VTRC
Virginia Transportation
Research Council

Leveraging Freely Available Remote Sensing and Ancillary Datasets for Semi-Automated Identification of Potential Wetland Areas Using a Geographic Information System (GIS)

http://www.virginiadot.org/vtrc/main/online_reports/pdf/16-r21.pdf

BENJAMIN R. FELTON
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Virginia

JONATHAN L. GOODALL, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering
University of Virginia

G. MICHAEL FITCH, Ph.D.
Associate Principal Research Scientist
Virginia Transportation Research Council

Final Report VTRC 16-R21

VIRGINIA TRANSPORTATION RESEARCH COUNCIL
530 Edgemont Road, Charlottesville, VA 22903-2454
www.VTRC.net

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 16-R21	2. Government Accession No.:	3. Recipient's Catalog No.:
4. Title and Subtitle: Leveraging Freely Available Remote Sensing and Ancillary Datasets for Semi-Automated Identification of Potential Wetland Areas Using a Geographic Information System (GIS)		5. Report Date: June 2016 6. Performing Organization Code: VTRC 16-R21
7. Author(s): Benjamin R. Felton, Jonathan L. Goodall, Ph.D., and G. Michael Fitch, Ph.D.		8. Performing Organization Report No.: VTRC 16-R21
9. Performing Organization and Address: Virginia Center for Transportation Innovation and Research 530 Edgemont Road Charlottesville, VA 22903		10. Work Unit No. (TRAIS): 11. Contract or Grant No.: 106482
12. Sponsoring Agencies' Name and Address: Virginia Department of Transportation 1401 E. Broad Street Richmond, VA 23219	Federal Highway Administration 400 North 8th Street, Room 750 Richmond, VA 23219-4825	13. Type of Report and Period Covered: Final Contract 14. Sponsoring Agency Code:
15. Supplementary Notes:		

16. Abstract:

The purpose of this study was to develop a wetland identification tool that makes use of freely available geospatial datasets to identify potential wetland locations at a spatial scale relevant for transportation corridor assessments. The tool was developed to assist the Virginia Department of Transportation in wetland identification over large geographic regions. Wetland identification is an integral part of many construction projects performed by state departments of transportation. However, current methods for wetland identification in support of these activities are lacking in one or more of the following ways: inadequate use of ancillary data, little automation, failure to leverage freely available data, excessive computation times, high expense, or the requiring of software not typically available to state departments of transportation.

This study addressed these limitations through development of a GIS-based wetland screening tool with freely available data and automated geoprocessing workflows to assist in wetland identification over large geographic regions. The tool was designed as a screening tool able to identify potential wetland areas that would require further investigation by a trained wetland identification expert. Therefore, the tool was designed to minimize false negatives: cases where the tool incorrectly designates wetland as non-wetland.

Application of the tool to a study region with detailed wetland delineations showed that the tool correctly identified wetlands nearly 70% of the time, produced false positives 24% of the time, and produced false negatives only 6% of the time. The tool allows decision makers to adjust the sensitivity of the wetland identification algorithm in order to decrease false negatives at the expense of increasing the fraction of the study area identified as potential wetland. The tool, therefore, allows decision makers to balance trade-offs between the amount of area requiring more detailed wetland identification and the frequency with which wetland areas are misidentified by the screening tool as false negatives.

Although the wetland identification tool was shown to be effective, future studies will be required to calibrate and validate the tool further using a broader range of application areas. The study recommends that this be done by way of additional corridor analyses to facilitate further improvements to the tool.

17 Key Words: Wetland identification, remote sensing, GIS, automated geoprocessing	18. Distribution Statement: No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.		
19. Security Classif. (of this report): Unclassified	20. Security Classif. (of this page): Unclassified	21. No. of Pages: 29	22. Price:

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

FINAL REPORT

LEVERAGING FREELY AVAILABLE REMOTE SENSING AND ANCILLARY DATASETS FOR SEMI-AUTOMATED IDENTIFICATION OF POTENTIAL WETLAND AREAS USING A GEOGRAPHIC INFORMATION SYSTEM (GIS)

Benjamin R. Felton
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Virginia

Jonathan L. Goodall, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering
University of Virginia

G. Michael Fitch, Ph.D.
Associate Principal Research Scientist
Virginia Transportation Research Council

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

Charlottesville, Virginia

June 2016
VTRC 16-R21

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

ABSTRACT

The purpose of this study was to develop a wetland identification tool that makes use of freely available geospatial datasets to identify potential wetland locations at a spatial scale relevant for transportation corridor assessments. The tool was developed to assist the Virginia Department of Transportation in wetland identification over large geographic regions. Wetland identification is an integral part of many construction projects performed by state departments of transportation. However, current methods for wetland identification in support of these activities are lacking in one or more of the following ways: inadequate use of ancillary data, little automation, failure to leverage freely available data, excessive computation times, high expense, or the requiring of software not typically available to state departments of transportation.

This study addressed these limitations through development of a GIS-based wetland screening tool with freely available data and automated geoprocessing workflows to assist in wetland identification over large geographic regions. The tool was designed as a screening tool able to identify potential wetland areas that would require further investigation by a trained wetland identification expert. Therefore, the tool was designed to minimize false negatives: cases where the tool incorrectly designates wetland as non-wetland.

Application of the tool to a study region with detailed wetland delineations showed that the tool correctly identified wetlands nearly 70% of the time, produced false positives 24% of the time, and produced false negatives only 6% of the time. The tool allows decision makers to adjust the sensitivity of the wetland identification algorithm in order to decrease false negatives at the expense of increasing the fraction of the study area identified as potential wetland. The tool, therefore, allows decision makers to balance trade-offs between the amount of area requiring more detailed wetland identification and the frequency with which wetland areas are misidentified by the screening tool as false negatives.

Although the wetland identification tool was shown to be effective, future studies will be required to calibrate and validate the tool further using a broader range of application areas. The study recommends that this be done by way of additional corridor analyses to facilitate further improvements to the tool.

FINAL REPORT

LEVERAGING FREELY AVAILABLE REMOTE SENSING AND ANCILLARY DATASETS FOR SEMI-AUTOMATED IDENTIFICATION OF POTENTIAL WETLAND AREAS USING A GEOGRAPHIC INFORMATION SYSTEM (GIS)

Benjamin R. Felton
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Virginia

Jonathan L. Goodall, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering
University of Virginia

G. Michael Fitch, Ph.D.
Associate Principal Research Scientist
Virginia Transportation Research Council

INTRODUCTION

Wetlands are a vital natural feature inherently capable of many beneficial hydrological and environmental processes. Some of these benefits include stormwater runoff control, effluent and sediment control, and the provision of habitats for wildlife and plants. Many wetlands have been destroyed or repurposed for agricultural or development purposes (Ouyang et al., 2014). Because of this practice, approximately one-half of the original wetlands in the United States no longer exist (Klemas, 2011). The need to protect wetlands is now well known, and wetland protection is required by federal law and regulation.

As a result of these requirements, roadway development projects done by state departments of transportation (DOTs) often must consider a reasonable number of alternative designs within a corridor. Each of these potential designs is evaluated on a number of criteria, one of which is the corridor's environmental impact, in particular, the area of wetlands expected to be impacted during construction. This process is used to make an informed decision based on federal regulations protecting wetlands while also balancing the project's intended purpose and need as well as considerations of cost, existing technology, and logistics. DOTs must sufficiently prove that the selected corridor for a project meets these criteria, showing that it is the least environmentally damaging practical alternative (LEDPA) by, among other tasks, providing wetland delineations. The U.S. Army Corps of Engineers evaluates these corridors as the governing authority in wetland delineation. If the LEDPA corridor is selected and a permit is issued, the DOT will have federal approval, under Section 404 of the Clean Water Act, for construction. Other federal approvals beyond the LEDPA permit may also be required for some projects.

There are a number of different wetland types, such as marshes, swamps, bogs, and fens, with variations in soils, topography, climate, hydrology, water chemistry, and vegetation based on geographic locations (Cowardin et al., 1979; Federal Geographic Data Committee, 2013). Despite the large number of wetland types, they all share basic characteristics of soils, vegetation, and hydrology that are used to describe wetland areas (U.S. Army Corps of Engineers, 1987). Although detailed identification and field verification methods will always be necessary to identify wetlands conclusively, there is the potential to use datasets available through federal and state agencies within a geographic information system (GIS) to determine potential wetland areas and assist in the wetland identification process. Coordinating with the U.S. Army Corps of Engineers, DOTs can determine if alternative tools are appropriate for identifying potential wetlands when field delineations are unavailable. The National Wetlands Inventory (NWI) provides one example of doing such an analysis; however, it is widely acknowledged that NWI, being a national-scale data product, often lacks the accuracy required to support transportation decision-making.

Current practices for identifying wetlands for LEDPA assessments can range from simple methods such as referring to publicly available datasets such as the NWI, to more time-intensive efforts such as performing field delineations, to highly advanced and involved remote sensing methods such as using image analysis and geospatial software to execute a composite of different weighted classification techniques. These more advanced remote sensing methods for wetland identification often use much higher resolution data with a series of classification methods. As the resolution and intensity of the classification technique increase, computation time and cost for obtaining the required data resources increase as well. This can be problematic for streamlining projects by slowing the delivery of tasks and increasing costs for high-resolution data acquisition.

Overview of Wetland Identification Methods

Classification methods used in remote sensing involve identifying features from their spectral signature and characteristics. The classification process will designate certain pixels of a raster to a particular class based on the pixels' spectral properties and/or characteristics (Lu and Weng, 2007). Unsupervised classification finds statistical relationships within the data, whereas supervised methods use training data, generally "ground truth" data, to develop a characteristic signature for each land cover dataset for a particular region. To accomplish this, manually specified training datasets are designated for the supervised classification algorithm to reference (Lu and Weng, 2007; Lu et al., 2003; Tana et al., 2013). Supervised classification can also use object oriented neighborhood analysis to define the vegetative class of a pixel relative to adjacent pixels' classification (Yan et al., 2006).

Many classification and processing software packages are available to assist with the classification procedure. Some of the programs and classification approaches described in related literature are presented here. However, this is not an exhaustive list, and some software is capable of using a number of different classification methods. ERDAS, developed by Hexagon Geospatial, is capable of performing both unsupervised and supervised techniques using a number of variations of the maximum likelihood and fuzzy logic algorithms (Mwita et al., 2013). Mwita et al. (2013) classified multispectral images using the ERDAS unsupervised Iterative Self-Organizing Data Analysis (ISODATA) method. Other studies used Esri's Image Classification in ArcGIS. The Image classifier provides classification methods for Maximum Likelihood, Iso Cluster Unsupervised, Class Probability, and Principle Components classification. Trimble's Definiens Developer (eCognition) is capable of nearest neighbor object oriented classification using DELPHI 2, which classifies using a combination of nearest neighborhood and fuzzy functions (Nobrega et al., 2011). Exelis Visual Information Solutions (ENVI) can be used for the multispectral imagery geocorrection and atmospheric corrections and is capable of a number of classification methods (Sugumaran et al., 2004). PANCHROMA can be used for pan-sharpening and gap-filling imagery (Lee, 2011).

