

Using LiDAR Topographic Data and Machine Learning Techniques to Identify Near-Surface Soil Saturation for Improved Environmental Planning-Scale Wetland Mapping

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Gina Lisette O'Neil

August 2019

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: *Jina L. O'Neil*

This Dissertation has been read and approved by the examining committee:

Advisor: Jonathan L. Goodall

Committee Member: Lawrence E. Band

Committee Member: Madhur Behl

Committee Member: Teresa B. Culver

Committee Member: Matthew A. Reidenbach

Committee Member: _____

Accepted for the School of Engineering and Applied Science:



Craig H. Benson, School of Engineering and Applied Science

August 2019

Abstract

Wetlands are important ecosystems that are threatened by agricultural and development repurposing, pollutant runoff, and climate change. Accurate and widely-available wetland inventories have the potential to support wetland conservation and environmental planning entities. The growing collection of remote sensing data has created new opportunities for wetland identification, and Light Detection and Ranging (LiDAR) data, specifically, has been widely embraced by the wetland science community. This dissertation aims to develop an open source wetland identification tool that leverages LiDAR elevation data and machine learning techniques to identify likely wetland areas at an environmental planning scale. The designed wetland identification tool is implemented and evaluated across four study areas in Virginia that encompass a range of ecoregion, built environment, and topographic characteristics. Key components of the wetland tool are developed and refined through three studies. The first study focuses on the identification and evaluation of LiDAR topographic metrics as indicators of near-surface soil moisture, using a Random Forest model. The second study focuses on the effects of alternative hydrologic terrain processing methods on wetland predictions and the Random Forest model used to generate them. The third study evaluates the potential for using deep learning for identification of wetlands from images that represent LiDAR-derived geomorphic characteristics, with relatively limited training data resources. Key research findings are as follows. i) The topographic wetness index, curvature, and cartographic depth-to-water index are successful wetland indicators for a range of landscapes, but there is potential to improve their abilities to distinguish wetted areas from dry uplands through site-specific modifications. ii) By applying a sophisticated LiDAR DEM preprocessing workflow, these topographic indices are better able to model wetlands and, therefore, considerably improve prediction accuracy for all sites studied. iii) Accurate wetland predictions can be produced using a basic deep learning architecture and imagery composed of the topographic indices and complementary vegetative information, despite being limited to training data sets that are significantly smaller relative to most deep learning applications. With the completion of model refinement through this research, the wetland tool has the potential to benefit the broader environmental planning and conservation community.

This dissertation is dedicated to my Abuelita, Tarcila Sotomayor. Thank you for being my greatest example of hard work, intelligence, and sincerity. Also, for the many snacks throughout my childhood. You are dearly missed.

Acknowledgements

This dissertation could not have been completed without the support and encouragement of many people.

Thank you to the great friends I've made in Charlottesville, the Link Lab, and the Hydroinformatics research group. To my research group members, I'm grateful to have shared this time with you and that I could be a part of such a friendly, supportive, and intelligent group of people.

Thank you to my undergraduate and MS advisor, Dr. Richard Miksad. Without witnessing, and eventually sharing, your enthusiasm for the Saqsaywaman project, I would not have considered pursuing graduate research. The positive experience I had during this project and your confidence in me were crucial in my decision to pursue my PhD.

Thank you to my research mentor and PhD advisor, Dr. Jonathan Goodall. Thank you for allowing me to be a part of the Hydroinformatics group. I feel lucky to have been guided through this process by an exemplary leader, standout researcher, and thoughtful teacher. Your encouragement and support over these few years made this dissertation possible.

Thank you to my parents, for trusting and supporting me through my academic career, beginning with my choice to attend UVa as an undergrad to unexpectedly extending my time here. I hope you're proud of how I've spent that time. And thank you to my Aunt Kathy and Uncle Tom. You've provided thoughtful advice and support during every big decision. I also thank my brother, Afa, for helping me keep work in perspective of bigger things in life, probably since kindergarten. Also, thank you to my sisters, Sasha and Kaitlyn. You're both too young to know what a dissertation is, but you would probably support me if you did and I appreciate that.

Finally, I thank Dan. I am immensely grateful to have had you by my side during this entire process. This dissertation would not be possible without your encouragement and reassurance during the countless self-doubting moments. By far, the greatest highlight of my time at UVa will have been meeting my future husband.

Table of Contents

1	Introduction.....	1
2	Evaluating the Potential for Site-Specific Modification of LiDAR DEM Derivatives to Improve Environmental Planning-Scale Wetland Identification using Random Forest Classification	4
2.1	Introduction.....	4
2.2	Study Areas.....	7
2.3	Input Data.....	9
2.3.1	LiDAR Elevation Data.....	9
2.3.2	Land Cover Data.....	9
2.3.3	National-Scale Datasets	10
2.3.4	VDOT Wetland Delineations.....	10
2.4	Methods.....	11
2.4.1	Preprocessing	11
2.4.2	Supervised Classification.....	19
2.4.3	Post Processing	19
2.5	Results and Discussion	20
2.5.1	Highest Performing Models	20
2.5.2	Response of Model to Input Data Modification.....	28
2.6	Conclusions.....	30
3	Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model.....	33
3.1	Introduction.....	33
3.2	Study Areas and Input Data	35
3.2.1	Study Areas.....	35
3.2.2	Input Data.....	37
3.3	Methods.....	37
3.3.1	Overview of the Wetland Identification Model	38
3.4	Results.....	45
3.4.1	Effects of Preprocessing Techniques on Model Performance	45
3.4.2	Characteristics of the Tuned RF Model	47
3.5	Discussion.....	48
3.5.1	Varying the Smoothing Scale and Method by Input Variable	48
3.5.2	Improvements to Wetland Predictions Due to Preprocessing Schemes.....	51
3.5.3	Comparison to Earlier Wetland Model Implementations	53

3.5.4	Approach Limitations	53
3.6	Conclusions.....	54
4	Deep Learning as a Wetland Classification Approach using LiDAR Topographic-Derived and NDVI Input Data.....	56
4.1	Introduction.....	56
4.2	Methodology	58
4.2.1	Study Areas.....	58
4.2.2	Input Data.....	60
4.2.3	Wetland Identification Method	61
4.2.4	Experimental Setup.....	66
4.3	Results.....	68
4.3.1	Performance of Site-Specific Models	68
4.3.2	Using Site-Specific Models to Predict Wetlands in Other Sites	69
4.3.3	Performance of Combined-Site Models.....	70
4.4	Discussion	71
4.4.1	Potential for Site-Specific Models	71
4.4.2	Potential for Combined-Site Models.....	73
4.4.3	Utility of the Proposed Input Data Configuration.....	76
4.4.4	Comparison of Deep Learning to a Random Forest Implementation	76
4.4.5	Future Work	77
4.5	Conclusions.....	78
5	Conclusions.....	80
6	References.....	82
7	Appendix.....	94

List of Figures

Figure 1. Overview of the wetland identification workflow, developed and refined through three targeted studies. ...	3
Figure 2. (A) Study site locations, outlined by watershed(s) used as site processing extent, spanning five of the seven ecoregions of Virginia, and (B) areas of each VDOT delineation site with orthoimagery corresponding to the time frame in which VDOT delineations were performed.....	8
Figure 3. Workflow followed to implement the wetland identification approach as an ArcGIS model consisting of preprocessing, supervised classification, and post processing phases to create model predictions and confusion matrices used for accuracy assessment.	11
Figure 4. Example calculation of the adjusted slope grid (solid line) for Site 1 where the β was set to a value of 2 and γ was calculated to be 11.42, corresponding to a representative slope value taken to be the 95th percentile of all underlying wetland slopes. These adjustments decrease slopes that are originally below 0.088 and increase slopes that are originally above 0.088, relative to a slope grid (dashed line) where γ and β are both equal to 1.....	15
Figure 5. Topographic input variables in Site 1, original TWI (A1), curvature (B1), and DTW (C1), compared to modified versions each variable, shown in A2, B2, and C2, respectively.	16
Figure 6. Example of the process, shown for Site 1, used to randomly separate VDOT delineations into training and testing datasets, consisting of four steps: (A) point creation, (B) point buffering, (C) value extraction, and (D) training data separation.....	18
Figure 7. Examples of NWI maps (A1 and B1) and model predictions (A2 and B2) for Site 1, both compared to VDOT delineations.....	22
Figure 8. Examples of NWI maps (A1 and B1) and model predictions (A2 and B2) for Site 2, both compared to VDOT delineations.....	23
Figure 9. Examples of NWI maps (A1 and B1) and model predictions (A2 and B2) for Site 3, both compared to VDOT delineations.....	24
Figure 10. Examples of NWI maps (A1, B1, and C1) and model predictions (A2, B2, and C2) for Site 4, both compared to VDOT delineations.	25
Figure 11. Wetland, non-wetland, and overall accuracy produced by the best performing model predictions, compared to accuracy produced by NWI maps.	28
Figure 12. Four study areas spanning four level III ecoregions in Virginia, USA (a). Each study area includes the wetland survey limits, referred to as study sites, and the encompassing HUC 12 watershed, used as the processing extent (b).....	36
Figure 13. Workflow of the wetland identification model created through this research. Each combination of preprocessing techniques (bold font) was executed for this analysis. Green shapes indicate input data, grey shapes indicate processes, yellow shapes indicate intermediate output, and red shapes indicate final output.	38
Figure 14. Wetland precision and recall resulting from each preprocessing technique combination across all study sites. Note the differences in x-scale and y-scale range.....	46
Figure 15. Effect of varying smoothing method and scale on wetland model accuracy.....	49
Figure 16. Wetland likelihoods resulting from different preprocessing configurations: worst generalized preprocessing, as described in Section 4.1 (a), best performing generalized preprocessing, as described in section 4.1 (b), and A* conditioning and best performing individualized smoothing, as described in section 5.1 (c).	52
Figure 17. Four study areas spanning four level III ecoregions in Virginia, USA (a). Each study area includes the wetland survey limits, referred to as study sites, and the encompassing HUC 12 watershed, used as the processing	

extent (b). Reprinted from “Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model” by O’Neil et al., 2019, Water Resources Research, 55 (5), 4343-4363.....59

Figure 18. Overview of the proposed wetland identification method. Green shapes indicate input data, grey shapes indicate processes, yellow shapes indicate intermediate output, and red shapes indicate final model output.61

Figure 19. Methodology followed for the four experiments designed to address the study research questions.68

Figure 20. Wetland mapping accuracy resulting from Experiment 1, where site-specific models were created using several training data sizes depending on site availability.69

Figure 21. Wetland mapping accuracy resulting from Experiment 2, where the best performing site-specific models were used to predict wetlands in other sites.....70

Figure 22. Wetland mapping accuracy resulting from Experiment 3, which used training data from all sites to create a general model, and Experiment 4 which used training data only from sites within the same ecoregion (sites 2 and 3) to create an ecoregion model.....71

Figure 23. Comparison of wetland predictions produced by site-specific models created from (column A) the smallest training dataset and (column B) the largest training dataset available for the site. Also shown are wetland predictions produced by models trained only with the largest training dataset for Site 4 (column C).73

Figure 24. Comparison of wetland predictions resulting by (column A) the best performing site-specific models (i.e., those trained on 70% of the validation area), (column B) the general model, and (column C) the ecoregion model..75

List of Tables

Table 1. Conditions of the processing extent and VDOT delineation area for each study site; upper portion describes conditions of the processing extent and lower portion describes conditions of the VDOT delineation area.	9
Table 2. Modification parameters for topographic indices, and soil thematic maps determined to be relevant for each study site.	12
Table 3. Statistics describing the training and testing data for each site.	19
Table 4. Input data that produced the highest performing wetland identification model in each site, in terms of overall accuracy, as well as variable importance and rank of each input variable according to the ESRI Classifier Definition file.	21
Table 5. Confusion matrices used to assess the accuracy of NWI maps (left) and best performing model predictions (right) compared to the test data, where predicted values are represented along rows and actual values are represented along columns.	27
Table 6. Wetland, non-wetland, and overall accuracy achieved by iterations of RF classification for each site.	29
Table 7. Characteristics of each study site, including dominate land cover, topographic characteristics, and surveyed wetland distributions.	37
Table 8. Characteristics of each study site, including dominate land cover, topographic characteristics, and surveyed wetland distributions. Reprinted from “Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model” by O’Neil et al., 2019, Water Resources Research, 55 (5), 4343-4363.	60
Table 9. Minimum and maximum values used to scale each input feature to a range of 0 to 1. Minimum and maximum values were assumed for the TWI, curvature, and DTW from statistical analyses.	64
Table 10. Parameters for the DeepNets implementation used in all performed experiments.	65
Table 11. Maximum number of training images available per site when randomly sampling 70% of the eligible validated area. Each labeled image used for training has a resolution of 320 x 320 pixels.	67

Chapter 1

Introduction

Wetlands are important ecosystems that are threatened by agricultural and development repurposing, pollutant runoff, and climate change (Klemas, 2011). It is estimated that approximately half of the Earth's wetlands have been destroyed since 1900 (Davidson, 2014) as well as half of wetlands in the conterminous U.S. since 1600 (Dahl et al., 1991). In the U.S., federal regulations play an important role in the protection of remaining wetlands. Specifically, Section 404 of the Clean Water Act requires environmental impact assessments prior to land development and water resources projects (Page & Wilcher, 1990). Compliance with this law entails environmental planning entities to provide detailed wetland delineations to the U.S. Army Corps of Engineers (USACE). Manual surveying by trained analysts will always be the most accurate method to delineate wetlands, however carrying out detailed field surveys can be time-consuming and costly. There is potential for computational models to streamline the delineation process by providing accurate wetland inventories that limit manual surveying to areas likely to contain wetlands.

A widely available wetland inventory with the reliability necessary to assist in the wetland permitting process is an unmet need. The existing national-scale wetland inventory in the U.S., the National Wetland Inventory (NWI), is insufficient for assisting in the permitting process. Despite being one of the most commonly used sources of wetland data in the U.S., NWI maps were never intended to map federally regulated wetlands (Cowardin & Golet, 1995; US Corps of Engineers, 1987). Research has shown that relying solely on the NWI may fail to protect a considerable fraction of wetlands (Morrissey & Sweeney, 2006). Moreover, development of wetland inventories for this purpose can be expensive and technically challenging due to the complexity of wetland features (Kloiber et al., 2015).

Remote sensing data offer an accurate and cost-effective way to observe common wetland characteristics at varying scales (Guo et al., 2017; Lang et al., 2013; Lang & McCarty, 2014). Multispectral imagery, radar, and Light Detection and Ranging (LiDAR) have all been shown to be useful for wetland identification, and most researchers incorporate multispectral imagery (Guo et al., 2017). However, incorporating multispectral imagery data can present barriers to wetland permitting applications as national-scale products are relatively coarse (~30 m) and higher resolution imagery are costly to collect. Furthermore, preliminary work reinforced that freely-available Landsat 8 multispectral imagery data are too coarse for these purposes (Felton et al., 2019). Given the utility of spectral imagery to describe wetland features, however, these data would be valuable additions to wetland models if and when high-resolution versions become widely available. Alternatively, Light Detection and Ranging (LiDAR) emerges as a superior candidate within the wetland science community for its high resolution (< 2 m) and availability (Kloiber et al., 2015; Lang & McCarty, 2014). LiDAR sensors provide detailed information on the

Earth's surface and data availability has increased rapidly over the past 20 years with ongoing federal efforts to hasten its collection throughout the entire U.S. (Snyder & Lang, 2012). LiDAR returns can be interpolated to create high-resolution digital elevation models (DEMs), which allow for detailed mapping of topographic metrics that describe flow convergence (Lang & McCarty, 2014). Research has shown that LiDAR-derived topographic metrics are successful wetland indicators that are able to model spatial patterns of near-surface soil moisture (Lang et al., 2013; Lang & McCarty, 2014; Millard & Richardson, 2015; O'Neil et al., 2018). While conventional DEMs and their derivatives have been shown to be useful for wetland delineation as well (Hogg & Todd, 2007), LiDAR DEMs allow for more detailed mapping of topographic metrics (Lang & McCarty, 2014). In addition, using LiDAR topography as inputs to machine learning algorithms has the potential to produce sophisticated models capable of identifying the complex characteristics of wetlands quickly and efficiently.

Previously, researchers have demonstrated the ability to delineate wetlands by coupling remote sensing data and machine learning techniques. However, researchers have yet to propose and thoroughly evaluate workflows that i) fully leverage the existing freely-available data resources, ii) consider alternative preprocessing methods that make the best use of the input data, and iii) evaluate both traditional and emerging machine learning algorithms. This dissertation research aims to develop an open source wetland identification tool that utilizes LiDAR data and machine learning techniques to identify likely wetland areas at an environmental planning-scale. The designed wetland model is implemented and evaluated across four study areas in Virginia that encompass a range of ecoregion, built environment, and topographic characteristics. Key components of the wetland tool are studied and refined through three studies (Figure 1).

The first study evaluates the potential for LiDAR DEM derivatives, combined with ancillary national-scale soil data, to identify wetland areas across four study sites in Virginia. The approach was implemented as a model in ArcGIS and performed a Random Forest (RF) classification of input variables that were modified to provide distinct wetland and non-wetland signatures. Model predictions were assessed against field-mapped testing data, provided by the Virginia department of transportation (VDOT), and compared to commonly-used NWI maps.

The second study reexamines the hydrologic terrain processing workflow to optimize the performance of established LiDAR wetland indicators. The literature shows that the replacement of conventional DEMs with LiDAR DEMs requires changes to the traditional hydrologic terrain processing workflow: smoothing and conditioning. Choice of smoothing and conditioning techniques can significantly impact accuracy of hydrologic parameters. However, a thorough analysis of the compound effects of smoothing and conditioning on wetland delineations and the RF model used to generate them has yet to be performed. This study tests the response of the wetland model to proposed preprocessing techniques as well as reconfigure the wetland model as a Python-based, open source workflow.

The third study explores an alternative machine learning approach for the wetland identification model. Deep learning for wetland identification is an emerging field that shows promise for advancing these efforts. Deep learning is unique to traditional machine learning

techniques for its ability to consider the spatial context of object characteristics within a landscape scene. However, remote sensing researchers are deterred from applying deep learning by the typical need for very large datasets in creating these models. Using four study sites across Virginia, we contribute to this research area by providing insight into the potential for deep learning of wetlands from limited, but typical, training data. Our proposed workflow performs a wetland semantic segmentation using DeepNets (Audebert et al., 2018 [ISPRS Journal of Photogrammetry and Remote Sensing]), a deep learning architecture for remote sensing data, and an input dataset consisting of high-resolution topographic indices and the Normalized Difference Vegetation Index. This work offers new insights into the potential for deep learning of environmental planning-scale wetlands given data resources that are typical for wetland remote sensing researchers.

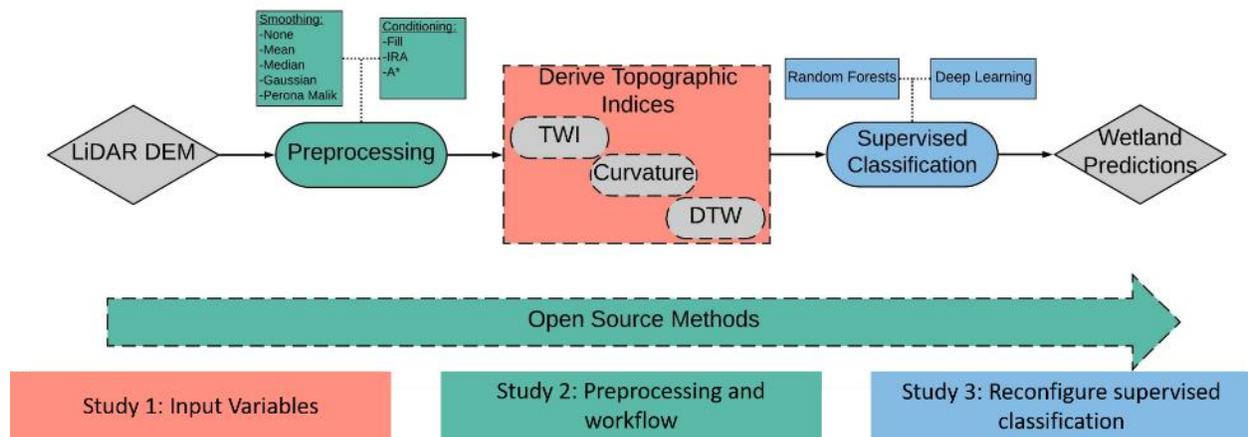


Figure 1. Overview of the wetland identification workflow, developed and refined through three targeted studies.

Chapter 2

Evaluating the Potential for Site-Specific Modification of LiDAR DEM Derivatives to Improve Environmental Planning-Scale Wetland Identification using Random Forest Classification

2.1 Introduction

Wetlands are important ecosystems that not only provide habitat for many plant and animal species, but also improve water quality, recharge groundwater, and ease flood and drought severity (Guo et al., 2017). Despite the ecological value of wetlands, their quality and presence are threatened by agricultural or development repurposing, pollutant runoff, and climate change (Klemas, 2011). Current estimates are that roughly 50% of wetlands have been lost globally since 1900 (Davidson, 2014) and approximately 53% of wetlands of the conterminous U.S. have been lost since the early 1600s (Dahl et al., 1991). The historic loss of wetlands and sustained threat to remaining wetlands has motivated increased efforts by scientists and government to protect and maintain these ecosystems.

U.S. federal regulations play an important role in the abatement of further wetland loss. One of the most important policies in support of this effort is Section 404 of the Clean Water Act, which protects the nation's waters, including wetlands. According to Page and Wilcher (1990), this law states that environmental planning entities must identify and assess environmental impact due to land development and water resource projects. This requires environmental planning entities, such as state departments of transportation (DOTs), to provide wetland delineations that are ultimately jurisdictionally confirmed by the U.S. Army Corps of Engineers (USACE). The USACE Wetlands Delineation Manual states that wetlands can be identified by environmental characteristics shared among the many wetland types. The USACE guidelines for wetland delineations use these common features and are based on the presence of hydrologic conditions that inundate the area, vegetation adapted for life in saturated soil conditions, and hydric soils (US Corps of Engineers, 1987).

Manual surveying by trained analysts will always be the most accurate method to delineate wetlands, however carrying out detailed field surveys can be time consuming and costly. According to estimates provided by representatives from the Virginia DOT (VDOT) Environmental Division, the costs of these delineations range from \$60 to \$140 per acre (~0.4 ha)

(personal communication, November 28, 2017). These estimates are based on recent VDOT projects and can vary widely across projects. To offset these costs, the wetland permitting process could potentially be streamlined by supplementing and guiding the manual delineations with accurate digital wetland inventories. However, developing and updating wetland inventories can be expensive and technically challenging due to the complexity of wetland features (Kloiber et al., 2015). Furthermore, the existing national-scale wetland inventory in the U.S., the National Wetland Inventory (NWI), is not ideal for assisting in the permitting process. Despite being one of the most commonly used sources of wetland data in the U.S., NWI maps were never intended to map federally regulated wetlands (Cowardin & Golet, 1995; US Corps of Engineers, 1987) and research has shown that relying solely on the NWI may fail to protect a considerable fraction of wetlands (Morrissey & Sweeney, 2006). Thus, a wetland inventory with the reliability necessary to assist in the wetland permitting process is an unmet need.

Remote sensing has long been recognized as a powerful tool for identifying wetlands (US Corps of Engineers, 1987) and may offer an accurate and cost-effective way to fulfill this need (Guo et al., 2017; Lang et al., 2013; Lang & McCarty, 2014). Past studies have incorporated remote sensing data such as multispectral imagery, radar, and Light Detection and Ranging (LiDAR) for wetland identification. A review of wetland remote sensing studies of the past 50 years shows that most researchers incorporate multispectral imagery in wetland classifications (Guo et al., 2017). However, the incorporation of multispectral imagery can weaken the potential for use during the wetland permitting process by introducing issues of resolution or accessibility. For example, the commonly used Landsat multispectral imagery is freely available on a national scale, but the 30 m resolution of this data can be too coarse to detect wetlands at a scale relevant to environmental planning entities, which can require a spatial accuracy of at least 1.5 m (VDOT Environmental Division, personal communication, November 28, 2017). While studies have shown higher resolution, multispectral data can result in accurate wetland classifications (e.g., Kloiber et al., 2015) these data can be inaccessible due to cost. Alternatively, LiDAR is remote sensing data that has been rapidly endorsed by the wetland science and management community for its growing availability and technological benefit to wetland mapping (Kloiber et al., 2015; Lang & McCarty, 2014). LiDAR sensors provide detailed information on the Earth's landscape and bare surface by collecting x, y, and z data that can then be interpolated to create digital elevation models (DEMs) (Lang & McCarty, 2014). LiDAR data availability has increased rapidly over the past 20 years, and although current coverage in the conterminous U.S. is at about one third, there is an ongoing effort by multiple federal agencies to hasten the collection of LiDAR data throughout the entire U.S. (Snyder & Lang, 2012). LiDAR derived DEMs have the ability to map wetlands by identifying areas of inundation based on topographic drivers of flow convergence and offer widely available, high-resolution data that could be utilized during the wetland permitting process. While conventional DEMs and their derivatives have been shown to be useful for wetland delineation (e.g., Hogg & Todd, 2007), LiDAR DEMs allow for more detailed mapping of topographic metrics (Lang & McCarty, 2014).

Previous research has shown that DEM derivatives have the potential to model spatial patterns of saturated areas, and that LiDAR DEM derivatives improve the ability of these metrics to do so (e.g., Hogg & Todd, 2007; Lang et al., 2013; Millard & Richardson, 2013). Among the DEM derivatives found to be useful for this purpose are curvature, Topographic Wetness Index (TWI) and the Cartographic Depth-to-Water index (DTW) (e.g., Ågren et al., 2014; Lang et al., 2013; Murphy et al., 2009, 2011; Sangireddy et al., 2016). Curvature is defined as the second derivative of the input surface and can describe the degree of convergence and acceleration of flow (Moore et al., 1991). The TWI, developed by Beven and Kirkby (1979), relates the tendency of a site to receive water to the tendency of a site to evacuate water and is defined as

$$TWI = \ln\left(\frac{\alpha}{\tan(\beta)}\right), \quad (1)$$

where α is the specific catchment area, or contributing area per unit contour length, and $\tan(\beta)$ is the local slope. The DTW is a soil moisture index developed by Murphy et al. (2007) that is based on an assumption that soils very close in elevation to their assigned surface water are more likely to be saturated. The DTW model in grid form is calculated as

$$DTW (m) = \left[\sum \left(\frac{dz_i}{dx_i} \right) a \right] * x_c, \quad (2)$$

where $\frac{dz}{dx}$ is the downward slope of a pixel, i is a pixel along a calculated least cost (i.e., slope) path to the assigned source pixel, a is 1 when the flow path is parallel to pixel boundaries or $\sqrt{2}$ when the flow crosses diagonally, and x_c is the pixel length (Murphy et al., 2007).

Although many studies have shown the benefit of using topographic indices to identify wetted areas, and the added benefit of deriving these indices at higher resolutions, there are unique challenges inherent to using LiDAR DEMs. Researchers have noted that LiDAR DEMs used for purposes related to modelling landform characteristics must be resampled to coarser resolutions and smoothed to overcome issues of increased “noise” from excessive topographic detail (MacMillan et al., 2003), with this topographic noise arising from DEMs on the order of 1 m pixel size (Richardson et al., 2009). Moreover, variations in DEM resolution result in significantly different spatial and statistical distributions of contributing areas and downslope flow path lengths (Woodrow et al., 2016), and at high resolutions, micro-topographic features can lead to highly variable slope values and provide unrealistic estimates of hydraulic gradients (Grabs et al., 2009; Lanni et al., 2011). Previous studies have acknowledged the negative effect that these micro-topographic features have on the ability of curvature (e.g., Sangireddy et al., 2016) and TWI (e.g., Sørensen & Seibert, 2007) to identify hydrologic features of interest. For example, Ågren et al. (2014) found that high-resolution DEMs (< 2 m) caused local TWI variations that are too strong to separate wetlands from uplands, whereas deriving the index from coarser (> 24 m) DEMs reduced these variations but resulted in poorly delineated flow channels and local depressions. In contrast, the researchers also concluded that DTW derivations are not sensitive to scale, but have suggested that the DTW could be further optimized (Ågren et al., 2014).

LiDAR DEM data and other remote sensing data are commonly used to map wetlands through supervised classification algorithms. Random Forest (RF) classification is a relatively new supervised classification method that is widely used for its ability to handle both continuous and

categorical, high-dimensional data and produce descriptive variable importance measures (Millard & Richardson, 2015; Rodriguez-Galiano et al., 2012). RF has been shown to produce higher accuracies than other classification techniques, such as maximum likelihood, when incorporating multisource data (Duro et al., 2012; Miao et al., 2012; Rodriguez-Galiano et al., 2012). Furthermore, studies have shown that LiDAR DEM metrics are suitable input variables for the RF approach (e.g., Deng et al., 2017; Kloiber et al., 2015; Zhu & Pierskalla, 2016), and that using this classifier has strong potential to improve mapping and imagery classification of wetlands (e.g., Millard & Richardson, 2013).

Many previous studies have relied primarily on ecological factors and spectral indices provided by multispectral imagery to classify wetlands, and fewer studies have evaluated the predictive power of LiDAR DEM data alone for this purpose. The primary objective of this study was to further advance the application of LiDAR DEM derivatives to wetland mapping by evaluating the potential of modified TWI, DTW, and curvature grids to address limitations noted by researchers and identify small (i.e., environmental planning-scale) wetlands across varying ecoregions. RF classifications of original and modified TWI, curvature, and DTW, where the TWI and curvature were modified via smoothing and the DTW was modified via adjustments to the input slope grid, along with ancillary national-scale soil data were assessed against field-mapped test data and compared to NWI maps to identify the best performing models. Accuracy assessments of these classifications provided a measure of the benefits and costs of modifying these input data. This approach was applied to four study sites across varying ecoregions of Virginia and implemented in ArcGIS with the potential for further refinement and utility by environmental planning entities.

2.2 Study Areas

The four sites in this study were selected due to availability of VDOT wetland delineations and LiDAR DEMs, and to have applications of this approach across varying ecoregions of Virginia. As seen in Figure 2, the study sites span five of the seven level III EPA ecoregions of Virginia: the Piedmont (45), the Mid-Atlantic Coastal Plain (63), the Northern Piedmont (64), the Southeastern Plains (65), and the Ridge and Valley (67). According to the EPA (Wilken et al., 2011), the Piedmont ecoregion is considered the non-mountainous region of the Appalachians Highland and is comprised of transitional areas between the mountainous Appalachians to the northwest and the relatively flat coastal plain to the southeast. The soils in this region tend to be finer textured than in ecoregions 63 and 65. The Mid-Atlantic Coastal Plain is characterized by low, nearly flat plains with many swamps, marshes, and estuaries. The region has a mix of coarse and finer textured soils and poorly drained soils are common here. The Northern Piedmont consists of low rounded hills, irregular plains, and open valleys. It is a transitional region between the low mountains in ecoregion 66 and the flat coastal area of ecoregions 63 and 65. The Southeastern Plains are irregular and have a mosaic of cropland, pasture, woodland, and forest. The subsurface is predominantly sands, silts, and clays. The Ridge and Valley ecoregion is relatively low-lying

and characterized by alternating forested ridges and agricultural valleys. Additional information describing the conditions of each study site can be found in Table 1.

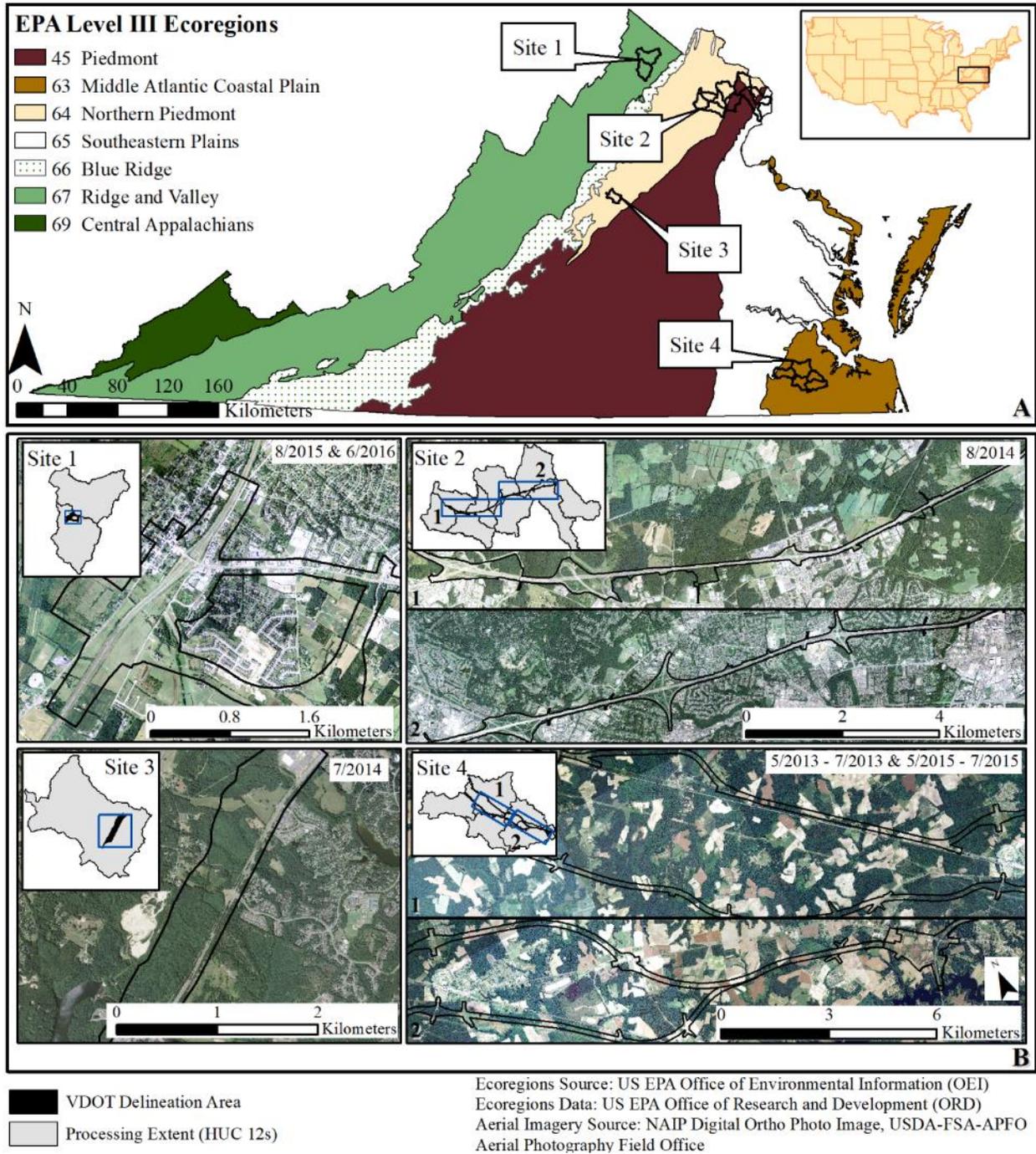


Figure 2. (A) Study site locations, outlined by watershed(s) used as site processing extent, spanning five of the seven ecoregions of Virginia, and (B) areas of each VDOT delineation site with orthoimagery corresponding to the time frame in which VDOT delineations were performed.

Table 1. Conditions of the processing extent and VDOT delineation area for each study site; upper portion describes conditions of the processing extent and lower portion describes conditions of the VDOT delineation area.

	Site 1	Site 2	Site 3	Site 4
Processing Extent (HUC 12s) (km ²)	273	1208	65	547
LiDAR DEM Resolution (m)	1.00	1.50	0.76	0.76
HUC12 Max. Elevation (m)	458	417	223	37
HUC12 Min. Elevation (m)	140	0	96	0
HUC12 Mean Slope (%)	9.5	7.0	12.6	3.7
VDOT Delineation Total Area (km ²)	2.98	7.87	1.82	12.17
VDOT Delineation Max. Elevation (m)	241	147	178	34
VDOT Delineation Min. Elevation (m)	210	47	101	3
VDOT Delineation Mean Slope (%)	7.2	9.4	14.7	3.2
VDOT Wetland to Non-Wetland Ratio	0.02	0.02	0.02	0.42

2.3 Input Data

Freely available LiDAR elevation data, land cover data, national-scale hydrography data, national-scale soil data, and VDOT wetland delineations were used as inputs to the wetland identification model.

2.3.1 LiDAR Elevation Data

LiDAR-derived elevation data used in this study were provided by the Virginia Information Technologies Agency (VITA) in raster format (VITA, 2016). VITA LiDAR data products were freely available and included hydro-flattened, bare-earth DEMs. The LiDAR DEMs used in this study were collected and processed between 2010 and 2015 and have horizontal resolutions ranging from 0.76 m to 1.5 m. Tiles with different resolutions were merged and resampled to the coarsest resolution using the bilinear resampling method in ArcGIS, following the approach previously done by Ågren et al. (2014). Site 2 was unique in that LiDAR data were unavailable for approximately 230 km² (23%) of the processing extent and 0.8 km² (12%) of the VDOT delineation area. To fill the missing areas, 3 m elevation data from the National Elevation Dataset were used (USGS, 2019) and resampled to 1.5 m to match the dominating LiDAR data. While resampling to finer resolutions is not ideal, maintaining consistency in the application of highest resolution LiDAR data across all study sites was prioritized over the error introduced in the relatively small portion of the processing extent, and even smaller portion of the delineation area.

2.3.2 Land Cover Data

Land cover data were used for post classification filtering. Land cover data used in this study were provided by VITA in raster format (VITA, 2016). VITA land cover data were derived from the Virginia Base Mapping Program 4 band orthophotography, collected between 2011 and 2014. These data provided 12 land cover classifications with 85-95% accuracy and have a horizontal resolution of 1 m (WorldView Solutions Inc., 2016).

2.3.3 National-Scale Datasets

National-scale soil and hydrography data were incorporated in the classification as ancillary data. Soil data used in this study were obtained from the Soil Survey Geographic database (SSURGO) and distributed by the Natural Resources Conservation Service's Web Soil Survey in polygon vector format (Soil Survey Staff, 2017b). The SSURGO hydric rating, depth to water table, hydrologic soil group, surface texture, and soil drainage class were used as indicators of saturated conditions. According to the Soil Survey Staff (2017a), the hydric rating attribute indicates the percentage of a map unit that meets the criteria for hydric soils. Hydric soils are characteristic of wetlands and are defined as soil that is formed under conditions of saturation, flooding, or ponding long enough during the growing season to develop anaerobic conditions in the upper horizon (Soil Survey Staff, 2017a). The surface texture attribute describes the representative texture class according to percentage of sand, silt, and clay in the fraction of the soil that is less than 2 mm in diameter. The drainage class attribute identifies the natural drainage conditions of the soil and refers to the frequency of wet periods without considering alterations of the water regime by human activities, unless they have significantly changed the morphology of the soil. The hydrologic soils group assignment is based on estimates of the rate of water infiltration when the soils are not protected by vegetation, are thoroughly wet, and receive precipitation from long-duration storms. The depth to water table attribute indicates the representative depth to the saturated zone in the soil.

Hydrography data used in this study were provided by the National Hydrography Dataset (NHD) in polygon vector format (USGS, 2019b). NHD HUC 12 watersheds intersected by the limits of VDOT delineations were combined to be used as the processing extent for each study site in order to encompass the hydrologically connected area around VDOT delineations. NHD streams and waterbodies within these processing extents were also used.

