



## **Predictive mapping of wetland soil types in the Canadian Prairie Pothole Region using high-resolution digital elevation model terrain derivatives**

Authors: Kiss, Jeremy, Bedard-Haughn, Angela, and Sorenson, Preston

Source: Canadian Journal of Soil Science, 103(1) : 21-46

Published By: Canadian Science Publishing

URL: <https://doi.org/10.1139/cjss-2022-0034>

---

BioOne Complete ([complete.BioOne.org](https://complete.BioOne.org)) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](https://www.bioone.org/terms-of-use).

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

# Predictive mapping of wetland soil types in the Canadian Prairie Pothole Region using high-resolution digital elevation model terrain derivatives

Jeremy Kiss , Angela Bedard-Haughn , and Preston Sorenson 

Department of Soil Science, University of Saskatchewan, Saskatoon, SK S7N 5A8, Canada

Corresponding author: Jeremy Kiss (email: [jeremy.kiss@usask.ca](mailto:jeremy.kiss@usask.ca))

## Abstract

Wetland soil types, which can be distinguished based on calcium carbonate content, vary in their effect on ecosystem functions like phosphorus retention, salinity contributions, and greenhouse gas forcing. These soil types may be predictively mapped with machine learning models that use terrain derivatives calculated from high-resolution digital elevation models. Soil profiles from three Saskatchewan study sites were classified into three functional categories—upland, calcareous wetland, or noncalcareous wetland—and used to train random forest models for predictive soil mapping. Multiple terrain derivatives were included as predictor variables to capture local- and landscape-scale morphometry and hydrology influences, including five derivatives developed for this study. Models were developed at three spatial resolutions: 2, 5, and 10 m, and tested via internal cross-validation and independent validation with datasets from previous studies. Predictive accuracies were highest when mapping at 2 m resolution (independent validation accuracy range = 64%–100%) but also successful when mapping at 5 and 10 m resolutions (independent validation accuracy range = 63%–100%); however, visual inspection determined that the maps generated at 10 m resolution were less detailed and occasionally featured questionable discontinuous soil distributions. Three of the five terrain derivatives developed for this study were among the most important predictor variables (first, second, and 10th most important). Models trained using only data from a specific site had slightly better performance than models trained using data from all sites, except in regions where training data were lacking.

**Key words:** Predictive soil mapping, wetlands, Prairie Pothole Region, hydrogeology, calcium carbonate

## Résumé

On différencie les sols des terres humides d'après leur teneur en carbonate de calcium et ces sols influent de diverses manières sur l'écosystème (rétention du phosphore, hausse de la salinité, forçage des gaz à effet de serre). Il est possible d'en établir une carte prédictive en recourant aux modèles à apprentissage automatique qui utilisent les dérivées du terrain venant de la modélisation numérique des hauteurs à haute résolution. Les auteurs ont classé le profil du sol de trois sites expérimentaux de la Saskatchewan en trois groupes fonctionnels – hautes terres, terres humides calcaires ou terres humides non calcaires –, puis ils s'en sont servis pour former les modèles à forêt aléatoire servant à tracer les cartes prédictives. Ils ont ajouté au modèle de nombreuses dérivées du terrain comme variables explicatives afin de saisir les influences morphométriques et hydrologiques locales ou propres au terrain. En faisaient partie les cinq dérivées élaborées dans le cadre de la présente étude. Les modèles ont été développés à trois résolutions spatiales : deux, cinq et dix mètres. Ensuite, les auteurs ont testé les modèles par validation croisée interne et par validation indépendante avec les jeux de données d'études antérieures. L'exactitude des prévisions atteint un maximum à la résolution de deux mètres (exactitude de la validation indépendante de 64 à 100 %), mais les prévisions étaient également correctes pour les cartes de la résolution de cinq ou de dix mètres (exactitude de la validation indépendante de 63 à 100 %). Une inspection visuelle a toutefois révélé que les cartes produites à la résolution de dix mètres sont moins détaillées et présentent parfois une distribution discontinue des sols douteuse. Trois des cinq dérivées du terrain développées dans le cadre de l'étude figuraient parmi les principales variables explicatives (la première, la deuxième et la dixième en importance). Les modèles formés avec les données d'un seul site fonctionnent légèrement mieux que ceux formés avec les données émanant de tous les sites, sauf dans les régions pour lesquelles on ne possède aucune donnée pour former le modèle. [Traduit par la Rédaction]