Overview of Key Datasets

Many freely available geospatial datasets can be leveraged to identify wetlands. The following geospatial datasets are described in the literature as common datasets used in wetland identification procedures: (1) digital elevation models (DEM) and light detection and ranging (LIDAR) measurements to characterize wetland topographical aspects, particularly the slope, curvature, canopy height, and depression locations; (2) multi- and hyper-spectral satellite and aerial imagery data to provide supporting detail about plant vegetation type and soil moisture using the specified bands available; (3) National Resources Conservation Service's Soil Survey Geographic (SSURGO) data to characterize soils, particularly hydric soils; (4) the U.S. Geographical Survey's (USGS) National Hydrography Dataset (NHD) to identify bodies of water; and (5) the U.S. Fish and Wildlife Service's NWI (O'Hara, 2002; Stein et al., 2012).

Current State of Practice

Many DOTs have explored the use of geospatial software to automate the process of identifying potential wetland locations. The North Carolina DOT focuses on the use of high resolution LIDAR, SSURGO, and the National Land Cover Database (NLCD) using ArcGIS to automate the process to accomplish this. Although the North Carolina DOT is focused on automating this process, the primary governing dataset used is LIDAR, which lacks the multi- or hyper-spectral imagery that has been shown to increase accuracy (Laymon et al., 2001).

The Mississippi DOT uses satellite imagery, aerial photographs, land use and land cover (LULC) data, and DEM data within ERDAS Imagine and Definiens' eCognitio to accomplish this. However, the DOT's methods make it difficult to automate the process by using multiple software packages and requiring users to tend to the workflow from step to step. This method also uses multispectral imagery that is not freely available (Repaka et al., 2004).

The Colorado DOT exercises the most extensive use of multi- and hyper-spectral imagery by using National Agriculture Imagery Program (NAIP) Landsat 7 ETM+, Terra ASTER, and EO-1 Hyperion/ALI datasets. Although using three different spectral imageries may increase accuracy, this would result in increased computational time costs. This method also lacks ancillary datasets that have been shown to increase accuracy in wetland identification (Stein et al., 2012).

The Michigan DOT has developed a tool using SSURGO datasets and datasets derived from multispectral imagery. Although this tool is close to the level of autonomy and accuracy desired, the need for isolated derived datasets as input causes this method to hinder the usability of the tool (Shuchman and Court, 2009).

PURPOSE AND SCOPE

The purpose of this study was to develop a wetland identification tool that makes use of freely available geospatial datasets to identify potential wetland locations at a spatial scale relevant for transportation corridor assessments. The tool was to be developed to assist the Virginia Department of Transportation (VDOT) in wetland identification over large geographic regions. Studies have shown an opportunity to improve the wetland identification process used by DOTs by leveraging newly available remote sensing techniques and GISs (Ghobadi et al., 2012; Ozesmi and Bauer, 2002).

This study advanced this past work by creating a tool that (1) uses only freely available public datasets and (2) automates many of the data processing steps required to transform input datasets into a wetland screening map. If these tools can be used by DOTs for wetland screening early in the planning phase of a project, it could offer several benefits such as reducing personnel-hours in the field (and associated costs of field studies), targeting expert wetland identification efforts on locations with a higher probability of containing wetlands, and expediting approval processes in order to streamline project delivery.

METHODS

Study Area

The study area is an approximate 17-mile corridor in Virginia surrounding U.S. Route 460 between the town of Zuni and the city of Suffolk (Figure 1). The analysis was done for the 26 12-digit hydrologic unit codes (HUCs) that intersect this corridor for a total area of approximately 597,780 acres. The corridor falls within the Coastal Plains, one of five physiographic regions in Virginia. The general wetland composition of the study area was forested wetlands and included cypress gum, swamp tupelo, and mineral soil pine flats.

The corridor also falls within the Middle Atlantic Coastal Plain, one of seven ecoregions in Virginia. The U.S. Environmental Protection Agency describes this ecoregion as a flat plain with many swampy or marshy areas. Forest cover consists primarily of loblolly-shortleaf pine mixed with patches of oak, gum, and cypress near major streams. The central and southwestern portions of this region are poorly drained soils, whereas the northeastern portions are not as poorly drained. The central and southwestern regions account for approximately 15% cropland coverage, whereas the northeastern can range from 20% to 40% cropland coverage.

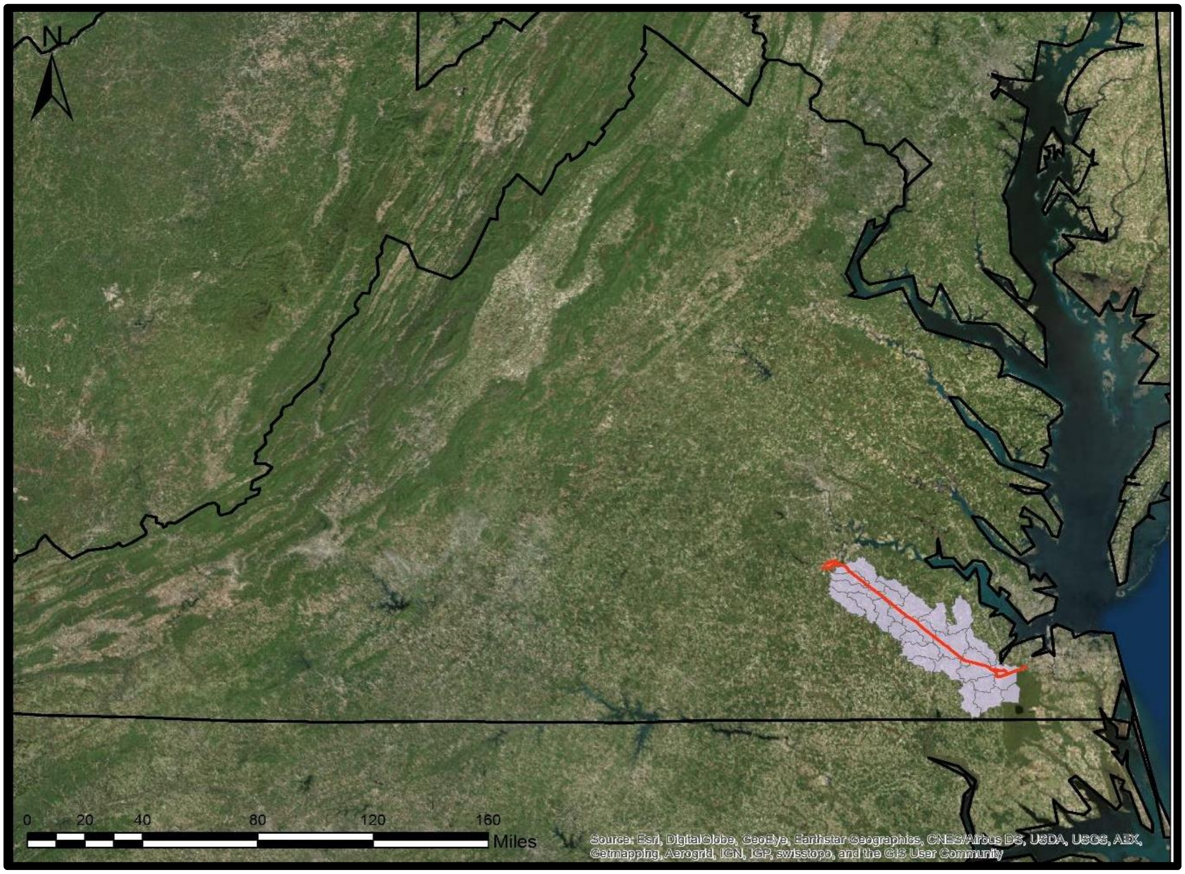

Figure 1. Site Location for Study Showing U.S. 460 Segment and 12-Digit HUCs

Data Preparation

Key datasets useful for wetland identification (Figure 2) were obtained, processed, and organized for use in the wetland screening tool. All datasets were organized into a specific structure with a defined naming convention for use by the wetland screening tool (Figure 3). Datasets were projected to the same coordinate system and clipped to the same boundary. This study used the NAD 1983 Virginia South State Plane coordinate system, and the data were clipped to the previously mentioned 12-digit HUCs. The datasets and formats required for the tool were as follows: DEM as raster data, 100-year floodplain map as polygon vector data, Landsat 8 Operational Land Imager (OLI) multispectral satellite imagery as raster data (applicable to Bands 2 through 7 from the OLI sensor), SSURGO as polygon vector data, NHD as polyline vector data, NLCD as raster data, NWI as polygon vector data, Watershed Boundary Dataset (WBD) HUC areas as polygon vector data, and training data as polygon vector data. These data are explained here.

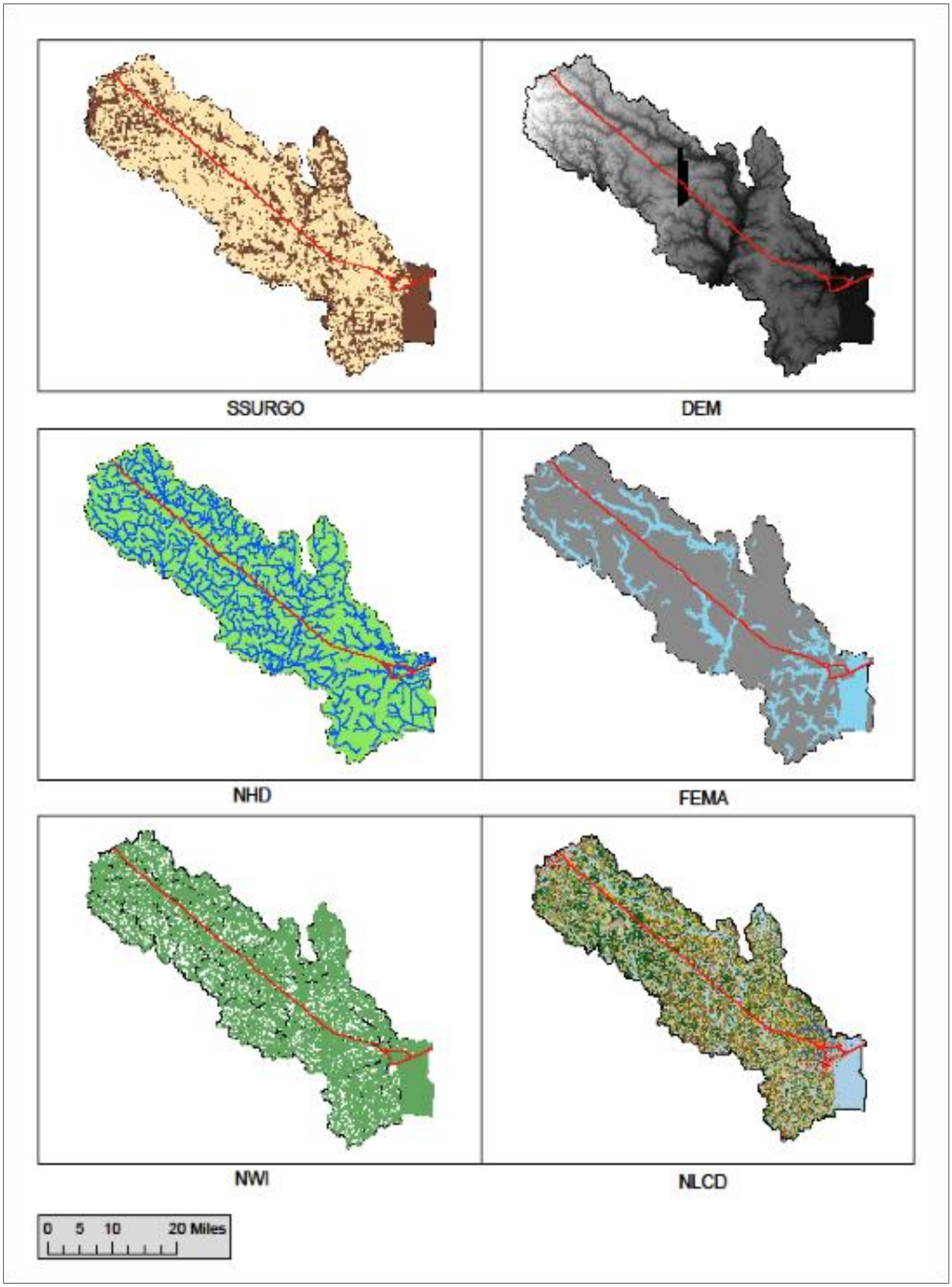

Figure 2. Key Input Geospatial Datasets (Excluding Landsat 8 Imagery) Used in Wetland Screening Tool