2.3.4 VDOT Wetland Delineations

Wetland delineations for each site were provided by VDOT and were used to create training and testing datasets. The VDOT delineations in Site 2, Site 3, and Site 4 were jurisdictionally confirmed by the USACE, and all study sites were produced through field surveys conducted by professional wetland scientists. For these reasons, the VDOT delineations were considered to be ground truth for the purpose of training and testing the wetland identification model. VDOT delineations were provided in polygon vector format and included both wetlands and streambeds. Both were included in subsequent processing because both are considered waters of the state and therefore must be delineated during the wetland permitting process. Although the delineations were categorized by wetland type by VDOT analysts, all areas were merged into a single "wetland" category before application in this study. Additionally, limits of delineations were used to identify true non-wetland areas.

2.4 Methods

The workflow followed to implement the wetland identification approach consisted of three main parts: preprocessing, supervised classification, and post processing (Figure 3). The workflow was implemented in ArcGIS 10.4 and the ModelBuilder tool was used to automate processes that did not require user intervention. Outputs of the workflow were model predictions and confusion matrices used to assess the accuracy of those predictions. Components of the workflow are described in more detail in the following sections.

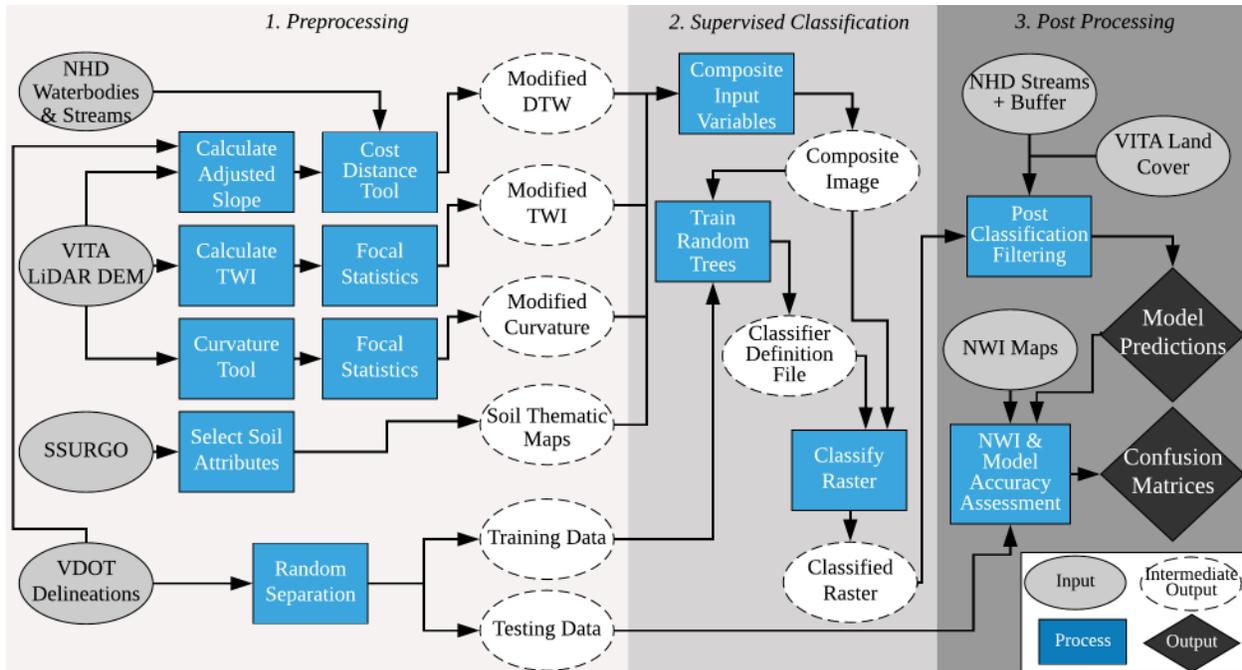


Figure 3. Workflow followed to implement the wetland identification approach as an ArcGIS model consisting of preprocessing, supervised classification, and post processing phases to create model predictions and confusion matrices used for accuracy assessment.

2.4.1 Preprocessing

The preprocessing phase consisted of a combination of automated and semi-automated processes that required user intervention. Preprocessing steps not explicitly shown in Figure 3 include projection of input data to the appropriate North or South Virginia State Plane coordinate system, clipping data to the HUC 12 processing extent, rasterizing input data originally in polygon vector format by using the site LiDAR data as the pixel size constraint, and filling sinks within the LiDAR DEM. Rasterizing the polygon vector layers mapped at larger scales assumes that the information provided at the original scale (ranging from 1:24,000 to 1:12,000) is true for each pixel of the output grid (ranging from 0.76 to 1.52 m). The LiDAR DEM was filled using the depression filling algorithm of Planchon and Darboux (2002) that is implemented in ArcGIS. Intermediate outputs created by the preprocessing phase were calibrated input variables, training data, and testing data.

2.4.1.1 Modified Input Variable Creation

Input variables included the modified TWI, modified curvature, modified DTW, and selected soil thematic maps. Input variables were modified based on site characteristics and information provided by VDOT delineations in order to produce distinct wetland and non-wetland signatures, and user intervention was necessary to execute some of the calibration processes. Summarized modification parameters for topographic indices and information relevant to their calculation are shown in Table 2 and the methods used to calculate these parameters are described in the following sections.

Table 2. Modification parameters for topographic indices, and soil thematic maps determined to be relevant for each study site.

	Site 1	Site 2	Site 3	Site 4
<i>LiDAR DEM Resolution (m)</i>	1.0	1.52	0.76	0.76
TWI Focal Statistic Window size (# pixels)	5	3	7	7
Curvature Focal Statistic Window size (# pixels)	5	3	7	7
TWI Focal Statistic Type	Median	Median	Median	Median
Curvature Focal Statistic Type	Mean	Mean	Mean	Mean
<i>Maximum Underlying Wetland Slope Value (m/m)</i>	0.751	1.134	1.652	1.403
<i>Representative Wetland Slope (m/m)</i>	0.088	0.168	0.41	0.115
DTW γ	11.42	5.95	2.44	8.70
DTW β	2	2	2	2
Hydrologic Soil Group	X	X		
Depth to Water Table	X			X
Surface Texture	X	X		
Hydric Rating		X		X
Soil Drainage Class		X		

Note: Site characteristics relevant to the calculation of modification parameters are italicized and inclusion of a soil layer is indicated by an "X."

2.4.1.1.1 TWI Modifications

The modified TWI grid is based on the TWI as defined in Eq. (1). The TWI was created in ArcGIS as a Map Algebra expression. The inputs required for this calculation were a flow accumulation grid, to represent the α term, and a slope grid, to represent the $\tan(\beta)$ term, both derived from the filled LiDAR DEM. The D8 method (Jenson & Domingue, 1988) was used to generate flow direction and flow accumulation grids. A slope grid was generated with the ArcGIS slope tool, calculated as the steepest downhill descent from each pixel in units of m/m (Burrough et al., 2015). A constant equal to 1 was added to flow accumulation grids so that every pixel received flow from itself as well as upslope pixels to avoid undefined TWI values, and a constant equal to 0.0001 (m/m) was added to slope grids to avoid dividing by zero. An example of the resulting TWI grid, overlaid with VDOT wetland areas, for a portion of Site 1 is shown in Figure 5 (panel A1). This TWI grid models the presence wetter areas (high TWI values) in locations of high flow accumulation and low slopes, and drier areas (low TWI values) in locations of steep

slopes and less flow accumulation. Larger clusters of relatively high TWI values align with the VDOT delineated wetlands, however there is also a scattering of high TWI values outside of these wetland boundaries, corroborating the challenges of high-resolution TWIs previously described in the literature (e.g., Ågren et al., 2014; Sørensen & Seibert, 2007). Some researchers recommend deriving TWIs from coarser DEMs (e.g., Ågren et al., 2014), but doing so would sacrifice the rich detail provided by LiDAR DEMs that may be needed to precisely model shape and size of environmental planning-scale wetlands.

Although these scatterings of relatively high TWI values may be modelling true micro-topographic features, their location outside of the field-mapped wetlands suggest these flow channels are not large enough to result in saturated conditions. Rather than lose hydrologic detail of the LiDAR data by resampling, anomalous local variations were smoothed by applying a low-pass filter over a moving $N \times N$ window to create the modified TWI. Applying a low-pass filter searches over a user-defined window in which every pixel is replaced with the statistical value from the surrounding pixels within the $N \times N$ window, as done by Ali et al. (2014), Buchanan et al. (2014), and Lanni et al. (2011). The window size for the smoothing operation is significant in that it is usually set with consideration of the average size of the feature of interest (Sangireddy et al., 2016). In this study we estimated that areas of interest must be at least 5 m in width based on the size of VDOT delineated wetlands. Therefore, window sizes were set to smooth over a total area of approximately 25 m² (5 m x 5 m) with this window size varying slightly across study sites depending on pixel length of the LiDAR data. Additionally, a median filter was chosen to perform smoothing rather than the mean filter. Visual assessment of both statistic types showed that the median filter better retained VDOT wetland edge features while removing scattered high TWI values outside of these boundaries. TWI smoothing was implemented in the ArcGIS model using the Focal Statistics tool. Window sizes used to calculate the modified TWI grid for each site are shown in Table 2, and an example of applying this modification for a portion of Site 1 is shown in Figure 5, panel A2. Compared to the unmodified TWI (panel A1), this scene shows the larger cluster of relatively high TWI values within VDOT delineated wetlands were maintained, but the discrete, small flow channels outside of the true wetland boundaries have been smoothed via replacement of these pixels with relatively lower TWI values.

2.4.1.1.2 Curvature Modifications

Curvature grids, as defined by Moore et al. (Moore et al., 1991) were created from the filled LiDAR DEM using the ArcGIS Curvature tool. Curvature has been shown to be a key component in the process of identifying likely channelized pixels indicating flow convergence (Ågren et al., 2014; Hogg & Todd, 2007; Kloiber et al., 2015; Millard & Richardson, 2013; Sangireddy et al., 2016). It was anticipated that the high resolution of the LiDAR-derived curvature grids would assist in separating small differences in concavity between nearly flat roadways and shallow local depressions. However, visual assessment of the LiDAR-derived curvature grids showed a similar issue of topographic noise as seen in the TWI, in that micro-topographic channels were also mapped. An example of the output curvature grid for a portion of Site 1 is shown in Figure 5, panel

B1. This image shows negative and zero curvature values within VDOT wetland extents, which correspond to concave and flat areas, respectively.

Similar to modified TWI creation, the curvature was modified by applying a statistical smoothing process to curvature grids, following the approach of Sangireddy et al. (2016). When choosing the window size for this calculation, the assumption of the average size of features of interest was kept consistent with that of the TWI (i.e., at least 5 m in width). In this case a mean filter was chosen to smooth the curvature data rather than a median filter due to a visual inspection and perceived improvement in VDOT wetland edge retention resulting from the mean smoothing. The modified curvature grid was created by applying the ArcGIS Focal Statistics tool. Window sizes used to calculate the modified curvature grid for each site are shown in Table 2 and an example of applying this modification for a portion of Site 1 is shown in Figure 5, panel B2. In this image one can see that the modified curvature grid has a smoother appearance but maintains significant areas of concavity.

2.4.1.1.3 DTW Modifications

The modified DTW grid is based on the DTW as defined in Eq. (2). This iterative function finds the cumulative slope value along the least downward slope (i.e., “cost”) path to the nearest surface water (i.e., “source”) pixel with which it is most likely to be hydrologically connected (Murphy et al., 2009). To calculate the DTW, two input grids are required: a grid of slope values and a grid of areas of open water (Murphy et al., 2009). In this study, slope grids were derived from the filled LiDAR DEM using the ArcGIS slope function, as done in the original formulation of the DTW model (Murphy et al., 2007, 2009, 2011), and the source grids were created from rasterized NHD waterbodies and streams. While the publicly available NHD was chosen in this study to maintain consistency between the four sites, there are alternatives for researchers without publicly available open water data. The source grid can also be generated directly from elevation data by deriving streams based on a designated flow accumulation threshold (Murphy et al., 2009) or use of open source channel extraction software, such as GeoNet (Sangireddy et al., 2016). The effects and limitations of using the relatively coarsely mapped NHD as the source grid for the DTW are discussed in Section 5.2 of this paper. The ArcGIS Cost Distance tool was used to evaluate Eq. (2) within the model using the slope and NHD source grids as inputs. It was also necessary to add a small constant (0.0001 m/m) to all pixels in the slope grid to differentiate from source grid pixels, which are assigned a value of zero for the calculation. An example of the resulting DTW grid for a portion of Site 1 is shown in panel C1 of Figure 5. As expected, low wetness (high DTW values) occurred in areas further and higher along the terrain from surface water, and high wetness (low DTW values) occurred in areas of low slopes that are closer to surface water. While wetted areas calculated by the DTW correspond to VDOT delineated wetlands, the transition from wet to dryer areas is gradual. We found this to result in lower non-wetland accuracy, or an overestimation of wetlands, when using only the original DTW formulation to identify wetland areas.

Therefore, a modified DTW was created to accelerate the gradual transition from wetlands to uplands in an effort to better distinguish wet from dry locations. The method outlined above

was used to calculate the modified DTW, except that the input slope grid was replaced with an adjusted slope grid, defined as,

$$Y = \gamma * X^\beta, \quad (3)$$

where X is the slope (with a small constant added to all values, as described earlier), and γ and β are calculated slope adjustment parameters. This adjustment to the slope values was intended to create two distinct ranges of low-cost areas, where wetlands are likely to exist, and high cost areas, where wetlands are unlikely to exist, based on the observed distribution of wetland slope values in each site. The γ parameter allows users to control the cutoff between the low and high cost slope values, which corresponds to a designated representative wetland slope value. The β parameter allows users to control the rate of increase in cost as the slopes increase throughout the site. In this study, β was set to a value of 2 for all sites while γ was individually calibrated. We hypothesized that setting the wetland slope value equal to the 95th percentile of all underlying VDOT wetland slope values would result in a γ parameter that further flattens the terrain (i.e., reduces the cost) where most wetlands exist, disregarding assumed outliers, and further steepening the terrain (i.e., increasing the cost) elsewhere. Representative slope values were calculated by extracting slope values within VDOT wetland boundaries, and calculating the 95th percentile of each array with the Numpy Python library. Figure 4 shows an example of this adjusted slope calculation and describes the effect of this adjustment for Site 1, where the 95th percentile was 0.088 m/m, which corresponded to a γ value of 11.42.

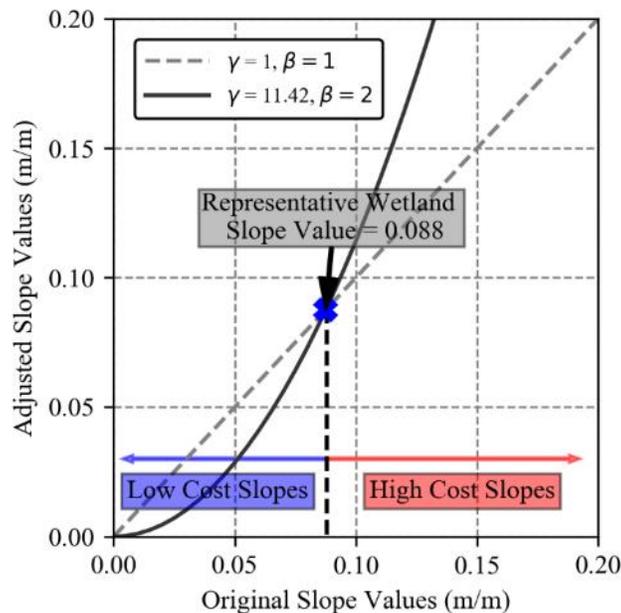


Figure 4. Example calculation of the adjusted slope grid (solid line) for Site 1 where the β was set to a value of 2 and γ was calculated to be 11.42, corresponding to a representative slope value taken to be the 95th percentile of all underlying wetland slopes. These adjustments decrease slopes that are originally below 0.088 and increase slopes that are originally above 0.088, relative to a slope grid (dashed line) where γ and β are both equal to 1.

Note: Although maximum wetland slope value in Site 1 was 0.751 m/m, a smaller range of values is shown here for clarity.

With the adjustments to the slope grid applied, Eq. (2) becomes

$$\text{Modified DTW (m)} = \left[\sum \gamma \left(\frac{dz_i}{dx_i} \right)^2 a \right] * x_c, \quad (4)$$

where γ and $\beta=2$ are introduced. Slope adjustment parameters and relevant site characteristics used to calculate these parameters are shown for each site in Table 2. An example of the effect of modifying the DTW in Site 1 using this calculation is shown in panel C2 of Figure 5. In this figure, the modified DTW (C2) shows relatively wetter areas within VDOT wetland boundaries and an accelerated increase to drier values moving away from VDOT wetlands, compared to the original DTW (C1).

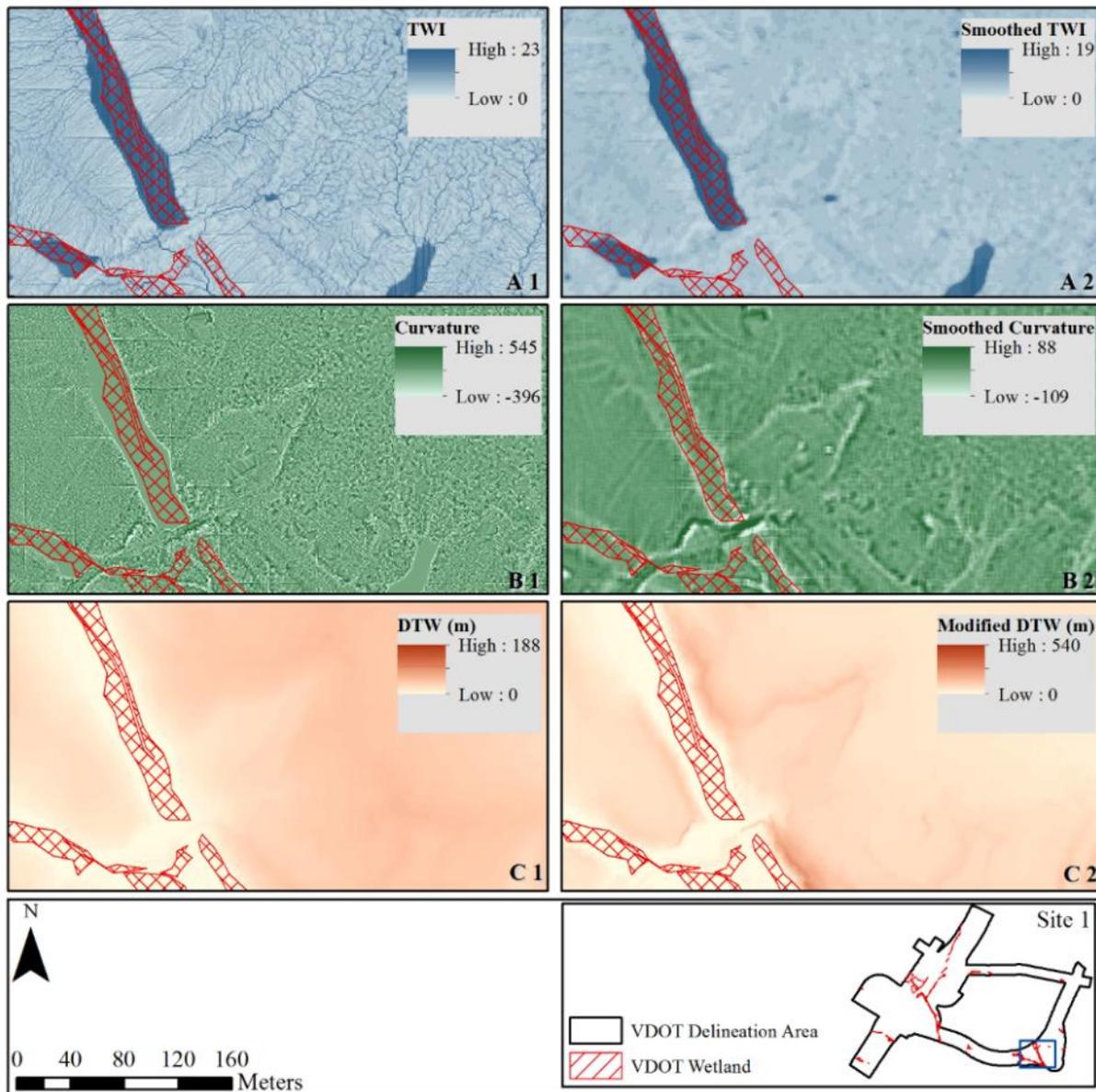


Figure 5. Topographic input variables in Site 1, original TWI (A1), curvature (B1), and DTW (C1), compared to modified versions each variable, shown in A2, B2, and C2, respectively.

Note: Panels A1 and B1 highlight anomalies in elevation data that are likely artifacts of LiDAR tile merging during original processing of raw data.

2.4.1.1.4 Soil Thematic Maps

The final input variables created in the preprocessing phase were soil thematic maps. Soil thematic maps were created from the extensive SSURGO database using the Soil Data Viewer ArcMap extension (USDA, 2019). Although the Soil Data Viewer creates soil thematic maps automatically, combinations of soil layers were manually chosen for each site based on correspondence of the soil data to the current physical landscape. This correspondence was assessed by visual comparison to VDOT delineations and VITA land cover data. Soil layers that appeared too coarse, i.e. generally did not vary enough within the VDOT delineated area to describe features of interest, were not selected.

2.4.1.2 Training and Testing Data

An automated process was used to randomly designate 10% of VDOT delineation area to train the classifier and reserve the remaining 90% to test the classification results. It has been noted that statistical classifiers and machine learning algorithms may be sensitive to imbalanced training data or cases where rare classes are being classified (such as most cases of wetland identification), and the sensitivity of RF, specifically, to training class proportions was investigated by Millard and Richardson (2015). The researchers found that when training samples were disproportionately higher or lower than the true distribution of that feature, the final classification over or under predicted that class, respectively. They concluded that using a sampling strategy that ensures representative class proportions, and minimal spatial autocorrelation, minimized proportion-error in their results (Millard & Richardson, 2015). In this study we took into account the findings of Millard and Richardson (2015) when designing the methodology to randomly separate VDOT delineations into training and testing data. This process consisted of 4 steps: random point creation, point buffering, value extraction, and training data separation (Figure 6). A stratified random sampling method was used in the first step to distribute a designated number of training sample locations proportionately between wetland and non-wetland areas (panel A). These randomly generated points were then buffered to create circle polygons with an area of approximately 100 m² each (panel B). In the value extraction step (panel C), training data, composed of approximately 10% of the delineated area and with representative class proportions, were produced by rasterizing the buffered polygons with pixel values extracted from VDOT delineations to correct cases of buffered polygons falling into both wetland and non-wetland classes. The testing data were created by separating the training data from the VDOT delineations, leaving approximately 90% of the delineated area to be used for accuracy assessment (panel D). Statistics describing the training and testing datasets for each site are found in Table 3.

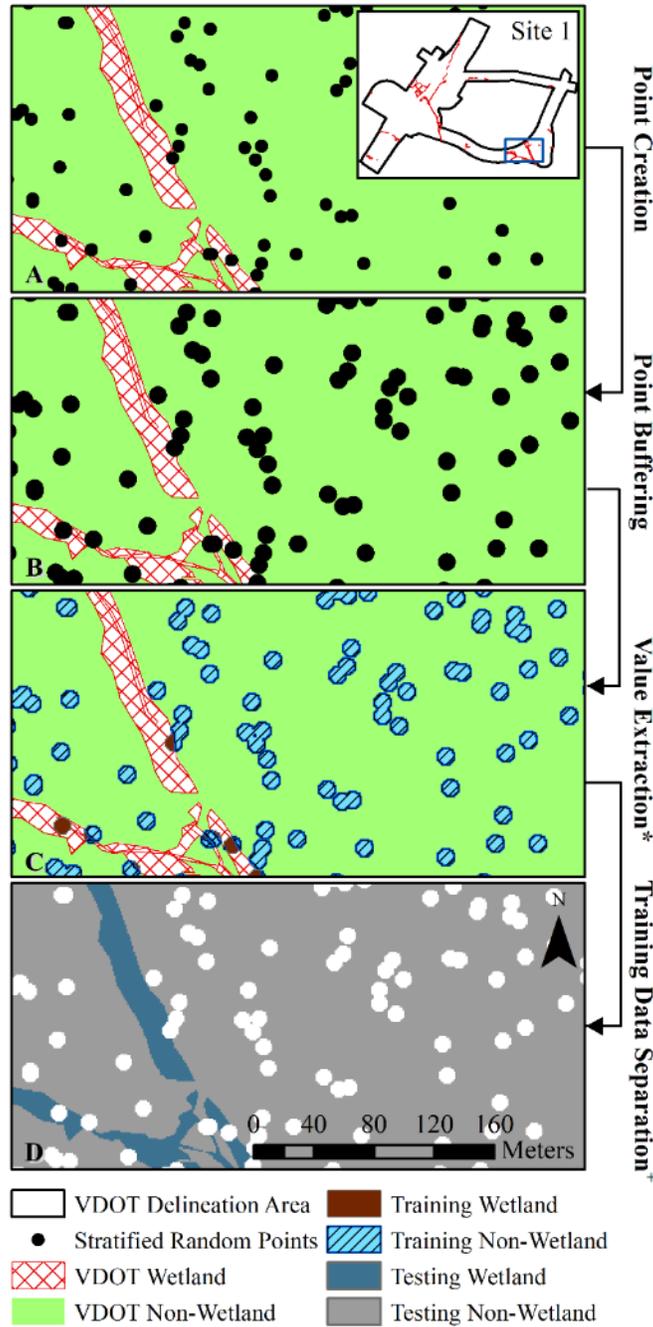


Figure 6. Example of the process, shown for Site 1, used to randomly separate VDOT delineations into training and testing datasets, consisting of four steps: (A) point creation, (B) point buffering, (C) value extraction, and (D) training data separation.

Note: Asterisk indicates the phase in which training data are created and superscript “+” indicates the phase in which testing data are created.

Table 3. Statistics describing the training and testing data for each site.

	Site 1	Site 2	Site 3	Site 4
Training Wetlands (km ²)	0.007	0.015	0.003	0.347
Training Non-Wetlands (km ²)	0.271	0.745	0.172	0.816
Training Wetland to Non-Wetland Ratio	0.03	0.02	0.02	0.43
Training Area to VDOT Delineation Area Ratio	0.09	0.09	0.10	0.10
Testing Area (VDOT Delineation - Training Area) (km ²)	2.71	7.11	1.65	11.00

2.4.2 Supervised Classification

In the first phase of the supervised classification portion of the workflow, the input variables created during preprocessing were combined into a multidimensional, composite image where each dimension stores an independent input variable. Wetland and non-wetland signatures were extracted from this composite image and used to perform the supervised classification. RF classification was chosen as the supervised classification algorithm for its noted advantages in similar studies, as described previously (e.g., Duro et al., 2012; Miao et al., 2012; Millard & Richardson, 2013; Rodriguez-Galiano et al., 2012). RF is an ensemble classifier that produces many Classification and Regression-like trees where each tree is generated from different bootstrapped samples of training data, and input variables are randomly selected for generating trees (Breiman, 2001). This algorithm also produces variable importance information, which measures the mean decrease in accuracy when a variable is not used in generating a tree.

The RF classification was executed in ArcGIS with the Train Random Trees and Classify Raster tools (ESRI, 2016). The Train Random Trees tool utilizes the OpenCV implementation of the RF algorithm (Bradski & Kaehler, 2000). Using Train Random Trees, the training data were used to extract class signatures from the dimensions (i.e., input variables) of the composite image, creating an ESRI Classifier Definition file with variable importance measures. The Classifier Definition file was subsequently used to classify the remainder of the composite image. The result of these operations is a grid where each pixel has been classified as wetland or non-wetland. As the focus of this study was to analyze the response of classification models to input data, the RF parameters were not varied or calibrated to study sites. For this reason, the default values of maximum number of trees, maximum tree depth, and maximum numbers of samples per class were held constant at the recommended default values of 50, 30, and 1000, respectively. Future work should perform a sensitivity analysis to test the effect of adjusting these parameters.

2.4.3 Post Processing

The first phase of post processing was post classification filtering. The objective of the post classification filtering was to account for areas that may be susceptible to water accumulation due to its local topography, but cannot be wetland areas due to impervious land cover. The post classification filtering algorithm first used a logical statement to determine if a classified wetland pixel overlaps VITA land cover designated as impervious. If this was false, the pixel classification was unchanged. If this was true, a second logical statement was used to account for cases where wetlands may exist under bridges by determining if classified wetland pixels are within 30 m of

NHD streams. The 30-m buffer distance was an estimated value based on visual inspection, and more precise measurements would increase effectiveness of post classification filtering. If this second statement was false, the pixel was reclassified as non-wetland, otherwise it was left unchanged. This process produced the model predictions.

The second phase of post processing was accuracy assessment. The model predictions and NWI map for the study area were assessed for accuracy in terms of agreement with the test dataset. Accuracy assessments were evaluated with confusion matrices, which summarized the areas of wetland agreement, non-wetland agreement, false negative predictions (cases where true wetland areas were predicted to be non-wetland), and false positive predictions (cases where true non-wetland areas were predicted to be wetland). Confusion matrices for the model predictions and NWI maps were used to calculate wetland accuracy, non-wetland accuracy, and overall accuracy using Eqs. (5-7),

$$\text{Wetland Accuracy} = \frac{\text{wetland agreement (km}^2\text{)}}{\Sigma \text{test (actual) wetland (km}^2\text{)}} \quad (5)$$

$$\text{NonWetland Accuracy} = \frac{\text{nonwetland agreement (km}^2\text{)}}{\Sigma \text{test (actual) nonwetland (km}^2\text{)}} \quad (6)$$

$$\text{Overall Accuracy} = \frac{\text{wetland agreement (km}^2\text{)} + \text{nonwetland agreement (km}^2\text{)}}{\Sigma \text{test (actual) area (km}^2\text{)}}. \quad (7)$$

The use of these metrics to assess wetland classifications is common in literature (e.g., Ågren et al., 2014; Millard & Richardson, 2013).

2.5 Results and Discussion

2.5.1 Highest Performing Models

To determine the highest performing models, classifications varying only topographic inputs were first performed and assessed, and the input data that resulted in highest overall accuracy were combined with relevant soil layers, if any. In the coming sections, the following results are discussed: (1) scenes for each site comparing highest performing models and their level of agreement with VDOT delineations, compared to NWI maps, (2) variable importance of highest performing input data, and (3) the accuracy assessment of highest performing models compared to the NWI. The input data used to produce the best performing models and the importance of these inputs according to the ESRI Classifier Definition file are listed in Table 4. Although accuracy assessments for each site only extend to testing dataset limits, scenes depicting predictions and VDOT delineations prior to the separation process are shown for clarity.

Table 4. Input data that produced the highest performing wetland identification model in each site, in terms of overall accuracy, as well as variable importance and rank of each input variable according to the ESRI Classifier Definition file.

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8
Site 1	TWI*	Curvature*	DTW*	HSG ¹	Depth to WT ²	ST ³	-	-
VI ⁺	0.087	0.111	0.333	0.131	0.182	0.156	-	-
Rank	6	5	1	4	2	3	-	-
Site 2	TWI*	Curvature*	DTW	HSG ¹	-	ST ³	HR ⁴	DC ⁵
VI ⁺	0.078	0.107	0.156	0.208		0.126	0.177	0.150
Rank	7	6	3	1		5	2	4
Site 3	TWI*	Curvature*	DTW*		-	-	-	-
VI ⁺	0.158	0.325	0.516					
Rank	3	2	1		-	-	-	-
Site 4	TWI*	Curvature*	DTW*		Depth to WT ²		HR ⁴	
VI ⁺	0.076	0.114	0.215		0.338		0.257	
Rank	5	4	3		1		2	

Note: Topographic inputs with an asterisk indicate the application of modifications using parameters from Table 2.

2.5.1.1 Site 1 Results

Wetland predictions and NWI data for Site 1 are shown in Figure 7. Both of the NWI scenes (A1 and B1) exemplify the tendency of the NWI to underestimate the size of VDOT delineated wetlands by mapping wetlands primarily along streams. While the narrow NWI wetlands precisely map the wetland areas that are in agreement with VDOT delineations, the NWI fails to match the contours or the size of larger wetland zones. These larger wetland zones were more fully mapped by wetland predictions produced by the model (A2 and B2). However the model also produced relatively higher overestimation of wetlands. Overestimation of wetlands is especially prevalent in location 1. Underlying input variables indicated that overestimation here was due to a depression that was filled to become a large, zero-slope area. This flat zone resulted in a corresponding generalized area of high wetness values in the modified TWI and modified DTW. In addition, the surface texture input indicated that silty clay loam, which have relatively slow infiltration rates (~0.5 cm/h) (Soil Survey Staff, 2017a), was also present in this overestimated area, likely contributing to the wetland predictions here. It is possible that the results in this site could be improved by using an alternative to the pit filling (i.e., ArcGIS Fill) algorithm to avoid creation of generalized, flat areas, more severe adjustments to the slope grid for the modified DTW, or higher resolution SSURGO data. Panel B2 shows more precise model wetland predictions, represented by conformity of predicted wetlands to the curvature of VDOT delineated wetlands. This panel encompasses the scene in Figure 5 (C2) where the modification to the DTW was shown to more precisely map wetland areas. For that reason, we attribute the relatively precise mapping of wetlands in B2 in part to the modifications used for the DTW in this site. Location 2 shows one small wetland that was undetected by the model. This may indicate a wetland formed due to conditions more strongly driven by vegetation rather than topography or proximity to surface water.

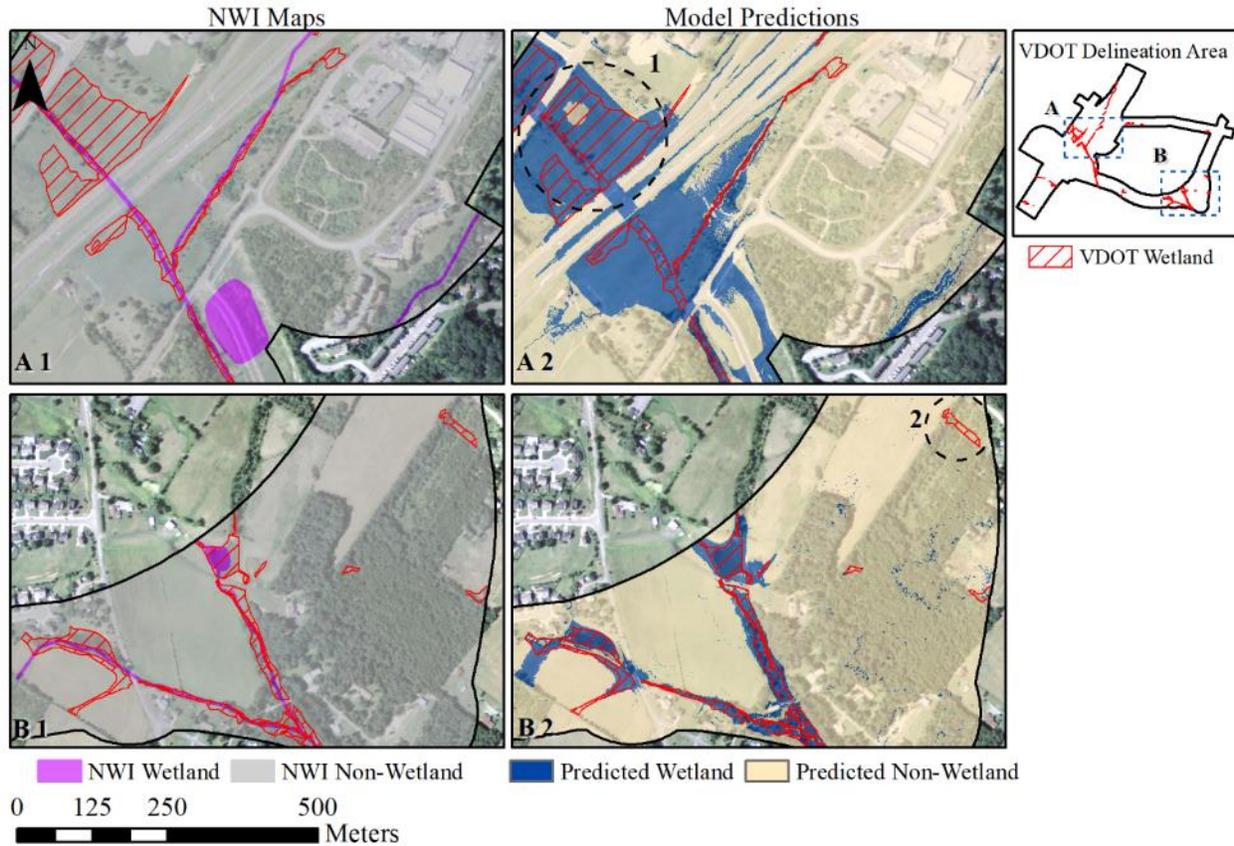


Figure 7. Examples of NWI maps (A1 and B1) and model predictions (A2 and B2) for Site 1, both compared to VDOT delineations.

2.5.1.2 Site 2 Results

Two scenes of the model predictions and NWI maps for Site 2 are shown in Figure 8. In panels A1 and A2, the NWI dataset and model predictions both show similar overestimation of wetland area, although the model resulted in higher overestimation. The false positive predictions in this area were due to flow convergence indicated by the topographic inputs, and the presence of hydric soils indicated by the SSURGO data. Also, many false positive predictions in this site were in locations overlapping road features (e.g., location 1). This may indicate a need for alternate modifications to topographic inputs, especially curvature, to better differentiate channelized areas due road features from channelized areas that are wetlands, as proposed by Sangireddy et al. (2016). Panel B1 shows another example of NWI wetland delineations following along streams, but failing to capture the extents of larger wetland zones. For this same area, the model predicted wetlands further from the streambeds due to the gradual slopes surrounding them and better encompassed VDOT delineated wetlands (locations 2 and 3).

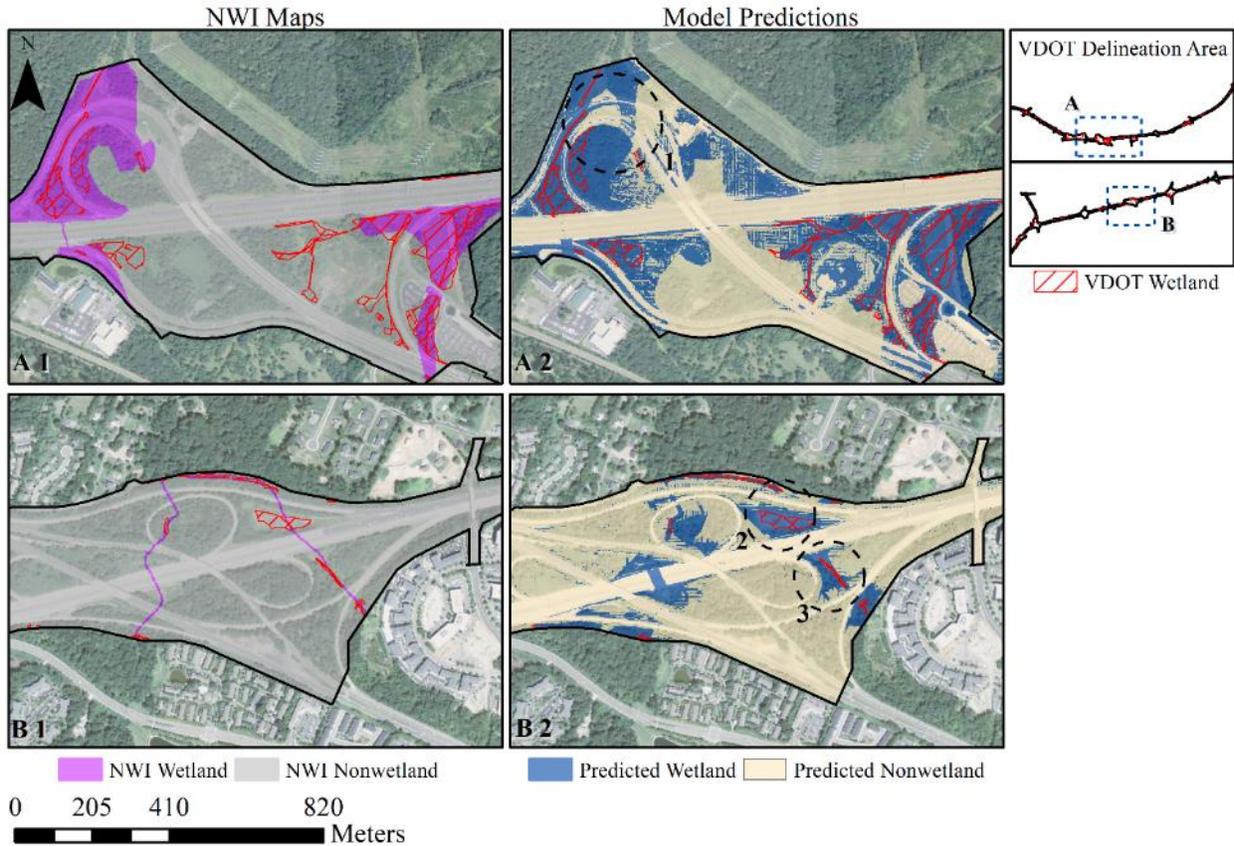


Figure 8. Examples of NWI maps (A1 and B1) and model predictions (A2 and B2) for Site 2, both compared to VDOT delineations.

