Deep Learning Using Physically-Informed Input Data for Wetland Identification

Gina L. O'Neil[1], Jonathan L. Goodall[1,2], Madhur Behl[2, 1], and Linnea Saby[1]

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

[2]Department of Computer Science, University of Virginia, Charlottesville, VA 22904, USA.

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

Highlights

- A deep learning workflow is described, able to segment wetlands from geospatial data

- Wetlands are segmented at a high spatial resolution for environmental planning

- The workflow resulted in accurate wetland predictions (91% recall and 57% precision)

- Training using physically-informed indicators outperformed using elevation directly

Software and Data Availability

Software created through this research along with documentation is available under an MIT license from https://github.com/uva-hydroinformatics/wetland_id. All input data required to run the model are publicly available through federal and state data providers. Wetland delineation

26 datasets used for training and evaluation was made available to the researchers through a

27 relationship with the Virginia Department of Transportation. To retrain the model for a new

28 landscape, similar wetland delineation data for that area may be required.

29

30 **Abstract**

31   Automated and accurate wetland identification algorithms are increasingly important for

32 wetland conservation and environmental planning. Deep learning for wetland identification is an

33 emerging field that shows promise for advancing these efforts. Deep learning is unique to

34 traditional machine learning techniques for its ability to consider the spatial context of object

35 characteristics within a landscape training the algorithms, which limits their application for many

36 environmental applications including wetland identification. Using four study sites across Virginia

37 with field delineated wetlands, we provide insight into the potential for deep learning for wetland

38 detection from limited, but typical, wetland delineation training data. Our proposed workflow

39 performs a wetland semantic segmentation using DeepNets, a deep learning architecture for remote

40 sensing data, and an input dataset consisting of high-resolution topographic indices and the

41 Normalized Difference Vegetation Index. Results show that models trained and evaluated for a

42 single site were able to achieve high accuracy (up to 91% recall and 56% precision) and similar

43 accuracy can be obtained for models trained across multiple sites (up to 91% recall and 57%

44 precision). Through this analysis we found that, across all sites, input data configurations taking

45 advantage of hydrologic properties derived from elevation data consistently outperformed models

46 using the elevation data directly, showing the benefit of physically-informed inputs in deep

47 learning training for wetland identification. By refining the wetland identification workflow

2

48   presented in this paper and collecting additional training data across landscapes, there is potential

49   for deep learning algorithms to support a range wetland conservation efforts.

50

Introduction

52   Wetlands are important ecosystems that are threatened by development, climate change,

53   and pollution (Klemas, 2011). Wetland loss is both a global (Davidson, 2014) and national

54   problem, as half of the wetlands of the conterminous U.S. have been lost since 1600 (Dahl et al.,

55   1991). In the U.S., federal regulations, such as Section 404 of the Clean Water Act, play an

56   important role in wetland protection. Laws require environmental impact assessments prior to land

57   development and water resources projects, which entails the creation of detailed wetland surveys

58   (Page & Wilcher, 1990). Conducting these surveys with the level of spatial resolution and accuracy

59   needed to abide by federal regulations and meet the goal of avoiding adverse impact to wetlands

60   can be time-consuming and costly. To support these efforts, methods for more rapidly identifying

61   wetland locations are needed. Although manual surveys will continue to be the most accurate

62   method to map wetlands, there is potential for supporting these efforts by using machine learning

63   approaches, including deep learning, to identify wetland features at varying scales (Guo et al.,

64   2017; Lang et al., 2013; Lang & McCarty, 2014).

65   Despite the many types of protected wetlands that exist, all wetlands can be identified by

66   common features. These include the presence of hydrologic conditions that inundate the area,

67   vegetation adapted for life in saturated soil conditions, and hydric soils (US Corps of Engineers,

68   1987). Researchers have demonstrated the ability to detect these features from multispectral

69   imagery, radar, and Light Detection and Ranging (LiDAR) data (Guo et al., 2017). Multispectral

70   imagery are the most commonly applied data in wetland studies (Guo et al., 2017; Klemas, 2011);

71   however, spectral variables alone may be unable to distinguish wetlands due to spectral confusions

72   from reflectance and backscattering (Dronova, 2015; Kim et al., 2011). LiDAR data are well-suited

73   to complement multispectral analyses due to its wide, and growing, availability and demonstrated

74   benefit to wetland mapping (Guo et al., 2017; Klemas, 2011; Kloiber et al., 2015; Lang & McCarty,

75   2014; Snyder & Lang, 2012). LiDAR returns can be interpolated to create high-resolution digital

76   elevation models (DEMs), from which wetland indicators based on flow convergence and near-

77   surface soil moisture can be derived (Lang et al., 2013; Lang & McCarty, 2014; Millard &

78   Richardson, 2013, 2015; O'Neil et al., 2018, 2019). Moreover, researchers have shown the benefit

79   of LiDAR DEM metrics as input variables to traditional machine learning techniques, such as

80   random forests, for wetland mapping and classification (e.g., Deng et al., 2017; Kloiber et al.,

81   2015; Millard & Richardson, 2013; Millard & Richardson, 2015; O'Neil et al., 2018, 2019; Zhu

82   & Pierskalla, 2016).

83   The successful coupling of LiDAR and multispectral imagery with traditional machine

84   learning techniques for wetland identification is well-documented. However, deep learning for

85   remote sensing studies, including wetland identification, is a new application space (Ma et al.,

86   2017; Zhang et al., 2016) that shows promise for fulfilling the unmet need for wetland inventory

87   creation. Deep learning architectures are modeled after the architecture of the mammal brain (Serre

88   et al., 2007), where inputs are perceived and processed through multiple layers of abstraction.

89   Convolutional neural networks (CNNs) (LeCun et al., 1998) are a representative form of deep

90   learning that is used for visual recognition. CNNs utilize the spatial context of detected features to

91   identify objects and classify scenes. The distinguishing element of CNN architectures are the

92   convolutional layers, which convolve spatial filters over input images to identify patterns that are

4

93   characteristic of target classes. Deep convolutional neural networks (DCNNs) (He et al., 2016;

94   Krizhevsky et al., 2017; Simonyan & Zisserman, 2014) and fully convolutional neural networks

95   (FCNs) (Long et al., 2015) are extensions of the CNN framework that can output dense pixel-wise

96   classifications within images (i.e., semantic segmentation), where each pixel of the input image is

97   assigned a class.

98   Since the formalization of the concept in 2006 (Hinton et al., 2006), deep learning has

99   advanced the fields of speech recognition, medical diagnosis, and autonomous driving

100   applications, and has since motivated new applications in environmental and water resources

101   management (Liu et al., 2018; Pan et al., 2019; Shen, 2018; Zhang et al., 2016). Researchers have

102   shown the ability of DCNNs, FCNs, and other CNN extensions to delineate urban and natural

103   landscape classes using multispectral imagery and topographic data (Audebert et al., 2017, 2018),

104   multispectral imagery and LiDAR point clouds (Xu et al., 2018), and multispectral imagery alone

105   (Hu et al., 2018; Kemker, Gewali, et al., 2018; Kemker, Salvaggio, et al., 2018; Scott et al., 2017).

106   Few researchers have applied DCNNs and FCNs specifically to wetland classification. These

107   include Liu et al. (2018), who applied orthoimagery and elevation information to deep learning

108   models for wetland segmentation. In addition, Rezaee et al. (2018) used multispectral imagery in

109   a wetland deep learning model, and posited that predictions would improve with the incorporation

110   of physical information from radar or LiDAR sources.

111   The typical need for massive validation sets to train deep learning models is a significant

112   deterrent to environmental and water resources researchers (Shen, 2018; Zhang et al., 2016), as

113   reliable training data is often lacking in these applications. This issue is especially prevalent for

114   wetland identification that is intended to inform conservation and permitting efforts, where training

115   data for computational models are ideally manually derived and confirmed by regulatory entities.

116   The effects of training data limits for wetland semantic segmentation have been investigated by

117   Liu et al. (2018), where comparisons were drawn for a single study area using DCNNs, FCNs,

118   random forests, and support vector machines, with privately contracted aerial imagery and surface

119   elevation information as input features. While this is an important stride in gaining insight into the

120   training data needs for deep learning of wetlands, an analysis has yet to be done that utilizes freely-

121   available data and is completed over multiple geographic regions.

122   The growing research area of deep learning for remote sensing applications shows promise

123   for advancing wetland mapping. Although researchers have begun to show the potential for

124   wetland identification at a high resolution using deep learning approaches, research gaps remain.

125   Specifically, analyses are needed to identify the deep learning performance potential for different

126   geographic regions when limited to relatively small quantities of verification data and freely

127   available input data, which are typical in practice. We aim to contribute to this field by presenting

128   a novel wetland identification methodology that implements a basic semantic segmentation

129   architecture and is generalizable because it leverages freely-available geospatial and remote

130   sensing data. Our input data configuration consists of LiDAR DEM derivatives that describe

131   geomorphologic and hydrologic contributors to wetland formation, as well as a commonly-used

132   vegetative index. Using four study sites across Virginia, we build and evaluate several wetland

133   models to demonstrate the potential for wetland semantic segmentation given typical training data

134   resources. Through this research, we seek to answer the following questions.

135   i.   Across geographically distinct study sites, what wetland prediction accuracy is
136   achievable by building site-specific models from typically available amounts of
137   wetland delineation training data?
138   ii.   What is the potential for a single, combined-site model trained using data from across
139   geographic regions to predict wetlands at each individual site?

