2, ctd.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4518076	1.3	3.0	53.1	6.8	40.1	183.2	149.3	16.6	2934	2926.4	7172	7155.1	21516.0	21465.3	8802.0	8779.1
4518092	103.6	2.7	40.1	6.3	53.4	171.5	362.9	16.1	3106	3032.3	7545	7393.6	20730.2	20314.3	8533.9	8331.3
4518094	104.0	2.7	40.2	6.3	53.3	170.3	364.0	16.1	3107	3032.2	7545	7393.5	20730.0	20313.8	8536.5	8331.1
4518096	2.8	3.0	19.0	8.5	72.5	170.1	118.3	12.6	2982	2969.6	7289	7260.6	21867.0	21781.9	8946.0	8908.8
4518098	12.0	3.0	56.8	6.4	36.4	157.4	193.7	17.9	2969	2947.8	7261	7209.2	21783.0	21627.7	8907.0	8843.3
4518100	1.6	3.0	22.0	9.4	68.6	158.6	171.4	12.4	2953	2942.7	7222	7196.0	21666.0	21588.1	8859.0	8828.2
4518104	94.8	2.7	41.2	6.3	52.3	171.5	362.9	16.1	3106	3037.4	7545	7404.4	20559.3	20176.2	8463.5	8276.6
4518106	12.0	3.0	56.8	6.4	36.4	156.1	194.9	17.9	2969	2947.8	7261	7209.2	21783.0	21627.7	8907.0	8843.3
4518108	0.9	3.0	70.8	12.4	16.8	184.9	130.7	19.9	2914	2909.3	7121	7107.3	21363.0	21321.8	8742.0	8728.0
4518110	314.0	2.7	55.6	6.9	36.7	156.0	396.0	16.6	3266	3069.3	8040	7475.2	21816.5	20283.9	8862.3	8328.5
4518124	110.9	2.8	40.2	6.3	53.3	160.6	373.7	16.2	3106	3028.5	7545	7385.8	20827.7	20388.1	8574.0	8359.9
4518126	0.5	3.0	47.7	11.7	40.6	188.0	100.5	17.6	2889	2884.8	7046	7038.1	21138.0	21114.3	8667.0	8654.4
4518128	110.9	2.8	40.2	6.3	53.3	160.8	373.5	16.2	3106	3028.5	7545	7385.9	20823.0	20384.0	8572.1	8358.3
4518130	1.2	3.0	56.2	7.0	36.7	183.5	121.6	16.4	2903	2896.9	7085	7069.9	21255.0	21209.6	8709.0	8690.6
4518132	4.6	3.0	42.8	5.9	51.2	151.6	188.2	14.9	2949	2933.6	7210	7172.5	21630.0	21517.4	8847.0	8800.8
4518136	1.1	3.0	38.9	3.3	57.8	172.3	160.4	16.2	2911	2902.3	7106	7086.3	21318.0	21258.8	8733.0	8707.0
4518140	4.4	3.0	43.6	6.7	49.7	150.3	166.0	16.5	2937	2923.3	7179	7144.1	21537.0	21432.3	8811.0	8769.8
4518146	1.6	3.0	55.1	7.0	38.0	167.9	137.1	17.4	2903	2896.1	7085	7067.6	21255.0	21202.8	8709.0	8688.4
4518148	1.3	3.0	65.8	8.3	25.9	168.1	120.8	19.0	2894	2887.6	7060	7045.4	21180.0	21136.3	8682.0	8662.9

Table E.2, ctd.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4518150	1.1	3.0	38.6	7.3	53.2	236.1	159.6	14.7	3139	3131.6	7580	7566.2	22740.0	22698.5	9417.0	9394.9
4518152	340.6	2.7	55.4	6.9	37.0	149.1	402.9	16.8	3265	3059.1	8038	7452.6	21932.2	20335.0	8908.8	8346.9
4518164	6.0	3.2	90.3	9.6	0.0	258.4	243.1	20.7	3228	3201.1	7851	7768.3	25458.8	25190.6	10467.6	10380.5
4518166	1.5	3.0	46.7	7.8	45.5	168.4	134.0	21.7	2874	2869.1	7000	6985.5	21000.0	20956.5	8622.0	8607.4
4518168	0.8	3.0	48.3	5.8	45.8	178.7	122.7	19.9	2884	2880.0	7030	7018.5	21090.0	21055.6	8652.0	8640.1
4518170	3.4	3.0	32.0	8.4	59.5	146.6	169.1	15.5	2931	2916.2	7160	7123.6	21395.2	21286.4	8758.3	8714.0
4518172	347.7	2.7	55.2	6.8	37.2	144.5	407.5	16.8	3266	3056.4	8040	7446.5	21924.0	20305.7	8906.0	8334.5
4518174	4.0	3.0	38.8	6.6	54.6	145.1	188.6	15.9	2923	2907.9	7135	7099.8	21366.5	21261.1	8753.2	8707.8
4518188	197.8	2.7	64.9	7.4	26.6	156.8	395.1	16.9	3266	3094.9	8040	7531.1	21565.5	20200.4	8760.3	8301.2
4518192	1.8	2.7	24.8	11.1	62.6	232.9	205.9	12.6	3150	3142.9	7604	7586.9	20822.3	20775.5	8625.7	8606.3
4518194	352.8	2.7	55.1	6.8	37.3	143.7	408.3	16.8	3266	3054.4	8040	7441.9	21933.4	20301.7	8909.8	8332.6
4518196	18.0	2.7	80.7	7.0	12.1	232.3	269.2	19.7	3228	3178.2	7851	7689.1	21324.3	20884.6	8767.6	8632.5
4518200	1.2	2.7	59.0	3.8	36.6	198.6	198.5	18.2	2972	2962.3	7272	7245.5	19364.0	19293.5	7913.9	7888.0
4518202	2.8	2.8	51.6	3.1	45.1	161.9	235.2	17.0	2983	2968.8	7303	7263.2	20471.0	20359.6	8361.6	8321.9
4518204	180.8	2.7	67.1	7.4	24.3	167.2	384.7	17.0	3265	3104.9	8039	7551.5	21466.9	20165.2	8718.7	8291.1
4518208	190.9	2.7	65.0	7.5	26.4	161.8	390.1	16.8	3266	3099.3	8040	7540.5	21567.0	20227.1	8760.9	8313.7
4518210	180.8	2.7	67.1	7.4	24.3	167.4	384.5	17.0	3265	3104.9	8039	7551.6	21466.5	20164.9	8718.5	8291.0
4518212	183.9	2.7	66.9	7.3	24.6	163.5	388.4	17.0	3265	3102.9	8039	7547.6	21457.5	20146.0	8714.9	8282.3
4518214	383.9	2.7	53.0	6.9	39.3	137.2	414.8	16.5	3265	3045.8	8039	7422.5	21973.4	20288.3	8924.4	8325.3