2.5.1.3 Site 3 Results

Examples of model predictions and NWI data for Site 3 are shown in Figure 9. As seen in Table 4, Site 3 was unique in that no soil layers were included in the best performing model. Visual assessment of relevant soil layers in this area showed that the SSURGO data did not vary in a way that effectively differentiated between features of interest. Site 3 was also unique for its wetlands which were typically narrow and located along small flow channels, rather than in larger wetland zones. The NWI data shown either do not conform to the bends along the length of wetlands (A1), or failed to map a number of wetlands in these channelized areas (B1). The model predicted a larger portion of the VDOT delineated wetlands in both scenes, however the wetland predictions often extended too far on either side of the narrow wetlands (A2). Location 1 shows another example of a local depression filled to become a generalized, flat area, resulting in an overestimation due to the modified TWI and modified DTW indices. Additionally, both scenes A2 and B2 show that the model detected road edges and road medians as wetland areas. This is a shortcoming of the model that was observed in other sites, such as Site 2, and indicates a need for further modification to topographic indices.

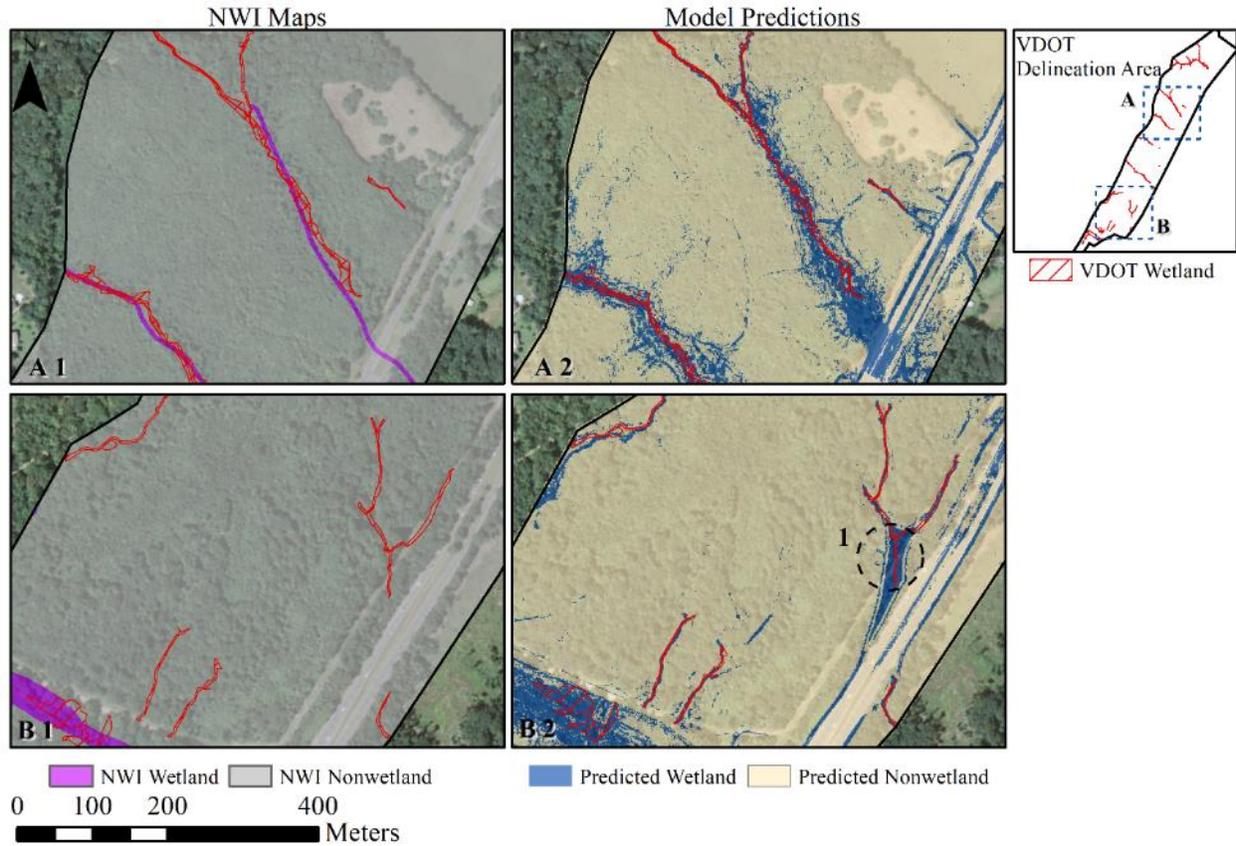


Figure 9. Examples of NWI maps (A1 and B1) and model predictions (A2 and B2) for Site 3, both compared to VDOT delineations.

2.5.1.4 Site 4 Results

Figure 10 shows three scenes from the NWI maps and model predictions for Site 4, which was the largest site studied. Site 4 was also unique for having the largest distribution of VDOT delineated wetlands, covering more than 40% of the surveyed area, as well as the mildest average slope (see Table 1). NWI maps underestimated a large portion of VDOT delineated wetlands, and the portions of these wetlands that were mapped were delineated with less precision than typically seen by the NWI (e.g., location 2). The model predictions also resulted in a large number of false negative predictions and imprecise wetland delineations. The well-defined contours of model predictions (e.g., locations 1, 3, and 4) exemplify the heavy reliance of the model on soil thematic layers. In these scenes, the primary drivers for wetland prediction were the presence of hydric soils and shallow depth to water table, which both outlined the same contours as these wetland predictions. The relatively lower reliance on topographic indices in this site is likely due to the unchanging topography of the area, which is characteristic of the Mid-Atlantic Coastal Plain, as there was often little to no flow convergence indicated by the topographic indices where VDOT delineated wetlands were mapped. It is possible that alternative filtering techniques or more severe adjustments to the slope grid could increase the effectiveness of topographic indices to detect wetted areas, however the correspondence of the model to the soil layers used and the relatively

high occurrence of false negative predictions imply that vegetation data would also be valuable in this region.

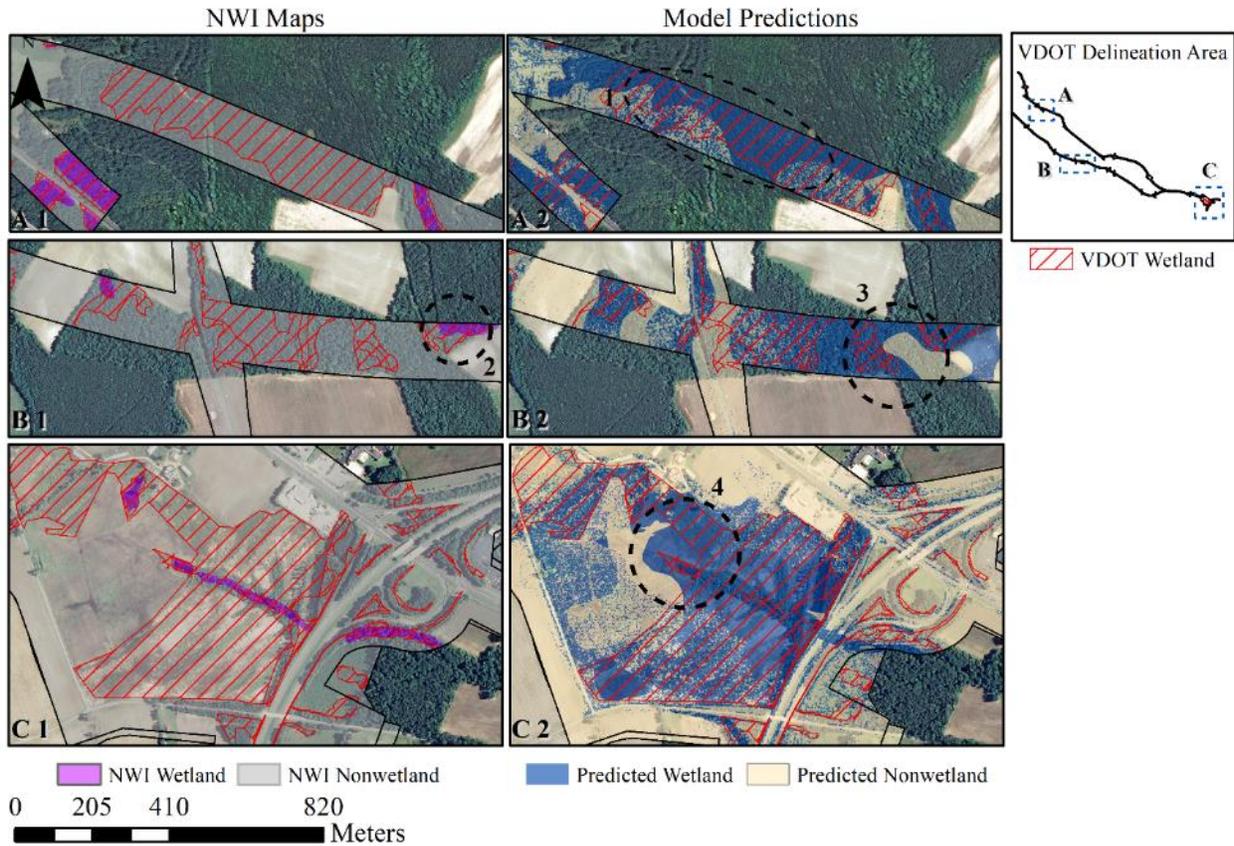


Figure 10. Examples of NWI maps (A1, B1, and C1) and model predictions (A2, B2, and C2) for Site 4, both compared to VDOT delineations.

2.5.1.5 Variable Importance

An important output from the RF classification was the ESRI Classifier Definition file, which provided the variable importance of each input used in classifications (see Table 4). Variable importance measures were used to gauge the ability of input variables to provide unique, significant information to the classifier. Table 4 shows that in Site 1, Site 3, and Site 4, the modified DTW was the most important topographic index, and in Site 2 the original DTW was the most important topographic index. In contrast, the modified TWI was the overall least important input variable in every study site. The low ranking of the modified TWI relative to the modified and original DTW suggests that some information was duplicated by these inputs, but that the modified DTW provided more robust wetland and non-wetland signatures. This corresponds to the findings of previous studies (e.g., Ågren et al., 2014; Murphy et al., 2009) which stated that wet TWI values were restricted to discrete lines of flow accumulation within wetted areas, whereas the DTW model effectively encompassed wetted areas as a whole and was therefore more robust. For this same reason, it was unexpected that for Site 3 the modified DTW ranked higher than the modified TWI, as the VDOT delineated wetlands here were primarily restricted to narrow lines of flow

accumulation. Soil data were among the most important variables in all sites that included them. In Site 1 and Site 2, this is likely due to the heavy presence of road features and the ability of the soil information to better distinguish these from wetland features relative to the topographic indices, which were observed to detect water accumulation near these features. The higher importance of soil layers in Site 4 is likely due to the flat terrain, and is in line with the wetland predictions seen in Figure 10, which were dictated primarily by areas of hydric soil and shallow depth to water table. The low importance of the topographic indices in Site 4 also reinforces the claim that topographic indices that are static and assume the local slope is an adequate proxy subsurface flow patterns, such as the TWI and DTW, are less suitable in flat areas due to undefined flow directions that are likely to change over time (Grabs et al., 2009). The lower importance of modified curvature relative to DTW inputs in all sites may indicate that our application of the curvature was limited by the ArcGIS fill operation and smoothing, which generalized potentially significant terrain features, since curvature has been shown to strongly determine flow convergence in flat topography (Sangireddy et al., 2016).

2.5.1.6 Accuracy Assessment

The accuracy of model predictions was assessed using the testing data, and compared to the accuracy achieved by the NWI maps. Table 5 shows the confusion matrices produced for the best performing model and the NWI maps across all study sites. In each confusion matrix, test data are represented along columns and NWI and model predictions are represented along rows. Categorized pixels (expressed as total km²) in Table 5 were used to calculate wetland accuracy, non-wetland accuracy, and overall accuracy using Eqs. (5-7). It is important to note that the accuracy assessment only extended to the limits of the testing data, which as previously described, are randomly selected subsets of the original VDOT delineations, and the effect of varying testing and training data separation on model accuracy was not assessed.

Table 5. Confusion matrices used to assess the accuracy of NWI maps (left) and best performing model predictions (right) compared to the test data, where predicted values are represented along rows and actual values are represented along columns.

Site 1									
		Test Data (actual)					Test Data (actual)		
		Wetland (km ²)	Non-Wetland (km ²)	Σ=			Wetland (km ²)	Non-Wetland (km ²)	Σ=
NWI (predicted)	Wetland (km ²)	0.012	0.034	0.05	Model (Predicted)	Wetland (km ²)	0.056	0.202	0.26
	Non-Wetland (km ²)	0.053	2.605	2.66		Non-Wetland (km ²)	0.009	2.441	2.45
	Σ=	0.07	2.64	2.7		Σ=	0.07	2.64	2.7
Site 2									
		Test Data (actual)					Test Data (actual)		
		Wetland (km ²)	Non-Wetland (km ²)	Σ=			Wetland (km ²)	Non-Wetland (km ²)	Σ=
NWI (predicted)	Wetland (km ²)	0.064	0.280	0.34	Model (Predicted)	Wetland (km ²)	0.127	1.038	1.16
	Non-Wetland (km ²)	0.084	6.673	6.76		Non-Wetland (km ²)	0.021	5.912	5.93
	Σ=	0.15	6.95	7.1		Σ=	0.15	6.95	7.1
Site 3									
		Test Data (actual)					Test Data (actual)		
		Wetland (km ²)	Non-Wetland (km ²)	Σ=			Wetland (km ²)	Non-Wetland (km ²)	Σ=
NWI (predicted)	Wetland (km ²)	0.010	0.022	0.03	Model (Predicted)	Wetland (km ²)	0.026	0.203	0.23
	Non-Wetland (km ²)	0.020	1.592	1.61		Non-Wetland (km ²)	0.004	1.411	1.41
	Σ=	0.03	1.61	1.6		Σ=	0.03	1.61	1.6
Site 4									
		Test Data (actual)					Test Data (actual)		
		Wetland (km ²)	Non-Wetland (km ²)	Σ=			Wetland (km ²)	Non-Wetland (km ²)	Σ=
NWI (predicted)	Wetland (km ²)	1.052	0.116	1.16	Model (Predicted)	Wetland (km ²)	2.648	1.717	4.37
	Non-Wetland (km ²)	2.220	7.596	9.81		Non-Wetland (km ²)	0.625	6.005	6.63
	Σ=	3.27	7.71	11.0		Σ=	3.27	7.71	11.0

Note: Wetland, non-wetland, and overall accuracy rates are derived from values in the confusions matrices using Eqs (5-7). Values shown are rounded for clarity.

Figure 11 summarizes the accuracy achieved by the best performing model predictions and NWI maps. In the context of the wetland permitting process, it is important to have high values for all accuracy metrics. To uphold the objective of protecting existing wetlands, wetland accuracy is of high importance, and in order to provide realistic estimates of potentially impacted wetland areas in transportation and environmental planning, non-wetland accuracy is also necessary. However, it is important to be aware of the potential for overall accuracy, which measures the portion of the entire area that is correctly classified regardless of class, to be misleading due to the uneven distribution of landscape classes. For example, the consistently conservative wetland mapping by the NWI is reflected by the high average non-wetland accuracy (98.0%). Due to the uneven distribution of wetland and non-wetland classes in all but one of the study sites, the conservative nature of the NWI predictions also translated into high average overall accuracy (92.0%), despite an average wetland accuracy of 32.1 %. In contrast, the model predictions resulted in significantly higher average wetland accuracy (84.9%), but at the expense of moderately lower average non-wetland and overall accuracy (85.6% and 85.6%, respectively). As previously

discussed, Site 4 was the lowest performing site. The low wetland accuracy here may be due to a lack of vegetative signatures to distinguish wetland from upland area, especially in this excessively flat area where terrain indices were found to be less important.

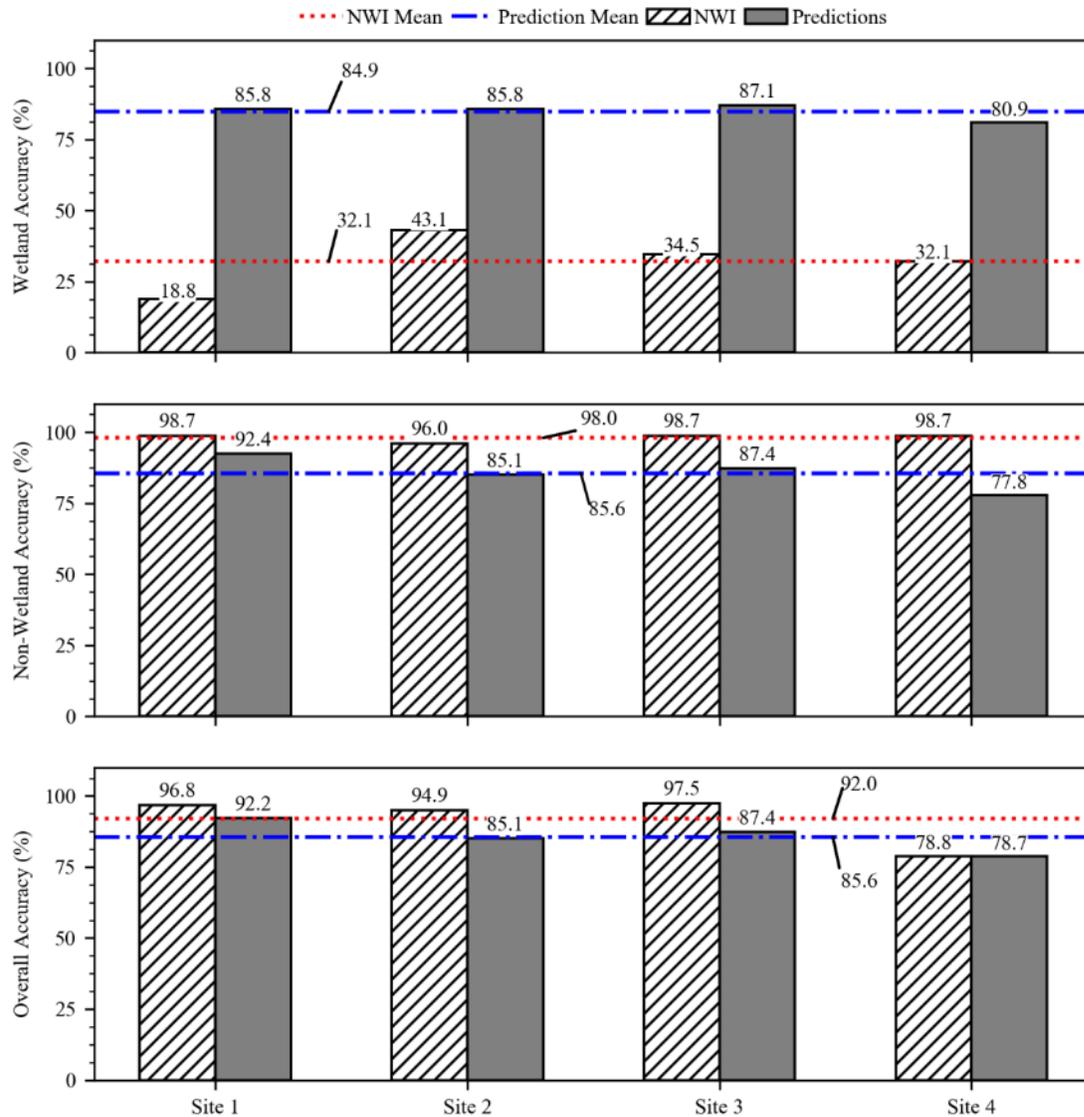


Figure 11. Wetland, non-wetland, and overall accuracy produced by the best performing model predictions, compared to accuracy produced by NWI maps.

2.5.2 Response of Model to Input Data Modification

Iteration results in terms of wetland, non-wetland, and overall accuracy highlight the benefit and cost of applying the modifications described here, as well as including the coarser mapped (1:24,000 to 1:12,000) SSURGO data. Results of the analysis of model responses to classification iterations are shown in Table 6, where the highest performing iteration per accuracy metric, not including iteration 5 which built off of top performing topographic inputs, is indicated with a “+” superscript and modified topographic inputs are indicated with an asterisk.

Table 6. Wetland, non-wetland, and overall accuracy achieved by iterations of RF classification for each site.

Iteration:		1	2	3	4	5
Input Data:		TWI, Curvature, DTW	TWI*, Curvature*, DTW*	TWI, Curvature, DTW*	TWI*, Curvature*, DTW	Best Performing of 1-4, plus soils
Site 1	Wetland Accuracy (%)	86.26 ⁺	83.65	84.47	85.97	85.84
	Non-Wetland Accuracy (%)	88.34	90.45 ⁺	87.77	89.15	92.36
	Overall Accuracy (%)	88.29	90.29 ⁺	87.69	89.08	92.20
Site 2	Wetland Accuracy (%)	67.57	69.85	71.33 ⁺	69.50	85.78
	Non-Wetland Accuracy (%)	83.58	83.87	81.14	84.26 ⁺	85.06
	Overall Accuracy (%)	83.25	83.58	80.94	84.13 ⁺	85.08
Site 3	Wetland Accuracy (%)	82.72	87.12	83.88	88.10 ⁺	-
	Non-Wetland Accuracy (%)	85.20	87.40 ⁺	83.49	86.72	-
	Overall Accuracy (%)	85.16	87.40 ⁺	83.50	86.74	-
Site 4	Wetland Accuracy (%)	55.15	57.11	62.67 ⁺	60.74	80.91
	Non-Wetland Accuracy (%)	69.31	78.03 ⁺	64.44	71.97	77.76
	Overall Accuracy (%)	65.09	71.80 ⁺	63.91	68.63	78.70

Note: Asterisk indicates modifications with parameters from Table 2 were applied and “+” superscript indicates highest performing iteration per accuracy metric.

Shown in Table 6, non-wetland accuracy and overall classification accuracy from iteration 1, where the original versions of all indices were used, improved in every site as a result of modifying all topographic indices (iteration 2). In addition, for three of the four sites, modifying all topographic indices resulted in the highest overall accuracy. These results suggest there is a benefit to applying the modifications presented here rather than using the indices as they are traditionally calculated, where this benefit is a reduction in false positive predictions and increase in overall accuracy. Furthermore, in every site that relevant soil layers were applicable, the inclusion of these soil layers with top performing topographic indices (i.e., iteration 5) further improved the RF classification. From this, we conclude that in these sites, the soil data provided important information to the classifier, despite its relatively coarse scale. Both Site 2 and Site 4 saw relatively high increases in wetland accuracy resulting from iteration 5, which suggests the topographic indices were not effective in encompassing flow convergence or subsurface moisture conditions in order to detect wetlands. Iterations 3 and 4 were performed to determine the effect of individual modifications on the classification. Note that for this evaluation, modified TWI and modified curvature were generalized into a single category of modifications because of their similar adjustment parameters and methods.

The purpose of modifying topographic indices was largely to reduce false positive predictions in that TWI and curvature grids were modified to reduce unrealistic flow convergence due to excess topographic detail, and the DTW was modified to accelerate the transition from wetland to upland areas. Results in Table 6 show that the effect of modifying only the TWI and curvature grids (iteration 4 vs. iteration 1) was an increase in non-wetland accuracy in every study site, as well as an increase in wetland accuracy in all but Site 1. The decrease in wetland accuracy in this site may indicate unintentional smoothing of some features of interest (i.e., too large of a window size), and it is possible that in this study site a mean filter or smaller window would have performed better. In sites 2, 3 and 4, results of iteration 4 suggest the statistic type and window size were effective. Despite the improvements to classifications with these modifications, the modified TWI and curvature grids can be further advanced. The current approach should be

expanded to test the effects of varying window sizes of smoothing filters and statistic type, as well as the TWI formulation.

The effect of modifying only the DTW (iteration 3 vs. iteration 1) appeared to be an increase in wetland accuracy in sites 2, 3, and 4, and an unexpected decrease in non-wetland accuracy in every site. This suggests that while the modified DTW was effective in increasing non-wetland accuracy when combined with modified TWI and modified curvature, the DTW modification alone may not be sufficient for reducing false positive predictions. The limited improvements provided by the DTW modification could be due to the designation of the representative wetland slope value, which may not apply an effective cut off between low and high cost areas. Additionally, improvements to the original DTW calculation before applying modifications may enhance the results of iteration 3. The DTW calculation can be improved first through slope calculation on a DEM corrected with an alternate method, and second by deriving the source grid by extracting surface water features directly from the LiDAR data. In this study, DTW source grids were generated from rasterized NHD data, which are mapped at a coarser scale (1: 24,000 – 1: 12,000) compared to the LiDAR data and therefore, do not capture precise curvature and locations of streams and open water.

2.6 Conclusions

This study evaluated the potential for modification of LiDAR DEM derivatives, combined with ancillary national-scale soil data, to improve a RF classification of wetland areas at a scale relevant for the wetland permitting process, over four study sites in Virginia. The approach was implemented as a model in ArcGIS and performed a RF classification of input variables that were modified to provide distinct wetland and non-wetland signatures. Model predictions were assessed against field-mapped testing data, provided by the Virginia DOT, and compared to NWI maps. Accuracy assessments showed that compared to NWI maps, the highest performing models produced significantly higher average wetland accuracy (84.9% and 32.1%, respectively), while resulting in moderately lower average non-wetland accuracy (85.6% and 98.0%, respectively) and overall accuracy (85.6% and 92.0%, respectively).

Through multiple iterations of input variable combinations, we concluded that there is potential to improve classifications through modification of topographic indices. In every site, the highest performing models included modified topographic indices, and the addition of available soil layers further improved these classifications. Assessment of the variable importance of the highest performing models showed that DTW inputs were of higher importance, compared to the modified TWI in all study sites. This finding supports conclusions of previous studies (e.g., Ågren et al., 2014; Murphy et al., 2009), which state the DTW model provides more robust flow convergence information compared to the TWI. The low variable importance of the TWI relative to the DTW also suggests that there is duplicate information provided between these two indices. In addition, the heavy reliance of the model in Site 4 on soil data reinforces previous findings that topographic indices like the TWI and DTW are less effective in flat areas due to undefined flow directions that are likely to change over time, whereas these indices typically model static

conditions and assume local slope describes subsurface flow patterns (Grabs et al., 2009; Murphy et al., 2009). Through classification iterations, we found that non-wetland and overall classification accuracy increased in all sites when all topographic indices were modified, compared to the accuracy achieved by using the original versions of these indices. While modifications to the DTW alone did not reduce false positive predictions, modifications to only the TWI and curvature did have this effect. However, we believe the DTW modification approach could be further improved on. In addition, iteration accuracies varied by small margins in many cases, and it is important to note that that RF parameters and training and testing data separation were not varied or calibrated to sites in this study. Completing this additional calibration step may produce different outcomes of iteration comparisons.

Results from this study offer a starting point to the enhancement of the model implementation in ArcGIS to include the capability of modifying LiDAR DEM derivatives based on site characteristics to map small-scale wetlands in support of environmental planning and conservation efforts. The results while successful, have also highlighted shortcomings that should be addressed to further enhance the approach and model implementation. We found that the topographic indices were limited by the use of the ArcGIS fill function, which removed local depressions in the LiDAR DEM by creating larger areas of flat terrain. Studies have shown that high-resolution elevation data could be filtered with more sophisticated methods (e.g., Besl et al., 1989; Haralick et al., 1983; Mainguy et al., 1995; Sangireddy et al., 2016; Woodrow et al., 2016), and exploring these methods could improve the accuracy of the topographic indices, especially in low relief areas. The TWI modification can be further advanced on by assessing model responses to alternate TWI formulations such as the D-infinity method for deriving flow accumulation (Tarboton, 1997a) and the Soil Topographic Index formulation which has been shown to improve modelling of soil moisture patterns through inclusion of relevant soil properties (e.g., Buchanan et al., 2014; Lanni et al., 2011). Alternate curvature modifications should also be explored, as this index has been shown to effectively model flow convergence in low-relief and engineered landscapes by applying automated filtering techniques (Sangireddy et al., 2016). Improvements to the DTW modification should include deriving source data directly from LiDAR DEMs through calibrated flow initiation thresholds, as shown by Ågren et al. (2014), and deriving flow accumulation using the D-infinity method (Murphy et al., 2009, 2011), or incorporating the use of other channel extracting software, such as GeoNet (Sangireddy et al., 2016). Furthermore, variable importance indicated that the DTW and TWI may provide duplicate information in many cases, and efforts should be made to effectively combine these indices through a mathematical relationship to reduce feature space for the classifier. Future work should also address the excessive computation times needed to process the high-resolution LiDAR data. Implementing this approach using parallel computing could allow for reductions in runtime needed to calculate γ and β parameters through an iterative calibration to study sites in the DTW modification process. Alternative implementations of the RF algorithm should be tested as well, as the ArcGIS implementation is limited in output data provided to users. Lastly, the approach presented here should be applied to additional study areas to begin to identify modification parameters that can

be effectively generalized by site characteristics. While the prototype model has produced more accurate wetland predictions for the study sites compared to NWI, these improvements would strengthen the potential for this approach to be a useful tool for wetland identification in support of environmental planning decision making in areas where wetland maps are currently unavailable.

Chapter 3

Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model

3.1 Introduction

Wetlands are important ecosystems that are threatened by anthropogenic pressures and climate change (Klemas, 2011). It is estimated that over half of the Earth's wetlands have been destroyed since 1900 (Davidson, 2014). In the conterminous U.S., half of the wetlands have been destroyed since 1600 (Dahl et al., 1991) due to agricultural or development repurposing, pollution, and climate change (Klemas, 2011). In the U.S., federal regulations play an important role in the protection of remaining wetlands. Specifically, Section 404 of the Clean Water Act requires environmental impact assessments prior to land development and water resources projects (Page & Wilcher, 1990). This law requires environmental planning entities to provide detailed wetland delineations to the U.S. Army Corps of Engineers (USACE), which can be time-consuming and costly to produce. There is potential for computational models to streamline the delineation process by providing accurate wetland inventories that limit manual surveying to likely wetland areas.

Wetlands can be identified by common features, including the presence of hydrologic conditions that inundate the area, vegetation adapted for life in saturated soil conditions, and hydric soils (Environmental Laboratory, 1987). Remotely sensed data offer new opportunities to accurately and rapidly observe these features at varying scales (Guo et al., 2017; Lang et al., 2013; Lang & McCarty, 2014). Multispectral imagery, radar, and Light Detection and Ranging (LiDAR), data have proven useful for a range of wetland conservation applications, including wetland mapping (Guo et al., 2017). However, availability of multispectral imagery and radar at resolutions fine enough to detect small-scale wetlands is lacking, and obtaining these data can be costly. Alternatively, LiDAR emerges as a candidate for wetland identification, especially on large scales, due to its wide, and growing, availability and demonstrated benefit to wetland mapping (Kloiber et al., 2015; Lang & McCarty, 2014; Snyder & Lang, 2012). LiDAR returns can be interpolated to create high-resolution digital elevation models (DEMs), from which topographic metrics can be derived that describe flow convergence and near-surface soil moisture to indicate wetlands (e.g., Lang et al., 2013; Lang & McCarty, 2014; Millard & Richardson, 2013, 2015; O'Neil et al., 2018). Additionally, studies have demonstrated the benefit of LiDAR DEM metrics as input variables to the Random Forest (RF) classification approach (Breiman, 2001) for wetland mapping and classification (e.g., Deng et al., 2017; Kloiber et al., 2015; Millard & Richardson, 2013, 2015; O'Neil et al., 2018; Zhu & Pierskalla, 2016). Deriving topographic metrics from higher resolution DEMs (i.e., < 2 m) has been shown to increase accuracy of saturation extent mapping (Hogg & Todd, 2007; Lang et al., 2013; Millard & Richardson, 2015). However, the replacement of

conventional DEMs with LiDAR DEMs requires changes to the traditional hydrologic terrain processing workflow: smoothing and hydrologic conditioning (Lidberg et al., 2017; Passalacqua, Do Trung, et al., 2010; Sangireddy et al., 2016; Woodrow et al., 2016).

DEM smoothing addresses microtopographic noise, which is ubiquitous in high-resolution DEMs and can be the product of erroneous data or true variations in the elevation of the vegetated ground surface (Jyotsna & Haff, 1997). Identifying and filtering noisy data is challenging as it risks artificially modifying the true land surface or degrading features of interest, and no widely-agreed upon approach currently exists (Passalacqua et al., 2015; Pelletier, 2013; Richardson et al., 2009). Although many smoothing techniques have been proposed, this study focuses on methods commonly used in related studies: mean, median, Gaussian, and Perona-Malik filtering. Mean and median filtering have been shown to improve hydrologic parameter extraction from high-resolution DEMs (e.g., Buchanan et al., 2014; O'Neil et al., 2018; Sangireddy et al., 2016; R. Sørensen et al., 2006), whereas Gaussian and Perona-Malik filtering are commonly incorporated into stream localization models (e.g., Hooshyar et al., 2016; Lashermes et al., 2007; Passalacqua et al., 2012; Passalacqua, Do Trung, et al., 2010; Passalacqua, Tarolli, et al., 2010; Pelletier, 2013; Sangireddy et al., 2016).

DEM conditioning resolves topographic depressions prior to calculating flow paths and flow accumulation (Jenson & Domingue, 1988; O'Callaghan & Mark, 1984). Topographic depressions can represent both erroneous data and actual features (Lindsay & Creed, 2005), and their presence interferes with overland flow path modeling by accumulating water, creating flow path discontinuities, and negatively influencing modeled watershed processes (Grimaldi et al., 2007; Lindsay, 2016; Lindsay & Creed, 2005). Furthermore, sensitivity of hydrologic parameter extraction to conditioning technique increases significantly with DEM resolution, making an evaluation of their effects on hydrologic model outcomes especially important for LiDAR DEM applications (Woodrow et al., 2016). Common conditioning techniques include traditional depression filling, breaching, stream burning, and least-cost path algorithms. In this study, evaluated techniques are narrowed to those that require only elevation data and have been used for related studies (e.g., Lidberg et al., 2017; Metz et al., 2011): traditional depression filling (Fill), impact reduction approach (IRA), which combines filling and breaching, and least-cost path search (A*).

The choice of smoothing and conditioning techniques can significantly impact the accuracy of derived hydrologic parameters, however, there is a research gap regarding the compound effects of these processes on subsequent wetland identification. Related studies focusing on either smoothing or conditioning have been largely limited to stream delineation applications. For example, Passalacqua, Do Trung, et al. (2010) found that, compared to Gaussian smoothing, the Perona-Malik method was more advantageous for extraction of channel networks and cross sections, especially in low slope areas. Pelletier (2013) found Perona-Malik, Gaussian, and an additional method, Optimal Weiner, filtering all to be effective in suppressing high-resolution DEM noise for channel network mapping, with tradeoffs between the three depending on the landscape and application. Moreover, Metz et al. (2011) compared the abilities of the Fill, IRA, and A* methods to resolve depressions in coarser, radar-base DEMs, and found that the A* approach provided more accurate drainage networks. In a related study, Lidberg et al. (2017) concluded that, compared to filling techniques, breaching created the most accurate stream

networks from LiDAR DEMs and that differences increased with DEM resolution. A key difference in stream network delineation and wetland delineation is that the former emphasizes connected linear features, whereas wetlands are areal features that may contain irregular topography (e.g., hummocks and hollows), and therefore have irregular and diffuse boundaries.

In this study, we address this research gap by performing a thorough analysis of the compound effects of smoothing and conditioning on wetland delineations and the RF model used to generate them. We test the response of a LiDAR DEM-based RF wetland model to unique combinations of preprocessing techniques for a range of ecoregion, topography, and built environments for four sites of Virginia. We examine the sensitivity of our model to mean, median, Gaussian, Perona-Malik, and no filter, as well as Fill, IRA, and A* conditioning techniques. We train and test the RF model, tuned for the imbalanced wetland and nonwetland distributions in each site, using manually surveyed wetlands provided by the Virginia Department of Transportation (VDOT).

3.2 Study Areas and Input Data

3.2.1 Study Areas

This analysis was completed for four study areas in Virginia, USA (Figure 12a). For each study area, the available data includes the extents of wetland surveys and the HUC 12 watershed (USGS, 2019b) that encompasses the surveys (Figure 12b). The HUC 12 watersheds served as the processing extents for model inputs and surveyed areas delimit the extents of verification data and, therefore, model output. Surveyed areas are referred to as the study sites. The study areas span four level III ecoregions of Virginia. Site 1 is located in the Ridge and Valley ecoregion (67), located between mountainous regions and is characterized by forested ridges and lowland agricultural valleys. Site 2 and Site 3 are located in the Northern Piedmont ecoregion (64), which is a transitional region between low mountains and the flat, coastal Piedmont area. Site 4 spans the Southeastern Plains (65) and the Mid-Atlantic Coastal Plain (63). The Southeastern Plains are comprised of cropland, pasture, woodland, and forest, and the subsurface is predominantly sands, silts, and clays. The Mid-Atlantic Coastal Plain is characterized by low, nearly flat plains and poorly drained soils, and swampy and marshy areas are common (Wilken et al., 2011). Table 7 provides additional characteristics for the study sites. Site 1 and Site 2 contain more impervious area than the other two sites, which are dominated by forested land. The steepest slopes are found in Site 3, where the average slope (0.14 m/m) is nearly twice as steep as or steeper than the average slope for the other sites. In contrast, Site 4 has the mildest slopes with the 90th percentile slope value (0.06 m/m) being less than the average slope in the other sites. While sites 1, 2, and 3 have highly imbalanced wetland to nonwetland distributions, wetlands are much more widespread in Site 4, which is characteristic of the Mid-Atlantic Coastal Plain. While there is a mix of wetland types across sites, Site 3 contains the largest distribution of streams or riverine wetlands, followed by Site 1. Note that all surveyed wetland types were merged into a single wetland category prior to use as verification data.