Methodology

Study Areas

Four study areas across Virginia, USA are used in this analysis (Figure 1a). Data for each study area include the extents of wetland surveys and the surrounding Hydrologic Unit Code (HUC) 12 watershed (USGS, 2019b) (Figure 1b). The HUC 12 watersheds were used as processing extents and surveyed areas provided the validation data, also referred to as the study sites. The study areas span four level-III ecoregions. As shown in Table 1, the sites also vary by size, land cover, and topographic characteristics. Notable differences include the higher rate of development in sites 1 and 2, and the mild topography of Site 4. In addition, wetlands are much more abundant in Site 4, where the wetland to nonwetland ratio is 0.42, compared to less than 0.1 in the other sites. Note that all surveyed wetland types were merged into a single wetland category prior to use as verification data.
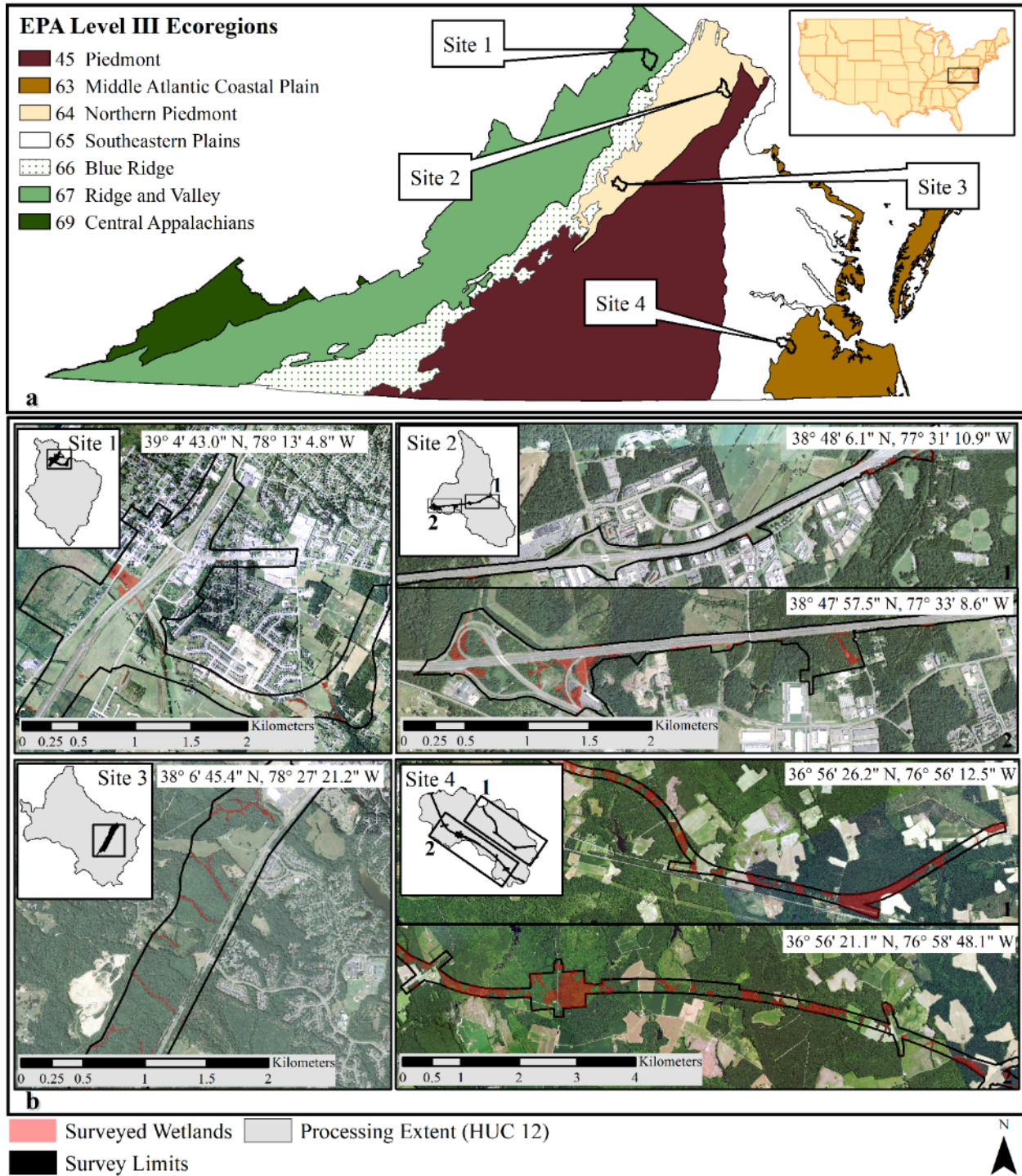
Figure 1. Four study areas spanning four level III ecoregions in Virginia, USA (a). Each study area includes the wetland survey limits, referred to as study sites, and the encompassing HUC 12 watershed, used as the processing extent (b). Reprinted from "Effects of LiDAR DEM Smoothing and Conditioning Techniques on a Topography-Based Wetland Identification Model" by O'Neil et al., 2019, Water Resources Research, 55 (5), 4343-4363.

157   Ecoregion data source: US EPA Office of Environmental Information. Aerial imagery data source: NAIP Digital

158   Ortho Photo Image.

159

160

	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
10th Percentile Slope[c] (m/m)	0.02	0.01	0.04	0.01
90th 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.

165

Input Data

167   This study used publicly available LiDAR DEMs, National Agriculture Imagery Program

168 (NAIP) aerial imagery, and field-mapped wetland surveys. LiDAR DEMs were obtained from the

169 Virginia Information Technologies Agency (VITA) (VITA, 2016) as hydro-flattened, bare-earth

170 DEMs. The LiDAR data used were collected and processed between 2010 and 2015 and have

171 horizontal resolutions ranging from 0.76 m to 1.5 m. NAIP imagery are provided by the United

172 Sates Department of Agriculture (Farm Service Agency, 2017). NAIP imagery were used to derive

173 the NDVI. NAIP imagery contain four spectral bands (red, green, blue, and near-infrared) at a 1

174 m spatial resolution. Imagery used in this study were collected near the dates of wetland surveying,

175   and images were resampled to match the resolution of the LiDAR DEMs, if necessary. Wetland

176   delineations and survey limits were provided by the Virginia Department of Transportation

177   (VDOT) in polygon vector format and served as validation data for this study. All verification

178   wetlands were manually surveyed during summer months (May – August) between 2013 and 2016

179   by professional wetland scientists in compliance with transportation planning permitting. Wetland

180   delineations for sites 2, 3, and 4 were also jurisdictionally confirmed by the US Army Corps of

181   Engineers (USACE). Binary wetland/nonwetland geotiffs were created from these data, with

182   resolutions matching those of the site LiDAR DEMs. Visual analyses of Google Earth images

183   showed that the study site landscapes changed minimally between LiDAR acquisition and wetland

184   delineation timeframes.

185   **Wetland Identification Method**

186   The wetland identification method consists of three main parts: preprocessing, feature

187   creation, and semantic segmentation and accuracy assessment (Figure 2). Input data required

188   include high-resolution DEM data, four-band aerial imagery, and validated wetland/nonwetland

189   distribution data, all in geotiff format. From these data, topographic indices (curvature,

190   Topographic Wetness Index, and Cartographic Depth-to-Water index) and the Normalized

191   Difference Vegetation Index are calculated. These input features are merged into a single four-

192   band composite grid. Smaller image tiles are created from the composite grid and validation data,

193   and the pairs of corresponding image tiles are randomly separated into training and testing datasets.

194   Finally, dense pixel-wise wetland predictions are made using a deep learning architecture created

195   for remote sensing data, DeepNets for Earth Observation (Audebert et al., 2018), and the accuracy

196   of wetland predictions is assessed. The main outputs are geotiff wetland predictions for each image

197   tile and an accuracy report for the entire validation data area. The method was implemented using

198   open source Python libraries and is available under an MIT license (see Software Availability

199   section).
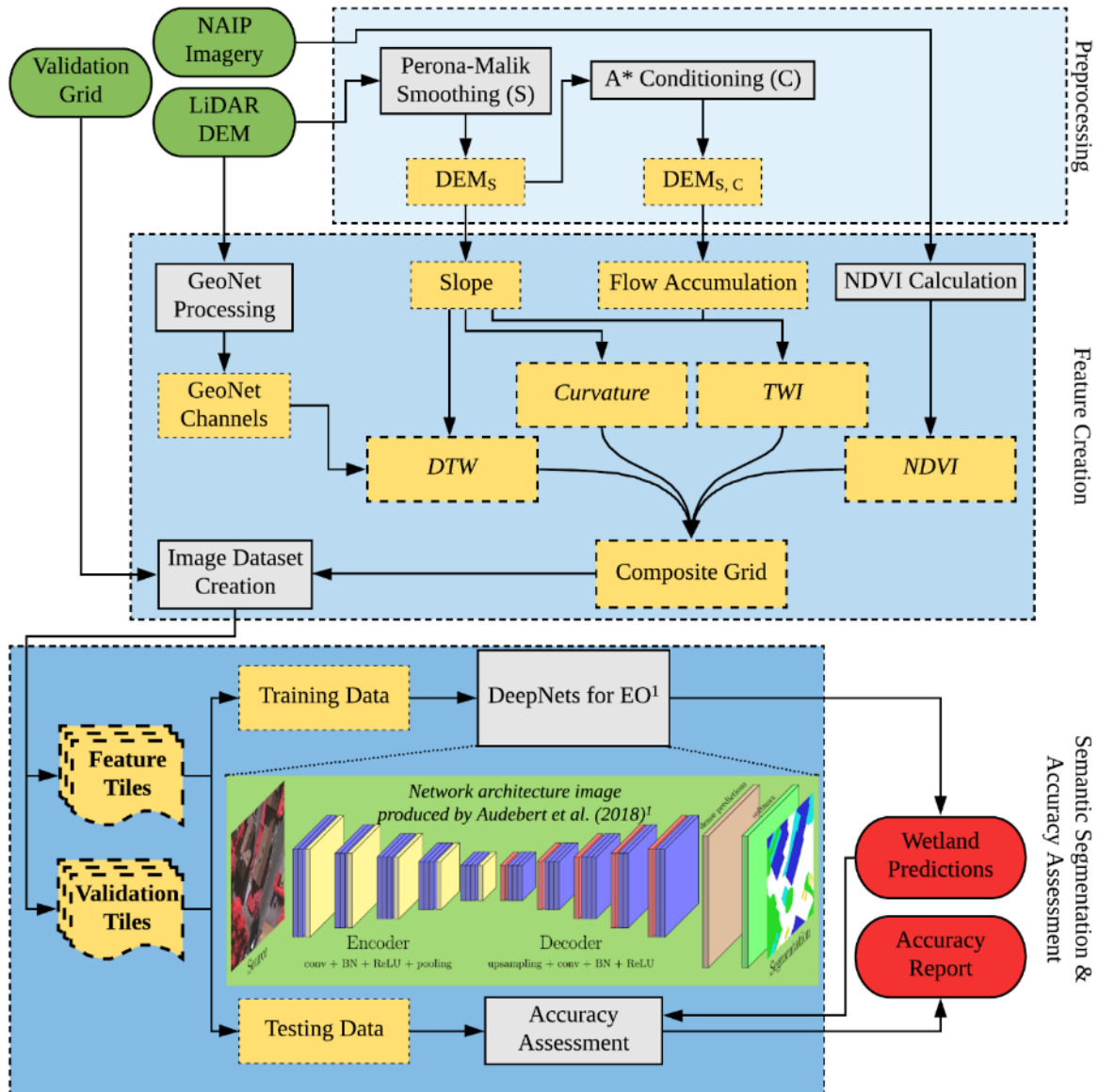
200
201   *Figure 2. Overview of the proposed wetland identification method. Green shapes indicate input data, grey shapes*

202   *indicate processes, yellow shapes indicate intermediate output, and red shapes indicate final model output.*

203   *[1]Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with*

204   *multimodal deep networks. ISPRS Journal of Photogrammetry and Remote Sensing, 140, 20-32.*

12

Preprocessing

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

228 sinks, thus avoiding unnecessary modification of the input DEM (Metz et al., 2011). The A*

229 conditioning method was executed using the GRASS GIS *r.watershed* module (GRASS

230 Development Team, 2017; Metz et al., 2011).

231

232 **Feature Creation**

233 *Topographic Features*

234   In a prior study, we concluded that the curvature, Topographic Wetness Index (TWI) and

235 Cartographic Depth-to-Water index (DTW) are successful topographic metrics for wetland

236 identification for our study sites (O'Neil et al., 2018, 2019).

237   Curvature of a surface can describe the degree of convergence and acceleration of flow

238 (Moore et al., 1991), and studies have shown its capability to indicate saturated and channelized

239 areas (Ågren et al., 2014; Hogg & Todd, 2007; Kloiber et al., 2015; Millard & Richardson, 2015;

240 O'Neil et al., 2018, 2019; Sangireddy et al., 2016). Here we use laplacian curvature, defined as the

241 second derivative of the elevation grid. Laplacian curvature has been shown to favor the extraction

242 of natural channels rather than artificial drainage paths, and to more effectively identify channels

243 in flat, developed landscapes compared to alternative curvature forms (Passalacqua et al., 2012).

244 Thus, we found the laplacian curvature to be most suitable for our study areas which all encompass

245 corridor projects and are partially developed (O'Neil et al., 2019). The curvature grid is created

246 from the smoothed DEM using code adopted from PyGeoNet (Passalacqua, Do Trung, et al., 2010;

247 "PyGeoNet," 2019; Sangireddy et al., 2016).

248   The ability of the TWI to indicate saturated areas is well-documented in the literature

249 (Ågren et al., 2014; Lang et al., 2013; Millard & Richardson, 2015; Murphy et al., 2009; O'Neil

250 et al., 2018, 2019). The TWI relates the potential for an area to accumulate water to its tendency

251 to drain water, defined as

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

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

253 local slope (Beven & Kirkby, 1979). The TWI was created from the smoothed, conditioned DEM

254 using the *r.watershed* program of GRASS GIS. This module calculates the α term using the

255 multiple flow direction algorithm (Holmgren, 1994) and the β term using a GRASS GIS-calculated

256 slope.