**Mots-clés :** carte prédictive des sols, terres humides, région des fondrières des Prairies, hydrogéologie, carbonate de calcium

## Introduction

Wetlands are recognized as a key element of the landscape for maintaining ecological integrity through the myriad of ecosystem services they provide. The degree to which wetlands provide certain ecosystem services depends in part on their biogeochemical characteristics. A key distinction in the biogeochemical characteristics for wetland soils in the Prairie Pothole Region (PPR) is their calcium carbonate ( $\text{CaCO}_3$ ) content, which is controlled by their hydrogeological development (Pennock et al. 2014). Calcareous and noncalcareous wetland soils are likely to vary in many functional respects, including their ability to reduce phosphorus (P) mobility, their effect as sources of salinization, and their contributions to greenhouse gas emissions.

In terms of watershed nutrient loading, wetlands are known to work as filters to reduce P mobility within the PPR (Badiou et al. 2018) and  $\text{CaCO}_3$  content has been identified as an important control on the availability of P in wetland soils. In a study by Brown et al. (2017b), calcareous and noncalcareous wetland soils had similar total P contents, but the calcareous soils had six times less available P than the noncalcareous soils. Interactions between wetland soil  $\text{CaCO}_3$  content and P mobility are highly complex (McFarlan 2021) and require further study to be quantified. With respect to soil salinization, wetland soil  $\text{CaCO}_3$  accumulation and general soil salinity accumulation develop through the same hydrological mechanisms, so where calcareous wetland soils are found, saline soils are also typically found (Pennock et al. 2014). In terms of greenhouse gas forcing, wetlands with soils enriched with  $\text{CaCO}_3$  often have increased sulfate content in the wetland pond water (Pennock et al. 2014), which can decrease the methane emitted by a wetland by providing an alternative electron acceptor to carbon dioxide during anaerobic oxidation of organic matter (Segers 1998).

Successful predictive mapping of calcareous and noncalcareous wetland soils would enable targeted sampling for projects focused on studying the differences in characteristics and services between these wetland soil types. Once relationships between these wetland soil types and ecosystem functions are established, predictive mapping of these soil types would help focus conservation and restoration efforts to manage ecosystem functions most effectively and could be used to establish broad-scale estimates of such functions. Developing successful methodology for detailed mapping of wetland soil extents is useful for other purposes, including establishing wetland buffer zones to inform cultivation boundaries, assessing the proportion of arable crop per quarter section for land valuation assessments, and estimating wetland soil areas for upscaled estimates of carbon stocks.

A model was proposed in Kiss and Bedard-Haughn (2021) for predicting the spatial distributions of PPR wetlands that are expected to have the greatest accumulations of  $\text{CaCO}_3$ . The modelling in that study predicts solute-richness classes for individual wetlands but does not attempt to map the distribution of  $\text{CaCO}_3$  within the wetland soils. It is important to note that calcareous soils are not evenly distributed within PPR wetlands (Pennock et al. 2014). The hydrogeological influences on the distributions of wetland soil types are highly complex; they include soil forming processes occurring over

landscape scales as well as within individual wetlands and are largely dependent on the wetland's relative position in the landscape and its relationship with groundwater (Fig. 1).