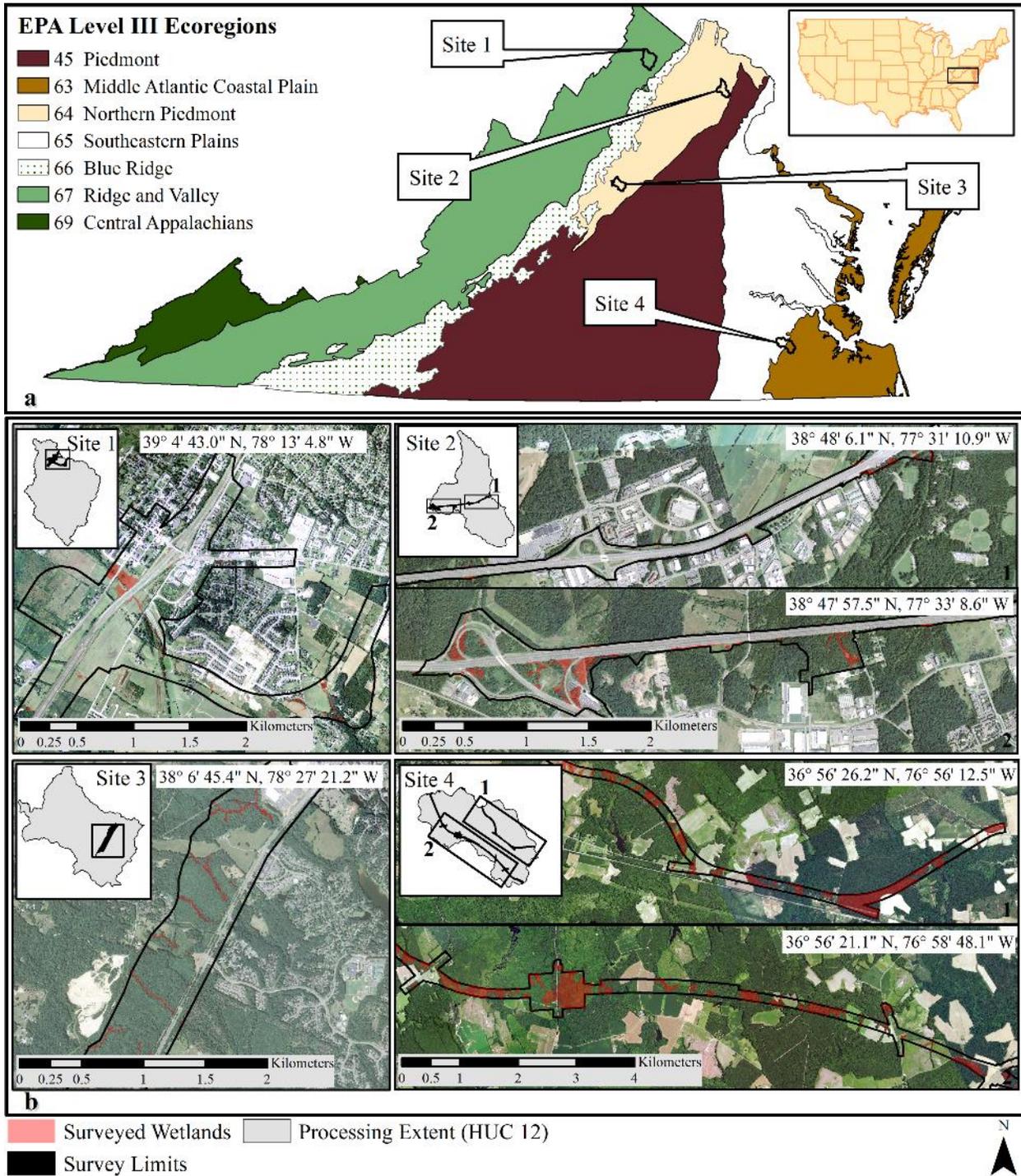


Figure 12. Four study areas spanning four level III ecoregions in Virginia, USA (a). Each study area includes the wetland survey limits, referred to as study sites, and the encompassing HUC 12 watershed, used as the processing extent (b).

Ecoregion data source: US EPA Office of Environmental Information
 Aerial imagery data source: NAIP Digital Ortho Photo Image.

Table 7. Characteristics of each study site, including dominate land cover, topographic characteristics, and surveyed wetland distributions.

	Site 1	Site 2	Site 3	Site 4
Dominating Land Cover ^a	Turf Grass (35%), Developed (22%), Cultivated (20%), Forested (19%)	Developed (36%), Turf Grass (31%), Forested (21%)	Forested (73%), Developed (9%), Cultivated (9%)	Forested (66%), Cultivated (18%), NWI Wetland (9%)
Verification Area (km ²)	2.8	1.6	1.8	5.6
Min. Elevation ^b (m)	209	46	101	10
Max. Elevation (m)	241	107	178	42
10 th Percentile Slope ^c (m/m)	0.02	0.01	0.04	0.01
90 th Percentile Slope ^c (m/m)	0.14	0.20	0.26	0.06
Mean Slope ^c (m/m)	0.07	0.08	0.14	0.03
Wetland : Nonwetland (m ² /m ²)	0.03	0.06	0.02	0.42
Dominating Cowardin Wetland Type(s) ^d	Palustrine Emergent (50%), Streams (20%) ^e	Palustrine Forested (44%), Palustrine Emergent (33%)	Palustrine Forested (56%), Streams (43%)	Palustrine Forested (88%), Palustrine Shrub (9%)

^a Source: Virginia Information Technologies Agency (VITA) Land Cover classifications

(<https://www.vita.virginia.gov/integrated-services/vgin-geospatial-services/land-cover/>).

^b In sites 1, 2, and 4, verification area varied slightly due to edge effects of applying filtering to DEMs.

^c Slope information was calculated from LiDAR DEMs resampled to a 5 m resolution to reduce effect of raw DEM noise on slope information.

^d Values are approximate and according to VDOT wetland surveying reports.

^e Wetland type for remaining 30% of wetland area was not reported.

3.2.2 Input Data

This study used publicly available LiDAR DEMs obtained from the Virginia Information Technologies Agency (VITA) (VITA, 2016). VITA LiDAR DEMs are provided in geotiff format and are hydro-flattened, bare-earth DEMs. The LiDAR data used were collected and processed between 2010 and 2015 and have horizontal resolutions ranging from 0.76 m to 1.5 m. Verification data for this study were provided by VDOT in the form of georeferenced wetland delineations and survey limits, in polygon vector format. All verification wetlands were manually surveyed during summer months (May – August) between 2013 and 2016 by professional wetland scientists in compliance with transportation planning permitting. Wetland delineations for sites 2, 3, and 4 were also jurisdictionally confirmed by the USACE. Binary wetland/nonwetland geotiffs were created from these data, with resolutions matching those of the site LiDAR DEMs. Visual analyses of Google Earth images showed that the study site landscapes changed minimally between LiDAR acquisition and wetland delineation timeframes.

3.3 Methods

The wetland identification algorithm was executed for each unique combination of smoothing and conditioning, producing 15 results for each site. In the following sections, we first outline the wetland identification workflow and then describe the workflow processes and parameters in greater detail.

3.3.1 Overview of the Wetland Identification Model

The wetland identification model is an open source, automated workflow consisting of three main parts: preprocessing, input variable calculation, and classification and accuracy assessment (Figure 13). Input data required include high-resolution DEM data and wetland delineations to serve as verification data, both in geotiff format. Final model outputs are geotiff wetland predictions and an accuracy report. In the preprocessing phase, the input DEM is first smoothed and then conditioned by the set of methods listed in Figure 13. Both the smoothed DEM (DEM_S) and the smoothed, conditioned DEM ($DEM_{S,C}$) are used for calculation of the topographic wetness index (TWI), curvature, and cartographic depth-to-water index (DTW). Training data are derived from the wetland delineations given a user-defined parameter indicating the proportion of wetlands and nonwetlands to sample. These data are used to train the RF model from the merged input variables. The remaining verification data are used to perform an accuracy assessment (i.e., testing data). This workflow is implemented in Python and executed using GDAL, SciPy, GRASS GIS, Scikit-Learn, and PyGeoNet. The code for the wetland identification model is available from GitHub at https://github.com/uva-hydroinformatics/wetland_identification.

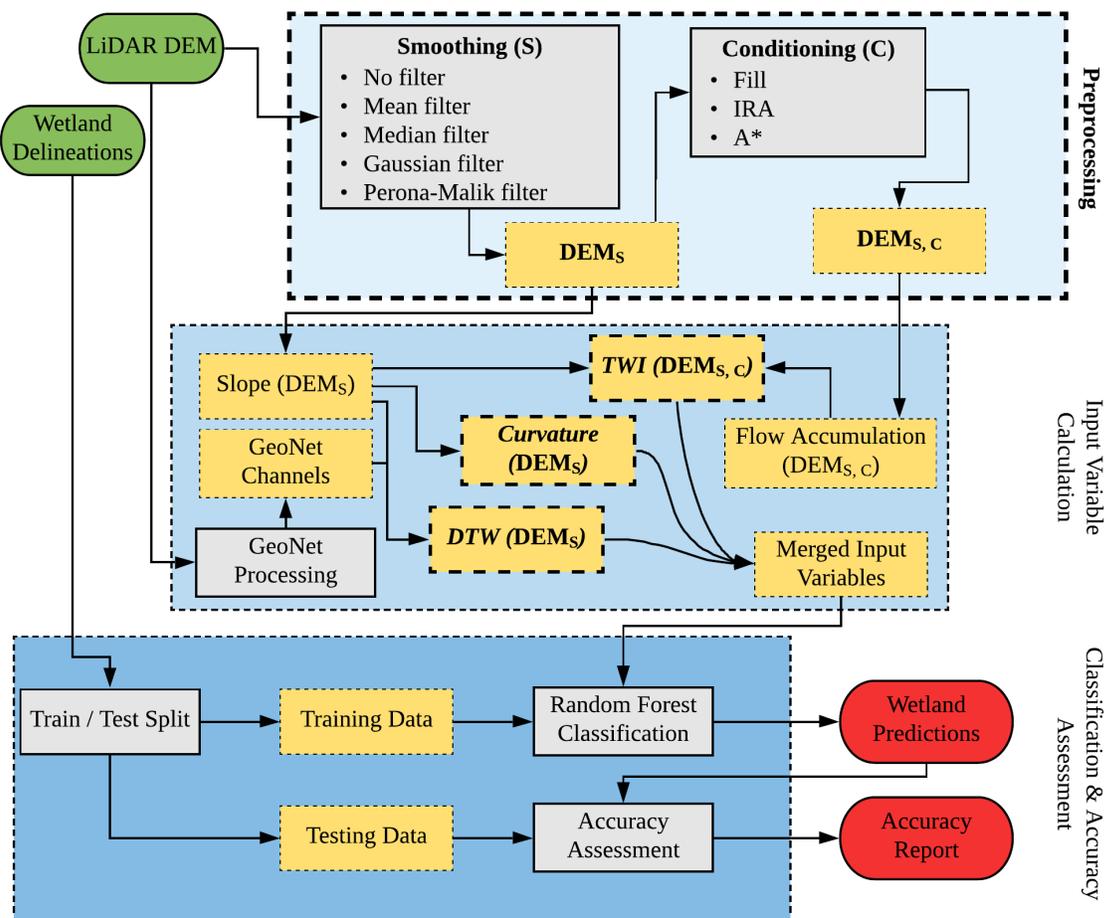


Figure 13. Workflow of the wetland identification model created through this research. Each combination of preprocessing techniques (bold font) was executed for this analysis. Green shapes indicate input data, grey shapes indicate processes, yellow shapes indicate intermediate output, and red shapes indicate final output.

3.3.1.1 Preprocessing

3.3.1.1.1 DEM Smoothing Methods

In addition to no smoothing, mean, median, Gaussian, and Perona-Malik filters were used. Any DEM smoothing should be physically meaningful and serve the purpose of preserving features of interest while smoothing areas smaller than the features of interest (Passalacqua, Do Trung, et al., 2010; Passalacqua, Tarolli, et al., 2010; Sangireddy et al., 2016). As a first step for the analyses, a generalized smoothing scheme was used where constant smoothing scales were applied to all input variables.

It was assumed that features smaller than a 5m by 5m area were insignificant, as the majority (over 90%) of verification wetlands were larger than 25m². This assumption translated to preliminary smoothing scales for mean, median, and Gaussian smoothing. Mean filtering performs a linear convolution on a user-defined N by N window, where the center pixel value is replaced with the mean of all pixels within the window. A mean filter was executed using the `ndimage.uniform_filter` module of the SciPy Python library (Jones et al., 2001). Similar to the mean smoothing method, median filtering is executed by replacing the center pixel value of an N by N window with the median of all pixels within the window. Unlike mean filters, median filters are minimally affected by outliers and are typically well-suited to remove salt-and-pepper type noise. Median filtering was executed using the `ndimage.median_filter` method of SciPy. Gaussian filtering is unique in that the scale of features smoothed is determined by a Gaussian kernel and it ensures causality. This means no spurious features are generated because any features at a coarse resolution must have a cause at finer resolutions, thus guaranteeing noise reduction as the resolution is coarsened (Koenderink, 1984; Passalacqua, Do Trung, et al., 2010). The Gaussian filter is defined as

$$h(x, y, \sigma) = h_o(x, y) * G(x, y; \sigma), \quad (8)$$

where h_o represents the unfiltered elevation at location (x, y) , $*$ represents the convolution operation, and $G(x, y; \sigma)$ represents the Gaussian kernel with standard deviation σ . The Gaussian kernel is defined as

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(x^2+y^2)}{2\sigma^2}\right], \quad (9)$$

where larger standard deviations result in coarser output landscapes (Passalacqua et al., 2015). In line with methods used by Lashermes et al. (2007), the standard deviation parameter was calculated to be one quarter of the smoothing widths. The wetland model applied a Gaussian filter using the `ndimage.gaussian_filter` method of SciPy.

Unlike the above filters, which smooths data equally in all directions, Perona-Malik filtering performs a nonlinear, anisotropic diffusion. The Perona-Malik filter applied here is based on the diffusion equation initially proposed by Perona and Malik (1990),

$$\partial_t h(x, y, t) = \nabla \cdot [c(x, y, t)\nabla h], \quad (10)$$

where $h(x, y, t)$ is the elevation at time t , c is the diffusion coefficient, and ∇ is the gradient operator. Eq. (10) is a configuration of the linear, isotropic diffusion equation (Koenderink, 1984), in which the diffusion coefficient is constant in space and time. The Perona-Malik implementation varies c in space and time in order to preserve feature edges to achieve preferential smoothing (Passalacqua, Do Trung, et al., 2010; Passalacqua, Tarolli, et al., 2010). While there are two possible forms of c , here we implemented

$$c = \frac{1}{1 + \left(\frac{|\nabla h|}{\lambda}\right)^2}, \quad (11)$$

where λ is the edge stopping threshold (Perona & Malik, 1990). We chose the form of c in Eq. (11) because it was found to result in more consistent degrees of smoothing when applied to natural and urban landscapes compared to results using the alternate edge stopping function (Sangireddy et al., 2016). In addition, λ was calculated to be the 90th percentile of the gradient (i.e., slope) distribution to provide a simple first estimate of feature edges based on elevation change, as proposed by Perona and Malik (1990) and implemented by Sangireddy et al. (2016) and Passalacqua, Do Trung, et al. (2010) for channel network extraction. The time of forward diffusion (t in Eq. (10)) controls the rate of smoothing in the Perona-Malik method, and a higher number of iterations results in coarser smoothing. However, unlike the other smoothing methods included in this study, this smoothing parameter has no unique and uniform equivalent spatial scale (Passalacqua, Do Trung, et al., 2010). We preliminarily set t to a value of 50 iterations, which has been shown to sufficiently remove small-scale variability from high-resolution DEMs for stream delineation (Hooshyar et al., 2016; Passalacqua, Do Trung, et al., 2010; Sangireddy et al., 2016). To execute Perona-Malik smoothing, code from the PyGeoNet nonlinear filtering module, `pyGeoNet_nonlinear_fitter.py`, was implemented into the wetland model. PyGeoNet is the Python implementation of GeoNet, an open source software for automatic channel network extraction using elevation input data (Passalacqua, Do Trung, et al., 2010; Sangireddy et al., 2016).

3.3.1.1.2 DEM Conditioning Methods

Hydrologic conditioning techniques are defined by their method to remove depressions to enforce downstream flow and connect flowpath grid cells (Woodrow et al., 2016). Comparisons of Fill, IRA, and A* conditioning techniques were included in this analysis for their common application and dependence solely on elevation data.

Fill is perhaps the most commonly used and widely implemented conditioning technique. However, it has been suggested that it is incompatible with LiDAR data due to the inherent assumption that depressions are erroneous data points, rather than reflective of true surface features (Rieger, 1998; Woodrow et al., 2016). Fill removes depressions by adjusting the elevation of a depression pixel to match the elevation of the surrounding pixels (Jenson & Domingue, 1988; Planchon & Darboux, 2002; Wang & Liu, 2007). Fill was executed in the wetland model using TauDEM (Tarboton & Ames, 2001; Tesfa et al., 2011), which allowed for parallelization of the computations.

Although Fill has been used to preprocess LiDAR DEMs within hydrologic workflows (e.g., Hooshyar et al., 2016; O'Neil et al., 2018; Richardson et al., 2009), more advanced techniques have become popular, such as the IRA method. Depending on which method has the least impact on the DEM, IRA addresses depressions by either filling or breaching, which lowers pixels adjacent to depression pixels to carve channels out of sinks and through obstacles (Lindsay & Creed, 2005). The IRA approach was implemented using the GRASS GIS `r.hydrodem` module (GRASS Development Team, 2017; Lindsay & Creed, 2005).

The A* least-cost path algorithm (Hart et al., 1968) offers an alternative to modifying elevation data by determining the least-cost drainage paths through unaltered terrain and out of sinks (Metz et al., 2011). A* handles pixels draining to depressions by routing flow along the steepest downhill slope to the bottom of the depression and then continuing along the least steep

uphill slope (Metz et al., 2011). The A* conditioning method was executed using the GRASS GIS *r.watershed* module (GRASS Development Team, 2017; Metz et al., 2011).

3.3.1.2 Input Variable Calculation

Previous development and implementation of the wetland identification model, which included the study areas used here, concluded that curvature, TWI and DTW are useful topographic metrics for RF wetland identification (O’Neil et al., 2018). It is important to note that in this workflow, the DTW and curvature grids were affected only by the smoothing operation, whereas TWI grids were affected by both the smoothing and conditioning operations. While it would have been possible to derive all input variables from DEMs subject to both operations, we strived to alter the LiDAR surface as little as possible. Following the calculation of the curvature, TWI, and DTW grids, the input variables were merged into a multiband grid, where each band stores data for a single input variable, using the GDAL *gdal_merge.py* module (GDAL/OGR contributors, 2019).

Curvature can be used to describe the degree of convergence and acceleration of flow (Moore et al., 1991), making it a useful indicator of saturated and channelized areas (Ågren et al., 2014; Hogg & Todd, 2007; Kloiber et al., 2015; Millard & Richardson, 2015; O’Neil et al., 2018; Sangireddy et al., 2016). We use laplacian curvature, defined as the second derivative of the elevation grid. Laplacian curvature has been shown to assign a higher value of positive curvature to more convergent features, leading it to favor extraction of natural channels rather than artificial drainage paths (Passalacqua et al., 2012). In addition, Passalacqua et al. (2012) found that compared to geometric curvature, laplacian curvature more effectively identified channels in flat and human-impacted landscapes, which can describe our study sites that all encompass corridor projects. In the wetland model, curvature was calculated from the smoothed DEM using code adopted from PyGeoNet, which utilizes NumPy operations (Oliphant, 2007).

TWI has been successfully used to map saturated areas (Ågren et al., 2014; Lang et al., 2013; Millard & Richardson, 2015; Murphy et al., 2009; O’Neil et al., 2018). Developed by Beven and Kirkby (1979), TWI relates the tendency of an area to receive water to its tendency to drain water, defined as

$$TWI = \ln\left(\frac{\alpha}{\tan \beta}\right), \quad (12)$$

where α is the specific catchment area (contributing area per unit contour length) and $\tan(\beta)$ is the local slope. The TWI was calculated two ways depending on the conditioning method used. For DEMs conditioned by Fill or IRA, TauDEM D-Infinity methods were used (Tarboton, 1997b), with the slope parameter calculated using NumPy. Alternatively, for DEMs conditioned using A*, a TWI grid was output directly from the same *r.watershed* program of GRASS GIS. This method used the multiple flow direction algorithm (Holmgren, 1994) and a GRASS GIS-calculated slope.

The DTW has been shown to accurately indicate saturated areas as well (e.g., Murphy et al., 2007, 2009, 2011; O’Neil et al., 2018; Oltean et al., 2016; White et al., 2012)(Murphy et al., 2007, 2009, 2011; O’Neil et al., 2018; Oltean et al., 2016; White et al., 2012) . The DTW, developed by Murphy et al. (2007), is a soil moisture index based on the assumption that soils closer to surface water, in terms of distance and elevation, are more likely to be saturated. When calculated for a grid, the DTW is defined as

$$DTW (m) = \left[\sum \left(\frac{dz_i}{dx_i} \right) a \right] * x_p, \quad (13)$$

where $\frac{dz}{dx}$ is the downward slope of pixel i , calculated along the least-cost (i.e., slope) path to the nearest surface water pixel, a is either 1 or $\sqrt{2}$ depending on parallel or diagonal paths across pixel boundaries, and x_p is the pixel resolution (Murphy et al., 2007). DTW calculation requires a slope grid to represent cost and a surface water grid to represent the source from which to calculate distance. Although national-scale streamline data, the National Hydrography Dataset (NHD), exists for the study sites, these data are generated at relatively coarser resolutions (1:12,000-1:24,000 scales) (USGS, 2019b). Instead, the surface water grid was generated using PyGeoNet (Passalacqua, Do Trung, et al., 2010; “PyGeoNet,” 2019; Sangireddy et al., 2016). PyGeoNet employs a statistical analysis of curvature, and geodesic minimization principles to extract channel networks from elevation data (Passalacqua, Do Trung, et al., 2010; Sangireddy et al., 2016). Visual analyses based on aerial imagery were performed to compare the accuracy of PyGeoNet streams, NHD streams, and streams generated using the flow initiation threshold method (Band, 1986; O’Callaghan & Mark, 1984; Tarboton, 1991). These analyses showed that PyGeoNet channels aligned with aerial imagery better than NHD streams and resulted in less overestimation of streams in developed areas compared to implementing the flow initiation threshold method with several accumulation area thresholds. We found that using parameters suggested for engineered landscapes (see Sangireddy et al., 2016) produced accurate results across all study sites. The DTW grid was created using the GRASS GIS *r.cost* module (GRASS Development Team, 2017).

3.3.1.3 Classification and Accuracy Assessment

The classification and accuracy assessment workflow involved splitting the verification dataset into training and testing subsets, initializing a RF model, training the model, performing the classification, and then an accuracy assessment. As shown in Table 7, the verification distributions of wetland and nonwetland area in the study sites can be considered slightly imbalanced (Site 4) or highly imbalanced (sites 1, 2, and 3). Imbalanced datasets can be problematic for RF models, because these models aim to minimize the overall error rate, resulting in more predictions of the majority (i.e., nonwetland) class and fewer predictions of the minority (i.e., wetland) class (Branco et al., 2016; Chen et al., 2004; Zhu & Pierskalla, 2016). Addressing this issue is nontrivial and we tested two proposed methods to improve minority class detection prior to generating final results: undersampling the majority class when creating training data and increasing the minority class weight. The Scikit-learn Python library (Pedregosa et al., 2011) was used to execute this workflow segment.

3.3.1.3.1 Training and Testing Data Creation

Creating greater balance between training classes has been shown to be an effective solution for imbalance-related prediction issues (Batuwita & Palade, 2010; Branco et al., 2016; Estabrooks et al., 2004; Fernández et al., 2008, 2010). The effect of training data characteristics has been explored for wetland classification applications by Millard and Richardson (2015), who found that wetland models performed best when training class proportions reflected the true land cover proportions. To test the effect of this method on model accuracy, all preprocessing combinations were classified using the training sampling scheme suggested by Millard and Richardson (2015). Of these results, the model achieving the highest accuracy was used to perform

classification tests where the nonwetland training data size was reduced by varying extents. Final results for all other preprocessing combinations were then obtained by applying the training class proportions that resulted in the highest accuracies. For each analysis, the subset of verification data remaining after training data separation became the testing dataset used for accuracy assessment. To conduct this testing, a Python module using Numpy array masking methods and random indices selection was written, which allowed user-defined fractions of verification wetland and nonwetland pixels to be selected for training.

3.3.1.3.2 *RF Classifier*

For each model iteration, a RF model was initialized given a set of user-defined parameters, including class weights. The weighted RF method has been proposed to combat imbalance issues, as this method entails assigning custom weights to classes that modify the penalty for misidentifying that class (Chen et al., 2004; Zhu & Pierskalla, 2016). Zhu and Pierskalla (2016) used class weights to avoid favoring majority class predictions for their imbalanced RF classification of karst sinkholes. They found that the best results were produced by weighing the positive, minority class four times higher than the negative, majority class. We tested the efficacy of applying these class weights, as well as a series of more severely deviating weights, for tuning the RF model for the imbalanced datasets. For these analyses, training class proportions were held constant at 15% of verification wetlands and 15% of verification nonwetlands sampled for training. Other RF model parameters included the number of trees and maximum tree depth. We used 300 trees for all models, as suggested by Zhu and Pierskalla (2016), who found that this number was sufficient to stabilize errors. The maximum tree depth was set to “None,” which expands nodes until all leaves are pure (Scikit-learn Developers, 2017a). Additionally, a fixed random state was used to obtain a deterministic behavior during training across all model runs. All other parameters were left at their default setting.

After initializing the RF model, the training dataset and corresponding merged input variable pixels were used to build the forest of trees. This trained model was subsequently used to classify the remaining input variable pixels, resulting in binary wetland/nonwetland predictions, i.e., the hard classification. The trained model was also used to output the probabilities of each pixel belonging to the wetland class. While pixels with probabilities greater than 50% for either class correspond to the hard classification output, this continuous range of class probabilities can provide valuable information about model performance and allow users to vary the decision threshold for classifications based on the intended application and the user-defined balance between detection and overprediction. The RF classification also output variable importance measures, defined as the mean decrease in accuracy resulting from the omission of variables. The hard classification, wetland class probabilities, and importance measures were used for model analysis and accuracy assessment. The Scikit-learn ensemble.RandomForestClassifier module (Scikit-learn Developers, 2017a) was used for the RF classification.

3.3.1.3.3 *Accuracy Assessment*

Accuracy metrics were selected considering that true positive (i.e., wetland) predictions should be rewarded more heavily than true negative (i.e., nonwetland) predictions for the intended environmental planning and permitting application, and the varying degrees of class imbalance among the study sites. Model performance was evaluated using confusion matrices, wetland recall

and wetland precision (referred to as recall and precision), precision recall (PR) curves, and receiver operating characteristic (ROC) curves. The sklearn.metrics module was used to calculate these accuracy metrics (Scikit-learn Developers, 2017b).

Recall and precision are common metrics used to compare model performance between sites. Recall, also known as the true positive rate, represents the proportion of true wetlands that were identified and is defined as

$$\text{Recall} = \frac{\text{True wetland predictions}}{\text{Total true wetlands}}. \quad (14)$$

Considering the emphasis on the minority wetland class, recall can be considered the priority indicator of model performance, a practice supported by statistical literature on imbalanced class evaluation (Branco et al., 2016; Chen et al., 2004; Sun et al., 2007). To account for model overprediction, we chose precision because, unlike the commonly used specificity (or, true negative rate), it is not biased by large numbers of true negative instances. For this reason, precision is considered more representative for imbalanced scenarios (Branco et al., 2016; Sun et al., 2007). Precision represents the proportion of correct wetland predictions and is defined as

$$\text{Precision} = \frac{\text{True wetland predictions}}{\text{Total wetland predictions}}. \quad (15)$$

Precision can account for model overprediction because, unlike the commonly used specificity (or, true negative rate), it is not biased by large numbers of true negative instances. For this reason, precision is considered more representative for imbalanced scenarios (Branco et al., 2016; Sun et al., 2007).

PR curves and ROC curves were used to summarize model performance and improvement within individual sites. In cases like Site 4, where there is less class imbalance, false positive rate is an adequate metric to account for model overprediction (Branco et al., 2016). For this reason, the ROC curve was used here, which plots recall versus false positive rate for each predictive threshold of a class. The area under the ROC curve (AUROC) was used to summarize Site 4 models. The baseline of AUROC values is 0.5, representing a random classifier; the closer AUROC values are to 1, the better a model is at distinguishing between two classes (Branco et al., 2016). For the highly imbalanced sites 1, 2, and 3, PR curves were used instead. PR curves and the area under PR curves are commonly used to summarize the performance of models where the positive class is the minority class (Davis & Goadrich, 2006; Keilwagen et al., 2014). PR curves plot precision versus recall for each predictive threshold of a class. The baseline of a PR curve is represented by the horizontal line equal to the true percentage of positive classes, and an area under a PR curve closer to 1 indicates a better performing model. However, the standard area under curve calculation has been shown to provide overly-optimistic measures from PR curves (Davis & Goadrich, 2006). Instead, we use the Average Precision (AP) score, which is strongly correlated to the area under PR curves (Aslam et al., 2005). AP is defined as

$$AP = \sum_n (R_n - R_{n-1})P_n, \quad (16)$$

where P_n and R_n are the precision and recall at the n^{th} threshold.

We found these metrics to be more suitable for this study than commonly used options, such as overall accuracy, Kappa statistic, and Matthews Correlation Coefficient (MCC). When using overall accuracy, the impact of the rare class is lower than that of the majority class (Branco et al., 2016; Chen et al., 2004), allowing a wetland model predicting all nonwetland instances to appear very accurate. The Kappa statistic is highly dependent on sample size, and can increase as

the proportion of wetlands to non-wetlands increases, even if recall decreases (Ali et al., 2014; Byrt et al., 1993). Overall accuracy and the Kappa statistic have been omitted from similar studies for these reasons (e.g., Ali et al., 2014; Zhu & Pierskalla, 2016). Lastly, the MCC metric has been shown to be suitable for imbalanced scenarios (e.g., Boughorbel et al., 2017), however its calculation includes number of true negative samples. Testing the MCC result for three trials of sites 1, 2 and 3 that achieved the same recall and precision, we found that MCC scores varied likely due to differences in wetland to nonwetland ratios.

3.4 Results

3.4.1 Effects of Preprocessing Techniques on Model Performance

Figure 14 shows the precision and recall for each combination of smoothing and conditioning (15 trials for each study site). Note that for these results, the same smoothing parameters were applied for all inputs. There was a large difference in accuracy between model results in sites 1, 2, and 3 compared to those in Site 4. In sites 1, 2, and 3, the majority of testing wetlands were identified, represented by high recall, but a minority of the wetland predictions were correct, represented by low precision. Even though these models were prone to overprediction, which is a less costly error than underprediction for wetland permitting, their high rate of wetland detection would make them useful as preliminary tools for subsequent manual investigation. In contrast, model results for Site 4 had a relatively higher precision and lower recall, reflecting fewer wetland predictions, which were also mostly incorrect. Furthermore, there were no significant improvements Site 4 when increasing the proportion of verification data used for training, further suggesting the topographic metrics and the applied preprocessing methods cannot sufficiently distinguish wetlands in this landscape.

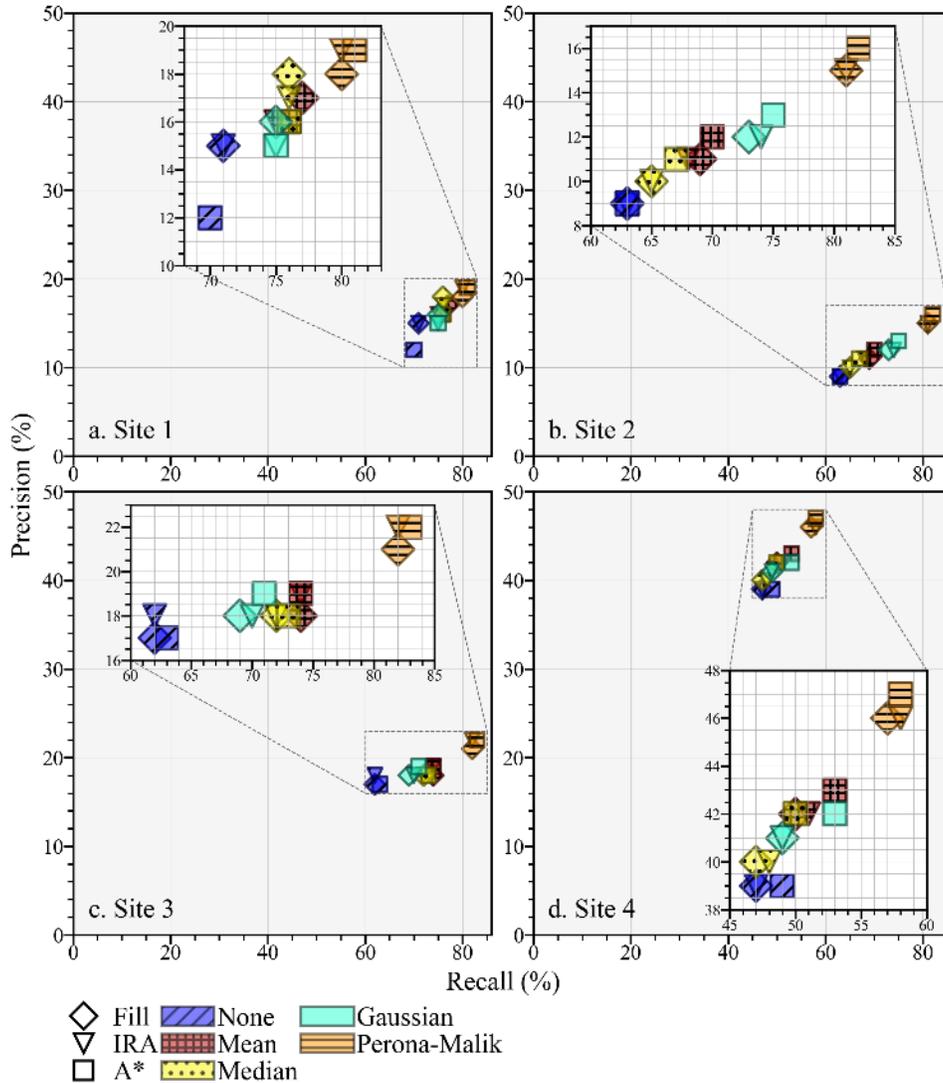


Figure 14. Wetland precision and recall resulting from each preprocessing technique combination across all study sites. Note the differences in x-scale and y-scale range.

Common trends in model performance due to smoothing and conditioning emerged despite differences in the accuracies. As seen in Figure 14, results were more consistently grouped by smoothing method than conditioning method for all sites, indicating that smoothing had a more significant impact on the wetland model. The highest precision and recall scores were achieved by the Perona-Malik and A* combination for all sites. No filtering and Fill resulted in the lowest precision and recall scores for all sites, except Site 1, where no filtering and A* resulted in the lowest scores. For sites 1, 3, and 4 the DTW was the most important variable in the best performing models. For Site 2, the most important variable was the DTW in the worst performing model and the TWI in the best performing model. The changes in variable importance due to preprocessing technique combinations are depicted in Figure A1 in the Appendix.

For sites 1, 2 and 3, all models using no filter produced the overall lowest precision and recall scores, and in Site 4 these models resulted in the lowest precision and among the lowest recall (Figure 14). Visual analyses showed that models resulting from unsmoothed DEMs had the

largest distribution of scattered false wetland predictions, many of which were located in impervious areas. Conversely, models incorporating the Perona-Malik filter achieved the highest precision and recall scores in all study sites. The Perona-Malik smoothing resulted in considerable removal of scattered wetland predictions and false positives surrounding developed areas. Perona-Malik smoothing also best represented natural drainage patterns, as demonstrated by increased wetland predictions within true wetland extents. Other smoothing methods resulted in somewhat similar performance in terms of recall and precision with the exception of Site 2, for which there was a clear difference between the filtering techniques (Figure 14). Mean, median, and Gaussian smoothing consistently reduced scattered false wetland pixels and better represented wetlands in natural areas, relative to unsmoothed models. However, median smoothing was noticeably less effective in doing so in vegetated areas. Gaussian and mean smoothing results were typically very similar in all land types. It was unexpected that Gaussian smoothing did not consistently outperform the relatively simpler mean and median methods since the Gaussian method guarantees causality. Additionally, an example of the effect of smoothing methods on curvature derivation for a wetland transect can be seen in Figure A2.

Models incorporating the A* technique and those using Fill consistently resulted in the highest and lowest accuracies within groups of common smoothing, respectively (Figure 14). Visual analyses showed that in developed areas, Fill created larger areal false wetlands along roads whereas IRA and A* methods resulted in smaller false positives in more linear patterns. In vegetated areas, Fill conditioning resulted in the largest distribution of scattered false wetlands within local depressions and A* conditioning the smallest. Moreover, flow routing for DEMs conditioned by the IRA method required 5+ hours when running on 20 cores on high performance computing resources, whereas this step for filled DEMs required less than one hour using the same resources. This substantial increase in computational cost did not correspond to notable differences in prediction accuracy (Figure 14). In contrast, generating the A* outputs required less than one hour on a desktop computer with no parallelization. Lastly, it is important to note that improved implementations of the traditional Fill algorithm have been recently proposed (e.g., Barnes et al., 2014), and this may perform better than the traditional method examined here. An example of the effect of conditioning on TWI calculation for a wetland transect is also provided in Figure A3.

3.4.2 Characteristics of the Tuned RF Model

Undersampling the majority class for training data selection improved wetland prediction accuracy more notably than adjusting the class weights (Figure A4). Increasing the wetland class weight while maintaining a nonwetland class weight of one resulted in small accuracy changes and did not consistently lead to improved wetland detection. This was also true when applying wetland to nonwetland weight ratios of 4:1, as recommended by Zhu and Pierskalla (2016), and when setting the wetland class weight as high as 1,000 (trial not shown in A4a). For that reason, the class weights parameter was set to “balanced,” which automatically adjusted weights to be inversely proportional to the class distribution (Scikit-learn Developers, 2017a); however, small changes in model results were observed when compared to equal class weights of one. Conversely, varying the ratio of training wetlands to training nonwetlands greatly affected precision and recall. As expected, precision decreased and recall increased as less nonwetlands were sampled for training, but with varying tradeoffs. Our testing consisted of sampling fewer nonwetlands until the

loss in precision outweighed the gain in recall. Sampling equal percentages from both classes, as proposed by Millard and Richardson (2015), did not result in levels of recall that are acceptable for wetland permitting. For the highly imbalanced sites, the best training dataset consisted of 15% of surveyed wetlands and only 1% of surveyed nonwetlands. The model performance for the slightly imbalanced Site 4 was very poor when sampling as little as 5% of nonwetlands (trial not shown in A4b), so it was necessary to test less severe undersampling schemes. Site 4 model results still improved due to less severe majority class undersampling, with the best performing training set consisting of 15% of surveyed wetlands and 8% of surveyed nonwetlands. Furthermore, we tested the effect of increasing the overall training data quantity while maintaining best performing sampling ratios, and found that there were no notable benefits to model performance.

