Watershed modeling for Evaluation of the Effectiveness of Conservation Practices in Agricultural Watersheds in Tennessee: Pilot Study

Final Report

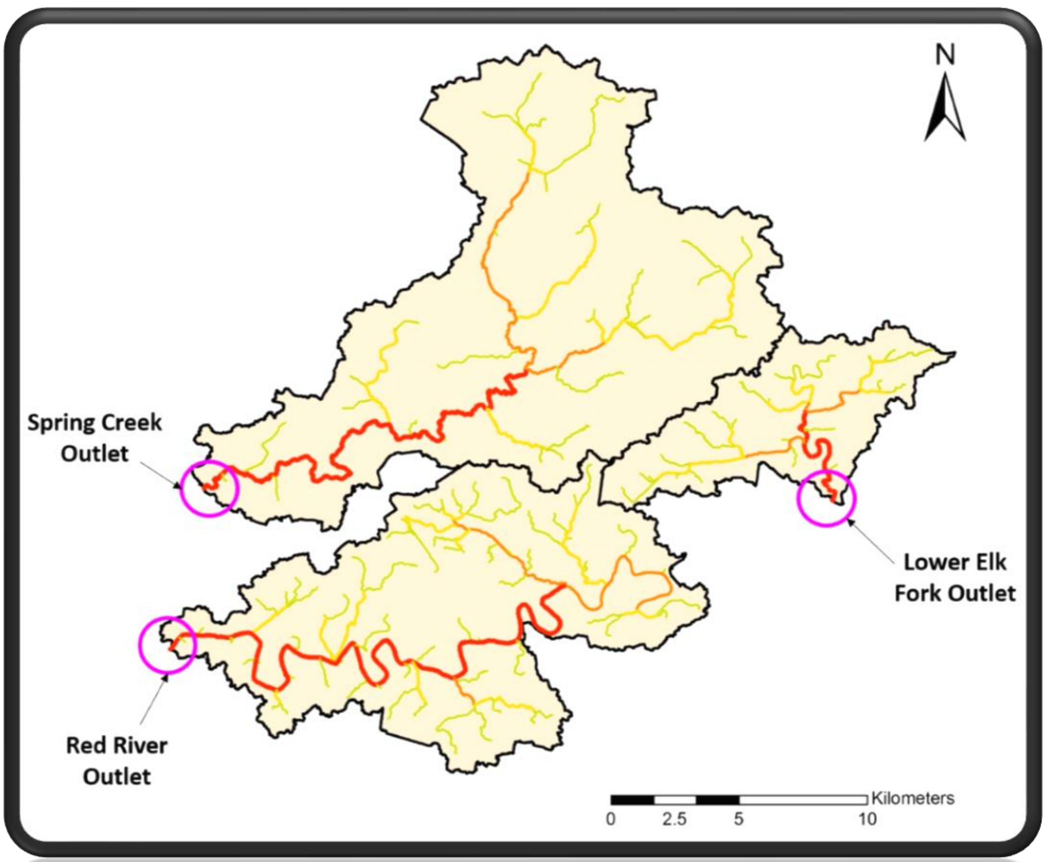


Table of Contents

List of Figures

List of Tables

1 Introduction

This project includes two main investigations: (1) **Watershed modeling:** The Annualized AGricultural Non-Point Source (AnnAGNPS) watershed pollution model was used to estimate long-term averages of non-point source pollution at watershed scale as an initial spatial screening tool for spatially identifying candidate locations for conservation practices. The findings from this pilot study support the long-term NRCS's goals by generating information to complement existing NRCS efforts in guiding future conservation strategies development, enhancements, and expansion. The final generated products include a quantitative estimate and assessment of the effectiveness of various conservation practices relatively to the status quo. (2) **GIS tool for riparian vegetation delineation:** LiDAR and GIS were used to derive the spatial distribution of the existing riparian buffers in the investigated watersheds. The final generated products include a riparian buffer identification tool/procedure that can be applied in other watersheds outside the scope of the project and riparian buffer maps for the investigated watersheds.

This report describes the methodology that have been applied to achieve these two goals as well as the results of the investigation. Section 2 describes the studied watersheds. Section 3 describes the watershed modeling datasets and steps performed. Section 4 describes the conservation practices simulations and results. Section 5 describes the approach used to build the riparian buffer tool. Section 6 is an executive summary of the entire methodology (sections 1-5). Appendices 1 and 2 are an attachment to section 3, appendix 1 showcases examples of the hydrological correction that was performed on the digital elevation models of the study areas, and appendix 2 displays the same modeling steps over system 1 since we used system 2 to describe the watershed modeling methodology in detail under section 3. Appendix 3 is the generated tutorial that can be used to derive the riparian buffer of any watershed from LiDAR data. Appendix 4 lists the generated datasets accompanying the report.

2 Study Areas

Two watershed systems are investigated to understand and evaluate the effectiveness of conservation practices in Tennessee. System 1 includes four watersheds in Northern Middle Tennessee and are all part of the 2019 National Water Quality Initiative (NWQI). System 2 includes six watersheds in West Tennessee and are all part of the Mississippi River Basin Initiative (MRBI) (Figure 1).

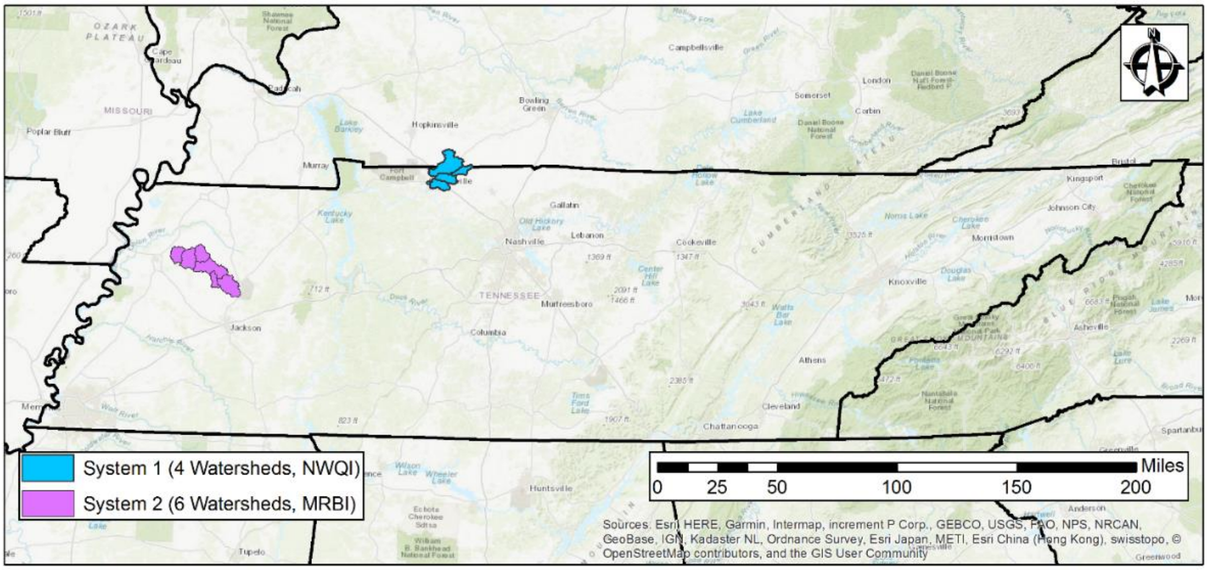


Figure 1: Watershed systems selected for this study.

2.1) System 1:

System 1 consists of three subsystems: the Lower Elk Fork, the Spring Creek (both of which have single HUC 12 subdivisions), and the Red River (which is divided into two HUC 12 subdivisions) (Figure 2). This system encompasses a karst region with exceptional geological features including a multitude of sink holes and cave systems and supports a significant black bass fishery and unique wildlife. Due to the karst topography many tributaries disappear underground and reappear at lower elevations as springs or glades.

2.1.1) Lower Elk Fork

The Lower Elk Fork subsystem is located in Robertson and Montgomery Counties in Tennessee and Logan and Todd Counties in Kentucky. Approximately 17 km of the Elk Fork Creek runs through this sub-watershed and the total area of this subsystem is 12,242 acres. Predominant land uses within this subsystem include agricultural cropland (70%), forest (17%), pastures (7%), and developed land (4%). Elevation ranges from 406.79 to 618.62 meters above mean sea level with an average of 556.39 meters above mean sea level. The slope ranges from 0 to 84.7 degrees with an average of 9.2 degrees. Soil in the area is mostly comprised of 78 classes with silt loam being the most dominant texture.

2.1.2) Spring Creek

The Spring Creek subsystem is located in both Montgomery County, Tennessee and Todd County, Kentucky. Approximately 36 km of the Spring Creek runs through this sub-watershed, and the total area of this subsystem is 51,466 acres. Predominant land uses within this subsystem include agricultural cropland (60%), forest (17%), pastures (8%), and developed land (13%). Elevation ranges from 368.82 to 658.77 meters above sea level with an average of 558.79 meters above sea level. The slope ranges from 0 to 84.7 degrees with an average of 7.6 degrees. Soils for the area are comprised of 63 classes with silt loam being the most dominant texture.

2.1.3) Red River

The Red River subsystem consists of two HUC 12 divisions (City of Kirkwood-Red River and Dunbar Lake-Red River) which are located in Montgomery County in Tennessee. The total acreage of these hydrological units are 18,196 and 18,154 acres respectively.

For the City of Kirkwood-Red River watershed, the predominant land uses within this subsystem include agricultural cropland (46%), forest (18%), and pastures (13%). For the Dunbar Lake-Red River watershed, the predominant land uses are agricultural cropland (14%), forest (33%), pastures (10%), and developed land (40%).

Elevation ranges from 140.20 to 657.07 meters above mean sea level with an average of 514.81 meters above mean sea level. The slope ranges from 0 to 88.53 degrees with an average of 13.95 degrees. Soil for the area is comprised of 66 classes with silt loam being the most dominant soil texture.

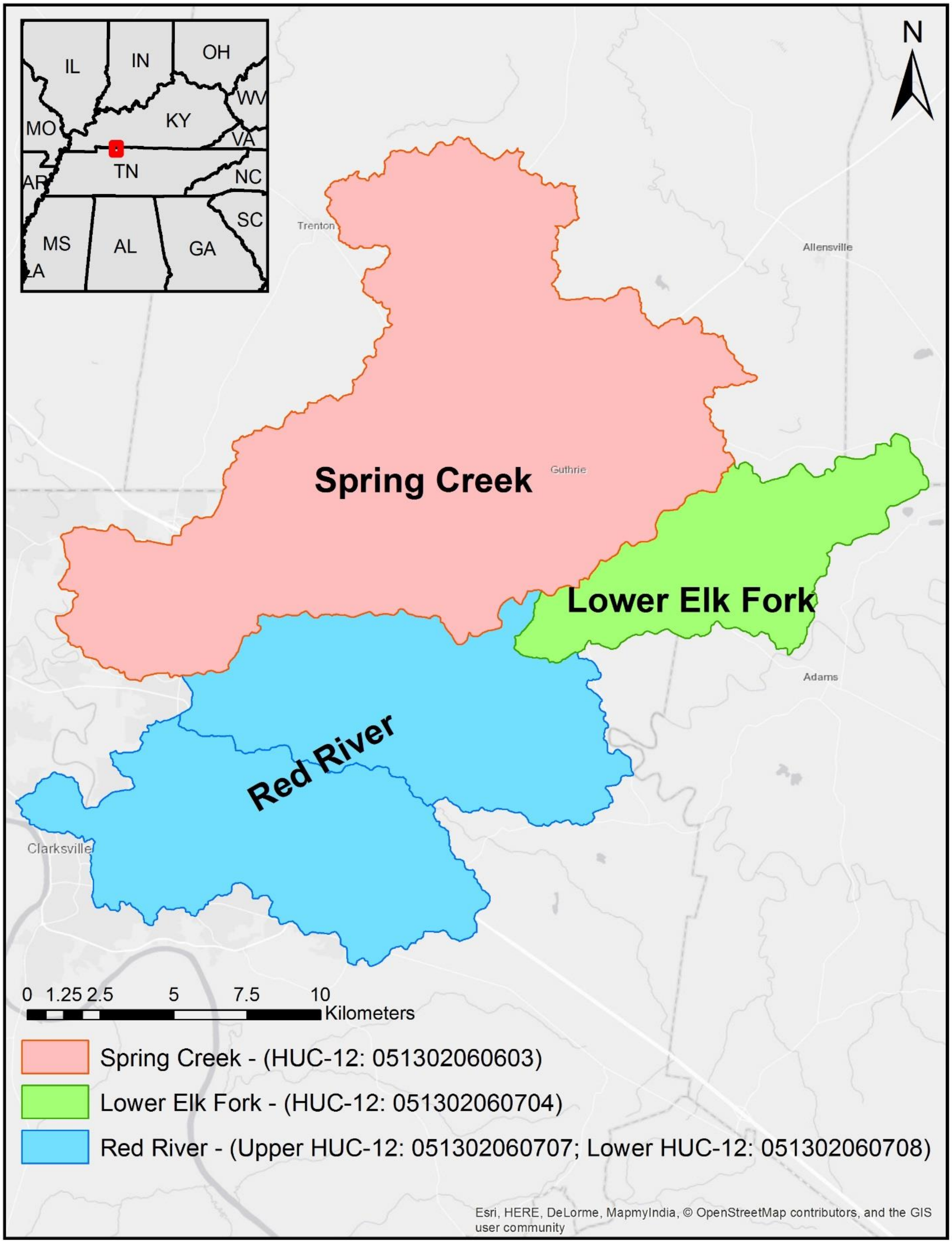


Figure 2: System 1 hydrological units.

2.2) System 2:

System 2 is located in Gibson and Dyer Counties in West Tennessee and consists of six HUC-12 (Figure 3). The total area is 156,000 acres and it drains to the Forked Deer River.

2.2.1) North Fork Forked Deer River Upper

The North Fork Forked Deer River upper, is the southeastern-most HUC 12 division and is located in Gibson County in western Tennessee. Its total area covers 36,522 areas, and is the second largest watershed in the system. Predominant land uses within this subsystem include agricultural cropland (38%), forest (36%), pastures (19%), and developed land (7%). Elevation ranges from 95.90 to 172.27 meters above mean sea level with an average of 123.22 meters above mean sea level. The slope ranges from 0 to 71.9 degrees with an average of 3.9 degrees. Soil for the area is comprised of 23 classes with silt loam being the most dominant soil texture.

2.2.2) Cain Creek

Cain Creek, is in the southern region of the system and is the smallest of the HUC 12 divisions with a total area of 10,764 acres, and it is located in Gibson County in western Tennessee. Predominant land uses within this subsystem include agricultural cropland (59%), forest (18%), pastures (10%), and developed land (13%). Elevation ranges from 92.9 to 148.1 meters above sea level with an average of 115.18 meters above mean sea level. The slope ranges from 0 to 52.5 degrees with an average of 3.4 degrees. Soil for the area is comprised of 27 classes with silt loam being the most dominant soil texture.

2.2.3) Mud Creek

Mud Creek, is the northern-most HUC 12 division and is located in Gibson County in western Tennessee. Its total area covers 20,916 acres, and it is the third smallest watershed in the system. Predominant land uses within this subsystem include agricultural cropland (78%), forest (11%), pastures (4%), and developed land (7%). Elevation ranges from 85.1 to 129.8 meters above sea level with an average of 129.8 meters above sea level. The slope ranges from 0 to 61.3 degrees with an average of 2.8 degrees. Soil for the area is comprised of 23 classes with silt loam being the most dominant soil texture.

2.2.4) North Fork Forked Deer River Middle

The North Fork Forked Deer River middle, is the largest of the HUC-12 divisions with a total area of 39,815 acres, and it is located in Gibson County in western Tennessee. Predominant land uses within this subsystem include agricultural cropland (51%), forest (33%), pastures (8%), and developed land (7%). These land cover characteristics were determined from the United States Department of Agriculture (USDA)'s National Agricultural Statistics Service (NASS) land cover/land use database for 2018. Elevation ranges from 85.0 to 160.4 meters above mean sea level with an average of 105.2 meters above mean sea level. The slope ranges from 0 to 69.1 degrees with an average of 3.2 degrees. Soil for the area is comprised of 28 classes with silt loam being the most dominant soil texture.

2.2.5) Doakville Creek

The Doakville Creek is the western-most HUC 12 division and is located in Dyer County in western Tennessee. Its total area covers 19,149 acres, and is the second smallest watershed in the system. Predominant land uses within this subsystem include agricultural cropland (80%), forest (7%), pastures (7%), and developed land (6%). Elevation ranges from 83.0 to 122.5 meters above mean sea level with an average of 98.8 meters above sea level. The slope ranges from 0 to 41.7 degrees with an average of 2.4 degrees. Soil for the area is comprised of 19 classes with silt loam being the most dominant soil texture.

2.2.6) North Fork Forked Deer River Lower

The North Fork Forked Deer River lower and is located in both Gibson and Dyer Counties in western Tennessee. Its total area covers 28,512 acres and it is the third largest watershed in the system. Land uses within this subsystem include agricultural cropland (73%), forest (19%), pastures (4%), and developed land (4%). Elevation ranges from 83.0 to 126.6 meters above mean sea level with an average of 95.8 meters above mean sea level. The slope ranges from 0 to 44.2 degrees with an average of 2.2 degrees. Soil for the area is comprised of 37 classes with silt loam being the most dominant soil texture.

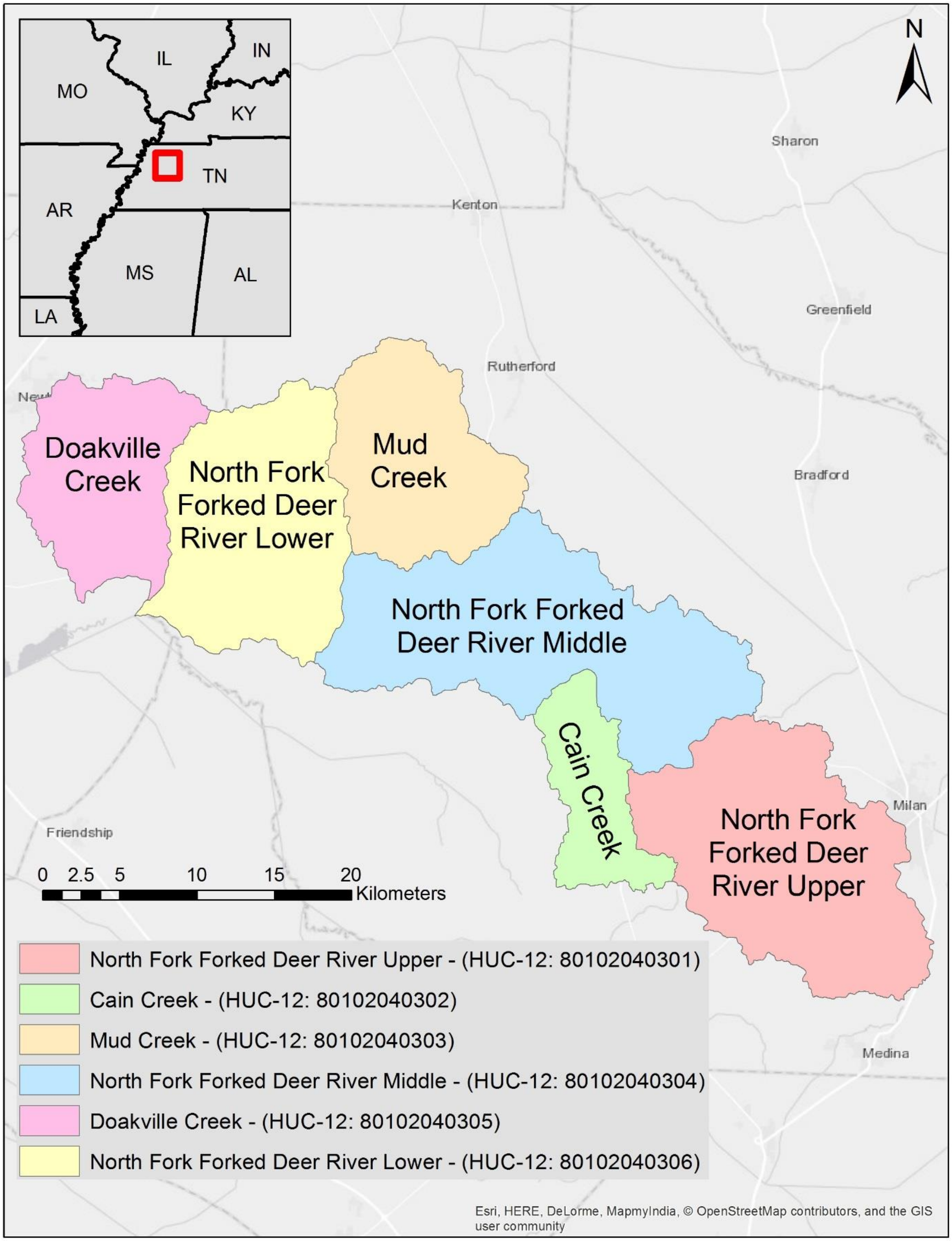


Figure 3: System 2 hydrological units.

3 Watershed Modeling

Identification and quantification of all different sources and sinks of non-point source pollution in agricultural watersheds is a difficult task. This is due to their spatiotemporal dynamic nature and many complex and interconnected driving forces, including natural (e.g. weather, climate, soils) and anthropogenic (e.g. farming and conservation practices) factors. This process becomes more complicated to decipher at large scales and over long periods of time. Alternatives for addressing such complexities include the use of watershed modeling technology. The latter allows the simulation of hydrological processes in a holistic approach while accounting for their temporal and spatial variation and their interrelationships. Watershed models allow the creation of simulations describing existing conditions and alternative scenarios as well, which is a convenient and cost-effective approach to obtain the response of a watershed to a multitude of physical and anthropogenic variations. In this study, we use the Annualized AGricultural Non-Point Source (AnnAGNPS) watershed pollution model to construct watershed simulations for our study areas. We use System 2 in this section to describe the performed operations and tasks, but the same procedures were applied on System 1 (Appendix 2).

3.1 The AnnAGNPS Watershed Pollution Model

The AnnAGNPS watershed pollution model is the result of a partnership between two branches of the U.S. Department of Agriculture (USDA), the Agricultural Research Service (research branch) and the Natural Resource Conservation Service (action branch). The AnnAGNPS model was designed and implemented to quantitatively assess the long-term impact of farming practices to the generation and transport of non-point source pollutants in un-gauged agriculture-dominated watersheds. The model contains procedures to account for the complex spatiotemporal interactions between sinks and sources allowing for the evaluation of best management practices, not only locally where the conservation practices are implemented, but also their inter-related effect to the overall watershed pollutant loads.

The watershed is represented by two basic modeling units: concentrated surface flow paths (referred to as reaches) and sub-catchments (referred to as AnnAGNPS cells). AnnAGNPS cells are hierarchically connected by reaches depicting how surface and shallow subsurface flow throughout the watershed. Runoff, sediment, and pollutants transported out from cells are routed through the reaches, while accounting for the inter-relation between upland and channel hydrological and pollutant transport processes.

The AnnAGNPS model performs long-term continuous simulations of mixed-land use watersheds on a daily time step to model farming management practice impacts on runoff and sediment/nutrient/pesticide detachment, transportation, and deposition. The hydrology of the watershed is based on a daily water balance considering surface runoff, evapotranspiration (ET), and percolation of water through the soil profile. Detachment, transportation, and deposition of sediment and attached and dissolved chemicals are determined using an integrated approach. Landscape erosion processes are calculated using the Revised Universal Soil Loss Equation (RUSLE) for estimation of sheet and rill erosion while accounting for land cover and farming management practices. The delivery of multiple particle sizes of eroded sediments to concentrated flow is calculated using the Hydro-geomorphic Universal Soil Loss Equation (HUSLE).

An important characteristic of the AnnAGNPS model is the capability of describing farming practices on a daily basis. This is accomplished by the development of databases representing unique farming management schedules and respective farming operations. These databases provide the means to estimate crop growth, canopy cover, root mass, fall height, and soil disturbance by farming equipment. Additionally, the AnnAGNPS model contains components to characterize ephemeral gully emergence, evolution, and erosion rates at scales smaller than sub-catchments, and components to estimate the effects of conservation practices such as riparian vegetated filters and sediment retention wetlands.

3.2 Watershed Characterization

3.2.1 Topography

Determination of how surface water flows through the landscape is of vital importance to quantitative estimation of soil erosion, transport, and deposition throughout the watershed. Topographic features determine how water will concentrate and flow through the field. This analysis was performed using standard flow routing algorithms applied to topographic information represented as raster grid file format, referred to as Digital Elevation Model (DEM). The majority of the off-the-shelve commercial GIS software packages contain the necessary functions to perform such analyses. However, in this study, the AnnAGNPS GIS component (TopAGNPS) was used because it streamlines most of the GIS steps and also generates the two topographic input data sections needed by the AnnAGNPS watershed pollution model.

3.2.1.1 The TopAGNPS GIS Tool

The TopAGNPS GIS software package is a subset of the topographic parameterization (TOPAZ) computer program. The TopAGNPS software package uses digital elevation models in raster grid format to identify and measure topographic features, define surface drainage spatial extent, and channel network pattern in order to support watershed hydrological modeling and analysis. This software package contains necessary GIS functions needed for hydrological analysis and commonly found in standard GIS software packages, such as filtering, re-sampling, pit filling, flow direction, flow accumulation, distance to channel, distance to watershed outlet, and watershed divide delineation. Additionally, TopAGNPS contains functions to calculate LS-factor, subdivide the watershed into reaches and AnnAGNPS cells, and generate the AnnAGNPS reach and cell input data sections in which characterizes each cell and reach topographically.

Two user-provided parameters control how the watershed will be sub-divided into AnnAGNPS cells and reaches, the critical source area (CSA) and the maximum source channel length (MSCL). The CSA is defined as the drainage area in hectares at the bottom of a flow path where a first-order channel begins. The MSCL is defined as the threshold flow path length in meters for a source channel to be considered as a channel. Larger CSA and MSCL values will result in lower drainage density and higher generalization of watershed features. Conversely, a smaller value of the CSA and MSCL results in generation on a high drainage density.

3.2.1.2 Data Collection and Processing

Topographic information was obtained from available public data repositories. Topographic datasets for the state of Tennessee were collected from the Tennessee Department of Finance and Administration (https://www.tn.gov/finance/sts-gis/gis/data.html) and for the state of Kentucky from the Kentucky Division for Geographic Information (http://kymartian.ky.gov). These datasets were provided as 1-km tilled raster grids (DEMs) and at different spatial resolution, 1-m for Tennessee and 1.5-m for Kentucky. A total of 290 tiles were used for System 1 and 814 tiles were used for System 2 (Figure 4). Additionally, topographic information used in this study included several surveys (varying time periods and vendors). Tilled information was merged into continuous datasets covering each studied system. Additionally, all datasets were resampled into 3-m spatial resolution.

Datasets were merged to cover an area larger than the USGS HUC12 boundaries. The initial study area encompassed a 500-m buffer zone of the USGS HUC12 boundaries to assure the watershed delineation based on flow routing algorithms would not be influenced by the HUC12 boundaries. Additionally, all datasets were re-projected to NAD83 UTM16N coordinate system.

Elevation for each system were initially processed using the TopAGNPS software package in reduced mode to generate intermediate GIS layers (i.e. filtering, pit filling, flow direction, and flow accumulation) to aid in the determination of the watershed outlet. Information generated by TopAGNPS was contrasted with high-resolution aerial photographs and HUC12 boundaries to assure proper outlet selection.

Once the watershed outlet was determined, TopAGNPS software package was re-used in full mode to delineate the watershed boundaries and to generate the channel network (reaches) and sub-catchments (AnnAGNPS cells) (Figure 4). Upon visual inspection of the generated channel network, it was observed that the DEM datasets obtained from state agencies were only partially hydrologically enforced. Only major man-made structures were artificially removed. This limitation forced us to perform hydrologically enforcing procedures. An iterative approach was applied (Figure 5). Datasets generated by the TopAGNPS computer program were visually analyzed and compared to high-resolution imagery and auxiliary GIS layers to determine whether manmade obstructions would hinder the flow routing algorithm causing these structures to work as pseudo-damns resulting in incorrect surface flow network and/or increase ponding beyond normal levels (Figure 6 and Appendix 1). In areas where it was challenging to assess the proper water routes, field visits were conducted to determine the locations of culverts, routing channels, and sinkholes that may be affecting the water flow (Figure 7).

A custom computer program (written in Python programming language) has been developed to modify user-selected regions in the DEM to enforce surface flow. Using standard GIS software package, the user identifies locations where man-made structures should be breached to allow water to flow by drawing a line connecting the two low points on either side of the obstruction. The modified DEM is then processed once again in TopAGNPS and a new set of channel network and watershed boundary layers are produced. The new channel network raster is then evaluated and, if additional corrections are needed to be made, then this process is repeated. Once the flow of the channel network is modeled as accurately as possible, the outputs are preserved for processing subsequent operations (Figures 8 and 9).

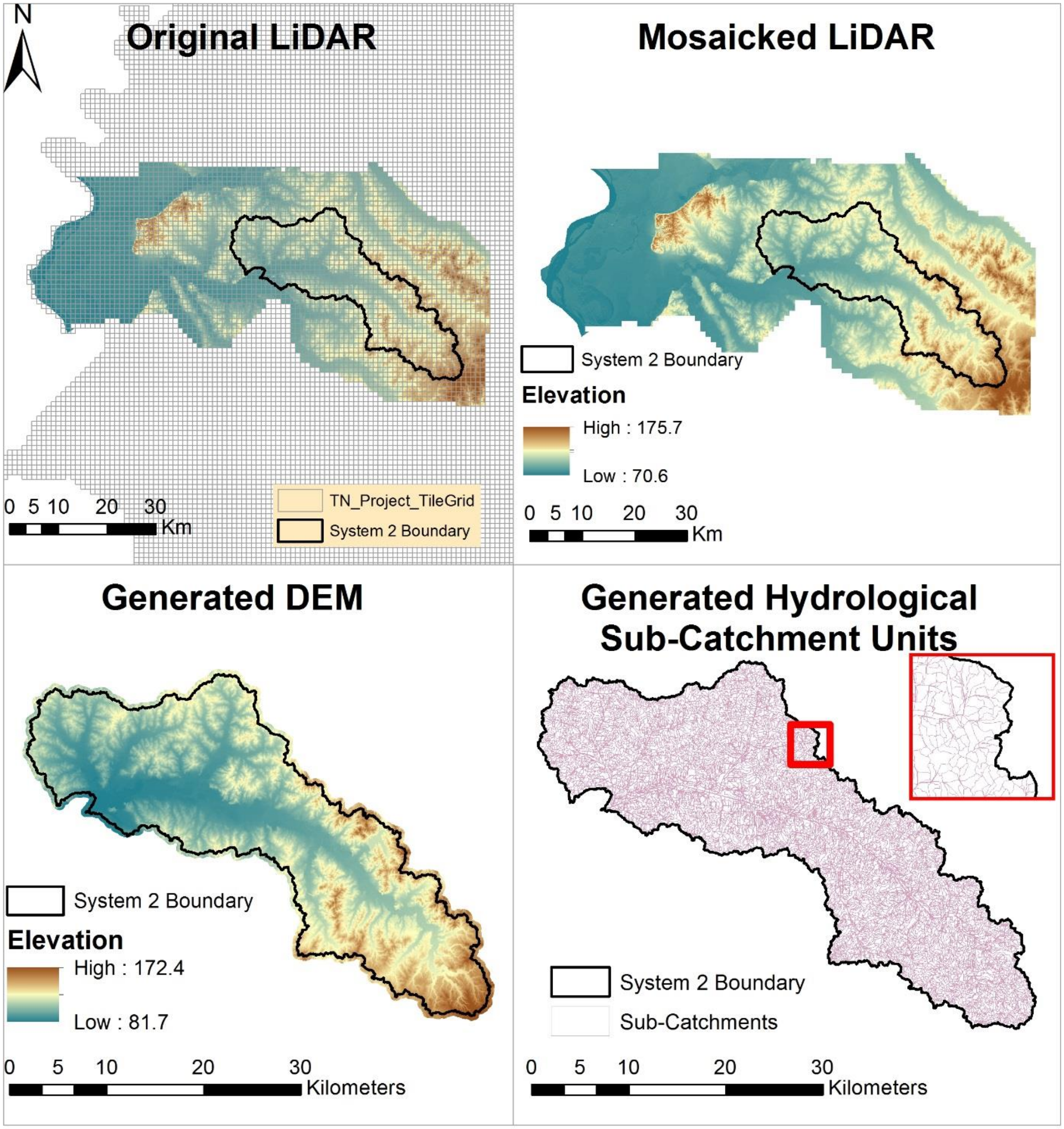


Figure 4: DEM pre-processing steps and the generated watershed subdivision (AnnAGNPS cells) for system 2

