Preprint for the published paper:
Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography‐Based Wetland Identification Model
G. L. O'Neil L. Saby L. E. Band, and J. L. Goodall
Water Resources Research, 55(5), doi:10.1029/2019WR024784, May 2019

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

Gina L. O'Neil[1], Linnea Saby[1], Lawrence E. Band[1, 2], and Jonathan L. Goodall[1]

[1]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA 22904, USA.

[2]Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904, USA.

Corresponding author: Jonathan L. Goodall (goodall@virginia.edu)

Key Points:

- For four sites, we tested the effects of terrain preprocessing on a Random Forest model that uses LiDAR to delineate wetlands.
- Perona-Malik smoothing and A* conditioning performed best in all sites, and models further improved by individualizing smoothing by input.
- For all sites, the model detected most wetlands (81-91%) but with varying precision (22-69%), indicating its best use as a screening tool.

Abstract

Accurate and widely-available wetland inventories are needed for wetland conservation and environmental planning. We propose an open source, automated wetland identification model that relies primarily on Light Detection and Ranging (LiDAR) digital elevation models (DEMs). LiDAR DEMs are increasingly available and provide the resolution needed to map detailed topographic metrics and areas of likely soil saturation, but the choice of smoothing and conditioning techniques can significantly impact 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 (TWI), curvature, and cartographic depth-to-water index (DTW) 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 datasets 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.

Plain Language Summary

Accurate wetland inventories are needed for wetland protection and conservation. We propose an automated tool that locates wetlands using Light Detection and Ranging (LiDAR) digital elevation models (DEMs). LiDAR DEMs are increasingly available and show elevation changes that likely affect soil saturation. However, the ability of LiDAR DEMs to describe saturated areas is affected by smoothing and conditioning. Smoothing blurs DEMs to remove elevation changes that are too small to indicate features of interest, and conditioning ensures accurate simulation of hydrologic flow paths. The effects of different smoothing and conditioning methods on wetland mapping have not been studied. We tested how our wetland tool is influenced by five smoothing techniques and three conditioning techniques for four sites in Virginia, USA. We found that Perona-Malik smoothing and A* conditioning improved predictions and reduced tool runtime for all sites. Also, we found predictions could be further improved by varying smoothing parameters specific to each input. Using only elevation information, the wetland tool predicted 81-91% of true wetlands across our sites. The proportion of wetland predictions that were correct varied (ranging from 22 to 69% across sites). Overall, the results suggest strong potential for the model to support environmental groups to delineate wetlands.

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; Millard & Richardson, 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; Millard & Richardson, 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 et al., 2010a; 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; 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., 2010a, 2010b, 2012; 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., Metz et al., 2011; Lidberg et al., 2017): 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 et al. (2010a) 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).

2. Study areas and input data

2.1. Study areas

This analysis was completed for four study areas in Virginia, USA (Figure 1a). For each study area, the available data includes the extents of wetland surveys and the HUC 12 watershed (USGS, 2013) that encompasses the surveys (Figure 1b). 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 (EPA, 2013). Table 1 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.