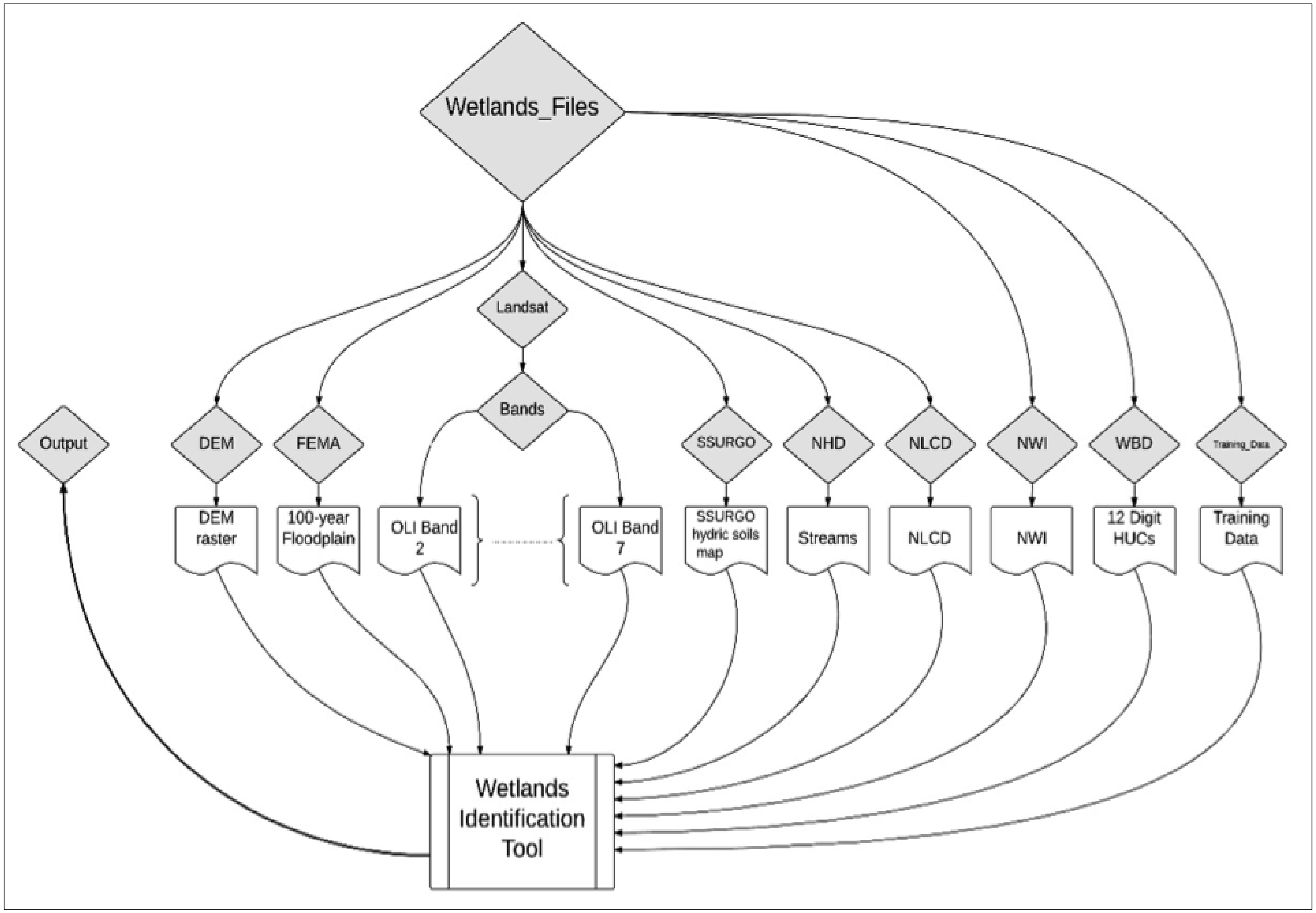

Figure 3. Image of Data Structure Hierarchy Required by Tool

Digital Elevation Model (DEM)

DEMs provide topographical information that can be used to derive regions where there is a high likelihood of pooled water. DEMs were downloaded from the USGS's National Map Viewer (USGS, 2016a). A DEM was created using the 1/9th and 1/3rd arc-sec National Elevation Dataset (NED), which correlates to resolutions of approximately 3.14 and 10.22 meters for the study area, respectively, by downloading and merging tiles for the area of interest. The 1/3rd resolution DEM data were resampled to match the 1/9th resolution before merging. Although DEM data were readily available for this study's area of interest, a 7,720-acre section found through 3 HUCs was left void because of the lack of 1/9th and 1/3rd resolution elevation data. This can be seen in Figure 3 toward the center of the study's area of interest. This section could have been supplemented with 1 arc-sec data, but it was left blank and excluded from analysis since this study was focused on the use of higher resolution elevation datasets.

Federal Emergency Management Agency (FEMA) Floodplain Maps

Floodplain maps are used to identify areas of water inundation for heavy storm or flood events: 100-year floodplain maps were downloaded from FEMA's Flood Map Service Center (FEMA, 2016). The 1% annual chance flood zone designations of Zone A, Zone AO, Zone AH, Zones A1-A30, Zone AE, Zone A99, Zone AR, Zone AR/AE, Zone AR/AO, Zone AR/A1-A30, Zone AR/A, Zone V, Zone VE, and Zones V1-V30 are all referred to as 100-year floodplain, or base flood, zones. All other zones categorized 500-year or more flood events.

Landsat 8 Operational Land Imager (OLI) Multispectral Satellite Imagery

Multispectral imagery is used to classify wetland spectral signatures and derive vegetative indices and vegetation analysis transformations from training data provided by image analysts. Landsat 8 OLI imagery and derived products were downloaded from USGS's Earth Resources Observation and Science Center's Science Processing Architecture On Demand interface (USGS, 2016b). This service provides a multitude of derived datasets that include conversion from digital numbers to top of the atmosphere and surface reflectance values using the Second Simulation of a Satellite Signal in the Solar System (6S) radiative transfer models and atmospheric corrections using MODIS correction routines. The Landsat 8 satellite follows the World Reference System (WRS-2) near-polar, sun-synchronous orbit. One orbit is approximately 99 minutes and provides a temporal resolution of complete coverage of the Earth every 16 days. Table 1 provides the OLI bands, wavelengths, and resolutions.

LandsatLook Viewer was used to identify the appropriate scenes required for this study, which is imagery from July 6 and August 14, 2014. The goal in scene selection is to identify dates within or near the wet season. However, since precipitation rates are evenly distributed throughout the year, July was isolated as the time frame of interest because it was historically the wettest month for the area of this study. The Scene IDs are designated as LC80140352014187LGN00 located on Path 14 and Row 35 with 10.23% cloud cover and LC80150342014226LGN00 located on Path 15 and Row 35 with 0.88% cloud cover. From manual image interpretation, no cloud cover was present over the portion of the imagery covering the study area.

Table 1. Landsat 8 Operational Land Imager (OLI) Band Details

Band	Wavelength (µm)	Resolution (m)	Description
1	0.43-0.45	30	Coastal Aerosol
2	0.45-0.51	30	Blue
3	0.53-0.59	30	Green
4	0.64-0.67	30	Red
5	0.85-0.88	30	NIR
6	1.57-1.65	30	SWIR 1
7	2.11-2.29	30	SWIR 2
8	0.50-0.68	15	Panchromatic
9	1.36-1.38	30	Cirrus
10	10.60-11.19	100	TIRS 1
11	11.50-12.51	100	TIRS 2

Soil Survey Geographic (SSURGO) Database

The SSURGO database provides valuable information about soil moisture content. Wetland regions generally consist of hydric soils. These hydric soils are provided by the SSURGO database in the form of polygon shapefiles. SSURGO datasets were downloaded from the Natural Resources Conservation Service's Web Soil Survey (U.S. Department of Agriculture, 2016). Data were downloaded on a per county basis for the following counties: Isle of Wight County (VA093), Prince George County (VA149), Southampton County (VA175), Surry County (VA181), Sussex County (VA183), Chesapeake City (VA550), Dinwiddie County (VA653), and City of Suffolk (VA800). The associated Access database (.mdb) was opened and linked to the associated tabular folder, which builds and loads data within the database. After building and filling the database was completed, ArcMap was used to import the soilsmu_a_va### polygon shapefile. The "Join" command was used to connect this polygon with data from the component table. Symbology was altered to represent the hydricating parameter within the component table. This parameter distinguishes hydric soils from non-hydric soils, which can be declared as *Yes*, *No*, or *Unknown.* This symbolized data layer was exported to retain the hydric classification from the join. Each of the county layers was then merged into a single polygon shapefile.

Other National-Scale Datasets

The USGS National Map Viewer was also used to obtain the statewide NHD, which provides stream location data found in the NHDFlowlines subset. The NLCD was downloaded from the Multi-Resolution Land Characteristics Consortium (MRLC, 2016). The NWI for the entire state of Virginia was downloaded from the U.S. Fish and Wildlife Service (2016). Finally, the WBD for the entire United States was downloaded from the USGS (USGS, 2016c) where the 12-digit HUCs covering this study's area of interest were exported into a new shapefile.

Training Data

Training data were manually created with the use of aerial imagery and the NWI (Figure 3). By the use of these data, known wetland areas were delineated throughout the study area to encompass a variety of wetland types and characteristics. These wetland areas were used to "train" or develop the classification algorithm. The classification algorithm makes use of information about wetland signatures extracted from each of the ancillary datasets within the areas designated as wetlands. Figure 4 shows a sample location of the training data depicting inland wetlands as green, river wetlands as purple, and non-wetlands as yellow.