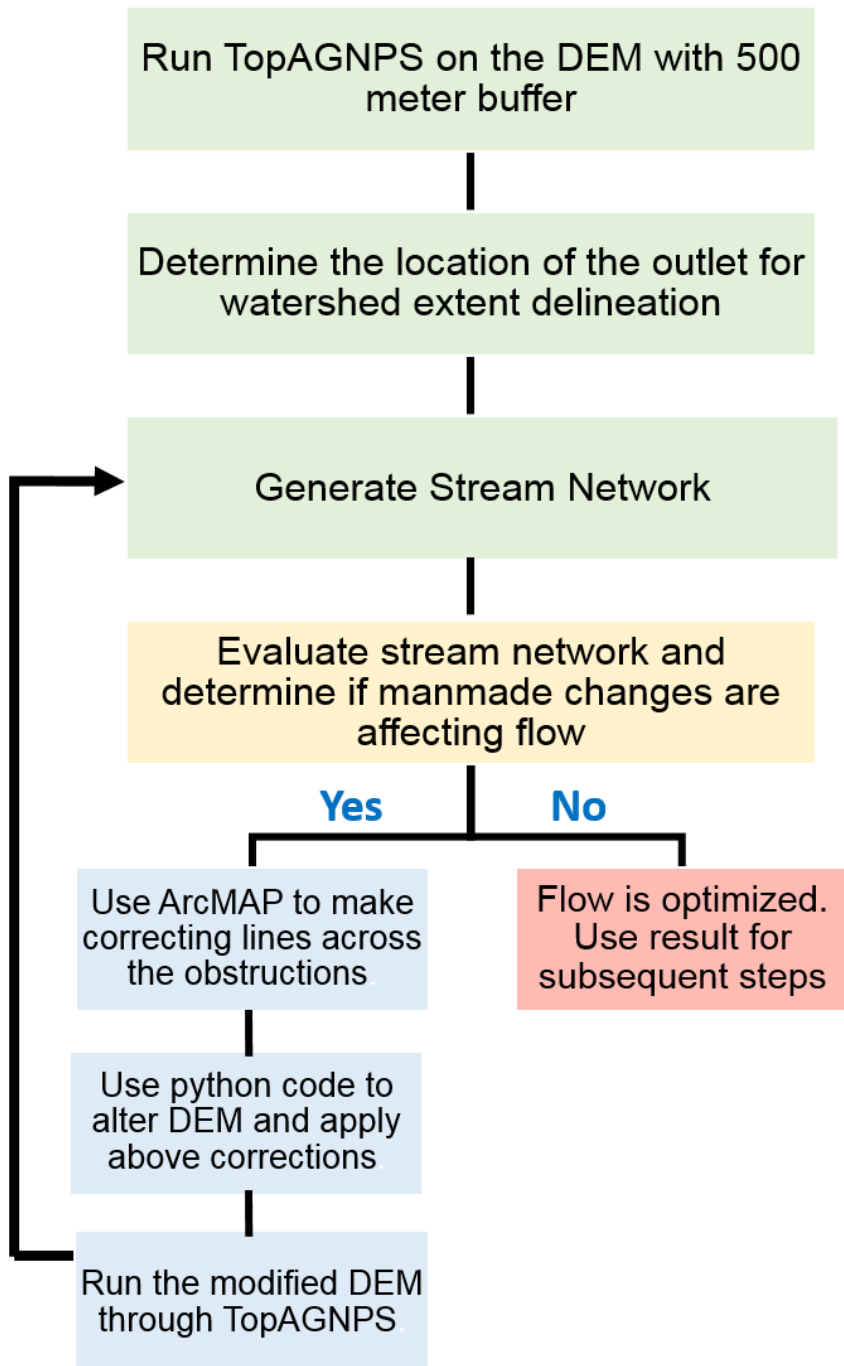


Figure 5: Flowchart describing the DEM post-processing steps for hydrological enforcement and features generation

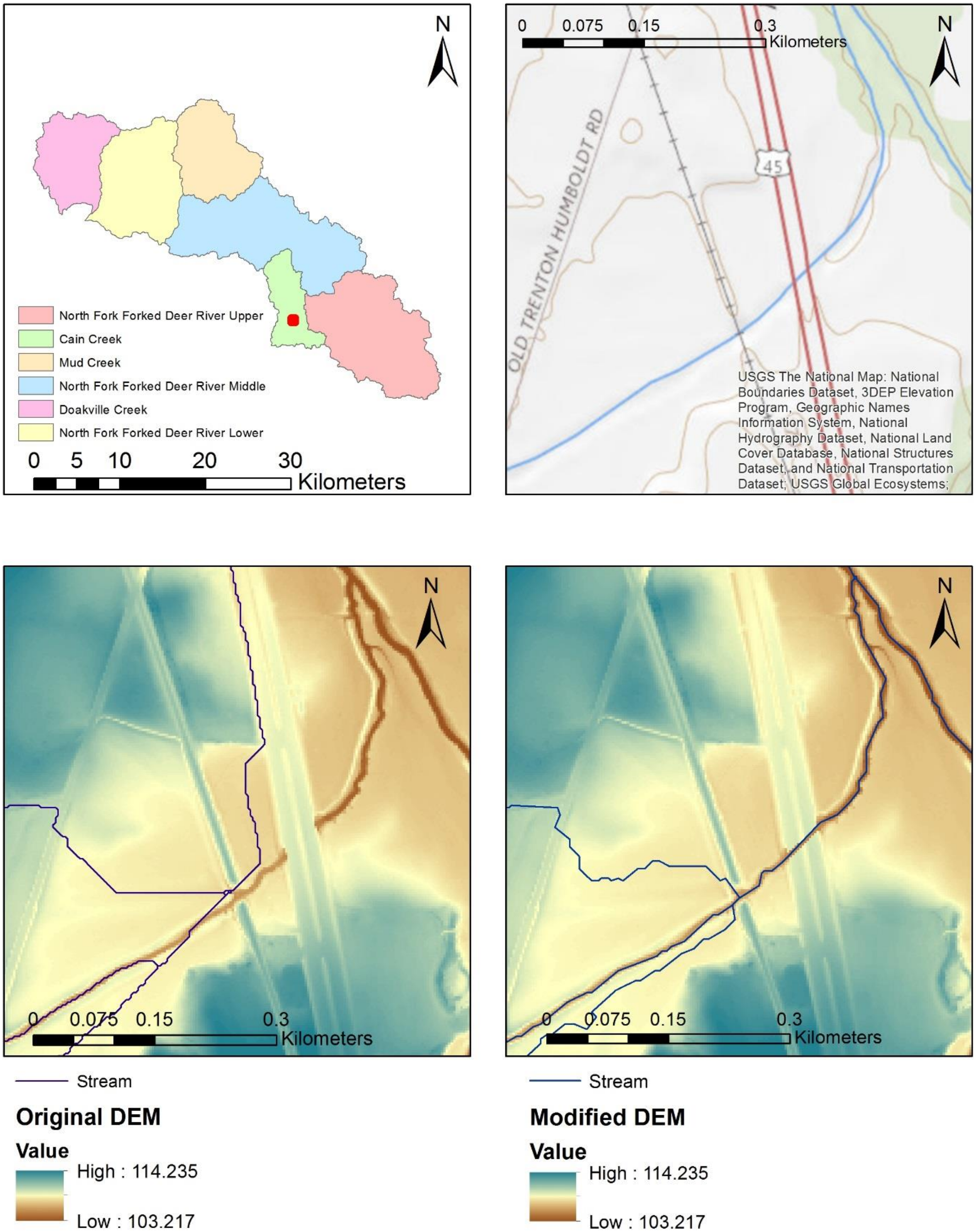


Figure 6: Hydrological Correction of DEM Example. Other examples are included in Appendix 1

Figure 7: Examples of field trip observations that were used in the hydrological correction process: (A) Ernest Rust road that is heavily wooded with man-made features not clear in satellite images. (B) Series of stops identifying the complex drainage network under I-24. (C) and (D) Field location of false positive drainage points.

The final topographical statistics for System 2 AnnAGNPS cells are summarized in Table. 1.

Table 1: The final statistics for System 2 AnnAGNPS cells.

HUC-12 Name	Number of AnnAGNPS Cells	Average Cell Area (Ha)	Average Cell Slope (%)	Average Cell Elevation (m)	Average Concentrated Flow Length (m)
North Fork Forked Deer River Upper	2946	5.00	0.08	120.51	46.45
Cain Creek	904	4.81	0.08	112.56	245.89
Mud Creek	1682	5.04	0.06	102.50	255.84
North Fork Forked Deer River Middle	3165	5.10	0.05	102.91	259.92
Doakville Creek	1561	4.95	0.06	96.97	243.84
North Fork Forked Deer River Lower	2311	4.99	0.04	94.18	259.56

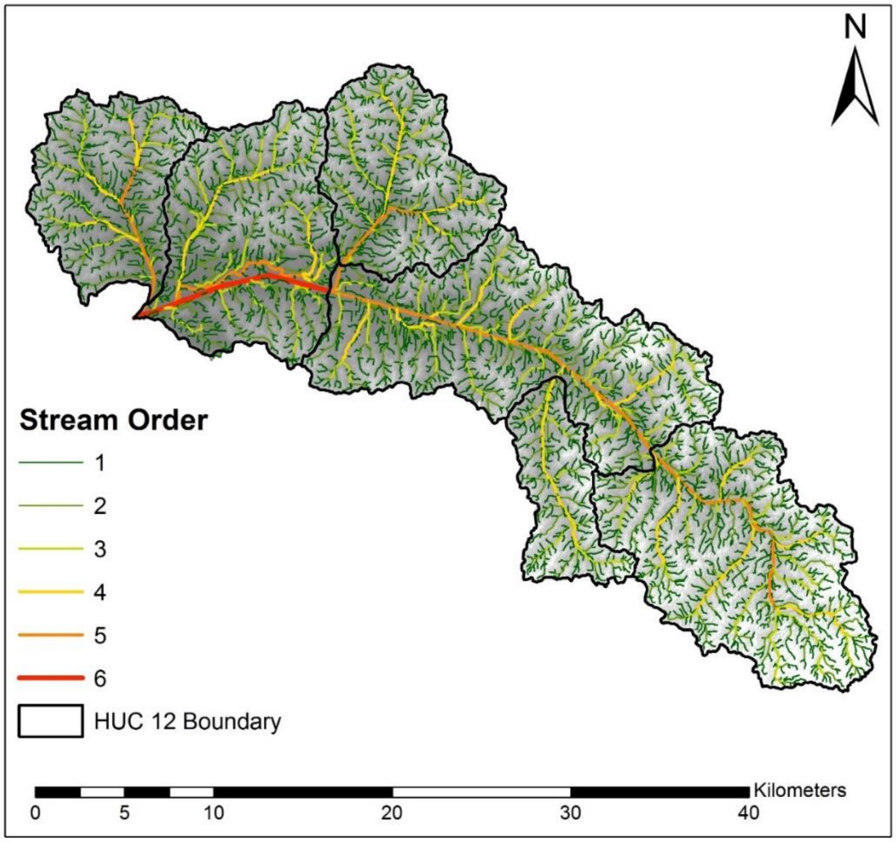


Figure 8: Generated stream network for System 2

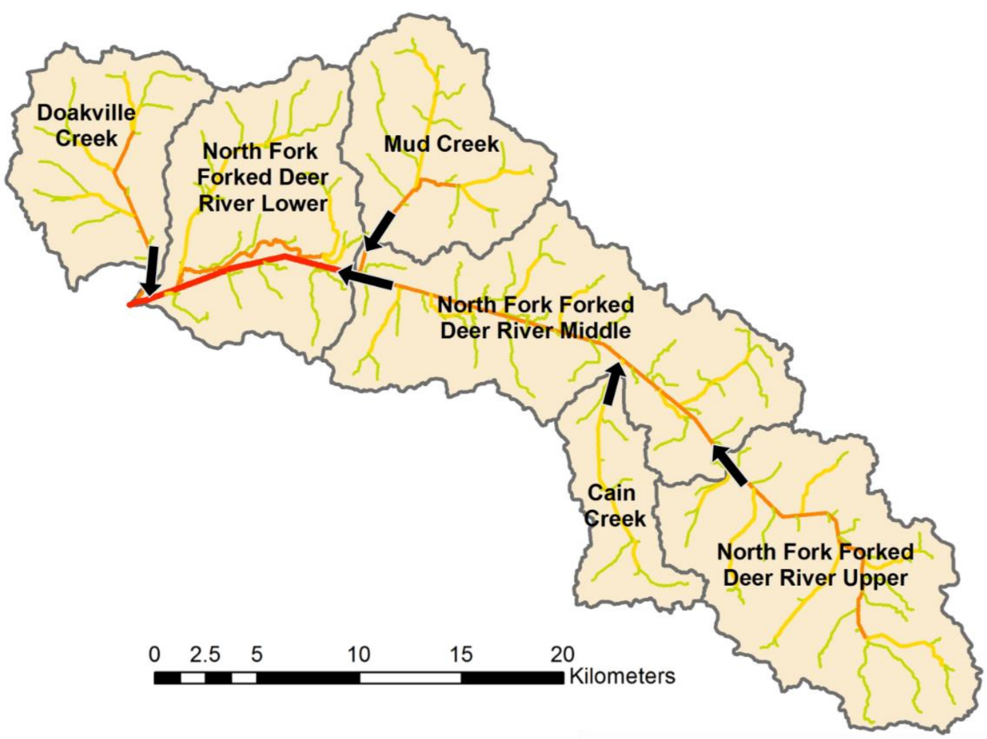


Figure 9: Sub-watersheds and their connections for System 2

3.2.2 Weather and Climate

Climate datasets from 2008 to 2018 were obtained from the U.S. National Oceanic and Atmospheric Administration (NOAA) for Montgomery, Robertson, Todd, and Logan Counties for system 1 and Gibson and Dyer Counties for system 2. These datasets contained climate elements that included the date, latitude, longitude, elevation, name, station, precipitation (mm), and the minimum and maximum air temperature (degrees Celsius). The data was acquired and imported into Microsoft Excel for pre-processing and data quality control using a custom algorithm (written in Python programming language).

The data was then spatially analyzed to identify the weather stations with the greatest temporal data coverage and their location in relation to the watershed. However, most of these stations were located outside of the study area boundaries. Hence, neighboring stations from outside the watersheds were used to fill data gaps in the stations located within the study area (referred to as secondary stations). Decision on which station to draw information was based on spatial zones of influence using Thiessen polygons. The primary station was defined as the centroid of each study area (Figure 10). Once these secondary stations record was filled, the stations were evaluated to ensure that the data distribution followed the climatic range for the closest cities in the region. The evaluation procedure also removed data anomalies. In the AnnAGNPS model, each sub-catchment (AnnAGNPS cell) is assigned to a secondary climate station. If data in a secondary station is missing data for a specific date, the missing information is then draw from the primary station.

Synthetic weather characteristics not available form historic observations were also generated using AGNPS Climate Generator (agGEM) software package, including dew point, sky cover, wind speed, and solar radiation. The AgGEM software package generates synthetic data for these four parameters using long-term patterns and records for different regions in the US.

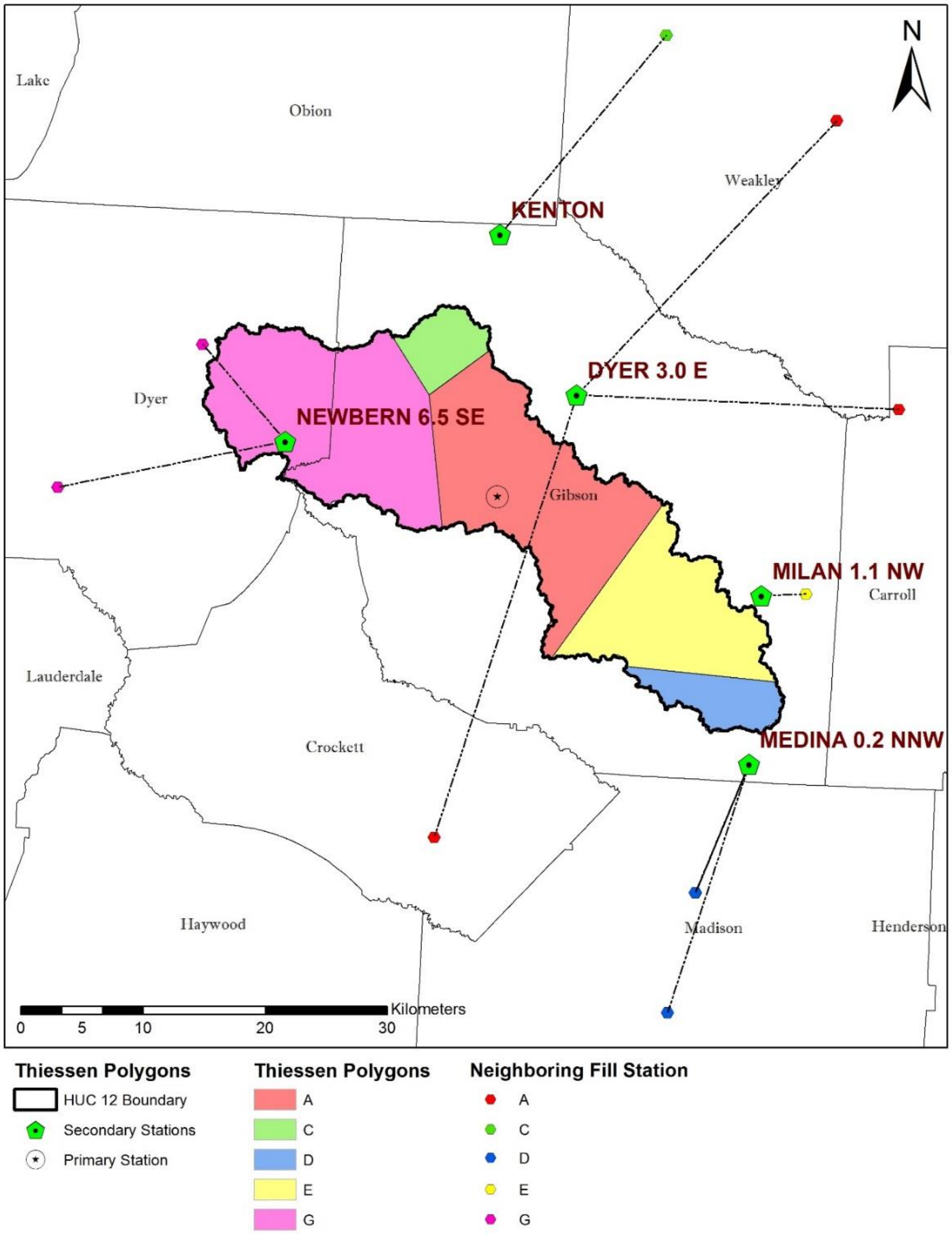


Figure 10: Weather/climate stations for System 2.

3.2.3 Soil

Soil spatial data was retrieved from the Web Soil Survey (WSS) using the Natural Resources Conservation Service's (NRCS) website for counties Montgomery, Robertson, Todd, and Logan for System 1 and counties Gibson and Dyer for System 2 (Figure 11). Using auxiliary information in concert with soil classification, soil polygons identified as flooded or submerged were removed.

Complementary soil description of physical and chemical properties in tabular format were retrieved from the USDA Soil Data Access website. The SQL script provided with the AnnAGNPS model distribution was used to query and retrieve soil information needed for the model. These datasets were post processed using the NASIS Import to AnnAGNPS (NITA) software package to ensure the accuracy of the soil characteristic table. This procedure was performed one county at the time (Figure 11). Once the data was quality controlled, the soil characteristic data table was joined to the attribute table of the original soil shapefile.

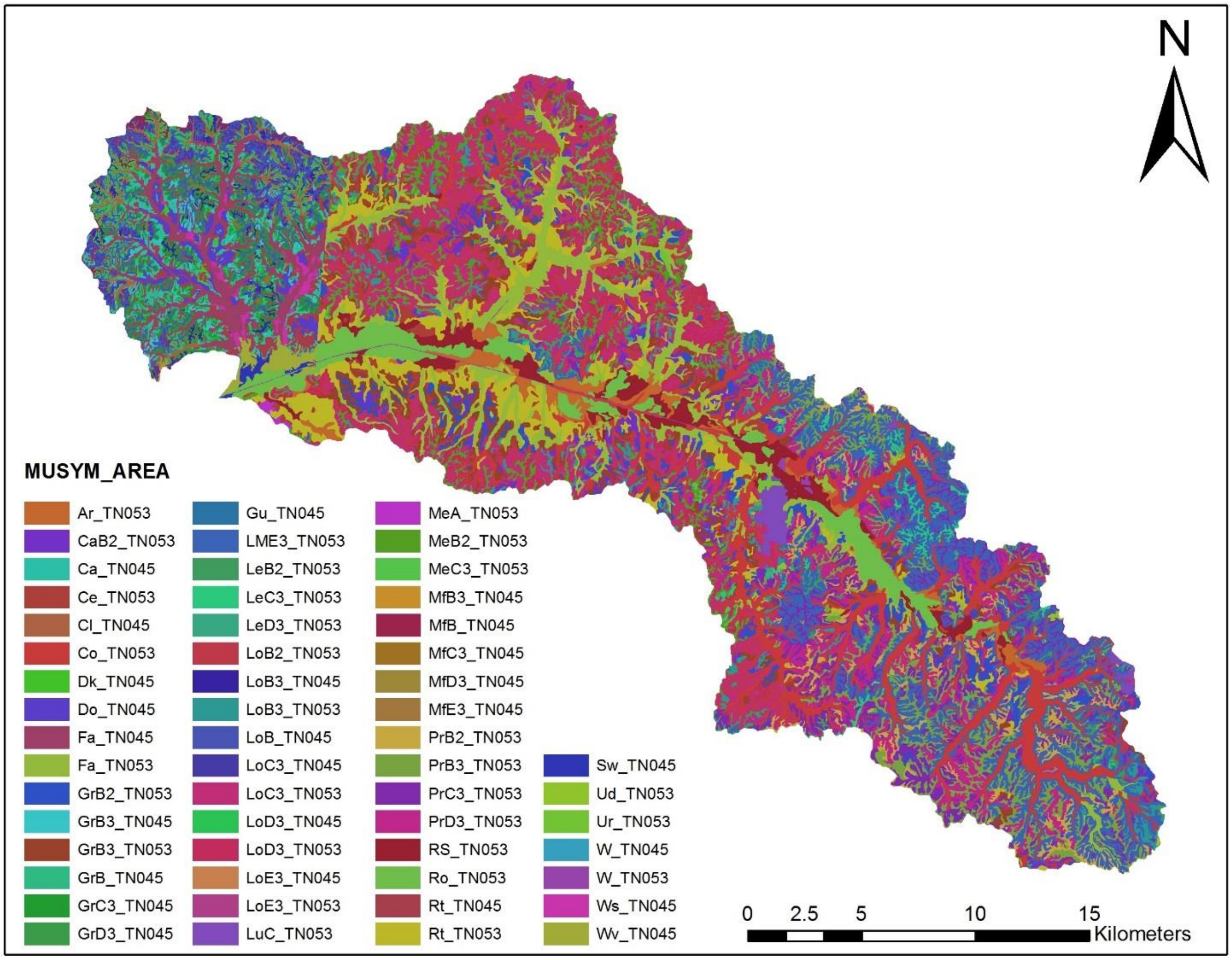


Figure 11: Soil map for System 2

3.2.4 Land Use and Land Cover

Annual land use/land cover data describing crop type from 2008 to 2018 was obtained from the National Agricultural Statistics Service's Cropland Data Layer (CDL) as raster grid files. These datasets were collected for Montgomery, Robertson, Todd, and Logan Counties for System 1 and Gibson and Dyer Counties for System 2. These CDL raster grid files were projected to the UTM 16N coordinate system. Statistical analyses were performed to determine dominant crop types based on datasets for years 2008, 2013, and 2018. Nine major classes were ascertained from the original land use classes. The nine dominant consistent classes are corn, cotton, winter wheat/soybean, forest, developed, grass/pasture, soybeans, woody wetlands, and water. These classes represented more than 90% of the watershed. Crops that were less conventional, such as pumpkins or Christmas trees, were classified under “grass/pasture,” given that they have a percentage of less than 0.01 of the total study area, and their effect will be very minimal on the results but their incorporation in the management schedule and operations will significantly increase the processing time and complexity.

The original land use/land cover raster grids for each system and for all years were resampled to the main nine land use classes. Additionally, the raster grids were resampled from 30 to 1-meter spatial resolution for improved results of spatial zonal statistic GIS analysis. This method was used to assign the representative land use to each AnnAGNPS cell (one dominant land use per year). Differences between the three main steps: original data, reclassified data and discretized data for the year 2008 are presented in Figure 12. The same procedure was applied for all other years, but we used 2008 for illustration purposes. Temporal distribution of the main nine classes during the period of investigation are shown in Figure 13.

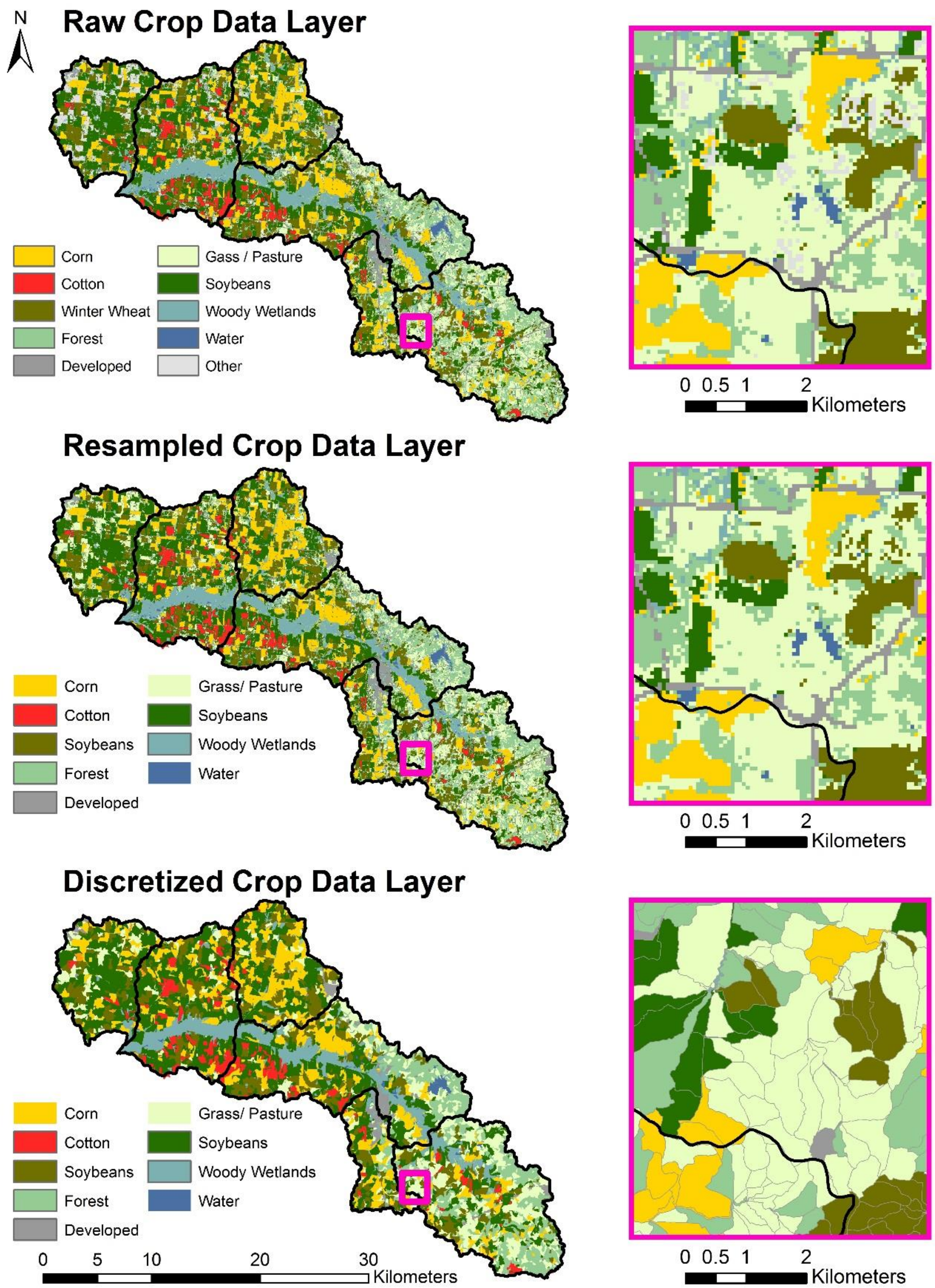


Figure 12: The three main processing steps for the crop data layer: original data, reclassified data and discretized data for the year 2008. The same procedure was applied for all other years, but only 2008 is included for display purposes.

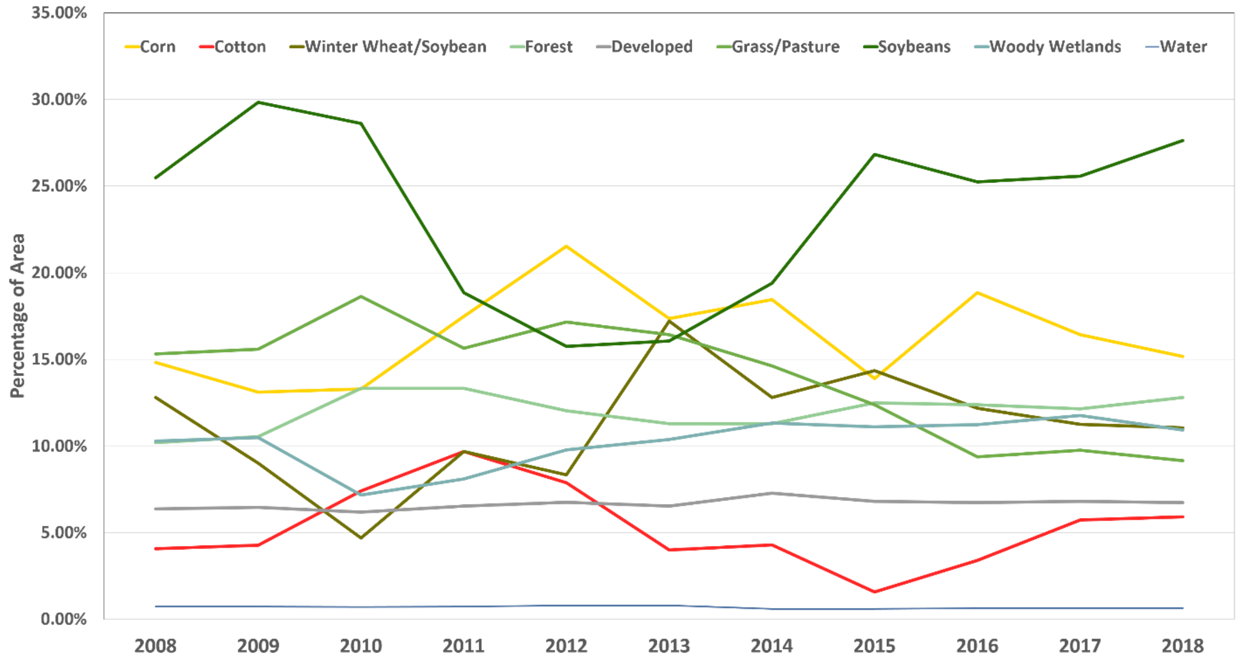


Figure 13: Temporal distribution of the main nine dominant land cover in system 2

3.2.5 Management Practices

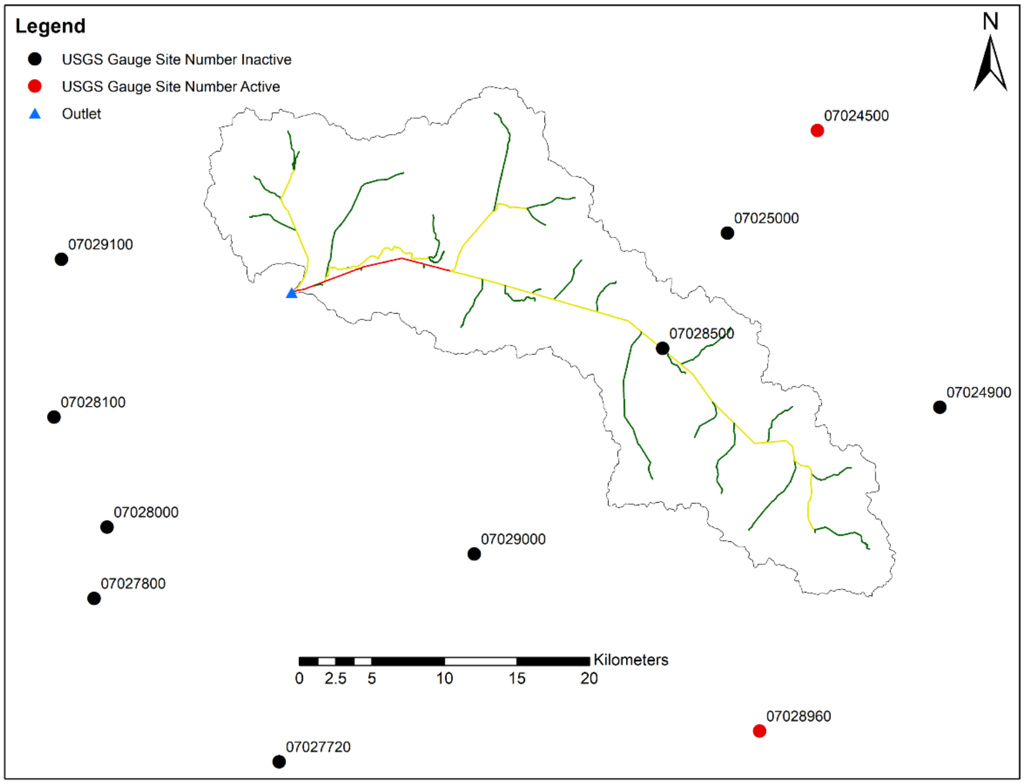
Farming management practices were generated by integrating spatiotemporal crop type information at raster grid cell scale (from CDL), average crop yield at county scale (from USDA-NASS), and one-year farming management schedule (from USDA-NRCS). The latter represents typical farming operations and schedules for a particular crop type in this region. This information was mapped to crop managements in the Revised Universal Soil Loss Equation (RUSLE2) database. The computer program RUSLE2 Import to AnnAGNPS (RITA) was used to convert crop management information from the RUSLE2 database into AnnAGNPS management schedule and operation data sections. Specifically, the management schedule includes information about when tilling occurs, when the plants are sowed, when the fertilizer and insecticide are applied, and when the harvest and fallow are conducted. Nine management templates describing farming schedule and operations in AnnAGNPS input file format were generated (i.e. – Corn, Cotton, Winter Wheat/Soybean, Mixed Forest, Developed, Grass/ Pasture, Soybean, Woody Wetland, and Water).

The land use management files were then processed using two custom algorithms (Python programming language). The first code generated a management sequence for each AnnAGNPS cell (sub-catchment) e.g. - C; WwS; C; WwS; C; WwS; C; WwS; C; WwS; C in which each abbreviation represents a land use type and each sequence covers the 11 year period from 2008 to 2018. This sequence of land use was used to generate a database of unique management sequence (crop rotation) so that all AnnAGNPS cells of the same rotation sequence will be referenced to the same management.