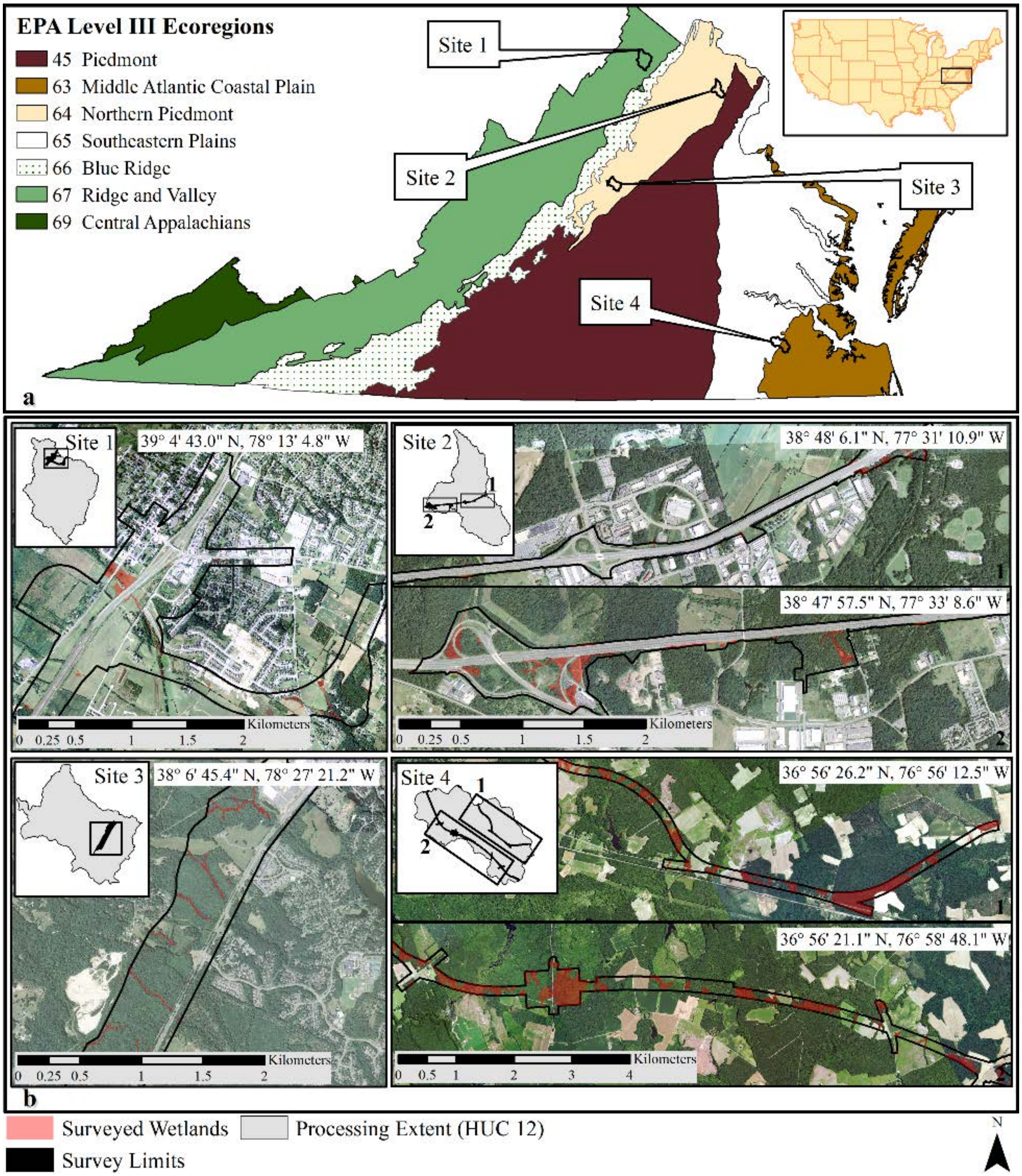

Figure 1. Four study areas spanning four level III ecoregions in Virginia, USA (a). Each study area includes the wetland survey limits, referred to as study sites, and the encompassing HUC 12 watershed, used as the processing extent (b).
Ecoregion data source: US EPA Office of Environmental Information
Aerial imagery data source: NAIP Digital Ortho Photo Image.

Table 1. 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.

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. 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.

3.1. 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 2). 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 2. 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.