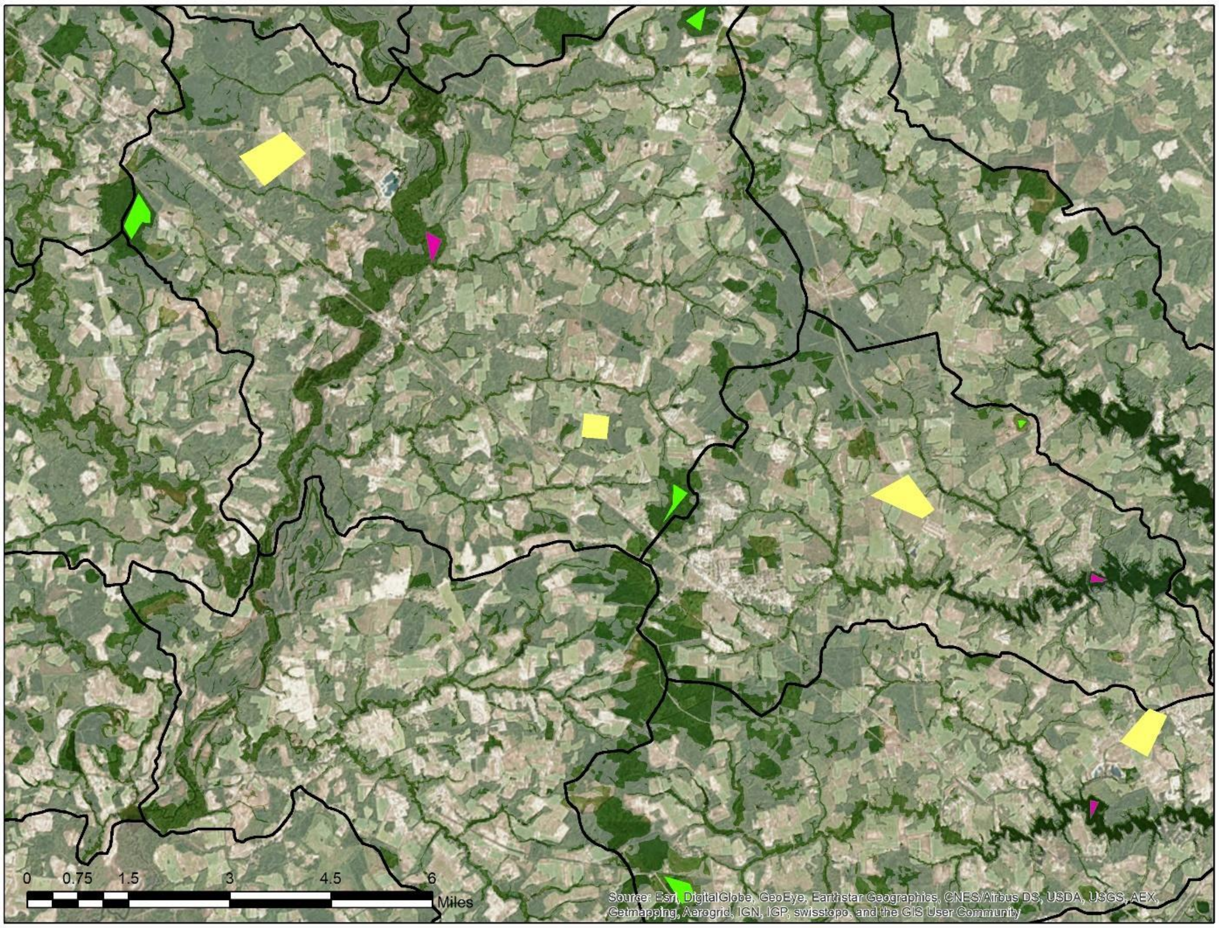

Figure 4. Sample Location of Training Data Depicting Inland Wetlands As Green, River Wetlands As Purple, and Non-Wetlands As Yellow. Training data were used to train the classification algorithm.

Verification Data

VDOT provided wetland areas created by wetland scientists for two corridor alternatives for U.S. 460 within the study area. These data were considered to be ground truth and used to assess the accuracy of the screening tool. The data were manually delineated by a trained image analyst using color-infrared imagery, land cover maps, NWI, SSURGO, NHD, LIDAR-derived DEM, and historical orthophotography, where analysts would pan the entire corridor looking at the visual cues indicating an area is wetland.

Tool Algorithm

Because Esri’s ArcGIS software is widely used and available to VDOT through a site license, it was selected as the platform for the development and incorporation of the potential wetland identification tool. To automate the geoprocessing steps, researchers used the

ModelBuilder tool in ArcGIS. ModelBuilder links a number of tools to allow users to run a series of processes without requiring user input.

Although ArcGIS is capable of a number of classification methods, Maximum Likelihood was selected as it is the more accepted method for classification. The Maximum Likelihood classification requires a manually generated training dataset in order for the classification to build a spectral profile for land cover types. Alternative classification methods include Iso Cluster, which is capable of classification without training data; however, output is organized into statistically clustered groups that then need intensive manual post-processing to merge clusters into appropriate land use / land class categories. The tool uses built-in ArcMap functions to execute a workflow that results in a final land use / land cover raster that identifies wetland locations.

Figure 5 depicts the overall workflow used by the wetland screening tool. The tool is segmented into four main sections: satellite imagery processing, DEM raster processing, riparian zone processing, and tertiary processing for all other ancillary datasets.

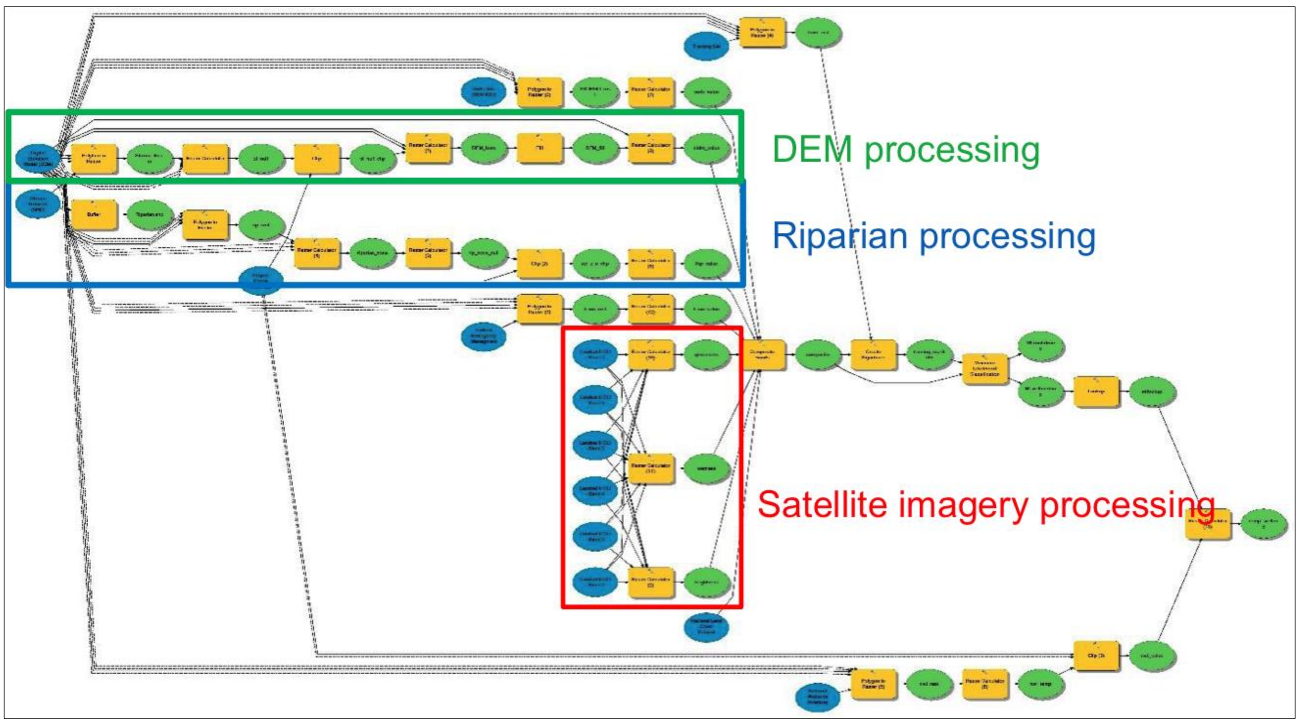

Figure 5. Flow Diagram for Potential Wetland Identification Tool. Figures 6 through 8 provide zoom-in views of the key components of the workflow that are highlighted with dashed lines in this figure.

The satellite imagery processing section (Figure 6) consists of the generation of three descriptive indices created from the use of a Tasseled Cap Transformation on the Landsat 8 OLI imagery for Bands 2 through 7. Bands 2, 3, and 4 are channels found in the visible spectrum, and Bands 5, 6, and 7 are channels found in the shortwave infrared and near infrared spectrum. The tool uses these bands and condenses them into the three indices, which describe the greenness, wetness, and brightness of an area. All of the rasters were created using Raster Calculator and Equations 1 through 3 describing the weighted sum of comments method for generating these rasters. Equations 1 through 3 describe the operations performed in the raster calculator tools in Figure 6. Table 2 provides the scalars used for weighting each band (Hasan Ali Baig et al., 2014).

$$Brightness = \sum_{i=2}^{7} (w_{1i} * band_i) \qquad \text{[Eq. 1]}$$

$$Greenness = \sum_{i=2}^{7} (w_{2i} * band_i) \qquad \text{[Eq. 2]}$$

$$Wetness = \sum_{i=2}^{7} (w_{3i} * band_i) \qquad \text{[Eq. 3]}$$

where w_{1i}, w_{2i}, and w_{3i} are the Tasseled Cap Transformation weighting factors for brightness, greenness, and wetness, respectively, for $band_i$, and $band_i$ is the spectral signature value for $band_i$ at a given pixel location.