The second algorithm developed (Python programming language) merges the sequence of land use with one-year management and operation templates to develop 11 years crop rotations in the required AnnAGNPS file format. Once both codes are applied, this management schedule and operation is run through AnnAGNPS.

3.3 Baseline conditions results:

Watershed modeling allows the simulation of hydrological processes in a holistic approach while accounting for natural and anthropogenic processes temporal and spatial variation and their interrelationships. In this study, the Annualized Agricultural Non-Point Source (AnnAGNPS) watershed pollution model was used to generate watershed simulations for our study areas. The following sections represent baseline conditions for systems 1 and 2 without simulating the potential impact of any conservation practice. The model was applied in un-gauged mode given the lack of stream gauge data in the studied watersheds (Figure 14). The generated results depict relative changes of long-term spatiotemporal trends in non-point source sources and sinks between the simulation considering baseline conditions and multiple simulations of alternative scenarios considering the adoption of conservation practices (section 4). Results are expressed as (1) annual average sediment yield per unit of area for individual AnnAGNPS cell to characterize spatial variation and (2) monthly sediment load to the watershed outlet representing the temporal overall watershed response.



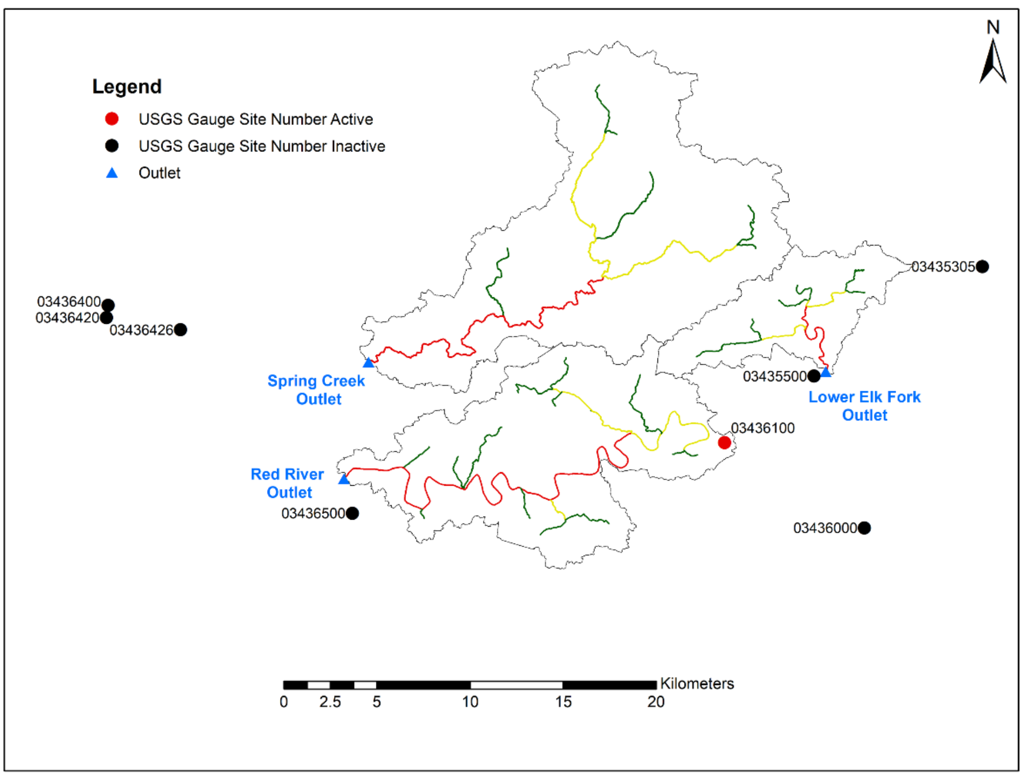


Figure 14: Stream gauge availability in Systems 1 and 2 during the period of study. The maps show that no active gauge is available in system 2 (top map) and only one active gauge is available near the headwaters of Red River in system 1 (bottom map), making it not suitable for calibrating flow at the outlet of the watershed.

3.3.1 Spatial Distribution:

- System 1:

Simulation results representing sediment yield (sediment eroded and leaving the edge of the field) were expressed as annual average per unit of area for different sediment particle sizes (Figures 15-17). To facilitate the visualization of spatial variation of sediment yield, a standardized approach was employed to all watersheds. Results were expressed as deviations from the mean, allowing for the relative identification and classification of high/low sediment producing AnnAGNPS cells.

Results from the Spring Creek simulation indicated a high spatial variability in sediment yield for clay and silt size particles but with reduced sediment yield of sand size particle sizes (Figure 15). The spatial distribution of high sediment producing sub-catchments does not show a defined pattern. For Lower Elk Fork watershed, most sediment yield simulated is of silt particle size (Figure 16) with identified clusters of high sediment producing sub-catchments (red polygons in Figure 16). Results for the Red River simulation indicated the lowest sediment load in all sediment categories but identified smaller clusters of high sediment producing sub-catchments (Figure 17).

In System 2, simulation results indicate small overall yield of sand size particle sizes but for both silt and clay particle sizes sediment yield are high especially in Mud Creek, North Fork Forked Deer River Lower and Doakville Creek in comparison to Cain Creek and North Fork Forked Deer River Upper (Figure 18). Simulation results of sediment yield agreed spatially with land cover, where lower values were estimated for forested, urban, and pasture and higher values for crop lands.

Additionally, simulation results characterized the complexity of processes acting as sources and sink of sediment in time and space based on detailed representation of variations of land cover, climate, and farming practices. These maps could serve as a guideline for the development of targeted implementation of the appropriate conservation practices in critical locations.

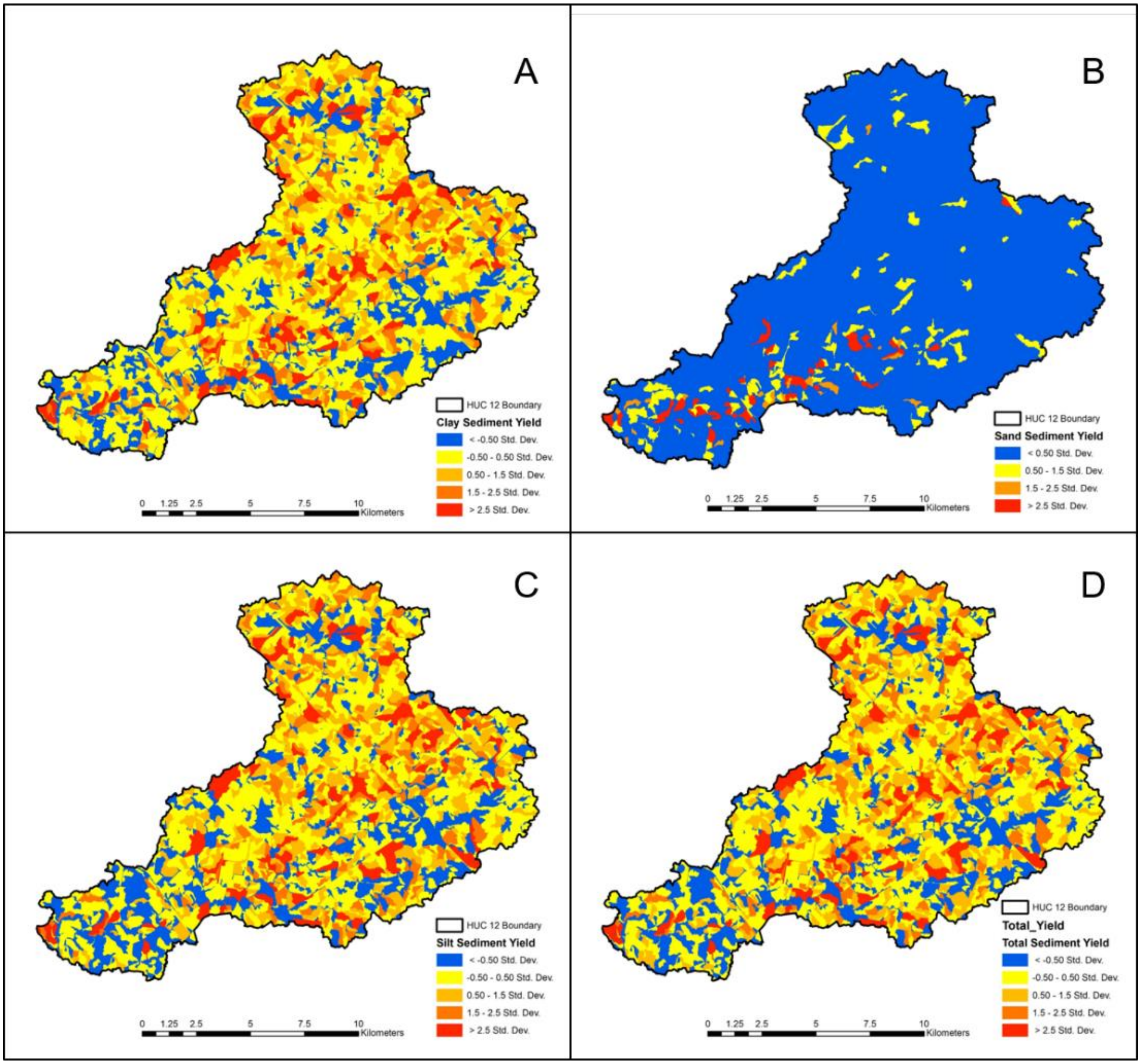


Figure 15: Spatial distribution of sediment yield in Spring Creek (A: Clay, B: Sand, C: Silt, and D: Total Sediment)

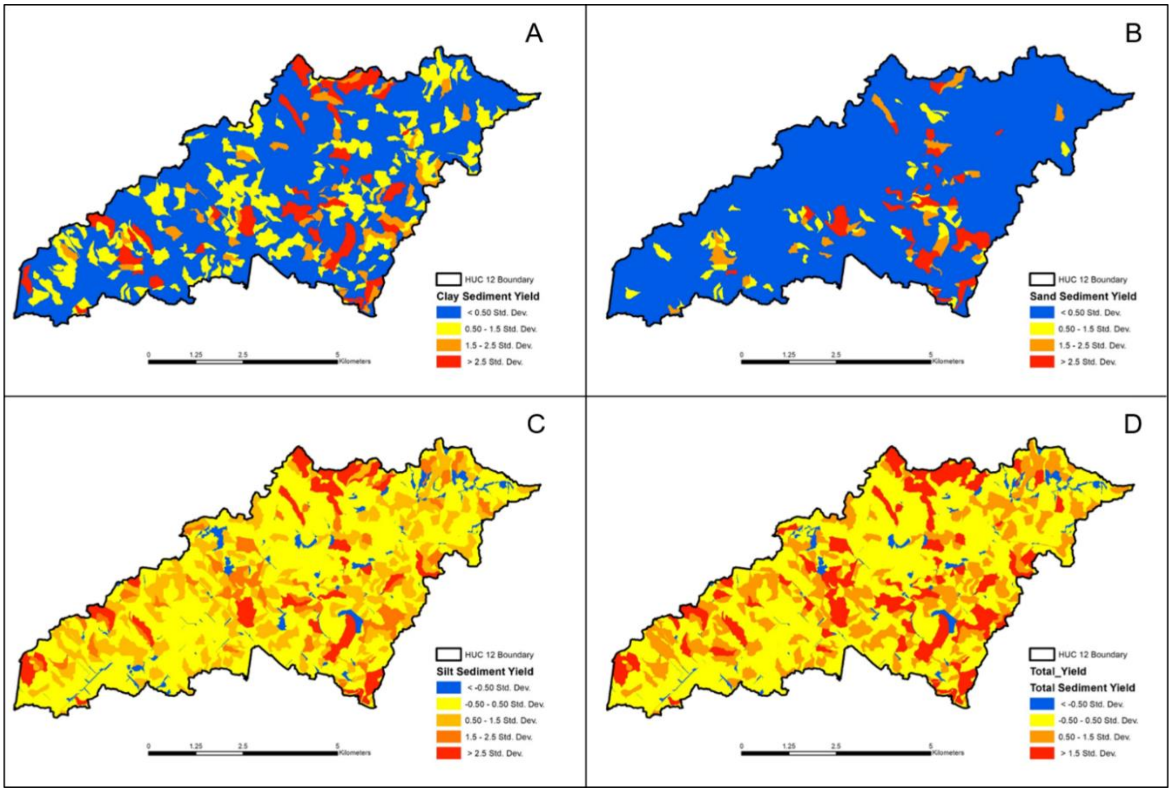


Figure 16: Spatial Distribution of Sediment Yield in Lower Elk Fork (A: Clay, B: Sand, C: Silt, and D: Total Sediment)

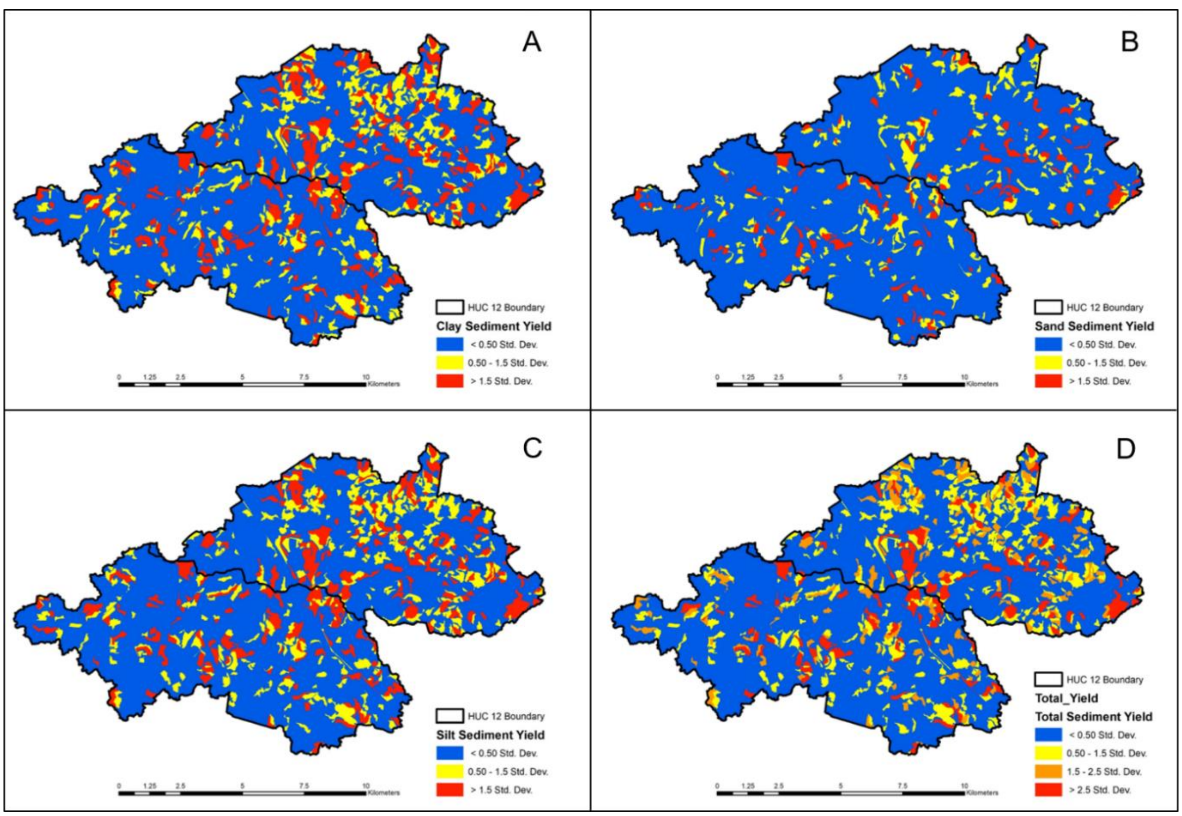


Figure 17: Spatial Distribution of Sediment Yield in Red River (A: Clay, B: Sand, C: Silt, and D: Total Sediment)

- System 2:

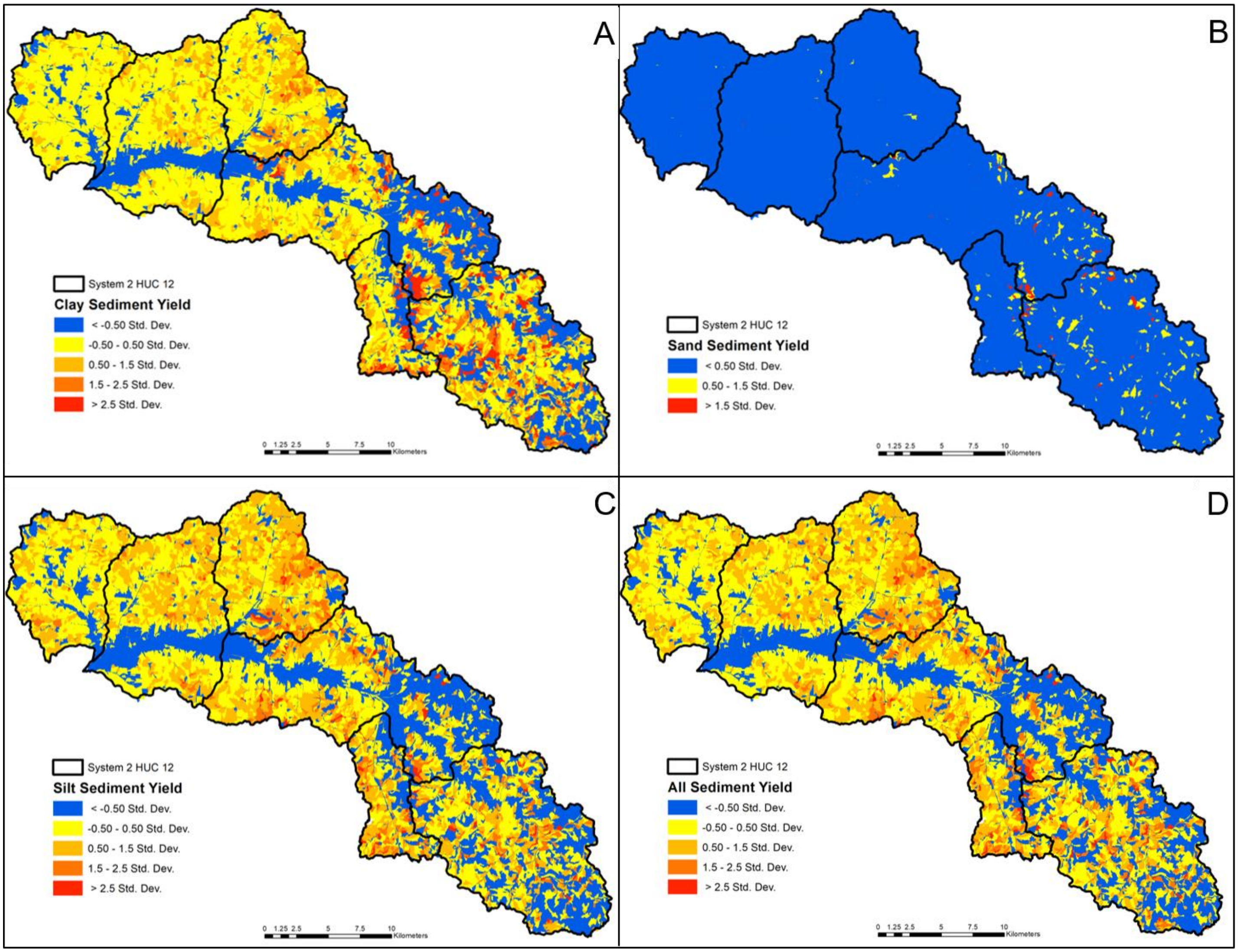


Figure 18: Spatial Distribution of Sediment Yield in System 2 (A: Clay, B: Sand, C: Silt, and D: Total Sediment)

3.3.2 Temporal Distribution:

- System 1:

The temporal distribution was evaluated as sediment load at the watershed outlet resampled from daily to monthly total values for Lower Elk Fork subsystem (Figure 19), Red River subsystem (Figure 20), Spring Creek (Figure 21) and system 2 (Figure 22). Sand, silt and clay loads are correlated in all subsystems of system 1 with major peaks in 2018 for Lower Elk Fork, in 2010 for Red River and in both 2010 and 2018 for Spring Creek. In system 2, there is a temporal correlation between Silt and Clay yield but not with sand load, with a major peak in 2010. Sand yield peaked in 2016 in system 2. It is noticeable that estimates of sediment load of sand size particles are orders of magnitude smaller than silt and clay for all systems.

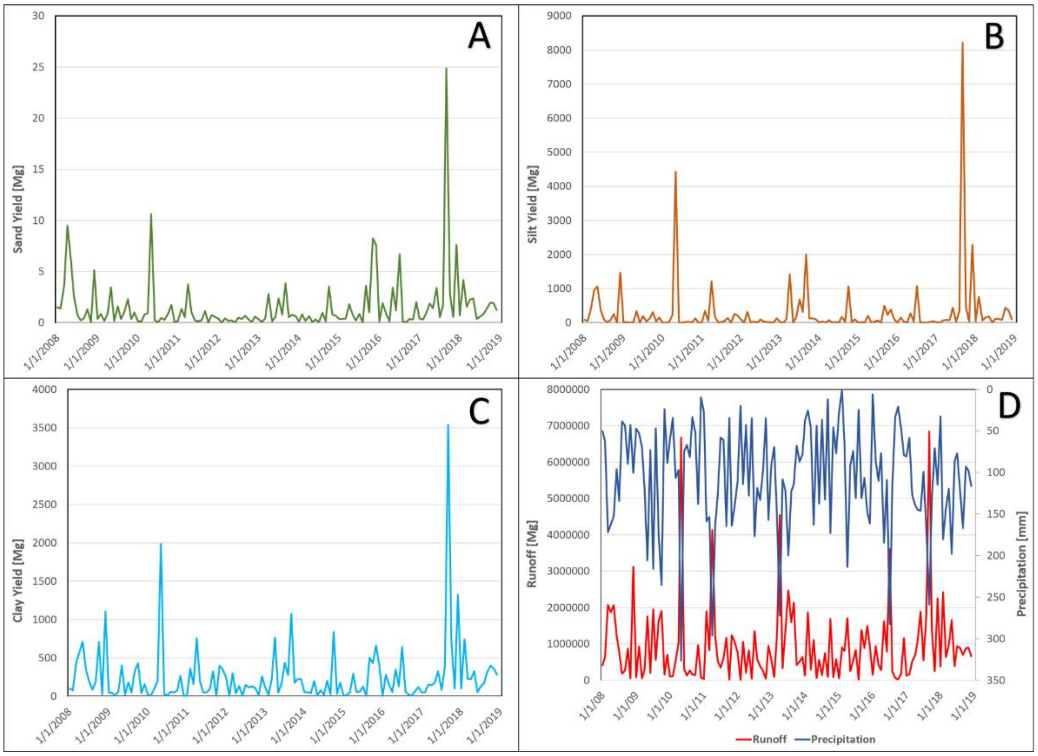


Figure 19: Temporal Distribution of Runoff and Sediment Load at the Outlet of Lower Elk Fork – System 1 (A: Sand, B: Silt, C: Clay and D: Runoff)

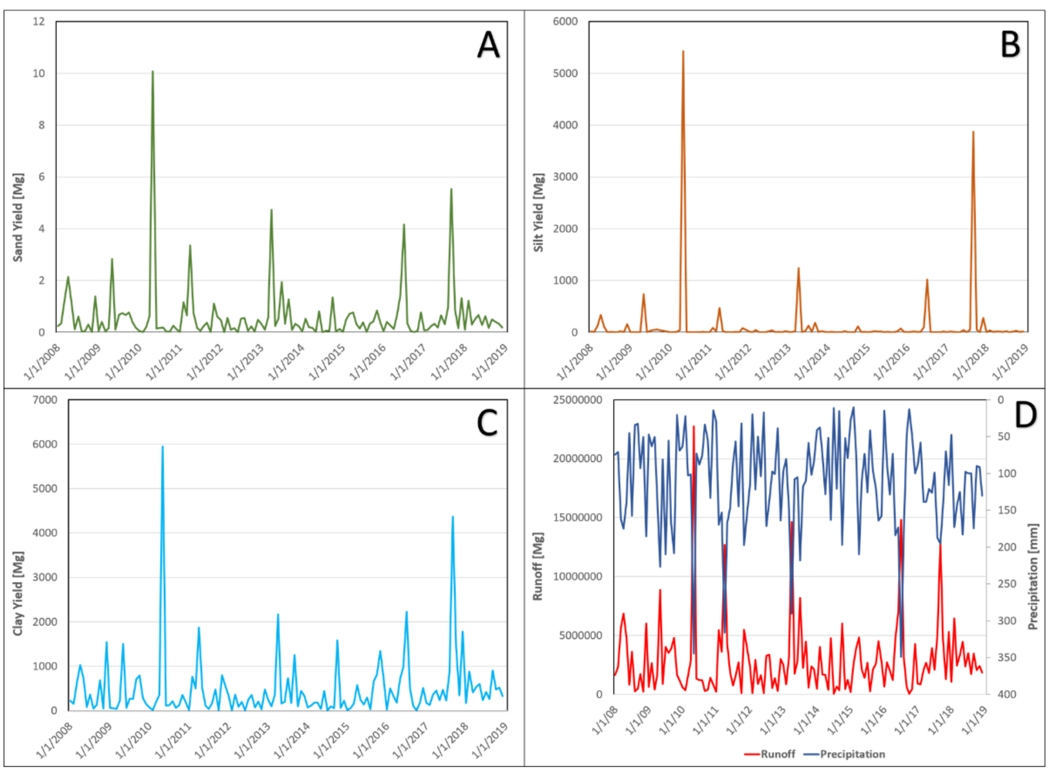


Figure 20: Temporal Distribution of Runoff and Sediment Load at the Outlet of Red River-System 1 (A: Sand, B: Silt, C: Clay and D: Runoff)

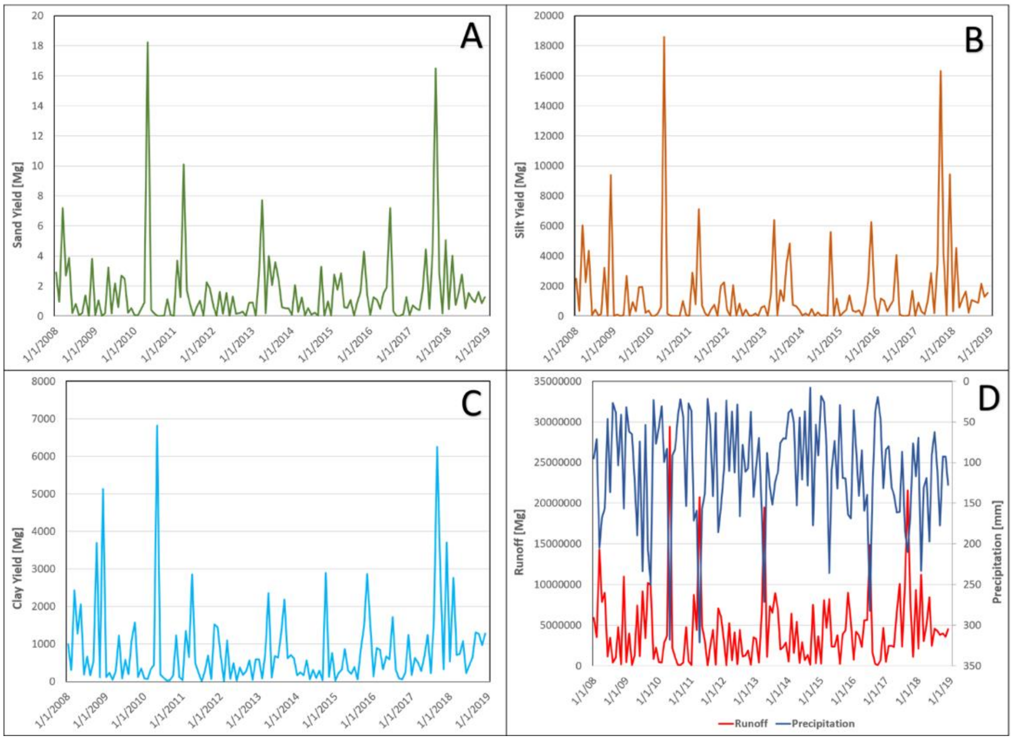


Figure 21: Temporal Distribution of Runoff and Sediment Load at the Outlet of Spring Creek - System 1 (A: Sand, B: Silt, C: Clay and D: Runoff)

- System 2:

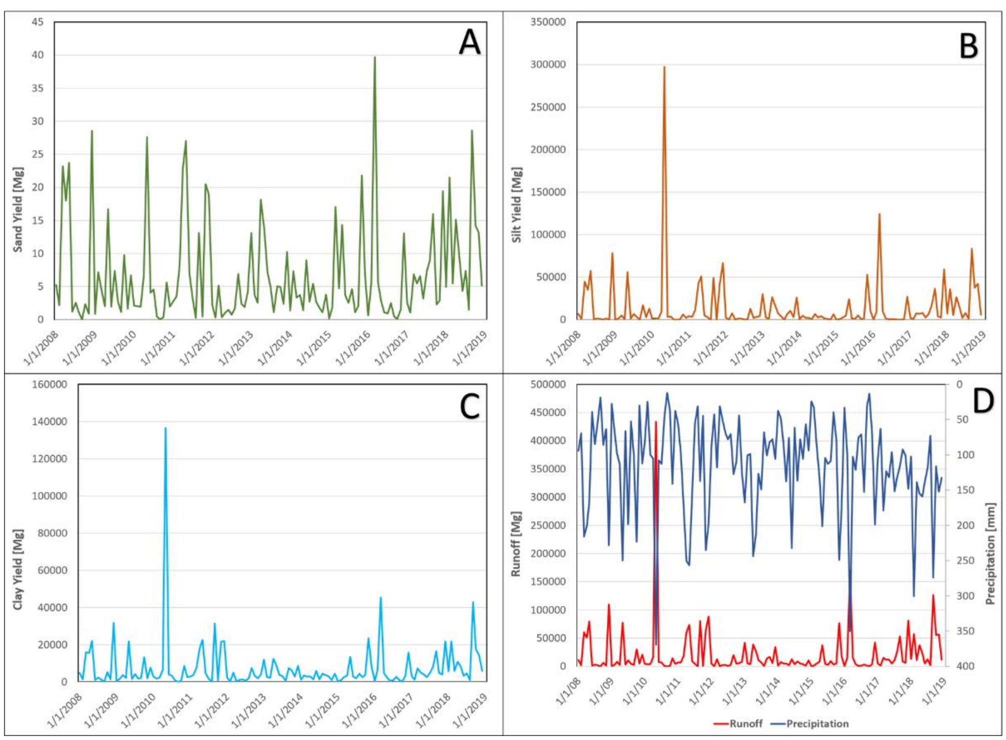


Figure 22: Temporal Distribution of Runoff and Sediment Load at the Outlet of System 2 (A: Sand, B: Silt, C: Clay and D: Runoff)

4 Evaluation of the Effectiveness of Conservation Practices Simulations

4.1 Riparian Forest Buffer (Code 391)

4.1.1 Description of the practice:

A riparian forest buffer is an area location adjacent to and up gradient from a water body predominantly covered by trees and/or shrubs. The riparian forest buffer applies to areas associated with ground water recharge such as permeant or intermittent streams, lakes, ponds and wetlands. The presence of riparian forest buffers reduces the transport of sediments, chemicals, pesticides and pathogens to surface water. The dominant vegetation in these buffers includes existing or planted trees and shrubs with grasses and forbs that grow in naturally; the planted vegetation is selected carefully to fit the site characteristics and ecosystem (Figure 23).

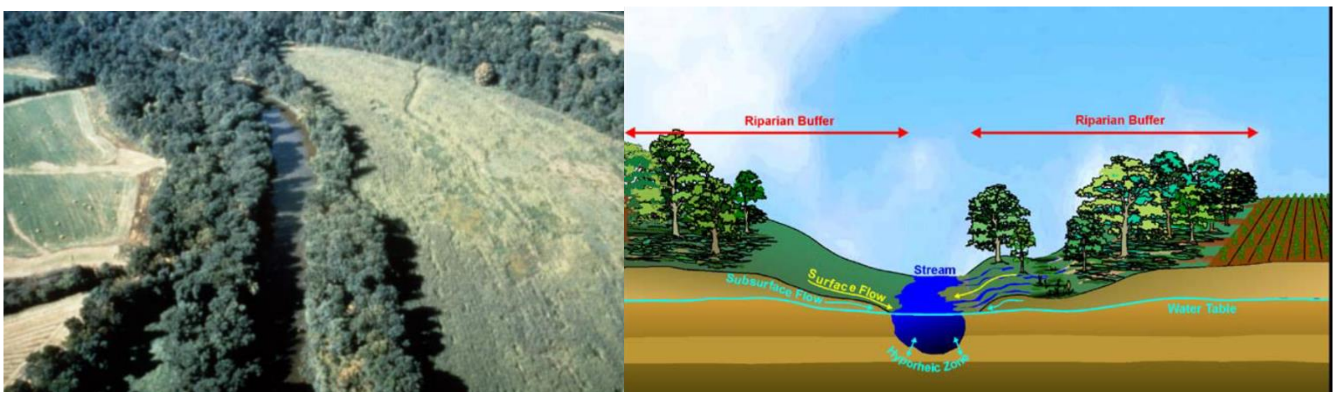


Figure 23: Conservation Practice 391. Left Image Source: USDA – NRCS Conservation Practice Job Sheet, Right Image Source: EPA EPA/600/R-05/118 report

4.1.2 Considered Scenarios:

First, we delineated the existing riparian buffer using the RCL tool that was generated by this project (see section 3 of this report). We simulated the impact of the existing riparian vegetation as potential buffer on the sediment load. Subsequently, we evaluated the effectiveness of constructed riparian buffer scenarios ranging in size and spatial extent.

The riparian buffer sizes that were considered are 10m, 30m and 60m. We selected these buffer widths, given that the EPA defines narrow buffer width as 1 – 15 meters and wide buffer width as higher than 50 meters. State and federal guidelines range from seven to 200 meters.