257   Researchers have demonstrated the capability of the DTW to capture saturated areas as

258 well (Murphy et al., 2007, 2009, 2011; O'Neil et al., 2018, 2019; Oltean et al., 2016; White et al.,

259 2012). The DTW assumes that the likelihood for soil to be saturated increases with its proximity

260 to surface water, in terms of distance and elevation (Murphy et al., 2007). Calculated on a per-

261 pixel basis, the DTW is defined as

$$DTW\ (m) = \left(\sum\left(\frac{dz_i}{dx_i}\right)a\right) * x_p \tag{2}$$

262 where $\frac{dz}{dx}$ is the downward slope of pixel i along the least-cost (i.e., slope) path to the nearest

263 surface water pixel, a is a factor accounting for flow moving parallel or diagonal across pixel

264 boundaries, and x_p is the pixel resolution (Murphy et al., 2007). Inputs required to calculate the

265 DTW include a slope grid, representing cost, and a surface water grid, representing the source from

266 which distance is calculated. We create the surface water grid directly from the LiDAR DEM using

267 PyGeoNet, which performs a statistical analysis of curvature and uses geodesic minimization

268 principles to predict stream lines (Passalacqua, Do Trung, et al., 2010; Sangireddy et al., 2016).

269 Visual analyses showed that streams created by PyGeoNet better aligned with aerial imagery,

270 compared to national hydrography data (i.e., NHD streams) and streams generated from the flow

271 initiation threshold method (Band, 1986; O'Callaghan & Mark, 1984; Tarboton, 1991) that is

272 commonly used. PyGeoNet was executed using parameters suggested for engineered landscapes

273 (see Sangireddy et al., 2016), which was found to produce accurate results across all sites in prior

274 wetland model development (O'Neil et al., 2019). The PyGeoNet streams and slope grid were used

275 as inputs to the GRASS GIS r.cost module (GRASS Development Team, 2017) to create the DTW

276 grid.

277 *NDVI*

278   The NDVI is a commonly-used spectral index that relates plant biomass and stress and

279 separates wet versus dry areas (Klemas, 2011; Ozesmi & Bauer, 2002). Researchers have used the

280 NDVI as a wetland indicator in traditional machine learning frameworks (Corcoran et al., 2013;

281 Dronova, 2015; Dronova et al., 2011; Guo et al., 2017; Mui et al., 2015; Rampi et al., 2014; Tian

282 et al., 2016), as well as for general land cover classifications using deep learning frameworks

283 (Audebert et al., 2017, 2018; Lee et al., 2019; Xu et al., 2018). The NDVI utilizes the red and the

284 near-infrared bands (Carlson & Riziley, 1997), defined as

$$NDVI = \frac{Infrared - Red}{Infrared + Red} \tag{3}$$

285 The red band indicates surface layer chlorophyll, and therefore surface conditions of plants, and

286 the near-infrared band is reflected from the inner leaf cell structure, indicating the abundance of

287 plant tissue (Klemas, 2011). To calculate the NDVI, Eq. (3) was executed using NumPy operations

288 and the appropriate NAIP imagery bands.

289 *Image Dataset Creation*

290   The image dataset creation produces two sets of image tiles: i) feature tiles representative

291 of the composite grid of input features, and ii) validation tiles representative of ground truth

292 wetland and nonwetland locations. Due to the irregular shapes of the field surveys, NoData pixels

16

293   existed within the rectangular extent of the validation data. Rather than reduce our validation data

294   to an extent without unverified area, NoData pixels were treated as an additional target landscape

295   class. Thus, all pixels in the validation data were categorized as NoData (0), nonwetland (1), or

296   wetland (2) as a first step in the image dataset creation process.

297   To build the dataset of feature tiles, each band of the composite grid is rescaled to a range

298   of 0 to 1, per the requirements of the DeepNets algorithm. Rescaling the NDVI band was

299   nontrivial, as these values have global minimum and maximum of -1 and 1. Conversely, the range

300   of values for each of the topographic features depends on the landscape they are calculated from,

301   therefore it was necessary to assume global minimum and maximum values. The range of each

302   topographic input was analyzed across the study sites, and global minimum and maximum values

303   that encompassed roughly 90% of the values were chosen. Note that only global maximum values

304   had to be assumed for the TWI and DTW, which both have global lower bounds of 0 or nearly 0.

305   Although this step generalizes portions of the study areas, this occurs only where there are extreme

306   topographic features that occur infrequently. In addition, by limiting the range applied to each

307   topographic input feature rather than choosing extreme, but encompassing, values, the significance

308   of the relative distance between values is minimally affected. The minimum and maximum values

309   used to rescale topographic features and the NDVI to a range of 0 to 1 are shown in Table 2.

310
311   *Table 2. Minimum and maximum values used to scale each input feature to a range of 0 to 1. Minimum and*

312   *maximum values were assumed for the TWI, curvature, and DTW from statistical analyses.*

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

313

314   Following these steps, the categorized validation grid and scaled composite grid were each

315   separated into image tiles of size 320 x 320 pixels. We chose the 320-pixel size constraint to

316   balance the desire to use image tiles large enough to depict heterogeneous landscapes and the need

317   to separate the study site into enough images to sample training and testing tiles that were randomly

318   dispersed. Feature and labeled image tiles sets were not considered for either training or testing if

319   more than 80% of the area was populated with NoData pixels.

320   **Semantic Segmentation Model: DeepNets for Earth Observation**

321   Our model performs a semantic segmentation of input images, where each pixel of an input

322   image is labeled as either NoData, nonwetland, or wetland. That is, a trained semantic

323   segmentation model will assign a class prediction to each pixel in an image, however different

324   instances of target class objects are not defined (i.e., instance segmentation).

325   As an initial step in developing a deep learning wetland model, the current work is intended

326   to demonstrate the suitability of a CNN to identify planning-scale wetlands in the landscape. We

327   implemented a multimodal deep network, DeepNets for Earth Observation, for semantic

328   segmentation classification (Audebert et al., 2017). DeepNets has emerged as a state-of-the-art

329   tool for segmentation of high-resolution remote sensing data (Demir et al., 2018), and has been

330   implemented and validated for automating segmentation of remote sensing data (Audebert et al.,

331   2016, 2017, 2018).

332   Although DeepNets was chosen as a vehicle to address the guiding research questions of

333   this work, it is among several deep learning architecture currently achieving competitively in

334   semantic segmentation of satellite imagery. Ghosh et al. (2018) applied a Stacked U-Nets

335   architecture to achieve high-quality satellite imagery segmentation with relatively few prediction

336   parameters. Volpi and Tuia (2016) use a CNN to segment very high-resolution imagery to achieve

337 F1 scores of about 85%. Marmanis et al. (2018) propose a downsample-upsample and achieve

338 similar results. While each of these approaches are likely to achieve good results with wetlands

339 segmentation, DeepNets achieved slightly higher results on segmentation of benchmark imagery

340 datasets (Demir et al., 2018), thus it was adopted for this study. An important future step in

341 progressing this research would be to perform a comparative analysis of other emerging deep

342 learning techniques for wetland segmentation.

343   As a starting point in the development of our deep learning wetland model, the baseline

344 DeepNets architecture is implemented here (Audebert et al., 2018, 2019). DeepNets builds on the

345 SegNet architecture (Badrinarayanan et al., 2017) and is implemented using PyTorch (Paszke et

346 al., 2017). SegNet produces predictions with the same resolution as the input image by using an

347 encoder-decoder structure, making it well-suited for classification of landscape objects from

348 georeferenced images (Audebert et al., 2018; Badrinarayanan et al., 2017). The encoder portion of

349 SegNet is based on the convolutional layers of VGG-16 (Simonyan & Zisserman, 2014), and

350 consists of convolutional layers, batch normalization, a rectified linear unit, and max-pooling. As

351 shown in the inset image (defined by Audebert et al., 2018) in Figure 2, the decoder is structurally

352 symmetrical to the encoder. Pooling layers are replaced with unpooling layers that relocate pixel

353 activations from the smaller feature maps to corresponding indices of zero-padded upsampled

354 images. Convolution blocks are then used to densify the sparse pixel activations. This sequence of

355 unpooling and convolutions is repeated until feature maps reach the original spatial resolution.

356 Following this, a softmax layer is used to compute multinomial logistic loss. Another feature of

357 the DeepNets approach is the generation of predictions at several resolutions, and the calculation

358 of loss at these intermediate resolutions. In doing so, the DeepNets model predicts a semantic map

359 at full resolution as well as smaller resolutions, which are averaged together to obtain a final full-

360   resolution semantic prediction. Lastly, a sliding window approach is used to extract smaller

361   patches within each input image, which acts as data augmentation. For further details on the

362   DeepNets architecture, we direct readers to Audebert et al. (2018).

363   Following procedures demonstrated by Audebert et al. (2016, 2017, 2018), we incorporate

364   the NDVI and elevation data into our DeepNets model. However, rather than using the original

365   elevation grid as an input, we guide the learning of the model by deriving specific geomorphic and

366   hydrologic features from the DEM as inputs. This strategy was chosen following a hypothesis that

367   wetland predictions would improve if a deep learning model trained from explanatory variables

368   that are specific to wetlands. In our implementation of DeepNets, we also applied class weights,

369   which are related to the importance of correct predictions for a specific class when calculating the

370   loss. We used this feature to account for the imbalance between the wetland and nonwetland

371   classes across all sites, as well as to decrease the importance of NoData areas. Lastly, we allow for

372   data augmentation in the form of mirroring images and flipping the orientation. Parameters for the

373   DeepNets model incorporated into our wetland model workflow are given in Table 3. Note that

374   these parameters were chosen as starting points to be later refined through additional model testing.

375   *Table 3. Parameters for the DeepNets implementation used in all performed experiments.*

Image tile size (# pixels)	320, 320
Sliding window size (# pixels)	64, 64
Sliding window stride (# pixels)	8
Base learning rate	0.01
Momentum	0.9
Weight decay	0.005
Training epochs	100

Class weights [NoData, Nonwetland, Wetland]	[0.02, 0.08, 0.9]

376

Accuracy Assessment

377

378   In line with the intended environmental planning and permitting application, accuracy

379   metrics were selected considering the higher importance of true positive (i.e., wetland) predictions

380   versus true negative (i.e., nonwetland) predictions to wetland conservation. Model performance

381   was evaluated in terms of wetland recall and wetland precision, calculated using the Scikit-learn

382   Python library (Scikit-learn Developers, 2017).

383   Recall, also known as the true positive rate, represents the percentage of true wetlands that

384   were predicted, and is defined as

$$Recall = \frac{True\ wetland\ predictions}{Total\ true\ wetlands}. \tag{4}$$

385   Recall can be considered the priority indicator of model performance given the importance of the

386   minority wetland class, a choice also supported by statistical literature (Branco et al., 2016; Chen

387   et al., 2004; Sun et al., 2007). Precision is used to account for model overprediction. Unlike the

388   commonly-used specificity, precision is not biased by large numbers of true negative instances,

389   and therefore can be considered more representative for imbalanced scenarios (Branco et al., 2016;

390   Sun et al., 2007). Precision represents the percentage of wetland predictions made that were

391   correct, defined as

$$Precision = \frac{True\ wetland\ predictions}{Total\ wetland\ predictions}. \tag{5}$$

392   It should be noted that the appropriate selection of accuracy metrics remains an open problem

393   not only for semantic segmentation, but for classification tasks in general, and additional criteria

394   have been proposed and widely used. We found recall and precision to be more suitable for model

395   assessment compared to commonly used options, such as overall accuracy, Kappa statistic, and

21

396 Matthews Correlation Coefficient (MCC). When using overall accuracy, detection rate of the

397 minority class has a lower impact than that of the majority class (Branco et al., 2016; Chen et al.,

398 2004), misrepresenting a wetland model predicting all nonwetland instances as very accurate.

399 Moreover, the Kappa statistic is biased by sample size, and can increase as the wetlands to

400 nonwetlands ratio increases, even if wetland recall decreases (Ali et al., 2014; Byrt et al., 1993).