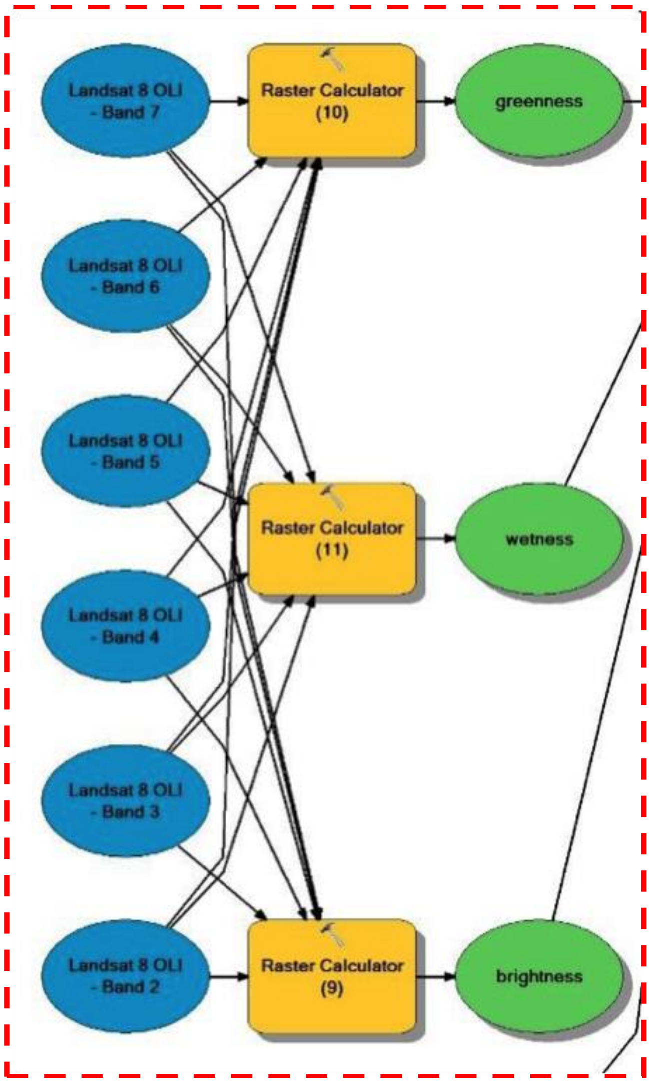

Figure 6. Satellite Imagery Processing

Table 2. Tasseled Cap Transformation (TCT) Coefficients for Landsat 8 OLI Sensor

Landsat 8	Blue	Green	Red	NIR	SWIR1	SWIR2
TCT	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
Brightness	0.3029	0.2786	0.4733	0.5599	0.5080	0.1872
Greenness	-0.2941	-0.2430	-0.5424	0.7276	0.0713	-0.1608
Wetness	0.1510	0.1973	0.3283	0.3407	-0.7117	-0.4559

NIR = near infrared; SWIR = shortwave infrared.

The DEM processing section (Figure 7) uses DEM data to compute a sink raster, which identifies depressions throughout the topography. The DEM is first conditioned using NHDFlowline converted to a raster. The NHDFlowline raster is multiplied by a large value, in this case 100 feet, and is subtracted from the DEM. This process is known as burning in streams. The Fill tool is then used to fill any depressions within the burned DEM. The filled and original DEMs are then used in Raster Calculator; pixels with changes in elevation are designated a value of 1, and pixels with the same elevation are designated a value of 0. The Fill tool is generally used to remove small imperfections in topography for flow path analysis; however, here it is used to identify depressed areas. Burning in streams is required to avoid cases where Fill may consider an extremely large area a depression. For example, a bridge crossing a stream would register at a higher elevation than the stream it is crossing and the Fill tool would fill all contributing areas up to the bridge, which would incorrectly lead to identifying these areas as depressions.

The riparian zone processing section (Figure 8) creates a buffer of 100 feet surrounding each NHDFlowline and converts the polygons to a raster, where Raster Calculator was then used to compute a binary raster where pixels within the 100-foot riparian zone are designated a value of 1 and all other areas are designated a value of 0. This process assumes wetlands are more likely to be within the riparian zone around rivers.

All tertiary processing in the overall workflow revolves around the creation of binary rasters that represent wetland traits. The tool assumes that the DEM is the highest resolution raster and is used as a processing constraint for the resolution, cell size, snap raster, and processing extent for all tools that involve conversion of vector data to raster data or the resampling of lower resolution data. FEMA data are converted from polygon to raster. Raster Calculator is then used to compute a binary raster where pixels within the floodplain are designated a value of 1 and all other areas a value of 0. SSURGO data are converted from polygon to raster. Raster Calculator is then used to compute a binary raster where pixels containing hydric soils are designated a value of 1 and non-hydric soils a value of 0. NWI data are converted from polygon to raster. Raster Calculator is then used to compute a binary raster where pixels containing wetland areas delineated by the U.S. Fish and Wildlife Service are designated a value of 1 and non-wetland areas a value of 0. The NLCD is incorporated as is. User-generated training data are converted from polygon to raster, specifying pixels with the appropriate land use land cover designation specified.

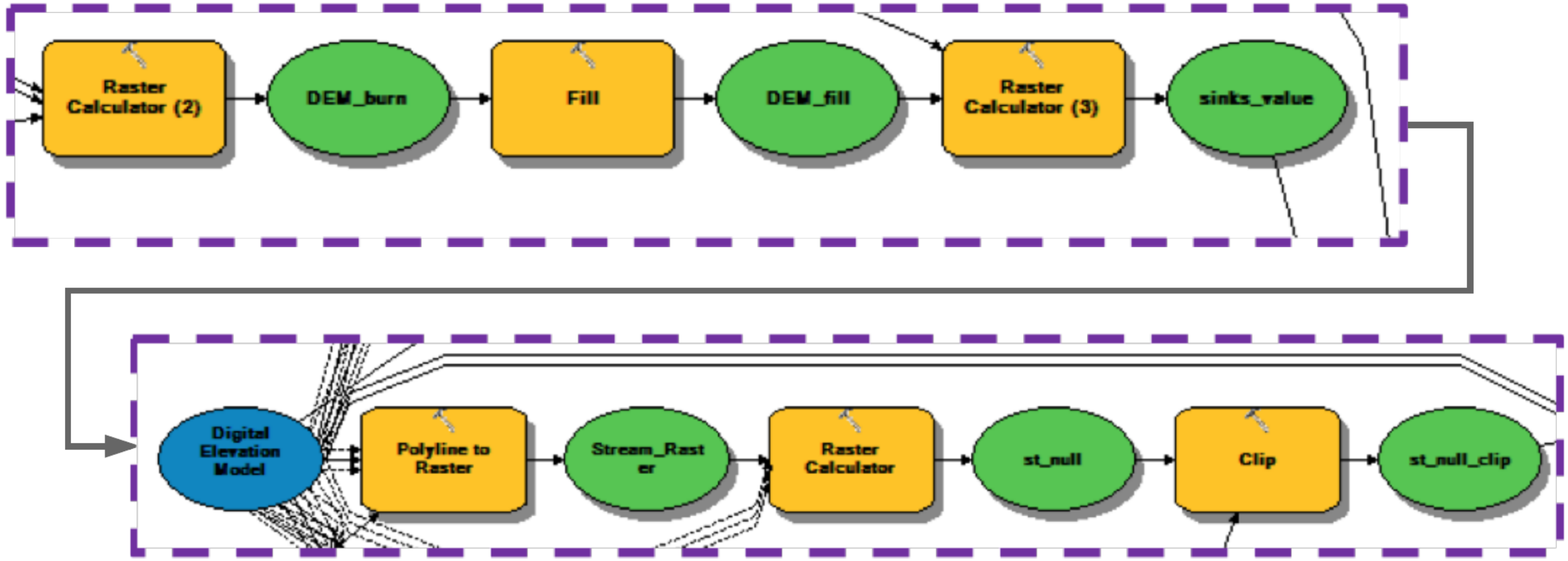

Figure 7. DEM Processing Workflow

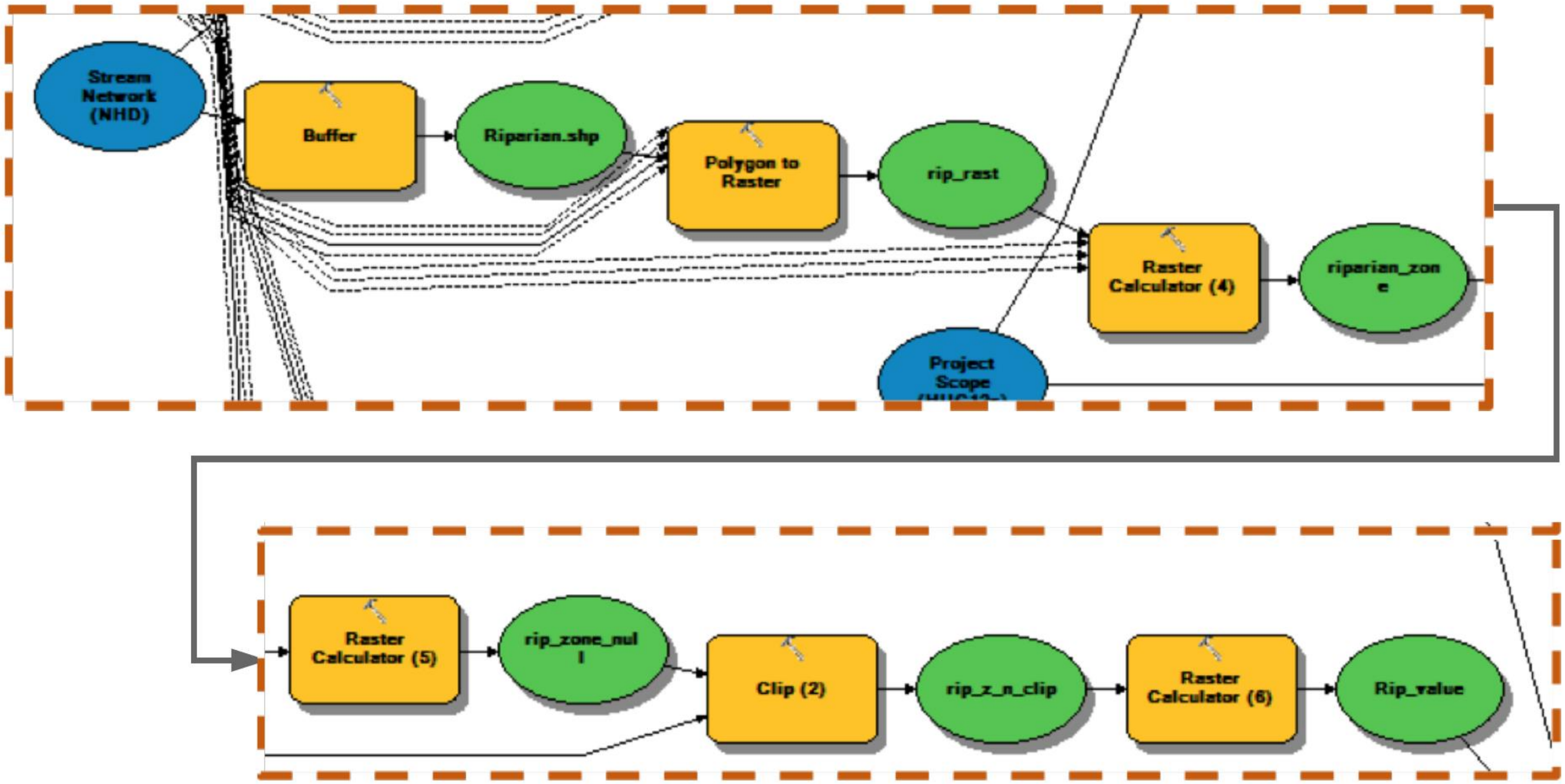

Figure 8. Riparian Processing Workflow

After all previously mentioned processing is complete, the tool composites each of these rasters into a comprehensive image, with the exception of the training data and NWI data. The rasterized training areas and composite image are then used to develop signatures for the particular known land classes found throughout the image. This builds a library of spectral signatures that is then used for the Maximum Likelihood classification; in this study's case, the land cover designations are river wetlands, inland wetlands, and non-wetlands. The result of these operations is a land use land cover raster mapping the user-specified land classes and a confidence raster describing the certainty of the Maximum Likelihood classification for 14 levels of confidence. The tool then merges the river wetlands and inland wetlands into one class and uses Raster Calculator to include NWI-designated wetlands.