Table E.2, ctd.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4518216	389.7	2.7	52.9	6.9	39.4	137.2	414.8	16.5	3266	3043.7	8040	7417.3	22008.8	20304.2	8940.4	8331.7
4518220	27.3	2.8	27.7	7.4	64.6	144.1	242.8	11.9	3008	2955.5	7369	7222.2	20722.7	20310.0	8458.9	8311.4
4518226	11.4	2.8	46.8	7.7	45.2	168.2	325.3	14.1	3050	3026.0	7482	7413.2	20639.2	20449.5	8413.5	8347.3
4518228	14.2	2.9	22.5	7.3	69.8	182.8	204.1	10.7	3008	2969.5	7370	7263.3	21030.2	20725.7	8583.3	8473.4
4518230	82.4	2.6	62.4	7.0	30.1	176.5	375.5	16.3	3266	3123.2	8040	7600.6	20665.7	19536.2	8394.8	8027.7
4518232	4.1	3.0	28.3	6.9	64.7	131.6	170.7	17.6	2883	2872.5	7013	6991.1	21039.0	20973.2	8649.0	8617.5
4518234	169.4	2.7	68.4	7.4	22.9	167.4	384.5	17.2	3266	3110.4	8040	7561.4	21428.4	20152.8	8704.6	8290.0
4518236	24.1	2.9	22.2	7.3	70.1	165.6	221.3	10.7	3008	2960.4	7369	7236.1	21111.8	20731.1	8617.8	8481.3
4518238	4.2	3.0	33.9	6.7	59.3	132.6	168.3	17.4	2890	2881.2	7038	7017.2	21114.0	21051.7	8670.0	8643.7
4518240	400.7	2.7	52.6	6.9	39.7	131.5	420.4	16.6	3266	3039.4	8040	7406.9	22010.5	20277.2	8941.1	8320.8
4518244	396.3	2.7	52.8	6.9	39.5	132.5	419.4	16.6	3266	3041.3	8040	7411.4	21987.4	20268.2	8931.7	8317.1
4518246	21.8	2.9	19.6	7.1	73.0	170.4	216.5	10.2	3008	2963.3	7369	7244.7	21279.4	20920.3	8686.2	8557.1
4518250	73.3	2.8	77.1	8.2	12.3	176.4	369.5	18.3	3228	3114.4	7851	7553.2	21895.9	21065.3	9002.7	8685.8
4518252	1.3	3.0	21.8	3.6	74.6	152.9	88.0	17.7	2863	2858.9	6959	6946.1	20877.0	20838.4	8589.0	8576.8
4518254	2.7	2.5	64.5	8.1	27.1	167.8	191.0	17.0	2931	2920.4	7151	7120.4	17925.2	17848.5	7347.0	7320.4
4518258	1.2	3.0	12.9	11.9	75.0	171.0	57.4	8.4	2940	2932.9	7174	7153.8	21522.0	21461.5	8820.0	8798.8
4518264	5.2	3.0	17.4	10.4	72.2	163.5	91.5	11.4	3018	3002.6	7395	7353.8	22185.0	22061.3	9054.0	9007.7
4518266	3.0	3.0	14.2	9.6	75.3	174.9	80.2	10.1	3027	3016.0	7419	7388.4	22257.0	22165.2	9081.0	9047.9
4518270	4.0	3.0	30.3	10.4	59.2	128.6	182.8	16.9	2877	2865.9	6988	6967.7	20622.2	20562.4	8490.3	8457.6