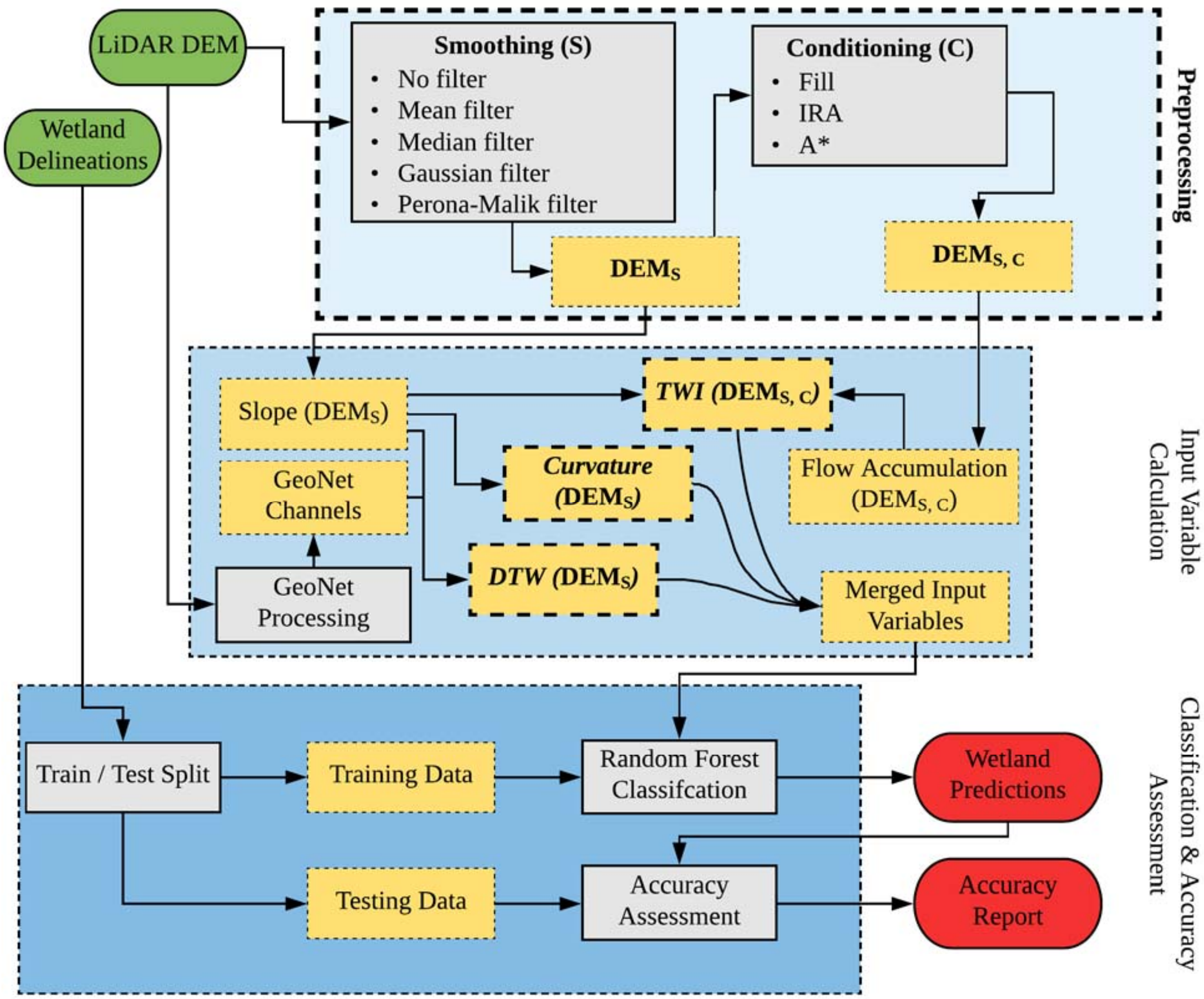

Figure 2. Workflow of the wetland identification model created through this research. Each combination of preprocessing techniques (bold font) was executed for this analysis. Green shapes indicate input data, grey shapes indicate processes, yellow shapes indicate intermediate output, and red shapes indicate final output.