The managed riparian buffer locations that were considered are:

- Around all the streams
- Around all the streams adjacent to agricultural fields (Figures 24–27-part **A**)
- Around all the streams adjacent to agricultural fields that have a sediment yield higher than mean – standard deviation (referred to as "> medium") (Figures 24-27 – part **B**)
- Around all the streams adjacent to agricultural fields that have a sediment yield higher than mean (referred to as "> high") (Figures 24-27 – part **C**)
- Around all the streams adjacent to agricultural fields that have a sediment yield higher than mean + standard deviation (referred to as "> Very High") (Figures 24-27 – part **D**)

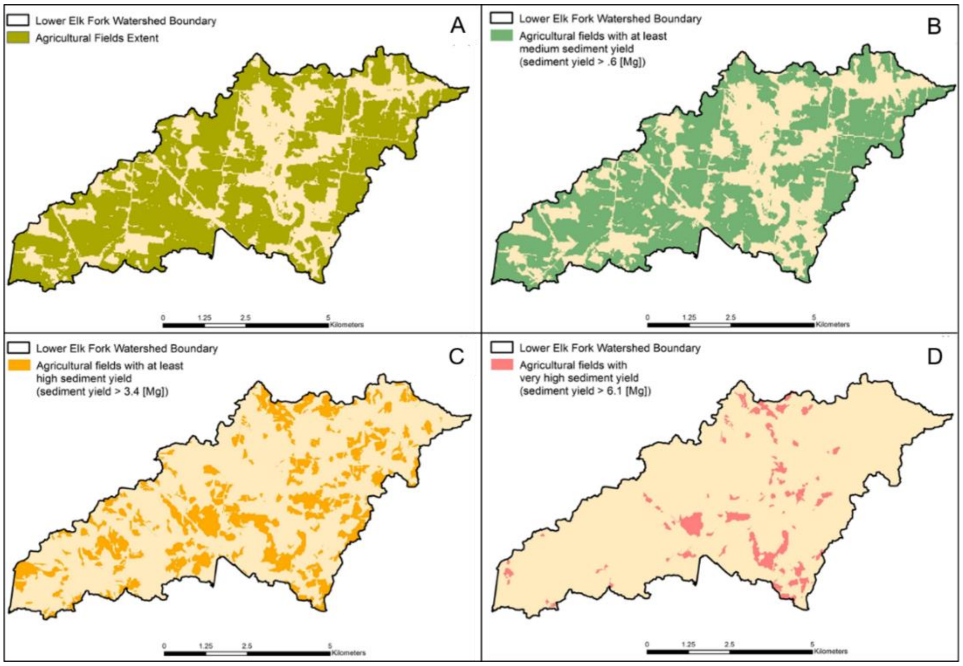


Figure 24: Lower Elk agricultural fields classified by sediment yield (A: all agricultural fields, B: Agricultural fields with at least medium sediment yield, C: Agricultural fields with at least high sediment yield and D: Agricultural field with at least very high sediment yield). These four classes are used as scenarios in our final simulations to estimate the contribution of using a riparian buffer in reducing sediment yield (from least to most restrictive)

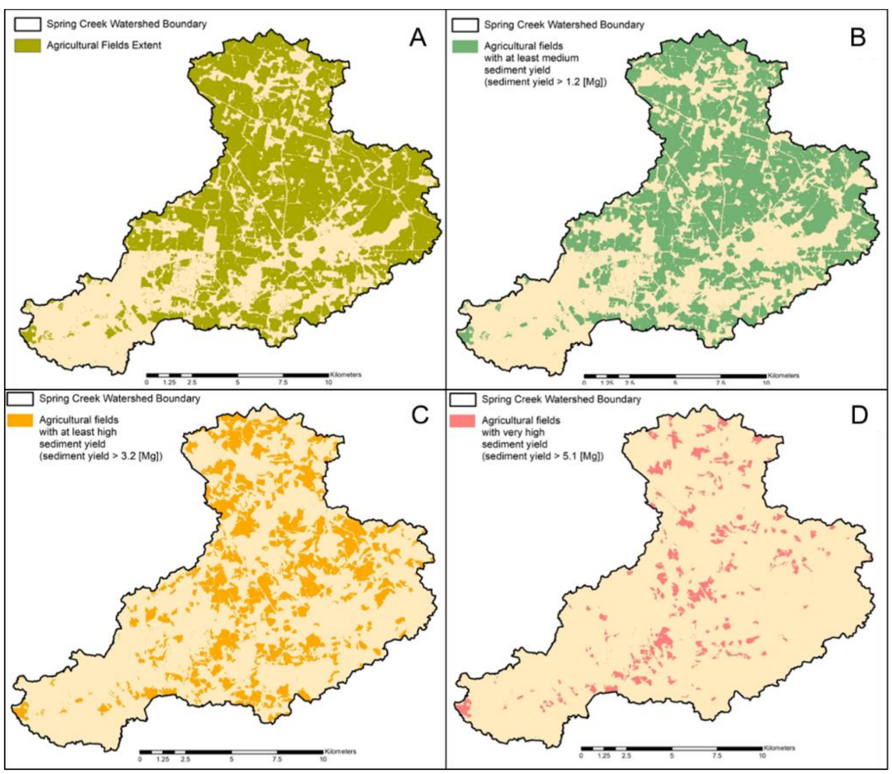


Figure 25: Spring Creek agricultural fields classified by sediment yield (A: all agricultural fields, B: Agricultural fields with at least medium sediment yield, C: Agricultural fields with at least high sediment yield and D: Agricultural field with at least very high sediment yield). These four classes are used as scenarios in our final simulations to estimate the contribution of using a riparian buffer in reducing sediment (from least to most restrictive)

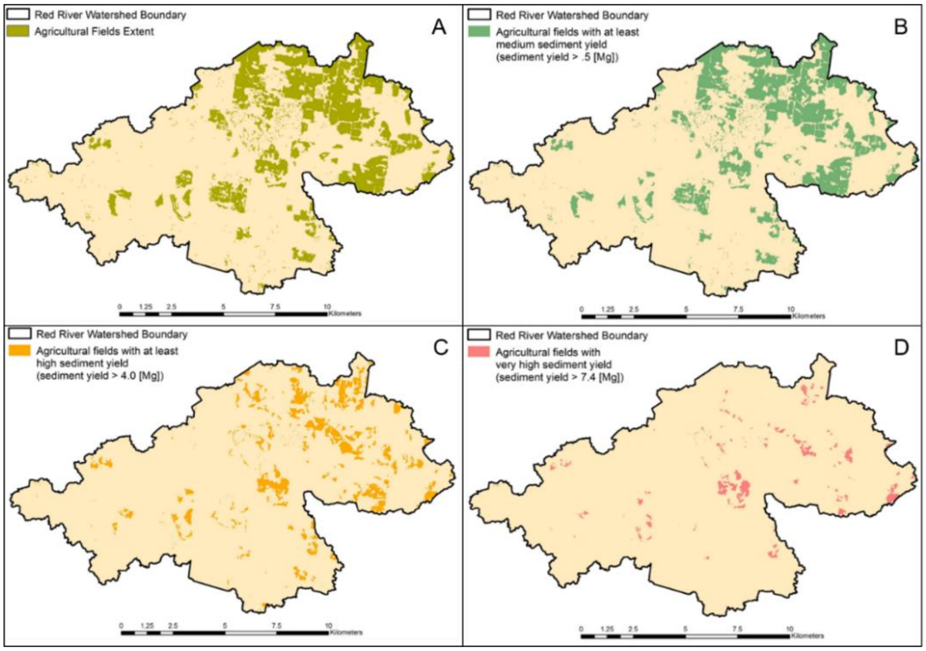


Figure 26: Red River agricultural fields classified by sediment yield (A: all agricultural fields, B: Agricultural fields with at least medium sediment yield, C: Agricultural fields with at least high sediment yield and D: Agricultural field with at least very high sediment yield). These four classes are used as scenarios in our final simulations to estimate the contribution of using a riparian buffer in reducing sediment yield (from least to most restrictive)

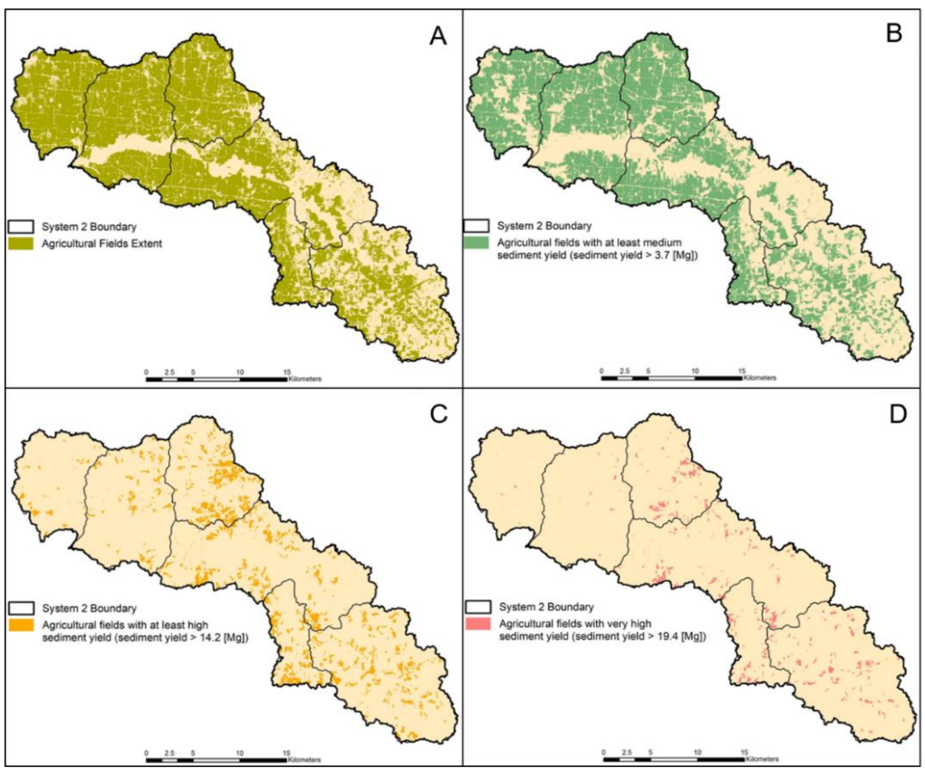


Figure 27: System 2 agricultural fields classified by sediment yield (A: all agricultural fields, B: Agricultural fields with at least medium sediment yield, C: Agricultural fields with at least high sediment yield and D: Agricultural field with at least very high sediment yield). These four classes are used as scenarios in our final simulations to estimate the contribution of using a riparian buffer in reducing sediment yield (from least to most restrictive)

These various location classes (Figures 24 -27) were selected to evaluate the effectiveness of constructed riparian buffer in a spectrum of locations ranging in one hand between just the very productive areas in terms of sediment production and everywhere in the watersheds. This will allow us to identify the optimal scenario in terms of sediment conservation.

By combining these different conditions (i.e. sediment yield class, all or just agricultural areas, buffer width) we run 16 simulations representing 16 scenarios summarized in Figure 28.

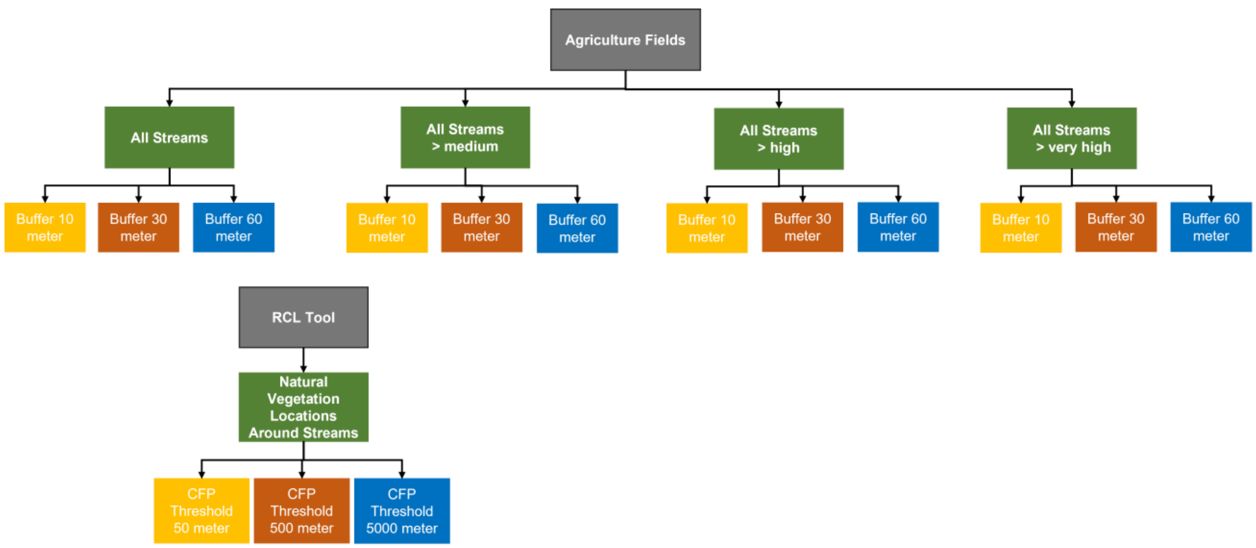


Figure 28: Flow Chart summarizing the various generated scenarios

4.1.3 Simulation Results

The results of the various riparian buffer scenarios are summarized in Tables 2-5 and Figures 29-30.

Riparian buffers work by slowing surface flow and promoting infiltration and fine sediment deposition. Our simulations indicate that they are a very effective tool with a potential reduction of up to 81% in System 2, 76% in Lower Elk, 60% in Red River, and 81% in Spring Creek. It is important to note that these estimates represent the maximum potential reduction by this approach, and we understand that it may not be feasible to implement a buffer around every stream of the watershed. However, those 12 simulations (3 for each watershed) provide an estimate the maximum that can be reduced by this conservation practice. A total of 76 AnnAGNPS simulations were performed for various riparian buffer conditions, with varying computer run time between 24 to 120 hours per simulation. The goal was to assess the impact of riparian buffers in a wide range of conditions.

Simulations of existing riparian buffer conditions indicate that in System 2, Lower Elk, and Spring Creek are under-served in terms of riparian buffer. Simulations of the existing buffer in System 2 indicate that it reduces sediment yield by 13% as opposed to a maximum potential of 81%. The existing buffer in Lower Elk reduces sediment yield by 24% as opposed to a maximum potential of 76%. The existing buffer in Spring Creek reduces sediment yield by 21% as opposed to a maximum potential of 81%. System 2 is the most underserved. In Red River, the existing riparian buffer, reduces sediment by 25% with the maximum being 60%, and therefore it is the most served. This is partially due to the low acreage of agriculture in Red River in comparison to the other watersheds. It is also one of the most actively managed watersheds in the region.

Figure 30 show a clear correlation between the surface area of the buffer (width and length) and sediment reduction. However, the length of the buffer is a more impactful factor than the width, even when the total surface area of the buffer is the same. For instance, cluster A - Figure 29 represents three scenarios with almost equal surface area (15,000,000 m^2), but the three scenarios vary in effectiveness: 51% for both LEF_MED_60 and LEF_AG_60, and 76% for LEF_ALL_30. We see a similar example in Spring Creek watershed. Cluster B showcases that effectiveness can vary between 49% to 80% for an almost equal surface area of the buffer.

It is possible that riparian vegetation contains concentrated flow paths (CFPs) through them (often referred to as “short-circuits”) which prevents the spreading and slowing down of surface flow and therefore reducing their buffer sediment trapping efficiency. In the AnnAGNPS simulation process, each flow path passing through the buffer is evaluated and if upstream drainage area of a flow path is equal to or greater than the user-provided threshold value, the flow path’s local TE is set to zero, indicating that the flow path is simulated as CFPs. The impact of concentrated flow paths through riparian buffers were investigated by considering three alternative scenarios: upstream drainage area threshold of 50m^2, 5000m^2, and 5000m^2. Results indicate that the effectiveness of the buffer very between 25% to 29% for an upstream drainage area threshold that ranges between 50m^2 and 5000m^2, making it a less impactful variable in comparison to the riparian width that can increase the sediment trapping efficiency from 65% (for 10m width) to 81% (for 60m width) in the case of system 2 (Table 5).

Another useful representation of results is the use of ranked ratios of annual average sediment load per unit of area versus corresponding ranked drainage area. For example, in Figure 29, for the condition of no buffer (black line), 40% of the watershed contributes with approximately 75% of sediment yield. However, implementing constructed 10m riparian buffers in locations with high sediment producing AnnAGNPS cells (yellow line in Figure 29 for system 2) reduces the overall sediment load by approximately 15%. The presence of concentrated flow paths increases the overall watershed sediment load by approximately 25% (comparison between blue line and gray lines in Figure 29 for system 2). Additionally, it is important to note the differences in response between watersheds when same conservation scenario is applied.

Table 2: Riparian Buffer Simulation Results – System 1 - Lower Elk Fork

Simulation ID	*Location Description*	*Riparian Buffer Width*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
S2_ALL _10	All Streams	10m	2,090.147	**68%**
S2_ALL_30	All Streams	30 m	1,551.973	**76%**
S2_ALL_60	All Streams	60 m	1,555.262	**76%**
S2_AG_10	All streams adjacent to agricultural fields.	10 m	3,534.267	**46%**
S2_AG_30	All streams adjacent to agricultural fields.	30 m	3,228.49	**51%**
S2_AG_60	All streams adjacent to agricultural fields.	60 m	3,212.527	**51%**
S2_MED_10	All streams adjacent to agricultural fields with a sediment yield higher than 0.6 Mg	10 m	3,564.962	**46%**
S2_MED_30	All streams adjacent to agricultural fields with a sediment yield higher than 0.6 Mg	30 m	3,259.758	**50%**
S2_MED_60	All streams adjacent to agricultural fields with a sediment yield higher than 0.6 Mg	60 m	3,239.027	**51%**
S2_HIGH_10	All streams adjacent to agricultural fields with a sediment yield higher than 3.4 Mg	10 m	4,502.538	**32%**
S2_HIGH_30	All streams adjacent to agricultural fields with a sediment yield higher than 3.4 Mg	30 m	4,310.971	**35%**
S2_HIGH_60	All streams adjacent to agricultural fields with a sediment yield higher than 3.4 Mg	60 m	4,309.017	**35%**
S2_VH_10	All streams adjacent to agricultural fields with a sediment yield higher than 6.1 Mg	10 m	5,480.914	**17%**
S2_VH_30	All streams adjacent to agricultural fields with a sediment yield higher than 6.1 Mg	30 m	5,400.888	**18%**
S2_VH_60	All streams adjacent to agricultural fields with a sediment yield higher than 6.1 Mg	60 m	5,378.768	**18%**

Natural	Existing Buffer as delineated with the RCL tool	Variable	4,975.65	**24%**
Natural – 50m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 50 m^2	Variable	4,943.722	**25%**
Natural – 500m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 500 m^2	Variable	4,861.738	**26%**
Natural – 5000m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 5000 m^2	Variable	4,661.183	**29%**
Baseline Conditions	No buffer is integrated into the model	0 m	6,582.258	**0%**

Table 3: Riparian Buffer Simulation Results – System 1 – Red River

Simulation ID	*Location Description*	*Riparian Buffer Width*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
S2_ALL _10	All Streams	10m	3,298.30	**54%**
S2_ALL_30	All Streams	30 m	2,868.26	**60%**
S2_ALL_60	All Streams	60 m	2,880.13	**60%**
S2_AG_10	All streams adjacent to agricultural fields.	10 m	5,606.89	**22%**
S2_AG_30	All streams adjacent to agricultural fields.	30 m	5,492.71	**23%**
S2_AG_60	All streams adjacent to agricultural fields.	60 m	5,470.37	**24%**
S2_MED_10	All streams adjacent to agricultural fields with a sediment yield higher than 0.5 Mg	10 m	5,622.32	**22%**
S2_MED_30	All streams adjacent to agricultural fields with a sediment yield higher than 0.5 Mg	30 m	5,516.73	**23%**
S2_MED_60	All streams adjacent to agricultural fields with a sediment yield higher than 0.5 Mg	60 m	5,493.78	**23%**
S2_HIGH_10	All streams adjacent to agricultural fields with a sediment yield higher than 4.0 Mg	10 m	6,106.80	**15%**

S2_HIGH_30	All streams adjacent to agricultural fields with a sediment yield higher than 4.0 Mg	30 m	6,044.89	**16%**
S2_HIGH_60	All streams adjacent to agricultural fields with a sediment yield higher than 4.0 Mg	60 m	6,019.21	**16%**
S2_VH_10	All streams adjacent to agricultural fields with a sediment yield higher than 7.4 Mg	10 m	6,649.54	**7%**
S2_VH_30	All streams adjacent to agricultural fields with a sediment yield higher than 7.4 Mg	30 m	6,605.30	**8%**
S2_VH_60	All streams adjacent to agricultural fields with a sediment yield higher than 7.4 Mg	60 m	6,590.89	**8%**
Natural	Existing Buffer as delineated with the RCL tool	Variable	5,407.39	**25%**
Natural – 50m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 50 m^2	Variable	177,239.98	**14%**
Natural – 500m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 500 m^2	Variable	5,377.42	**25%**
Natural – 5000m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 5000 m^2	Variable	5,152.61	**28%**
Baseline Conditions	No buffer is integrated into the model	0 m	7,179.81	**0%**

Table 4: Riparian Buffer Simulation Results – System 1 – Spring Creek

Simulation ID	*Location Description*	*Riparian Buffer Width*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
S2_ALL _10	All Streams	10m	8,330.43	69%
S2_ALL_30	All Streams	30 m	5,477.88	80%
S2_ALL_60	All Streams	60 m	5,185.11	81%
S2_AG_10	All streams adjacent to agricultural fields.	10 m	15,240.99	44%
S2_AG_30	All streams adjacent to agricultural fields.	30 m	14,030.85	48%

S2_AG_60	All streams adjacent to agricultural fields.	60 m	13,804.65	49%
S2_MED_10	All streams adjacent to agricultural fields with a sediment yield higher than 1.2 Mg	10 m	15,572.45	42%
S2_MED_30	All streams adjacent to agricultural fields with a sediment yield higher than 1.2 Mg	30 m	14,446.03	47%
S2_MED_60	All streams adjacent to agricultural fields with a sediment yield higher than 1.2 Mg	60 m	14,237.86	47%
S2_HIGH_10	All streams adjacent to agricultural fields with a sediment yield higher than 3.2 Mg	10 m	18,910.93	30%
S2_HIGH_30	All streams adjacent to agricultural fields with a sediment yield higher than 3.2 Mg	30 m	18,179.08	33%
S2_HIGH_60	All streams adjacent to agricultural fields with a sediment yield higher than 3.2 Mg	60 m	18,009.42	33%
S2_VH_10	All streams adjacent to agricultural fields with a sediment yield higher than 5.1 Mg	10 m	23,262.73	14%
S2_VH_30	All streams adjacent to agricultural fields with a sediment yield higher than 5.1 Mg	30 m	23,040.19	15%
S2_VH_60	All streams adjacent to agricultural fields with a sediment yield higher than 5.1 Mg	60 m	22,943.50	15%
Natural	Existing Buffer as delineated with the RCL tool	Variable	21338.02	21%
Natural – 50m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 50 m^2	Variable	20881.77	23%
Natural – 500m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 500 m^2	Variable	20723.02	23%
Natural – 5000m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 5000 m^2	Variable	20262.37	25%

Baseline Conditions	No buffer is integrated into the model	0 m	27,080.50	0%

Table 5: Riparian Buffer Simulation Results - System 2

Simulation ID	*Location Description*	*Riparian Buffer Width*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
S2_ALL _10	All Streams	10m	71,312.717	**65%**
S2_ALL_30	All Streams	30 m	44,970.192	**78%**
S2_ALL_60	All Streams	60 m	39,745.744	**81%**
S2_AG_10	All streams adjacent to agricultural fields.	10 m	113,389.27	**45%**
S2_AG_30	All streams adjacent to agricultural fields.	30 m	99,602.65	**52%**
S2_AG_60	All streams adjacent to agricultural fields.	60 m	96,321.708	**53%**
S2_MED_10	All streams adjacent to agricultural fields with a sediment yield higher than 3.7 Mg	10 m	118,246.48	**43%**
S2_MED_30	All streams adjacent to agricultural fields with a sediment yield higher than 3.7 Mg	30 m	105,205.77	**49%**
S2_MED_60	All streams adjacent to agricultural fields with a sediment yield higher than 3.7 Mg	60 m	101,905.84	**51%**
S2_HIGH_10	All streams adjacent to agricultural fields with a sediment yield higher than 14.2 Mg	10 m	178,229.71	**13%**
S2_HIGH_30	All streams adjacent to agricultural fields with a sediment yield higher than 14.2 Mg	30 m	175,074.39	**15%**
S2_HIGH_60	All streams adjacent to agricultural fields with a sediment yield higher than 14.2 Mg	60 m	174,123.99	**15%**
S2_VH_10	All streams adjacent to agricultural fields with a sediment yield higher than 19.4 Mg	10 m	195,271.08	**5%**
S2_VH_30	All streams adjacent to agricultural fields with a sediment yield higher than 19.4 Mg	30 m	194,101.35	**6%**

S2_VH_60	All streams adjacent to agricultural fields with a sediment yield higher than 19.4 Mg	60 m	193,756.37	**6%**
Natural	Existing Buffer as delineated with the RCL tool	Variable	178,956.221	**13%**
Natural – 50m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 50 m^2	Variable	177,239.98	**14%**
Natural – 500m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 500 m^2	Variable	175,173.82	**15%**
Natural – 5000m2	Existing Buffer as delineated with the RCL tool with a drainage area minimum of 5000 m^2	Variable	169,296.10	**18%**
Baseline Conditions	No buffer is integrated into the model	0 m	205,880.221	**0%**

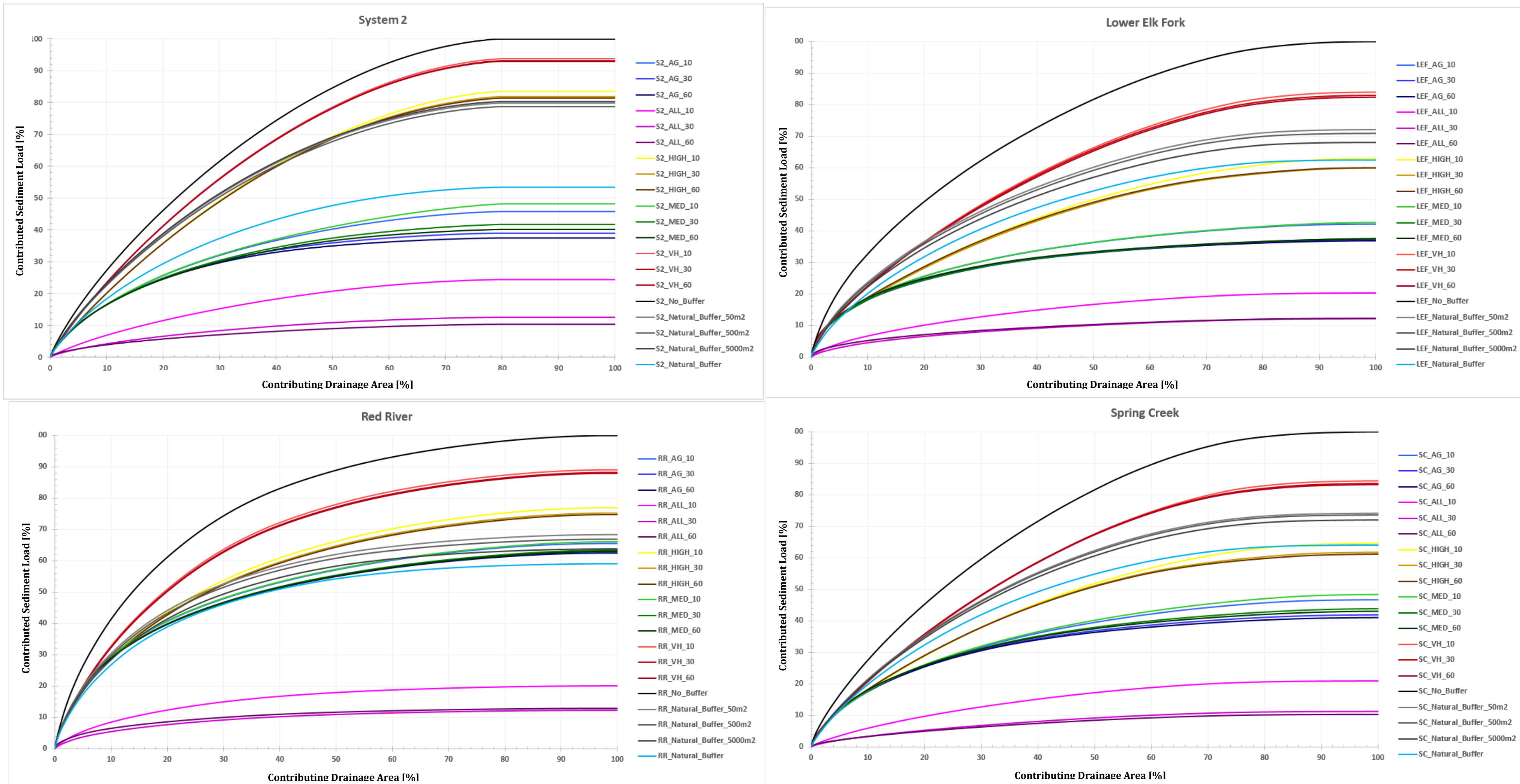


Figure 29: Simulated Riparian Buffer Scenarios Results Graphical Representation

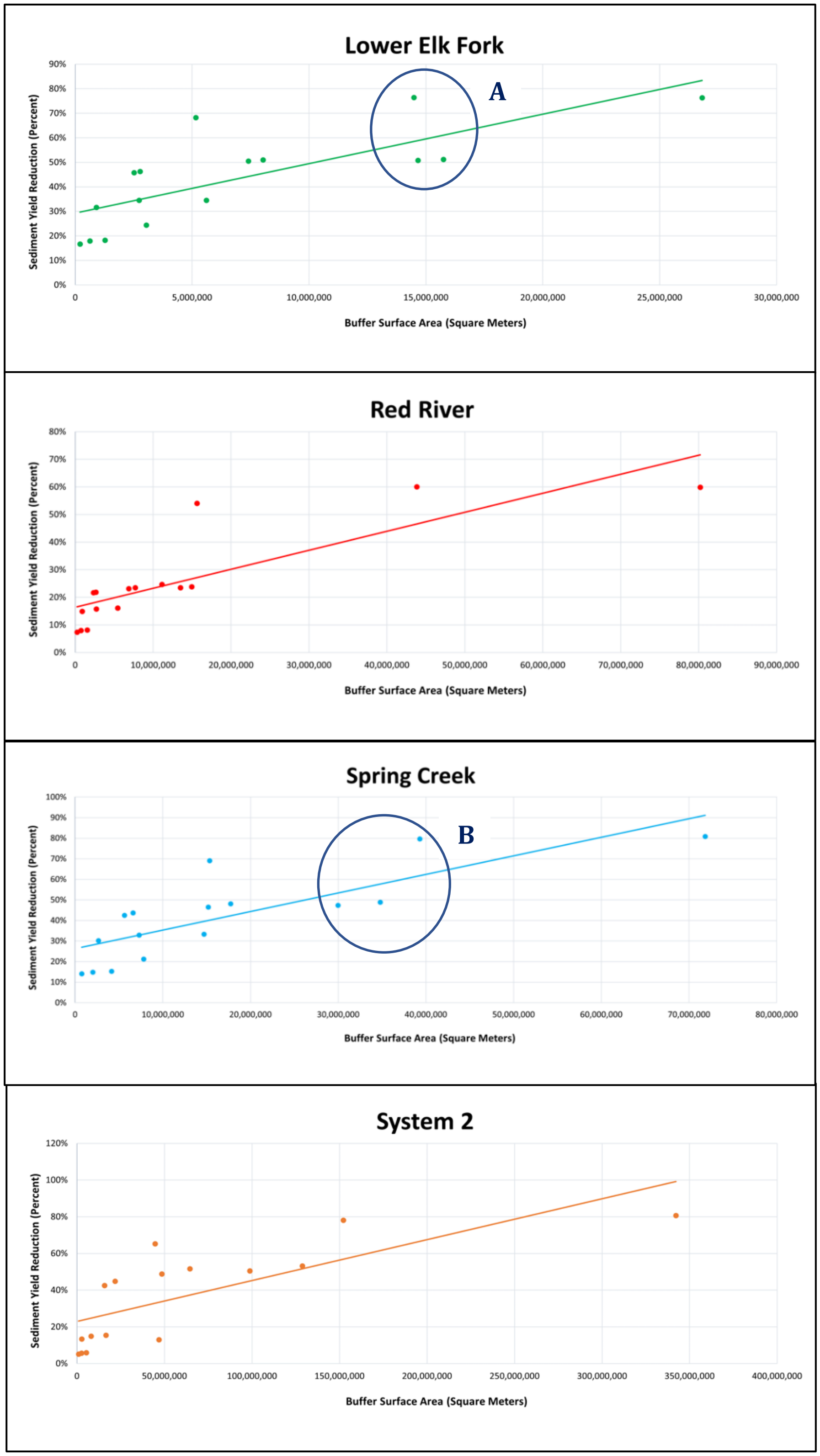


Figure 30: Buffer surface area and sediment reduction relationship in the studied watersheds)

4.2 Sediment Basin (code 350)

4.2.1 Description of the practice:

NRCS describes the sediment basin code 350 as a "basin constructed with an engineered outlet, formed by an embankment or excavation or a combination of the two." A sediment basin is built to help capture sediment runoff for a sufficient length of time to allow settlement of sediment and other suspended solids (Figure 31). These basins are constructed in agricultural and other disturbed areas. Sediment basins act as the last line of defense for capturing sediment.