Table E.2, ctd.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4518272	407.2	2.7	52.4	6.9	39.9	128.5	423.5	16.6	3266	3036.9	8040	7400.5	22017.5	20266.1	8943.9	8316.5
4518276	0.6	3.0	6.5	5.5	88.0	194.7	33.5	7.0	2940	2934.8	7174	7158.6	21522.0	21475.9	8820.0	8804.3
4518278	16.5	2.9	20.9	7.1	71.6	182.4	204.5	10.3	3008	2968.1	7369	7258.9	21166.7	20850.3	8640.2	8525.4
4518280	10.9	2.7	47.1	7.4	45.1	174.8	318.7	14.2	3050	3027.0	7482	7415.8	20556.6	20374.6	8379.8	8316.6
4518282	1.3	3.0	0.7	21.2	78.0	210.1	44.9	6.9	3008	2995.8	7369	7335.6	22107.0	22006.9	9024.0	8987.4
4518284	24.8	3.0	86.5	7.2	0.6	204.8	341.1	20.2	3152	3097.8	7567	7518.5	22884.0	22737.5	9532.2	9368.3
4518286	2.5	3.0	14.7	6.7	78.5	124.0	117.0	15.2	2866	2860.3	6959	6945.8	20722.0	20682.6	8534.2	8517.1
4518288	11.2	3.2	79.0	9.3	0.6	266.4	256.5	18.7	3152	3122.2	7567	7540.1	24171.6	24085.7	10068.6	9973.5
4518290	5.4	2.6	77.0	8.2	1.1	334.8	188.1	17.1	3152	3130.5	7567	7547.8	19426.9	19377.7	8092.2	8037.0
4518292	1.8	3.0	13.6	15.8	70.6	200.9	54.1	9.4	3017	3004.4	7395	7359.5	22185.0	22078.4	9051.0	9013.1
4518296	415.4	2.7	52.2	6.9	40.1	123.9	428.0	16.6	3266	3033.4	8040	7391.5	22005.6	20230.5	8939.1	8302.4
4518298	1.7	3.0	11.7	10.6	77.7	198.3	56.9	8.7	3026	3018.1	7418	7395.8	22254.0	22187.3	9078.0	9054.4
4518302	3.5	2.5	84.0	4.9	11.0	197.3	295.8	19.1	3050	3037.4	7481	7444.4	18730.7	18639.1	7636.5	7604.9
4518306	425.4	2.7	51.7	6.9	40.5	117.2	434.8	16.6	3265	3029.1	8039	7380.3	21957.3	20158.1	8917.9	8273.5
4518312	12.1	2.8	93.0	5.5	0.5	266.6	279.3	20.2	3104	3075.9	7529	7499.9	21412.1	21329.5	8827.6	8747.8
4518318	3.4	2.0	93.3	6.7	0.0	336.2	203.2	18.9	3077	3064.1	7515	7491.3	15030.0	14982.5	6154.0	6128.2
4518672	2.1	2.2	36.9	9.1	54.0	265.1	217.0	16.3	3065	3054.1	7455	7435.8	16453.6	16411.2	6764.6	6740.7
4518676	24.9	2.6	59.5	4.7	35.5	226.6	307.8	17.6	3106	3080.9	7545	7487.2	19251.6	19104.1	7925.2	7861.2
4518680	116.0	2.8	39.8	6.3	53.8	156.8	377.5	16.2	3107	3025.8	7545	7380.0	20875.1	20418.7	8596.3	8371.5

Table E.2, ctd.

COMID	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
4518682	1.6	3.0	61.1	6.4	32.6	170.4	145.9	19.0	2926	2917.0	7149	7128.7	21447.0	21386.2	8778.0	8750.9
4518684	328.4	2.7	55.6	6.9	36.7	151.8	400.2	16.7	3266	3064.0	8040	7463.7	21897.4	20327.7	8895.1	8345.0
4518688	54.8	2.5	69.6	5.9	23.9	211.1	340.9	17.0	3266	3154.1	8040	7674.8	20334.8	19411.2	8260.4	7977.3
4518690	55.0	2.5	69.4	6.0	24.0	210.7	341.3	17.0	3266	3153.9	8040	7674.2	20347.6	19421.9	8265.6	7981.8
4518692	342.9	2.7	55.4	6.8	36.9	146.5	405.5	16.8	3266	3058.4	8040	7451.1	21918.5	20313.1	8903.7	8337.8
4518694	34.8	2.7	71.7	9.2	18.7	204.1	297.4	17.4	3228	3149.7	7851	7616.1	21461.4	20819.4	8824.0	8609.9
4518696	5.2	2.9	25.2	4.9	69.0	191.7	193.6	10.2	2994	2968.0	7330	7259.9	20899.5	20699.7	8536.6	8462.4
4518698	6.0	3.0	3.5	10.8	85.3	191.7	63.4	7.5	3008	2981.5	7369	7296.0	22107.0	21887.9	9024.0	8944.6
4518764	6.0	2.0	85.2	13.6	0.3	295.5	239.0	16.2	3266	3229.4	8040	7928.5	16089.6	15866.6	6535.9	6462.7
4518800	6.2	3.2	89.8	9.8	0.1	256.2	245.3	20.7	3228	3200.5	7851	7765.8	25037.2	24765.5	10294.3	10206.6
4518842	1.9	3.0	12.7	10.4	75.6	195.5	59.6	9.1	3026	3018.0	7418	7395.1	22254.0	22185.4	9078.0	9054.0

Table E.3. Binary watershed characteristic (independent variable) model inclusion matrix. Rows were summed to develop the frequency of significance for each model variable.

Independent Variable	Exceedance Probability (%)														Significance Frequency
	99	90	70	50	40	30	25	20	10	5	2.5	1	0.1	0.01	
DA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
HSG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PCT_FOR	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
PCT_DEV	1	1	1	1	0	0	0	0	0	0	1	1	1	1	8
PCT_AG	0	0	0	0	0	1	1	1	0	0	0	0	0	0	3
ELEV_MIN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ELEV_RNGE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLOPE_MEAN	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
2yr24MAX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2yr24hrMEAN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100yr24MAX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100yr24MEAN	0	0	1	1	1	1	1	1	0	0	0	0	0	0	6
HSG*100max	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HSG*100mean	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
HSG*2yrmax	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HSG*2yrmean	0	0	0	0	0	0	0	0	0	1	1	1	1	1	5

Table E.4. Independent variable matrix used in the regression. The last four columns represent the actual independent variable matrix, while the first column represents the site numbers of the gauges used in the regression.