3.5 Discussion

3.5.1 Varying the Smoothing Scale and Method by Input Variable

Results showed that smoothing had a larger impact on model performance than conditioning for all sites. This is likely due, in part, to the fact that DEM smoothing was included in the calculation of all input variables whereas DEM conditioning was only required for the TWI calculation. In addition to this, smoothing has been shown to impact the scale of hydrologic patterns captured, as modeled soil moisture distributions and groundwater table gradients depend on the level of detail of topographic variations (Burt & Butcher, 1986; Rodhe & Seibert, 1999; Seibert et al., 1997; R. Sørensen et al., 2006; Zinko et al., 2005), and both smoothing method and scale are important. While the smoothing method determines the distinction between features of interest and noisy data, the smoothing scale determines the scale of these features. By extension, the best smoothing scale and method may vary by input variable as they each capture unique hydrologic characteristics. To further explore the effect of smoothing on wetland identification, we performed additional analyses where input variables were derived from DEMs with a range of smoothing methods and scales applied. Classifications were executed for each input variable derived from the individualized smoothing schemes (“single input models”). Input variables used in the best performing single input models were merged into a three-band grid and classified (“wetland model”), following our proposed approach. For mean, median, and Gaussian smoothing, we tested 2m, 10m, 25m, 50m, and 100m smoothing scales, as done in studies evaluating TWI and DTW for wet soil mapping (Ågren et al., 2014; Murphy et al., 2011). For the Perona-Malik method, 20 and 100 iterations were tested, similar to analyses performed by Passalacqua, Tarolli, et al. (2010) for channel extraction. Single input models were compared first by precision and recall and then by AP score (sites 1, 2, and 3) or AUROC score (Site 4) if needed (Figure 15). A* conditioning was applied to all TWI models.

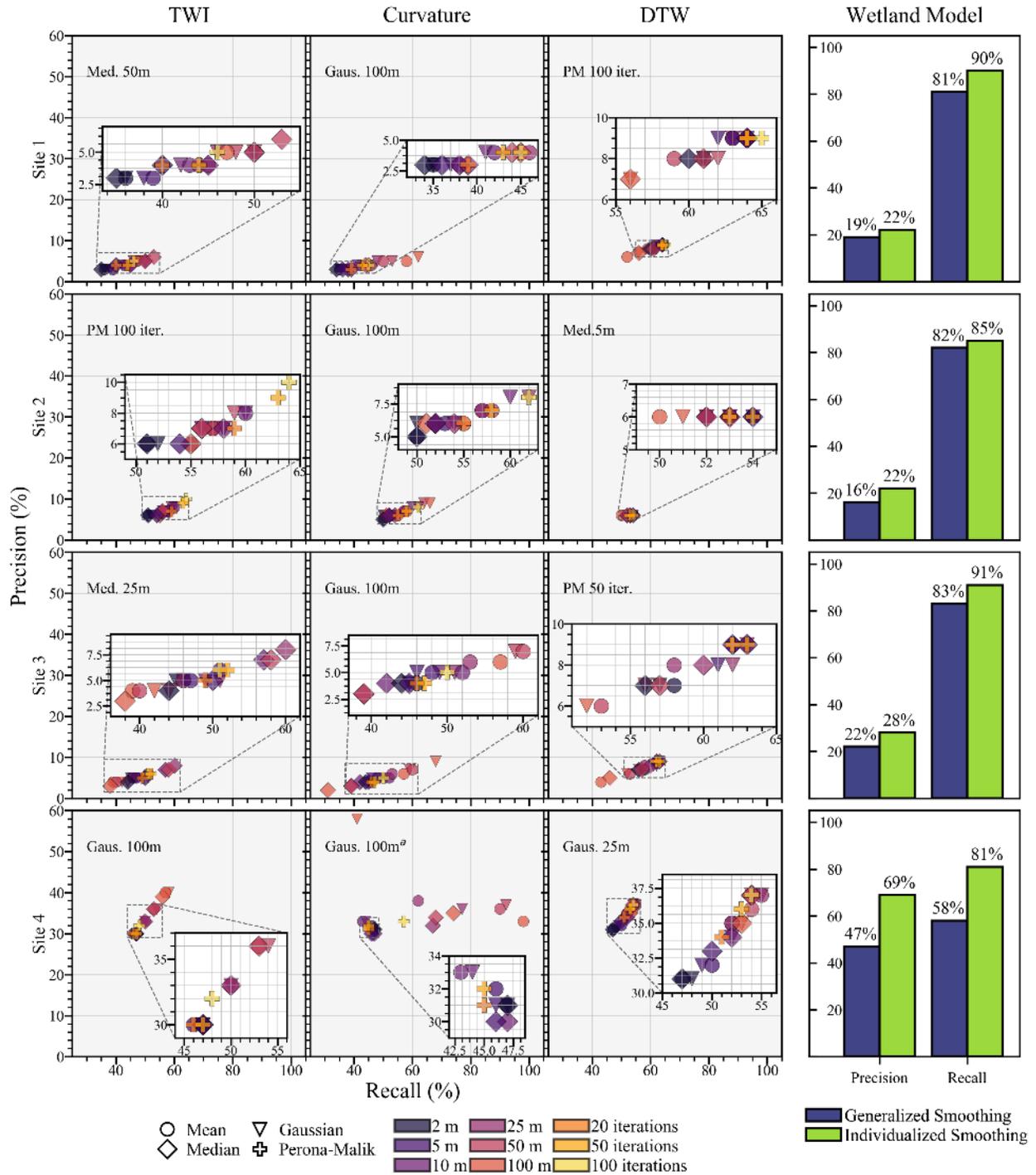


Figure 15. Effect of varying smoothing method and scale on wetland model accuracy. Scatter plots show the results for models trained on a single input, and annotations indicate the best performing smoothing formulation for that input. Bar plots show the results of wetland models (i.e., trained on three inputs) when applying the individualized smoothing formulation vs. the smoothing formulation generalized across all inputs. Note the differences in x-scale and y-scale range. ^aGaussian 100m, Gaussian 50m, and Mean 100m were considered in determining the best performing curvature formulation for Site 4.

For all sites, varying the smoothing scale and method affected the accuracy of input variables and applying the best performing individualized smoothing scheme improved the wetland model performance. While we can gain insight from the trends depicted in Figure 15, it is important to note that relatively small accuracy margins separated results in many cases, and determination of the best performing models was based on differences of AP scores as low as 0.002 and AUROC score as low as 0.02. It would be useful to expand the testing performed here with additional study sites and repeated trials to more clearly establish best performing smoothing formulations for each input variable by landscape.

The best performing TWI smoothing method varied across sites, but coarser smoothing scales generally performed better than finer-scale models, with the exception of Site 3. According to the literature, this is likely because the TWI is effective in modeling saturation correlated to groundwater table gradients, which are better described by macrotopographic patterns (Ågren et al., 2014; Grabs et al., 2009; Murphy et al., 2009, 2011; Rasmus Sørensen & Seibert, 2007). However, the fluvial landscape in Site 3 required finer-scale indications of flow accumulation and convergence to capture riverine wetlands and riparian corridors. TWI models for Site 4 that incorporated Perona-Malik smoothing resulted in the lowest accuracies regardless of the number of iterations (i.e., rate of smoothing) used. This suggests that in the very flat study site, wetlands are characterized by gradually sloping and diffuse boundaries rather than sharper ones that would be estimated by the Perona-Malik method.

Similar to TWI, curvature models typically improved as scales became coarser. In addition, for all sites the best performing smoothing formulation was Gaussian at a 100m scale. In determining the best performing curvature model in Site 4, we considered Gaussian 50m and mean 100m, which resulted in the highest recall, and Gaussian 100m, which resulted in the highest precision. Because none of these formulations resulted in both the highest precision and recall, and because precision in Site 4 can be considered more important relative to other sites due to greater class balance, Gaussian 100m was chosen as it resulted in the highest AUROC score. The high accuracies for curvature models using Gaussian 100m shows that curvature was consistently more successful in identifying wetland depressions when coarser smoothing allowed smaller depressions such as roadsides and culverts to be degraded. It is also possible that larger Gaussian kernels would have further improved models in some of the sites. Curvature also became the most important variable in sites 3 and 4, rather than the DTW. Rank of the most important variables did not change in sites 1 and 2.

DTW models in sites 1, 2 and 3 followed an opposite trend in which accuracy generally increased as smoothing scale became finer. This is likely because the DTW has been found to be scale invariant and therefore use detailed topographic information to capture riparian wetted areas (Ågren et al., 2014; Murphy et al., 2009, 2011). In Site 4, finer-scale smoothing applied to the DTW tended to result in lower accuracy than coarser scales. This may reflect the higher distribution of large depression wetlands in the area, which are better represented by gradual slope gradients rather than those modeled by microtopography. DTW models filtered by the Perona-Malik with 50 iterations (i.e., the best performing generalized smoothing scheme) resulted in high accuracy for all sites. This indicates that this Perona Malik formulation is effective for DTW calculations for a range of landscapes, and that changes to DTW smoothing schemes had little effect on complete wetland model improvements.

3.5.2 Improvements to Wetland Predictions Due to Preprocessing Schemes

3.5.2.1 Applying the Best Performing Generalized Scheme

Between the worst and best performing generalized preprocessing schemes, as described in Section 4.1, AP scores (sites 1-3) and the AUROC score (Site 4) increased by 0.16 in Site 1, 0.18 in Site 2, 0.07 in Site 3, and 0.09 in Site 4 (see Figure 16 curves). The improvements from the worst performing models (Figure 16, a1-a4) are likely due to the ability of the Perona-Malik filter to enhance feature edges, allowing for more distinct transitions between converging and diverging areas. This feature resulted in higher wetland probabilities within surveyed wetland boundaries and abrupt transitions between high and low probability areas (Figure 16, b1-b4). For Site 2 there was a drastic decrease in wetland likelihood within impervious areas compared to the worst performing model (Figure 16, b2 vs. a2). No filter and A* conditioning did not result in a similar model output for Site 2, showing that the reduction of convergent areas detected on roadways was a product of the Perona-Malik filtering. Improvements between the best and worst generalized preprocessing methods were relatively subtle in Site 4 (Figure 16, b4 vs. a4). Despite slightly more accurate wetland predictions, the persistent random dispersion of probabilities point to an inability to identify wetlands among the mild slopes and complex subsurface of Site 4 when preprocessed using a generalized Perona-Malik smoothing and A* conditioning.

3.5.2.2 Applying an Individualized Scheme

Across all sites, the wetland model further improved as a result of individualizing the smoothing technique and scale to each input variable. Performance curves given in Figure 16 show that the AP scores increased in sites 1-3 (+0.13, +0.11, and +0.11, respectively) and the AUROC score increased in Site 4 (+0.18) relative to the best performing generalized models (Figure 16, b1-b4). Individualized smoothing in Site 1 and Site 3 resulted in fewer instances of hydrologic paths surrounding true wetland boundaries contributing to overprediction (Figure 16, c1 and c3). In Site 1, this is likely due to deriving the TWI grid from coarser median smoothing (50m scale), which degraded smaller slope variations and removed salt-and-pepper noise. In Site 3, deriving the curvature grid from coarser Gaussian smoothing (100m) likely highlighted wider and general channelized areas that more robustly encompassed true wetlands. In Site 2, individualized smoothing improved the model by eliminating flow accumulation in developed areas (Figure 16, c2 vs. b2). The coarser curvature (Gaussian 100m) likely contributed to filtering out narrow, convergent zones surrounding roadways and thereby decreasing overprediction. Applying individualized smoothing resulted in the greatest accuracy improvement in Site 4 despite the more complex subsurface of the area. TWI and DTW contributions to the improvements in Site 4 can be summarized as generalized slope patterns modeled by coarse, Gaussian smoothing that better represented hydraulic gradients that contribute to wetland formation (Figure 16, c4 vs. b4). It is clear that the most significant contributions to the complete wetland model resulted from the individualized curvature smoothing formulation. The improved wetland detection due to the curvature grid suggests the wetlands in the study site are well represented by large, isolated intrusions into the groundwater table. Overall, the consistent improvements to the wetland models due to individualizing smoothing suggest it would be useful to expand the testing performed here with additional study sites and trials to more clearly establish best performing smoothing formulations for each input variable by landscape. Additional scenes from these improved wetland

models are provided in figures A5-A8 and corresponding confusion matrices are given in tables A1-A4.

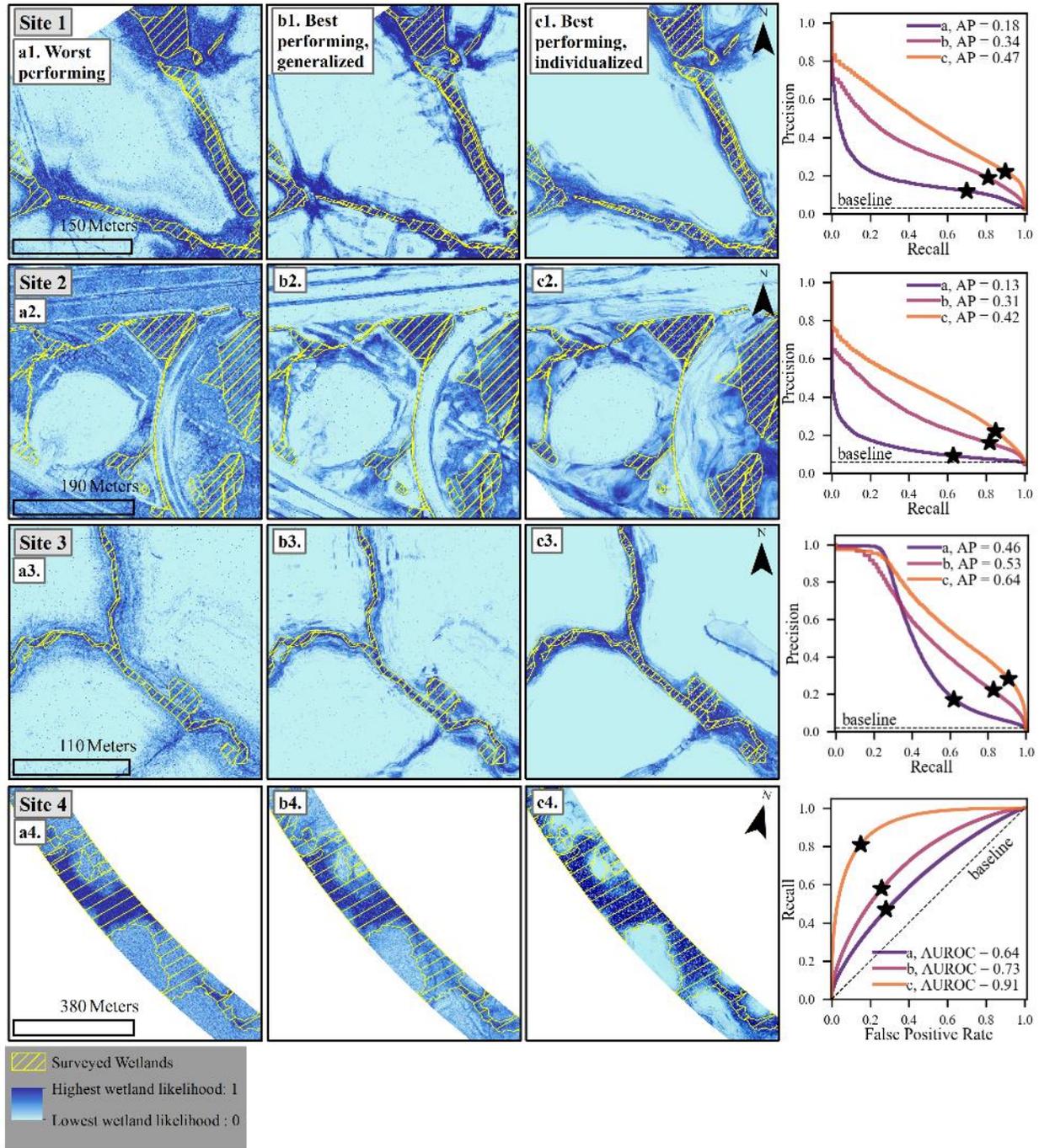


Figure 16. Wetland likelihoods resulting from different preprocessing configurations: worst generalized preprocessing, as described in Section 4.1 (a), best performing generalized preprocessing, as described in section 4.1 (b), and A* conditioning and best performing individualized smoothing, as described in section 5.1 (c). PR curves and ROC curve are shown to the right, with the accuracy for the hard classifications starred. Note the differences in results extents between panels are due to edge degradation caused by coarser smoothing scales.

3.5.3 Comparison to Earlier Wetland Model Implementations

As mentioned, the sites were previously studied using an earlier version of the wetland identification model (see O'Neil et al., 2018). The earlier model included Soil Survey Geographic Database (SSURGO) soil data (Soil Survey Staff, 2017b) in addition to TWI, curvature, and DTW. Soil data were omitted from this analysis to isolate the effects of DEM smoothing and conditioning techniques on the model accuracy. However, soil data were reintroduced where available to provide a comparison to the earlier wetland model where the input data per site are the same while the processing techniques and classification parameters differ. Following the procedure of O'Neil et al. (2018), input datasets were created that included relevant SSURGO layers and topographic input variables with best performing individualized preprocessing applied. For Site 1, incorporating soil data resulted in 38% precision and 96% recall, which were improvements from 22% precision and 92% recall using the earlier wetland model. For Site 2, the addition of soil data resulted in precision and recall scores of 34% and 92%, respectively. Compared to the earlier approach, this represents an improvement from 15% precision and a small decrease from 93% recall. For Site 3, where soil information was insufficient and therefore omitted in both wetland model versions, precision increased from 11% to 28% and recall increased from 87% to 91%. This comparison was not extended to Site 4 due to lack of overlap between verification data limits. The improvements in accuracy from the earlier model show that applying the more sophisticated terrain processing techniques resulted in higher quality wetland predictions that eliminated erroneous predictions while identifying more of the true wetlands, or only slightly fewer. In addition, model improvements in sites 1 and 2 show the ability of the polygonal, categorical soil information to describe soil characteristics relevant to wetland formation that are not captured by surface topographic patterns.

3.5.4 Approach Limitations

With the exception of Site 4, the wetland identification tool produced high wetland accuracy but relatively low precision (22-28%) when using only LiDAR-derived input variables. This low precision paired with high recall demonstrates the model configuration to identify convergent areas that are likely to become saturated, which will include wetlands as well as other areas with these characteristics. Although some of the overprediction occurred in concave, impervious areas, other predictions with consistently high wetland probabilities occurred in vegetated areas that surround surveyed wetlands, according to recent aerial imagery. It is possible that these overpredictions represent the diffuse boundaries of seasonally saturated areas while the surveyed wetlands, which were all delineated in summer months, were limited to areas saturated during most of the year. Topographic metrics are considered to be seasonally-averaged indicators of soil saturation, thus it is not surprising that models using these indices alone overpredicted wetlands according to surveys conducted during summer months. In addition, overpredictions surrounding developed structures or representing roadside ditches may be due to a lack of built drainage network representation. The current flow routing implementation does not reflect drainage through artificial structures. Including these flow paths by modifying the DEM near built drainage inlets would more realistically represent water accumulation in developed areas, thus reducing overprediction.

A shortcoming of the model common to all study sites was scattered, isolated wetland predictions, which is expected from a pixel-based classification. Pixel-based classifications do not

take increased wetland probability into account for adjacent similar classifications. Thus, the RF classification ignores that wetlands exist as distinct landscape units bound by geomorphic features. Although we found that including object-based soil data begins to address this issue, alternative techniques may allow the model to still rely solely on DEM data. For example, incorporating object-based image analysis (OBIA), where pixels are segmented into similar landscape groups prior to classification, may be useful. Many studies have demonstrated the ability of OBIA to address data heterogeneity and noise in wetland classifications (Dronova, 2015), and researchers have shown the benefits of applying OBIA specifically to DEM data (e.g., Kloiber et al., 2015; Richardson et al., 2009; Serran & Creed, 2016). Using deep learning networks, rather than RF, may also address this issue. Deep learning networks identify objects based on contextual spatial patterns and, although an emerging field (Zhang et al., 2016), they show promise for improving wetland identification from various remote sensing data (Liu et al., 2018; Ma et al., 2017; Rezaee et al., 2018).

While it is valuable to test the technical limits of LiDAR topography for its wide availability and high resolution, wetland predictions could be improved by incorporating additional remote sensing data. Multispectral data have been shown to be useful for determining vegetation extent optically and radar data have been used to identify water extent and flooded vegetation without being hindered by cloud cover (Guo et al., 2017). Researchers have demonstrated the ability of these data to identify wetlands in geographic regions where topographic information is less effective due to mild topographic variations and glacial or coastal influence (e.g., Allen et al., 2013; Behnamian et al., 2017; Corcoran et al., 2013; Kloiber et al., 2015; Millard & Richardson, 2013). Thus, a more robust set of wetland characteristics may be detected by including multispectral imagery and radar data to supplement the LiDAR topography used in this analysis. When these data become widely available at adequate resolutions, it would be valuable to incorporate them into our proposed framework to improve predictions while maintaining accessibility for environmental planning decision makers.

3.6 Conclusions

Accurate and widely-available wetland inventories are an important resource to aid wetland conservation and environmental planning. We outline an automated, open source wetland identification model that uses LiDAR DEM-derived topographic wetness index (TWI), curvature, and cartographic depth-to-water index (DTW) as input variables to a Random Forest (RF) model. The use of high-resolution DEMs allows for more detailed mapping of topographic features, but also requires more sophisticated smoothing and conditioning techniques. We tested the effects of smoothing (none, mean, median, Gaussian, and Perona-Malik) and conditioning (Fill, Impact Reduction Approach (IRA), and A* least-cost path analysis) techniques on our wetland model results for four sites in Virginia that encompass a range of topography, built environment, and ecoregions.

We conclude the following from our results.

1. For all sites, Perona-Malik smoothing followed by A* conditioning resulted in the best performing models, in terms of wetland precision and recall.

2. Applying Perona-Malik smoothing can enhance the input variable calculations in a way that wetland locations can be modeled.
3. The A* conditioning method can improve the accuracy of the TWI for wetland identification and decrease calculation runtime compared to Fill and IRA implementations.
4. The accuracy of wetland predictions improved considerably by individualizing smoothing method and scale to each input variable, most notably for a very flat site located in the Coastal Plain.
5. Without the data required to perform individualized smoothing testing for a new area, we recommend applying the generalized Perona-Malik smoothing scheme and A* conditioning as these methods greatly improved wetland identification for a range of landscapes.
6. Varying the training class distribution more effectively addressed wetland underprediction due to class imbalance, compared to varying class weights, and wetland accuracy improved for all sites by undersampling the nonwetland training class.

Using the individualized smoothing schemes and the best performing A* conditioning, our models resulted in high recall (81-91%) but lower precision (22-69%), and our proposed framework improved results compared to earlier wetland model implementations. These best performing models may not yet be adequate as definitive wetland delineation sources due to the low precision. However, recall can be considered more important than precision for wetland screening applications meant to guide subsequent field surveys. Wetland predictions produced by the current model would lead field surveyors to portions of most, if not all, wetlands, while saving resources by avoiding nonwetland areas. Thus, the proposed framework has strong potential to act as a preliminary screening tool based on its high rate of wetland detection.

Chapter 4

Deep Learning as a Wetland Classification Approach using LiDAR Topographic-Derived and NDVI Input Data

4.1 Introduction

Wetlands are important ecosystems that are threatened by development, climate change, and pollution (Klema, 2011). Wetland loss is both a global (Davidson, 2014) and national problem, as half of the wetlands of the conterminous U.S. have been lost since 1600 (Dahl et al., 1991). In the U.S., federal regulations, such as Section 404 of the Clean Water Act, play an important role in wetland protection. Laws require environmental impact assessments prior to land development and water resources projects, which entails the creation of detailed wetland surveys (Page & Wilcher, 1990). Conducting these surveys with the level of accuracy needed to abide by federal regulations and meet the goal of avoiding adverse impact to wetlands can be time-consuming and costly. To support these efforts, a method for reducing the total area that requires detailed field analyses is needed, where the most likely wetland locations can be identified accurately and efficiently. Although manual surveys are the most accurate method to map wetlands, there is potential for supporting these efforts by using remote sensing data and machine learning to identify wetland features at varying scales (Guo et al., 2017; Lang et al., 2013; Lang & McCarty, 2014).

Despite the many types of protected wetlands that exist, all wetlands can be identified by common features. These include the presence of hydrologic conditions that inundate the area, vegetation adapted for life in saturated soil conditions, and hydric soils (US Corps of Engineers, 1987). Researchers have demonstrated the ability to detect these features from multispectral imagery, radar, and Light Detection and Ranging (LiDAR) data (Guo et al., 2017). Multispectral imagery are the most commonly applied data in wetland studies (Guo et al., 2017; Klema, 2011). However, spectral variables alone may be unable to distinguish wetlands due to spectral confusions from reflectance and backscattering (Dronova, 2015; Kim et al., 2011). LiDAR data are well-suited to complement multispectral analyses due to its wide, and growing, availability and demonstrated benefit to wetland mapping (Guo et al., 2017; Klema, 2011; Kloiber et al., 2015; Lang & McCarty, 2014; Snyder & Lang, 2012). LiDAR returns can be interpolated to create high-resolution digital elevation models (DEMs), from which wetland indicators based on flow convergence and near-surface soil moisture can be derived (Lang et al., 2013; Lang & McCarty, 2014; Millard & Richardson, 2013, 2015; O'Neil et al., 2018, 2019). Moreover, researchers have shown the benefit of LiDAR DEM metrics as input variables to traditional machine learning techniques, such as Random Forest, for wetland mapping and classification (e.g., Deng et al., 2017; Kloiber et al., 2015; Millard & Richardson, 2013; Millard & Richardson, 2015; O'Neil et al., 2018, 2019; Zhu & Pierskalla, 2016).

The successful coupling of LiDAR and multispectral imagery with traditional machine learning techniques for wetland identification is well-documented. However, deep learning for remote sensing studies, including wetland identification, is an emerging field (Ma et al., 2017; Zhang et al., 2016) that shows promise for fulfilling the unmet need for wetland inventory creation. Deep learning architectures are modeled after the architecture of the mammal brain (Serre et al., 2007), where inputs are perceived and processed through multiple layers of abstraction. Convolutional neural networks (CNNs) (LeCun et al., 1998) are a representative form of deep learning that is used for visual recognition. CNNs utilize the spatial context of detected features to identify objects and classify scenes. The distinguishing element of CNN architectures are the convolutional layers, which convolve spatial filters over input images to identify patterns that are characteristic of target classes. Deep convolutional neural networks (DCNNs) (He et al., 2016; Krizhevsky et al., 2017; Simonyan & Zisserman, 2014) and fully convolutional neural networks (FCNs) (Long et al., 2015) are extensions of the CNN framework that can output dense pixel-wise classifications within images (i.e., semantic segmentation), where each pixel of the input image is assigned a class.

Since the formalization of the concept in 2006 (Hinton et al., 2006), deep learning has advanced the fields of speech recognition, medical diagnosis, and autonomous driving applications, and has recently motivated applications in remote sensing and water resources (Liu et al., 2018; Pan et al., 2019; Shen, 2018; Zhang et al., 2016). Researchers have shown the ability of DCNNs, FCNs, and other CNN extensions to delineate urban and natural landscape classes using multispectral imagery and elevation grids (Audebert et al., 2017, 2018), multispectral imagery and LiDAR point clouds (Xu et al., 2018), and multispectral imagery alone (Hu et al., 2018; Kemker, Gewali, et al., 2018; Kemker, Salvaggio, et al., 2018; Scott et al., 2017). Few researchers have applied DCNNs and FCNs specifically to wetland classification. For example, Liu et al. (2018) applied orthoimagery and elevation information to deep learning models for wetland segmentation. Also, Rezaee et al. (2018) used multispectral imagery in a wetland deep learning model, and posited that predictions would improve with the incorporation of physical information from radar or LiDAR sources.

The typical need for massive validation sets to train deep learning models is a significant deterrent to remote sensing and water resources researchers (Shen, 2018; Zhang et al., 2016), as reliable training data is often lacking in these applications. This issue is especially prevalent for wetland identification that is intended to inform conservation and permitting efforts, where training data for computational models are ideally manually derived and confirmed by regulatory bodies. Liu et al. (2018) has investigated this issue for wetland semantic segmentation, comparing results for a single study area using DCNNs, FCNs, Random Forest, and support vector machines, with privately contracted aerial imagery and surface elevation information as input features. While this is an important stride in gaining insight into the training data needs for deep learning of wetlands, an analysis has yet to be done that utilizes freely-available data and is completed over multiple geographic regions.

The growing research area of deep learning for remote sensing applications shows promise for advancing wetland mapping. Researchers have shown that wetlands can be identified at a high resolution using deep learning and that this method can be advantageous over traditional machine learning. However, research gaps remain pertaining to the potential for leveraging freely-available

data in these frameworks and the accuracy possible given typical training data resources. We aim to contribute to this field by presenting a novel wetland identification methodology that implements a basic semantic segmentation architecture and leverages freely-available remote sensing data. Our input data configuration consists of LiDAR DEM derivatives that describe geomorphologic and hydrologic contributors to wetland formation, as well as a commonly-used vegetative index. Using four study sites across Virginia, we build and evaluate several wetland models to demonstrate the potential for wetland semantic segmentation given typical training data resources. Mainly, we answer the following research questions:

- i. Across four geographic locations in Virginia, what wetland prediction accuracy is achievable by building site-specific models from typically-available amounts of training data?
- ii. What is the potential for combined-site model training using data from different geographic regions to create a single model able to predict wetlands across sites?

4.2 Methodology

4.2.1 Study Areas

Four study areas across Virginia, USA are used in this analysis (Figure 17a). Data for each study area include the extents of wetland surveys and the surrounding HUC 12 watershed (USGS, 2019b) (Figure 17b). The HUC 12 watersheds were used as processing extents and surveyed areas provided the validation data, also referred to as the study sites. The study areas span four level-III ecoregions. As shown in Table 8, the sites also vary by size, land cover, and topographic characteristics. Notable differences include the higher rate of development in sites 1 and 2, and the mild topography of Site 4. In addition, wetlands are much more abundant in Site 4, where the wetland to nonwetland ratio is 0.42, compared to less than 0.1 in the other sites. Note that all surveyed wetland types were merged into a single wetland category prior to use as verification data.

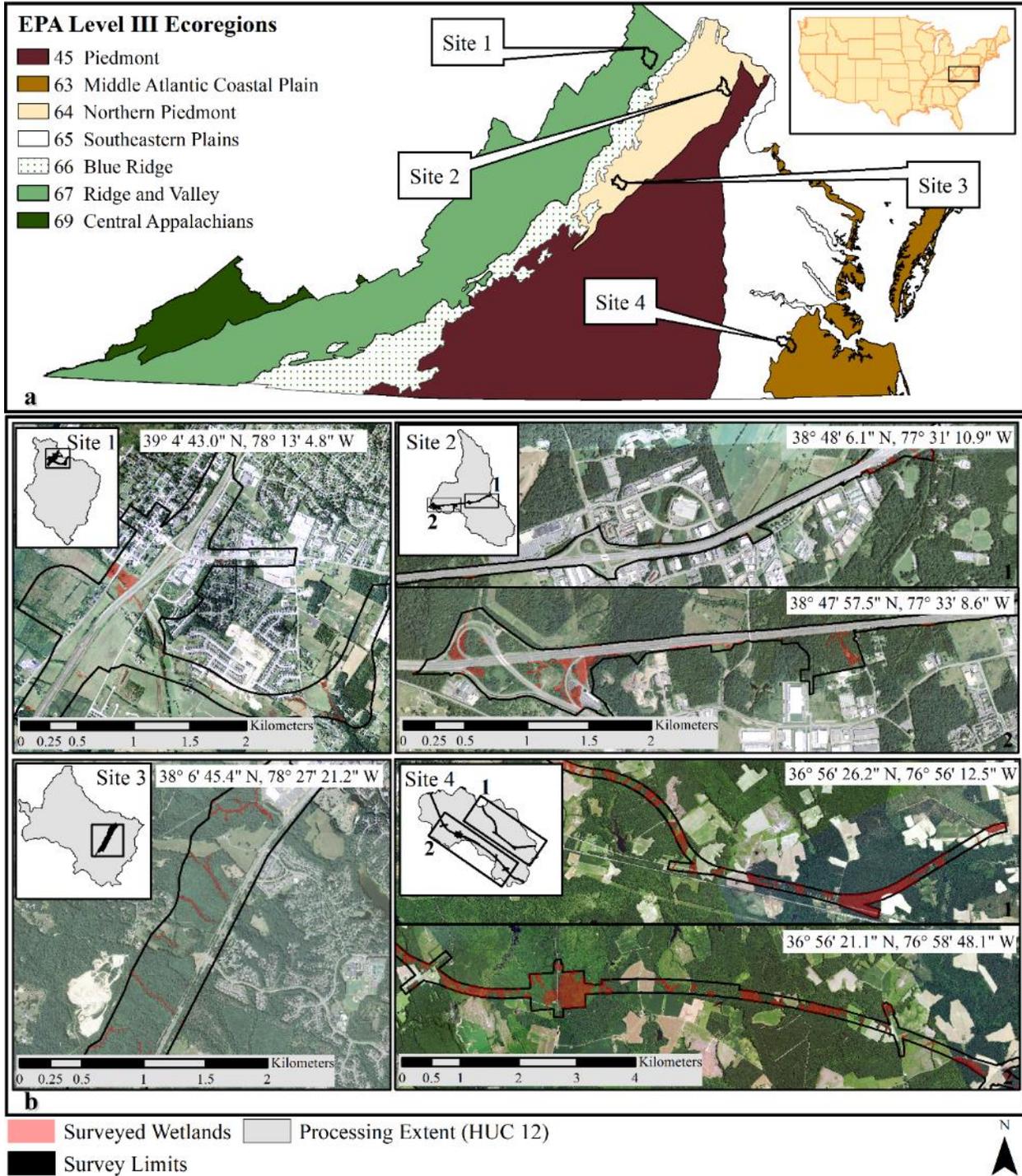


Figure 17. Four study areas spanning four level III ecoregions in Virginia, USA (a). Each study area includes the wetland survey limits, referred to as study sites, and the encompassing HUC 12 watershed, used as the processing extent (b). Reprinted from “Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model” by O’Neil et al., 2019, *Water Resources Research*, 55 (5), 4343-4363. Ecoregion data source: US EPA Office of Environmental Information. Aerial imagery data source: NAIP Digital Ortho Photo Image.

Table 8. Characteristics of each study site, including dominate land cover, topographic characteristics, and surveyed wetland distributions. Reprinted from “Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model” by O’Neil et al., 2019, *Water Resources Research*, 55 (5), 4343-4363.

	Site 1	Site 2	Site 3	Site 4
Dominating Land Cover ^a	Turf Grass (35%), Developed (22%), Cultivated (20%), Forested (19%)	Developed (36%), Turf Grass (31%), Forested (21%)	Forested (73%), Developed (9%), Cultivated (9%)	Forested (66%), Cultivated (18%), NWI Wetland (9%)
Verification Area (km ²)	2.8	1.6	1.8	5.6
Min. Elevation ^b (m)	209	46	101	10
Max. Elevation (m)	241	107	178	42
10 th Percentile Slope ^c (m/m)	0.02	0.01	0.04	0.01
90 th Percentile Slope ^c (m/m)	0.14	0.20	0.26	0.06
Mean Slope ^c (m/m)	0.07	0.08	0.14	0.03
Wetland : Nonwetland (m ² /m ²)	0.03	0.06	0.02	0.42
Dominating Cowardin Wetland Type(s) ^d	Palustrine Emergent (50%), Streams (20%) ^e	Palustrine Forested (44%), Palustrine Emergent (33%)	Palustrine Forested (56%), Streams (43%)	Palustrine Forested (88%), Palustrine Shrub (9%)

^a Source: Virginia Information Technologies Agency (VITA) Land Cover classifications (<https://www.vita.virginia.gov/integrated-services/vgin-geospatial-services/land-cover/>).

^b In sites 1, 2, and 4, verification area varied slightly due to edge effects of applying filtering to DEMs.

^c Slope information was calculated from LiDAR DEMs resampled to a 5 m resolution to reduce effect of raw DEM noise on slope information.

^d Values are approximate and according to VDOT wetland surveying reports.

^e Wetland type for remaining 30% of wetland area was not reported.

4.2.2 Input Data

This study used publicly available LiDAR DEMs, National Agriculture Imagery Program (NAIP) aerial imagery, and field-mapped wetland surveys. LiDAR DEMs were obtained from the Virginia Information Technologies Agency (VITA) (VITA, 2016) as hydro-flattened, bare-earth DEMs. The LiDAR data used were collected and processed between 2010 and 2015 and have horizontal resolutions ranging from 0.76 m to 1.5 m. NAIP imagery are provided by the United States Department of Agriculture (Farm Service Agency, 2017). NAIP imagery were used to derive the NDVI. NAIP imagery contain four spectral bands (red, green, blue, and near-infrared) at a 1 m spatial resolution. Imagery used in this study were collected near the dates of wetland surveying, and images were resampled to match the resolution of the LiDAR DEMs, if necessary. Wetland delineations and survey limits were provided by the Virginia Department of Transportation (VDOT) in polygon vector format and served as validation data for this study. All verification wetlands were manually surveyed during summer months (May – August) between 2013 and 2016 by professional wetland scientists in compliance with transportation planning permitting. Wetland delineations for sites 2, 3, and 4 were also jurisdictionally confirmed by the US Army Corps of Engineers (USACE). Binary wetland/nonwetland geotiffs were created from these data, with resolutions matching those of the site LiDAR DEMs. Visual analyses of Google Earth images showed that the study site landscapes changed minimally between LiDAR acquisition and wetland delineation timeframes.

4.2.3 Wetland Identification Method

The wetland identification method consists of three main parts: preprocessing, feature creation, and semantic segmentation and accuracy assessment (Figure 18). Input data required include high-resolution DEM data, four-band aerial imagery, and validated wetland/nonwetland distribution data, all in geotiff format. From these data, topographic indices (curvature, Topographic Wetness Index, and Cartographic Depth-to-Water index) and the Normalized Difference Vegetation Index are calculated. These input features are merged into a single four-band composite grid. Smaller image tiles are created from the composite grid and validation data, and the pairs of corresponding image tiles are randomly separated into training and testing datasets. Finally, dense pixel-wise wetland predictions are made using a deep learning architecture created for remote sensing data, DeepNets for Earth Observation (Audebert et al., 2018), and the accuracy of wetland predictions is assessed. The main outputs are geotiff wetland predictions for each image tile and an accuracy report for the entire validation data area. The method was implemented using open source Python Libraries and the code is available from GitHub at https://github.com/uva-hydroinformatics/wetland_identification.

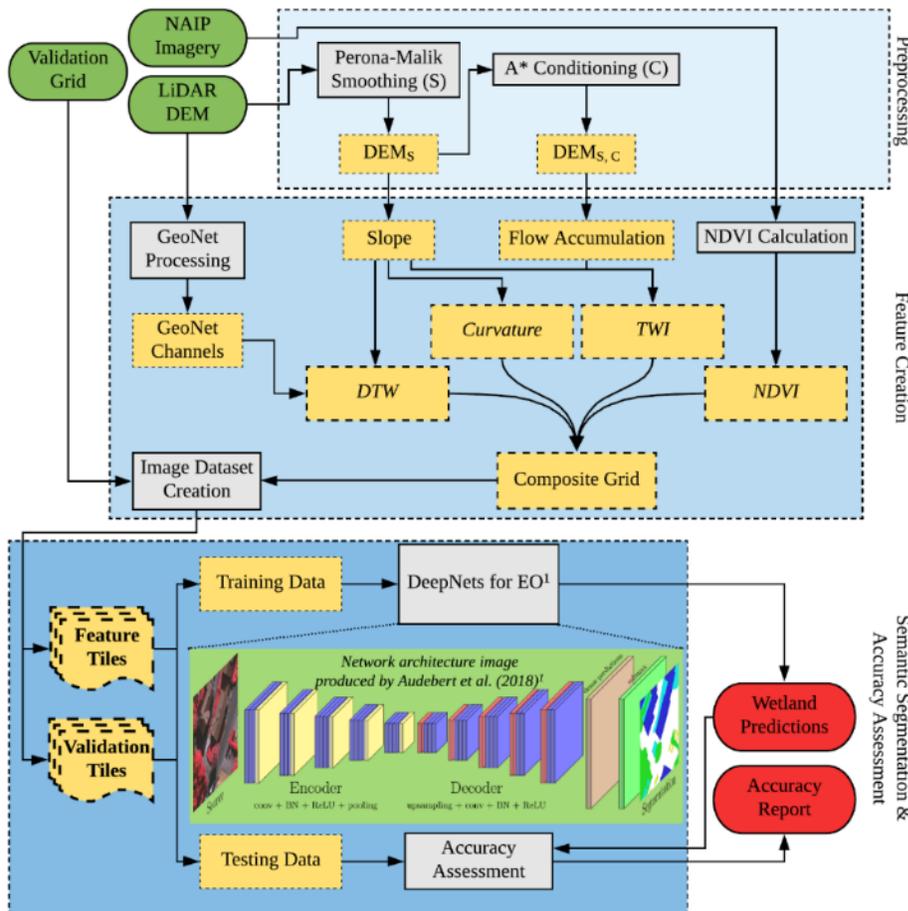


Figure 18. Overview of the proposed wetland identification method. Green shapes indicate input data, grey shapes indicate processes, yellow shapes indicate intermediate output, and red shapes indicate final model output.