401 Both overall accuracy and the Kappa statistics have been omitted from wetland classification

402 studies for these reasons (Baig et al., 2014; Zhu & Pierskalla, 2016). Although the MCC metric

403 has been shown to be suitable for imbalanced scenarios (e.g., Boughorbel et al., 2017), its takes

404 into account the number of true negative samples.

Experimental Setup

Addressing Research Question 1: Creating Site-Specific Models

407   Experiments 1 and 2 (Figure 3A) were designed to offer insight into potential wetland

408 accuracy given varying sizes of reliable training sets, evaluated over four geographic regions. In

409 Experiment 1, we created models that sample training images from the area to be mapped (i.e.,

410 site-specific models). For each site, 70% of eligible image sets were randomly selected, producing

411 the maximum training set size available, which varied based on site size (Table 4). To compare

412 how models of different ecoregions perform given the same training resources, site-specific

413 models were created and evaluated at each threshold of training set size. Experiment 2 applied the

414 site-specific models created through Experiment 1 (those using the maximum training set size) to

415 predict wetlands in the other sites. Thus, Experiment 2 represents the scenario where a pretrained

416 wetland model is applied for a new area for which training data is unavailable.

Table 4. Maximum number of training images available per site when randomly sampling 70% of the eligible validated area. Each labeled image used for training has a resolution of 320 x 320 pixels.

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

Addressing Research Question 2: Creating Combined-Site Models

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

428
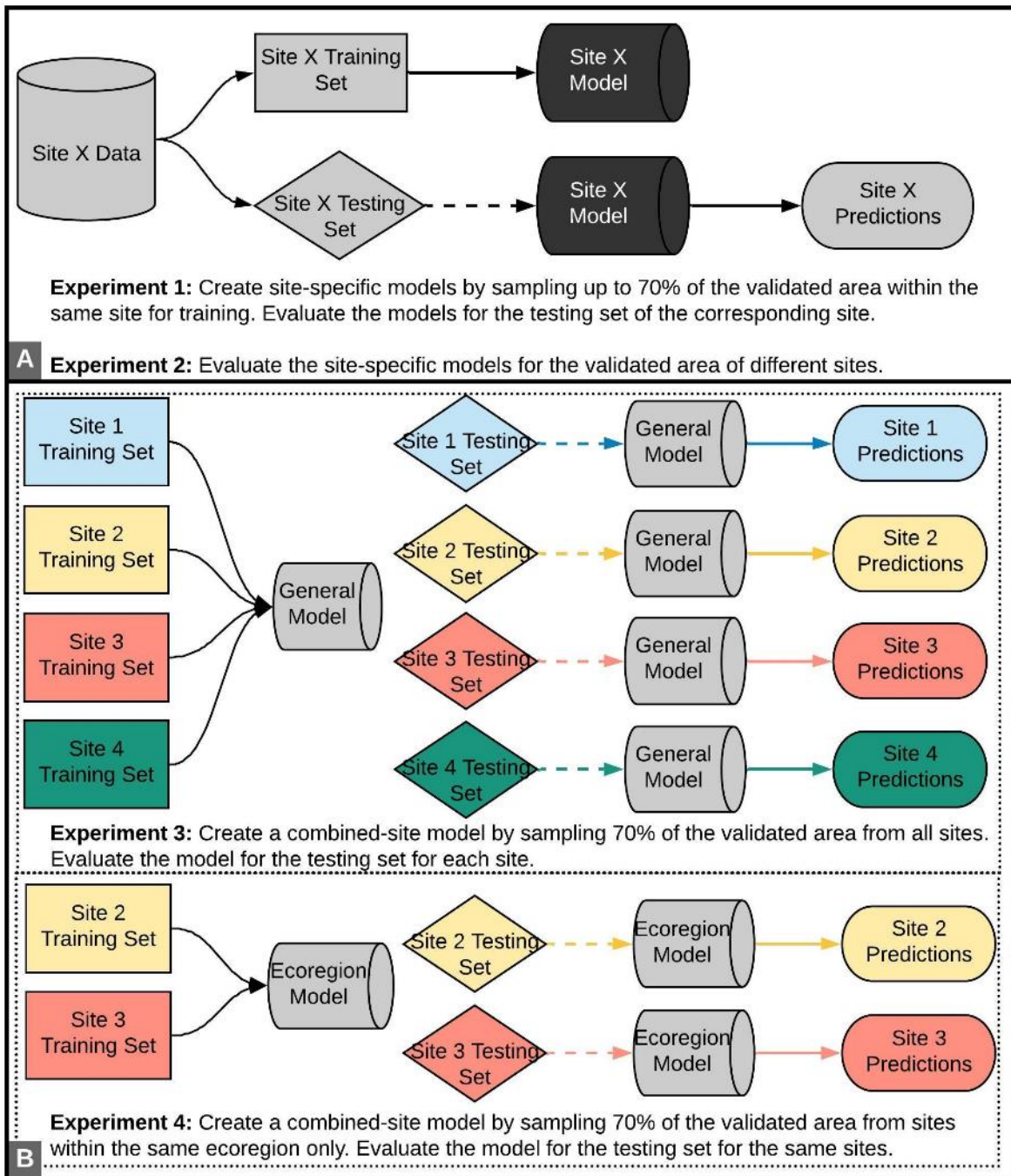
Figure 3. Methodology followed for the four experiments designed to address the study research questions.

430

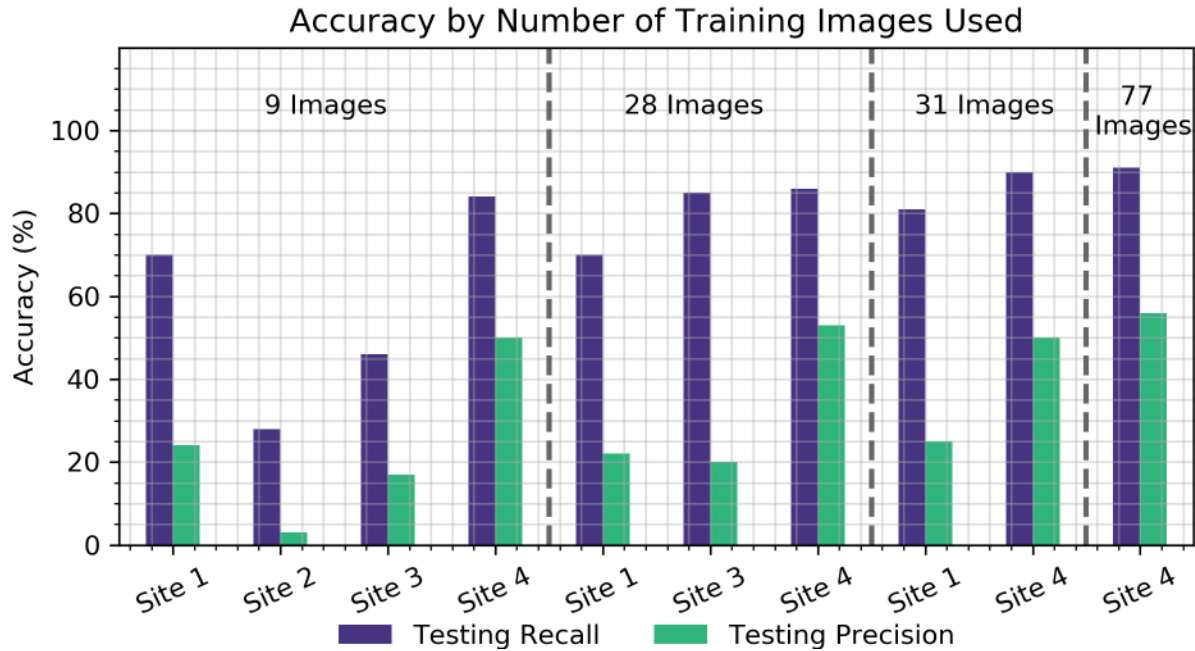## Results

Performance of Site-Specific Models

For Experiment 1, site-specific models were built using training data quantities ranging from 9 to 77 images, depending on validation data extents (Figure 4). The resulting 10 sets of wetland predictions were evaluated for the testing area complementing the training data quantity used. Results show that the best performing models for each site were those trained using the maximum training set size available, equal to 70% of the validation area. Conversely, the lowest performing models across all sites occurred when using the fewest training data, nine images. The Site 4 model trained with 77 images achieved the highest wetland recall and precision across all site models. The Site 4 model also outperformed other sites when limited to the same number of training images (Figure 4). The overall lowest performing model was built for Site 2, which also had the smallest training dataset available, only nine images.

While the improvements in prediction accuracy as training data increased were expected, intermediate changes in accuracy were inconsistent. For Site 3, recall increased considerably (46% to 85%) and precision increased slightly (17% to 20%) when increasing training images from 9 to 28. However, changes in model accuracy were less significant for Site 1, where the most notable accuracy improvement occurred when increasing training data from 28 to 31 images, which increased recall from 70% to 81% and precision from 22% to 25%. Models built for Site 4 performed consistently, maintaining high performance regardless of training set sizes ranging from 9 to 77 images. For Site 4, recall only varied between 84% and 91% and precision between 50% and 56%. It was unexpected that Site 4 did not improve more notably when increasing the training dataset from 31 to 77 images, as this was the largest increase in training set studied. This may be

453   due to the fact Site 4 has the most balanced wetland to non-wetland areas, so fewer training images

454   are needed to create an accurate model.

455

456   *Figure 4. Wetland mapping accuracy resulting from Experiment 1, where site-specific models were created using*

457   *several training data sizes depending on site availability.*

Using Site-Specific Models to Predict Wetlands in Other Sites

459   Experiment 2 resulted in an additional 12 sets of results, where the best performing site-specific

460   models (i.e., those trained with the maximum training data set size) were used to predict wetlands

461   in the other sites. The evaluation of these trials represents wetland prediction accuracy for the

462   entirety of the site validation area, and the results achieved by applying the site-specific models

463   for their own areas are also shown for reference (Figure 5). In most cases, utilizing training

464   information from a different area, even if this represented a greater quantity of data, did not

465   improve predictions compared to those resulting from a model trained for its own area. Site 2 was

466   the exception for this trend, as both recall and precision improved when using any of the models

467   built for other sites, compared to using the Site 2 model. Moreover, the Site 2 model produced

468   more accurate wetland predictions when applied to the other sites, compared to its own testing

469   area. Although the predictions for others sites resulting from the Site 2 model were still among the

470   lowest accuracies per site, this suggests there may be topographic or spectral confusion between

471   Site 2 training and testing data. Also, there was an unexpected increase in precision when applying

472   the Site 1 model versus the Site 4 model for Site 4 predictions. However, since both wetland

473   precision and wetland recall should be considered when summarizing model performance, the

474   significantly greater recall achieved by the Site 4 model leads us to conclude that the Site 4 model

475   outperformed the Site 1 model here. Lastly, the Site 4 model resulted in the highest recall scores

476   and among the lowest precision scores across all trials for sites 1, 2, and 3. This reflects a tendency

477   of the Site 4 model to overpredict wetlands in other sites. This may be because Site 4 includes

478   large, areal wetlands common in the coastal plain given its low relief topography, but uncommon

479   in the other three sites that are outside of the coastal plain.