3.2. Preprocessing

3.2.1. 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 et al., 2010a, 2012; 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 25m^2. 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 et al., 2010a). The Gaussian filter is defined as

$$h(x, y, \sigma) = h_o(x, y) * G(x, y; \sigma), \quad (1)$$

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

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

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,

$$\partial_t h(x, y, t) = \nabla \cdot [c(x, y, t) \nabla h], \quad (3)$$

where $h(x, y, t)$ is the elevation at time *t*, *c* is the diffusion coefficient, and ∇ is the gradient operator (1990). Eq. (3) 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 et al., 2010a, 2010b). While there are two possible forms of *c*, here we implemented

$$c = \frac{1}{1+\left(\frac{|\nabla h|}{\lambda}\right)^2}, \tag{4}$$

where λ is the edge stopping threshold (Perona & Malik, 1990). We chose the form of c in Eq. (4) 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 et al. (2010a) for channel network extraction. The time of forward diffusion (t in Eq. (3)) 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 et al., 2010a). 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 et al., 2010a; Sangireddy et al., 2016). To execute Perona-Malik smoothing, code from the PyGeoNet nonlinear filtering module, *pyGeoNet_nonlinear_fitler.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 et al., 2010a; Sangireddy et al., 2016).

3.2.2. 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. 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 Development Team, 2018).

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, 2006).

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

$$TWI = \ln\left(\frac{\alpha}{\tan\beta}\right), \tag{5}$$

where α is the specific catchment area (contributing area per unit contour length) and tan(β) 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, 1997), 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). 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

$$DTW\ (m) = \left[\sum \left(\frac{dz_i}{dx_i}\right) a\right] * x_p, \tag{6}$$

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, 2013). Instead, the surface water grid was generated using PyGeoNet (Version 2.0; Sangireddy et al., 2016). PyGeoNet employs a statistical analysis of curvature, and geodesic minimization principles to extract channel networks from elevation data (Passalacqua et al., 2010a; 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).

3.4. 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 1, 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.

3.4.1. 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.

3.4.2. 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, 2017b) was used for the RF classification.

3.4.3. 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

$$Recall = \frac{True\ wetland\ predictions}{Total\ true\ wetlands}. \tag{7}$$

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

$$Precision = \frac{True\ wetland\ predictions}{Total\ wetland\ predictions}. \quad (8)$$

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

$$AP = \sum_n (R_n - R_{n-1})P_n, \quad (9)$$

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.

4. Results

4.1. Effects of preprocessing techniques on model accuracy

Figure 3 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.