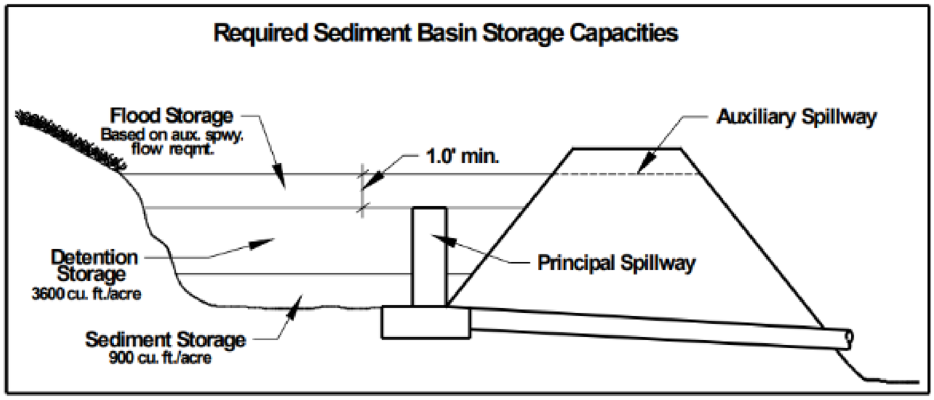


Figure 31: Sediment Basin Practice Standard. Image source: NRCS-Conservation Practice Standard (nrcs.usda.gov) Code 350

This practice was not simulated in system 1 due to the high concentration of Karst features including sinkholes. This conservation practice was simulated in system 2 using the AnnAGNPS model based on GIS characterization using the AgWET model.

4.2.2 Description of AgWET:

AgWet is an extension of the TopAnGNPS computer program (distributed with the AGNPS system). It is a GIS based tool that is used to identify and describe natural or artificial sediment basins, ponds, or wetlands. The tool can automate the characterizations of basins at a watershed-scale. The extra inputs needed for AgWet are the basin barrier locations and the basin barrier height. The output of the tool is a basin data section that can be used by AnnAGNPS. AgWet was used to determine the existing basin location's area and corresponding reach.

4.2.3 Delineation of existing sediment basins:

The existing sediment basins were digitally delineated and described using the AgWet tool. The delineation was conducted by examining and researching all the water bodies in the watershed using remotely sensed imagery. In system 2, 57 existing ponds met the sediment basin criteria. Once the sediment basins were described using the AgWet program, they were compared to the manually delineated basins for calibration. AgWet produces the surface area for each basin. For most of the sediment basins, AgWet simulated the correct area. However, some of the sediment basins areas were underestimated using topography-based simulations. The area for each basin was determined from the combination of both remotely sensed and AgWet calculations. Figure 32 shows the distribution of the existing sediment basins in system 2.

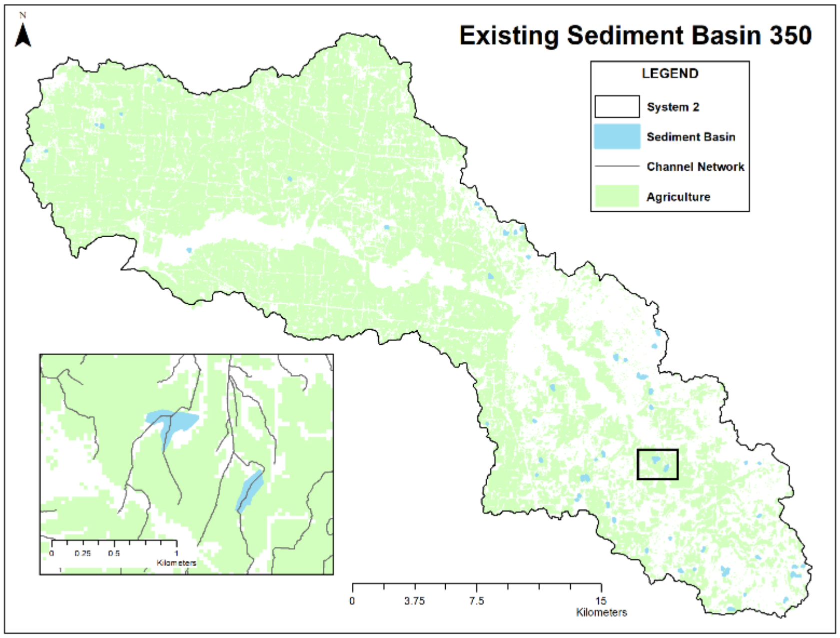


Figure 32: Distribution of Existing Sediment Basins in System 2

4.2.4 *Determination of optimal locations for future sediment basins:*

The potential sediment basin locations were selected using three criteria (flowchart in Figure 33).

1. Channel stream order (1 or 2)
2. Channel stream length.
3. AnnAGNPS sub-catchment (representing field) sediment yield into streams

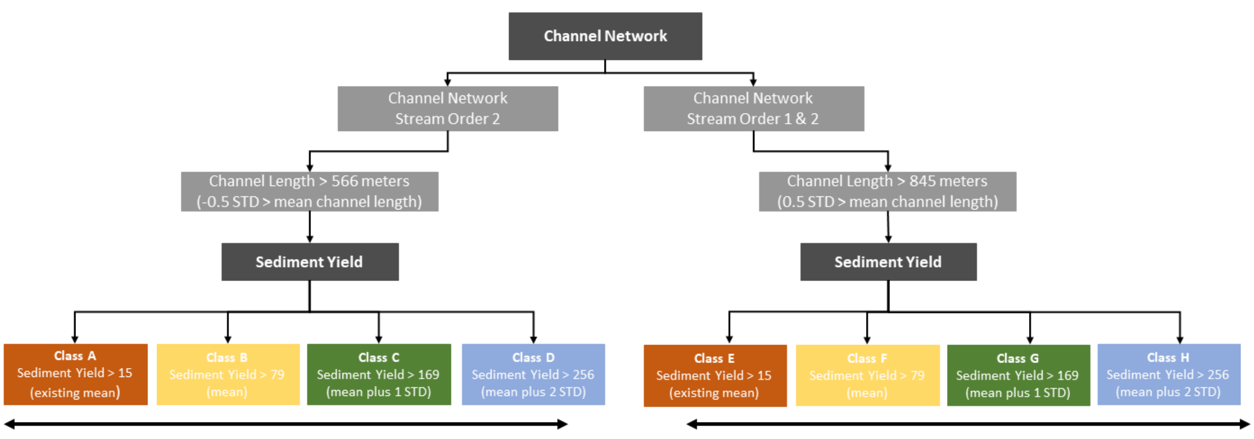


Figure 33: Flow Chart summarizing the used criteria for the determination of future sediment basin locations and their corresponding scenario

Sediment yield was classified based on the mean sediment yield of existing sediment basins. The goal is to replicate the same conditions of existing sediment basins and expand their impact by selecting more locations (Figures 34 and 35).

The least restrictive classes (i.e. classes A and E) contain all the locations that are greater than 15 cell-yield. The next classes (i.e. B and F) contain all the locations that are greater than 79 cell-yield. The third classes (i.e. C and G) contain all the locations that are greater than 169 cell-yield. The last and most restrictive classes (D and H) contain all the locations that are greater than 256 cell-yield. The resulting locations number and upstream area per class are shown in Table 6.

Table 6: Proposed sediment basin number and upstream area per scenario (i.e. Classes from previous flowchart)

Scenario	Number of Suggested Sediment Basins	Upstream area for each class in Hectare
Class A	233	8,306.80
Class B	147	3,155.13
Class C	67	563.78
Class D	21	199.86
Class E	527	2,212.82
Class F	297	1,439.84
Class G	126	1,076.08
Class H	54	497.35

Basin characteristics (i.e. weir height and area) were determined as follow:

- Basin weir height: the weir height for each proposed sediment basin was determined manually using the local topography data and satellite imagery.
- Basin area: after running AgWet on the potential sediment basin sites, the area generated by AgWet was either too small or too large in a subset of the proposed basins relatively to the existing sediment basins. Hence, the existing basins area statistics were used as a guide. The slope of the best-fit line for the existing sediment basins’ area is 0.0375. The drainage area file generated by TopANGPS was used to calculate the drainage area for each potential sediment basin and was calibrated using the 0.0375 factor.

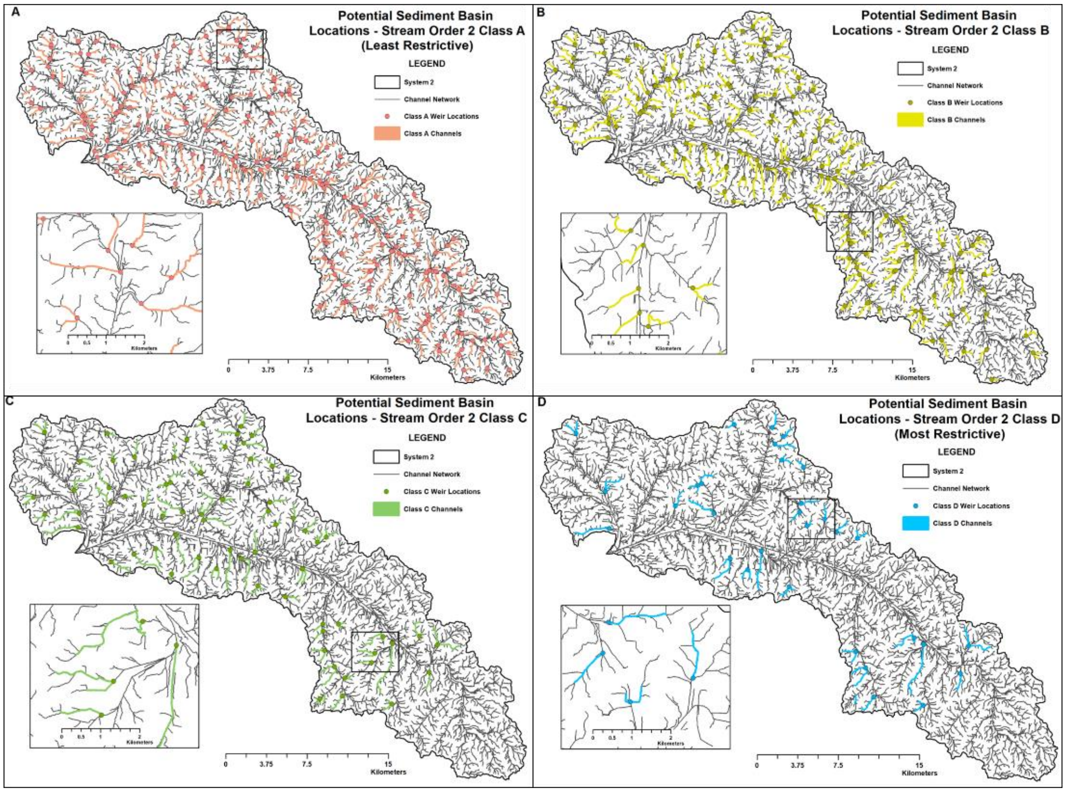


Figure 34: Selected sediment basin locations corresponding to scenarios A, B, C and D of the flowchart in figure 33

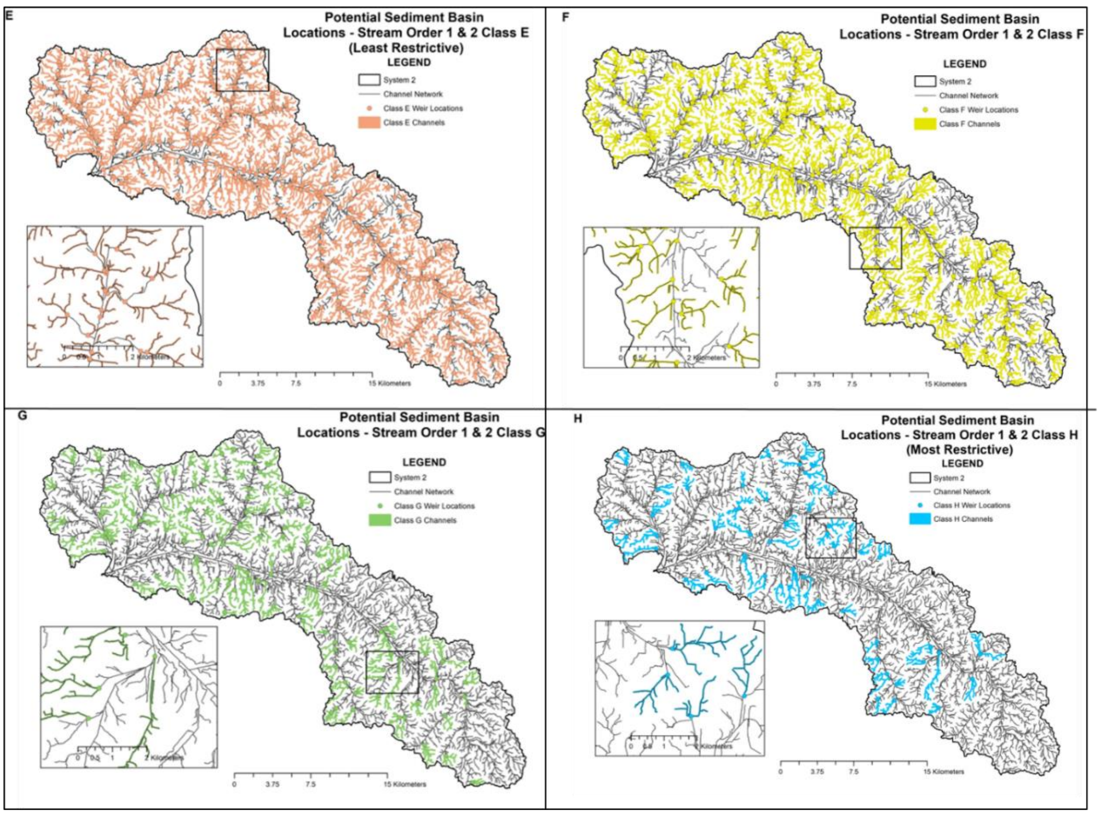


Figure 35: Selected sediment basin locations corresponding to scenarios E, F, G and H of the flowchart in figure 33

4.2.5 Simulation Results

Table 7 summarizes the results of the sediment basin simulations.

Sediment retention ponds work by reducing surface flow energy and promoting fine sediment deposition, however downsides to the approach is that they remove land out of production. Most sediment retention ponds are recommended to be placed in proximity of stream orders 1 and 2. The size of sediment retention ponds is usually associated with the drainage area upstream of it.

Our simulation results indicate the wide range of sediment load reduction from 3% to 95% showing the potential of sediment ponds in reducing sediment load (Table 7). Their efficiency is a function of the number of ponds and the selection of their placement throughout the watershed. Alternative scenarios A and B that focus on stream order 2 seem significantly more efficient than scenarios E and F. For example, scenario B was simulated using optimally located 147 ponds, leading to a reduction of 66% in sediment load whereas scenario F, based on more ponds (i.e. 297) leads to a reduction of only 13%. Based on these findings, it is suggested to prioritize the size of the upstream area (Table 6) when deciding where to place sediment retention ponds in the watershed.

Simulation results describing existing basins, which were delineated and digitized (57 basins), did not provide any reduction (about 0%). Another example of the importance of the appropriate selection of the location is that scenario C with 67 ponds provides a 10% reduction and scenario D (21 basins) provides a 3% reduction.

Table 7: Sediment basin simulation scenarios results for System 2

Scenario Classification (Figure 33)	Total Sediment Load (Mg)	Sediment Reduction (%)
Class A	10,699.12	**95%**
Class B	70,137.00	**66%**
Class C	184,607.95	**10%**
Class D	199,254.02	**3%**
Class E	164,705.46	**20%**
Class F	179,666.38	**13%**
Class G	179,073.71	**13%**
Class H	198,661.18	**4%**
Existing Ponds	204,992.04	**0%**
Baseline conditions	205,880.221	**0%**

4.3 Conservation Crop Rotation (code 328)

4.3.1 Description of the practice:

Conservation crop rotation is applied as a seasonal sequence of crops grown in the same field yielding a multi-crop rotation cycle. The crop rotation needs to include a minimum of two different crops. The crops in the rotation should include a high residue producing crop such as wheat or corn along with a low residue producing crop such as soybeans or vegetables. Conservation crop rotation has many benefits which include reduce sheet, rill, and wind erosion, increase soil health and organic matter content, and improve soil moisture efficiency.

4.3.2 Considered Scenarios:

The only major crop in our study area that allowed rotation is soybeans, so we simulated the soybeans/winter wheat rotations in four location scenarios (Figure 36):

- All soybean agricultural fields
- Soybean agricultural fields that have a sediment yield higher than mean - standard deviation (referred to as "> medium")
- Soybean agricultural fields that have a sediment yield higher than mean (referred to as "> high")
- Soybean agricultural fields that have a sediment yield higher than mean + standard deviation (referred to as "> Very High")

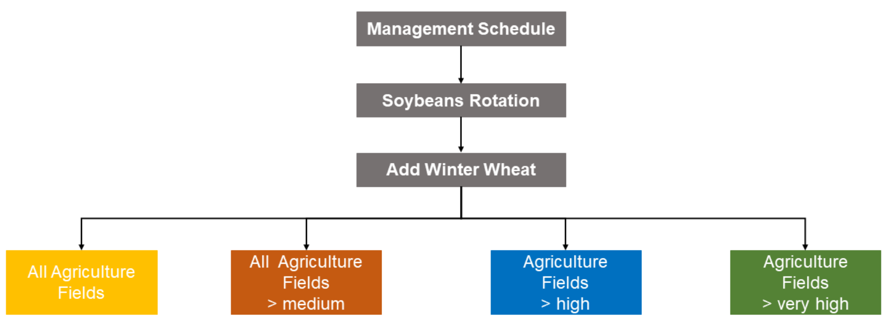


Figure 36: Flow Chart summarizing the used criteria for the determination of crop rotation scenarios

The number of AnnAGNPS cells that fit the selected criteria are shown in Table 8.

Table 8: Selected AnnAGNPS cell number per scenario for the crop rotation simulations

Number of cells selected to be altered	Lower Elk Fork	Red River	Spring Creek	System 2
All Agricultural Fields	1335	1094	1879	7340
Medium or higher Erosion Fields	1247	1028	1668	5689
High or higher Erosion Fields	478	431	854	1166
Very High Erosion Fields	120	134	273	424

4.3.3 Simulation Results:

Simulated results are summarized in Tables 9-12. Crop rotation between soybean and winter wheat would reduce sediment load by about 5% for Red River (while soybean fields represent about 6%

of the total area of the watershed), by about 7% for Lower Elk (while soybean fields represent about 6% of the total area of the watershed), by about 7% for Spring Creek (while soybean fields represent about 6% of the total area of the watershed) and by about 11 % for System 2 (while soybean fields represent about 27% of the total area of the watershed). Overall, it is a relatively small reduction, but a cost-effective alternative since it does not reduce production area.
Per discussion with NRCS collaborators, the only crop rotation identified in these watersheds was soybean with winter wheat. A more widespread application of crop rotation could further decrease the sediment yield in these systems.

Table 9: Crop rotation simulation results for Spring Creek

Simulation ID	***Location Description***	***Total Sediment Load Mg***	***Sediment Reduction (%)***
All Ag Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed	25,212.84	**7%**
Medium Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 1.2 Mg	25,578.39	**7%**
High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 3.2 Mg	25,210.62	**6%**
Very High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 5.1 Mg	26,498.64	**2%**
Baseline Conditions	No crop rotation is integrated into the model	27,080.50	**0%**

Table 10: Crop rotation simulation results for Red River

Simulation ID	***Location Description***	***Total Sediment Load Mg***	***Sediment Reduction (%)***
All Ag Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed	6,798.11	**5%**
Medium Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 0.5 Mg	6,916.69	**5%**

High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 4.0 Mg	6,800.18	4%
Very High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 7.4 Mg	7,038.45	2%
Baseline Conditions	No crop rotation is integrated into the model	7,179.81	0%

Table 11: Crop rotation simulation results for Lower Elk Fork

Simulation ID	***Location Description***	***Total Sediment Load Mg***	***Sediment Reduction (%)***
All Ag Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed	6,118.82	7%
Medium Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 0.6 Mg	6,261.46	7%
High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 3.4 Mg	6,119.13	5%
Very High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 6.1 Mg	6,457.86	2%
Baseline Conditions	No crop rotation is integrated into the model	6,582.258	0%

Table 12: Crop rotation simulation results for System 2

Simulation ID	***Description***	***Total Sediment Load Mg***	***Sediment Reduction (%)***
All Ag Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed	182,278.44	11%
Medium Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed	193,476.40	11%

	with a sediment yield higher than 3.7 Mg		
High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 14.2 Mg	182,969.31	6%
Very High Erosion Fields (S to WwS)	Crop rotation is applied to all soybean fields in the watershed with a sediment yield higher than 19.4 Mg	197,905.31	4%
Baseline Conditions	No crop rotation is integrated into the model	205,880.221	0%

4.4 Conservation Reserve Program (CRP)

4.4.1 Description of the practice

The Conservation Reserve Program (CRP) is a voluntary program implemented by USDA Farm Service Agency (FSA). The program works with agriculture producers to establish long-term conservation benefits to agriculture land. This land is not farmed or ranched but planted with native plant species that will improve the long-term environmental health and quality of the land. The goal of this program is to remove environmentally fragile land from agricultural production and overall improve water quality, wildlife habitat, and prevent soil erosion. The contract between agriculture producers and FSA lasts between ten to fifteen years with yearly rental payments and cost share assistance provided by FSA.

4.4.2 Considered Scenarios

The simulation period for this study is 11 years (Jan 2008 – Dec 2018). This period fits in the criteria of the CRP that ranges in time between 10 to 15 years. For each scenario (Figure 37), we changed the crop schedule and type to grass during the simulation period for the selected watershed. The selected watersheds under each scenario are:

- All agricultural fields (Figures 24-27 -part **A**)
- All agricultural fields that have a sediment yield higher than mean - standard deviation (referred to as "> medium") (Figures 24-27 – part **B**)
- All agricultural fields that have a sediment yield higher than mean (referred to as "> high") (Figures 24-27 – part **C**)
- All agricultural fields that have a sediment yield higher than mean + standard deviation (referred to as "> Very High") (Figures 24-27 – part **D**)

The agriculture field's entire crop rotation schedule was changed from the actual rotation to grass. The grass rotation simulates the land being returned to native plant species with not cultivation.

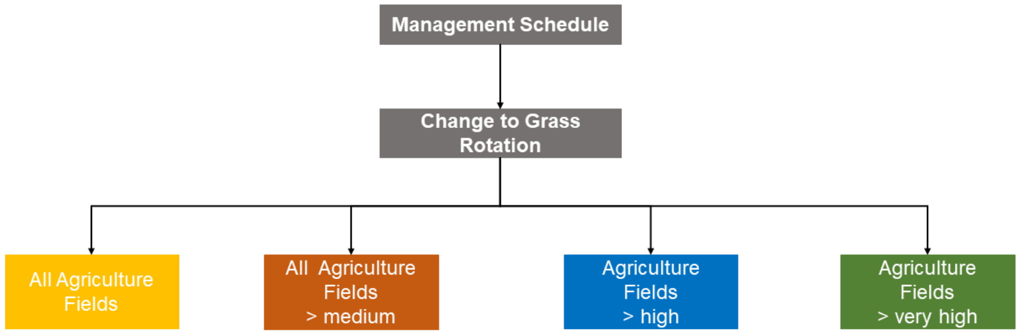


Figure 37: Flow Chart summarizing the used criteria for the CRP scenarios

4.4.3 Simulation Results

Simulated results are summarized in Tables 13-16. Based on how many fields are removed from production, sediment reduction ranges between 10 and 81% for system 2, between 28 and 80% for Lower Elk, between 22 and 73% for Spring Creek and between 10 and 33% for Red River. The CRP-based sediment reduction is relatively low in Red River due to the small percentage of agricultural land in Red River in comparison to the other watersheds (i.e. 20% for Red River as opposed to 53% for Spring Creek, 59% for Lower Elk Fork and 66% for system 2). In addition to the number of fields, their location selection is very significant. The identification of high sediment producing fields is crucial to the CRP implementation. For instance, it makes no difference in Red River to implement CRP in all fields (2686 hectares) or medium sediment yield fields (2128 hectares). The result in term of sediment relative reduction is the same.

Table 13: CRP simulation results for Spring Creek

Simulation ID	*Location Description*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
All Ag (11-year grass)	CRP is applied to all agricultural fields in the watershed	7,265.48	**73%**
Medium (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 1.2 Mg	13,387.94	**72%**
High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 3.2 Mg	7,592.09	**51%**

Very High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 5.1 Mg	21,134.13	**22%**
Baseline Conditions	No crop rotation is integrated into the model	27,080.50	**0%**

Table 14: CRP simulation results for Red River

Simulation ID	*Location Description*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
All Ag (11-year grass)	CRP is applied to all agricultural fields in the watershed	4,830.69	**33%**
Medium (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 0.5 Mg	5,687.41	**33%**
High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 4.0 Mg	4,840.74	**21%**
Very High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 7.4 Mg	6,459.54	**10%**
Baseline Conditions	No crop rotation is integrated into the model	7,179.81	**0%**

Table 15: CRP simulation results for Lower Elk Fork

Simulation ID	*Location Description*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
All Ag (11-year grass)	CRP is applied to all agricultural fields in the watershed	1,285.12	**80%**
Medium (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 0.6 Mg	2,893.13	**80%**
High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 3.4 Mg	1,294.60	**56%**
Very High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 6.1 Mg	4,749.58	**28%**

Baseline Conditions	No crop rotation is integrated into the model	6,582.26	**0%**

Table 16: CRP simulation results for System 2

Simulation ID	*Description*	*Total Sediment Load Mg*	*Sediment Reduction (%)*
All Ag (11-year grass)	CRP is applied to all agricultural fields in the watershed	38,598.09	**81%**
Medium (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 3.7 Mg	154,930.80	**78%**
High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 14.2 Mg	44,460.16	**25%**
Very High (11-year grass)	CRP is applied to all agricultural fields in the watershed with a sediment yield higher than 19.4 Mg	186,290.72	**10%**
Baseline Conditions	No crop rotation is integrated into the model	205,880.22	**0%**

5 Technical Details on the Riparian Classification from LiDAR (RCL) Tool and Tutorial

5.1 Goals

The main goal of this component of the project is to create a reproducible workflow for classifying landcover in the riparian buffer using only raw LiDAR point cloud as input. We have termed such a workflow as Riparian Classification from LiDAR (RCL) model. A specific pre-trained RCL model has been packaged for ease of use as an ArcGIS tool.

While both aerial imagery and LiDAR have been used for years to train landcover classification models, the specifics on how to properly create training data (reference data), implementing a machine-learning model, and then evaluate the quality of the output are typically unclear to those unfamiliar with landcover classification. Thus, we have endeavored to make the RCL model as robust as possible to variations in study area physiography and LiDAR collection methods so that the final model is generalizable. This allows the model to be agnostic of how or where the input data is collected and can produce a reliable output without exhaustive pre-input preparation and analysis.

5.2 Datasets

Light Detection and Ranging, or LiDAR, uses laser pulses to measure the distance between a sensor and a target. Upon striking the target, the laser pulse is reflected, and the sensor records the time between the pulse leaving the LiDAR apparatus and returning to the sensor. This information can be used to determine the distance between the sensors and object. In coordination with a GPS system and when mounted on aircraft, LiDAR systems can be used to generate detailed representations of the ground below the flightpath. These datasets, called *point clouds*, typically contain millions of points each representing a *return* (reflected laser pulse). Each point is associated with geographic information (northing, easting, elevation) as well as ancillary information, such as the *intensity* of the return (a measure of how reflective the target is) and the *return number*. Individual laser pulses from the LiDAR apparatus spread as they travel, resulting in an ever-widening *footprint*. If the footprint strikes an object with multiple elevation levels (such as a building edge or tree branches), then the originating pulse will return in multiple pulses rather than one discrete pulse; the order that the split pulses return determines the return number for each returned pulse. Both return splitting and intensity are sensitive to the type of LiDAR apparatus used and flight characteristics. LiDAR vendors will occasionally classify landcover using proprietary models, but the classification methods vary in quality and detail.

Traditionally three types of elevation models are generated from aerially collected LiDAR point clouds. A Digital Surface Model (DSM) is created by generating an interpolated surface using only the first returns in the point cloud, and roughly represents the elevation of the ground or, if present, aboveground structures such as trees and power lines. A Digital Elevation Model (DEM) is created by interpolating only the last returns in a point cloud, and roughly represents the elevation of the ground without any aboveground structures. Occasionally points classified as buildings will be excluded when generating a DEM, but vendor classifications cannot always be relied on for this. The third elevation product, the Digital Height Model, is the difference between the DSM and the DEM. This roughly represents the height of aboveground structures. Because the DHM is often used as a proxy for canopy height, it is sometimes called the Canopy Height Model (CHM).

These data products, as well as those derived from them such as slope models and Haralick textures can be used to train machine-learning algorithms to identify the unique morphometric signatures associated with different landcover types. Creating a reproducible workflow to do this is a key goal of this project.

5.3 Methods

A watershed in western Tennessee (HUC 080102040304) was the primary focus for this study. Data for a 2012 LiDAR mission entirely covering the watershed was obtained and used to create multiple data products, including but not limited to a DEM, DSM, DHM, and derived slope models. Approximately 6.4% of the watershed's landcover was manually classified into one of 19 categories (Table 17), which was then used along with the LiDAR-derived data to train a decision tree to classify landcover. This process was repeated for nine additional watersheds across the continental US (Table 18) in order to evaluate the effects of physiography and LiDAR vendor on model validity; training coverage accounted for approximately 9.2% of all land in the study areas. The general classification model was trained using a portion of the training data from seven of the 10 watersheds; the model was then validated against the unused portion of the training data from those seven watersheds and the three-naïve watersheds.

Table 17: Landcover classes used to train the model. The more specific categories (left column) were grouped together (right column) in order to improve model quality before training

Category	**Reclassification**
Forest	Trees
Linear Trees	Trees
Individual Trees/Small Clusters	Trees
Building Tops	Other
Building Edges	Other
Dirt/Bare Field	Other
Crops	Herbaceous Vegetation
Rough Vegetation	Herbaceous Vegetation
Other Impervious Surfaces	Other
Water	Other
Snow	Other
Bare Rock	Other
Sand	Other
Wetlands	Herbaceous Vegetation

Power Lines	Other
Charred Trees and Vegetation	*(excluded)*
Utility Easement	Herbaceous Vegetation
Large-Scale Urban	*(excluded)*
Canyon	Other

Table 18: Locations used to generate the model. The western TN HUC (080102040304) was of particular focus for this study, but other watersheds were used to ensure the applicability of the model in various environments

HUC12	**180500020905 San Francisco, CA**	**070801050901 Kelley, IA**	**130202090102 Alamo, NM**	**080102040304 Western TN**	**010500021301 Penobscot, ME**	**030902040303 Naples, FL**	**140801040103 Central CO**	**080902030201 New Orleans, LA**	**100301011309 Helena, MT**	**102901110304 Freeburg, MO**	**TOTAL**
EPSG	26910	26915	26913	3723	26919	2777	26913	26945	2818	26997	
Status	Trained	Trained	Trained	Trained	Trained	Trained	Trained	Naïve	Naïve	Naïve	
Type	Urban	Ag	Desert	Ag	Coast	Coast	Mts	Urban	Mts	Ag	
Size (sq km)	25.3	50.7	66.1	161.1	80.0	74.7	136.8	61.2	55.5	77.8	**789.2**
Trained (sq km)	1.3	8.8	7.7	10.3	10.4	8.1	13.2	4.0	3.2	6.1	**72.9**
Training Fraction	5.1%	17.3%	11.6%	6.4%	13.0%	10.8%	9.6%	6.5%	5.8%	7.8%	**9.2%**
F1-Score	81%	93%	91%	91%	95%	94%	87%	79%	88%	95%	**90% (Trained)** **87% (Naive)**

A TN-specific model trained only on data derived from the west TN watershed (HUC 080102040304) as well as a general model trained on all watersheds were created. The general model was allowed access to all derived datasets, but the TN model was restricted to datasets that can be easily reproduced using ArcGIS software and ArcPy. The full explanation of data types used in each model can be found in Table 19 and Figure 38. Additionally, though 18 landcover types were manually classified, model output was restricted to two classes (binary models: trees and all other) or three classes (ternary models: trees, herbaceous vegetation, and all other) due to insufficient differences in generalizable LiDAR signature between most classes.

Table 19: Summary of LiDAR-derived raster products used in the model. A total of 26 products were generated. Of which, only five derived products were selected in the final decision tree.