¹Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 20-32.

4.2.3.1 Preprocessing

DEM preprocessing was necessary to create an improved land surface representation from which to calculate indicators of wetland geomorphology. First, DEM smoothing is performed, which is necessary to address microtopographic noise. Microtopographic noise is common in high-resolution DEMs and can be representative of either erroneous data or true variations in the elevation of vegetated surfaces (Jyotsna & Haff, 1997). DEM conditioning is then executed, which is necessary prior to modeling hydrologic flow paths, as it addresses topographic depressions (Jenson & Domingue, 1988; O’Callaghan & Mark, 1984). Topographic depressions interfere with overland flow path modeling by creating discontinuities in flow paths and accumulating water, which negatively influences modeled watershed processes (Grimaldi et al., 2007; Lindsay, 2016; Lindsay & Creed, 2005). DEM conditioning is particularly important for hydrologic modeling from high-resolution DEMs, as researchers have found that sensitivity of hydrologic parameter extraction to conditioning technique increases significantly with DEM resolution (Woodrow et al., 2016). Although many techniques have been proposed for both DEM smoothing and conditioning, we apply the Perona-Malik smoothing and A* least-cost path conditioning. This preprocessing combination was found to considerably improve wetland identification for the study sites in prior study (see O’Neil et al., 2019). The Perona-Malik filter (Perona & Malik, 1990) performs a nonlinear, anisotropic diffusion that preserves feature edges by penalizing smoothing across estimated feature boundaries (Passalacqua, Do Trung, et al., 2010; Passalacqua, Tarolli, et al., 2010). Perona-Malik smoothing was implemented using code from the nonlinear filtering module from PyGeoNet, an open source software for automatic channel network extraction from DEMs (Passalacqua, Do Trung, et al., 2010; Sangireddy et al., 2016). The A* least-cost path algorithm (Hart et al., 1968) determines the least-cost drainage paths through unaltered terrain and out of sinks, thus avoiding unnecessary modification of the input DEM (Metz et al., 2011). The A* conditioning method was executed using the GRASS GIS *r.watershed* module (GRASS Development Team, 2017; Metz et al., 2011).

4.2.3.2 Feature Creation

4.2.3.2.1 Topographic Features

In a prior study, we concluded that the curvature, Topographic Wetness Index (TWI) and Cartographic Depth-to-Water index (DTW) are successful topographic metrics for wetland identification for our study sites (O’Neil et al., 2018, 2019).

Curvature of a surface can describe the degree of convergence and acceleration of flow (Moore et al., 1991), and studies have shown its capability to indicate saturated and channelized areas (Ågren et al., 2014; Hogg & Todd, 2007; Kloiber et al., 2015; Millard & Richardson, 2015; O’Neil et al., 2018, 2019; Sangireddy et al., 2016). Here we use laplacian curvature, defined as the second derivative of the elevation grid. Laplacian curvature has been shown to favor the extraction of natural channels rather than artificial drainage paths, and to more effectively identify channels in flat, developed landscapes compared to alternative curvature forms (Passalacqua et al., 2012). Thus, we found the laplacian curvature to be most suitable for our study areas which all encompass corridor projects and are partially developed (O’Neil et al., 2019). The curvature grid is created from the smoothed DEM using code adopted from PyGeoNet (Passalacqua, Do Trung, et al., 2010; “PyGeoNet,” 2019; Sangireddy et al., 2016).

The ability of the TWI to indicate saturated areas is well-documented in the literature (Ågren et al., 2014; Lang et al., 2013; Millard & Richardson, 2015; Murphy et al., 2009; O’Neil et al., 2018, 2019). The TWI relates the potential for an area to accumulate water to its tendency to drain water, defined as

$$TWI = \ln\left(\frac{\alpha}{\tan\beta}\right), \quad (21)$$

where α is the specific catchment area (contributing area per unit contour length) and $\tan(\beta)$ is the local slope (Beven & Kirkby, 1979). The TWI was created from the smoothed, conditioned DEM using the *r.watershed* program of GRASS GIS. This module calculates the α term using the multiple flow direction algorithm (Holmgren, 1994) and the β term using a GRASS GIS-calculated slope.

Researchers have demonstrated the capability of the DTW to capture saturated areas as well (Murphy et al., 2007, 2009, 2011; O’Neil et al., 2018, 2019; Oltean et al., 2016; White et al., 2012). The DTW assumes that the likelihood for soil to be saturated increases with its proximity to surface water, in terms of distance and elevation (Murphy et al., 2007). Calculated on a per-pixel basis, the DTW is defined as

$$DTW(m) = \left[\sum \left(\frac{dz_i}{dx_i} \right) a \right] * x_p, \quad (22)$$

where $\frac{dz}{dx}$ is the downward slope of pixel i along the least-cost (i.e., slope) path to the nearest surface water pixel, a is a factor accounting for flow moving parallel or diagonal across pixel boundaries, and x_p is the pixel resolution (Murphy et al., 2007). Inputs required to calculate the DTW include a slope grid, representing cost, and a surface water grid, representing the source from which distance is calculated. We create the surface water grid directly from the LiDAR DEM using PyGeoNet, which performs a statistical analysis of curvature and uses geodesic minimization principles to predict stream lines (Passalacqua, Do Trung, et al., 2010; Sangireddy et al., 2016). Visual analyses showed that streams created by PyGeoNet better aligned with aerial imagery, compared to national hydrography data (i.e., NHD streams) and employing the commonly-used flow initiation threshold method (Band, 1986; O’Callaghan & Mark, 1984; Tarboton, 1991). PyGeoNet was executed using parameters suggested for engineered landscapes (see Sangireddy et al., 2016), which was found to produce accurate results across all sites in prior wetland model development (O’Neil et al., 2019). The PyGeoNet streams and slope grid were used as inputs to the GRASS GIS *r.cost* module (GRASS Development Team, 2017) to create the DTW grid.

4.2.3.2.2 NDVI

The NDVI is a commonly-used spectral index that relates plant biomass and stress and separates wet versus dry areas (Klemas, 2011; Ozesmi & Bauer, 2002). Researchers have used the NDVI as a wetland indicator in traditional machine learning frameworks (Corcoran et al., 2013; Dronova, 2015; Dronova et al., 2011; Guo et al., 2017; Mui et al., 2015; Rampi et al., 2014; Tian et al., 2016), as well as for general land cover classifications using deep learning frameworks (Audebert et al., 2017, 2018; Lee et al., 2019; Xu et al., 2018). The NDVI utilizes the red and the near-infrared bands (Carlson & Rizley, 1997), defined as

$$NDVI = \frac{Infrared-Red}{Infrared+Red}. \quad (23)$$

The red band indicates surface layer chlorophyll, and therefore surface conditions of plants, and the near-infrared band is reflected from the inner leaf cell structure, indicating the abundance of

plant tissue (Klemas, 2011). To calculate the NDVI, Eq. (23) was executed using NumPy operations and the appropriate NAIP imagery bands.

4.2.3.2.3 Image Dataset Creation

The image dataset creation produces two sets of image tiles: i) feature tiles representative of the composite grid of input features, and ii) validation tiles representative of ground truth wetland and nonwetland locations. Due to the irregular shapes of the field surveys, NoData pixels existed within the rectangular extent of the validation data. Rather than reduce our validation data to an extent without unverified area, NoData pixels were treated as an additional target landscape class. Thus, all pixels in the validation data were categorized as NoData (0), nonwetland (1), or wetland (2) as a first step in the image dataset creation process.

To build the dataset of feature tiles, each band of the composite grid is rescaled to a range of 0 to 1, per the requirements of the DeepNets algorithm. Rescaling the NDVI band was nontrivial, as these values have global minimum and maximum of -1 and 1. Conversely, the range of values for each of the topographic features depends on the landscape they are calculated from, therefore it was necessary to assume global minimum and maximum values. The range of each topographic input was analyzed across the study sites, and global minimum and maximum values that encompassed roughly 90% of the values were chosen. Note that only global maximum values had to be assumed for the TWI and DTW, which both have global lower bounds of 0 or nearly 0. Although this step generalizes portions of the study areas, this occurs only where there are extreme topographic features that occur infrequently. In addition, by limiting the range applied to each topographic input feature rather than choosing extreme, but encompassing, values, the significance of the relative distance between values is minimally affected. The minimum and maximum values used to rescale topographic features and the NDVI to a range of 0 to 1 are shown in Table 9.

Table 9. Minimum and maximum values used to scale each input feature to a range of 0 to 1. Minimum and maximum values were assumed for the TWI, curvature, and DTW from statistical analyses.

	TWI	Curvature	DTW	NDVI
Global Minimum	0	-3	0	-1
Global Maximum	30	3	35	1

Following these steps, the categorized validation grid and scaled composite grid were each separated into image tiles of size 320 x 320 pixels. We chose the 320-pixel size constraint to balance the desire to use image tiles large enough to depict heterogeneous landscapes and the need to separate the study site into enough images to sample training and testing tiles that were randomly dispersed. Feature and labeled image tiles sets were not considered for either training or testing if more than 80% of the area was populated with NoData pixels.

4.2.3.3 Semantic Segmentation Model: DeepNets for Earth Observation

Our model performs a semantic segmentation of input images, where each pixel of an input image is labeled as either no data, nonwetland, or wetland. To perform the semantic labeling of images, we use DeepNets (Audebert et al., 2018), a deep fully convolutional network approach designed to leverage earth observation data, including LiDAR and multispectral imagery. This model has been previously validated for the semantic labeling of urban and natural landscapes (Audebert et al., 2018). Similar to the original study, we incorporate the NDVI and elevation data.

However, rather than using the original elevation grid as an input, we guide the learning of the model by deriving specific geomorphic and hydrologic features from the DEM as inputs. As a starting point in the development of our deep learning wetland model, the baseline DeepNets architecture is implemented here (Audebert et al., 2018, 2019).

DeepNets builds on the SegNet architecture (Badrinarayanan et al., 2017) and is implemented using PyTorch (Paszke et al., 2017). SegNet produces predictions with the same resolution as the input image by using an encoder-decoder structure, making it well-suited for classification of landscape objects from georeferenced images (Audebert et al., 2018; Badrinarayanan et al., 2017). The encoder portion of SegNet is based on the convolutional layers of VGG-16 (Simonyan & Zisserman, 2014), and consists of convolutional layers, batch normalization, a rectified linear unit, and max-pooling. As shown in the inset image (Audebert et al., 2018) in Figure 2, the decoder is structurally symmetrical to the encoder. Pooling layers are replaced with unpooling layers that relocate pixel activations from the smaller feature maps to corresponding indices of zero-padded upsampled images. Convolution blocks are then used to densify the sparse pixel activations. This sequence of unpooling and convolutions is repeated until feature maps reach the original spatial resolution. Following this, a softmax layer is used to compute multinomial logistic loss. Another feature of the DeepNets approach is the generation of predictions at several resolutions, and the calculation of loss at these intermediate resolutions. In doing so, the DeepNets model predicts a semantic map at full resolution as well as smaller resolutions, which are averaged together to obtain a final full-resolution semantic prediction. Lastly, a sliding window approach is used to extract smaller patches within each input image, which acts as data augmentation.

In our implementation of DeepNets, we also applied class weights, which are related to the importance of correct predictions for a specific class when calculating the loss. We used this feature to account for the imbalance between the wetland and nonwetland classes across all sites, as well as to decrease the importance of NoData areas. Lastly, we allow for data augmentation in the form of mirroring images and flipping the orientation. Parameters for the DeepNets model incorporated into our wetland model workflow are given in Table 10. Note that these parameters were chosen as starting points to be later refined through additional model testing. For further details on the DeepNets architecture, we direct readers to Audebert et al. (2018).

Table 10. Parameters for the DeepNets implementation used in all performed experiments.

<i>Image Tile Size (# pixels)</i>	320, 320
<i>Sliding window size (# pixels)</i>	64, 64
<i>Sliding window stride (# pixels)</i>	8
<i>Base learning rate</i>	0.01
<i>Momentum</i>	0.9
<i>Weight decay</i>	0.005
<i>Training epochs</i>	100
<i>Class weights: [no data, nonwetland, wetland]</i>	[0.02, 0.08, 0.9]

4.2.3.4 Accuracy Assessment

In line with the intended environmental planning and permitting application, accuracy metrics were selected considering the greater impact of true positive (i.e., wetland) predictions

versus true negative (i.e., nonwetland) predictions to wetland conservation. Model performance was evaluated in terms of wetland recall and wetland prediction, calculated using Scikit-learn (Scikit-learn Developers, 2017b).

Recall, also known as the true positive rate, represents the percentage of true wetlands that were predicted, and is defined as

$$Recall = \frac{True\ wetland\ predictions}{Total\ true\ wetlands}. \quad (24)$$

Recall can be considered the priority indicator of model performance given the importance of the minority wetland class, a choice also supported by statistical literature (Branco et al., 2016; Chen et al., 2004; Sun et al., 2007). Precision is used to account for model overprediction. Unlike the commonly-used specificity, precision is not biased by large numbers of true negative instances, and therefore can be considered more representative for imbalanced scenarios (Branco et al., 2016; Sun et al., 2007). Precision represents the percentage of wetland predictions made that were correct, defined as

$$Precision = \frac{True\ wetland\ predictions}{Total\ wetland\ predictions}. \quad (25)$$

We found these metrics to be more suitable for model assessment compared to commonly used options, such as overall accuracy, Kappa statistic, and Matthews Correlation Coefficient (MCC). When using overall accuracy, detection rate of the minority class has a lower impact than that of the majority class (Branco et al., 2016; Chen et al., 2004), misrepresenting a wetland model predicting all nonwetland instances as very accurate. Moreover, the Kappa statistic is biased by sample size, and can increase as the wetlands to nonwetlands ratio increases, even if wetland recall decreases (Ali et al., 2014; Byrt et al., 1993). Both overall accuracy and the Kappa statistics have been omitted from wetland classification studies for these reasons (Baig et al., 2014; Zhu & Pierskalla, 2016). Although the MCC metric has been shown to be suitable for imbalanced scenarios (e.g., Boughorbel et al., 2017), it takes into account the number of true negative samples.

4.2.4 Experimental Setup

4.2.4.1 Addressing Research Question 1: Creating Site-Specific Models

Experiments 1 and 2 (Figure 19A) were designed to offer insight into potential wetland accuracy given varying sizes of reliable training sets, evaluated over four geographic regions. In Experiment 1, we created models that sample training images from the area to be mapped (i.e., site-specific models). For each site, 70% of eligible image sets were randomly selected, producing the maximum training set size available, which varied based on site size (Table 11). To compare how models of different ecoregions perform given the same training resources, site-specific models were created and evaluated at each threshold of training set size. Experiment 2 applied the site-specific models created through Experiment 1 (those using the maximum training set size) to predict wetlands in the other sites. Thus, Experiment 2 represents the scenario where a pretrained wetland model is applied for a new area without available training data.

Table 11. Maximum number of training images available per site when randomly sampling 70% of the eligible validated area. Each labeled image used for training has a resolution of 320 x 320 pixels.

Site	Maximum training set size (# images)
Site 1	31
Site 2	9
Site 3	28
Site 4	77

4.2.4.2 Addressing Research Question 2: Creating Combined-Site Models

Experiments 3 and 4 (Figure 19B) aim to evaluate the potential for improving wetland accuracy by incorporating training data from different geographic regions into a single model. In Experiment 3, a wetland model is trained using the largest training sets available from each site (i.e., “general model”). In Experiment 4, a model is created using the maximum training data from two sites within the same ecoregion: Site 2 and Site 3 (i.e., “ecoregion model”). Both experiments were aimed at gaining insight into the change in wetland predictions when the model learns wetland characteristics that exist for a range of landscapes.

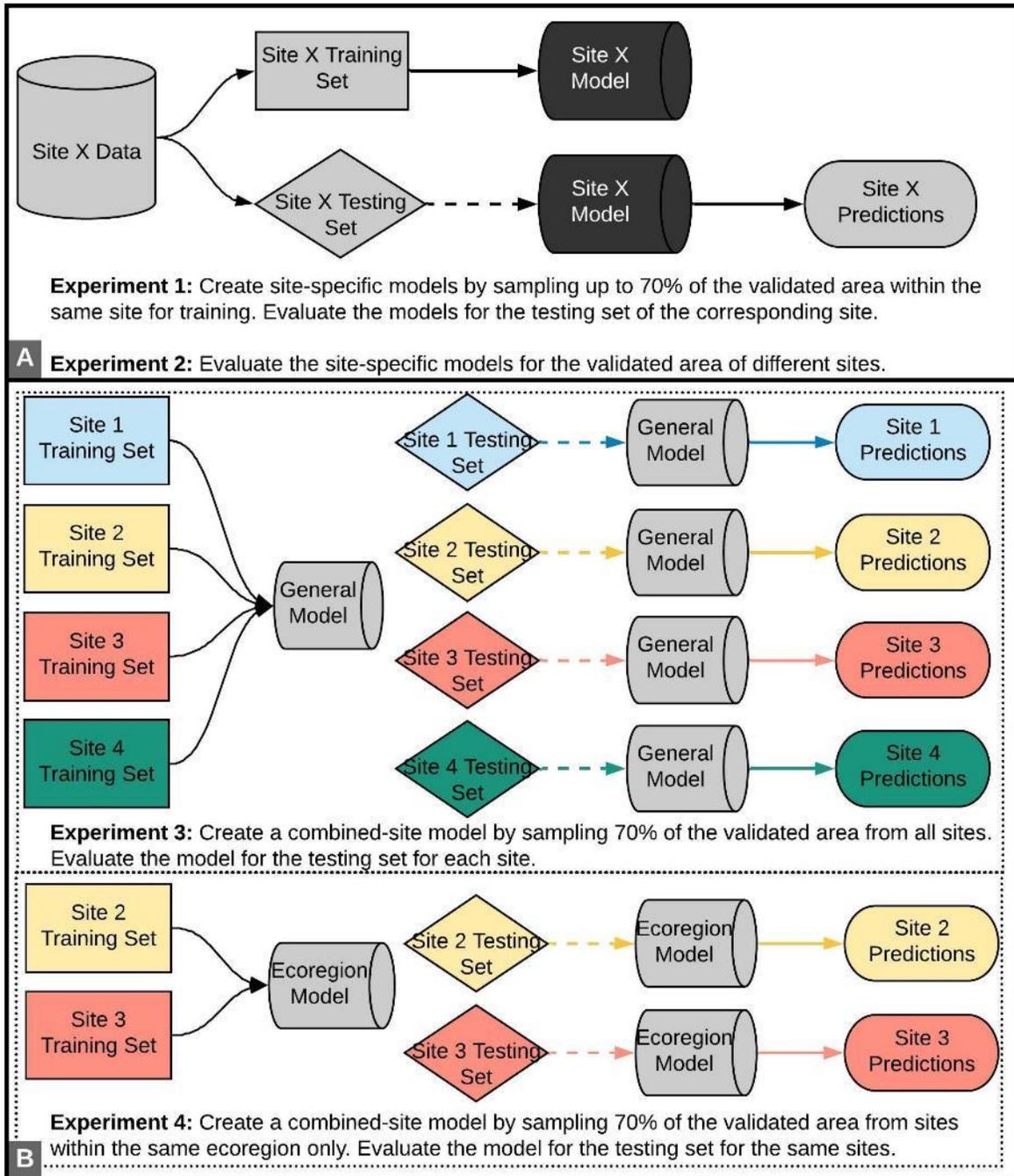


Figure 19. Methodology followed for the four experiments designed to address the study research questions.

4.3 Results

4.3.1 Performance of Site-Specific Models

For Experiment 1, site-specific models were built using training data quantities ranging from 9 to 77 images, depending on validation data extents (Figure 20). The resulting 10 sets of wetland predictions were evaluated for the testing area complementing the training data quantity used.

Results show that the best performing models for each site were those trained using the maximum training set size available, equal to 70% of the validation area. Conversely, the lowest performing models across all sites occurred when using the fewest training data, 9 images. The Site 4 model trained with 77 images achieved the highest wetland recall and precision across all site models. The Site 4 model also outperformed other sites when limited to the same number of training images (Figure 20). The overall lowest performing model was built for Site 2, which also had the smallest training dataset available, only 9 images.

While the improvements in prediction accuracy as training data increased were expected, intermediate changes in accuracy were not consistent. For Site 3, recall increased considerably (46% to 85%) and precision increased slightly (17% to 20%) when increasing training images from 9 to 28. However, changes in model accuracy were less significant for Site 1, where the most notable accuracy improvement occurred when increasing training data from 28 to 31 images, which increased recall from 70% to 81% and precision from 22% to 25%. Models built for Site 4 were generally constant from 9 to 77 training images, maintaining high performance regardless of training set size. For Site 4, recall only varied between 84% and 91% and precision between 50% and 56%. It was unexpected that Site 4 did not improve more notably when increasing the training dataset from 31 to 77 images, as this was the largest increase in training set studied. This may be due to the fact Site 4 has the most balanced wetland to non-wetland areas, so fewer training images are needed to create an accurate model.

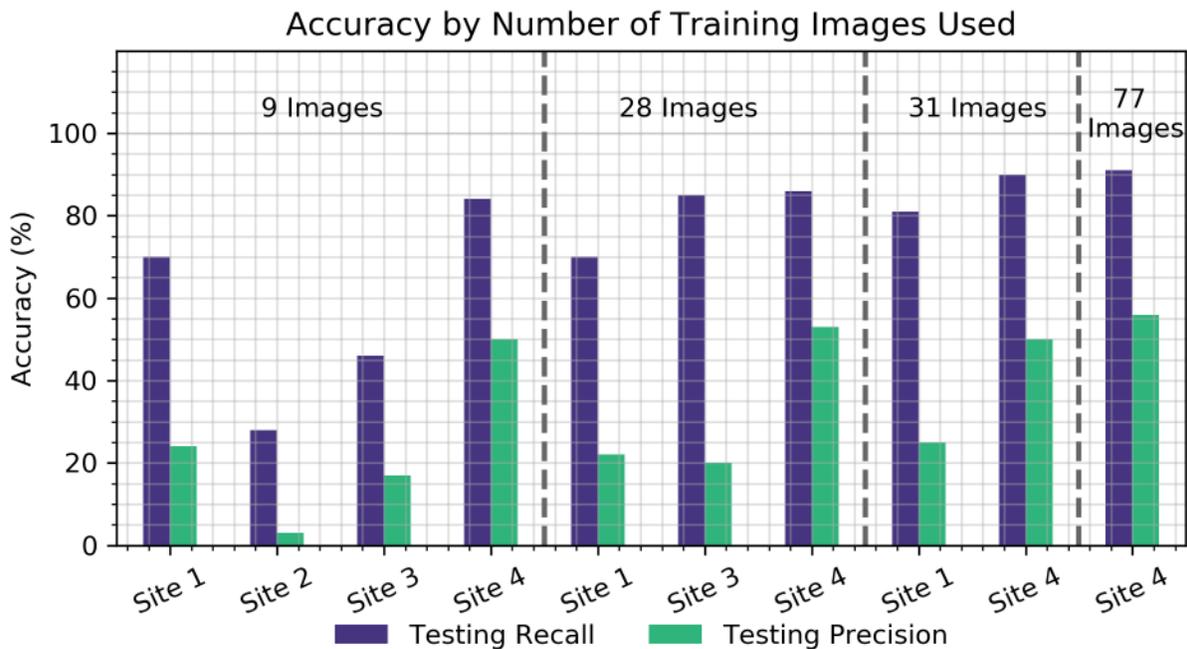


Figure 20. Wetland mapping accuracy resulting from Experiment 1, where site-specific models were created using several training data sizes depending on site availability.

4.3.2 Using Site-Specific Models to Predict Wetlands in Other Sites

Experiment 2 resulted in an additional 12 sets of results, where the best performing site-specific models (i.e., those trained with the maximum training data set size) were used to predict wetlands in the other sites. The evaluation of these trials represents wetland prediction accuracy for the entirety of the site validation area, and the results achieved by applying the site-specific models

for their own areas are also shown for reference (Figure 21). In nearly all sites, utilizing training information from a different area, even if this represented a greater quantity of data, did not improve predictions compared to those resulting from a model trained for its own area. Site 2 was the exception for this trend, as both recall and precision improved when using any of the models built for other sites, compared to using the Site 2 model. Moreover, the Site 2 model produced more accurate wetland predictions when applied to the other sites, compared to its own testing area. Although the predictions for others sites resulting from the Site 2 model were still among the lowest accuracies per site, this suggests there may be topographic or spectral confusion between Site 2 training and testing data. Lastly, the Site 4 model resulted in the highest recall scores and among the lowest precision scores across all trials for sites 1, 2, and 3. This reflects a tendency of the Site 4 model to overpredict wetlands in other sites. This may be because Site 4 includes large, areal wetlands common in the coastal plain given its low relief topography, but not common in the other three sites that not in the coastal plain.

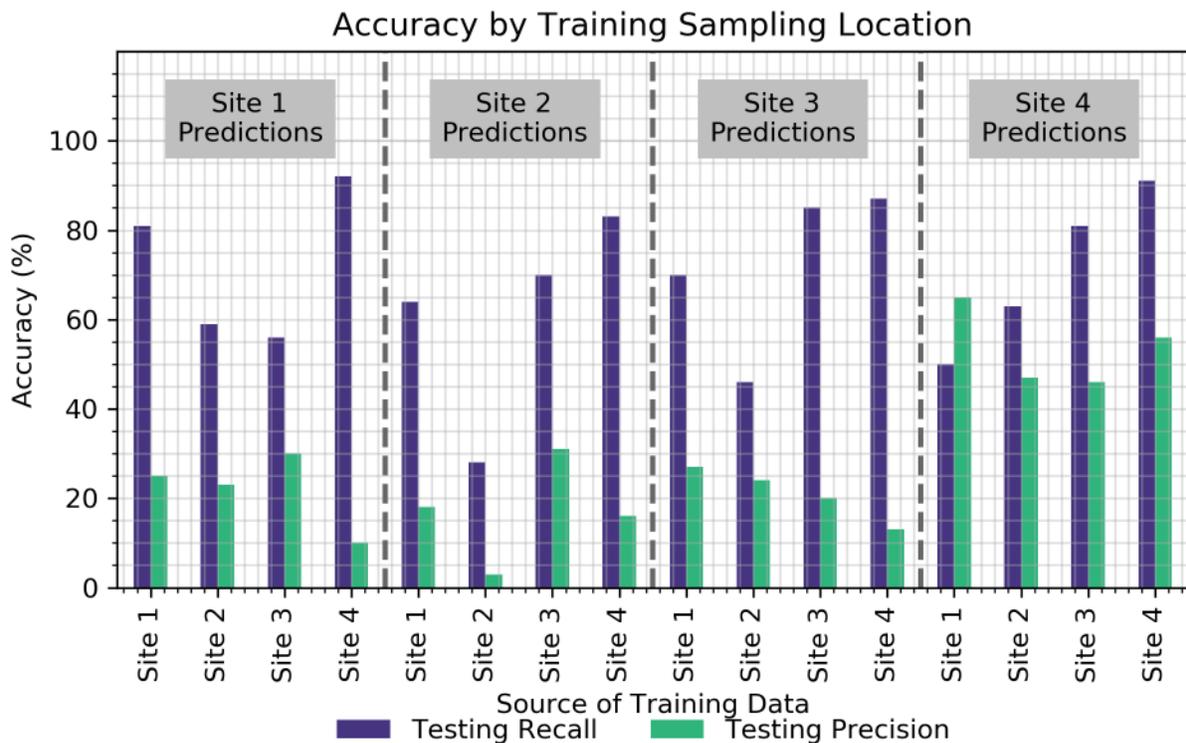


Figure 21. Wetland mapping accuracy resulting from Experiment 2, where the best performing site-specific models were used to predict wetlands in other sites.

4.3.3 Performance of Combined-Site Models

Experiment 3 resulted in the general model, trained with the maximum available training images from each site. When applying the general model to Site 1 testing areas, recall increased from to 81% to 89% and precision decreased from 25% to 18%, relative to the best performing Site 1 model (Figure 22). For Site 2 testing areas, the general model considerably improved wetland recall (28% to 40%) and minimally changed precision (3% to 2%), compared to the best performing site-specific model (Figure 22). The general model produced worse predictions than the site-specific model for Site 3, decreasing recall from 85% to 73% and precision from 20% to

15%. The general model performed nearly the same for Site 4 compared to the site-specific model, where recall remained high at 91% and precision increased by a small margin from 56% to 57%. This result suggests that a general model across all of the sites would not be a suitable way for wetland prediction, at least with the current methodology and available training and testing data.

Experiment 4 resulted in the ecoregion model, trained with the maximum available training images from sites 2 and 3, which are both in the Northern Piedmont ecoregion. This experiment tested the idea that a general wetland classification may be possible, but only within a single ecoregion and not across ecoregions as was attempted in Experiment 3. For Site 2, the ecoregion model produced worse predictions than the general model and the site-specific model, with recall decreasing to 21% and precision remaining nearly the same at 2% (Figure 22). In contrast, the ecoregion model improved wetland recall and precision for Site 3 (77% and 22%, respectively) compared to the general model, however this was not an improvement from the Site 3-specific model (Figure 22). This suggests that an accurate ecoregion-specific classification model may be possible, but with the available training and testing data, it was not able to improve on a site-specific model.

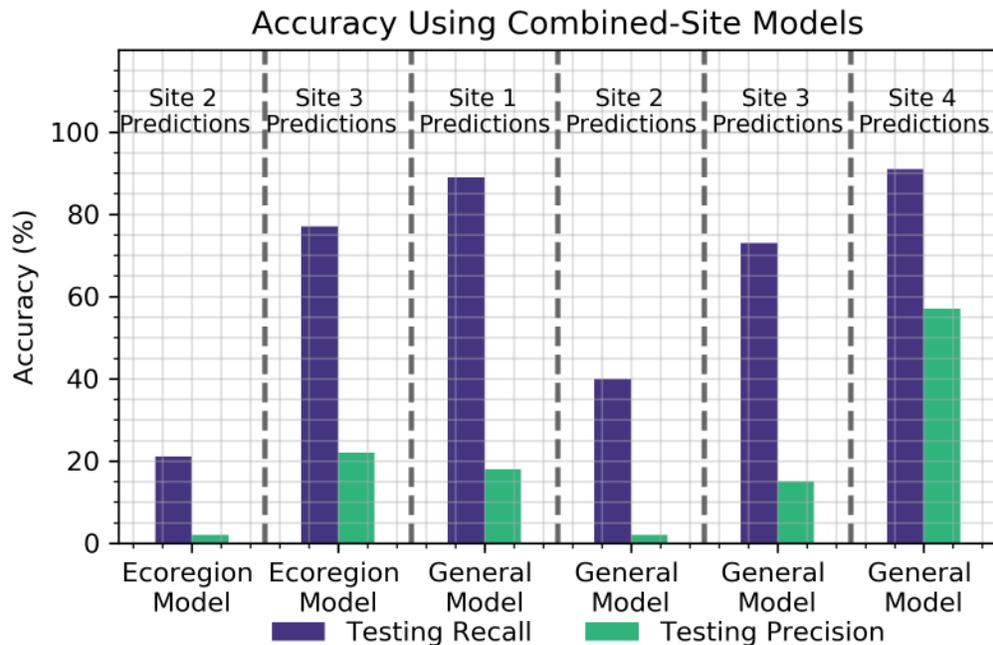


Figure 22. Wetland mapping accuracy resulting from Experiment 3, which used training data from all sites to create a general model, and Experiment 4 which used training data only from sites within the same ecoregion (sites 2 and 3) to create an ecoregion model.

4.4 Discussion

4.4.1 Potential for Site-Specific Models

We found that site-specific models improved as more training data was sampled from the area to be mapped, with the best models created from the maximum training datasets studied: 70% of the validation area. The improvements as training data increased are likely due to the ability of the model to learn a wider range of wetland characteristics that exist in the additional landscape scenes. Figure 23 demonstrates the changes in wetland predictions as a result of increasing training data from 9 training images (column A) to the maximum training images per site (column B). For

sites 1, 3, and 4, increased training data reduced wetland overprediction surrounding the extents of ground truth wetlands, most notably for narrow wetland segments in sites 1 and 3. In addition, wetland predictions for these sites encompassed more of the true wetland area, most apparent for Site 4, where predictions densified within a relatively large wetland after increasing training data. Figure 23 also exemplifies the poor performance of the Site 2 model. Although the Site 2 model predicts wetlands as small, linear features that are representative of the nature of ground truth wetlands in the area, the predictions are relatively sparse and incorrect. By visually examining the input features and testing data for Site 2, we found that validation wetlands existed underneath dense tree canopy along a road corridor. Topographic metrics in this area indicated values corresponding to wetness within the true wetland boundaries, however, the NDVI showed constant values for most of the forested area. The lack of distinction between values by the NDVI is likely due to the source imagery, the NAIP, which is collected during the growing season, with leaf-on conditions and is therefore affected by tree canopy. Moreover, the better performance for Site 4, even when using few training data, suggests that this landscape was particularly well-suited to the deep learning network. This may be due to the large distribution of wetlands in Site 4, leading to a higher quantity of wetlands in the entire training data set as well as more significant presence of wetlands in each training image.

Figure 23 also shows model predictions when using the Site 4-specific model. The Site 4 model produced predictions with the highest recall scores of all model trials for sites 1, 2, and 3. As indicated by the increases in recall, predictions resulting from the Site 4 more densely encompassed the ground truth wetlands (Figure 23, column C), relative to results for the site-specific models (Figure 23, columns A and B). Attributing to the lower precision scores also produced by the Site 4 model, wetland overprediction is apparent in the scenes for site 1, 2, and 3 (Figure 23, column C). The wetland predictions for these sites are also made at a coarse resolution within image tile extents, evident by the rectangular edges of wetland predictions in sites 1 and 3 (Figure 23, column C). In addition, segment of a narrow wetland feature is omitted for Site 3 when applying the model trained for Site 4. Overall, these shortcomings demonstrate the potential for bias to a specific landscape and wetland type in site-specific models, which may lead to decreased accuracies when applied to different landscapes. This may be overcome by changing the classification strategy away from a simple wetland/non-wetland classification to one that classifies different wetland types, although this strategy was not explored through this research. The increase in recall scores when using the Site 4 model, and the concentration of wetland overprediction occurring in the adjacent and surrounding areas of the ground truth wetlands, suggests the noted shortcomings may also be addressed by using a more balanced sampling of different wetland types.

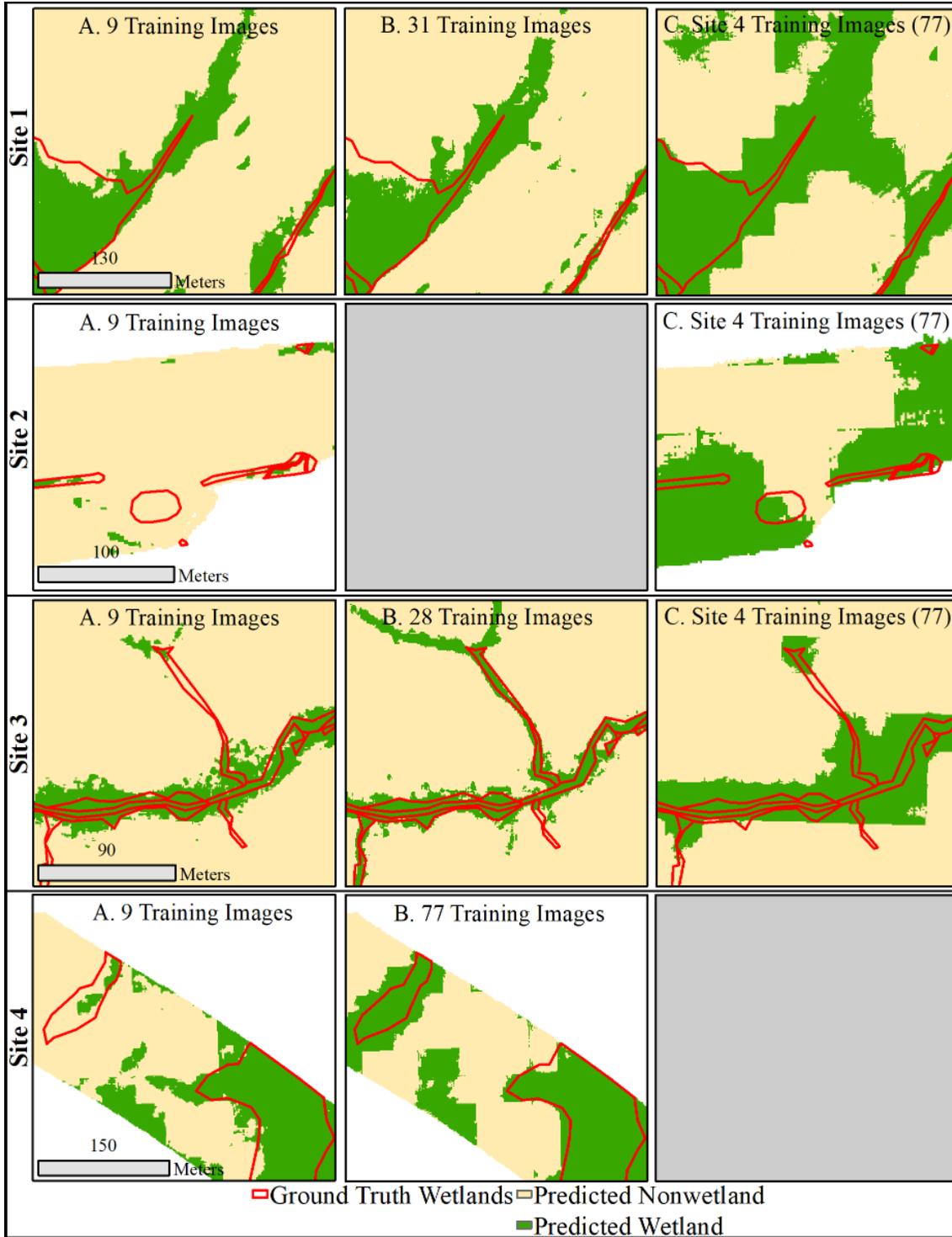


Figure 23. Comparison of wetland predictions produced by site-specific models created from (column A) the smallest training dataset and (column B) the largest training dataset available for the site. Also shown are wetland predictions produced by models trained only with the largest training dataset for Site 4 (column C).