Accuracy by Training Sampling Location
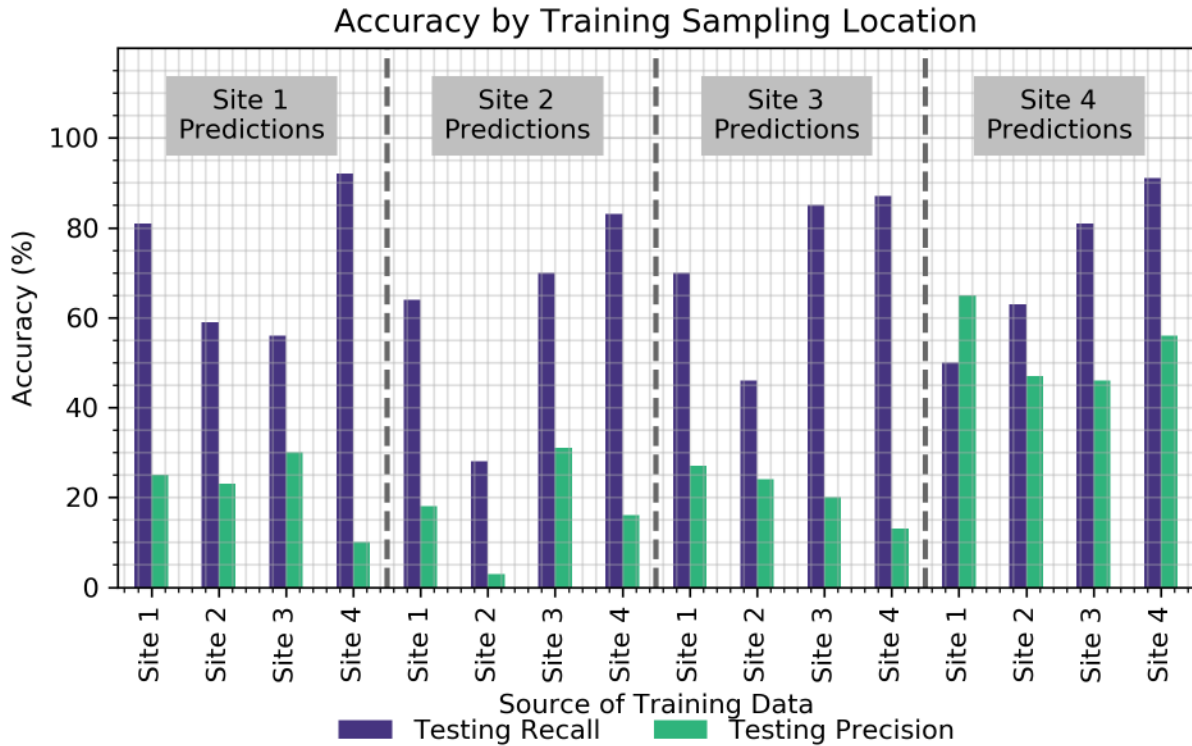
Figure 5. Wetland mapping accuracy resulting from Experiment 2, where the best performing site-specific models

were used to predict wetlands in other sites.

Performance of Combined-Site Models

Experiment 3 resulted in the general model, trained with the maximum available training images from each site. When applying the general model to Site 1 testing areas, recall increased from to 81% to 89% and precision decreased from 25% to 18%, relative to the best performing Site 1 model (Figure 6). For Site 2 testing areas, the general model considerably improved wetland recall (28% to 40%) and minimally changed precision (3% to 2%), compared to the best performing site-specific model (Figure 6). The general model produced worse predictions than the site-specific model for Site 3, decreasing recall from 85% to 73% and precision from 20% to 15%. The general model performed nearly the same for Site 4 compared to the site-specific model, where recall remained high at 91% and precision increased by a small margin from 56% to 57%. These

28

493   results suggest that a general model trained with data collected across all sites would not be a

494   suitable method for wetland prediction, at least with the current methodology and data availability.

495   Experiment 4 resulted in the ecoregion model, trained with the maximum available training

496   images from sites 2 and 3, which share the Northern Piedmont ecoregion. This experiment tested

497   the idea that a general wetland classification may be possible, but only within a single ecoregion

498   and not across ecoregions as was attempted in Experiment 3. For Site 2, the ecoregion model

499   produced worse predictions than the general model and the site-specific model, with recall

500   decreasing to 21% and precision remaining nearly the same at 2% (Figure 6). In contrast, the

501   ecoregion model improved wetland recall and precision for Site 3 (77% and 22%, respectively)

502   compared to the general model, however this was not an improvement from the Site 3-specific

503   model (Figure 6). This suggests that an ecoregion-specific classification model may be useful, but

504   not more so than a site-specific model given the data available here.

Accuracy Using Combined-Site Models
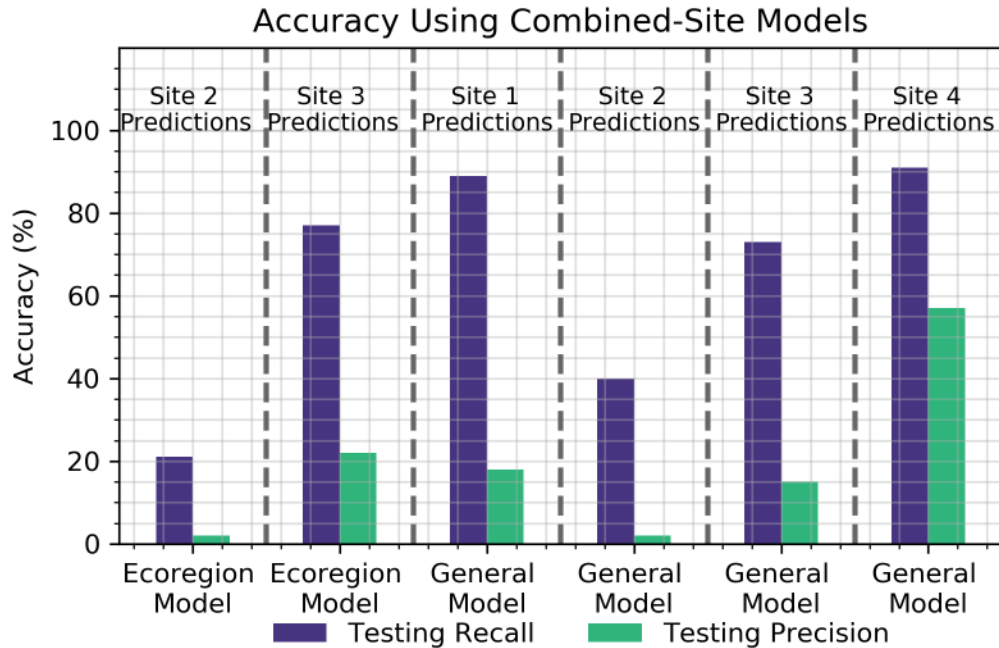
Figure 6. Wetland mapping accuracy resulting from Experiment 3, which used training data from all sites to create a general model, and Experiment 4 which used training data only from sites within the same ecoregion (sites 2 and 3) to create an ecoregion model.

Discussion

Potential for Site-Specific Models

We found that site-specific models improved as more training data was sampled from the area to be mapped, with the best models created from the maximum training datasets studied: 70% of the validation area. However, performance did not improve consistently for sites at the intermediate training data thresholds. This outcome exemplifies that model improvement is an issue of not only increasing the quantity of training data, but also the quality. The performance inconsistencies may be due to unequal wetland distributions in each training image. For example, the training images introduced for Site 1 when increasing the training data threshold from 9 to 28 images, may have provided very few wetland areas if the random selection included scenes with few or only small wetlands. In addition, it is possible that the random nature of the training image

520 set creation led to the introduction of some scenes with conflicting wetland/nonwetland signatures.

521 As there is a benefit to identifying a training area threshold that begins to improve model

522 performance across different sites, future work should include repeating this experiment with

523 quality-controlled training data images and thresholds. Evaluating model performance across sites

524 with training image thresholds at even increments of wetland and nonwetland area would result in

525 more conclusive insights as to the changes in model performance as more training data becomes

526 available. This being said, the overall improvements across the sites as training data increased to

527 the maximum available set are likely due to the ability of the model to learn a wider range of

528 wetland characteristics that exist in the additional landscape scenes.

529   Figure 7 demonstrates the changes in wetland predictions as a result of increasing training

530 data from nine training images (column A) to the maximum training images per site (column B).

531 For sites 1, 3, and 4, increased training data reduced wetland overprediction surrounding the

532 extents of ground truth wetlands, most notably for narrow wetland segments in sites 1 and 3. In

533 addition, wetland predictions for these sites encompassed more of the true wetland area, most

534 apparent for Site 4, where predictions densified for a relatively large wetland as a result of

535 increasing the training data. Figure 7 also exemplifies the poor performance of the Site 2 model.

536 Although the Site 2 model predicts wetlands as small, linear features that are representative of the

537 nature of ground truth wetlands in the area, the predictions are relatively sparse and incorrect. By

538 visually examining the input features and testing data for Site 2, we found that validation wetlands

539 existed underneath dense tree canopy along a road corridor. Topographic metrics in this area

540 indicated values corresponding to wetness within the true wetland boundaries, however, the NDVI

541 showed constant values for most of the forested area. The lack of distinction between values by

542 the NDVI is likely due to the source imagery, the NAIP, which is collected during the growing

31

543   season, with leaf-on conditions and is therefore affected by tree canopy. Moreover, the better

544   performance for Site 4, even when using few training data, suggests that this landscape was

545   particularly well-suited to the deep learning approach. This may be due to the large distribution of

546   wetlands in Site 4, leading to a higher quantity of wetlands in the entire training data set as well as

547   more significant presence of wetlands in each training image.

548   Figure 7 also shows model predictions when using the Site 4-specific model. The Site 4

549   model produced predictions with the highest recall scores of all model trials for sites 1, 2, and 3.

550   As indicated by the increases in recall, predictions resulting from the Site 4 more densely

551   encompassed the ground truth wetlands (Figure 7, column C), relative to results for the site-

552   specific models (Figure 7, columns A and B). Attributing to the lower precision scores also

553   produced by the Site 4 model, wetland overprediction is apparent in the scenes for site 1, 2, and 3

554   (Figure 7, column C). The wetland predictions for these sites are also made at a coarse resolution

555   within image tile extents, evident by the rectangular edges of wetland predictions in sites 1 and 3

556   (Figure 7, column C). In addition, a segment of a narrow wetland feature is omitted for Site 3 when

557   applying the model trained for Site 4. Overall, these shortcomings demonstrate the potential for

558   bias to a specific landscape and wetland type in site-specific models, which may lead to decreased

559   accuracies when applied to different landscapes. This may be overcome by changing the

560   classification strategy away from a simple wetland/non-wetland classification to one that classifies

561   different wetland types, although this strategy was not explored through this research. The increase

562   in recall scores when using the Site 4 model, and the concentration of wetland overprediction

563   occurring in the adjacent and surrounding areas of the ground truth wetlands, suggests the noted