USGS Gauge Site Number	Constant	DA*	PCT_DEV*	100yr24mean*
1439500	1	8.48	0.75	3.87
1440000	1	8.23	0.62	3.87
1442500	1	8.83	1.31	3.90
1443500	1	8.52	1.12	3.88
1448000	1	8.92	1.04	3.90
1449800	1	8.43	1.21	3.91
1450000	1	8.45	1.21	3.91
1450500	1	8.30	1.03	3.89
1468500	1	8.57	1.23	3.92
1470500	1	8.97	1.13	3.92
1470756	1	8.62	1.00	3.90
1534500	1	8.45	1.00	3.85
1536000	1	8.94	1.33	3.85
1539000	1	8.85	0.74	3.84
1546500	1	8.35	1.43	3.78
1547100	1	8.58	1.35	3.77
1547200	1	8.84	1.20	3.77
1547500	1	8.94	1.17	3.77
1555000	1	8.90	0.81	3.84
1556000	1	8.88	1.18	3.77
1558000	1	8.76	1.08	3.76
1559500	1	8.53	0.75	3.78
1562500	1	8.34	0.81	3.79
1564500	1	8.72	0.86	3.79
1565000	1	8.63	0.92	3.81
1566000	1	8.74	0.71	3.82
1568000	1	8.73	0.82	3.84
1573000	1	8.94	1.08	3.89

*Indicates the attribute and logarithmic transformations that took place on the variables prior to use in the regression tool.

Table E.5. Predicted FDCs for the ungauged reaches in the study watershed.

COMID	Q_{99} (ft^3/s)	Q_{90} (ft^3/s)	Q_{70} (ft^3/s)	Q_{50} (ft^3/s)	Q_{40} (ft^3/s)	Q_{30} (ft^3/s)	Q_{25} (ft^3/s)	Q_{20} (ft^3/s)	Q_{10} (ft^3/s)	Q_{5} (ft^3/s)	$Q_{2.5}$ (ft^3/s)	Q_{1} (ft^3/s)	$Q_{0.1}$ (ft^3/s)	$Q_{0.01}$ (ft^3/s)
4517826	-0.4	0.1	1.0	1.8	2.2	2.5	2.7	3.0	3.8	4.9	6.9	9.2	15.0	17.5
4517836	-0.9	-0.7	-0.3	0.2	0.4	0.6	0.8	1.0	1.5	2.2	3.4	4.9	9.1	9.6
4517838	-0.8	-0.6	-0.3	0.2	0.4	0.5	0.6	0.7	1.1	1.6	2.5	3.6	6.4	6.8
4517840	-0.9	-0.7	-0.4	-0.1	0.1	0.2	0.3	0.4	0.7	1.1	1.8	2.7	5.0	5.1
4517842	-0.9	-0.7	-0.2	0.5	0.9	1.3	1.5	1.8	2.7	3.8	5.7	8.3	16.1	17.2
4517848	-0.9	-0.7	-0.5	-0.2	-0.1	0.0	0.1	0.2	0.4	0.7	1.3	2.0	3.7	3.7
4517850	-0.9	-0.8	-0.6	-0.3	-0.2	-0.2	-0.1	0.0	0.1	0.4	0.9	1.4	2.7	2.7
4517856	-0.9	-0.9	-0.7	-0.4	-0.3	-0.1	-0.1	0.0	0.3	0.6	1.3	2.1	4.3	4.0
4517858	-0.9	-0.8	-0.3	0.3	0.7	1.1	1.4	1.6	2.6	3.7	5.7	8.5	17.2	17.8
4517862	-0.8	-0.5	-0.1	0.5	0.8	1.0	1.2	1.3	1.9	2.7	4.0	5.6	9.8	10.8
4517868	-0.8	-0.6	-0.3	0.1	0.2	0.4	0.5	0.6	0.9	1.3	2.1	3.0	5.4	5.7
4517870	-0.6	-0.2	0.7	1.9	2.6	3.2	3.6	4.0	5.6	7.4	10.5	14.6	26.4	31.7
4517872	0.1	1.2	3.4	6.3	7.9	9.5	10.7	12.0	16.3	21.6	29.8	41.2	75.7	101.8
4517874	-0.8	-0.6	-0.1	0.4	0.7	0.9	1.0	1.2	1.7	2.4	3.6	5.1	8.9	9.8
4517876	-0.9	-0.8	-0.4	0.3	0.7	1.0	1.3	1.6	2.6	3.7	5.7	8.6	17.8	18.3
4517878	-0.4	0.3	1.5	3.1	4.0	4.7	5.3	5.9	7.9	10.5	14.6	20.1	35.8	44.7
4517882	-0.7	-0.4	0.5	1.7	2.4	3.0	3.5	3.9	5.6	7.7	11.1	15.8	30.1	34.7
4517890	0.3	1.6	4.6	8.3	10.4	12.6	14.2	15.8	21.7	28.8	39.9	55.4	103.7	140.5
4517900	-0.7	-0.4	0.1	0.7	1.0	1.2	1.4	1.6	2.2	3.0	4.3	6.0	10.3	11.5
4517904	-0.6	0.0	1.4	3.4	4.6	5.7	6.6	7.4	10.5	14.3	20.3	28.9	56.1	69.2
4517914	-0.9	-0.9	-0.8	-0.7	-0.7	-0.7	-0.7	-0.6	-0.6	-0.6	-0.4	-0.3	-0.1	-0.2
4517920	-0.9	-0.8	-0.5	-0.3	-0.2	-0.2	-0.1	-0.1	0.1	0.3	0.8	1.3	2.4	2.4
4517930	0.5	2.0	5.1	9.2	11.7	14.2	16.0	17.9	24.7	33.1	45.7	63.5	119.7	168.0
4517932	-0.5	0.0	1.0	2.3	3.1	3.7	4.3	4.7	6.5	8.7	12.2	16.9	30.5	37.3
4517936	-0.9	-0.8	-0.6	-0.4	-0.3	-0.3	-0.2	-0.2	0.0	0.2	0.6	1.1	2.3	2.2

Table E.5, ctd.