4.4.2 Potential for Combined-Site Models

Compared to the site-specific models, the general model mostly resulted in more wetland overprediction, but in some cases increased coverage ground truth wetlands (Figure 24, column

B). This trend is likely due to the bias of the general model to favor wetland types present in the Site 4 landscape, as more than half of all the training images used were from Site 4. While the general model results do not present an improvement from the site-specific models, there are improvements compared to wetland predictions resulting from a model trained only on Site 4 (see Figure 23, column C). By supplementing the Site 4 training data with wetland information from other landscapes, we see finer, more precise wetland prediction boundaries (Figure 24, Site 1 B and Site 3 B). For Site 2, the general model produced a greater overall amount of wetland predictions compared to the site-specific model, but predictions were inaccurate (Figure 24, column A vs. column B). However, the erroneous wetland predictions for Site 2 were greater when using the Site 4 model versus the general model. It was expected that predictions for Site 4 would be mostly unchanged between the site-specific model and the general model, due to the significant presence of Site 4 training data. However, the weak training data influence from other sites did slightly improve precision for Site 4, demonstrated by finer-scale edges of wetland predictions (Figure 24 Site 4 A vs. Site 4 B).

The ecoregion model explored the potential for creating combined-site models that are specific to certain landscape characteristics by including training data only from within the same ecoregion (i.e., sites 2 and 3). Fewer wetland predictions were made overall for Site 2 using the ecoregion model (Figure 24, column C), which considerably reduced compared to the general model, but also resulted in sparser correct wetland predictions than the Site 2-specific model. For Site 3, the ecoregion model improved both precision and recall compared to the general model, but results were still less accurate than the site-specific model. Compared to general model predictions, the ecoregion model regained correct wetland predictions for narrow, riparian wetland features for Site 3 (Figure 24, column C). The ecoregion model also reduced wetland overprediction compared to the general and site-specific models in the scenes shown in Figure 24, representative of the higher precision produced by the ecoregion model (22% vs. 20% by the site-specific model and 15% by the general model). However, wetland predictions resulting from the ecoregion model encompassed less ground truth wetland area overall relative to the Site 3-specific model.

Although neither approach for creating a combined-site model was able to outperform site-specific models, results show potential to refine and improve these methods. A potential source of error for both the general and ecoregion models may be the unequal sampling of training data from the different geographic study areas. To test this hypothesis, the general model and the ecoregion model were recreated by limiting training data from sites to just 9 images each, balancing the representation from each site. For all sites, the general model built with equal, but limited training data performed worse than the proposed general model. For Site 3, the ecoregion model built with limited training data performed considerably worse, where recall decreased from 77% to 30% and precision improved slightly from 15% to 17%. For Site 2, however, the limited ecoregion model improved results slightly (recall increasing from 21 to 27% and precision remaining at 2%), but still not to an acceptable level of accuracy. Thus, improving the combined-site model approach may not just be a matter of equally sampling different landscapes, but also balancing an adequate amount of training data from different landscapes. Lastly, the lack of consistent improvement to Site 2 and Site 3 predictions when applying the ecoregion model suggests it would be beneficial to consider additional landscape similarities when building combined-site models. Landscape

characteristics to consider may be those that affect the distributions of topographic inputs, such as influence of built environment drainage and land cover.

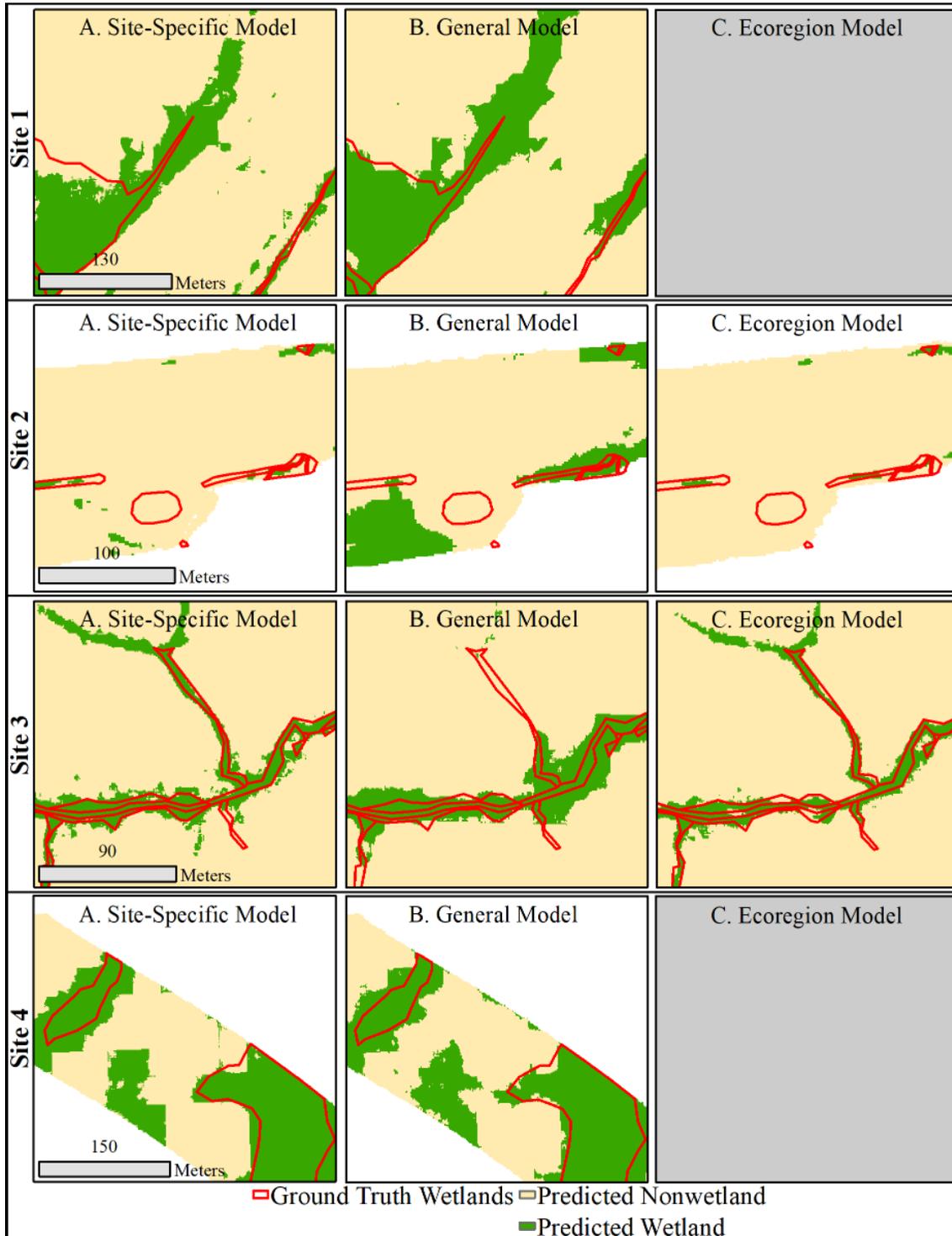


Figure 24. Comparison of wetland predictions resulting by (column A) the best performing site-specific models (i.e., those trained on 70% of the validation area), (column B) the general model, and (column C) the ecoregion model.

4.4.3 Utility of the Proposed Input Data Configuration

This study explored an input data configuration unique to most deep learning applications where topographic derivatives of the input “image” (i.e., LiDAR DEM) are predetermined and specific to the target object (i.e., wetlands). The hypothesis was that these predetermined elevation derivatives (TWI, DTW, and curvature) would improve wetland classification training by including hydrologic information, compared to training directly from the elevation data. To evaluate the efficacy of this method, we compared the accuracy achieved using our novel input data configuration versus two-band images composed of the LiDAR DEM and the NDVI, which is more representative of the common input data approach taken (e.g., Audebert et al., 2017, 2018; Latifovic et al., 2018; Liu et al., 2018; Silburt et al., 2018; Xu et al., 2018). The LiDAR DEMs used to create the two-band images were smoothed and hydrologically corrected, as suggested by O’Neil et al. (2019), and 70% of the areas were used for training for both model sets.

For sites 1, 2, and 3, the proposed input data configuration outperformed the typical approach in terms of both recall and precision. Wetlands predicted from only the DEM and NDVI for Site 1 achieved lower recall (73% vs. 81%) and precision (21% vs. 25%) compared to the models using the derived topographic indices and the NDVI. For Site 2, predictions learned from the DEM and NDVI encompassed only 12% of the ground truth wetlands with near 0% precision, compared to 28% recall and 3% precision achieved by the proposed approach. Wetland predictions for Site 3 lost considerable accuracy with the typical input data approach, producing 24% recall and 9% precision, whereas our approach resulted in 85% recall and 20% precision. For Site 4, this comparison showed that the model that learned from the DEM and NDVI alone produced a higher recall (96% vs. 91%) and lower precision (49% vs. 56%). While this indicates that more ground truth wetlands were detected using the typical approach, it is slightly outweighed by the loss in wetland precision. Considering the consistent improvement to the other three sites, the lack of significant change in Site 4 when applying only the DEM and NDVI may suggest that the deep learning model relies more heavily on the vegetative characteristics provided by the NDVI than the geomorphologic and hydrologic information that the elevation data offers. This is likely due to the fact that Site 4 had the least topographic relief, being within the coastal plain. Results for Site 4 using a Random Forest classification (see O’Neil et al., 2019) also support this idea, showing that the topographic input variables were insufficient for describing wetland characteristics unless preprocessing methods were calibrated specifically to the area. Thus, it is logical that wetlands in Site 4 are better described by vegetative characteristics than topography, explaining the lack of change in predictions when replacing the topographic inputs with the DEM and leaving the NDVI input unchanged.

4.4.4 Comparison of Deep Learning to a Random Forest Implementation

To examine the potential for deep learning to advance the more commonly used Random Forest approach for wetland classification (e.g., O’Neil et al., 2019), we compared the performance of the site-specific deep learning models to a Random Forest classification with the same set of input variables. The Random Forest implementation follows the approach of O’Neil et al. (2019), but with the addition of the NDVI to the original set of inputs: the TWI, curvature, and DTW. The training sampling used in the O’Neil et al. (2019) study was maintained, where training data consists of randomly dispersed pixels that encompass only 15% of the validated wetland area and up to 8% of the validated nonwetland area. However, accuracy assessments for both the deep

learning and Random Forest models were limited to the extents of the testing image tiles that correspond to the deep learning approach.

Compared to the Site 1 deep learning model, the Random Forest classification resulted in an improvement in recall from 81% to 91%, but a decrease in precision from 25% to 19%. For Site 2, Random Forest improved recall considerably, from 28% to 78%, and slightly improved precision from 3% to 5%. The Site 3 Random Forest model produced no change in recall (85%) and a slight decrease in precision (20% vs. 18%), compared to deep learning. Finally, the Site 4 Random Forest model considerably decreased recall from 91% to 70% and increased precision from 56% to 64%, relative to the deep learning model. With the exception of Site 2, these findings show that deep learning was able to perform similarly to Random Forests (e.g., Site 1 and Site 3), and arguably better in some cases (e.g., Site 4). The poor performance in Site 2 further supports that the deep learning model was not sufficiently able to learn characteristics of wetland features that were very small and sparse relative to the landscape scenes in each training image. Similarly, the Site 4 results again support the idea that deep learning is better suited to detecting wetlands where they are areal and large relative to the landscape scene. In addition, an evaluation of the entire testing areas corresponding to the Random Forest models shows that the inclusion of the NDVI as a wetland indicator improves on the O'Neil et al. (2019) approach. Compared to the Random Forest models using only the topographic inputs, the addition of the NDVI improved wetland recall and precision in Site 1 (81% vs. 88% and 19% vs. 24%), Site 2 (82% vs. 88% and 16% vs. 22%), Site 3 (83% vs. 86% and 22% vs. 25%), and Site 4 (58% vs. 68% and 47% vs. 54%).

Overall, it is important to note that the Random Forest models were able to achieve these accuracies by sampling much less training data than was required for deep learning models. However, this result also shows that deep learning models can approach the same accuracies using training data resources that are considerably smaller relative to most deep learning applications. In addition, the similar performance of deep learning to Random Forest in three of the study sites supports findings by other researchers that state deep learning can improve landscape segmentation accuracy over traditional machine learning, such as support vector machine, maximum likelihood classification, and Random Forest, given enough training data (e.g., Hu et al., 2018; Latifovic et al., 2018; Liu et al., 2018; Mahdianpari et al., 2018).

4.4.5 Future Work

Additional methods and analyses should be considered in the future refinement of this work. For example, incorporating Class Activation Mapping (CAM) (Zhou et al., 2016), which highlights scene elements that are most influential during classifications, would offer further insight into model learning. By utilizing CAM, model refinements could be made by quantifying the impact of the input data and identifying sources of error. Considering additional remote sensing data may also improve model performance. These may include LiDAR point clouds, which researchers have incorporated into 3-dimensional CNNs for wetland identification (e.g., Xu et al., 2018). Also, incorporating radar data may reduce errors where the NDVI is affected by tree canopy, as it is able to penetrate this layer and provide vegetation density and inundation information for wetland mapping (Allen et al., 2013; Behnamian et al., 2017; Corcoran et al., 2013; Kloiber et al., 2015; Millard & Richardson, 2013). Also on this point, the contribution of each input data source throughout the DeepNets workflow can be handled in a more sophisticated way.

This was demonstrated by Audebert et al. (2018) who proposed novel data fusing methods for elevation data and the NDVI within the DeepNets workflow to improve land cover classifications.

Additional training information that consists of accurately delineated wetlands from across different ecoregions should improve the deep learning classification results. Also, additional training data would make it possible to train models for specific wetland types rather than a simple, binary wetland/nonwetland classification. These training data are likely available from state and federal agencies given the need for wetland assessments under the Clean Water Act, but are not collected into a single, standardized repository. Future work could focus on building such a training and testing repository for wetland classification. Furthermore, to more efficiently make use of any amount of reliable training information available, applying more sophisticated data augmentation techniques may improve wetland predictions, as demonstrated by Stivaktakis et al. (2019).

Refinements to the current approach should also include more robust accuracy assessments. The current accuracy metrics are transparent and represent the two factors that are needed for reliable implementation: coverage of ground truth wetlands and limited overprediction. However, a single accuracy metric that encompasses both of these factors while also acknowledging the significantly higher importance of wetland recall would improve the interpretation of model results. Model evaluation improvements should also take into account the diffuse boundaries of wetlands which may fluctuate seasonally by penalizing overprediction less if it occurs adjacent to or surrounding defined ground truth wetland extents. Lastly, this study did not test the effect of tuning the DeepNets parameters. Among other parameter adjustments, future work should explore the benefit of adjusting window sizes based on target wetland size and the accuracy tradeoffs when training the model for more epochs.

4.5 Conclusions

We explore a wetland identification workflow that implements a basic semantic segmentation architecture and an input data configuration that consists of the NDVI and LiDAR DEM-derived indicators of wetland hydrology and geomorphology. The workflow was trained and evaluated using available data resources from four geographic regions across Virginia. From this work, we draw the following conclusions.

- i. Site-specific deep learning models created from relatively small training datasets can achieve accurate results. For three of the four study sites, wetland recall ranged from 81 to 91% and precision ranged from 20 to 56%, when training models with 70% of site area and testing on the remaining 30% of the site area.
- ii. Site-specific models were more successful for areas where wetlands are abundant and occupy a significant portion of training images. For a site with large, areal wetlands that were almost evenly balanced with nonwetland areas, high accuracy was achieved with 7.5 km² (70%) of training area (91% recall and 56% precision). Using a much smaller training area, 0.4 km² (10% of the study area), still resulted in a fairly accurate model (84% recall and 50% precision).
- iii. In most cases, accuracy decreased when using models trained for another site. However, the site-specific model trained with the largest area studied (7.5 km²)

increased wetland recall in all other sites. Although model predictions were imprecise and showed a bias towards the types of wetlands for which it was trained (i.e., large, areal wetlands), the correct localization of wetland predictions suggests there is potential for this approach if models are trained with sufficient data and for areas with similar landscapes.

- iv. Combined-site models can produce accurate wetland predictions, but training data contributions from the target landscapes should be balanced. The general model revealed the potential for bias towards landscape characteristics more heavily represented in the training data. However, the influence of less represented sites was still apparent, as wetland predictions were more inclusive of different wetland types compared to a model created without training data from these sites.
- v. Shared ecoregion alone may not offer sufficient landscape similarities to improve the training sampling approach for combined-site models. The ecoregion model showed accuracy improvements from the general model for one site. However, wetland predictions for the other site were less accurate. Future work should explore the benefit of creating combined-site models from areas that share additional characteristics that would affect the distributions of the topographic derivatives, such as level of development, land cover, and topography.
- vi. The proposed input data configuration improves wetland identification compared to a more typical approach of using the NDVI and the LiDAR DEM directly. By predetermining the derivatives of the DEM that are wetland indicators, rather than allowing the deep learning network to determine these through convolutions, wetland predictions were more accurate in three sites. For the remaining site, accuracy was nearly unchanged between the two approaches. However, analyses show that this is likely due to the greater importance of the NDVI for identifying wetlands in the topographically mild landscape.
- vii. Compared to a Random Forest approach, the best performing models produced comparable accuracy, using more training data than required for Random Forest, but still significantly less than what is typical in most deep learning applications.

Our results demonstrate the potential for deep learning to not only improve accuracy compared to traditional machine learning algorithms, but also provide flexible models that are accurate for a range of landscapes. Paramount to achieving this will be larger efforts within the research community to gather reliable training data and pretrained models stored as open source repositories, as has been done for established deep learning fields (e.g., Lecun, 1999; Lin et al., 2014). The wetland models created through this research may offer a starting point for creating a repository open to other researchers. By refining this implementation of the deep learning wetland workflow and further training the created models, there is potential for deep learning to support a range of wetland conservation efforts by producing accurate wetland inventories across many landscapes.

Chapter 5

Conclusions

This dissertation advances the application of freely-available remote sensing data, with an emphasis on LiDAR datasets, along with hydrologic terrain processing and machine learning algorithms, for wetland identification. Through the development and refinement of a wetland identification method and tool, innovative approaches are offered for the approximation of complex wetland features at an environmental planning-scale using commonly-available data and open source methods. Among the contributions to this research field are i) the identification and evaluation of LiDAR-derived topographic metrics that describe near-surface soil moisture across a range of landscapes, ii) a thorough analysis of the compound effects of smoothing and conditioning on wetland delineations and the Random Forest model used to generate them, and iii) a novel approach to performing deep learning for wetland mapping and new insights into the potential for using this approach while being limited to training data resources that are typical for remote sensing researchers.

Through this research, three LiDAR-derived indices have been thoroughly evaluated in terms of their ability to indicate likely wetland areas. In Chapter 2, the potential for improvement of wetland identification models through modification of LiDAR DEM derivatives, at a scale necessary for small-scale wetland delineations, is evaluated. A novel approach of flow convergence modelling is presented where Topographic Wetness Index (TWI), curvature, and Cartographic Depth-to-Water index (DTW), are modified to better distinguish wetland from upland areas, combined with ancillary soil data, and used in a Random Forest classifier. This approach is applied to four study sites in Virginia, implemented as an ArcGIS model. The model shows a significant improvement in average wetland accuracy compared to the commonly used National Wetland Inventory (84.9% vs. 32.1%), at the expense of a moderately lower average non-wetland accuracy (85.6% vs. 98.0%) and average overall accuracy (85.6% vs. 92.0%). From this, we conclude that modifying TWI, curvature, and DTW provides more robust wetland and non-wetland signatures to the models by improving accuracy rates compared to classifications using the original indices. The resulting ArcGIS model is a general tool able to modify these local LiDAR DEM derivatives based on site characteristics to identify wetlands at a high resolution.

This work also provides a novel evaluation of the effects of LiDAR DEM smoothing and conditioning on wetland extraction. In Chapter 3, we propose an open source, automated wetland identification model that relies primarily on LiDAR DEMs. Although LiDAR DEMs provide the resolution needed to map detailed topographic metrics and areas of likely soil saturation, the choice of smoothing and conditioning techniques can significantly impact the accuracy of hydrologic parameter extraction. So far, the effect of these preprocessing steps on wetland delineation has not been thoroughly analyzed. We test the response of a Random Forest wetland classifier, using topographic wetness index, curvature, and cartographic depth-to-water index as input variables, to

combinations of smoothing techniques (none, mean, median, Gaussian, and Perona-Malik) and conditioning techniques (Fill, Impact Reduction Approach, and A* least-cost path analysis) for four sites in Virginia, USA. The Random Forest model was configured to account for imbalanced data sets, and manually surveyed wetlands were used for verification. Applying Perona-Malik smoothing and A* conditioning yielded the highest accuracy across all sites and considerably reduced model runtime. We found that models could be further improved by individualizing the smoothing method and scale to each input variable. Using only topographic information, the wetland identification model could accurately detect wetlands in all sites (81-91% recall). Model overprediction varied across sites, represented by precision scores ranging from 22 to 69%. In its current form, the wetland model shows strong potential to support wetland field surveying by identifying likely wetland areas.

By exploring an alternative to the pixel-based Random Forest approach for wetland identification, this research also demonstrates the potential to use deep learning as a wetland identification approach using the LiDAR-derived indices and the NDVI. In Chapter 4, we propose a wetland identification workflow that implements a basic deep learning architecture that learns from LiDAR DEM-derived geomorphic and hydrologic wetland indicators, as well as the commonly-used NDVI. By implementing our workflow across four geographic regions of Virginia, we offer new insights pertaining to the potential for semantic segmentation of wetlands using training data resources that are representative of most remote sensing researchers. Results show that models trained and evaluated for a single site were able to achieve high accuracy (up to 91% recall and 56% precision), as well as models trained with data from multiple sites (up to 91% recall and 57% precision). Through this analysis, we found that, across all sites, our derived input data configuration that takes advantage of hydrologic principals outperformed models using the more typical approach of building the deep learning model directly from the elevation data. By refining the deep learning wetland identification workflow and collecting additional training data, there is potential for the created wetland models to support a range of wetland conservation efforts.

As climate change continues to be a prevalent issue in society and growing populations demand further agricultural and development repurposing, the importance of wetland management and protection will only increase. This research will improve the ability to accurately inventory remaining wetlands in a quick and cost-effective way. By designing the wetland identification model to require only freely-available input data and use open source computational methods, there is potential for the wetland tool to be implemented by a wide range of users. Using professionally surveyed wetlands encompassing different landscapes for model evaluation, we have demonstrated that the proposed wetland model outperforms the most commonly-used national-scale wetland inventory dataset, the National Wetland Inventory (NWI). With the completion of model refinement through this research, the wetland tool has the potential to benefit the broader environmental planning and conservation community.

References

- (VITA), V. I. T. A. (2016). Virginia Base Map Data Downloads. Retrieved August 1, 2016, from <http://vgin.maps.arcgis.com>
- Ågren, A. M., Lidberg, W., Strömberg, M., Ogilvie, J., & Arp, P. A. (2014). Evaluating digital terrain indices for soil wetness mapping—a Swedish case study. *Hydrology and Earth System Sciences*, 18(9), 3623–3634. <https://doi.org/10.5194/hess-18-3623-2014>
- Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Tarolli, P. (2014). A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions. *Earth Surface Processes and Landforms*, 39(3), 399–413. <https://doi.org/10.1002/esp.3506>
- Allen, T. R., Wang, Y., & Gore, B. (2013). Coastal wetland mapping combining multi-date SAR and LiDAR. *Geocarto International*, 28(7), 616–631. <https://doi.org/10.1080/10106049.2013.768297>
- Aslam, J. A., Yilmaz, E., & Pavlu, V. (2005). A geometric interpretation of r-precision and its correlation with average precision. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '05* (p. 573). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1076034.1076134>
- Audebert, N., Le Saux, B., & Lefèvre, S. (2017). Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks. In S.-H. Lai, V. Lepetit, K. Nishino, & Y. Sato (Eds.), *Computer Vision -- ACCV 2016* (pp. 180–196). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-54181-5_12
- Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 20–32. <https://doi.org/10.1016/j.isprsjprs.2017.11.011>
- Audebert, N., Le Saux, B., & Lefèvre, S. (2019). Deep Learning for Earth Observation. Retrieved January 1, 2019, from <https://github.com/nshaud/DeepNetsForEO>
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, 5(5), 423–431. <https://doi.org/10.1080/2150704X.2014.915434>
- Band, L. E. (1986). Topographic Partition of Watersheds with Digital Elevation Models. *Water Resources Research*, 22(1), 15–24. <https://doi.org/10.1029/WR022i001p00015>
- Barnes, R., Lehman, C., & Mulla, D. (2014). Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers and Geosciences*, 62(2), 117–127. <https://doi.org/10.1016/j.cageo.2013.04.024>

References

- Batuwita, R., & Palade, V. (2010). Efficient resampling methods for training support vector machines with imbalanced datasets. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). <https://doi.org/10.1109/IJCNN.2010.5596787>
- Behnamian, A., Banks, S., White, L., Brisco, B., Millard, K., Pasher, J., et al. (2017). Semi-automated surfacewater detection with synthetic aperture radar data: A wetland case study. *Remote Sensing*, *9*(12), 1–21. <https://doi.org/10.3390/rs9121209>
- Besl, P. J., Birch, J. B., & Watson, L. T. (1989). Robust window operators. *Machine Vision and Applications*, *2*(4), 179–191. <https://doi.org/10.1007/BF01215874>
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, *24*(1), 43–69. <https://doi.org/10.1080/02626667909491834>
- Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE*, *12*(6), 1–17. <https://doi.org/10.1371/journal.pone.0177678>
- Bradski, G., & Kaehler, A. (2000). OpenCV. *Dr. Dobb's Journal of Software Tools*, *3*.
- Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys*, *49*(2), 1–50. <https://doi.org/10.1145/2907070>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buchanan, B. P., Fleming, M., Schneider, R. L., Richards, B. K., Archibald, J., Qiu, Z., & Walter, M. T. (2014). Evaluating topographic wetness indices across central New York agricultural landscapes. *Hydrology and Earth System Sciences*, *18*(8), 3279–3299. <https://doi.org/10.5194/hess-18-3279-2014>
- Burrough, P. A., McDonnell, R., McDonnell, R. A., & Lloyd, C. D. (2015). *Principles of geographical information systems*. Oxford university press.
- Burt, T., & Butcher, D. (1986). Stimulation from simulation? A teaching model of hillslope hydrology for use on microcomputers. *Journal of Geography in Higher Education*, *10*(1), 23–39. <https://doi.org/10.1080/03098268608708953>
- Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, *46*(5), 423–429. [https://doi.org/10.1016/0895-4356\(93\)90018-V](https://doi.org/10.1016/0895-4356(93)90018-V)
- Carlson, T. N., & Riziley, D. A. (1997). On the Relation between NDVI , Fractional Vegetation Cover , and Leaf Area Index. *Remote Sensing of Environment*, *62*(3), 241–252.
- Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, (1999), 1–12. <https://doi.org/ley.edu/sites/default/files/tech-reports/666.pdf>

References

- Corcoran, J. M., Knight, J. F., & Gallant, A. L. (2013). Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota. *Remote Sensing*, 5(7), 3212–3238. <https://doi.org/10.3390/rs5073212>
- Cowardin, L. M., & Golet, F. C. (1995). U.S. Fish and Wildlife Service 1979 wetland classification: a review. *Vegetatio*, 118, 139–152. <https://doi.org/10.1007/BF00045196>
- Dahl, T. E., Johnson, C. E., & Frayer, W. E. (1991). *Wetlands, status and trends in the conterminous United States mid-1970's to mid-1980's*.
- Davidson, N. C. (2014). How much wetland has the world lost? Long-term and recent trends in global wetland area. *Marine and Freshwater Research*, 65(10), 934–941. <https://doi.org/10.1071/MF14173>
- Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, 233–240. <https://doi.org/10.1145/1143844.1143874>
- Deng, J., Smith, A. S., Davis, S., Weatherford, M., Paugh, L., & Wang, S.-G. (2017). *Identification of NC Wetland Types by LiDAR Data and Tree Based Machine Learning Methods*.
- Dronova, I. (2015). Object-based image analysis in wetland research: A review. *Remote Sensing*, 7(5), 6380–6413. <https://doi.org/10.3390/rs70506380>
- Dronova, I., Gong, P., & Wang, L. (2011). Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China. *Remote Sensing of Environment*, 115(12), 3220–3236. <https://doi.org/10.1016/j.rse.2011.07.006>
- Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, 259–272. <https://doi.org/10.1016/j.rse.2011.11.020>
- ESRI. (2016). Train Random Trees Classifier. Retrieved August 1, 2017, from <http://desktop.arcgis.com/en/arcmap/10.4/tools/spatial-analyst-toolbox/train-random-trees-classifier.htm>
- Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, 20(1), 18–36. <https://doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x>
- Farm Service Agency. (2017). National Geospatial Data Asset (NGDA) National Agriculture Imagery Program (NAIP) Imagery. Retrieved January 1, 2019, from <https://catalog.data.gov/dataset/national-geospatial-data-asset-ngda-naip-imagery>
- Felton, B. R., O'Neil, G. L., Robertson, M. M., Fitch, G. M., & Goodall, J. L. (2019). Using Random Forest Classification and Nationally Available Geospatial Data to Screen for Wetlands over Large Geographic Regions. *Water*, 11(6), 1158. <https://doi.org/10.3390/w11061158>

References

- Fernández, A., García, S., del Jesus, M. J., & Herrera, F. (2008). A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18), 2378–2398. <https://doi.org/10.1016/j.fss.2007.12.023>
- Fernández, A., del Jesus, M. J., & Herrera, F. (2010). On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets. *Information Sciences*, 180(8), 1268–1291. <https://doi.org/10.1016/j.ins.2009.12.014>
- GDAL/OGR contributors. (2019). GDAL/OGR Geospatial Data Abstraction software Library. Retrieved from <https://gdal.org>
- Grabs, T., Seibert, J., Bishop, K., & Laudon, H. (2009). Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model. *Journal of Hydrology*, 373(1–2), 15–23. <https://doi.org/10.1016/j.jhydrol.2009.03.031>
- GRASS Development Team. (2017). Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2. Retrieved from <http://grass.osgeo.org>
- Grimaldi, S., Nardi, F., Benedetto, F. Di, Istanbuluoglu, E., & Bras, R. L. (2007). A physically-based method for removing pits in digital elevation models. *Advances in Water Resources*, 30(10), 2151–2158. <https://doi.org/10.1016/j.advwatres.2006.11.016>
- Guo, M., Li, J., Sheng, C., Xu, J., & Wu, L. (2017). A review of wetland remote sensing. *Sensors (Switzerland)*, 17(4), 1–36. <https://doi.org/10.3390/s17040777>
- Haralick, R. M., Watson, L. T., & Laffey, T. J. (1983). The Topographic Primal Sketch. *The International Journal of Robotics Research*, 2(1), 50–72. <https://doi.org/10.1177/027836498300200105>
- Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). Formal Basis for the Heuristic Determination eijj ., *Systems Science and Cybernetics*, (2), 100–107.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *IEEE conference on computer vision and pattern recognition* (pp. 770–778). Retrieved from <http://arxiv.org/abs/1512.03385>
- Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7), 1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527>
- Hogg, A. R., & Todd, K. W. (2007). Automated discrimination of upland and wetland using terrain derivatives. *Canadian Journal of Remote Sensing*, 33(July), S68–S83. <https://doi.org/10.5589/m07-049>
- Holmgren, P. (1994). Multiple flow direction algorithms for runoff modelling in grid based elevation models: An empirical evaluation. *Hydrological Processes*, 8(4), 327–334. <https://doi.org/10.1002/hyp.3360080405>
- Hooshyar, M., Wang, D., Kim, S., Medeiros, S. C., & Hagen, S. C. (2016). Valley and channel networks extraction based on local topographic curvature and k-means clustering of contours. *Water Resources Research*, 52, 8081–8102. <https://doi.org/10.1002/2015WR018479>.Received

References

- Hu, Y., Zhang, Q., Zhang, Y., & Yan, H. (2018). A Deep Convolution Neural Network Method for Land Cover Mapping: A Case Study of Qinhuangdao, China. *Remote Sensing*, *10*(12), 2053. <https://doi.org/10.3390/rs10122053>
- Jenson, S. K., & Domingue, J. O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing*, *54*(11), 1593–1600. [https://doi.org/0099-1112/88/5411-1593\\$02.25/0](https://doi.org/0099-1112/88/5411-1593$02.25/0)
- Jones, E., Oliphant, T., Peterson, P., & others. (2001). SciPy: Open source scientific tools for Python. Retrieved from <http://www.scipy.org/>
- Jyotsna, R., & Haff, P. K. (1997). Microtopography as an indicator of modern hillslope diffusivity in arid terrain. *Geology*, *25*(8), 695–698. Retrieved from [http://dx.doi.org/10.1130/0091-7613\(1997\)025%3C0695:MAAIOM%3E2.3.CO](http://dx.doi.org/10.1130/0091-7613(1997)025%3C0695:MAAIOM%3E2.3.CO)
- Keilwagen, J., Grosse, I., & Grau, J. (2014). Area under precision-recall curves for weighted and unweighted data. *PLoS ONE*, *9*(3), 1–13. <https://doi.org/10.1371/journal.pone.0092209>
- Kemker, R., Salvaggio, C., & Kanan, C. (2018). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*(April), 60–77. <https://doi.org/10.1016/j.isprsjprs.2018.04.014>
- Kim, M., Warner, T. A., Madden, M., & Atkinson, D. S. (2011). Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *International Journal of Remote Sensing*, *32*(10), 2825–2850.
- Klemas, V. (2011). Remote Sensing of Wetlands: Case Studies Comparing Practical Techniques. *Journal of Coastal Research*, *27*(3), 418–427. <https://doi.org/10.2112/JCOASTRES-D-10-00174.1>
- Kloiber, S. M., Macleod, R. D., Smith, A. J., Knight, J. F., & Huberty, B. J. (2015). A Semi-Automated, Multi-Source Data Fusion Update of a Wetland Inventory for East-Central Minnesota, USA. *Wetlands*, *35*(2), 335–348. <https://doi.org/10.1007/s13157-014-0621-3>
- Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, *50*(5), 363–370. <https://doi.org/10.1007/BF00336961>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. <https://doi.org/10.1145/3065386>
- Lang, M., & McCarty, G. (2014). Light Detection and Ranging (LiDAR) for Improved Mapping of Wetland Resources and Assessment of Wetland Conservation Projects, (September), 7. Retrieved from http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1260970.pdf
- Lang, M., McCarty, G., Oesterling, R., & Yeo, I. Y. (2013). Topographic metrics for improved mapping of forested wetlands. *Wetlands*, *33*(1), 141–155. <https://doi.org/10.1007/s13157-012-0359-8>

References

- Lanni, C., McDonnell, J. J., & Rigon, R. (2011). On the relative role of upslope and downslope topography for describing water flow path and storage dynamics: A theoretical analysis. *Hydrological Processes*, 25(25), 3909–3923. <https://doi.org/10.1002/hyp.8263>
- Lashermes, B., Foufoula-Georgiou, E., & Dietrich, W. E. (2007). Channel network extraction from high resolution topography using wavelets. *Geophysical Research Letters*, 34(23), 2–7. <https://doi.org/10.1029/2007GL031140>
- Latifovic, R., Pouliot, D., & Campbell, J. (2018). Assessment of Convolution Neural Networks for Surficial Geology Mapping in the South Rae Geological Region, Northwest Territories, Canada. *Remote Sensing*, 10(2), 307. <https://doi.org/10.3390/rs10020307>
- Lecun, Y. (1999). THE MNIST DATABASE of handwritten digits. Retrieved from <https://ci.nii.ac.jp/naid/10027939599/en/>
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., & others. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Lee, C. suk, Sohn, E., Park, J. D., & Jang, J.-D. (2019). Estimation of soil moisture using deep learning based on satellite data: a case study of South Korea. *GIScience & Remote Sensing*, 56(1), 43–67. <https://doi.org/10.1080/15481603.2018.1489943>
- Lidberg, W., Nilsson, M., Lundmark, T., & Ågren, A. M. (2017). Evaluating preprocessing methods of digital elevation models for hydrological modelling. *Hydrological Processes*, 31(26), 4660–4668. <https://doi.org/10.1002/hyp.11385>
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–755).
- Lindsay, J. B. (2016). Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrological Processes*, 30(6), 846–857. <https://doi.org/10.1002/hyp.10648>
- Lindsay, J. B., & Creed, I. F. (2005). Removal of artifact depressions from digital elevation models: Towards a minimum impact approach. *Hydrological Processes*, 19(16), 3113–3126. <https://doi.org/10.1002/hyp.5835>
- Liu, T., Abd-Elrahman, A., Morton, J., & Wilhelm, V. L. (2018). Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience and Remote Sensing*, 55(2), 243–264. <https://doi.org/10.1080/15481603.2018.1426091>
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).
- Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 277–293. <https://doi.org/10.1016/j.isprsjprs.2017.06.001>

References

- MacMillan, R. A., Martin, T. C., Earle, T. J., & McNabb, D. H. (2003). Automated analysis and classification of landforms using high-resolution digital elevation data: applications and issues. *Canadian Journal of Remote Sensing*, 29(5), 592–606. <https://doi.org/10.5589/m03-031>
- Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., & Zhang, Y. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing*, 10(7), 1119. <https://doi.org/10.3390/rs10071119>
- Mainguy, Y., Birch, J. B., & Watson, L. T. (1995). A robust variable order facet model for image data. *Machine Vision and Applications*, 8(3), 141–162. <https://doi.org/10.1007/BF01215810>
- Metz, M., Mitasova, H., & Harmon, R. S. (2011). Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrology and Earth System Sciences*, 15(2), 667–678. <https://doi.org/10.5194/hess-15-667-2011>
- Miao, X., Heaton, J. S., Zheng, S., Charlet, D. A., & Liu, H. (2012). Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data. *International Journal of Remote Sensing*, 33(6), 1823–1849.
- Millard, K., & Richardson, M. (2013). Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR-SAR fusion using a random forest classifier. *Canadian Journal of Remote Sensing*, 39(4), 290–307. <https://doi.org/10.5589/m13-038>
- Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing*, 7(7), 8489–8515. <https://doi.org/10.3390/rs70708489>
- Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1), 3–30. <https://doi.org/10.1002/hyp.3360050103>
- Morrissey, L. a., & Sweeney, W. R. (2006). Assessment of the National Wetlands Inventory: Implications for wetlands protection. *Geographic Information Systems and Water Resources IV AWRA Spring Specialty Conference, Houston, TX. May 8-10, 2006*, 2006(802), 1–6.
- Mui, A., He, Y., & Weng, Q. (2015). An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 30–46. <https://doi.org/10.1016/j.isprsjprs.2015.08.005>
- Murphy, P. N. C., Ogilvie, J., Connor, K., & Arp, P. A. (2007). Mapping wetlands: A comparison of two different approaches for New Brunswick, Canada. *Wetlands*, 27(4), 846–854. [https://doi.org/10.1672/0277-5212\(2007\)27\[846:MWACOT\]2.0.CO;2](https://doi.org/10.1672/0277-5212(2007)27[846:MWACOT]2.0.CO;2)
- Murphy, P. N. C., Ogilvie, J., & Arp, P. (2009). Topographic modelling of soil moisture conditions: a comparison and verification of two models. *European Journal of Soil Science*, 60(1), 94–109. <https://doi.org/10.1111/j.1365-2389.2008.01094.x>