RESULTS AND DISCUSSION

Model Prediction Validation

Figure 9 depicts two datasets that contain mapped wetlands: this study's model output and VDOT-identified wetlands resulting from a survey by trained image analysts. The color depictions for each of the datasets are as follows: model output for wetlands is green, model output for non-wetlands is beige, and the VDOT-delineated wetlands have a red outline.

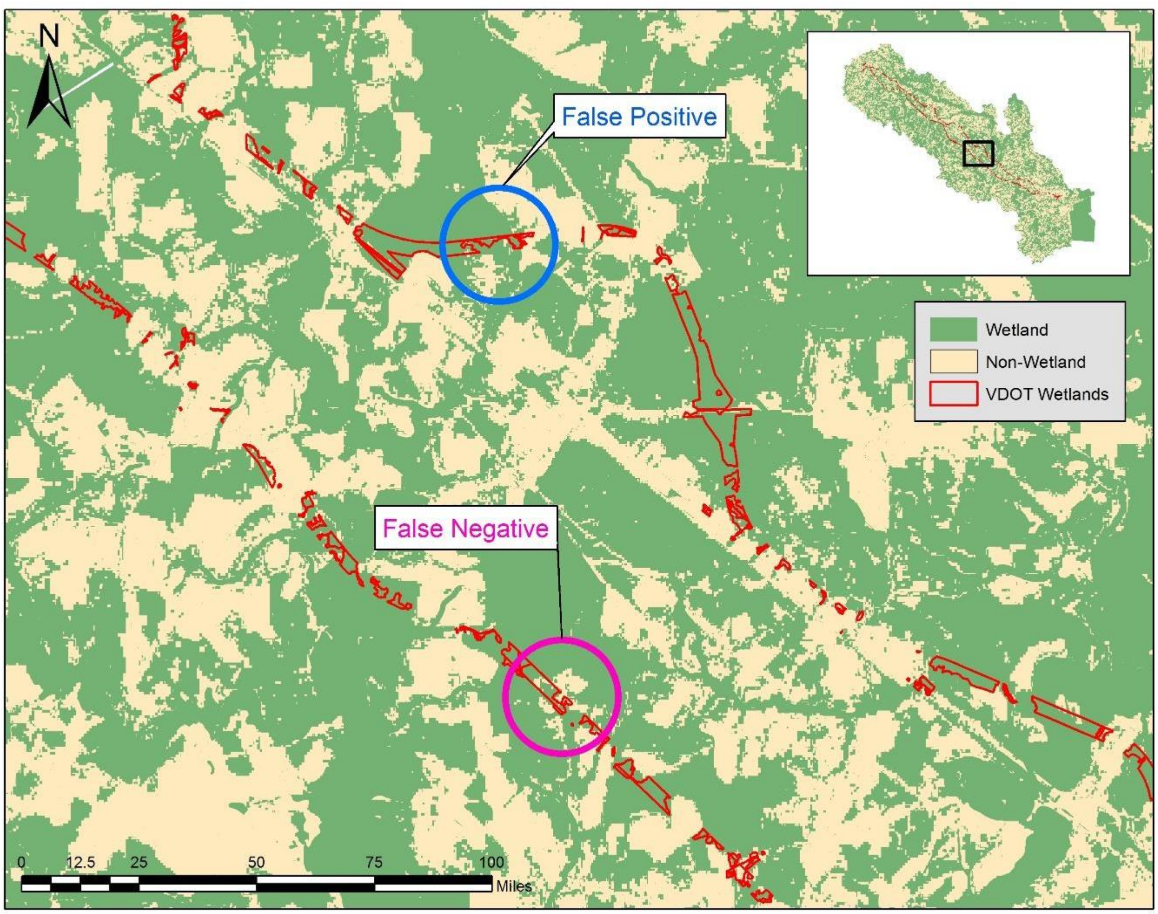

Figure 9. Survey vs. Model Predicted Wetlands

The model performed well for the region shown in Figure 10, with a low percentage of false negatives. However, there were still areas where false positives appeared. The purple outline identifies a false negative where the model predicted no wetland but the VDOT delineation identified a wetland. The blue outline identifies a false positive where the model predicted a wetland but the VDOT delineation did not identify a wetland. In terms of this study's goals of creating a potential wetland identification map that can be used to focus survey-based identification efforts, false positives are less concerning than false negatives. It is important to note that the VDOT data extend only to the associated corridor; therefore, model verification does not extend beyond the corridor extent.

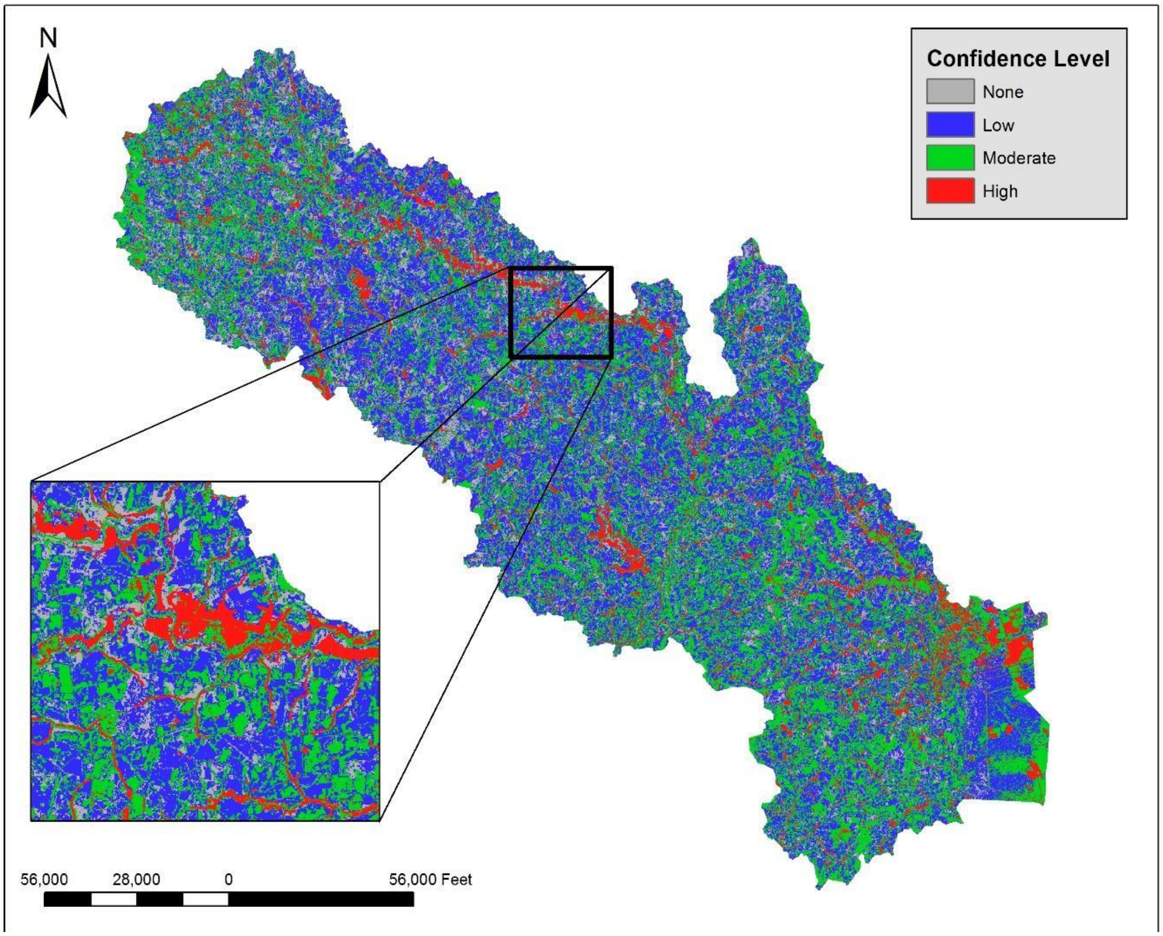

Figure 10. Grouped Confidence Raster

Confidence Level

Figure 10 presents the confidence raster associated with the model predictions. This raster is generated through ArcGIS's Maximum Likelihood classification and can be used to supplement the model projections for decision support applications. This raster depicts the Maximum Likelihood classification's confidence in classification of each particular pixel. The level of confidence ranges from values of 1 to 14, where lower values represent lower confidence and higher values represent higher confidence. These discrete levels were combined into four levels of confidence: none, low, moderate, and high. *None* spans values 1 to 2.9, *low* spans values 3 to 5.9, *moderate* spans values 6 to 9.9, and *high* spans values 10 to 14.

Accuracy Assessment

Figure 11 provides a comparison of the model output to VDOT-mapped wetlands, which are considered the ground truth for the accuracy assessment. The figure was generated by use of a raster difference calculation. For both the model output and VDOT-mapped wetlands, wetland locations are designated with a value of 1 and non-wetland areas with a value of 0. Raster Calculator was used to subtract the VDOT binary rasters from the model output binary raster. This results in false negatives being assigned a value of −1, shown in red; false positives being assigned a value of 1, shown in blue; and agreement between the two rasters being assigned a value of 0, shown in green.

The tool is configured to minimize false negatives (predicting no wetland when there is in fact a wetland) in order to focus survey efforts for wetland delineation on areas that have potential wetlands. For high levels of accuracy in identifying as many actual wetland locations as possible, reducing the number of false negatives is extremely important, whereas minimizing the number of false positives is much less important. The tool can be reconfigured to meet other objectives, such as minimizing both false positives and false negatives if simple prediction of actual wetlands is the primary need of the decision maker.