COMID	Q_{99} (ft^3/s)	Q_{90} (ft^3/s)	Q_{70} (ft^3/s)	Q_{50} (ft^3/s)	Q_{40} (ft^3/s)	Q_{30} (ft^3/s)	Q_{25} (ft^3/s)	Q_{20} (ft^3/s)	Q_{10} (ft^3/s)	Q_{5} (ft^3/s)	$Q_{2.5}$ (ft^3/s)	Q_{1} (ft^3/s)	$Q_{0.1}$ (ft^3/s)	$Q_{0.01}$ (ft^3/s)
4517940	-0.9	-0.8	-0.8	-0.7	-0.7	-0.7	-0.6	-0.6	-0.6	-0.5	-0.4	-0.3	0.0	-0.1
4517944	-0.9	-0.9	-0.7	-0.6	-0.6	-0.5	-0.5	-0.5	-0.4	-0.3	-0.1	0.2	0.7	0.6
4517956	-0.8	-0.7	-0.4	-0.1	0.0	0.1	0.2	0.3	0.5	0.8	1.4	2.1	3.8	3.9
4517958	-0.8	-0.7	-0.4	0.0	0.1	0.3	0.4	0.4	0.7	1.2	1.9	2.7	4.9	5.2
4517960	-0.9	-0.7	-0.5	-0.1	0.0	0.1	0.2	0.3	0.6	1.0	1.7	2.6	4.7	4.9
4517964	-0.9	-0.7	-0.5	-0.2	-0.1	0.0	0.1	0.1	0.3	0.6	1.2	1.8	3.3	3.4
4517966	-0.7	-0.4	0.2	1.0	1.4	1.7	2.0	2.2	3.0	4.1	5.9	8.1	14.2	16.4
4517968	-0.9	-0.8	-0.6	-0.3	-0.2	-0.1	-0.1	0.0	0.2	0.5	1.0	1.6	3.1	3.1
4517970	-0.9	-0.8	-0.6	-0.5	-0.4	-0.3	-0.3	-0.2	-0.1	0.1	0.5	0.9	1.9	1.8
4517982	0.9	3.0	7.3	13.3	16.9	20.8	23.6	26.6	37.3	50.3	69.6	97.8	189.6	277.7
4517984	-0.8	-0.6	-0.2	0.2	0.5	0.6	0.8	0.9	1.4	2.0	3.0	4.2	7.5	8.1
4517990	-0.9	-0.7	-0.5	-0.3	-0.2	-0.1	-0.1	0.0	0.2	0.4	0.9	1.4	2.7	2.7
4518000	-0.8	-0.7	-0.4	-0.1	0.1	0.2	0.3	0.3	0.6	1.0	1.6	2.4	4.3	4.3
4518002	0.9	3.1	8.1	15.0	19.2	23.8	27.0	30.6	43.3	58.8	81.6	116.0	229.7	332.4
4518004	-0.5	0.1	1.7	3.9	5.2	6.4	7.3	8.3	11.6	15.7	22.2	31.6	61.0	75.2
4518010	-0.6	-0.2	1.1	3.0	4.1	5.1	5.9	6.7	9.5	13.0	18.5	26.5	51.8	61.8
4518018	-0.9	-0.9	-0.8	-0.6	-0.5	-0.5	-0.4	-0.4	-0.2	-0.1	0.3	0.7	1.7	1.4
4518026	-0.9	-0.8	-0.7	-0.6	-0.5	-0.5	-0.4	-0.4	-0.3	-0.2	0.1	0.3	0.9	0.8
4518032	-0.5	0.0	1.0	2.3	3.0	3.6	4.1	4.6	6.2	8.3	11.6	15.9	28.4	34.9
4518034	-0.9	-0.8	-0.6	-0.3	-0.2	-0.1	-0.1	0.0	0.2	0.5	1.0	1.6	3.1	3.0
4518036	-0.9	-0.9	-0.7	-0.6	-0.5	-0.5	-0.4	-0.4	-0.3	-0.1	0.2	0.6	1.5	1.3
4518038	1.3	3.9	10.0	18.8	24.4	30.7	35.0	40.0	57.7	79.2	110.3	158.1	321.8	486.7
4518040	-0.7	-0.4	0.1	0.6	0.9	1.1	1.3	1.5	2.1	2.8	4.1	5.7	9.8	11.0
4518044	0.6	2.4	6.6	12.6	16.3	20.2	23.0	26.2	37.1	50.5	70.4	100.0	198.4	285.6
4518046	0.7	2.7	7.3	13.8	17.8	22.2	25.3	28.7	40.8	55.6	77.4	110.0	219.0	318.8

Table E.5, ctd.