References

- Murphy, P. N. C., Ogilvie, J., Meng, F. R., White, B., Bhatti, J. S., & Arp, P. A. (2011). Modelling and mapping topographic variations in forest soils at high resolution: A case study. *Ecological Modelling*, *222*(14), 2314–2332. <https://doi.org/10.1016/j.ecolmodel.2011.01.003>
- O’Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, *28*(3), 323–344. [https://doi.org/10.1016/S0734-189X\(84\)80011-0](https://doi.org/10.1016/S0734-189X(84)80011-0)
- O’Neil, G. L., Goodall, J. L., & Watson, L. T. (2018). Evaluating the potential for site-specific modification of LiDAR DEM derivatives to improve environmental planning-scale wetland identification using Random Forest classification. *Journal of Hydrology*, *559*, 192–208. <https://doi.org/10.1016/j.jhydrol.2018.02.009>
- O’Neil, G. L., Saby, L., Band, L. E., & Goodall, J. L. (2019). Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model. *Water Resources Research*, 2019WR024784. <https://doi.org/10.1029/2019WR024784>
- Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, *9*(3), 10–20. <https://doi.org/10.1109/MCSE.2007.58>
- Oltean, G. S., Comeau, P. G., & White, B. (2016). Linking the depth-to-water topographic index to soil moisture on boreal forest sites in Alberta. *Forest Science*, *62*(2), 154–165. <https://doi.org/10.5849/forsci.15-054>
- Ozesmi, S. L., & Bauer, M. E. (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management*, *10*(5), 381–402.
- Page, R. W., & Wilcher, L. S. (1990). Memorandum of Agreement Between the Environmental Protection Agency and the Department of the Army concerning the determination of mitigation under the Clean Water Act, Section 404 (b)(1) Guidelines. *Washington, DC, USA*.
- Passalacqua, P., Do Trung, T., Fofoula-Georgiou, E., Sapiro, G., & Dietrich, W. E. (2010). A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths. *Journal of Geophysical Research*, *115*(F1), F01002. <https://doi.org/10.1029/2009JF001254>
- Passalacqua, P., Tarolli, P., & Fofoula-Georgiou, E. (2010). Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape. *Water Resources Research*, *46*(11), 1–17. <https://doi.org/10.1029/2009WR008812>
- Passalacqua, P., Belmont, P., & Fofoula-Georgiou, E. (2012). Automatic geomorphic feature extraction from lidar in flat and engineered landscapes. *Water Resources Research*, *48*(3), 1–18. <https://doi.org/10.1029/2011WR010958>
- Passalacqua, P., Belmont, P., Staley, D. M., Simley, J. D., Arrowsmith, J. R., Bode, C. A., et al. (2015). Analyzing high resolution topography for advancing the understanding of mass and energy transfer through landscapes: A review. *Earth-Science Reviews*, *148*, 174–193. <https://doi.org/10.1016/j.earscirev.2015.05.012>

References

- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., et al. (2017). Automatic Differentiation in PyTorch. In *NIPS Autodiff Workshop*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Pelletier, J. D. (2013). A robust, two-parameter method for the extraction of drainage networks from high-resolution digital elevation models (DEMs): Evaluation using synthetic and real-world DEMs. *Water Resources Research*, *49*(1), 75–89. <https://doi.org/10.1029/2012WR012452>
- Perona, P., & Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(7), 629–639. <https://doi.org/10.1109/34.56205>
- Planchon, O., & Darboux, F. (2002). A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, *46*(2–3), 159–176. [https://doi.org/10.1016/S0341-8162\(01\)00164-3](https://doi.org/10.1016/S0341-8162(01)00164-3)
- PyGeoNet. (2019). Retrieved August 1, 2018, from <https://github.com/passaH2O/GeoNet>
- Rampi, L. P., Knight, J. F., & Pelletier, K. C. (2014). Wetland Mapping in the Upper Midwest United States. *Photogrammetric Engineering & Remote Sensing*, *80*(5), 439–448. <https://doi.org/10.14358/pers.80.5.439>
- Rezaee, M., Mahdianpari, M., Zhang, Y., & Salehi, B. (2018). Deep Convolutional Neural Network for Complex Wetland Classification Using Optical Remote Sensing Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(9), 3030–3039. <https://doi.org/10.1109/JSTARS.2018.2846178>
- Richardson, M. C., Fortin, M. J., & Branfireun, B. A. (2009). Hydrogeomorphic edge detection and delineation of landscape functional units from lidar digital elevation models. *Water Resources Research*, *45*(10), 1–18. <https://doi.org/10.1029/2008WR007518>
- Rieger, W. (1998). A phenomenon-based approach to upslope contributing area and depressions in DEMs. *Hydrological Processes*, *12*(6), 857–872. [https://doi.org/10.1002/\(SICI\)1099-1085\(199805\)12:6<857::AID-HYP659>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1099-1085(199805)12:6<857::AID-HYP659>3.0.CO;2-B)
- Rodhe, A., & Seibert, J. (1999). Wetland occurrence in relation to topography: A test of topographic indices as moisture indicators. *Agricultural and Forest Meteorology*, *98–99*, 325–340. [https://doi.org/10.1016/S0168-1923\(99\)00104-5](https://doi.org/10.1016/S0168-1923(99)00104-5)
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93–104.
- Sangireddy, H., Stark, C. P., Kladzyk, A., & Passalacqua, P. (2016). GeoNet: An open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data. *Environmental Modelling and Software*, *83*, 58–73. <https://doi.org/10.1016/j.envsoft.2016.04.026>

References

- Scikit-learn Developers. (2017a). Ensemble Methods. Retrieved from <http://scikit-learn.org/stable/modules/ensemble.html#forest>
- Scikit-learn Developers. (2017b). Model evaluation: quantifying the quality of predictions. Retrieved from http://scikit-learn.org/stable/modules/model_evaluation.html
- Scott, G. J., England, M. R., Starns, W. A., Marcum, R. A., & Davis, C. H. (2017). Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geoscience and Remote Sensing Letters*, *14*(4), 549–553. <https://doi.org/10.1109/LGRS.2017.2657778>
- Seibert, J., Bishop, K. H., & Nyberg, L. (1997). A test of TOPMODEL's ability to predict spatially distributed groundwater levels. *Hydrol. Process*, *11*(February 1996), 1131–1144. <https://doi.org/10.1021/cg200410b>
- Serran, J. N., & Creed, I. F. (2016). New mapping techniques to estimate the preferential loss of small wetlands on prairie landscapes. *Hydrological Processes*, *30*(3), 396–409. <https://doi.org/10.1002/hyp.10582>
- Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research*, *165*, 33–56. [https://doi.org/10.1016/S0079-6123\(06\)65004-8](https://doi.org/10.1016/S0079-6123(06)65004-8)
- Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research*, *54*(11), 8558–8593. <https://doi.org/10.1029/2018WR022643>
- Silburt, A., Ali-dib, M., Zhu, C., & Jackson, A. (2018). Lunar Crater Identification via Deep Learning.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.
- Snyder, G. I., & Lang, M. (2012). Significance of a 3D Elevation Program to wetland mapping. *National Wetlands Newsletter*, *34*(5), 11–15. Retrieved from <http://pubs.er.usgs.gov/publication/70193349>
- Soil Survey Staff. (2017a). Official Soil Series Descriptions. Retrieved August 1, 2017, from online
- Soil Survey Staff. (2017b). Soil Survey Geographic (SSURGO) Database for Virginia. Retrieved August 1, 2017, from online
- Sørensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*, *10*(1), 101–112. <https://doi.org/10.5194/hess-10-101-2006>
- Sørensen, Rasmus, & Seibert, J. (2007). Effects of DEM resolution on the calculation of topographical indices: TWI and its components. *Journal of Hydrology*, *347*(1–2), 79–89. <https://doi.org/10.1016/j.jhydrol.2007.09.001>

References

- Stivaktakis, R., Tsagkatakis, G., & Tsakalides, P. (2019). Deep Learning for Multilabel Land Cover Scene Categorization Using Data Augmentation. *IEEE Geoscience and Remote Sensing Letters*, 16(7), 1–5. <https://doi.org/10.1109/lgrs.2019.2893306>
- Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12), 3358–3378. <https://doi.org/10.1016/j.patcog.2007.04.009>
- Tarboton, D. G. (1991). On the extraction of channel networks from digital elevation data. *Hydrological Processes*, 5(1), 81–100. <https://doi.org/10.1002/hyp.3360050107>
- Tarboton, D. G. (1997a). A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models. *Water Resources Research*, 33(2), 309–319.
- Tarboton, D. G. (1997b). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, 33(2), 309–319. <https://doi.org/10.1029/96WR03137>
- Tarboton, D. G., & Ames, D. P. (2001). Advances in the Mapping of Flow Networks from Digital Elevation Data. *The World Water and Environmental Resources Congress*, 111(435), 166–175. [https://doi.org/10.1061/40569\(2001\)166](https://doi.org/10.1061/40569(2001)166)
- Tesfa, T. K., Tarboton, D. G., Watson, D. W., Schreuders, K. A. T., Baker, M. E., & Wallace, R. M. (2011). Extraction of hydrological proximity measures from DEMs using parallel processing. *Environmental Modelling and Software*, 26(12), 1696–1709. <https://doi.org/10.1016/j.envsoft.2011.07.018>
- Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sensing*, 8(11), 954. <https://doi.org/10.3390/rs8110954>
- US Corps of Engineers. (1987). Corps of Engineers Wetlands Delineation Manual and Regional Supplements. US Army Engineer Waterways Experiment Station Vicksburg, Mississippi. Retrieved from <http://www.epa.gov/cwa-404/section-404-clean-water-act-how-wetlands-are-defined-and-identified>
- USDA. (2019). Soil Data Viewer. Retrieved August 1, 2017, from <https://www.nrcs.usda.gov/>
- USGS. (2019a). USGS National Elevation Dataset (NED). Retrieved August 1, 2017, from <https://catalog.data.gov/dataset/usgs-national-elevation-dataset-ned>
- USGS. (2019b). USGS National Hydrography Dataset. Retrieved August 1, 2017, from <https://www.usgs.gov/core-science-systems/ngp/national-hydrography/national-hydrography-dataset>
- Wang, L., & Liu, H. (2007). An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science*, 20(2), 193–213. <https://doi.org/10.1080/13658810500433453>

References

- White, B., Ogilvie, J., Campbell, D. M. H., Hiltz, D., Gauthier, B., Chisholm, H. K., et al. (2012). Using the Cartographic Depth-to-Water Index to Locate Small Streams and Associated Wet Areas across Landscapes. *Canadian Water Resources Journal*, 37(4), 333–347. <https://doi.org/10.4296/cwrj2011-909>
- Wilken, E., Jiménez Nava, F., & Griffith, G. (2011). *North American Terrestrial Ecoregions—Level III. North American Commission for Environmental Cooperation*. Retrieved from <https://www.epa.gov/eco-research/ecoregions-north-america>
- Woodrow, K., Lindsay, J. B., & Berg, A. A. (2016). Evaluating DEM conditioning techniques, elevation source data, and grid resolution for field-scale hydrological parameter extraction. *Journal of Hydrology*, 540, 1022–1029. <https://doi.org/10.1016/j.jhydrol.2016.07.018>
- WorldView Solutions Inc. (2016). *Technical Plan of Operations: Virginia Statewide Land Cover Data Development*. Richmond, VA.
- Xu, Z., Guan, K., Casler, N., Peng, B., & Wang, S. (2018). A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144(August), 423–434. <https://doi.org/10.1016/j.isprsjprs.2018.08.005>
- Zhang, L., Zhang, L., & Du, B. (2016). Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40. <https://doi.org/10.1109/MGRS.2016.2540798>
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. Retrieved from <http://arxiv.org/abs/1512.04150>
- Zhu, J., & Pierskalla, W. P. (2016). Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *Journal of Hydrology*, 533, 343–352. <https://doi.org/10.1016/j.jhydrol.2015.12.012>
- Zinko, U., Seibert, J., Dynesius, M., & Nilsson, C. (2005). Plant species numbers predicted by a topography-based groundwater flow index. *Ecosystems*, 8(4), 430–441. <https://doi.org/10.1007/s10021-003-0125-0>

Appendix

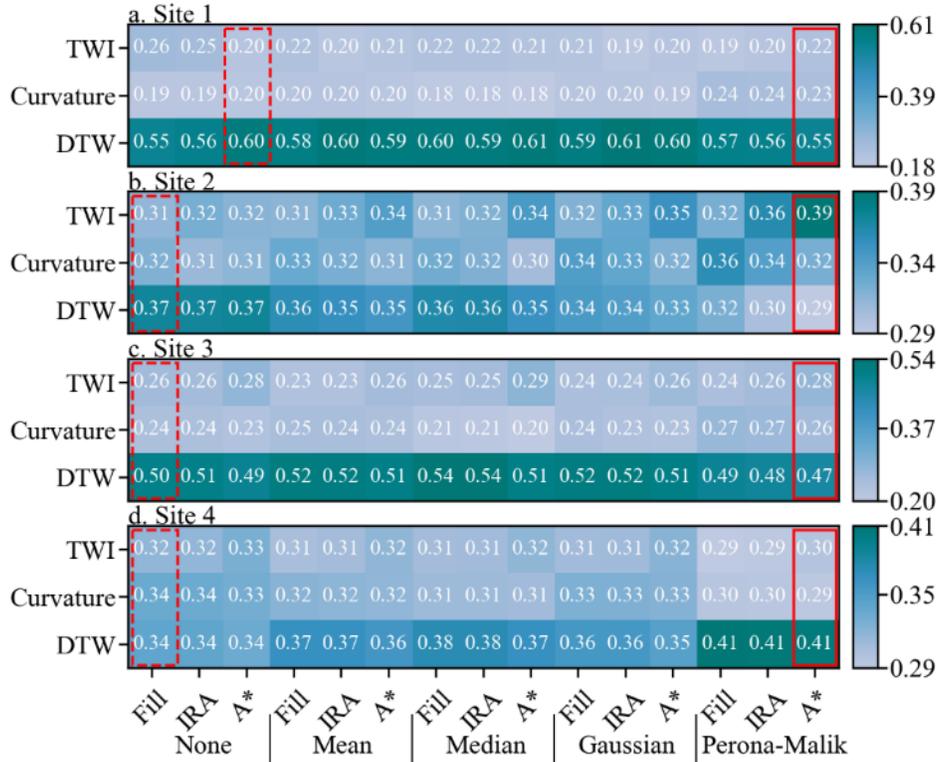


Figure A1. Variable importance (mean decrease in accuracy) for each conditioning and smoothing (generalized scheme) combination. Best performing (solid line) and worst performing (dashed line) models are boxed.

Appendix

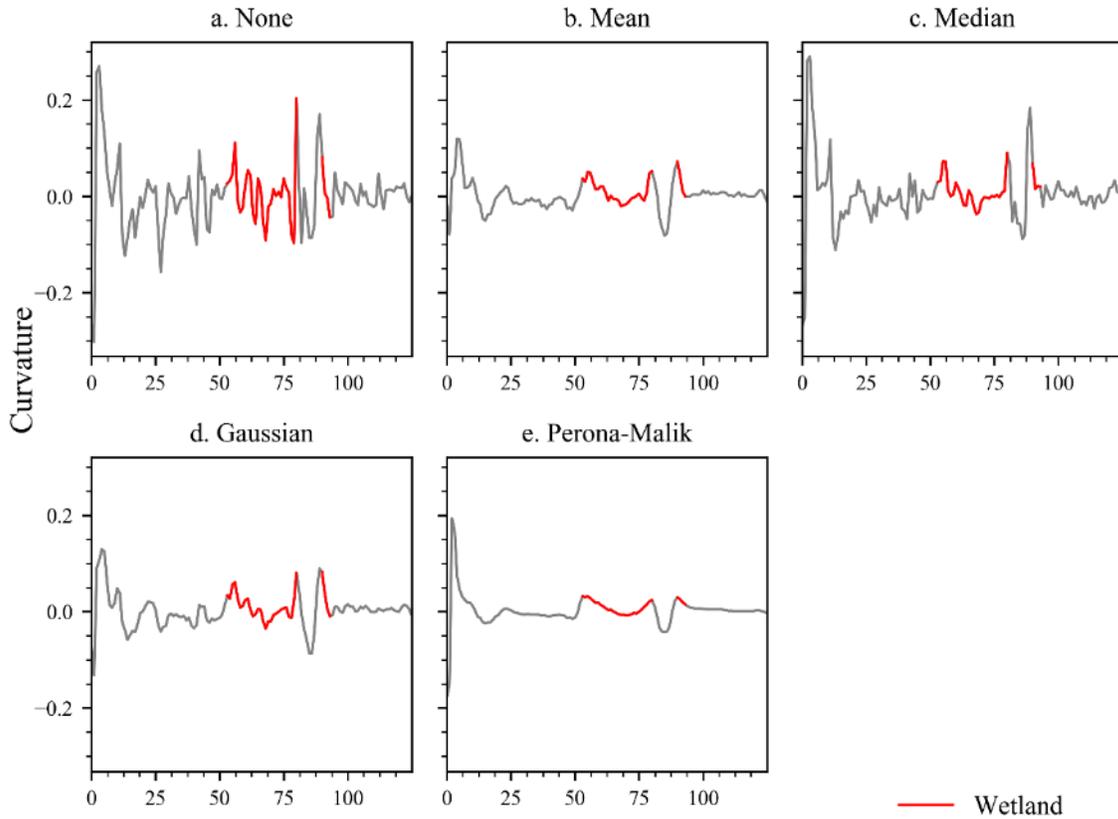


Figure A2. Curvature transect encompassing two wetlands in a portion of Site 1, calculated from unsmoothed (a) and smoothed DEMs (mean (b), median (c), Gaussian (d), and Perona-Malik (e)). The x-axis marks the position along a transect in meters.

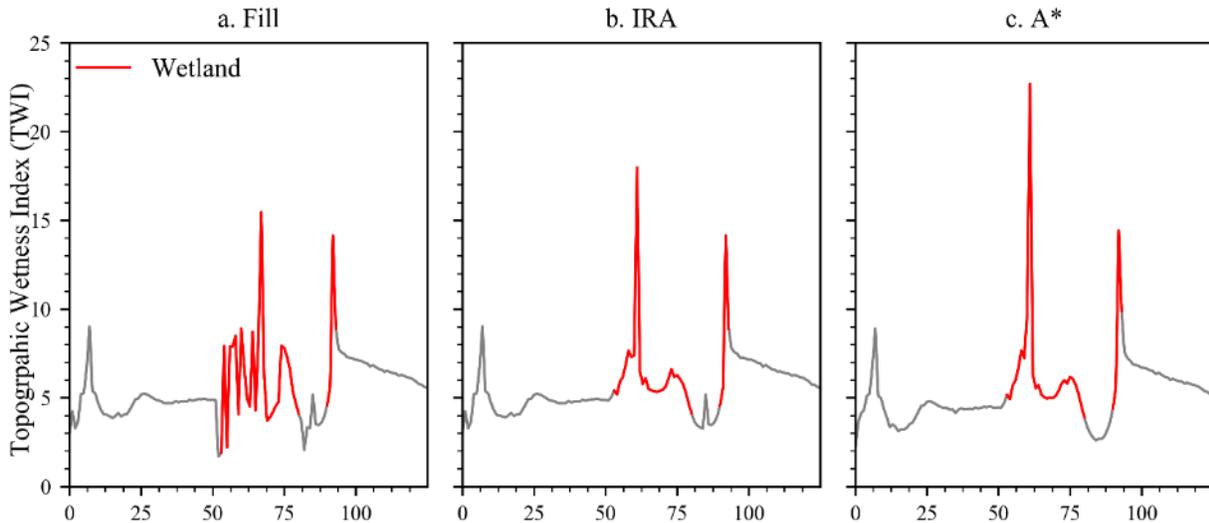


Figure A3. Topographic wetness index (TWI) transect encompassing two wetlands in a portion of Site 1, calculated from DEMs conditioned using Fill (a), Impact Reduction Approach (IRA) (b), and A* least-cost path analysis (c). All TWI calculations shown incorporated Perona-Malik filtered slope grids, and the X-axis marks the position along transect in meters.

Appendix

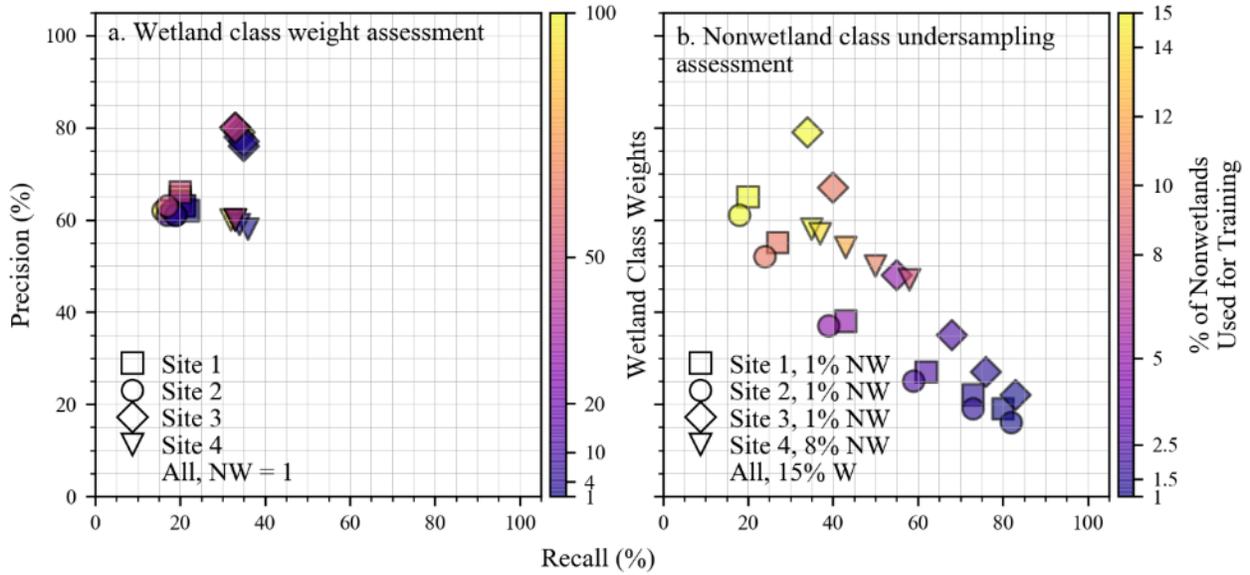


Figure A4. Response of the wetland identification model to increasing minority class weights (a) and undersampling the majority class (b). Colorbar ticks indicate the wetland class weights and nonwetland sampling percentages used to produce results in (a) and (b), respectively.

Appendix

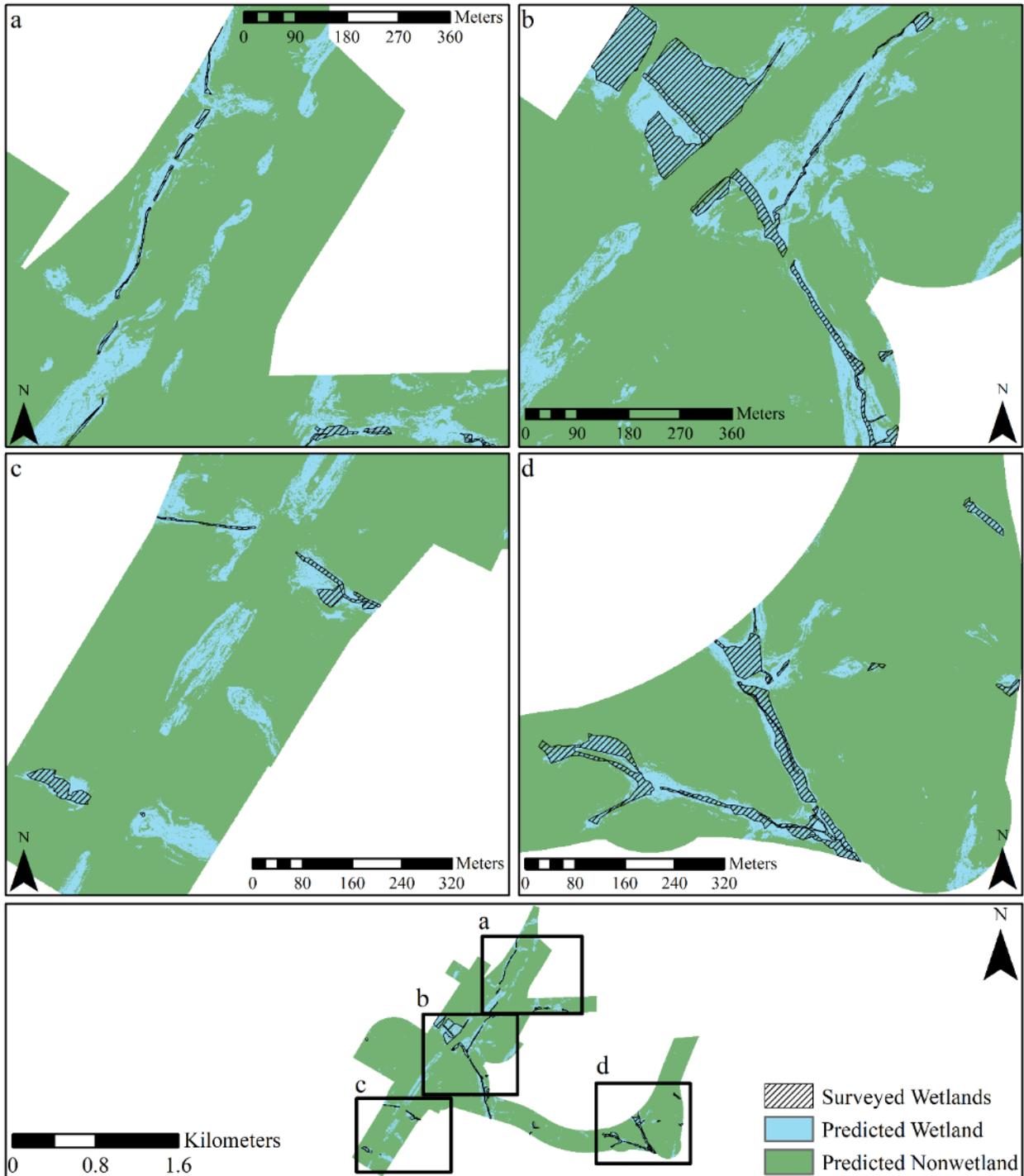


Figure A5. Extended results for Site 1 best performing model predictions, which are the product of median (50m scale) smoothing and A* conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and Perona-Malik (100 iterations) smoothing applied to DTW.

Appendix

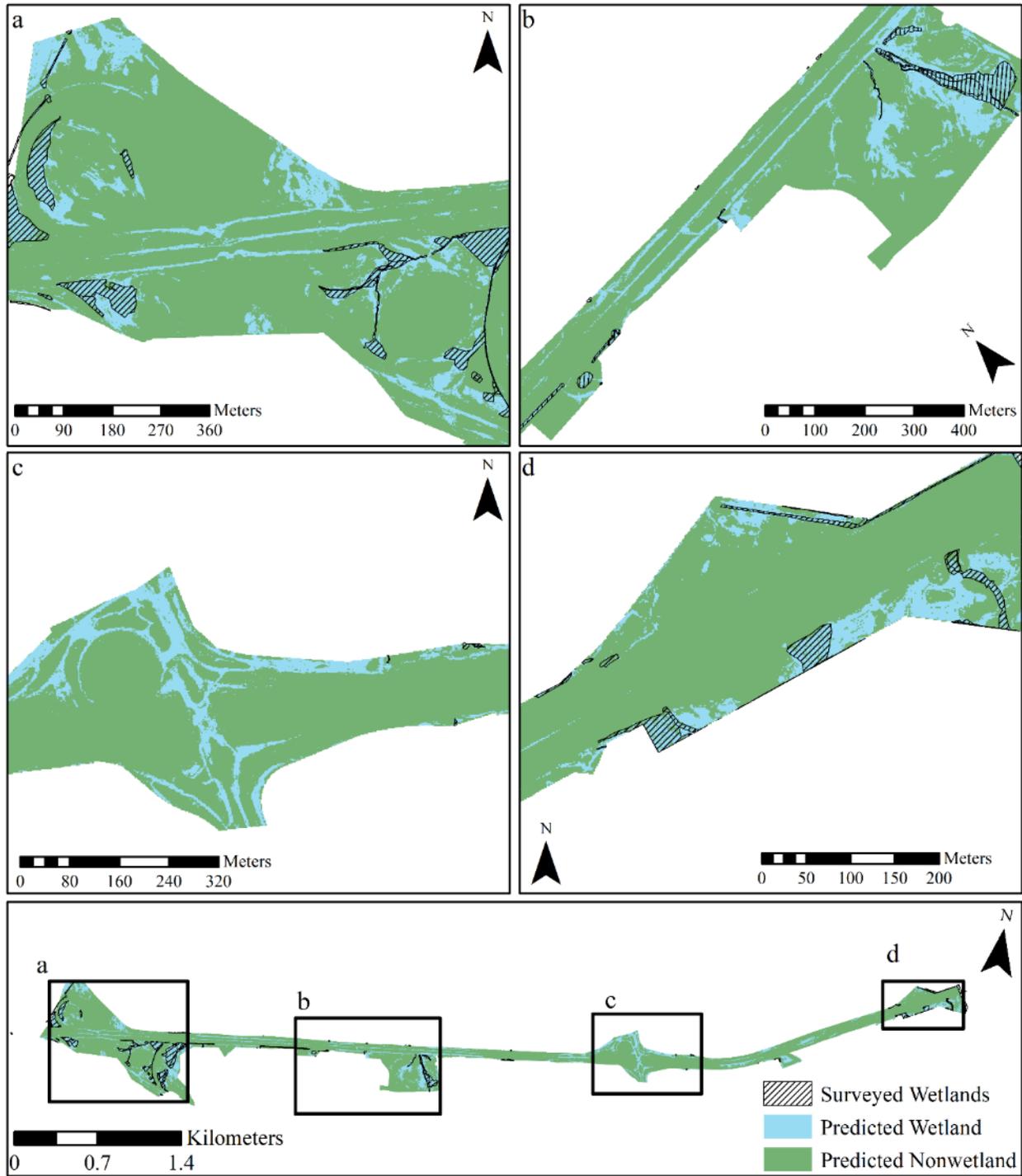


Figure A6. Extended results for Site 2 best performing model predictions, which are the product of Perona-Malik (100 iterations) smoothing and A^* conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and median (5m scale) smoothing applied to DTW.

Appendix

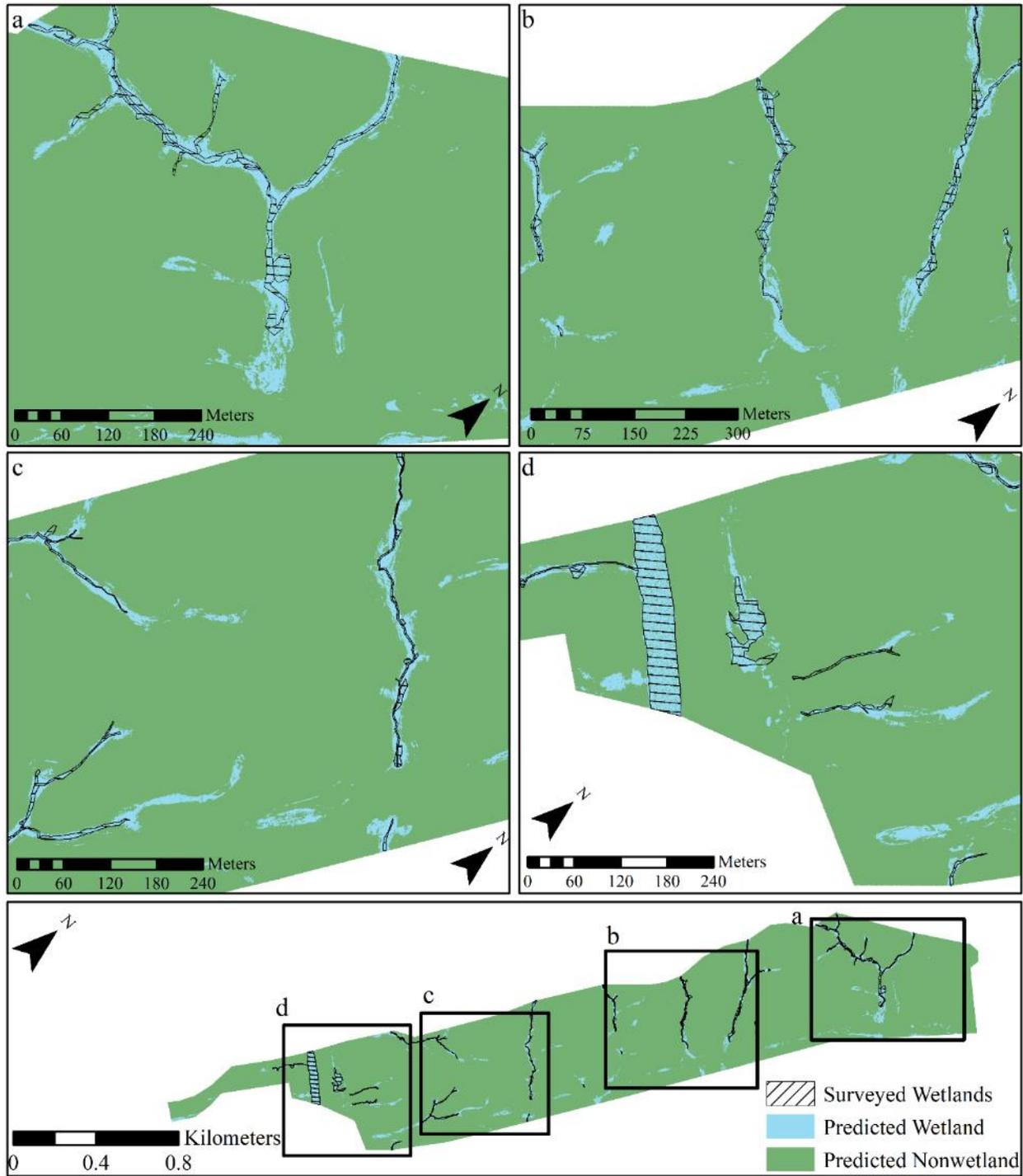


Figure A7. Extended results for Site 3 best performing model predictions, which are the product of median (25m scale) smoothing and A* conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and Perona-Malik (50 iterations) smoothing applied to DTW.

Appendix

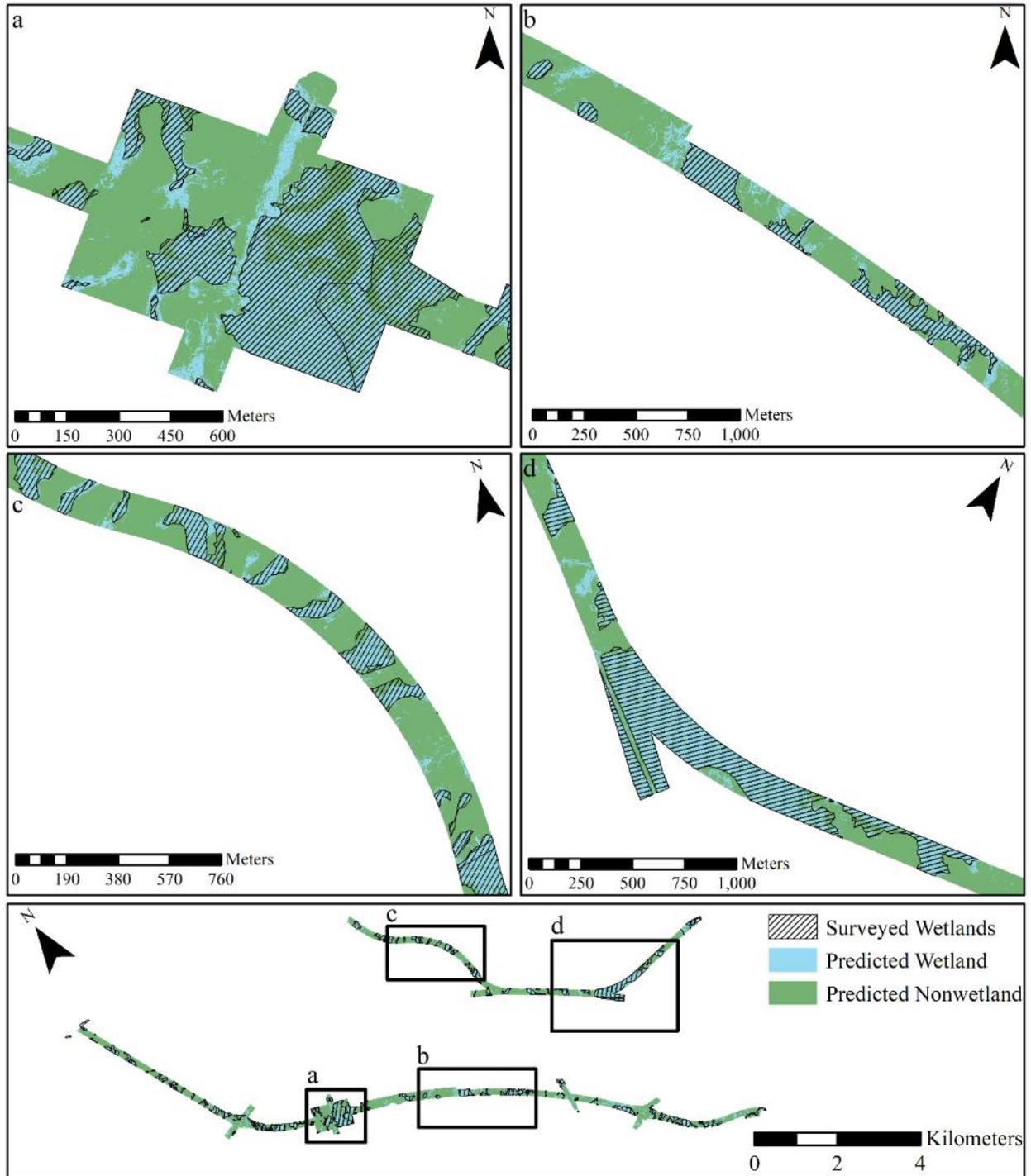


Figure A8. Extended results for Site 4 best performing model predictions, which are the product of Gaussian (100m scale) smoothing and A* conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and Gaussian (25m scale) smoothing applied to DTW.

Appendix

Table A1. Confusion matrix for Site 1 best performing model predictions, which are the product of median (50m scale) smoothing and A conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and Perona-Malik (100 iterations) smoothing applied to DTW.*

	Predicted wetlands (km ²)	Predicted nonwetlands (km ²)	Σ
Testing wetlands (km ²)	0.0596	0.0065	0.066
Testing nonwetlands (km ²)	0.2103	2.3651	2.575
Σ	0.270	2.372	2.64

Table A2. Confusion matrix for Site 2 best performing model predictions, which are the product of Perona-Malik (100 iterations) smoothing and A conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and median (5m scale) smoothing applied to DTW.*

	Predicted wetlands (km ²)	Predicted nonwetlands (km ²)	Σ
Testing wetlands (km ²)	0.0613	0.0109	0.072
Testing nonwetlands (km ²)	0.2114	1.1371	1.349
Σ	0.273	1.148	1.42

Table A3. Confusion matrix for Site 3 best performing model predictions, which are the product of median (25m scale) smoothing and A conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and Perona-Malik (50 iterations) smoothing applied to DTW.*

	Predicted wetlands (km ²)	Predicted nonwetlands (km ²)	Σ
Testing wetlands (km ²)	0.0256	0.0025	0.028
Testing nonwetlands (km ²)	0.0659	1.7044	1.770
Σ	0.092	1.707	1.80

Table A4. Confusion matrix for Site 4 best performing model predictions, which are the product Gaussian (100m scale) smoothing and A conditioning applied to TWI, Gaussian (100m scale) smoothing applied to curvature, and Gaussian (25m scale) smoothing applied to DTW.*

	Predicted wetlands (km ²)	Predicted nonwetlands (km ²)	Σ
Testing wetlands (km ²)	1.1256	0.2606	1.386
Testing nonwetlands (km ²)	0.5155	2.9113	3.427
Σ	1.641	3.172	4.81