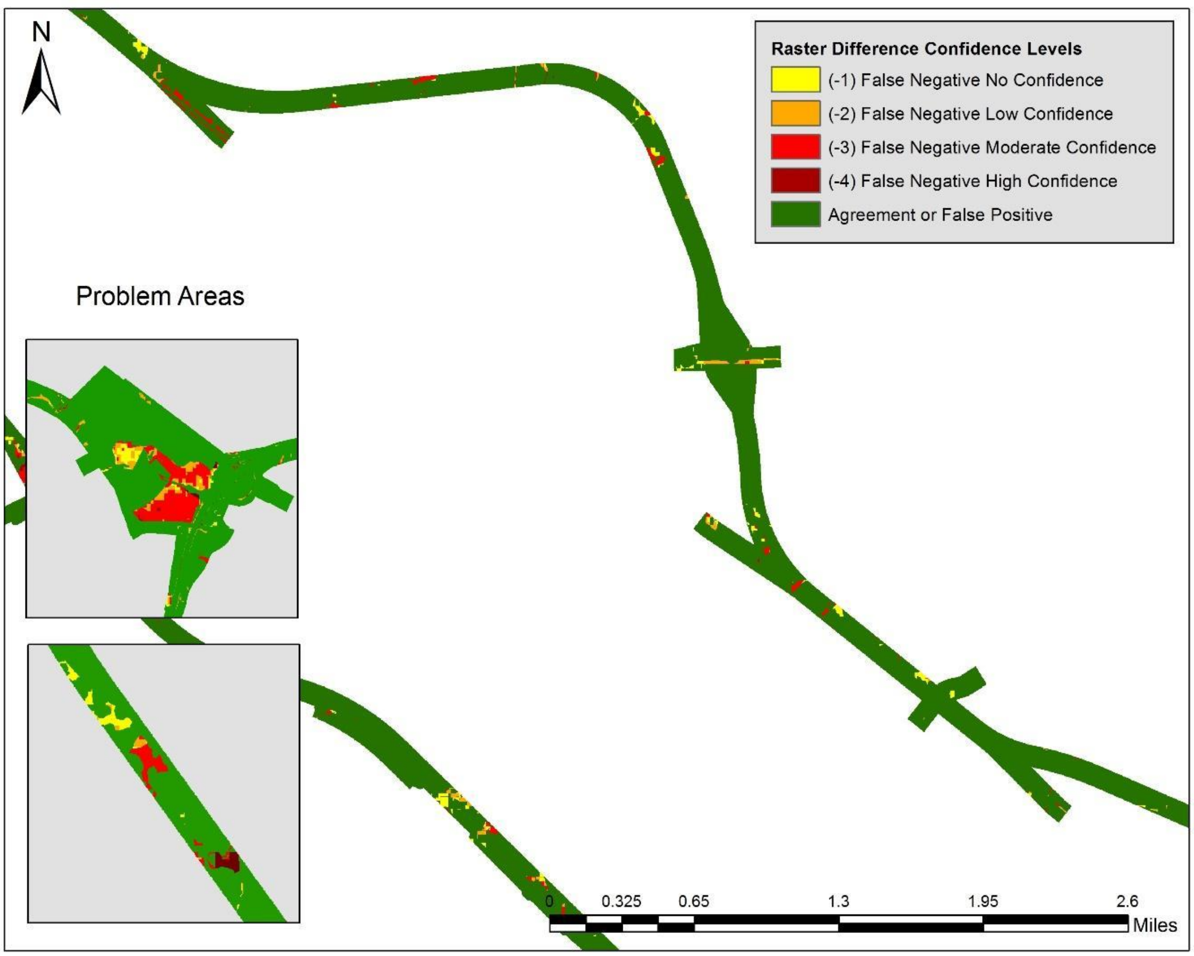

Figure 11. Confidence Levels for False Negative Model Predictions

In Figure 12, the bottom left images are focused on two prominent problem areas found within the model output where false negatives were high. Future work should be directed to achieving a better understanding of the reason for these clustered regions of false negative predictions. Potential reasons for these errors include missing information in the wetland identification algorithm. If so, there may be unique characteristics of these locations that could be incorporated into the prediction tool to remove these false negative predictions. It is also possible that there have been recent land changes in these regions that are not reflected in the underlying datasets used in the prediction tool.

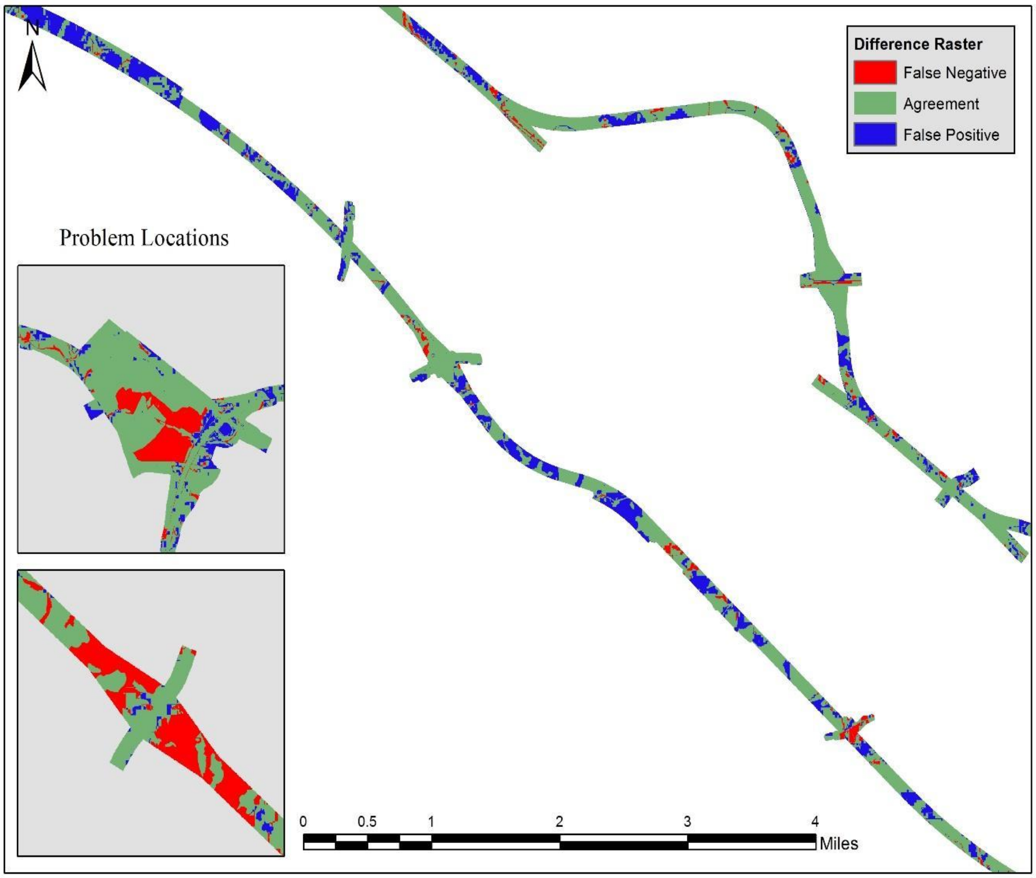

Figure 12. Locations of False Negative, Agreement, and False Positive Predictions Between the Wetland Screening Tool and a More Detailed Wetland Survey Effort

Table 3 outlines the total area for false negatives, false positives, and agreement pixels. The tool agreed with the VDOT wetland delineations for 69.3% of the study area. For the remaining portion of the study region, the majority of this area included false positive errors (24.3%) where the model predicted a wetland but no wetland existed according to the VDOT delineation. Only 6.4% of the study region had false negatives. Again, because the primary goal of the tool is to minimize false negatives, this 6.4% is an important performance metric for the model. If this tool had been used as a preliminary screening tool that focused wetland delineation efforts in potential wetland areas identified by the tool, then although this would streamline delineation efforts by reducing the area required for wetland delineation, 6.4% of the wetlands in the study area would have been missed in the original wetland impact estimate, potentially affecting decisions that are predicated on minimizing wetland impacts.

Table 3. False Negative, Agreement, and False Positive Areas for Both Corridors

Classification Result	Area (acre)	Percent Area (%)
False negatives	523	6.4
Agreement	5,669	69.3
False positives	1,984	24.3

Using Model Predictions With Confidence Level

Figure 11 presents a composite of the model predictions with the confidence level added for false negative areas. The confidence level provides important information to decision makers that can be used when determining the benefits and costs of focusing wetland delineation efforts. For example, if a high level of confidence is needed, then the decision maker may wish to survey all pixels that have a minimal, low, or moderate confidence level, even if the tool determined these pixels to be non-wetland areas. This would increase the total area that required surveying, therefore increasing the cost and time required to complete the surveying, but it would reduce the number of missed wetland areas. Table 4 provides the area within each confidence level for the false negative predictions along with its percentage of the total site area and the percentage of false positives in that confidence level compared to the total site area. These data show that if all non-wetland "minimal confidence" and "low confidence" regions were included in the wetland delineation survey, it would have resulted in the need to survey 2,584 more acres (31.6% of the study region). At the same time, it would have reduced the false negatives from 6.4% of the study region to 2.9% of the study region.

Table 4. Confidence Levels for False Negative Model Predictions

Classification Confidence	Area (acre)	Percent Area (%)	Percent False Positives (%)
Minimal confidence	781	9.5	1.90
Low confidence	1,803	22.1	1.56
Moderate confidence	1,834	22.4	2.86
High confidence	123	1.5	0.09

Building on this idea, Figure 13 summarizes this trade-off between reducing the number of false negative predictions and increasing the area of the study region that must be surveyed. Given that the area of the region that must be surveyed is a surrogate for the cost and time required to complete the wetland delineation, this figure illustrates the trade-off between error (false positives) and cost (survey area) for the study region. Given this information, a decision maker may elect to reduce the percentage of false negatives from 6.4% to 4.5% by surveying an additional 9.5% of the project site area. The additional 9.5% of the site area that would be surveyed are pixels that were classified as non-wetland but with lower confidence. This relationship between error and cost is likely to be specific to this study region, and further work applying the wetland identification tool and performing surveyed wetland delineations for other regions would be necessary to gain insight into the regional variability of the error vs. cost relationship. If a general relationship were found, it could be applied for sites without survey data to foster understanding of the potential trade-off between error and cost for the wetland identification tool.

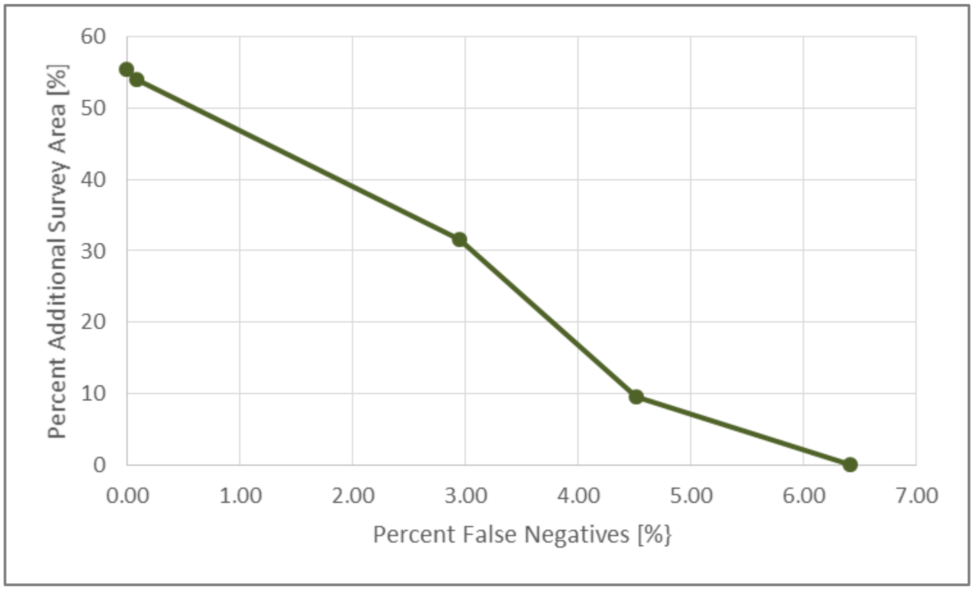

Figure 13. Relationship Between Level of Certainty in False Negative Predictions and Increased Area That Must Be Investigated Using More Detailed Wetland Identification Procedures

CONCLUSIONS

- *The automation of current wetland identification techniques using freely available datasets to isolate target sites for on the ground wetland delineation shows great promise.* With this approach, it was possible to obtain results in agreement with those from a trained image analyst's method for nearly 70% of the study region.