COMID	Q_{99} (ft^3/s)	Q_{90} (ft^3/s)	Q_{70} (ft^3/s)	Q_{50} (ft^3/s)	Q_{40} (ft^3/s)	Q_{30} (ft^3/s)	Q_{25} (ft^3/s)	Q_{20} (ft^3/s)	Q_{10} (ft^3/s)	Q_{5} (ft^3/s)	$Q_{2.5}$ (ft^3/s)	Q_{1} (ft^3/s)	$Q_{0.1}$ (ft^3/s)	$Q_{0.01}$ (ft^3/s)
4518050	0.8	2.8	7.7	14.5	18.8	23.5	26.7	30.4	43.4	59.3	82.6	117.7	235.7	345.0
4518052	1.3	3.9	10.1	19.0	24.6	30.9	35.3	40.4	58.2	79.9	111.3	159.4	324.4	491.4
4518054	1.2	3.5	8.5	15.3	19.5	24.0	27.2	30.7	43.2	58.4	80.8	113.5	221.3	330.1
4518062	1.2	3.7	9.5	18.0	23.4	29.4	33.5	38.3	55.1	75.7	105.5	151.1	307.0	461.8
4518064	-0.6	-0.1	0.9	2.2	3.0	3.7	4.3	4.8	6.7	9.0	12.8	17.8	33.0	40.3
4518066	-0.4	0.1	1.4	3.0	3.9	4.8	5.4	6.1	8.4	11.2	15.7	21.7	39.7	50.0
4518068	-0.9	-0.8	-0.7	-0.5	-0.5	-0.5	-0.4	-0.4	-0.3	-0.2	0.1	0.4	1.0	0.9
4518076	-0.9	-0.8	-0.6	-0.4	-0.3	-0.2	-0.2	-0.1	0.0	0.3	0.7	1.2	2.4	2.3
4518092	3.8	9.1	21.5	39.8	51.8	66.1	75.9	87.7	129.2	180.1	250.7	363.1	766.2	1278.5
4518094	3.8	9.1	21.5	39.9	51.9	66.3	76.2	88.0	129.7	180.8	251.8	364.7	770.0	1285.2
4518096	-0.7	-0.5	-0.1	0.3	0.6	0.7	0.9	1.0	1.4	2.0	3.0	4.2	7.2	7.9
4518098	-0.3	0.4	2.0	4.0	5.2	6.3	7.1	8.0	10.9	14.6	20.3	28.1	51.8	67.1
4518100	-0.8	-0.7	-0.5	-0.2	-0.1	-0.1	0.0	0.1	0.2	0.5	1.0	1.5	2.8	2.9
4518104	3.4	8.3	19.7	36.5	47.4	60.4	69.3	79.9	117.4	163.2	227.2	328.6	690.0	1137.7
4518106	-0.3	0.4	2.0	4.0	5.2	6.3	7.1	8.0	10.9	14.6	20.3	28.1	51.8	67.1
4518108	-0.9	-0.8	-0.6	-0.5	-0.5	-0.4	-0.4	-0.4	-0.3	-0.2	0.1	0.3	0.8	0.8
4518110	12.9	28.5	65.2	121.2	159.7	210.2	243.9	287.2	441.1	631.1	881.9	1307.1	2941.5	5629.3
4518124	4.1	9.7	22.8	42.5	55.3	70.8	81.4	94.2	139.3	194.5	271.0	393.4	834.7	1404.4
4518126	-0.9	-0.9	-0.8	-0.7	-0.7	-0.7	-0.7	-0.7	-0.6	-0.6	-0.4	-0.3	-0.1	-0.2
4518128	4.1	9.7	22.8	42.5	55.3	70.8	81.4	94.2	139.3	194.5	271.0	393.4	834.7	1404.4
4518130	-0.9	-0.8	-0.7	-0.5	-0.4	-0.3	-0.3	-0.2	-0.1	0.1	0.5	0.9	1.9	1.9
4518132	-0.7	-0.4	0.2	0.9	1.4	1.7	2.0	2.2	3.1	4.3	6.2	8.6	15.5	17.9
4518136	-1.0	-0.9	-0.8	-0.6	-0.5	-0.4	-0.4	-0.3	-0.1	0.1	0.5	1.0	2.3	2.1
4518140	-0.7	-0.4	0.2	0.9	1.3	1.6	1.8	2.0	2.9	3.9	5.6	7.8	13.8	15.9
4518146	-0.9	-0.8	-0.6	-0.3	-0.2	-0.1	0.0	0.0	0.2	0.5	1.0	1.6	3.1	3.1

Table E.5, ctd.

COMID	Q_{99} (ft^3/s)	Q_{90} (ft^3/s)	Q_{70} (ft^3/s)	Q_{50} (ft^3/s)	Q_{40} (ft^3/s)	Q_{30} (ft^3/s)	Q_{25} (ft^3/s)	Q_{20} (ft^3/s)	Q_{10} (ft^3/s)	Q_{5} (ft^3/s)	$Q_{2.5}$ (ft^3/s)	Q_{1} (ft^3/s)	$Q_{0.1}$ (ft^3/s)	$Q_{0.01}$ (ft^3/s)
4518148	-0.9	-0.8	-0.6	-0.4	-0.3	-0.2	-0.2	-0.1	0.0	0.2	0.7	1.1	2.2	2.2
4518150	-0.9	-0.8	-0.6	-0.5	-0.4	-0.3	-0.3	-0.2	-0.1	0.1	0.5	0.9	1.8	1.7
4518152	13.8	30.4	69.7	130.3	172.1	227.3	264.2	311.8	481.1	690.4	965.7	1435.0	3252.7	6289.5
4518164	-0.5	0.1	1.0	2.1	2.6	3.0	3.4	3.7	4.9	6.4	8.9	12.1	20.5	24.3
4518166	-0.9	-0.7	-0.5	-0.3	-0.2	-0.1	-0.1	0.0	0.2	0.5	0.9	1.5	2.8	2.9
4518168	-0.9	-0.9	-0.8	-0.7	-0.6	-0.6	-0.5	-0.5	-0.4	-0.3	-0.1	0.2	0.8	0.7
4518170	-0.7	-0.4	0.0	0.6	0.8	1.0	1.2	1.4	1.9	2.6	3.9	5.3	9.2	10.4
4518172	14.0	30.8	70.8	132.6	175.3	231.8	269.5	318.3	491.8	706.3	988.3	1469.6	3337.6	6470.5
4518174	-0.7	-0.5	0.1	0.7	1.1	1.3	1.6	1.8	2.5	3.4	5.0	6.9	12.3	14.1
4518188	8.7	19.3	43.6	79.1	102.8	132.7	152.8	178.1	266.7	375.7	522.4	763.2	1652.8	2989.2
4518192	-0.8	-0.6	-0.3	-0.1	0.1	0.1	0.2	0.3	0.5	0.8	1.3	2.0	3.4	3.5
4518194	14.1	31.2	71.6	134.3	177.7	235.1	273.4	323.0	499.6	717.9	1004.6	1494.7	3399.2	6602.3
4518196	0.1	1.3	3.9	7.2	9.1	10.9	12.3	13.8	18.8	25.1	34.7	48.3	90.0	120.3
4518200	-0.9	-0.9	-0.7	-0.5	-0.4	-0.3	-0.3	-0.2	0.0	0.2	0.6	1.2	2.6	2.4
4518202	-0.9	-0.8	-0.4	0.1	0.4	0.6	0.8	0.9	1.5	2.3	3.6	5.2	10.1	10.6
4518204	8.0	17.9	40.4	72.9	94.5	121.5	139.7	162.4	242.2	340.0	472.4	688.4	1479.8	2645.9
4518208	8.5	18.9	42.5	76.8	99.6	128.3	147.6	171.8	256.6	360.8	501.4	731.4	1577.3	2841.1
4518210	8.0	17.9	40.4	72.9	94.5	121.5	139.7	162.4	242.1	340.0	472.3	688.3	1479.6	2645.3
4518212	8.1	18.1	40.8	73.9	95.9	123.5	142.1	165.3	246.8	346.9	482.2	703.4	1516.0	2715.2
4518214	15.5	33.9	77.6	145.6	192.7	255.6	297.5	352.0	545.9	786.0	1100.0	1638.6	3742.0	7354.5
4518216	15.9	34.6	78.9	147.8	195.6	259.5	302.0	357.3	554.3	798.1	1116.8	1663.3	3798.3	7483.9
4518220	0.7	2.3	5.7	10.3	13.1	16.0	18.0	20.2	28.0	37.6	51.8	72.0	136.5	196.6
4518226	-0.2	0.6	2.1	4.1	5.2	6.2	7.0	7.7	10.4	13.7	19.0	26.1	46.9	60.3
4518228	-0.1	0.8	2.7	5.1	6.5	7.8	8.8	9.8	13.3	17.7	24.5	33.8	61.8	82.0
4518230	3.3	8.0	18.7	33.7	43.2	54.3	62.0	71.1	102.9	141.8	196.6	282.1	579.4	936.0