564   shortcomings may also be addressed by using a more balanced sampling of different wetland types.
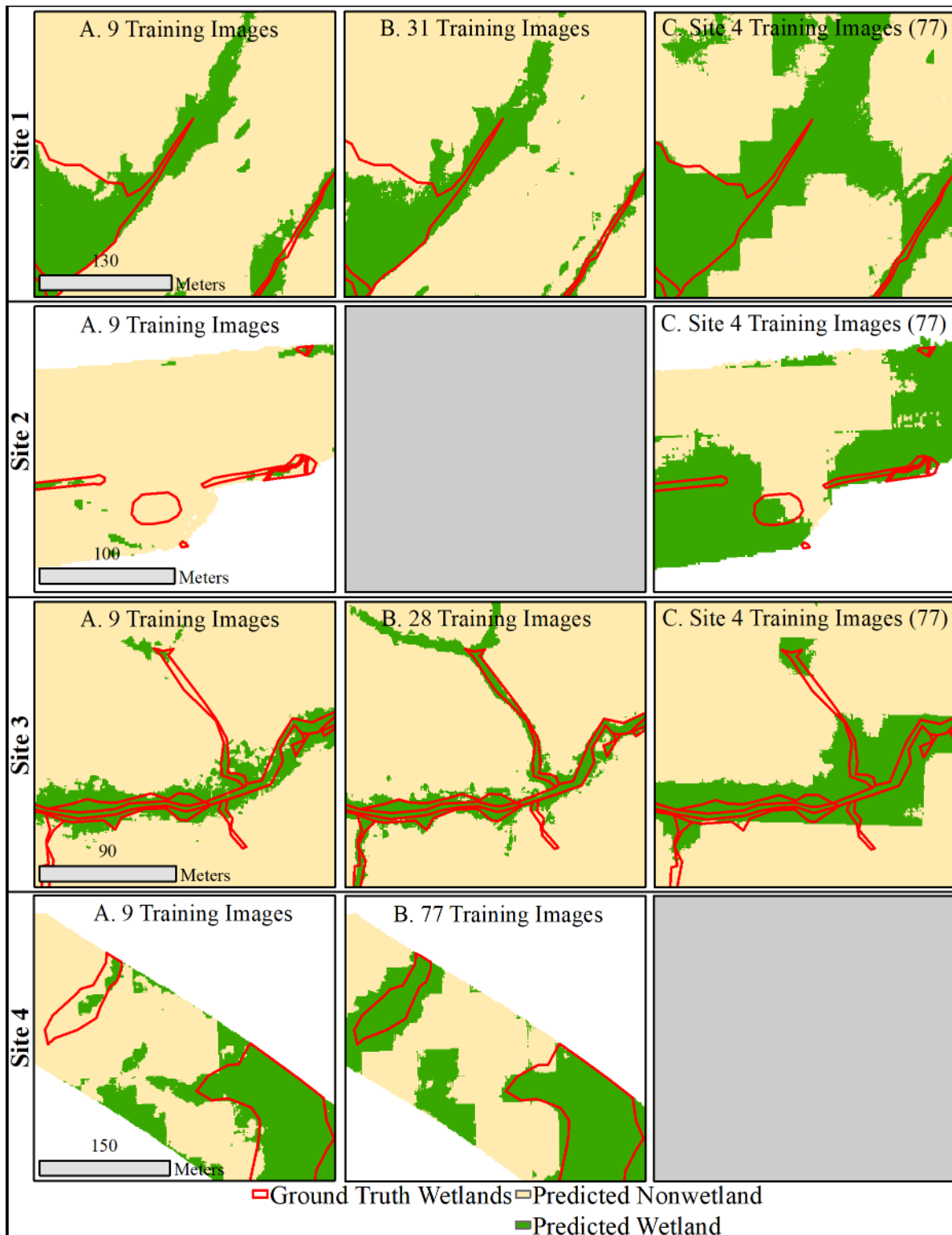
Figure 7. Comparison of wetland predictions produced by site-specific models created from (column A) the smallest training dataset and (column B) the largest training dataset available for the site. Also shown are wetland predictions produced by models trained only with the largest training dataset for Site 4 (column C).

569

## Potential for Combined-Site Models

571   Compared to the site-specific models, the general model mostly resulted in more wetland

572   overprediction, but in some cases increased coverage of ground truth wetlands (Figure 8, column

573   B). This trend is likely due to the bias of the general model to favor wetland types present in the

574   Site 4 landscape, as more than half of all the training images used were from Site 4. While the

575   general model results do not present an improvement from the site-specific models, there are

576   improvements compared to wetland predictions resulting from a model trained only on Site 4 (see

577   Figure 7, column C). By supplementing the Site 4 training data with wetland information from

578   other landscapes, we see finer, more precise wetland prediction boundaries (Figure 8, Site 1 B and

579   Site 3 B). For Site 2, the general model produced a greater overall amount of wetland predictions

580   compared to the site-specific model, but predictions were inaccurate (Figure 8, column A vs.

581   column B). However, the quantity of erroneous wetland predictions for Site 2 was greater when

582   using the Site 4 model versus the general model. It was expected that predictions for Site 4 would

583   be mostly unchanged between the site-specific model and the general model, due to the significant

584   presence of Site 4 training data. However, the weak training data influence from other sites did

585   slightly improve precision for Site 4, demonstrated by finer-scale edges of wetland predictions

586   (Figure 8 Site 4 A vs. Site 4 B).

587   The ecoregion model explored the potential for creating combined-site models that are

588   specific to certain landscape characteristics by including training data only from within the same

589   ecoregion (i.e., sites 2 and 3). Fewer wetland predictions were made overall for Site 2 using the

590   ecoregion model (Figure 8, column C), which considerably reduced recall compared to the general

591   model, but also resulted in sparser correct wetland predictions than the Site 2-specific model. For

592   Site 3, the ecoregion model improved both precision and recall compared to the general model,

593   but results were still less accurate than the site-specific model. Compared to general model

594   predictions, the ecoregion model regained correct wetland predictions for narrow, riparian wetland

595   features for Site 3 (Figure 8, column C). The ecoregion model also reduced wetland overprediction

596   compared to the general and site-specific models in the scenes shown in Figure 8, representative

597   of the higher precision produced by the ecoregion model (22% vs. 20% by the site-specific model

598   and 15% by the general model). However, wetland predictions resulting from the ecoregion model

599   encompassed less ground truth wetland area overall relative to the Site 3-specific model.

600   Although neither approach for creating a combined-site model was able to outperform site-

601   specific models, results show potential to refine and improve these methods. We found that the

602   relatively poor performances of the general and ecoregion models were not likely caused by the

603   unequal sampling of training data from the different geographic study areas. To investigate this

604   potential source of error, the general model and the ecoregion model were recreated by limiting

605   training data from sites to just nine images each, balancing the representation from each site. For

606   all sites, the general model built with equal, but limited training data performed worse than the

607   proposed general model. For Site 3, the ecoregion model built with limited training data performed

608   considerably worse, where recall decreased from 77% to 30% and precision improved slightly

609   from 15% to 17%. For Site 2, however, the limited ecoregion model improved results slightly

610   (recall increasing from 21 to 27% and precision remaining at 2%), but still not to an acceptable

611   level of accuracy. Thus, improving the combined-site model approach may not just be a matter of

612   equally sampling different landscapes, but also balancing an adequate amount of training data from

613   different landscapes. Lastly, the lack of consistent improvement to Site 2 and Site 3 predictions

614   when applying the ecoregion model suggests it would be beneficial to consider additional

615   landscape similarities when building combined-site models. Landscape characteristics to consider

616   may be those that affect the distributions of topographic inputs, such as influence of built

617   environment drainage and land cover.

618
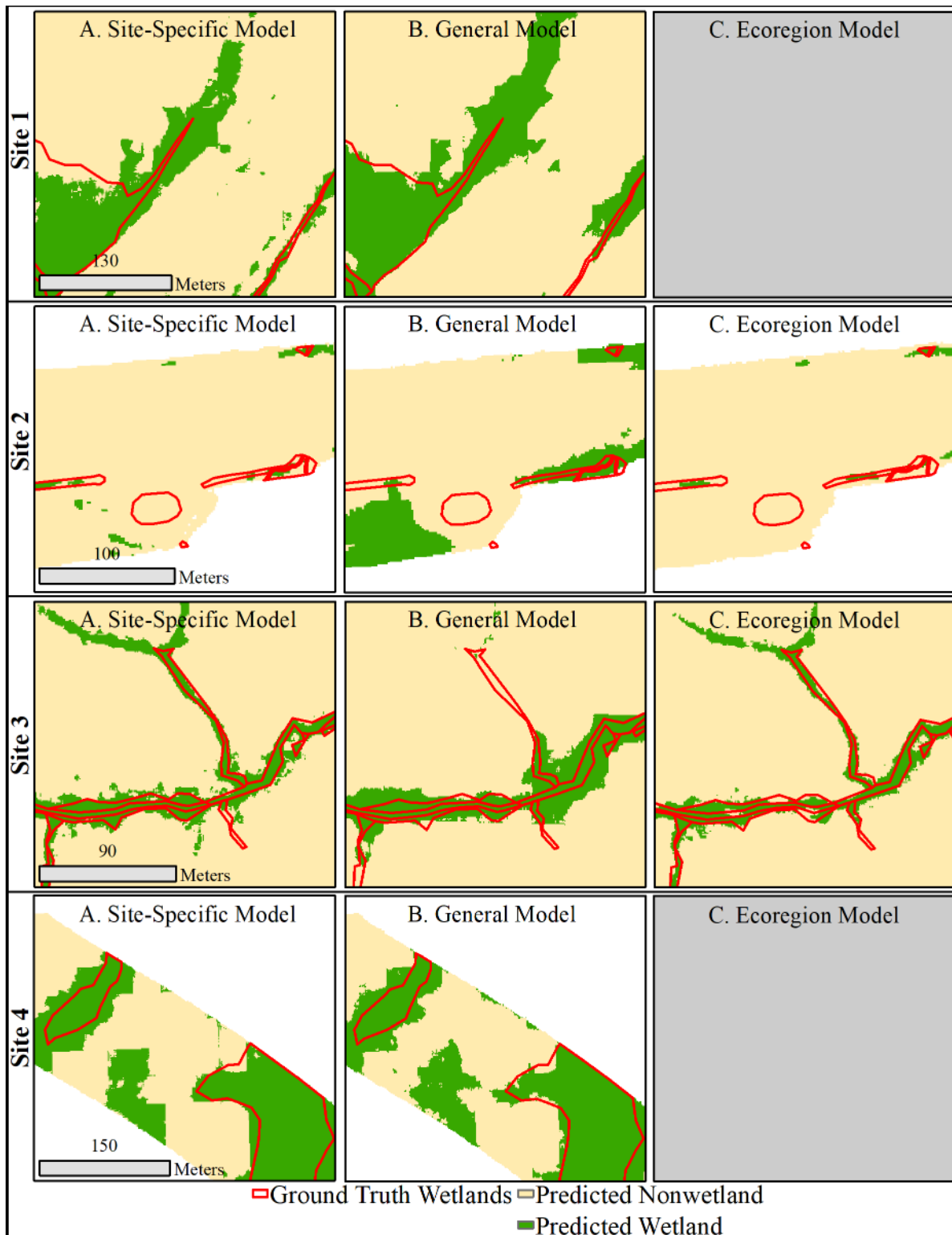
Figure 8. Comparison of wetland predictions resulting by (column A) the best performing site-specific models (i.e.,

those trained on 70% of the validation area), (column B) the general model, and (column C) the ecoregion model.

Utility of the Proposed Input Data Configuration

621

622   This study explored an input data configuration unique to most deep learning applications

623   where topographic derivatives of the input "image" (i.e., LiDAR DEM) are predetermined and

624   specific to the target object (i.e., wetlands). The hypothesis was that predetermined elevation

625   derivatives (TWI, DTW, and curvature) would improve wetland classification training by

626   including hydrologic information, compared to training directly from the elevation data. To

627   evaluate the efficacy of this method, we compared the accuracy achieved using our novel input

628   data configuration versus two-band images composed of the LiDAR DEM and the NDVI, which

629   is more representative of the common input data approach taken (e.g., Audebert et al., 2017, 2018;

630   Latifovic et al., 2018; Liu et al., 2018; Silburt et al., 2018; Xu et al., 2018). The LiDAR DEMs

631   used to create the two-band images were smoothed and hydrologically corrected, as suggested by

632   O'Neil et al. (2019), and 70% of the areas were used for training for both model sets.

633   For sites 1, 2, and 3, the proposed input data configuration outperformed the typical approach

634   in terms of both recall and precision. Wetlands predicted from only the DEM and NDVI for Site

635   1 achieved lower recall (73% vs. 81%) and precision (21% vs. 25%) compared to the models using

636   the derived topographic indices and the NDVI. This suggests that combining physical

637   understanding of the system, in this case hydrological and ecological characteristics of wetlands,

638   helps to guide the deep learning algorithm so that it is able to obtain increased predictive skill. For

639   Site 2, predictions learned from the DEM and NDVI encompassed only 12% of the ground truth

640   wetlands with near 0% precision, compared to 28% recall and 3% precision achieved by the

641   proposed approach. Wetland predictions for Site 3 lost considerable accuracy with the typical input

642   data approach, producing 24% recall and 9% precision, whereas our approach resulted in 85%

643   recall and 20% precision. For Site 4, this comparison showed that the model that learned from the

644 DEM and NDVI alone produced a higher recall (96% vs. 91%) and lower precision (49% vs. 56%).

645 While this indicates that more ground truth wetlands were detected using the typical approach, it

646 is slightly outweighed by the loss in wetland precision. Considering the consistent improvement

647 to the other three sites, the lack of significant change in Site 4 when applying only the DEM and

648 NDVI may suggest that the deep learning model relies more heavily on the vegetative

649 characteristics provided by the NDVI than the geomorphologic and hydrologic information that

650 the elevation data offers. This is likely due to the fact that Site 4 had the least topographic relief,

651 being within the coastal plain. Results for Site 4 using a random forest classification (see O'Neil

652 et al., 2019) also support this idea, showing that the topographic input variables were insufficient

653 for describing wetland characteristics unless preprocessing methods were calibrated specifically

654 to the area. Thus, it is logical that wetlands in Site 4 are better described by vegetative

655 characteristics than topography, explaining the lack of change in predictions when replacing the

656 topographic inputs with the DEM and leaving the NDVI input unchanged.