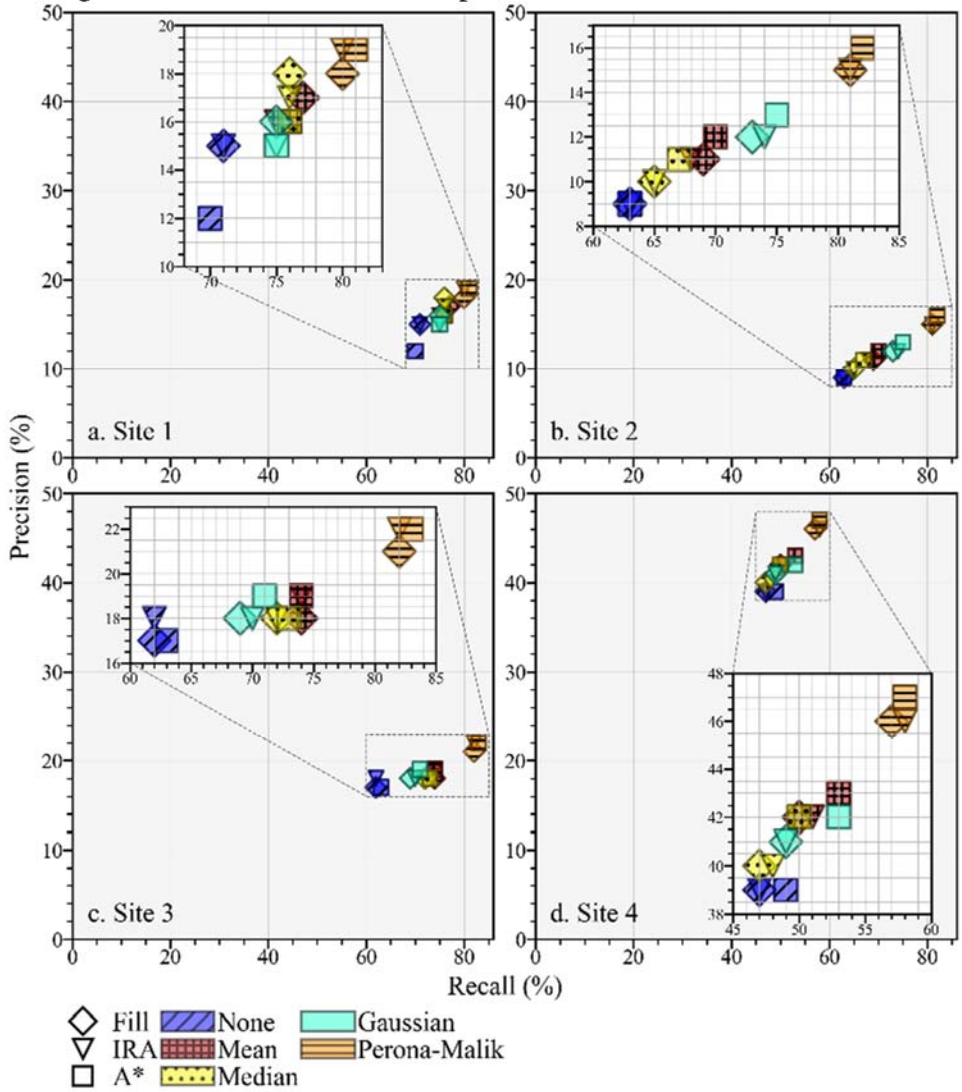

Figure 3. Wetland precision and recall resulting from each preprocessing technique combination across all study sites. Note the differences in x-scale and y-scale range.

Common trends in model performance due to smoothing and conditioning emerged despite differences in the accuracies. As seen in Figure 3, 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 S1.

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 3). 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 3). 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 S2.

Models incorporating the A* technique and those using Fill consistently resulted in the highest and lowest accuracies within groups of common smoothing, respectively (Figure 3). 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 3). 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 S3.

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 S4). 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 S4a). 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 S4b), 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.

5. Discussion

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; 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 et al. (2010b) 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 4). A* conditioning was applied to all TWI models.