Table E.5, ctd.

COMID	Q_{99} (ft^3/s)	Q_{90} (ft^3/s)	Q_{70} (ft^3/s)	Q_{50} (ft^3/s)	Q_{40} (ft^3/s)	Q_{30} (ft^3/s)	Q_{25} (ft^3/s)	Q_{20} (ft^3/s)	Q_{10} (ft^3/s)	Q_{5} (ft^3/s)	$Q_{2.5}$ (ft^3/s)	Q_{1} (ft^3/s)	$Q_{0.1}$ (ft^3/s)	$Q_{0.01}$ (ft^3/s)
4518232	-0.7	-0.4	0.1	0.7	1.1	1.4	1.6	1.8	2.5	3.4	5.0	7.0	12.2	14.2
4518234	7.5	16.8	38.0	68.5	88.7	113.8	130.8	151.8	225.7	316.4	439.5	639.5	1369.2	2427.2
4518236	0.5	1.9	5.0	9.0	11.5	13.9	15.7	17.7	24.4	32.7	45.1	62.6	118.1	167.2
4518238	-0.7	-0.4	0.1	0.8	1.2	1.5	1.7	1.9	2.7	3.6	5.3	7.3	13.0	15.0
4518240	16.2	35.3	80.7	151.4	200.6	266.4	310.2	367.3	570.7	822.6	1151.4	1716.4	3929.4	7770.5
4518244	16.0	35.0	79.9	149.9	198.6	263.7	307.0	363.3	564.3	813.0	1138.0	1695.9	3879.5	7660.1
4518246	0.3	1.6	4.4	8.1	10.2	12.5	14.1	15.8	21.8	29.2	40.3	56.0	105.3	147.0
4518250	3.4	8.0	17.7	30.7	38.9	48.2	54.8	62.4	88.8	121.1	166.8	236.4	472.3	759.5
4518252	-0.9	-0.9	-0.7	-0.5	-0.4	-0.3	-0.2	-0.2	0.0	0.3	0.8	1.3	2.9	2.8
4518254	-0.8	-0.6	-0.2	0.2	0.4	0.6	0.7	0.8	1.2	1.8	2.7	3.8	6.7	7.4
4518258	-0.8	-0.7	-0.6	-0.4	-0.3	-0.3	-0.2	-0.2	-0.1	0.1	0.4	0.8	1.5	1.6
4518264	-0.5	0.0	0.7	1.5	1.9	2.3	2.5	2.8	3.7	4.8	6.8	9.1	15.3	18.3
4518266	-0.7	-0.5	0.0	0.5	0.7	0.9	1.0	1.1	1.6	2.2	3.2	4.4	7.5	8.3
4518270	-0.6	-0.3	0.2	0.8	1.1	1.4	1.6	1.7	2.3	3.1	4.5	6.1	10.1	12.0
4518272	16.4	35.8	81.7	153.5	203.5	270.5	315.1	373.2	580.4	837.1	1171.9	1747.7	4006.2	7939.7
4518276	-1.0	-0.9	-0.8	-0.7	-0.7	-0.7	-0.6	-0.6	-0.6	-0.5	-0.3	-0.1	0.4	0.2
4518278	0.0	1.1	3.2	6.0	7.6	9.2	10.4	11.6	15.9	21.2	29.3	40.5	75.0	101.3
4518280	-0.3	0.5	2.0	3.8	4.9	5.9	6.6	7.3	9.9	13.1	18.2	24.9	44.8	57.3
4518282	-0.7	-0.5	-0.4	-0.2	-0.2	-0.2	-0.1	-0.1	-0.1	0.1	0.4	0.7	1.2	1.3
4518284	0.5	2.1	5.5	9.9	12.5	15.1	17.1	19.2	26.5	35.4	48.9	68.1	128.9	180.8
4518286	-0.8	-0.6	-0.3	0.1	0.3	0.4	0.6	0.7	1.1	1.6	2.5	3.5	6.4	6.9
4518288	-0.1	0.8	2.4	4.3	5.4	6.3	7.1	7.8	10.3	13.5	18.6	25.2	44.1	56.9
4518290	-0.6	-0.1	0.7	1.6	2.1	2.5	2.8	3.1	4.1	5.5	7.8	10.6	18.3	21.5
4518292	-0.7	-0.5	-0.2	0.0	0.1	0.2	0.2	0.3	0.4	0.7	1.2	1.8	2.9	3.2
4518296	16.8	36.5	83.2	156.3	207.3	275.7	321.2	380.5	592.2	854.5	1196.3	1784.4	4093.9	8138.1

Table E.5, ctd.