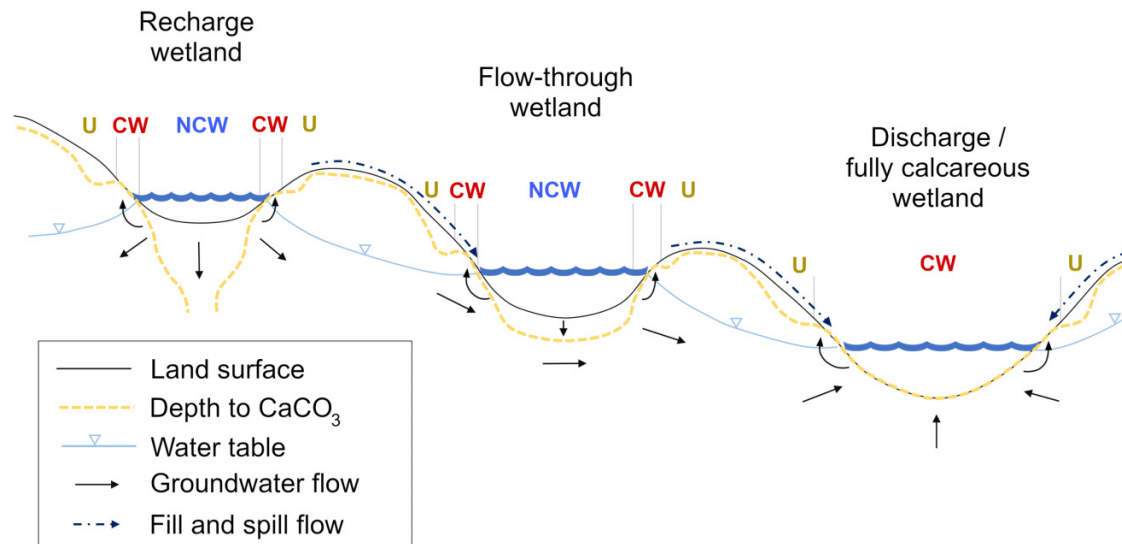
Wetlands are often classified as recharge, discharge, or flow-through (Arndt and Richardson 1988), where the dominant groundwater movement in recharge wetlands is downwards, causing  $\text{CaCO}_3$  to be leached out of the wetland soils and contrastingly, groundwater moves into discharge wetland soils, causing  $\text{CaCO}_3$  to accumulate within them (Pennock et al. 2014). Flow-through wetlands may recharge groundwater or receive groundwater discharge depending on water table levels (Winter and Rosenberry 1998) and often contain soils with reduced solute accumulations in the upper depths of their profile due to leaching during periods of dominantly downward movement of water (Arndt and Richardson 1989). In addition to deep groundwater controls, wetlands can receive significant inputs of solutes from fill and spill and shallow groundwater flow from upslope wetlands (Fig. 1) (Cook and Hauer 2007; Nachshon et al. 2013). This can result in fully calcareous wetlands that are characterized by  $\text{CaCO}_3$ -enriched soils throughout the wetland basin regardless of the wetland's relationship with groundwater (see Appendix A).

Regardless of wetland type (recharge, discharge, or flow through), a ring of calcareous discharge soils forms at the wetland fringe; this phenomenon has been observed at multiple sites across the climate gradient of the Saskatchewan region of the PPR (Pennock et al. 2014). Due to the increased hydraulic conductivity of the near-surface glaciolacustrine and oxidized tills compared to the deeper glacial tills, water within the wetland soil moves laterally outward from the wetland and then upward towards the soil surface at the wetland fringe through capillary rise (Fig. 1) (Knuteson et al. 1989; Hayashi et al. 1998; Heagle et al. 2013). Evaporation and evapotranspiration from plants in the wetland fringe contribute to this phenomenon (Hayashi et al. 1998). The upward movement of water to the wetland fringe redistributes  $\text{CaCO}_3$  to these positions causing the development of a ring of calcareous wetland soils surrounding the wetlands (Fig. 1) (Pennock et al. 2014).

These various hydrogeological processes may be approximated using terrain derivatives calculated from high-resolution light detection and ranging (LiDAR) derived digital elevation models (DEM) as relationships between such processes and local- and landscape-scale topographic characteristics have been observed (Pennock et al. 2014; Kiss and Bedard-Haughn 2021). Methodologies for calculating numerous terrain derivatives from DEMs have been developed in the last 30 years (Hengl and Reuter 2008); these have proven to be critical for predictive soil mapping (PSM) of a wide range of soil types and properties (McBratney et al. 2003). Pennock et al. (2014) and Bedard-Haughn and Pennock (2002) suggest a few terrain attributes that relate specifically to the distribution of PPR wetland soil types, such as specific dispersal area, elevation from wetland basin minimum, and position relative to maximum observed wetland water levels.

The spatial distributions of wetland soil types can likely be related to many terrain attributes and random forest models are well suited to determine relationships between soil types and large numbers of predictor variables, as they can determine relationships reflecting complex effects and inter-

**Fig. 1.** Schematic diagram of soil type distributions, depth to  $\text{CaCO}_3$ , groundwater table, and direction of groundwater and fill and spill flow based on diagrams from Van der Kamp and Hayashi (2009), Pennock et al. (2011), and Pennock et al. (2014). Soil types are indicated by the uppercase letters: U, upland; CW, calcareous wetland; NCW, noncalcareous wetland. Calcareous wetland and noncalcareous wetland soils are distinguished based on the presence of  $\text{CaCO}_3$  in the surficial depth of the profile. [Colour online.]