Feature type	**Raster grid name**	**Used to train model**	**Appears in decision tree**
Elevation-based features	Digital surface model	No	No
	Digital elevation model	No	No
	Digital height model	No	No
	Filtered digital height model	Yes	No
Non-elevation-	Intensity raster	No	No
	Return raster	No	No

based features	Slope rasters (from surface, elevation, height and filtered height models)	Yes	Yes (from DEM and DSM)
Textural features	Roughness rasters (from surface, elevation, height and filtered height models)	Yes	Yes (from fDHM and DSM)
	Laplacian filtered rasters (from surface, elevation, height and filtered height models)	Yes	Yes (from DSM)
	Haralick textures (generated for digital height model only)	No	No

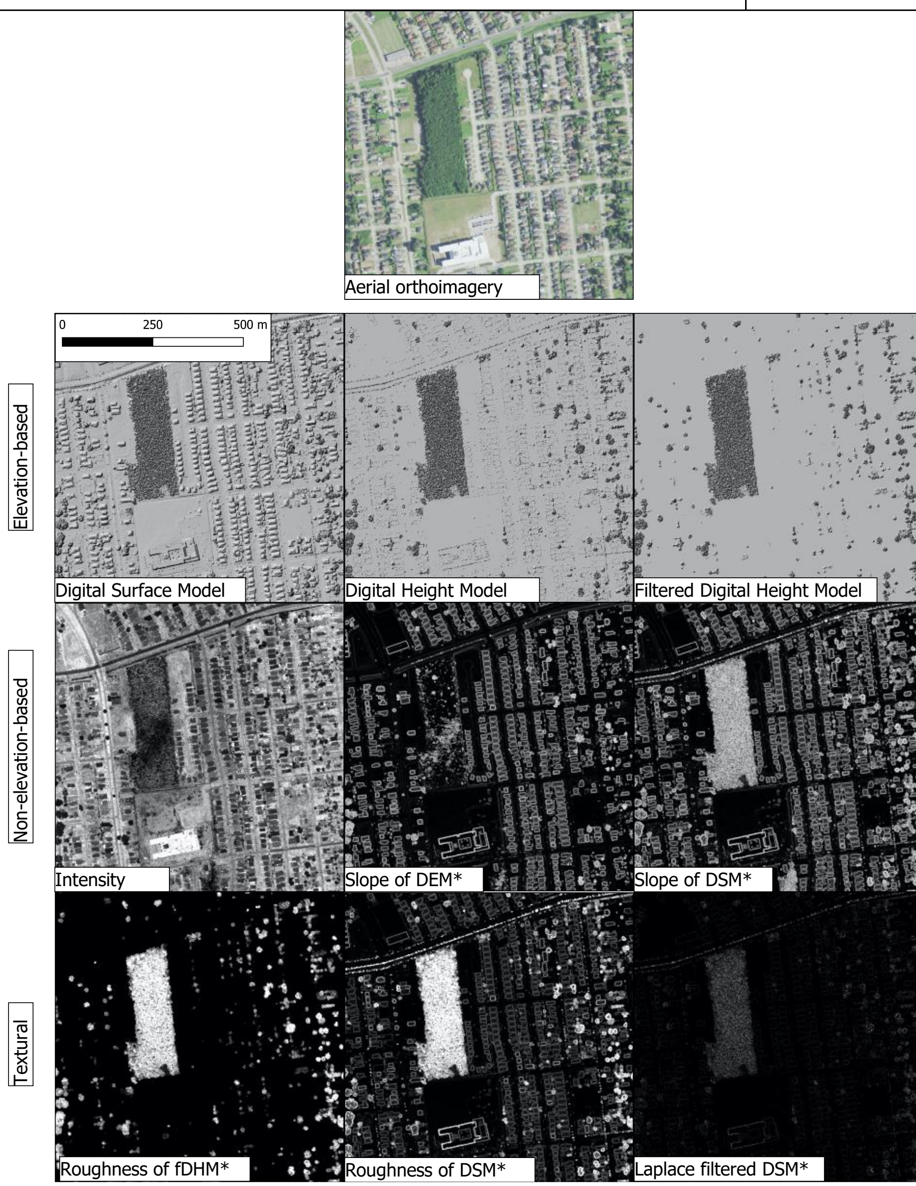


Figure 38: Representation of selected raster grids derived from LiDAR as an example of the generated inputs. Note that the elevation-based raster grids are displayed as hillshade

Though this model is primarily intended to classify riparian landcover, landcover throughout the entirety of the watersheds was classified. Doing so increases the amount of training data available to the model and diversifies the LiDAR-signatures encountered during training. The overall used methodology is summarized in Figure 39.

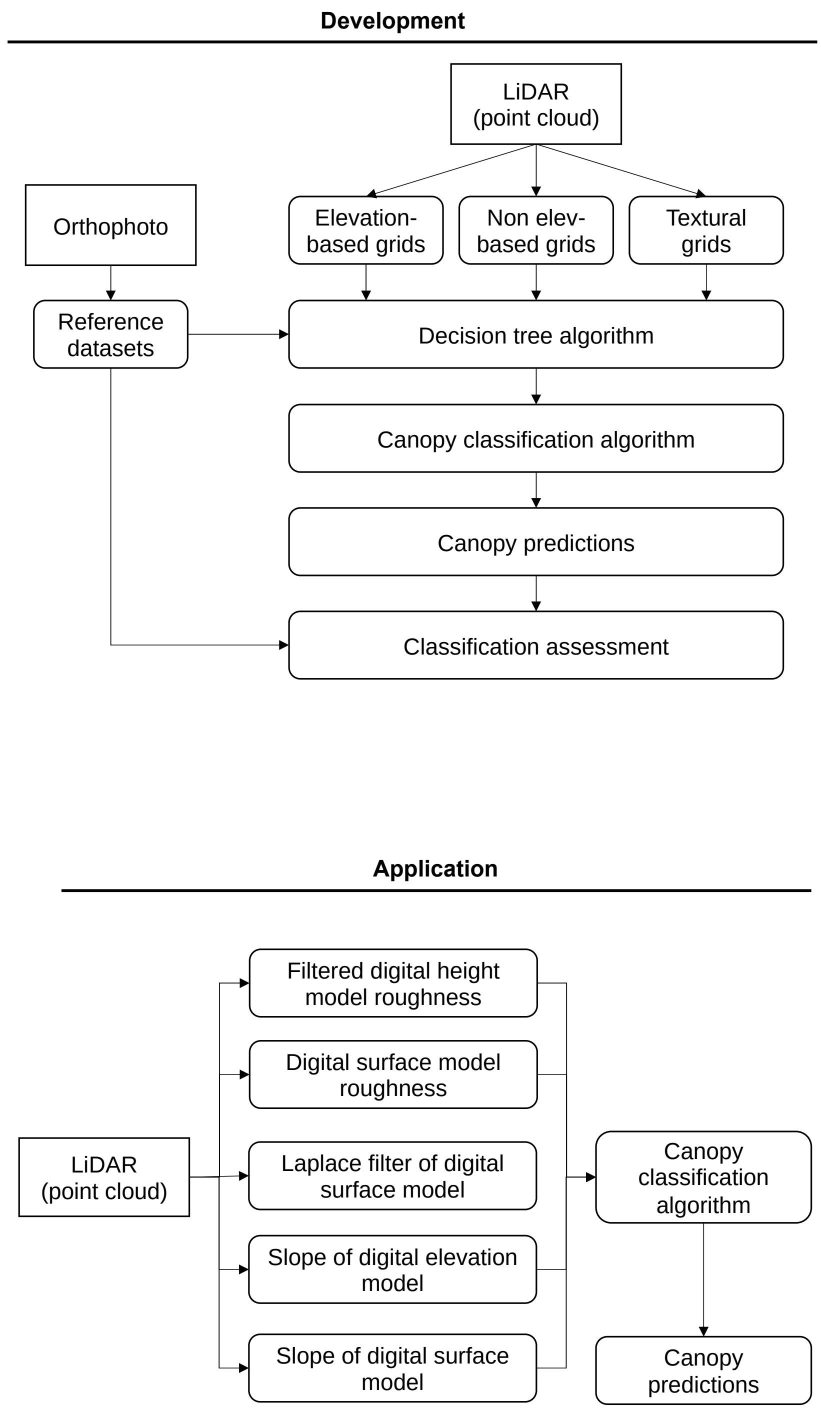


Figure 39: Summary of the design and implementation of the classification model

5.4 Results and Discussion

The decision trees for a ternary and binary, western Tennessee-specific (HUC 080102040304) ArcGIS-friendly models are shown in Figures 40 and 41. Decision trees for general models are not included in this report due to the size and complexity of the graphs. Quality metrics for the TN-specific models are available in Table 20. “ArcGIS-friendly” means the model uses data products that are easily generated using native ArcGIS tools. A sample classification is provided in Figure 42.

Table 20: Quality metrics for the TN-specific binary and ternary models. The TN model was training using only data from HUC 080102040304, while the general models were trained on 7 different watersheds. The TN model (both ternary and binary) are what is packaged in the RCL tool

TN, Ternary			
	Precision	**Recall**	**F1-Score**
Trees	97.3%	91.4%	94.3%
Herb. Veg.	49.5%	61.7%	55.0%
Other	91.2%	88.7%	89.9%
Accuracy		85.8%	
Macro Avg.	79.4%	80.6%	79.7%
Weighted Avg.	87.3%	85.8%	86.4%
TN, Binary			
	Precision	**Recall**	**F1-Score**
Trees	92.1%	97.3%	94.6%
Other	98.8%	96.2%	97.4%
Accuracy		96.5%	
Macro Avg.	95.4%	96.8%	96.0%
Weighted Avg.	96.7%	96.5%	96.6%

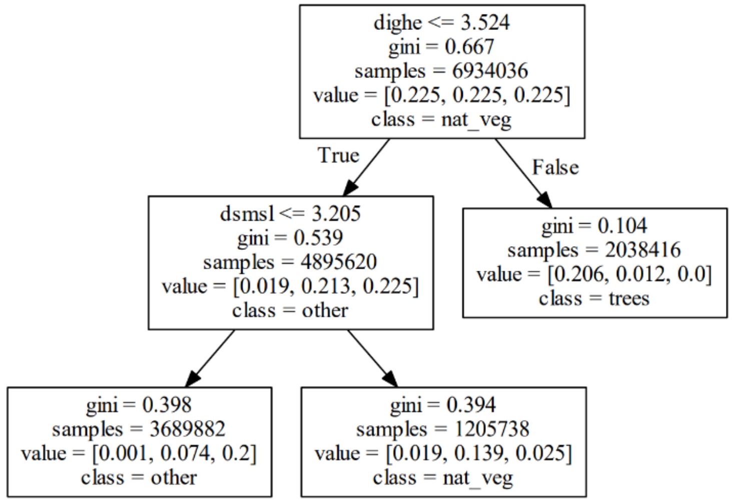


Figure 40: The decision tree for the ternary, western Tennessee specific, ArcGIS friendly model. This model and its binary counterpart are packaged in the RCL tool. dighe = DHM, dsmsl = slope of DSM

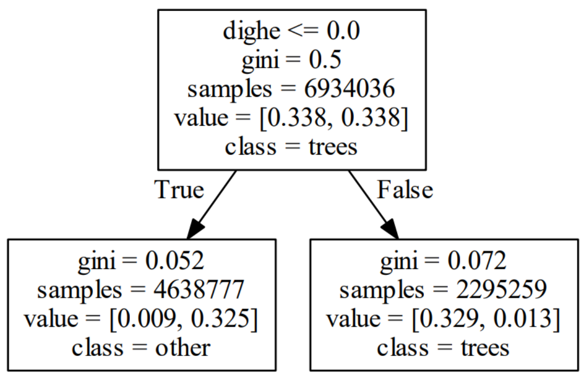


Figure 41: The decision tree for the binary, western Tennessee specific, ArcGIS friendly model. This model and its ternary counterpart are packaged in the RCL tool. dighe = DHM

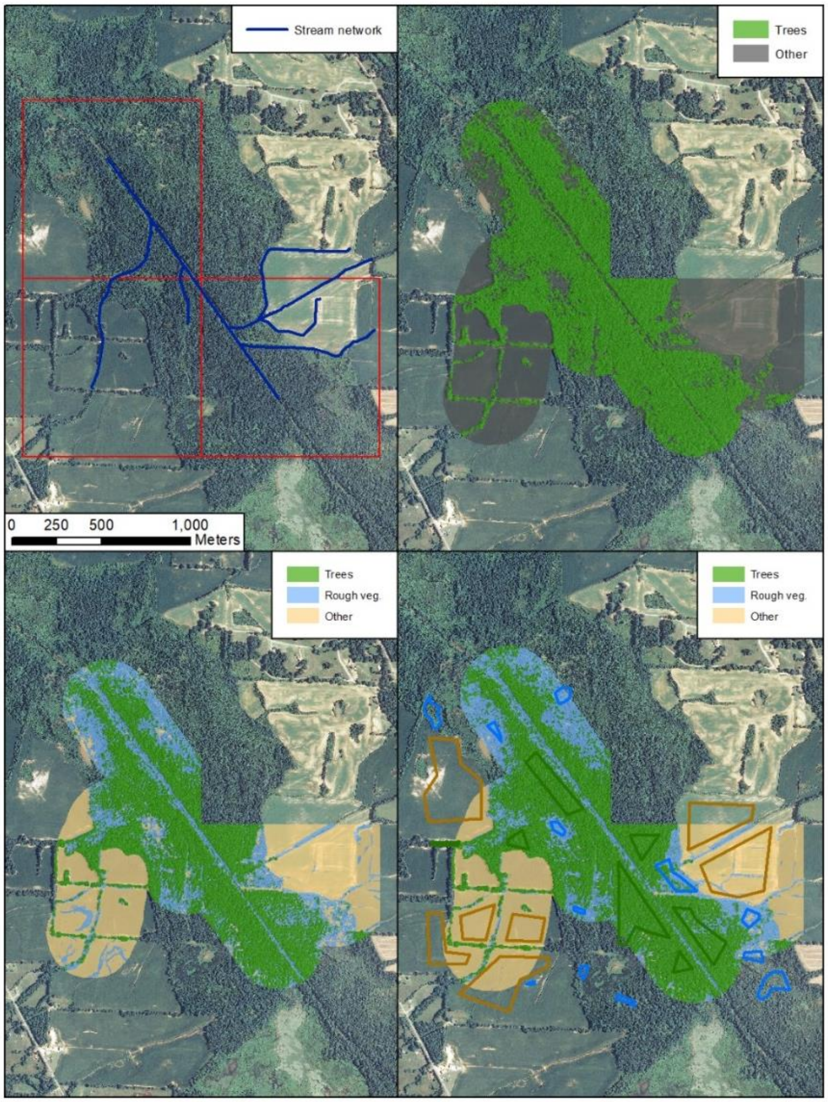


Figure 42: Sample outputs for various version of the RCL model. Output was restricted to the 250m buffer surrounding the stream network. Upper left: footprint of input .las files and stream network. Upper right: pretrained binary model. Bottom left: pretrained ternary model. Bottom right: custom trained model and training classes used to train model.

Both the binary general model and the binary TN model exhibit high F1-scores, indicating strong predictive power, and each do so using a decision tree with a single branch. This is not particularly surprising given that the binary models need only to differentiate tree and non-tree cover, but such binary classification schemes are still useful for many remote-sensing and watershed investigations. The significant contributor to model misclassification in these schemes are building edges and power lines, both of which produce return splitting that is similar to the splitting produced by trees. Though return numbers are not directly used in the model, this splitting causes artifacts in the DHM that are similar to the signature of trees. Misclassifications of this sort are difficult to eliminate due to the relative infrequency of building edges and power lines compared to other landcover types.

The ternary general and ternary TN models both exhibit acceptable predictive power in general but are relatively limited in their ability to identify vegetative cover. In general herbaceous produces a faint, inconsistent LiDAR signature compared to bare ground and often accounts for a small portion of total landcover during the leaf-off season when most LiDAR missions are flown, including those whose data is included in this study; this reduces the available training samples and likely causes training sample inconsistency. The TN model's predictive abilities are reduced further due to its data restrictions; both roughness and Haralick textures were excluded from this model due to the difficulty of generating them in ArcGIS, but the DHM roughness and height model cluster shade are relatively diagnostic of the presence or absence of vegetation. Like the binary models, some model inaccuracy can be attributed to misclassification of building edges and power lines, though significant inaccuracy is due to rough tilled soil, which produces a signature that is similar to the rough textures of herbaceous vegetation.

The binary and ternary TN models were packaged into an ArcGIS tool; these are pretrained models that can be used without supplying a training classification file. The binary model is expected to be relatively generalizable outside of urban areas because it classifies exclusively based on DHM signature. Because a signature on the DHM occurs due to return splitting, and because non-urban areas do not contain significant areas of return-splitting buildings or powerlines, it can be assumed that any DHM signature is due to a tree.

The ternary TN model may generalize within TN, but it is unknown how the model may be affected by differing LiDAR collection methods (such as sensor type of flight altitude). The general model appears robust to both changes in LiDAR collection methods and physiography, but this model could not be packaged within the RCL tool because it requires roughness and Haralick texture data products, both of which are difficult to generate natively in ArcGIS.

Overall, all models display similar classification quality inside and outside of the riparian corridor. However, because the riparian corridor typically lacks the anthropogenic structures and tilled soils that contribute to a significant portion of model misclassifications, it can generally be assumed that intra-riparian corridor classifications are of slightly higher quality than those falling outside the riparian corridor.

It is important to note that the models were trained primarily using leaf-off (fall-winter) LiDAR data, which comprises the bulk of publicly funded LiDAR missions. Thus, these models may not generalize to LiDAR collected during other seasons even if they generalize along other parameters such as physiography or LiDAR collection method. In order to allow more flexible classification schemes, an alternative version of the RCL tool has been created that allows the user to input their own training data to train a classification model. Because site-specific models are not required to be generalizable, they can take advantage of incomparable data (such as return intensity) than the generalizable models cannot use. Thus, these models have the potential for higher classification quality at the cost of requiring creation of training data.

6 Executive Summary

Estimation and mitigation of non-point source pollutants from agricultural fields depends on an understanding of complex interactions between anthropogenic and natural processes varying in time and space. This complexity poses a challenge to action agencies when developing conservation strategies to reduce non-point source pollutants at watershed scale, requiring decisions on what type of conservation practice to use and where, in the watershed to implement it. This pilot project supported such efforts by conducting two related studies: watershed modeling and development of GIS tool for riparian vegetation delineation.

In the watershed modeling component of the project, long-term averages of non-point source pollution were estimated using the USDA developed and supported Annualized AGricultrual Non-Point Source (AnnAGNPS) watershed management and pollution model. This model was chosen primally for its unique technology allowing for detailed description of farming management and conservation practices at field scale and for its capability of being applied to un-gaged watersheds. Quantitative analysis of results was performed through relative comparisons between AnnAGNPS simulations and informing about the potential impact of conservation practices to the overall reduction in non-point source pollution at the watershed scale. Additionally, spatial estimates of sources and sinks generated information of critical sediment yield producing locations.

This study covered two systems comprising of 10 HUC12s sub-watersheds (USGS classification). A total of 121 AnnAGNPS simulations were performed to evaluate alternative scenarios describing different conservation practices and their effect to reducing sediment yield/load at field and at watershed scale.

Comparison of simulation results between systems/watersheds depicted the importance of the development of tailored mitigation strategies for individual watersheds. Estimates from simulations containing the same conservation practice applied to different watersheds yielded different reduction amounts in annual average sediment loads when compared to the baseline conditions.

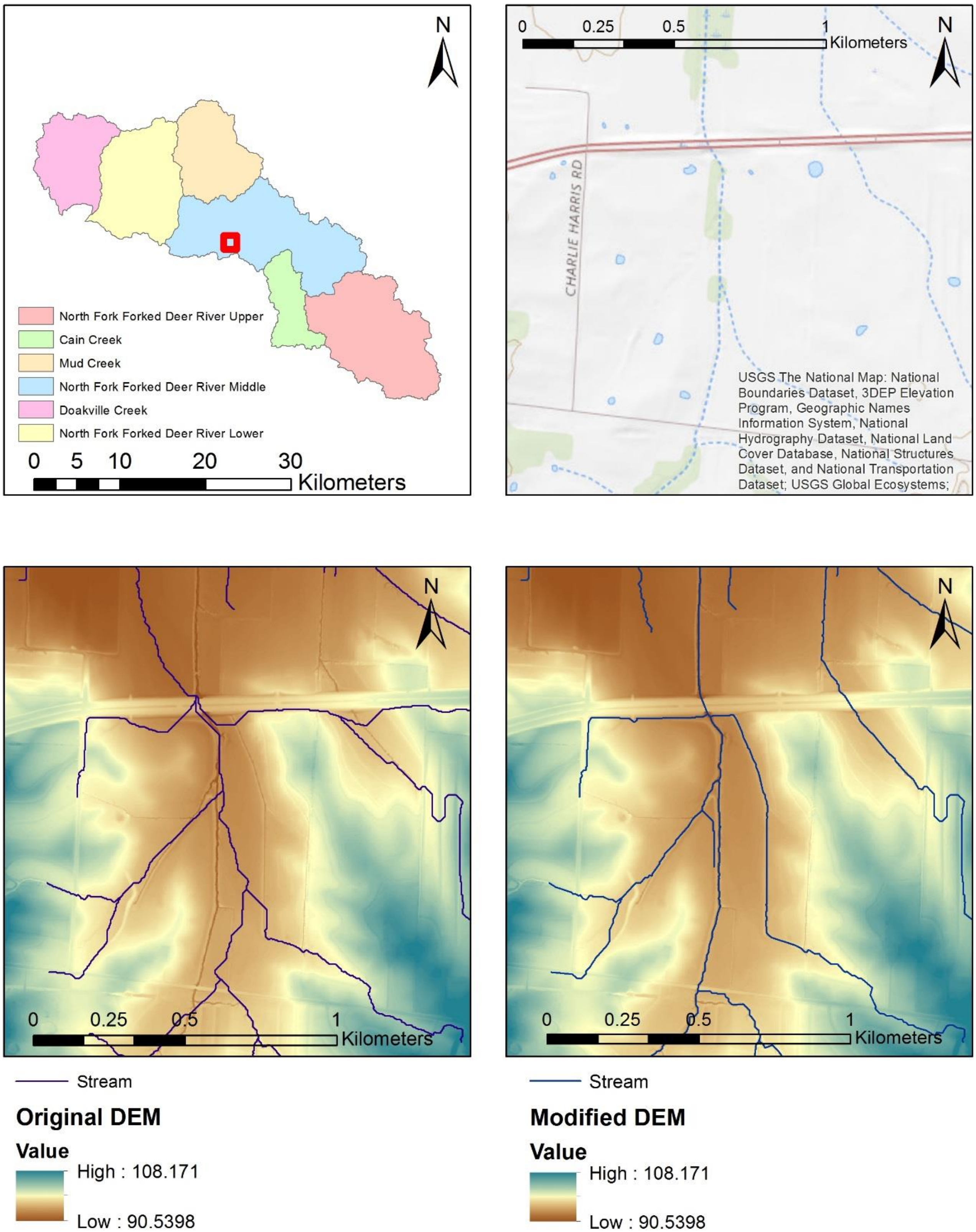
Baseline conditions considered no conservation practice. Additional alternative scenarios considering existing riparian vegetation (tree-like as determined by LiDAR analysis) and scenarios varying assumptions of the presence of concentrated flow path through existing riparian vegetation were also considered. These simulation results indicate the potential contribution of existing riparian vegetation of acting as filter strips, and therefore reducing sediment input into waterways. Existing conditions are expected to be between the simulation considering existing riparian vegetation and no concentrated flow path and simulations considering existing riparian vegetation but accounting for the presence of concentrated flow paths, as these are natural occurring vegetation and therefore not managed.

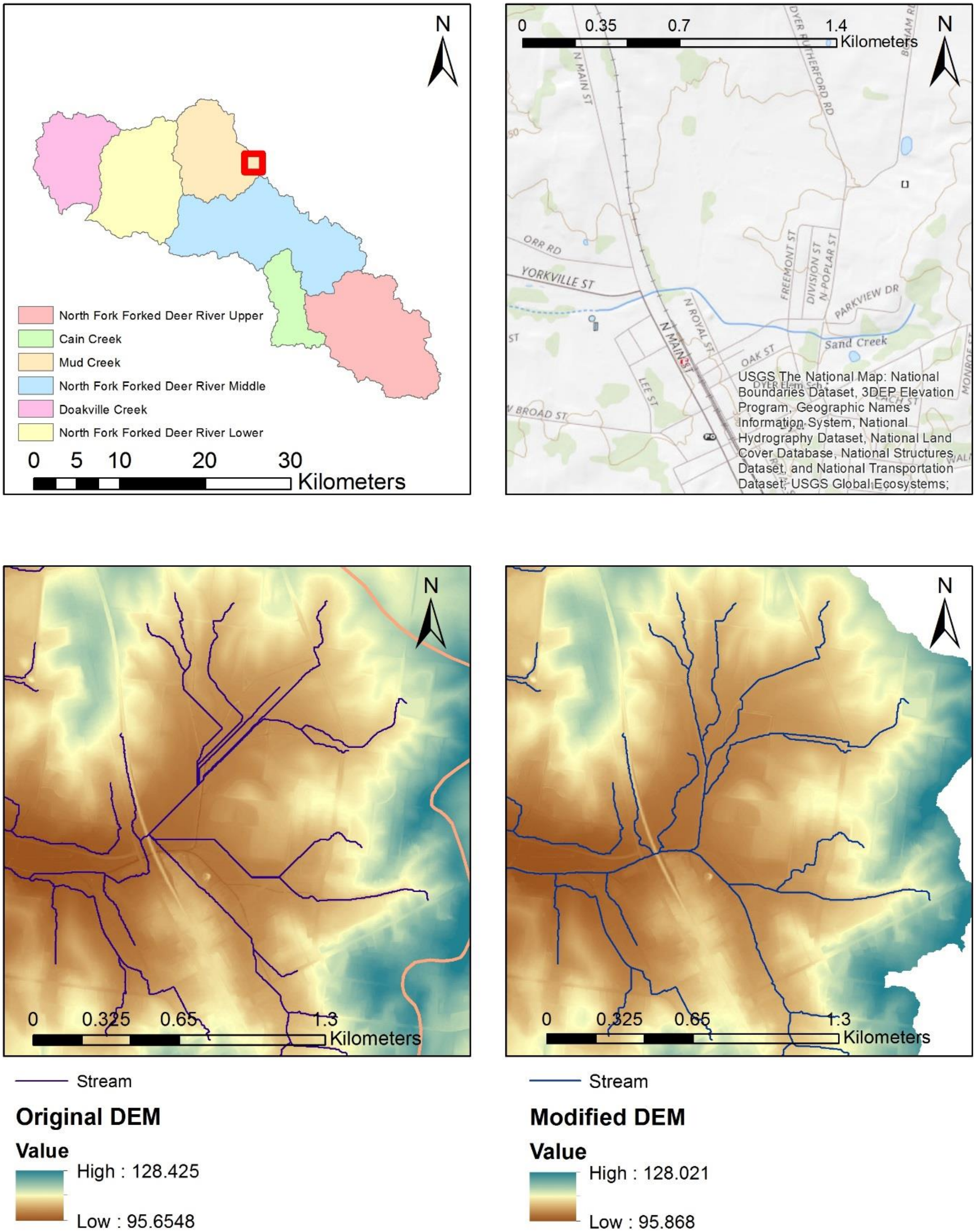
Assessment of the effectiveness of different scenarios considering managed conservation practices applied to the same watershed indicated the implementation of riparian buffers to be the most effective practice but at the cost of potentially removing area from production. However, simulation results suggest the possibility of implementation of conservation practices at strategic locations promoting sediment load reduction while minimizing the negative impact to production. Additionally, results also point to an alternative mitigation strategy comprised of multiple integrated conservation practices.

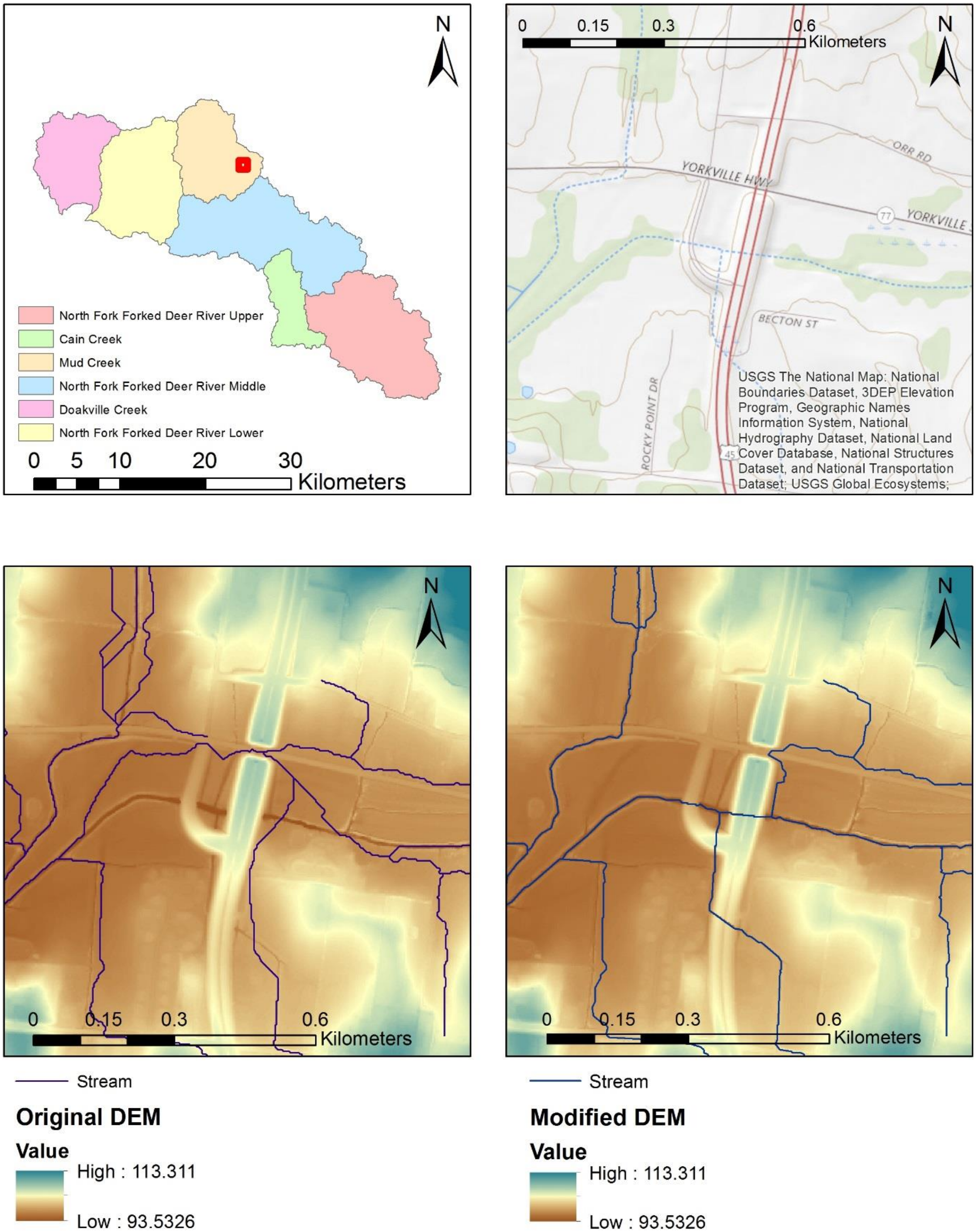
In the development of the Riparian Classification from LiDAR (RCL) Tool, machine learning was used to develop decision trees for the classification of trees (and tree-like vegetation) from LiDAR datasets. The decision of using decision trees as the machine learning algorithm stems from the fact that they generate human-readable solutions comprised of recursive if-then statements. Solutions were generated specific to be generalizable to different conditions and agnostic to LiDAR sampling intensity, vendor specific LiDAR characteristics, and based on standard procedures found in industry leading GIS software package (ArcGIS).

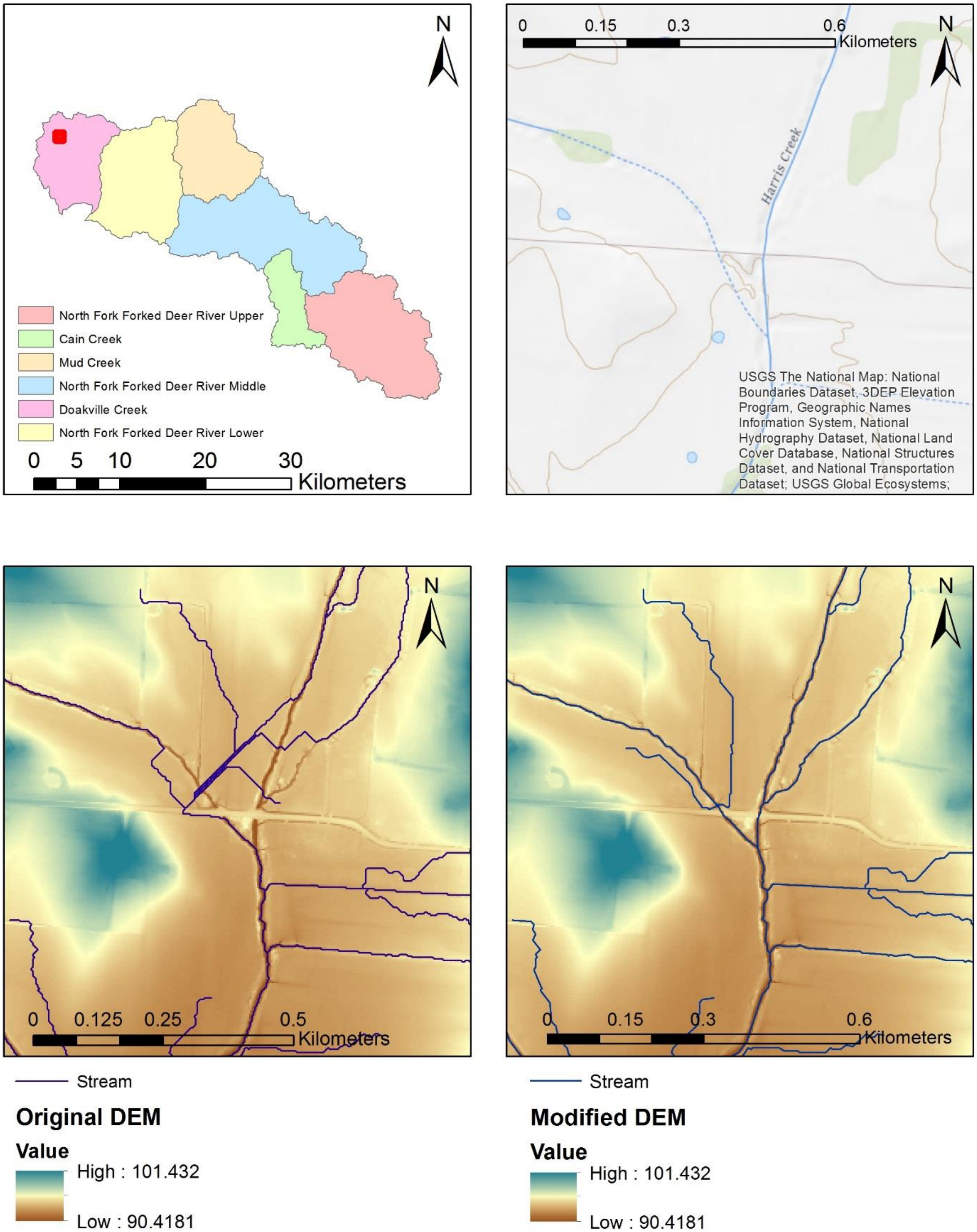
Methods developed and findings demonstrate the importance of these studies to the development of mitigation and conservation strategies at watershed scale but with the capability of spatially identifying candidate locations for conservation practices. The findings from this pilot study support the long-term NRCS's goals by generating information to complement existing NRCS efforts in guiding future conservation strategies development, enhancements, and expansion.