- *Because the wetland identification tool developed in this study was configured to minimize false negatives (predicting no wetland when there is in fact a wetland), the majority of disagreement between the results with the tool and the trained image analyst's method comprised cases where the tool identified a wetland and the analyst's method* did not *(false positives).* Only about 6% of the study region resulted in false negatives.

- *The tool's confidence raster provides important insights into the predictive capability of the tool and potential areas for improvement.* For example, the confidence raster indicates that the tool is easily able to identify lakes, rivers, and ocean pixels, which can be attributed to their characteristic spectral response of being highly absorptive in the near infrared bands, which is a stark contrast to other land cover types. Urbanized areas also generally had higher classification confidence levels because of their distinctive spectral response.

- *The coarse spatial resolution of Landsat 8 imagery appears to be the restrictive dataset that contributes to the tool's inaccuracy.* The pixelated mapping of the model output edges indicates that the multispectral imagery component of the model is governing, which is apparent in this imagery because of its coarse resolution in comparison to the other model components. Unfortunately, because of a poor spatial resolution of 30 meters, these areas can account for some error within the tool's classification. The spectral response for a single pixel is associated with an area of 900 square meters s, which may include a number of different land cover types, resulting in a pixel value that is not representative of this area and incorrect classification.

- *Although the wetland screening tool developed in this study was effective for the case study project, future studies will be needed to calibrate and validate the tool further using a broader range of application areas*.

RECOMMENDATIONS

1. *VDOT's Environmental Division, through its Environmental Research Advisory Committee, should decide if a comparison of the proposed model and VDOT's consultant model (using photointerpretation) is appropriate and feasible (considering the proprietary nature of the photointerpretation model).* The comparison could be done on either an existing or future corridor assessment and should concentrate on the accuracy, costs, and time requirements associated with each method.

2. *VDOT's Environmental Division should continue to explore potential areas of improvement in data and acquisition methods that could be used in the wetland screening process to improve the predictive capabilities of whatever method is ultimately used to identify wetlands.* In addition to freely available data as used in this study, there are some commercially available data products that may be worth exploring to improve the tool's predictive capabilities. For example, GeoEye's IKONOS satellite currently can provide multispectral imagery with resolutions up to 3.2 meters, but the data come at a cost (Satellite Imaging Corporation, 2016). Other data acquisition methods include the use of unmanned aerial vehicles in addition to satellites for data collection; LIDAR-derived digital elevation model datasets; and additional satellite-derived longer wavelength datasets such as that collected by the Soil Moisture Active Passive satellite (NASA, 2016).

BENEFITS AND IMPLEMENTATION

Benefits

Performing a direct comparison between the newly developed and currently used wetland identification methods will help VDOT select the most appropriate means of estimating the wetland impacts resulting from specific alignments as a part of corridor development studies. This will allow for the optimization of wetland identification methods used based on accuracy, cost, and timeliness, ultimately resulting in fewer problems and delays associated with alignment selection.

In addition, because of the rapid improvements in data acquisition technology, additional and better datasets will undoubtedly become available in the near future. Increases in spatial and spectral resolution are coming and, as a result, may also parallel increases in resolution with freely available imagery. In addition, newer satellites are offering hyperspectral imagery, which will drastically increase the number of available bands, giving the ability to develop more continuous spectral profiles. Keeping up to date on these new datasets will help make certain that VDOT continues to use the best methods available for corridor-level wetland identification.

Implementation

VDOT's Environmental Division will consider (1) the merits of comparing the wetland identification method developed in this study to its current method, and (2) the need to explore additional data acquisition methods through the project prioritization process of its Environmental Research Advisory Committee. This process will take place in November 2016. If it is determined that this additional analysis is warranted, it could be initiated as soon as spring 2017.

ACKNOWLEDGMENTS

This study was supported by funds from the Federal Highway Administration. The authors thank VDOT's Environmental Division for providing the data for this study and Angel Deem, Cooper Wamsley, Leo Snead, Scott Smizik, and Caleb Parks for their valuable guidance and feedback. This report also benefited greatly from the assistance of Catherine Harrison and Aaron Capelouto, summer undergraduate research interns, and Amy O'Leary and Linda Evans from the Virginia Transportation Research Council.

REFERENCES

Cowardin, L.M., Carter, V., Golet, F.C., and LaRoe, E.T. *Classification of Wetlands and Deepwater Habitats of the United States*. U.S. Fish and Wildlife Service, Washington, DC, 1979.

Federal Emergency Management Agency. FEMA Flood Map Service Center. 2016. https://msc.fema.gov/portal. Accessed June 23, 2016.

Federal Geographic Data Committee. *Classification of Wetlands and Deepwater Habitats of the United States.* FGDC-STD-004-2013. Second Edition. Wetlands Subcommittee, Federal Geographic Data Committee and U.S. Fish and Wildlife Service, Washington, DC, 2013.

Ghobadi, Y., Pradhan, B., Kabiri, K., Pirasteh, S., Shafri, H.Z.M., and Sayyad, G.A. Use of Multi-Temporal Remote Sensing Data and GIS for Wetland Change Monitoring and Degradation. In *IEEE Colloquium on Humanities, Science and Engineering Research,* Kota Kinabalu, Sabah, Malaysia, 2012, pp. 103-108.

Hasan Ali Baig, M., Zhang, L., Shuai, T., and Tong, Q. Derivation of a Tasselled Cap Transformation Based on Landsat 8 At-Satellite Reflectance. *International Journal of Remote Sensing*, Vol. 5, No. 5, 2014, pp. 423-431.

Klemas, V. Remote Sensing of Wetlands: Case Studies Comparing Practical Techniques. *Journal of Coastal Research*, Vol. 27, 2011, pp. 418-427.

Laymon, C., Cruise, J., Estes, M., and Howell, B. *Technology Guide: NCRSTE TG001: Assessing the Need for Remote Sensing Information to Conduct Environmental Assessment in Transportation.* National Consortium for Remote Sensing in Transportation– Environmental Assessment, Mississippi State University, Starkville, 2001.

Lee, S. *Detecting Wetland Change Through Supervised Classification of Landsat Satellite Imagery Within the Tunkwa Watershed of British Columbia.* University of Gävle, Gävle, Sweden, 2011.

Lu, D., and Weng, Q. A Survey of Image Classification Methods and Techniques for Improving Classification Performance. *International Journal of Remote Sensing*, Vol. 28, 2007, pp. 823-870.

Lu, D., Mausel, P., Brondízio, E., and Moran, E. Change Detection Techniques. *International Journal of Remote Sensing*, Vol. 25, No. 12, 2003, pp. 2365-2401.

Multi-Resolution Land Characteristics Consortium (MRLC). National Land Cover Database (NLCD). 2016. http://www.mrlc.gov. Accessed June 23, 2016.

Mwita, E., Menz, G., Misana, S., Becker, M., Kisanga, D., and Boehme, B. Mapping Small Wetlands of Kenya and Tanzania Using Remote Sensing Techniques. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 21, 2013, pp. 173-183.

NASA. Soil Moisture Active Passive Satellite. 2016. http://smap.jpl.nasa.gov. Accessed June 23, 2016.

Nobrega, R.A.A., O'Hara, C., and Stich, B. Top-Down Landscape-Based Approach Toward the Assess and Rank Watershed and Wetlands Impacted by Transportation Corridor. In *Transportation Research Record: Journal of the Transportation Research Board, No. 90*. Transportation Research Board of the National Academies, Washington, DC, 2011.

O'Hara, C.G. *Technology Guide: NCRSTE_TG003 Remote Sensing and Geospatial Application for Wetlands Mapping, Assessment, and Mitigation*. National Consortium for Remote Sensing in Transportation–Environmental Assessment, Mississippi State University, Starkville, 2002.

Ouyang, Z., Becker, R., Shaver, W., and Chen, J. Evaluating the Sensitivity of Wetlands to Climate Change with Remote Sensing Techniques. *Hydrologic Processes*, Vol. 28, 2014, pp. 1703-1712.

Ozesmi, S.L., and Bauer, M.E. Satellite Remote Sensing of Wetlands. *Wetlands Ecology and Management*, Vol. 10, 2002, pp. 381-402.

Repaka, S.R., Truax, D.D., Kolstad, E., and O'Hara, C.G. Comparing Spectral and Object Based Approaches for Classification and Transportation Feature Extraction From High Resolution Multispectral Imagery. In *ASPRS Annual Conference Proceedings*, ASPRS, Denver, Colorado, 2004.

Satellite Imaging Corporation. IKONOS Satellite Sensor. 2016. http://www.satimagingcorp.com/satellite-sensors/ikonos/. Accessed June 23, 2016.

Shuchman, R., and Court, G. Developing and Applying a Geospatial Decision Support Tool for Efficient Identification of Wetlands Mitigation Sites. In *Transportation Research Record*: *Journal of the Transportation Research Board, No. 89.* Transportation Research Board of the National Academies, Washington, DC, 2009, pp. 10-1870.

Stein, B.R., Zheng, B., Kokkinidis, I., Kayastha, N., Seigler, T., Gokkaya, K., Gopalakrishnan, R., and Hwang, W.H. An Efficient Remote Sensing Solution to Update the NCWI. *Photogrammetric Engineering and Remote Sensing*, June, 2012, pp. 537-547.

Sugumaran, R., Harken, J., and Gerjevic, J. *Using Remote Sensing Data to Study Wetland Dynamics in Iowa.* University of Northern Iowa, Cedar Falls, 2004.

Tana, G., Letu, H., Cheng, Z., and Tateishi, R. Wetlands Mapping in North America by Decision Rule Classification Using MODIS and Ancillary Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 6, December, 2013, pp. 2391-2401.

U.S. Army Corps of Engineers. *Corps of Engineers Wetlands Delineation Manual*. Environmental Laboratory, Technical Report Y-87-1, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, 1987.

U.S. Department of Agriculture. Web Soil Survey. 2016. http://websoilsurvey.sc.egov.usda.gov. Accessed June 23, 2016.

U.S. Fish and Wildlife Service. National Wetlands Inventory. 2016. https://www.fws.gov/wetlands. Accessed June 23, 2016.

U.S. Geologic Survey. U.S. Geologic Survey: About NED. 2016a. http://ned.usgs.gov/about.html. Accessed June 23, 2016.

U.S. Geologic Survey. ESPA Ordering Interface. 2016b. https://espa.cr.usgs.gov. Accessed June 23, 2016.

U.S. Geologic Survey. Watershed Boundary Dataset. 2016c. http://nhd.usgs.gov/wbd.html. Accessed June 23, 2016.

Yan, G., Mas, J.-F., Maathuis, B.H.P., Xiangmin, Z., and Van Dijk, P.M. Comparison of Pixel-Based and Object-Oriented Image Classification Approaches—A Case Study in a Coal Fire Area, Wuda, Inner Mongolia, China. *International Journal of Remote Sensing*, Vol. 27, September, 2006, pp. 4039-4055.