COMID	Q_{99} (ft^3/s)	Q_{90} (ft^3/s)	Q_{70} (ft^3/s)	Q_{50} (ft^3/s)	Q_{40} (ft^3/s)	Q_{30} (ft^3/s)	Q_{25} (ft^3/s)	Q_{20} (ft^3/s)	Q_{10} (ft^3/s)	Q_{5} (ft^3/s)	$Q_{2.5}$ (ft^3/s)	Q_{1} (ft^3/s)	$Q_{0.1}$ (ft^3/s)	$Q_{0.01}$ (ft^3/s)
4518298	-0.8	-0.6	-0.4	-0.1	0.0	0.1	0.1	0.2	0.4	0.7	1.2	1.8	3.2	3.4
4518302	-0.8	-0.6	-0.1	0.5	0.8	1.1	1.3	1.5	2.2	3.1	4.6	6.6	12.0	13.1
4518306	17.1	37.2	84.8	159.5	211.7	281.9	328.5	389.4	606.8	876.2	1227.0	1831.4	4209.5	8395.6
4518312	-0.4	0.3	2.0	4.2	5.4	6.6	7.5	8.4	11.7	15.6	21.9	30.6	57.5	73.0
4518318	-0.8	-0.5	0.0	0.6	0.9	1.1	1.3	1.5	2.1	2.9	4.3	5.9	10.5	11.5
4518672	-0.8	-0.6	-0.3	0.1	0.2	0.3	0.4	0.5	0.8	1.2	1.9	2.7	4.8	5.0
4518676	0.1	1.3	4.4	9.0	11.8	14.7	16.8	19.1	27.2	37.1	52.0	74.3	148.1	203.7
4518680	4.3	10.1	23.8	44.3	57.7	74.1	85.2	98.7	146.2	204.5	285.1	414.4	882.5	1493.0
4518682	-0.9	-0.8	-0.6	-0.3	-0.2	-0.1	0.0	0.0	0.3	0.5	1.1	1.7	3.3	3.3
4518684	13.4	29.5	67.7	126.2	166.5	219.5	254.9	300.6	462.8	663.1	927.1	1376.0	3108.3	5982.6
4518688	1.6	4.6	11.8	22.1	28.5	35.7	40.7	46.5	67.0	92.0	128.1	184.0	375.5	570.6
4518690	1.6	4.7	11.9	22.2	28.6	35.8	40.9	46.7	67.3	92.4	128.6	184.6	376.6	572.7
4518692	13.8	30.5	70.0	131.0	173.1	228.8	265.9	313.9	484.7	695.8	973.4	1447.1	3283.1	6352.9
4518694	1.5	4.0	8.9	15.0	18.7	22.5	25.2	28.2	38.7	51.5	70.5	97.6	182.9	269.4
4518696	-0.7	-0.4	0.2	1.2	1.7	2.1	2.4	2.7	3.9	5.3	7.6	10.7	19.7	22.8
4518698	-0.4	0.1	1.0	1.9	2.4	2.8	3.1	3.4	4.4	5.8	8.1	10.8	18.1	22.2
4518764	-0.3	0.4	1.4	2.4	2.9	3.2	3.5	3.8	4.9	6.2	8.6	11.4	18.4	22.2
4518800	-0.4	0.1	1.1	2.2	2.7	3.2	3.5	3.9	5.1	6.6	9.3	12.5	21.2	25.3
4518842	-0.8	-0.6	-0.3	-0.1	0.1	0.2	0.2	0.3	0.5	0.9	1.5	2.1	3.7	3.9

Table E.6. Validation gauge watershed characteristics.

SITENO	DA (km^2)	HSG	PCT_FOR (%)	PCT_DEV (%)	PCT_AG (%)	ELEV_MIN (m)	ELEV_RNGE (m)	SLOPE_MEAN (%)	2yr24MAX (in*1000)	2yr24hrMEAN (in*1000)	100yr24MAX (in*1000)	100yr24MEAN (in*1000)	HSG*100max (in*1000)	HSG*100mean (in*1000)	HSG*2yrmax (in*1000)	HSG*2yrmean (in*1000)
1470779	181.6	2.2	10.9	16.2	72.1	92.2	324.3	5.5	3157	3116.2	7580	7478.7	17029.4	16801.9	7092.6	7000.9
1554500	141.2	3.6	76.1	17.2	1.1	183.6	370.5	15.7	3126	2999.2	7705	7333.4	28085.7	26731.0	11394.7	10932.4
1555500	419.8	2.7	51.9	6.9	40.3	122.8	428.3	15.0	3262	3032.2	8035	7388.4	21997.3	20227.0	8930.3	8301.3
1563000	22.6	3.0	76.9	5.2	6.6	238.4	175.0	18.1	2676	2670.5	6006	5993.0	18018.0	17979.1	8028.0	8011.4
1570500	62339.5	2.7	65.7	6.6	23.0	83.3	871.0	14.9	3516	2688.5	8804	6145.4	23516.9	16415.5	9391.8	7181.4
1614500	1300.0	2.3	38.4	12.4	47.8	118.0	630.3	10.0	3357	2953.8	8322	6872.6	19230.2	15881.0	7757.3	6825.6