actions of the terrain derivatives as soil forming factors. Random forest models have been found to be excellent performers for mapping soil classes when compared to other model approaches (Brungard et al. 2015; Heung et al. 2016; Jeune et al. 2018; Assami and Hamdi-Aïssa 2019) and additionally have been successful in predictively mapping wetlands when using numerous predictor variables (Millard and Richardson 2014; Reschke and Hüttich 2014; Mahdianpari et al. 2017; Berhane et al. 2018; Jahncke et al. 2018; Goldman et al. 2020). No study has yet attempted to use terrain attributes as predictor variables in random forest models for the purpose of mapping wetland soil types in the PPR.

The objective of this study was to assess the suitability of PSM methodologies that incorporate terrain derivatives from high-resolution DEMs to predict the spatial distributions of wetland soil types in PPR landscapes. Soil types attempted for mapping include calcareous and noncalcareous wetland soil classes and a generalized upland soil class. This study was an attempt to quantitatively model the concepts established in the synthesis study by Pennock et al. (2014) through use of random forest models. This expands on the Kiss and Bedard-Haughn (2021) study on mapping wetland solute-richness classes by specifically mapping the distribution of soil types within wetlands. This study aimed to determine what existing terrain derivatives are most useful for predictively mapping wetland soil types and additionally test newly developed terrain derivatives created to better characterize the unique morphology of PPR landscapes. A secondary objective of this study was to assess the effectiveness of mapping wetland soil types using terrain attributes derived from DEMs of varying spatial resolution.