657 **Comparison of Deep Learning to a Random Forest Implementation**

658   To examine the potential for deep learning to advance the more commonly used random forest

659 approach for wetland classification (e.g., O'Neil et al., 2019), we compared the performance of

660 the site-specific deep learning models to a random forest classification with the same set of input

661 variables. The random forest implementation follows the approach of O'Neil et al. (2019), but with

662 the addition of the NDVI to the original set of inputs: the TWI, curvature, and DTW. The training

663 sampling used in the O'Neil et al. (2019) study was maintained, where training data consists of

664 randomly dispersed pixels that encompass only 15% of the validated wetland area and up to 8%

665 of the validated nonwetland area. However, accuracy assessments for both the deep learning and

666   random forest models were limited to the extents of the testing image tiles that correspond to the

667   deep learning approach.

668   Compared to the Site 1 deep learning model, the random forest classification resulted in an

669   improvement in recall from 81% to 91%, but a decrease in precision from 25% to 19%. For Site

670   2, Random Forest improved recall considerably, from 28% to 78%, and slightly improved

671   precision from 3% to 5%. The Site 3 random forest model produced no change in recall (85%) and

672   a slight decrease in precision (20% vs. 18%), compared to deep learning. Finally, the Site 4 random

673   forest model considerably decreased recall from 91% to 70% and increased precision from 56%

674   to 64%, relative to the deep learning model. With the exception of Site 2, these findings show that

675   deep learning was able to perform similarly to random forests (e.g., Site 1 and Site 3), and arguably

676   better in some cases (e.g., Site 4). The poor performance in Site 2 further supports that the deep

677   learning model was not sufficiently able to learn characteristics of wetland features that were very

678   small and sparse relative to the landscape scenes in each training image. Similarly, the Site 4 results

679   again support the idea that deep learning is better suited to detecting wetlands where they are areal

680   and large relative to the landscape scene. In addition, an evaluation of the entire testing areas

681   corresponding to the random forest models shows that the inclusion of the NDVI as a wetland

682   indicator improves on the O'Neil et al. (2019) approach. Compared to the random forest models

683   using only the topographic inputs, the addition of the NDVI improved wetland recall and precision

684   in Site 1 (81% vs. 88% and 19% vs. 24%), Site 2 (82% vs. 88% and 16% vs. 22%), Site 3 (83%

685   vs. 86% and 22% vs. 25%), and Site 4 (58% vs. 68% and 47% vs. 54%).

686   Overall, it is important to note that the random forest models were able to achieve these

687   accuracies by sampling much less training data than was required for deep learning models.

688   However, this result also shows that deep learning models can approach the same accuracies using

689   training data resources that are considerably smaller relative to most deep learning applications. In

690   addition, the similar performance of deep learning to random forests in three of the study sites

691   supports findings by other researchers that state deep learning can improve landscape segmentation

692   accuracy over traditional machine learning, such as support vector machine, maximum likelihood

693   classification, and random forests, given enough training data (e.g., Hu et al., 2018; Latifovic et

694   al., 2018; Liu et al., 2018; Mahdianpari et al., 2018).

695   **Limitations**

696   Limitations of this approach could be addressed through additional research. For example,

697   incorporating Class Activation Mapping (CAM) (Zhou et al., 2016), which highlights scene

698   elements that are most influential during classifications, would offer further insight into model

699   learning. By utilizing CAM, model refinements could be made by quantifying the impact of the

700   input data and identifying sources of error. Considering additional remote sensing data may also

701   improve model performance. These may include LiDAR point clouds, which researchers have

702   incorporated into 3-dimensional CNNs for wetland identification (e.g., Xu et al., 2018). Also,

703   incorporating radar data may reduce errors where the NDVI is affected by tree canopy, as it is able

704   to penetrate this layer and provide vegetation density and inundation information for wetland

705   mapping (Allen et al., 2013; Behnamian et al., 2017; Corcoran et al., 2013; Kloiber et al., 2015;

706   Millard & Richardson, 2013). Also on this point, the contribution of each input data source

707   throughout the DeepNets workflow can be handled in a more sophisticated way. This was

708   demonstrated by Audebert et al. (2018) who proposed novel data fusing methods for elevation data

709   and the NDVI within the DeepNets workflow to improve land cover classifications.

710   Additional training information that consists of accurately delineated wetlands from across

711   different ecoregions should improve the deep learning classification results. Also, additional

712   training data would make it possible to train models for specific wetland types rather than a simple,

713   binary wetland/nonwetland classification. These training data are likely available from state and

714   federal agencies given the need for wetland assessments under the Clean Water Act, but are not

715   collected into a single, standardized repository. Future work could focus on building such a

716   training and testing repository for wetland classification. Furthermore, to more efficiently make

717   use of any amount of reliable training information available, applying more sophisticated data

718   augmentation techniques may improve wetland predictions, as demonstrated by Stivaktakis et al.

719   (2019).

720   Refinements to the current approach should also include more robust accuracy assessments.

721   The current accuracy metrics are transparent and represent the two factors that are needed for

722   reliable implementation: coverage of ground truth wetlands and limited overprediction. However,

723   a single accuracy metric that encompasses both of these factors while also acknowledging the

724   significantly higher importance of wetland recall would improve the interpretation of model

725   results. Model evaluation improvements should also take into account the diffuse boundaries of

726   wetlands which may fluctuate seasonally by penalizing overprediction less if it occurs adjacent to

727   or surrounding defined ground truth wetland extents. Lastly, this study did not test the effect of

728   tuning the DeepNets parameters. Among other parameter adjustments, future work should explore

729   the benefit of adjusting window sizes based on target wetland size and the accuracy tradeoffs when

730   training the model for more epochs.

Conclusions

731

732   We explore a wetland identification workflow that implements a basic semantic

733   segmentation architecture and an input data configuration that consists of the NDVI and LiDAR

734   DEM-derived indicators of wetland hydrology and geomorphology. The workflow was trained and

735 evaluated using available data resources from four geographic regions of Virginia. From this work,

736 we draw the following conclusions.

737   i.   Site-specific deep learning models created from relatively small training datasets

738   can achieve accurate results. For three of the four study sites, wetland recall ranged

739   from 81 to 91% and precision ranged from 20 to 56%, when training models with

740   70% of site area and testing on the remaining 30% of the site area.

741   ii.   Site-specific models were more successful for areas where wetlands are abundant

742   and occupy a significant portion of training images. For a site with large, areal

743   wetlands that were almost evenly balanced with nonwetland areas, high accuracy

744   was achieved with 7.5 km^2 (70%) of training area (91% recall and 56% precision).

745   Using a much smaller training area, 0.4 km^2 (10% of the study area), still resulted

746   in a fairly accurate model (84% recall and 50% precision).

747   iii.   In most cases, accuracy decreased when using models trained for another site.

748   However, the site-specific model trained with the largest area studied (7.5 km^2)

749   increased wetland recall in all other sites. Although model predictions were

750   imprecise and showed a bias towards the types of wetlands for which it was trained

751   (i.e., large, areal wetlands), the correct localization of wetland predictions suggests

752   there is potential for this approach if models are trained with sufficient data and for

753   areas with similar landscapes.

754   iv.   Combined-site models can produce accurate wetland predictions, but training data

755   contributions from the target landscapes should be balanced. The general model

756   revealed the potential for bias towards landscape characteristics more heavily

757   represented in the training data. However, the influence of less represented sites

was still apparent, as wetland predictions were more inclusive of different wetland types compared to a model created without training data from these sites.

v. Shared ecoregion alone may not offer sufficient landscape similarities to improve the training sampling approach for combined-site models. The ecoregion model showed accuracy improvements from the general model for one site. However, wetland predictions for the other site were less accurate. Future work should explore the benefit of creating combined-site models from areas that share additional characteristics that would affect the distributions of the topographic derivatives, such as level of development, land cover, and topography.

vi. The proposed input data configuration improves wetland identification compared to a more typical approach of using the NDVI and the LiDAR DEM alone. By predetermining the derivatives of the DEM that are wetland indicators based on physical understanding of hydrology and wetland formation, rather than allowing the deep learning network to determine these through convolutions on raw data, wetland predictions were more accurate in three sites. This speaks to the benefit and power of combining physical understanding along with machine and deep learning algorithms for improved predictive skill. For the remaining site, accuracy was nearly unchanged between the two approaches. However, analyses show that this is likely due to the greater importance of the NDVI for identifying wetlands in the topographically mild landscape.

vii. Compared to a random forest approach, the best performing models produced comparable accuracy, using more training data than required for random forest, but still significantly less than what is typical in most deep learning applications.

781   Our results demonstrate the potential for deep learning to not only improve accuracy

782   compared to traditional machine learning algorithms, but also provide flexible models that are

783   accurate for a range of landscapes. Paramount to achieving this will be larger efforts within the

784   research community to gather reliable training data and pretrained models stored as open source

785   repositories, as has been done for established deep learning fields (e.g., Lecun, 1999; Lin et al.,

786   2014). The wetland models created through this research may offer a starting point for creating a

787   repository open to other researchers. By refining this implementation of the deep learning wetland

788   workflow and further training the created models, there is potential for deep learning to support a

789   range of wetland conservation efforts by producing accurate wetland inventories across many

790   landscapes.

References

(VITA), V. I. T. A. (2016). Virginia Base Map Data Downloads. Retrieved August 1, 2016, from http://vgin.maps.arcgis.com

Ågren, A. M., Lidberg, W., 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

Audebert, N., Le Saux, B., & Lefèvre, S. (2016). Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In *Asian conference on computer vision* (pp. 180-196). Springer, Cham.

Audebert, N., Le Saux, B., & Lefèvre, S. (2017). Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks. In S.-H. Lai, V. Lepetit, K. Nishino, & Y. Sato (Eds.), *Computer Vision -- ACCV 2016* (pp. 180–196). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-54181-5_12

Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, 20–32. https://doi.org/10.1016/j.isprsjprs.2017.11.011

Audebert, N., Le Saux, B., & Lefèvre, S. (2019). Deep Learning for Earth Observation. Retrieved January 1, 2019, from https://github.com/nshaud/DeepNetsForEO

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, *5*(5), 423–431. https://doi.org/10.1080/2150704X.2014.915434

Band, L. E. (1986). Topographic Partition of Watersheds with Digital Elevation Models. *Water Resources Research*, *22*(1), 15–24. https://doi.org/10.1029/WR022i001p00015

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

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

Carlson, T. N., & Riziley, D. A. (1997). On the Relation between NDVI , Fractional Vegetation
Cover , and Leaf Area Index. *Remote Sensing of Environment*, *62*(3), 241–252.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data.
University of California, Berkeley, (1999), 1–12.
https://doi.org/ley.edu/sites/default/files/tech-reports/666.pdf

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

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., ... & Raska, R. (2018,

June). Deepglobe 2018: A challenge to parse the earth through satellite images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 172-17209). IEEE.