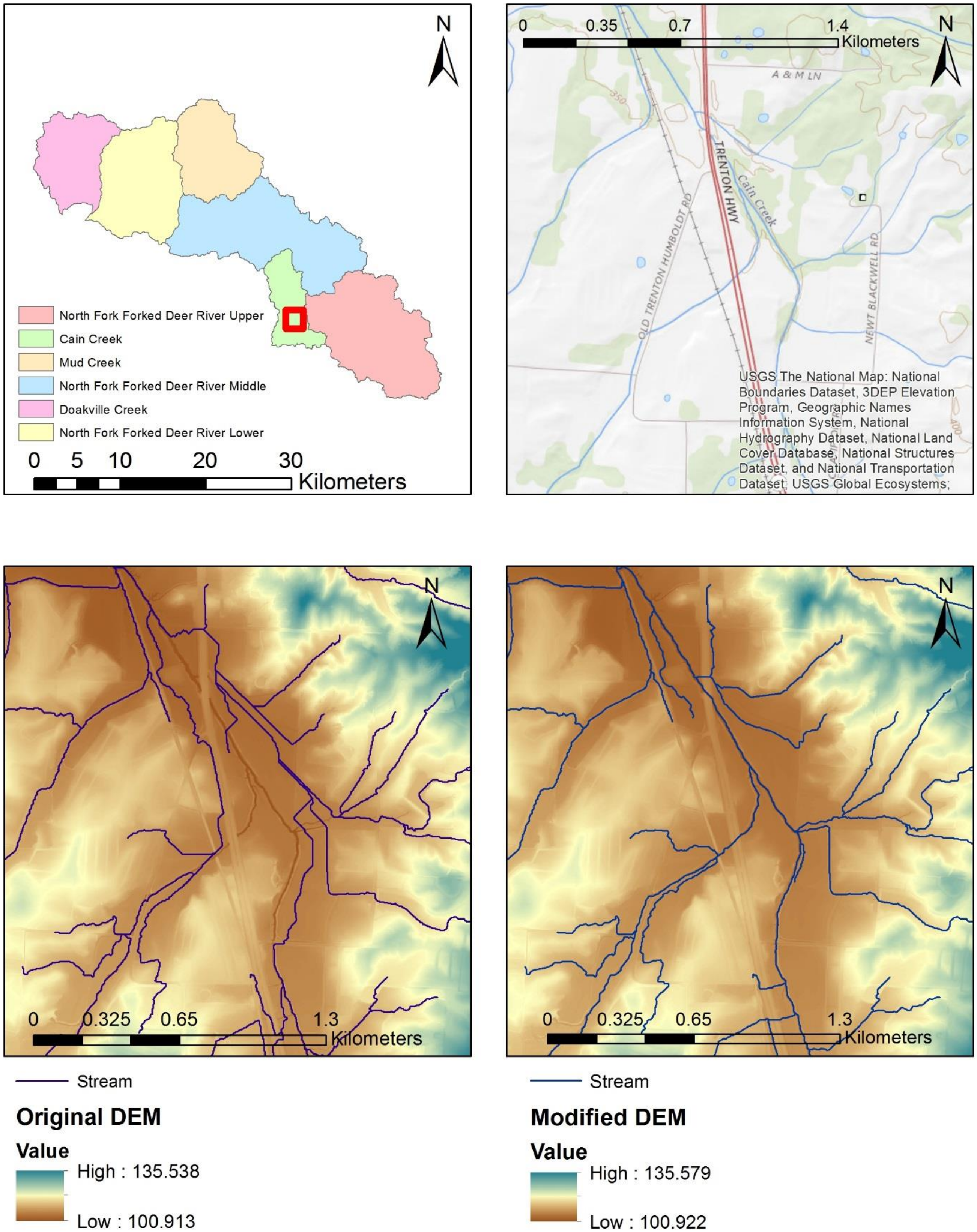
7 Appendix 1: DEM Hydrological Correction Examples

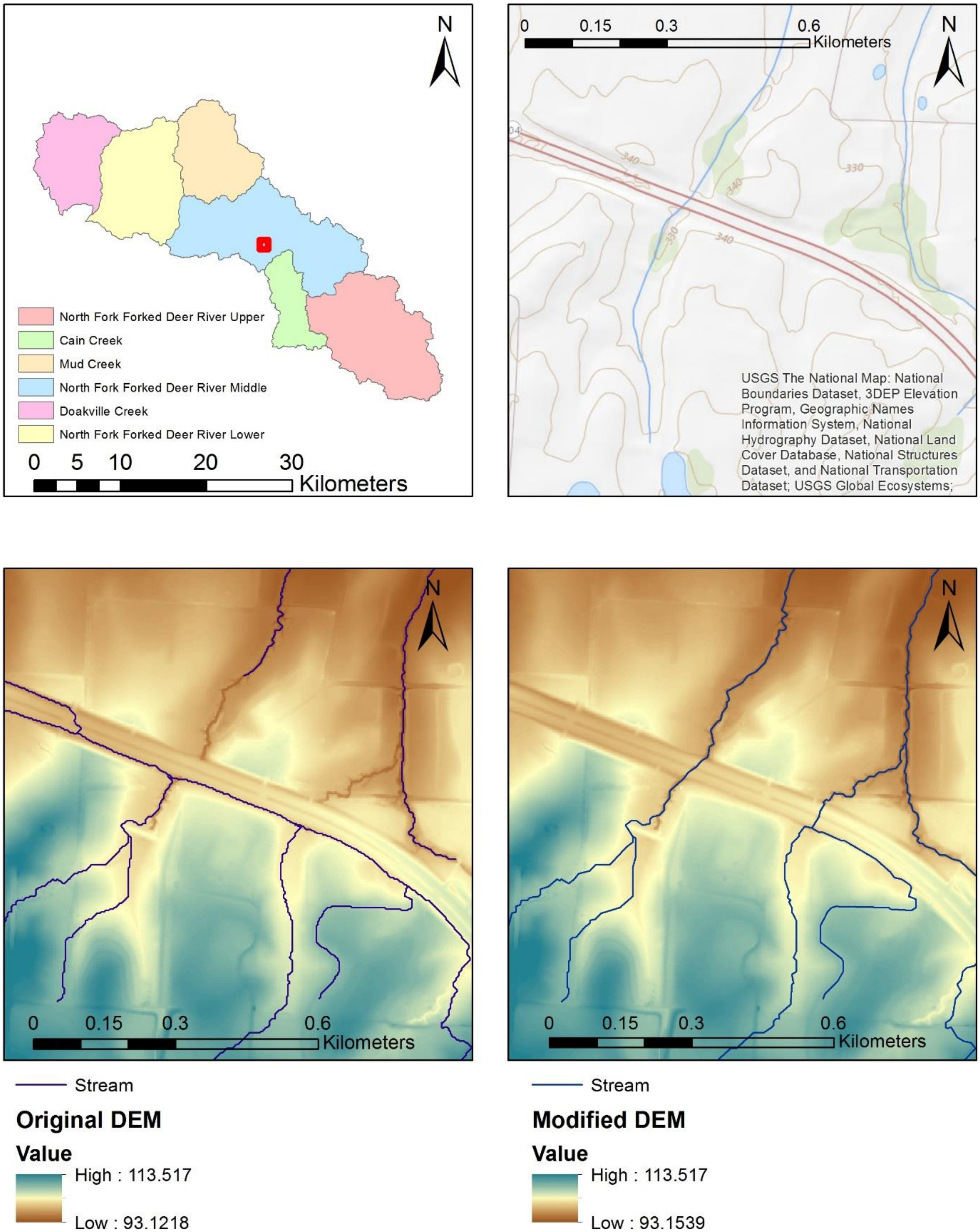


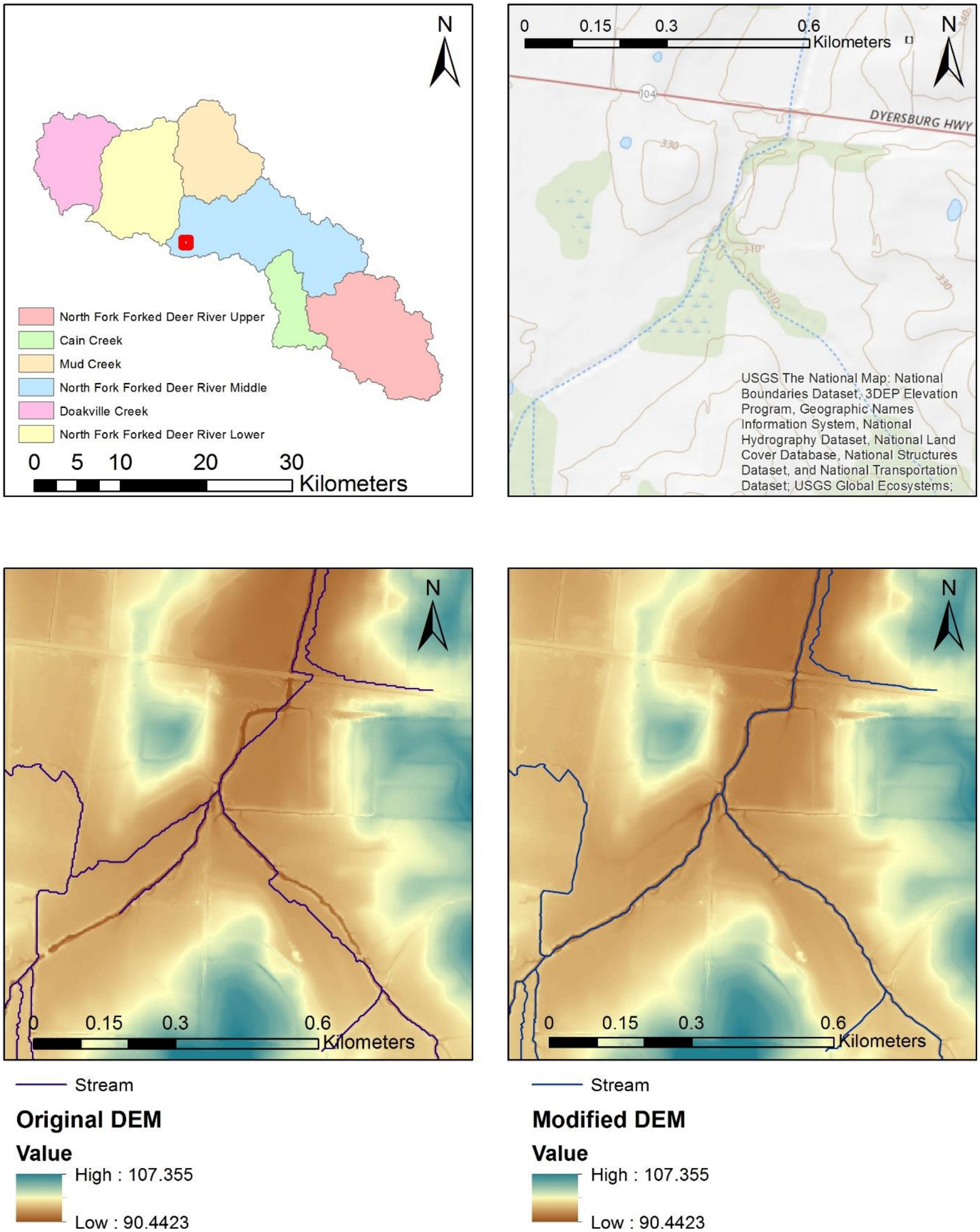

N
North Fork Forked Deer River Upper
Cain Creek
Mud Creek
North Fork Forked Deer River Middle
Doakville Creek
North Fork Forked Deer River Lower
0 5 10 20 30
Kilometers
0 0.35 0.7 1.4
Kilometers
N
N MAIN ST
DYER RUTHERFORD RD
ORR RD
YORKVILLE ST
FREEMONT ST
DIVISION ST
N POPLAR ST
PARKVIEW DR
N ROYAL ST
N MAIN ST
Sand Creek
OAK ST
LEE ST
BROAD ST
USGS The National Map: National Boundaries Dataset, 3DEP Elevation Program, Geographic Names Information System, National Hydrography Dataset, National Land Cover Database, National Structures Dataset, and National Transportation Dataset; USGS Global Ecosystems;
N
0 0.325 0.65 1.3
Kilometers
Stream
Original DEM
Value
High : 128.425
Low : 95.6548
N
0 0.325 0.65 1.3
Kilometers
Stream
Modified DEM
Value
High : 128.021
Low : 95.868


N
North Fork Forked Deer River Upper
Cain Creek
Mud Creek
North Fork Forked Deer River Middle
Doakville Creek
North Fork Forked Deer River Lower
0 5 10 20 30
Kilometers
0 0.15 0.3 0.6
Kilometers
YORKVILLE HWY
ORR RD
YORKVILLE
BECTON ST
ROCKY POINT DR
USGS The National Map: National Boundaries Dataset, 3DEP Elevation Program, Geographic Names Information System, National Hydrography Dataset, National Land Cover Database, National Structures Dataset, and National Transportation Dataset; USGS Global Ecosystems;
Stream
Original DEM
Value
High : 113.311
Low : 93.5326
Stream
Modified DEM
Value
High : 113.311
Low : 93.5326


N
North Fork Forked Deer River Upper
Cain Creek
Mud Creek
North Fork Forked Deer River Middle
Doakville Creek
North Fork Forked Deer River Lower
0 5 10 20 30
Kilometers
0 0.15 0.3 0.6
Kilometers
Harris Creek
USGS The National Map: National Boundaries Dataset, 3DEP Elevation Program, Geographic Names Information System, National Hydrography Dataset, National Land Cover Database, National Structures Dataset, and National Transportation Dataset; USGS Global Ecosystems;
0 0.125 0.25 0.5
Kilometers
0 0.15 0.3 0.6
Kilometers
Stream
Original DEM
Value
High : 101.432
Low : 90.4181
Stream
Modified DEM
Value
High : 101.432
Low : 90.4181


N
North Fork Forked Deer River Upper
Cain Creek
Mud Creek
North Fork Forked Deer River Middle
Doakville Creek
North Fork Forked Deer River Lower
0 5 10 20 30
Kilometers
0 0.35 0.7 1.4
Kilometers
A & M LN
TRENTON HWY
Cain Creek
OLD TRENTON HUMBOLDT RD
NEWT BLACKWELL RD
USGS The National Map: National Boundaries Dataset, 3DEP Elevation Program, Geographic Names Information System, National Hydrography Dataset, National Land Cover Database, National Structures Dataset, and National Transportation Dataset; USGS Global Ecosystems;
0 0.325 0.65 1.3
Kilometers
Stream
Original DEM
Value
High : 135.538
Low : 100.913
Stream
Modified DEM
Value
High : 135.579
Low : 100.922


N
North Fork Forked Deer River Upper
Cain Creek
Mud Creek
North Fork Forked Deer River Middle
Doakville Creek
North Fork Forked Deer River Lower
0 5 10 20 30
Kilometers
0 0.15 0.3 0.6
Kilometers
340
330
USGS The National Map: National Boundaries Dataset, 3DEP Elevation Program, Geographic Names Information System, National Hydrography Dataset, National Land Cover Database, National Structures Dataset, and National Transportation Dataset; USGS Global Ecosystems;
Stream
Original DEM
Value
High : 113.517
Low : 93.1218
Stream
Modified DEM
Value
High : 113.517
Low : 93.1539


N
North Fork Forked Deer River Upper
Cain Creek
Mud Creek
North Fork Forked Deer River Middle
Doakville Creek
North Fork Forked Deer River Lower
0 5 10 20 30
Kilometers
0 0.15 0.3 0.6
Kilometers
104
DYERSBURG HWY
330
310
USGS The National Map: National Boundaries Dataset, 3DEP Elevation Program, Geographic Names Information System, National Hydrography Dataset, National Land Cover Database, National Structures Dataset, and National Transportation Dataset; USGS Global Ecosystems;
Stream
Original DEM
Value
High : 107.355
Low : 90.4423
Stream
Modified DEM
Value
High : 107.355
Low : 90.4423

8 Appendix 2: System 1 Figures

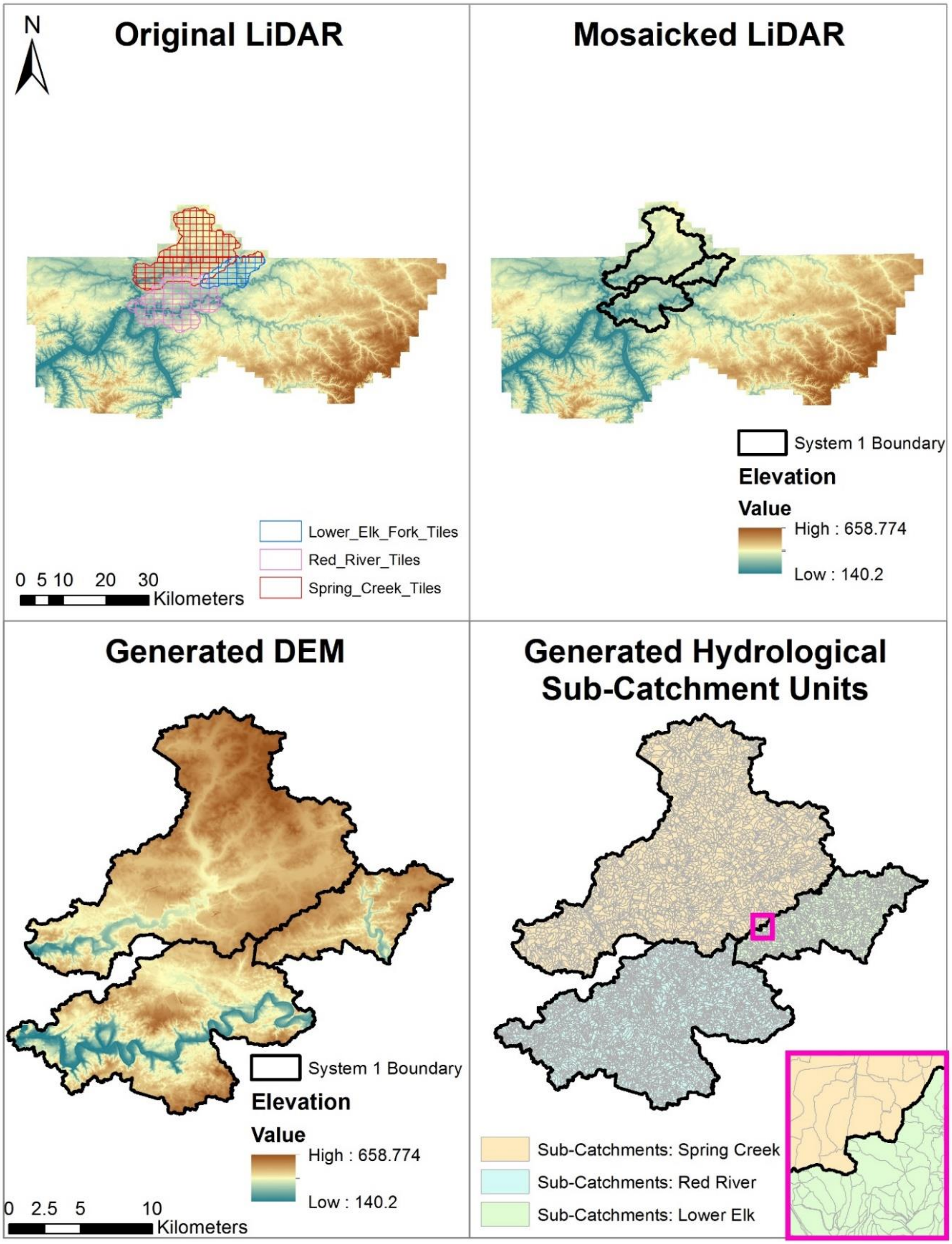


Figure 43: DEM pre-processing steps and the generated AnnAGNPS cells for system 1

Table 21: The final topographical statistics for System 1 AnnAGNPS cells

HUC-12 Name	Number of AnnAGNPS Cells	Average Cell Area (Ha)	Average Cell Slope (%)	Average Cell Elevation (m)	Average Concentrated Flow Length (m)
Lower Elk Fork	2132	2.28	0.15	550.53	129.47
Spring Creek	6221	2.20	0.23	502.29	124.61
Red River Upper	2887	2.18	0.19	514.79	126.64
Red River Lower	3201	2.20	0.26	491.79	121.95