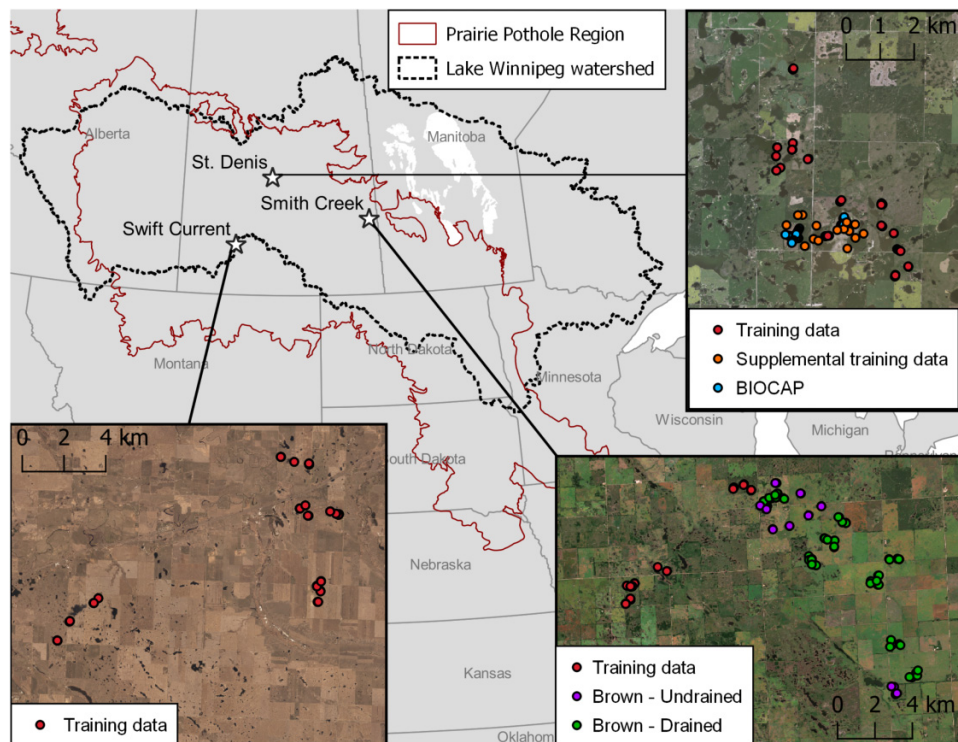
## Materials and methods

### Study areas

Soil profiles were described and sampled from three Saskatchewan study areas near Swift Current, St. Denis, and Smith Creek (Fig. 2). General characteristics for the study areas are summarized in Kiss and Bedard-Haughn (2021). The climate gradient spanned by the Swift Current, St. Denis, and Smith Creek study areas corresponds to the Brown, Dark Brown, and Black soil zones, respectively, which reflect general soil organic carbon storage trends influenced by the climatic gradient (Pennock et al. 2011). Similar soil catenas are found at the three sites, where Regosols and thinner Chernozems are commonly found in the eroded hillslope shoulder positions, Chernozems with thicker solums develop in mid-slope positions, and eluviated and gleyed Chernozems and Gleysols form in the depressional positions (Pennock et al. 2011). Soils in the Smith Creek area are commonly strongly enriched with  $\text{CaCO}_3$  due to the limestone-rich parent materials (Saskatchewan Soil Survey Staff 1991). LiDAR-derived DEMs were available for all three sites (meta-data for these are described in the supplemental material of Kiss and Bedard-Haughn (2021)). Data collection was conducted in fall 2015 and 2016 for the Swift Current study area; fall 2015 for the St. Denis study area; and fall 2016 for the Smith Creek study area. By fall, water levels within the wetlands were at their lowest. Data collection and predictive mapping were restricted to areas mapped in the soil survey (Saskatchewan Land Resource Unit 2009) as containing glacial till parent material as it is most representative of the PPR landscape.



**Fig. 2.** Study area locations within approximate extents of the Prairie Pothole Region (Ducks Unlimited 2005) and the Lake Winnipeg Watershed Basin (Government of Canada 2008). Training and testing sample point locations shown within each study area. Satellite RGB imagery for Swift Current from Agriculture and Agri-Food Canada (2009), satellite RGB imagery for St. Denis from Saskatchewan Geospatial Imagery Collaborative (2016), and satellite RGB imagery for Smith Creek from Copernicus (2017). [Colour online.]



## Sampling design

Wetland discharge rings represent a small portion of the overall landscape; randomly placed sampling points within a PPR landscape are likely to miss these features, so alternative sampling approaches were required. Wetland soil characteristics only develop within depressions, but not all depressions develop wetland soil characteristics. Individual wetlands could be interpreted from water and vegetative extents present in aerial imagery to inform the sampling design, but those can change drastically over short periods of time as a function of climate and management. Instead, the focus of this study was on the topographic characteristics that influence the development of wetland soil characteristics. By basing the sampling design on depressions rather than only on confirmed wetlands, the data could provide information on the differences between depressions that form wetland soil characteristics and those that do not.

Individual depression boundaries were based on closed topographic depressions determined from the DEM, which are groups of raster grid cells sharing internal drainage (Lindsay 2016). Depressions were selected for sampling using a stratified random sampling design. Depressions were stratified based on size (area and depth) and relative elevation in the watershed (upper and lower) (Table 1). Size classes were based on Millar's (1976) size classifications for wetlands in the Canadian PPR. The smallest depressions ( $\leq 4047 \text{ m}^2$ ) were further stratified based on depth of the closed topographic depres-

sion to capture very small depressions in the sampling design. Relative elevation was defined by two categories, upper and lower halves of the watershed, which were determined using Jenks natural breaks classification method (Jenks 1967).

Sampled depressions were located in sites with a range of land uses including cultivated, pasture, and native prairie. Some of the smaller depressions within the cultivated sites were tilled. The wetland soil characteristics of gleying and discharge rings are expected to persist regardless of land use. Tillage erosion can result in redistribution of  $\text{CaCO}_3$  to wetland positions, but this effect can be differentiated from natural, hydrologically influenced  $\text{CaCO}_3$  distributions with field inspections by determining if the  $\text{CaCO}_3$  distributions have abrupt horizontal boundaries that indicate they were redistributed through tillage. Wetland vegetation has been found to contribute to the formation of the discharge ring (Hayashi et al. 1998). However, the discharge rings developed over long periods of time (Knuteson et al. 1989), so the current vegetative characteristics do not necessarily reflect the long-term vegetative regimes for a depression. Wetlands that had undergone obvious mechanical drainage were avoided because the soils can be substantially disturbed through such processes. There has been extensive drainage of wetlands throughout the lower portion of the Smith Creek watershed and so sampling was done only in the upper half of that watershed (Table 1).

**Table 1.** Criteria for depression size classes and number of depressions sampled per study area per relative position in watershed and size stratifications.