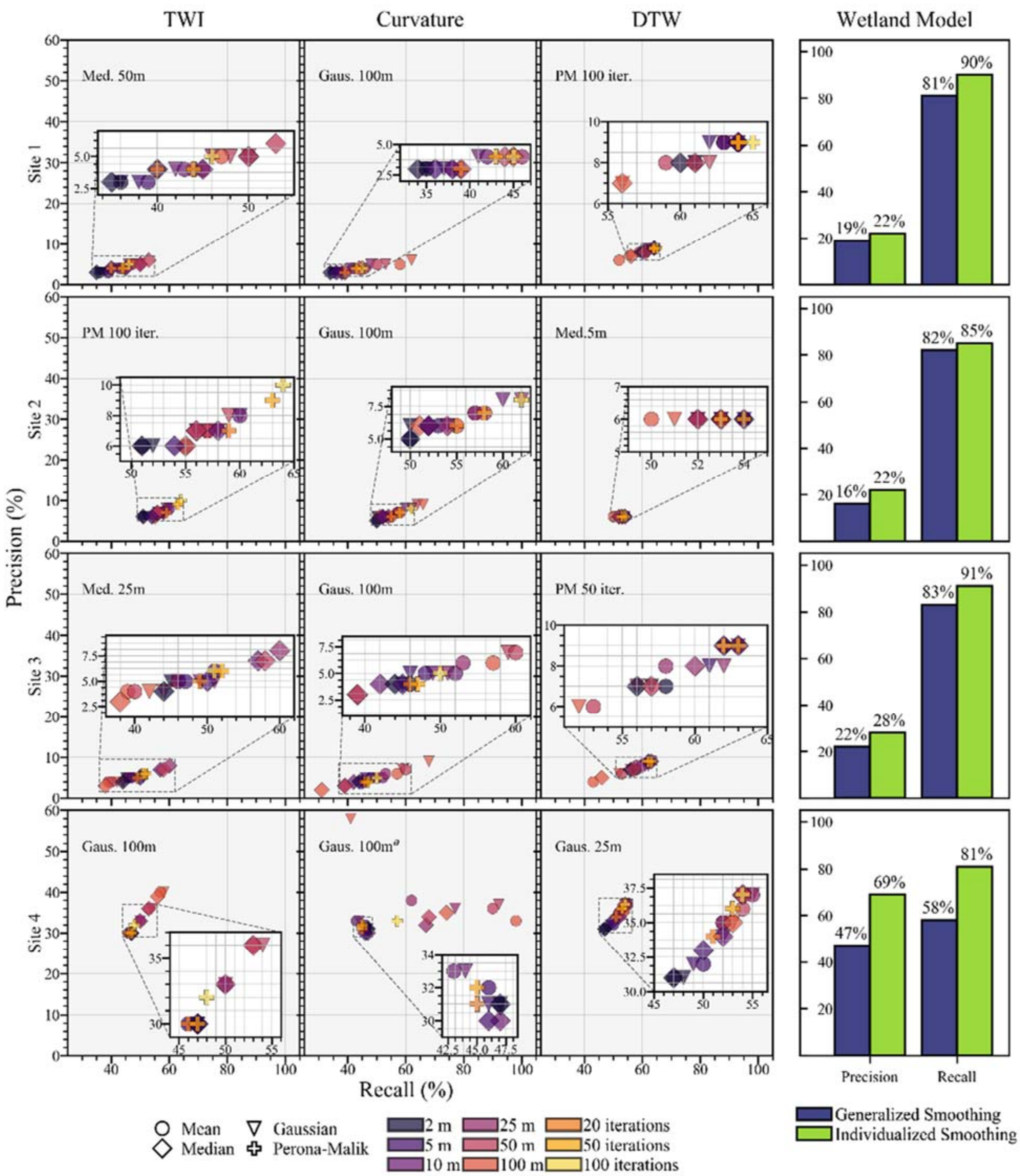

Figure 4. Effect of varying smoothing method and scale on wetland model accuracy. Scatter plots show the results for models trained on a single input, and annotations indicate the best performing smoothing formulation for that input. Bar plots show the results of wetland models (i.e., trained on three 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.

[a]Gaussian 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 4, 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; 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.

5.2. Improvements to wetland predictions due to preprocessing schemes

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 5 curves). The improvements from the worst performing models (Figure 5, 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 5, b1-b4). For Site 2 there was a drastic decrease in wetland likelihood within impervious areas compared to the worst performing model (Figure 5, 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 5, 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.

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 5 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 5, 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 5, 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 5, 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 5, 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 S5-S8 and corresponding confusion matrices are given in tables S1-S4.