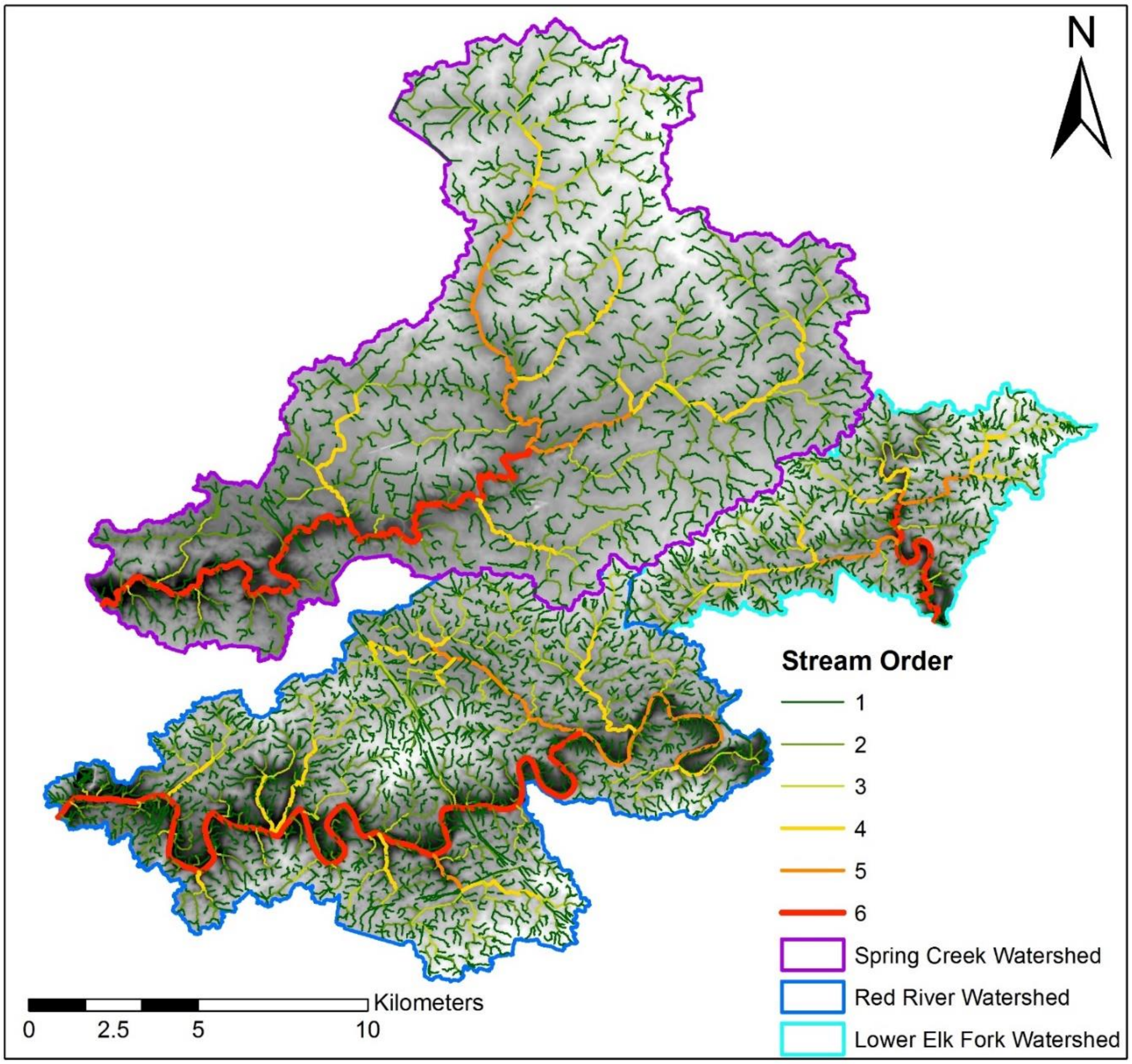


Figure 44: Generated Stream Network for System 1

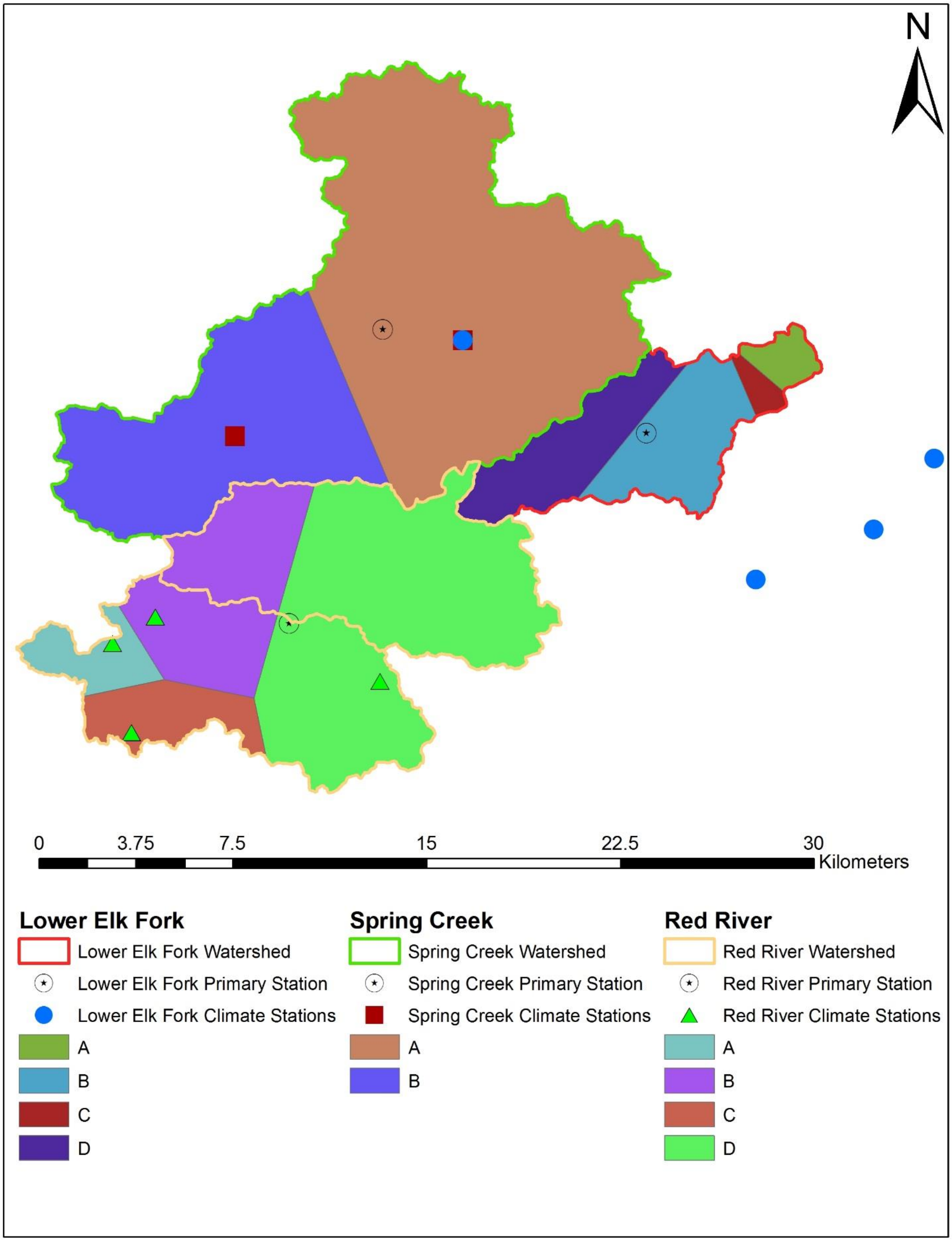


Figure 45: Generated Stream Network for System 1

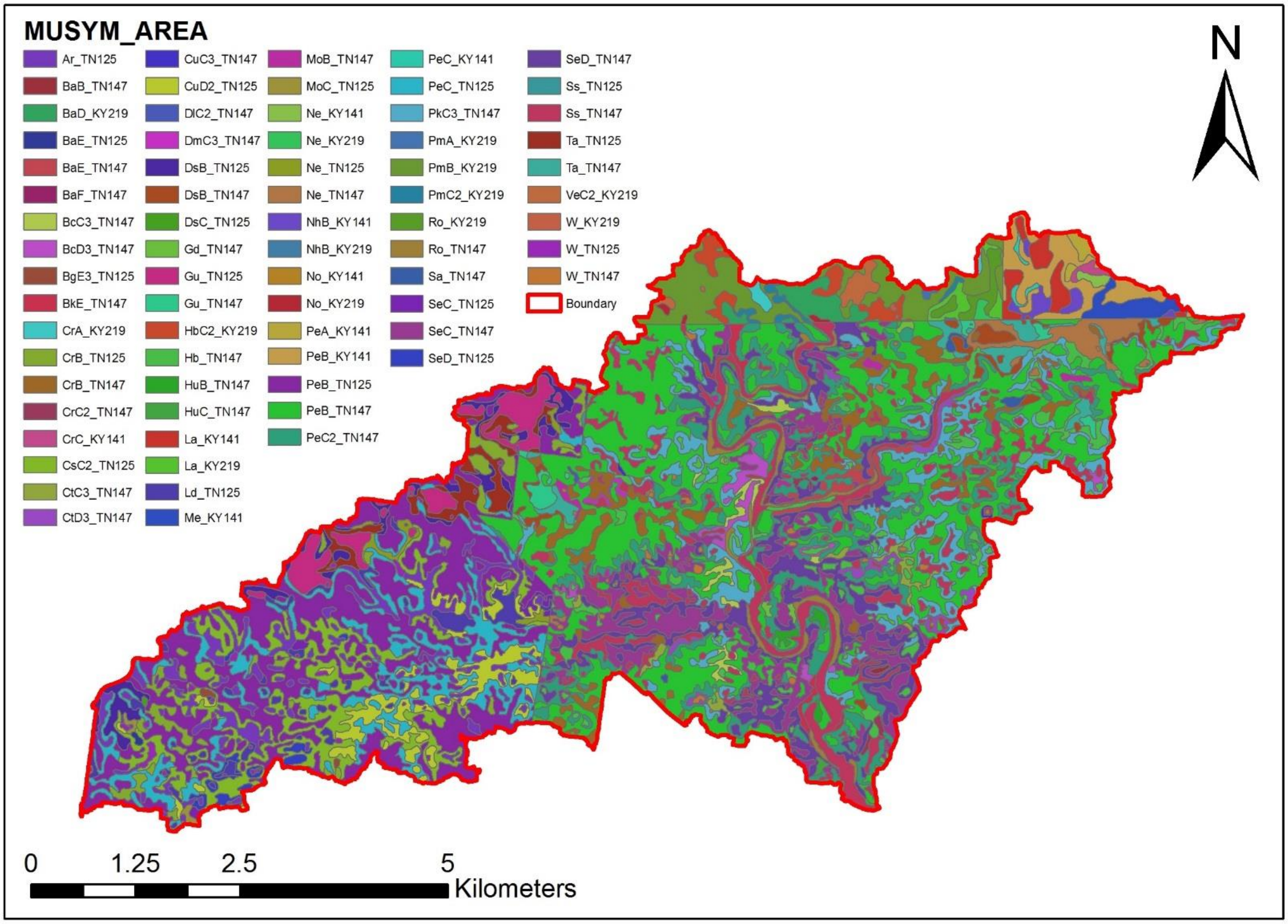


Figure 46: System 1 – Lower Elk Soils Map

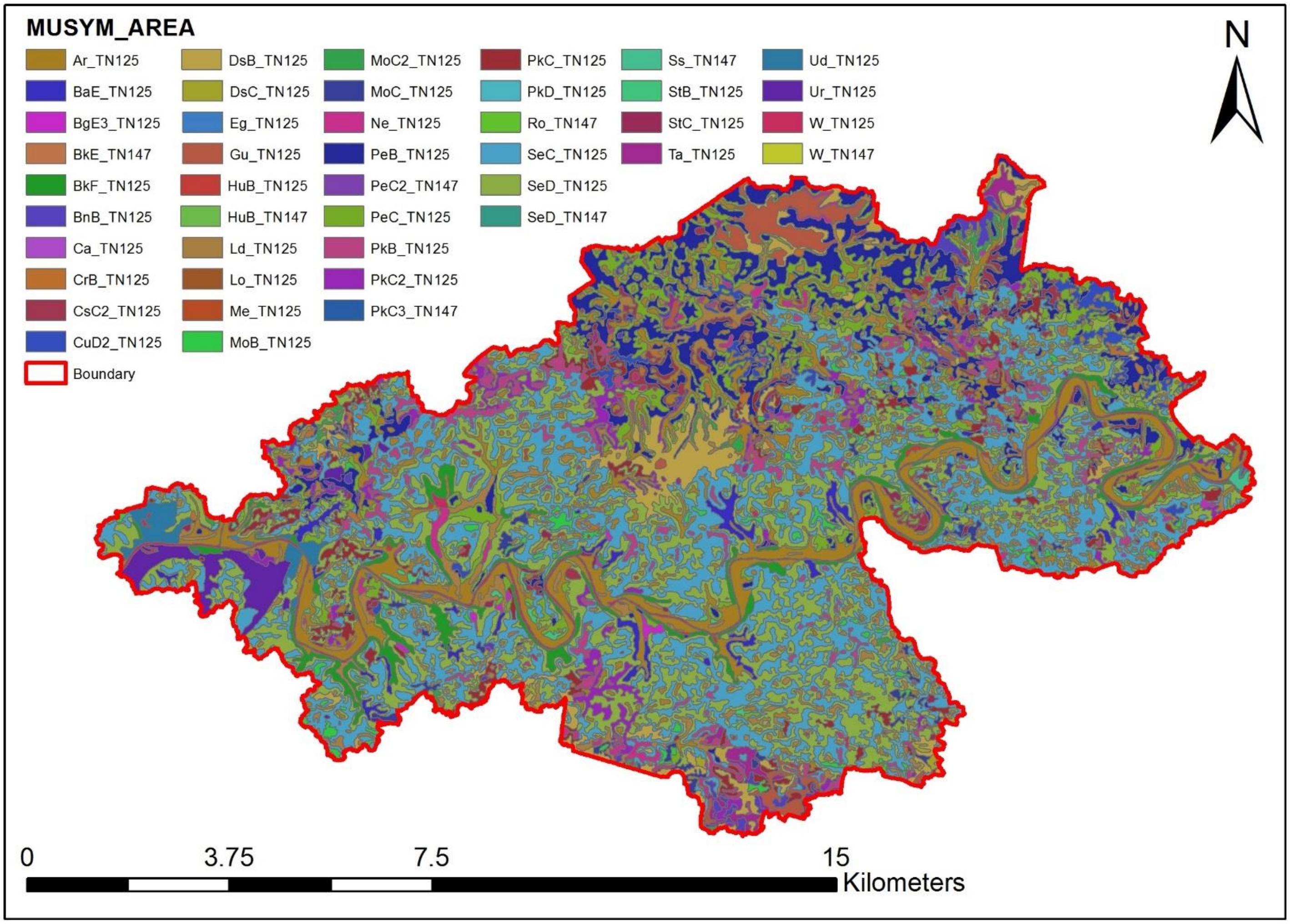


Figure 47: System 1 – Red River Soils Map

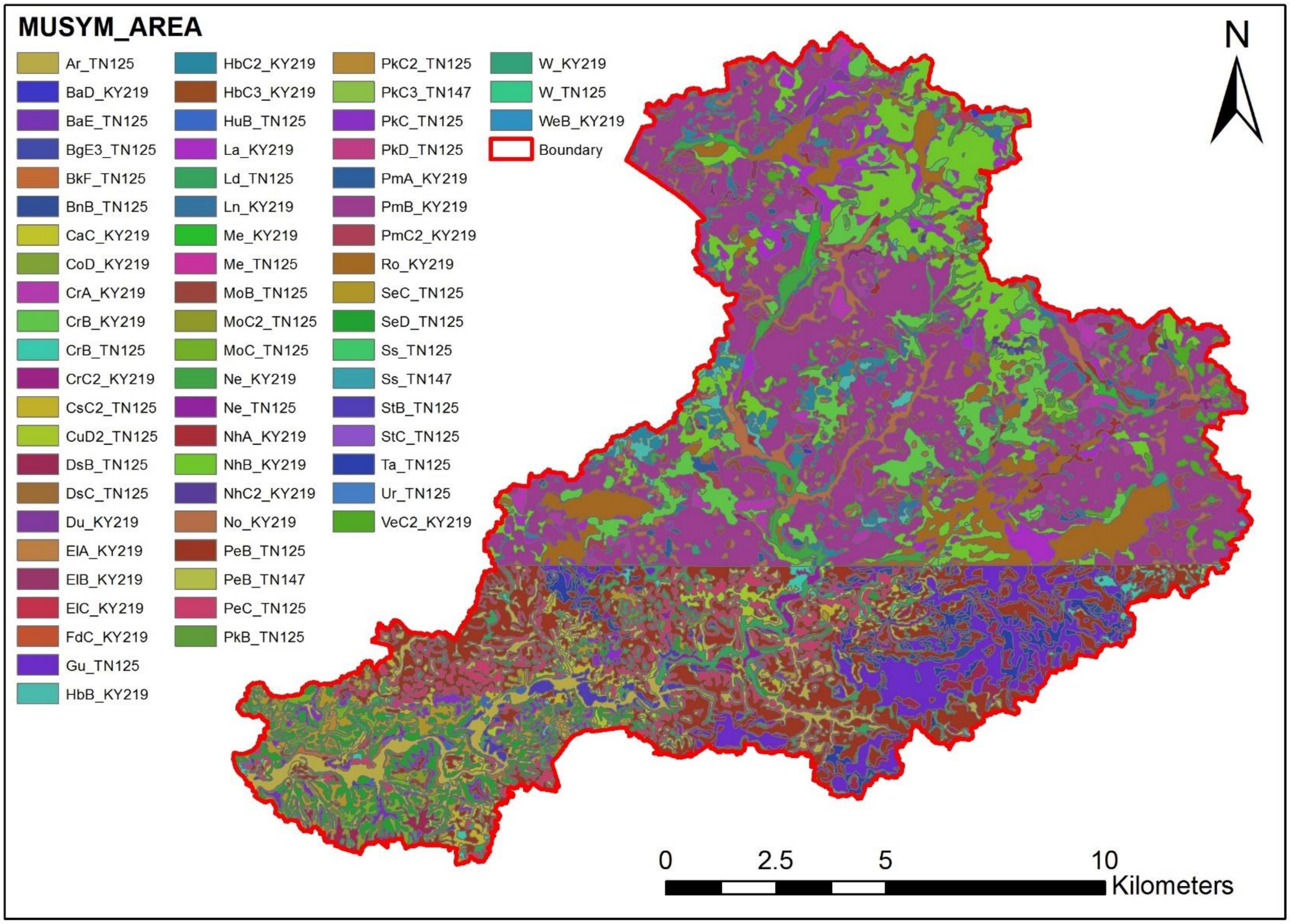


Figure 48: System 1 – Spring Creek Soils Map

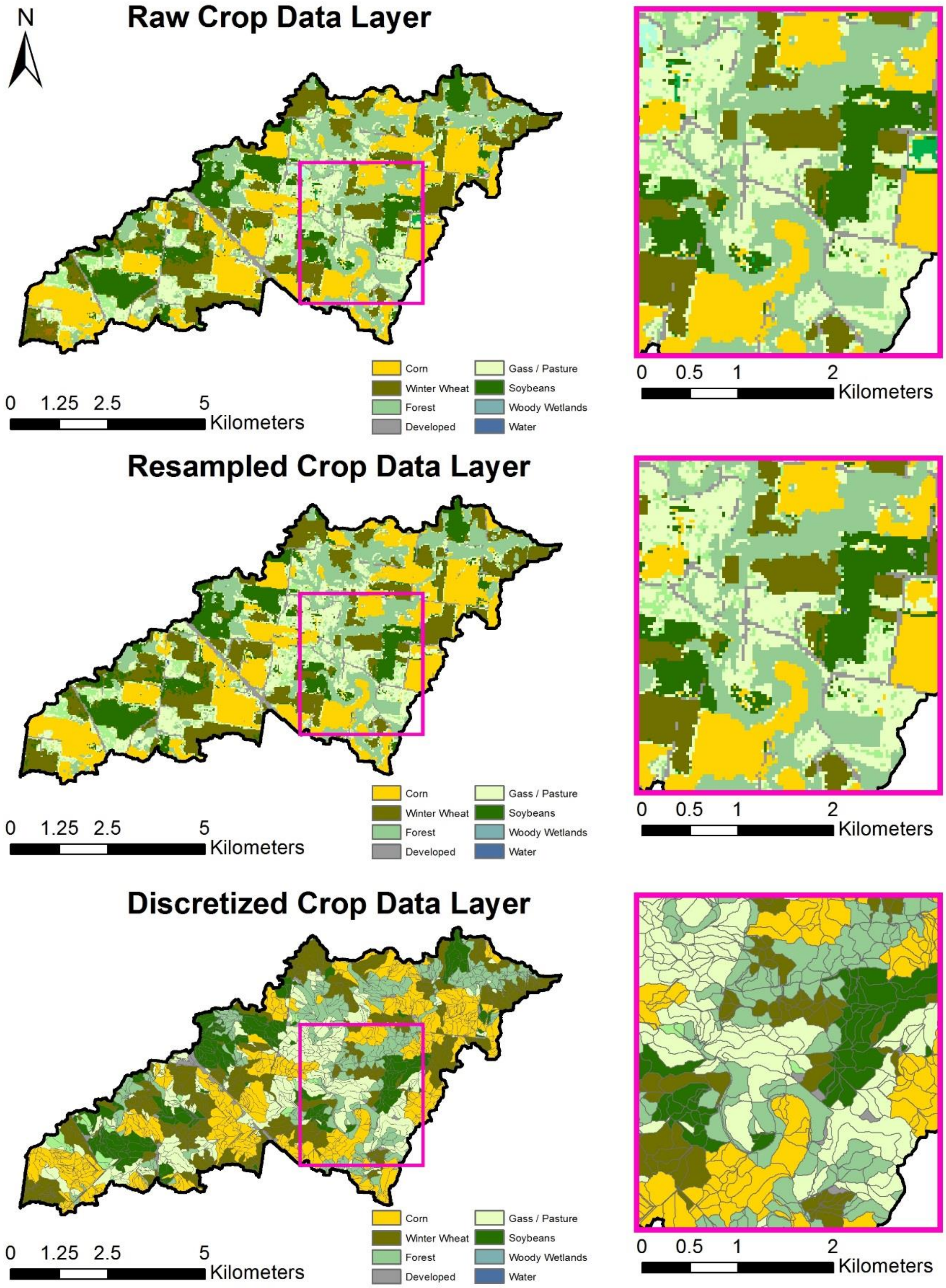


Figure 49: The three main processing steps for the crop data layer for System 1 – Lower Elk

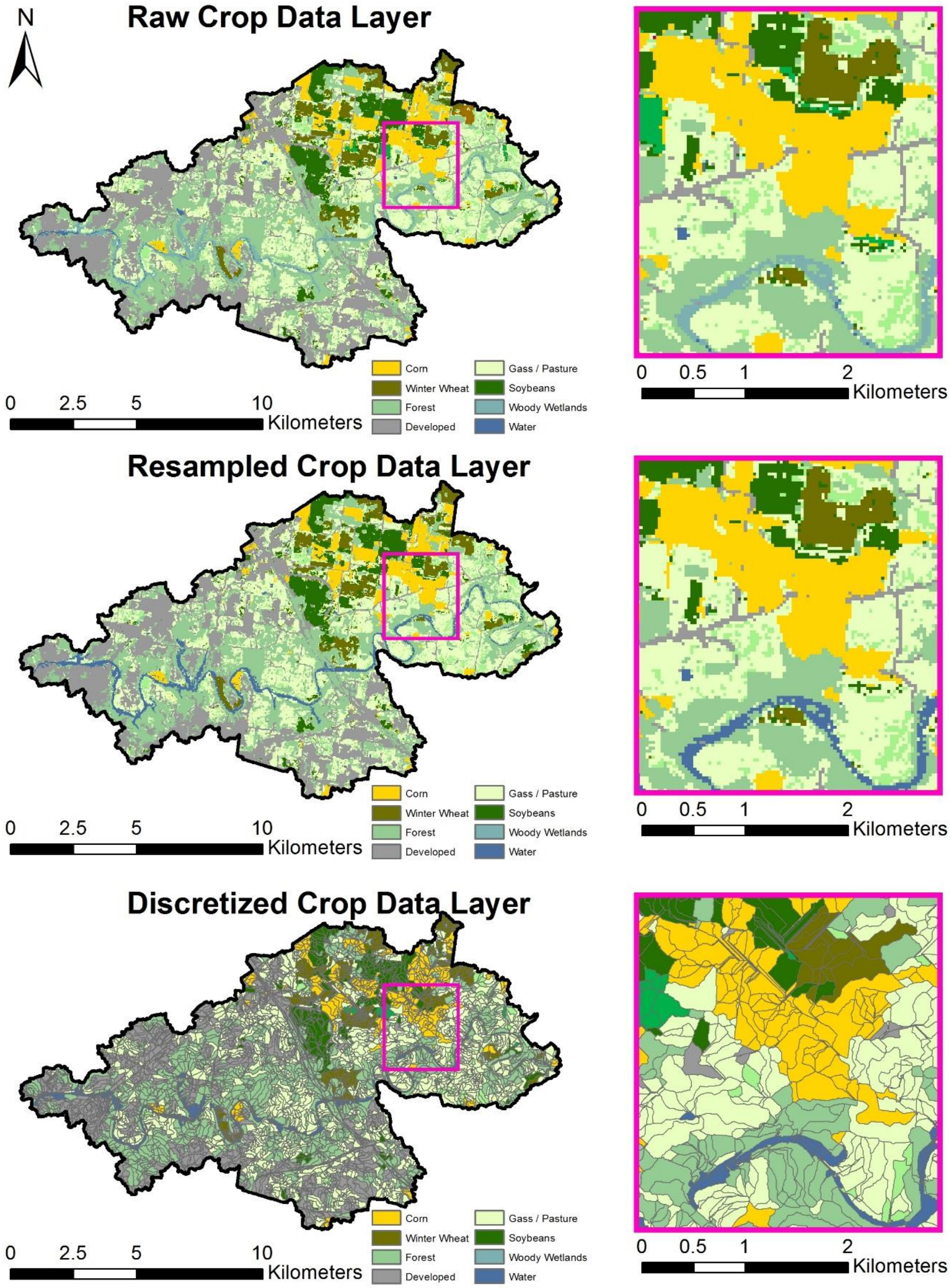


Figure 50: The three main processing steps for the crop data layer for System 1 – Red River

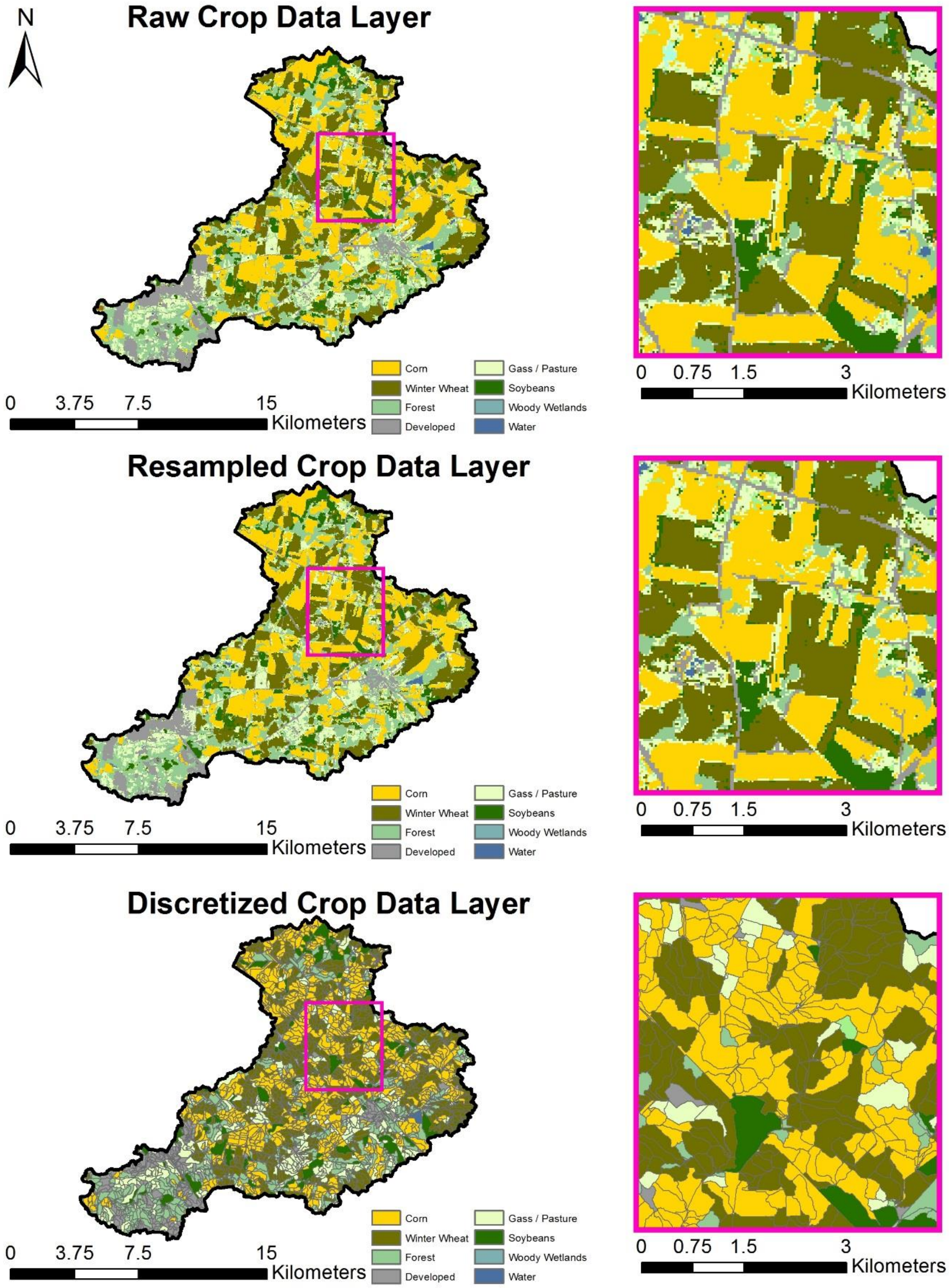


Figure 51: The three main processing steps for the crop data layer for System 1 – Spring Creek

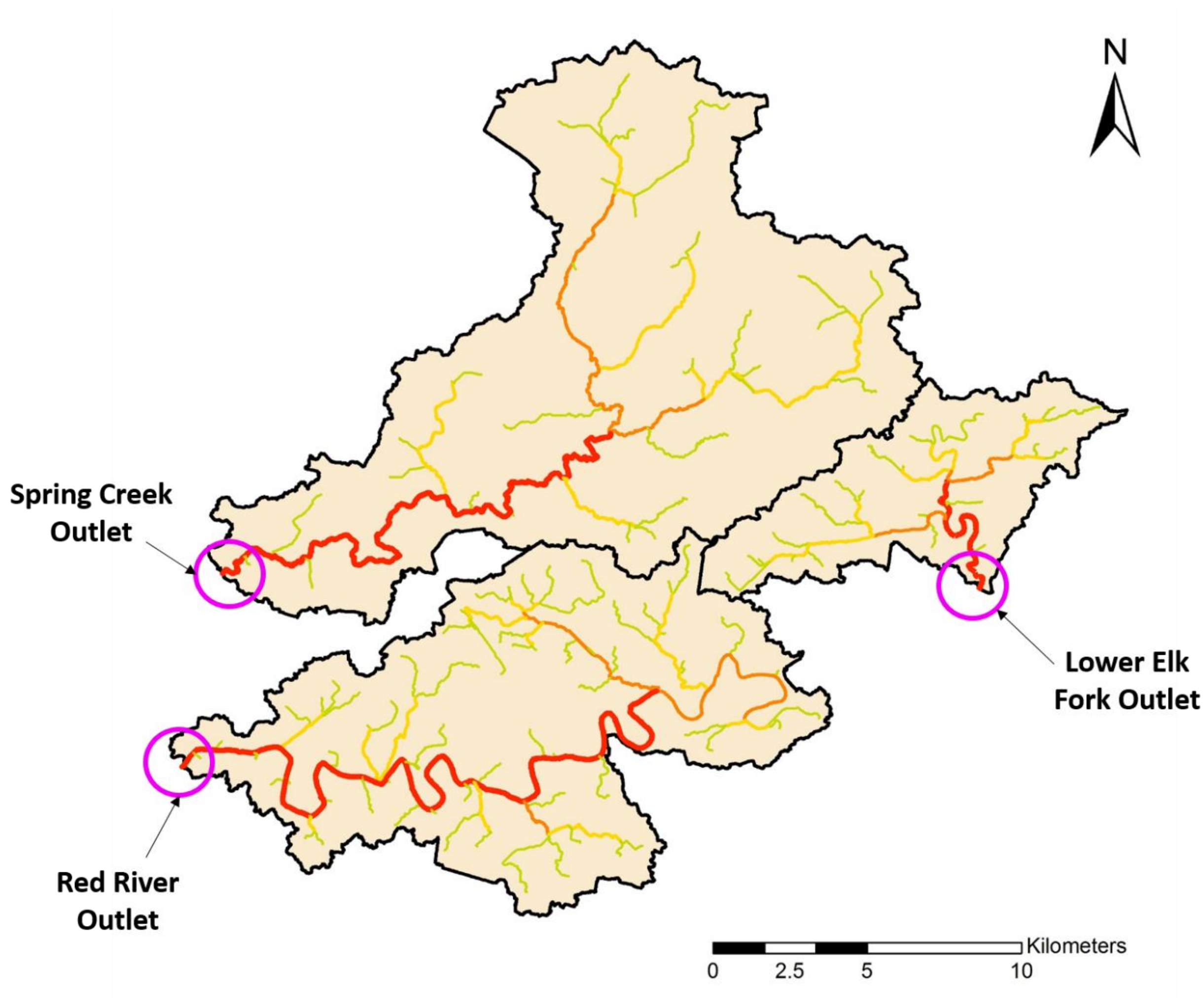


Figure 52: System 1 sub-systems

9 Appendix 3 – RCL Tutorial

9.1 Abstract

The Riparian Classification from LiDAR (RCL) tool is a Python script designed to classify riparian land cover. The tool accepts a folder of LiDAR binary (.las*) files and outputs a land use raster as well as supporting data files such as digital elevation and slope models. It is available as an ArcGIS script tool, meaning it can be called using the same graphical user interface that standard ArcGIS tools use. There are currently two versions of this tools: one that uses a pretrained model to classify land cover, and one that accepts training data to train a custom classification model. This document details how to set up and use both version of the RCL tool.*

9.2 Assumptions

This walkthrough assumes that the user is generally familiar with ArcMap and has already obtained LiDAR data in *.las* format for their study area, either by directly downloading *.las* files or unpacking *.laz* (compressed *.las*) or *.zlas* (ESRI compressed *.las*) files. Optimal results will be achieved with LiDAR data having a point density of over 1pt/m^2. It is not necessary that the user is familiar with any scripting languages. The *.las* files for the study area should be contained within a single folder with no other files.

The RCL tool requires ArcGIS 10.X and valid Spatial and 3D Analyst licenses to run.

9.3 Walkthrough

9.3.1 Downloading the RCL Tool

The RCL tool can be found at MTSU OnDrive.” Unpack the zipped folder in a location that can be easily found later. This folder contains the RCL Python script and the associated ESRI Toolbox file, as well as supporting documentation (including this walkthrough).

9.3.2 Running the RCL Tool

To open the RCL interface, open ArcCatalog. The Catalog can be accessed either directly in the ArcCatalog application or through ArcMap by clicking the “Catalog” button in the toolbar. In the Catalog navigate to the folder downloaded in the previous step. Double click on *rcl.tbx* to expand it, then double click the *Riparian Classification from LiDAR (pretrained)* or *Riparian Classification from LiDAR (custom)* tool to open the tool dialogue (Figure 53). A graphical user interface similar to standard ESRI tools will appear. Fill each field as instructed by the tooltips.

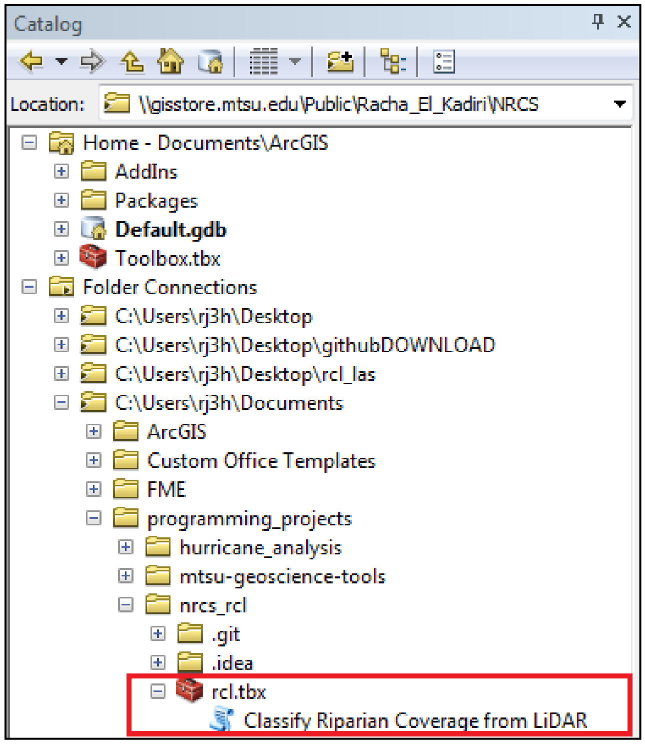


Figure 53: The directory structure of the RCL toolbox in the Catalog. Double-clicking Classify Riparian Coverage from LiDAR will launch the RCL tool user interface.

. If you are using the pretrained version of the tool, the only required input is the folder of *.las* files. You may optionally supply a file representing a flow network and buffer width to restrict the area of investigation; this typically shortens the tool's runtime. The units used to buffer the stream network will match those of the input flow network.

a. If you are using the custom version of the tool, both the folder of *.las* files and a shapefile of training data are required as input. The training shapefile should be created using ArcGIS's *Image Classification* toolbar, and each entry should represent a distinct landcover class (Figure 54). Though a stream network file can be specified for the custom RCL tool, this will not alter runtime because the full extent of the rasters are generated regardless for model training purposes.

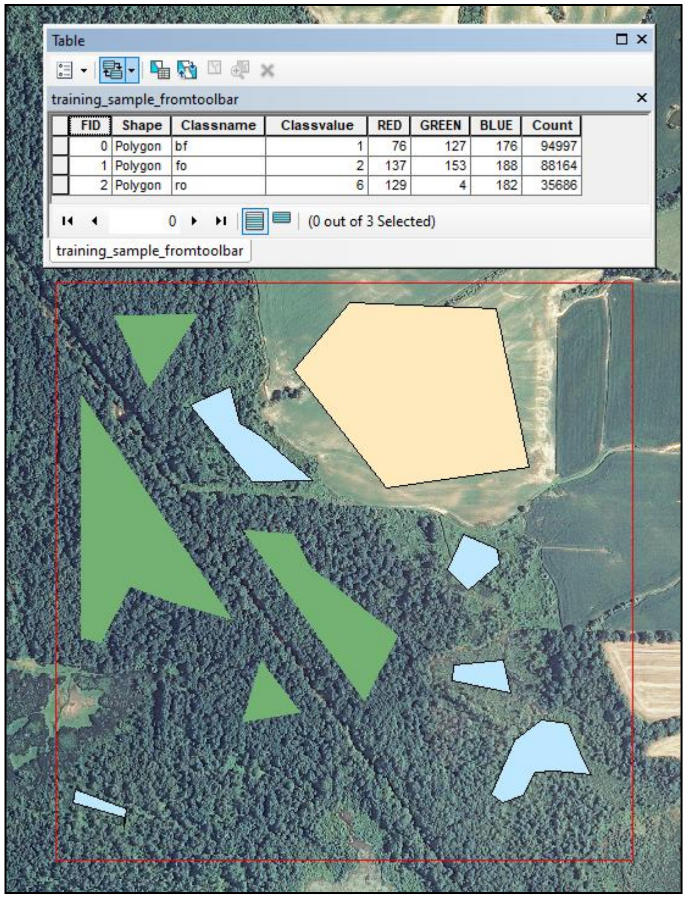


Figure 54: An example training shapefile for use with the custom version of the RCL tool, created using the Image Classification toolbar. Three classes (bare field, forest, and rough vegetation) are present. The landcover classification file created by the custom RCL tool will map onto the FID of the input training file (e.g., a GRIDCODE of 0 in the output shapefile corresponds to the entry with an FID of 0 in the input training file).

IMPORTANT: Your *.las* files MUST be in a projected coordinate system that uses either meters or feet as its XY units. Though the tool will automatically account for the XY units supplied; if you are using the pre-trained RCL tool, you must also supply a scale factor that converts the vertical units to meters if the vertical units are not meters. If the vertical units of the *.las* files are feet, supply 0.3048 as the scale factor. If you are using the custom tool, a z-factor is optional.

When ready, hit run. Processing time varies with study area size and LiDAR point density. For large study areas (*.las* files totaling over 5gb) processing times of 30 minutes and beyond are common if a stream network shapefile is not supplied. Tool progress can be monitored under “Messages” in the Results pane, which is accessible in the Geoprocessing dropdown menu. The tool will progress through steps similar to the following, depending on the model version and input:

Generating footprint and .lasd>Generating DSM>Generating DEM>

Generating DHM>Generating slope rasters>Classifying cover>Classification complete

Once the tool has finished running, the classification shapefile can be found in the output folder specified in the tool parameters along with a support folder full of derived data products[1]. The values in the classification file will depend on whether the classification scheme selected was “binary” or “ternary” (if the pretrained RCL tool was used) or the FID codes of the input training file (if the custom RCL tool was used).

9.4 Model Details and Limitations

It is recommended that all LiDAR data for the study area is coterminous. The tool generates interpolated rasters, meaning data gaps due to distant, unconnected LiDAR tiles will be interpolated across even if the distance is large. This results in expensive, unnecessary computations that slow classification speed.

Additionally, though this model is most accurate within the riparian corridor, it will output a raster that classified the entirety of the LiDAR input unless a clipping shapefile is supplied (Figure 55). Because classification done outside the riparian corridor has limited accuracy when using the pretrained RCL tool, it is recommended that the user either clip the output classification using a riparian buffer polygon or (preferably) supply a stream network shapefile at runtime. This is less important for the custom RCL tool, whose accuracy is largely dependent on the quality of the training data.

[1] The projection and units of the support rasters will match that of the input *.las* files. If the input files use meters as the XY unit, then the cellsize will be 1. However, if the input files use feet, then the cell size will be 3.28084, equivalent to a meter. Height values will be scaled by the z factor supplied, if it was supplied at all.

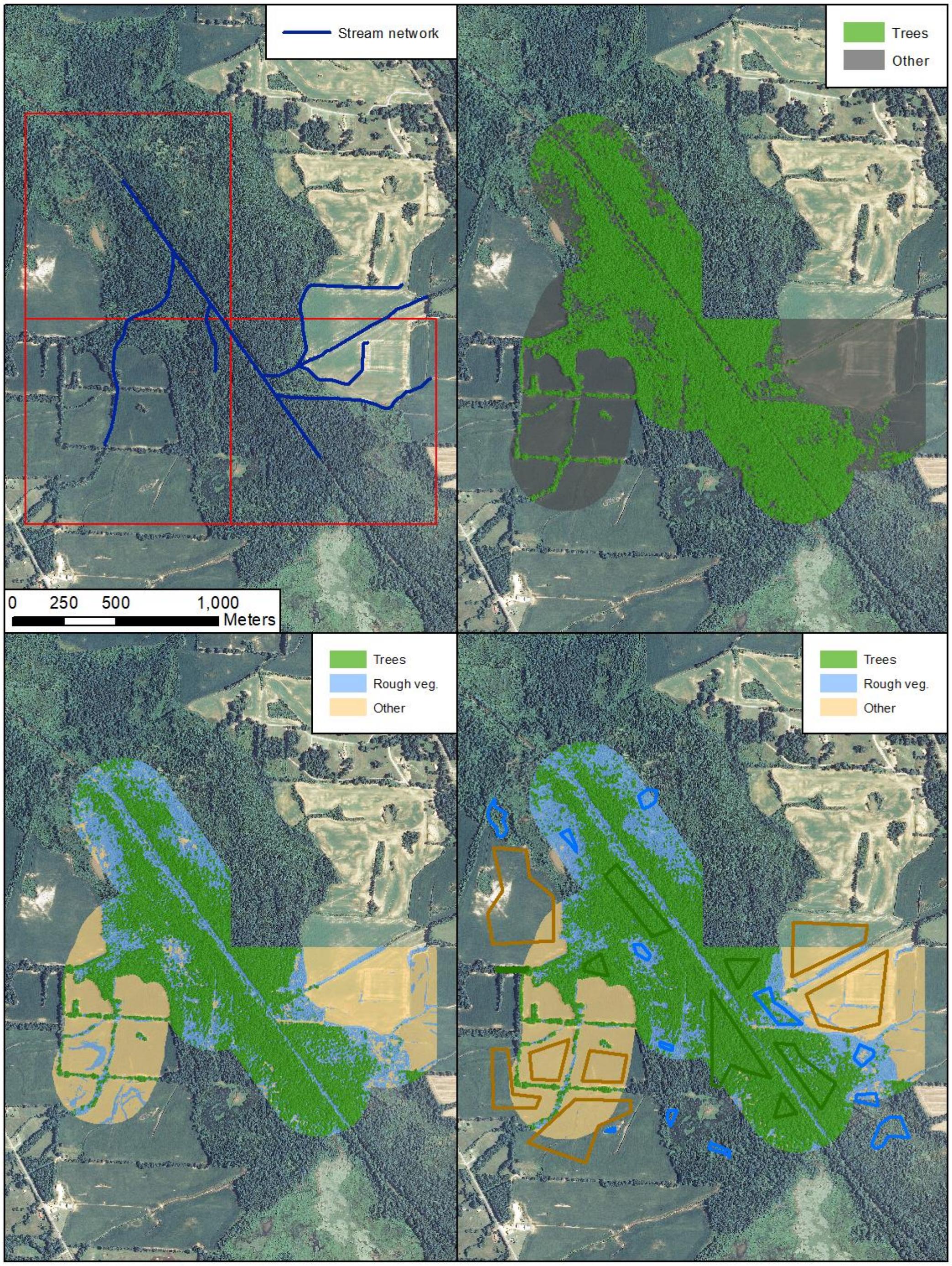


Figure 55: Sample outputs for various version of the RCL model. Output was restricted to the 250m buffer surrounding the stream network. Upper left: footprint of input .las files and stream network. Upper right: pretrained binary model. Bottom left: pretrained ternary model. Bottom right: custom trained model and training classes used to train model.

10 Appendix 4: Project Tasks Status and Deliverables

The proposal outlines the following target tasks. The table summarizes the status of completion of each task

Tasks (as featured in the proposal)	**Status**
Data Collection and Pre-Processing	Done (Report – Section 3)
Models Databases Development and Application	Done (Report – Sections 3 and 4)
Models Validation	Due to the unavailability of stream flow gauge data, an ungagged approach was adopted as explained in the proposal
Conservation Scenarios Simulation	Done (Report – Section 4)
Results Interpretation and Documentation	Done (Report – All Sections)
Riparian Buffer Tool and Datasets	Done (Report – Section 5 and Appendix 3)

In addition to the report, we are delivering to the funding agency the following datasets:

(1) Riparian buffer tool final version
    a. Tool
    b. Tutorial
    c. Sample data
        i. Lidar data
        ii. Training data

(2) Riparian buffer maps for the investigated watersheds
    a. System 1
    b. System 2

(3) EXTRA: Input Datasets for the 10 modeled watersheds (CDL, DEM, Climate, Soils and more)
    a. System 1
    b. System 2