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*. (No. 17-01199).

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

Dronova, I., Gong, P., & Wang, L. (2011). Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China. *Remote Sensing of Environment*, *115*(12), 3220–3236. https://doi.org/10.1016/j.rse.2011.07.006

Farm Service Agency. (2017). National Geospatial Data Asset (NGDA) National Agriculture Imagery Program (NAIP) Imagery. Retrieved January 1, 2019, from https://catalog.data.gov/dataset/national-geospatial-data-asset-ngda-naip-imagery

Ghosh, A., Ehrlich, M., Shah, S., Davis, L. S., & Chellappa, R. (2018, June). Stacked U-Nets for Ground Material Segmentation in Remote Sensing Imagery. In *CVPR Workshops* (pp. 257-261).

GRASS Development Team. (2017). Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2. Retrieved from http://grass.osgeo.org

Grimaldi, S., Nardi, F., Benedetto, F. Di, 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). Formal Basis for the Heuristic Determination eijj ,. *Systems Science and Cybernetics*, (2), 100–107.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *IEEE conference on computer vision and pattern recognition* (pp. 770–778). Retrieved from http://arxiv.org/abs/1512.03385

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hogg, A. R., & Todd, K. W. (2007). Automated discrimination of upland and wetland using terrain derivatives. *Canadian Journal of Remote Sensing*, *33*(July), S68–S83. https://doi.org/10.5589/m07-049

Holmgren, P. (1994). Multiple flow direction algorithms for runoff modelling in grid based elevation models: An empirical evaluation. *Hydrological Processes*, *8*(4), 327–334. https://doi.org/10.1002/hyp.3360080405

Hu, Y., Zhang, Q., Zhang, Y., & Yan, H. (2018). A Deep Convolution Neural Network Method for Land Cover Mapping: A Case Study of Qinhuangdao, China. *Remote Sensing*, *10*(12), 2053. https://doi.org/10.3390/rs10122053

Jenson, S. K., & Domingue, J. O. (1988). Extracting topographic structure from digital elevation

data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing*, *54*(11), 1593–1600. https://doi.org/0099-1112/88/5411-1593$02.25/0

Jyotsna, R., & Haff, P. K. (1997). Microtopography as an indicator of modern hillslope diffusivity in arid terrain. *Geology*, *25*(8), 695–698. Retrieved from http://dx.doi.org/10.1130/0091-7613(1997)025%3C0695:MAAIOM%3E2.3.CO

Kemker, R., Salvaggio, C., & Kanan, C. (2018). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*(April), 60–77. https://doi.org/10.1016/j.isprsjprs.2018.04.014

Kemker, R., Gewali, U. B., & Kanan, C. (2018). EarthMapper: A Tool Box for the Semantic Segmentation of Remote Sensing Imagery, 1–5. Retrieved from http://arxiv.org/abs/1804.00292

Kim, M., Warner, T. A., Madden, M., & Atkinson, D. S. (2011). Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *International Journal of Remote Sensing*, *32*(10), 2825–2850.

Klemas, V. (2011). Remote Sensing of Wetlands: Case Studies Comparing Practical Techniques. *Journal of Coastal Research*, *27*(3), 418–427. https://doi.org/10.2112/JCOASTRES-D-10-00174.1

Kloiber, S. M., Macleod, R. D., Smith, A. J., Knight, J. F., & Huberty, B. J. (2015). A Semi-Automated, Multi-Source Data Fusion Update of a Wetland Inventory for East-Central Minnesota, USA. *Wetlands*, *35*(2), 335–348. https://doi.org/10.1007/s13157-014-0621-3

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Lang, M., & McCarty, G. (2014). Light Detection and Ranging (LiDAR) for Improved Mapping of Wetland Resources and Assessment of Wetland Conservation Projects, (September), 7. Retrieved from

http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1260970.pdf

Lang, M., McCarty, G., Oesterling, R., & Yeo, I. Y. (2013). Topographic metrics for improved mapping of forested wetlands. *Wetlands*, *33*(1), 141–155. https://doi.org/10.1007/s13157-012-0359-8

Latifovic, R., Pouliot, D., & Campbell, J. (2018). Assessment of Convolution Neural Networks for Surficial Geology Mapping in the South Rae Geological Region, Northwest Territories, Canada. *Remote Sensing*, *10*(2), 307. https://doi.org/10.3390/rs10020307

Lecun, Y. (1999). THE MNIST DATABASE of handwritten digits. Retrieved from https://ci.nii.ac.jp/naid/10027939599/en/

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., & others. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lee, C. suk, Sohn, E., Park, J. D., & Jang, J.-D. (2019). Estimation of soil moisture using deep learning based on satellite data: a case study of South Korea. *GIScience & Remote Sensing*, *56*(1), 43–67. https://doi.org/10.1080/15481603.2018.1489943

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft

coco: Common objects in context. In *European conference on computer vision* (pp. 740–755).

Lindsay, J. B. (2016). Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrological Processes*, *30*(6), 846–857. https://doi.org/10.1002/hyp.10648

Lindsay, J. B., & Creed, I. F. (2005). Removal of artifact depressions from digital elevation models: Towards a minimum impact approach. *Hydrological Processes*, *19*(16), 3113–3126. https://doi.org/10.1002/hyp.5835

Liu, T., Abd-Elrahman, A., Morton, J., & Wilhelm, V. L. (2018). Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience and Remote Sensing*, *55*(2), 243–264. https://doi.org/10.1080/15481603.2018.1426091

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *130*, 277–293. https://doi.org/10.1016/j.isprsjprs.2017.06.001

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., & Zhang, Y. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing*, *10*(7), 1119.

https://doi.org/10.3390/rs10071119

Marmanis, D., Schindler, K., Wegner, J. D., Galliani, S., Datcu, M., & Stilla, U. (2018). Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, 158-172. http://dx.doi.org/10.1016/j.isprsjprs.2017.11.009

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

Mui, A., He, Y., & Weng, Q. (2015). An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *109*, 30–46. https://doi.org/10.1016/j.isprsjprs.2015.08.005

Murphy, P. N. C., Ogilvie, J., Connor, K., & Arp, P. A. (2007). Mapping wetlands: A

comparison of two different approaches for New Brunswick, Canada. *Wetlands*, *27*(4), 846–

854. https://doi.org/10.1672/0277-5212(2007)27[846:MWACOT]2.0.CO;2

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

O'Neil, G. L., Saby, L., Band, L. E., & Goodall, J. L. (2019). Effects of LiDAR DEM Smoothing

and Conditioning Techniques on a Topography-Based Wetland Identification Model. *Water

Resources Research*, 2019WR024784. https://doi.org/10.1029/2019WR024784

Oltean, G. S., Comeau, P. G., & White, B. (2016). Linking the depth-to-water topographic index

to soil moisture on boreal forest sites in Alberta. *Forest Science*, *62*(2), 154–165.

https://doi.org/10.5849/forsci.15-054

Ozesmi, S. L., & Bauer, M. E. (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management*, *10*(5), 381–402.

Page, R. W., & Wilcher, L. S. (1990). Memorandum of Agreement Between the Environmental Protection Agency and the Department of the Army concerning the determination of mitigation under the Clean Water Act, Section 404 (b)(1) Guidelines. *Washington, DC, USA*.

Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving Precipitation Estimation Using Convolutional Neural Network. *Water Resources Research*, *55*(3), 2301–2321. https://doi.org/10.1029/2018WR024090

Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G., & Dietrich, W. E. (2010). A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths. *Journal of Geophysical Research*, *115*(F1), F01002. https://doi.org/10.1029/2009JF001254

Passalacqua, P., Tarolli, P., & Foufoula-Georgiou, E. (2010). Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape. *Water Resources Research*, *46*(11), 1–17. https://doi.org/10.1029/2009WR008812

Passalacqua, P., Belmont, P., & 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

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., et al. (2017). Automatic Differentiation in PyTorch. In *NIPS Autodiff Workshop*.

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

PyGeoNet. (2019). Retrieved August 1, 2018, from https://github.com/passaH2O/GeoNet

Rampi, L. P., Knight, J. F., & Pelletier, K. C. (2014). Wetland Mapping in the Upper Midwest United States. *Photogrammetric Engineering & Remote Sensing*, *80*(5), 439–448. https://doi.org/10.14358/pers.80.5.439

Rezaee, M., Mahdianpari, M., Zhang, Y., & Salehi, B. (2018). Deep Convolutional Neural Network for Complex Wetland Classification Using Optical Remote Sensing Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(9), 3030–3039. https://doi.org/10.1109/JSTARS.2018.2846178

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

Scott, G. J., England, M. R., Starms, W. A., Marcum, R. A., & Davis, C. H. (2017). Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution

Imagery. *IEEE Geoscience and Remote Sensing Letters*, *14*(4), 549–553.

https://doi.org/10.1109/LGRS.2017.2657778

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative

theory of immediate visual recognition. *Progress in Brain Research*, *165*, 33–56.

https://doi.org/10.1016/S0079-6123(06)65004-8

Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for

Water Resources Scientists. *Water Resources Research*, *54*(11), 8558–8593.

https://doi.org/10.1029/2018WR022643

Silburt, A., Ali-dib, M., Zhu, C., & Jackson, A. (2018). Lunar Crater Identification via Deep

Learning.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image

recognition. *ArXiv Preprint ArXiv:1409.1556*.

Snyder, G. I., & Lang, M. (2012). Significance of a 3D Elevation Program to wetland mapping.

National Wetlands Newsletter, *34*(5), 11–15. Retrieved from

http://pubs.er.usgs.gov/publication/70193349

Stivaktakis, R., Tsagkatakis, G., & Tsakalides, P. (2019). Deep Learning for Multilabel Land

Cover Scene Categorization Using Data Augmentation. *IEEE Geoscience and Remote

Sensing Letters*, *16*(7), 1–5. https://doi.org/10.1109/lgrs.2019.2893306

Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for

classification of imbalanced data. *Pattern Recognition*, *40*(12), 3358–3378.

https://doi.org/10.1016/j.patcog.2007.04.009

Tarboton, D. G. (1991). On the extraction of channel networks from digital elevation data. *Hydrological Processes*, *5*(1), 81–100. https://doi.org/10.1002/hyp.3360050107

Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sensing*, *8*(11), 954. https://doi.org/10.3390/rs8110954

US Corps of Engineers. (1987). Corps of Engineers Wetlands Delineation Manual and Regional Supplements. US Army Engineer Waterways Experiment Station Vicksburg, Mississippi. Retrieved from http://www.epa.gov/cwa-404/section-404-clean-water-act-how-wetlands-are-defined-and-identified

USGS. (2019). USGS National Hydrography Dataset. Retrieved August 1, 2017, from https://www.usgs.gov/core-science-systems/ngp/national-hydrography/national-hydrography-dataset

Volpi, M and Tuia. D. (2016). Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing. http://dx.doi.org/10.1109/TGRS.2016.2616585

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

Xu, Z., Guan, K., Casler, N., Peng, B., & Wang, S. (2018). A 3D convolutional neural network

    method for land cover classification using LiDAR and multi-temporal Landsat imagery.

    ISPRS Journal of Photogrammetry and Remote Sensing, *144*(August), 423–434.

    https://doi.org/10.1016/j.isprsjprs.2018.08.005

Zhang, L., Zhang, L., & Du, B. (2016). Deep Learning for Remote Sensing Data: A Technical

    Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 22–

    40. https://doi.org/10.1109/MGRS.2016.2540798

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features

    for Discriminative Localization. Retrieved from http://arxiv.org/abs/1512.04150

Zhu, J., & Pierskalla, W. P. (2016). Applying a weighted random forests method to extract karst

    sinkholes from LiDAR data. *Journal of Hydrology*, *533*, 343–352.

    https://doi.org/10.1016/j.jhydrol.2015.12.012