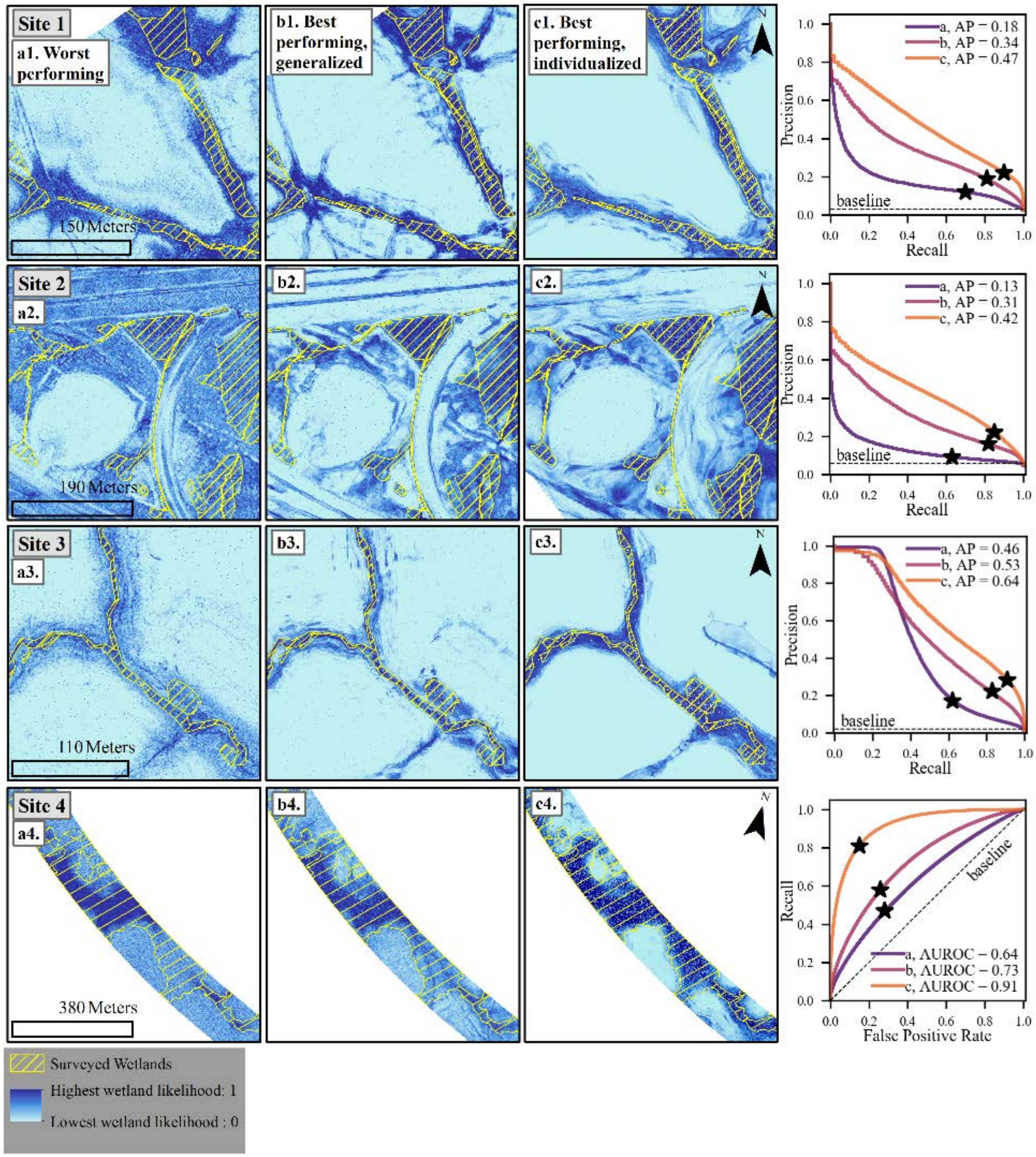

Figure 5. Wetland likelihoods resulting from different preprocessing configurations: worst generalized preprocessing, as described in Section 4.1 (a), best performing generalized preprocessing, as described in section 4.1 (b), and A* conditioning and best performing individualized smoothing, as described 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.

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, 2017) 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.

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 anticipate drainage through artificial structures. Including these flow paths by artificially lowering the DEM along built drainage paths and outlets 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.

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.

Acknowledgments

The authors wish to thank the Virginia Department of Transportation (VDOT) for supplying the wetland survey data (proprietary, may be available through personal communication with VDOT). Funding for this project was provided by the Department of Education through a Graduate Assistance in Areas of National Need (GAANN) grant. Computer resources were provided by Rivanna, the University of Virginia's high-performance computing system, and the Advanced Research Computing Services (ARCS) group. Python code used to execute the wetland identification tool is available on GitHub (https://github.com/uva-hydroinformatics/wetland_identification), and LiDAR data used in these analyses are available from http://vgin.maps.arcgis.com.

Table and figure captions

Table 1. Characteristics of each study site, including dominating 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.

Figure 1. 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.

Figure 2. 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.

Figure 3. Wetland precision and recall resulting from each preprocessing technique combination across all study sites. Note the differences in x-scale and y-scale range.

Figure 4. 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.
[a]Gaussian 100m, Gaussian 50m, and Mean 100m were considered in determining the best performing curvature formulation for Site 4.

Figure 5. 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.

References

Ågren, A. M., Lidberg, W., Strömgren, 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., 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)*. ACM, New York, NY, USA, 573-574. https://doi.org/10.1145/1076034.1076134

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

Batuwita, R., & Palade, V. (2010). Efficient resampling methods for training support vector machines with imbalanced datasets. Efficient resampling methods for training support vector machines with imbalanced datasets. *In The 2010 International Joint Conference on Neural Networks (IJCNN)*. Barcelona, Spain. 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

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

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

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

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

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

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*.

Environmental Laboratory. (1987). Corps of Engineers wetlands delineation manual. US Army Engineer Waterways Experiment Station Vicksburg, Mississippi.

Dronova, I. (2015). Object-based image analysis in wetland research: A review. *Remote Sensing*, *7*(5), 6380–6413. https://doi.org/10.3390/rs70506380

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

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 Development Team. (2018). GDAL - Geospatial Data Abstraction Library, Version 2.0.1., *Open Source Geospatial Foundation*. http://gdal.osgeo.org.

GeoNetPython [source code]. (2017). Retrieved January, 2018 from https://sites.google.com/site/geonethome/source-code.

GRASS Development Team. (2017). Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2., *Open Source Geospatial Foundation*. https://grass.osgeo.org.

Grimaldi, S., Nardi, F., Benedetto, F. Di, Istanbulluoglu, 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

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination. *IEEE transactions on Systems Science and Cybernetics*, 4(2), 100-107.

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

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

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. 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

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

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

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

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

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

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

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

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

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

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

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

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

Oliphant, T. E. (2006). A guide to NumPy, USA: Trelgol Publishing. http://www.numpy.org/

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

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., Foufoula-Georgiou, E., Sapiro, G., & Dietrich, W. E. (2010a). 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., & Foufoula-Georgiou, E. (2010b). 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., & Foufoula-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

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. http://scikit-learn.org.

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

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

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

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

Scikit-learn Developers. (2017a). Ensemble Methods. Retrieved August, 2018 from http://scikit-learn.org/stable/modules/ensemble.html#forest.

Scikit-learn Developers. (2017b). Model evaluation: quantifying the quality of predictions. Retrieved August 2018, from http://scikit-learn.org/stable/modules/model_evaluation.html.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. (2017). Web Soil Survey. Retrieved January 2017, from https://websoilsurvey.nrcs.usda.gov/.

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

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

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

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. (1997). 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

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

U.S. Environmental Protection Agency (EPA). (2013). Primary Distinguishing Characteristics of Level III Ecoregions of the Continental United States. https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states.

U.S. Geological Survey (USGS). (2013). National Hydrography Geodatabase: The National Map viewer. Retrieved December, 2017 from https://viewer.nationalmap.gov/viewer/nhd.html?p=nhd.

Virginia Information Technologies agency (VITA) (2016). Virginia Geographic Information Network (VGIN): Virginia LiDAR. Retrieved December, 2017 from http://vgin.maps.arcgis.com/apps/Viewer/index.html?appid=1e964be36b454a12a69a3ad0bc1473ce.

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

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

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

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

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