Relative elevation in watershed	Size class	Size class criteria		Study area (n)		
		Area* (m <sup>2</sup> )	Depth (m)	Swift Current	St. Denis	Smith Creek
Upper	1	0–4047	0.1–0.3	3	3	3
	2	0–4047	>0.3	2	2	2
	3	4047–40 470	>0.1	2	2	3
	4	>40 470	>0.1	2	1	3
Lower	1	0–4047	0.1–0.3	3	3	—
	2	0–4047	>0.3	2	2	—
	3	4047–40 470	>0.1	2	2	—
	4	>40 470	>0.1	2	2	—

\*Area criterion were based on Millar's (1976) wetland size classes described in acres.

Each depression was sampled with a single transect with points spaced according to specified changes in elevation (dependent on the depth of the depression) with an average of five points per depression. Refer to Appendix A for a detailed description of the transect sampling design. Transect points were sampled with a truck-mounted hydraulic corer (Giddings Machine Company Ltd., Windsor, CO) where possible. Otherwise, transect points were sampled by hand auger. Sample point locations were collected with a Trimble Geo-Explorer 2005 Series GeoXT GPS (Trimble, California, USA). At each point, sample cores were taken to at least 50 cm to facilitate classification and to 90 cm when possible. The profiles at each point were described and classified according to the Canadian System of Soil Classification (CSSC) (Soil Classification Working Group 1998) and were additionally classified as either calcareous wetland, noncalcareous wetland, upland, or buried/depositional soil types according to the criteria defined in Table 2, as the CSSC does not include classifications that reflect these soil types. Appendix A includes a detailed description and justification of the criteria that define the classifications used in this study.

## Model training datasets

The data collection took place over a period of relatively high precipitation for these areas (Brown et al. 2017a), which meant that the larger, more permanent ponds were often inundated with water. This limited sampling access to the wetland basin centers. An additional dataset from the St. Denis study area was used as supplementary training data to better reflect the soil class distributions in these landscape positions. This dataset was collected between 2000 and 2002 during a period of relative drought when many of the larger wetlands had dried out. A description of this dataset and the criteria used to classify its soil point observations according to the soil types of this study is provided in Appendix A.

Table 3 shows the number of observations per wetland soil type classification from each study area. These observations were used to train the predictive models. Buried/depositional soils were not considered in the models because it would not be possible to accurately predict their distribution based on the current elevation surface (Bedard-Haughn and Pennock 2002). There were fewer calcareous wetland soil type obser-

vations as compared to noncalcareous wetland soil type observations, which is representative as they are less abundant within a typical PPR landscape.

Models were trained using two training data approaches—a generalized approach and a site-specific approach. For the generalized approach, training data from all three sites were used to train a predictive model and the same model was applied to predictively map the soil types at the three sites. For the site-specific approach, models were trained only using the data from the site to be mapped.

## Independent validation datasets

Two datasets of soil point observations from previous studies were used as independent validation datasets to test model predictive accuracy. Independent validation soil point observations were sampled from wetland depressions that were not sampled as part of the model training datasets of this study. These included data from the BIOCAP study at St. Denis (Pennock et al. 2014) and the Brown et al. (2017a) study at Smith Creek. The BIOCAP dataset consisted of three wetlands sampled with two transects, each with 10–12 points per transect. The three sampled wetlands consisted of two recharge wetlands and a discharge wetland. This sample set was considered a rigorous test for the predictive models due to the high density of samples per wetland and the inclusion of a fully discharge wetland, which contained calcareous wetland soils throughout its basin.

The Brown dataset consisted of samples from 42 wetlands; however, 32 of those wetlands had been mechanically drained at some point. The data from the undrained and drained wetlands were separated into two validation sets. Each of the 42 wetlands was sampled with two points, one at the wetland toe-slope and one in a mid-slope position just above the discharge ring. The Undrained dataset did not include any calcareous wetland soil observations and the Drained dataset included only four. These datasets effectively tested the boundary surrounding the discharge ring. The number of observations per independent validation dataset are shown in Table 4. The point observations of the independent validation datasets included CSSC classifications (Soil Classification Working Group 1998) and profile descriptions that enabled their classifications into the target soil types

**Table 2.** Soil type classification criteria.

Soil type	CaCO <sub>3</sub> distribution	CSSC classification	Other criteria
Noncalcareous wetland	Weak or no CaCO <sub>3</sub> observed in top 15 cm of profile	<ul style="list-style-type: none"> <li>• Gleysols OR</li> <li>• Gleyed soils</li> </ul>	—
Calcareous wetland	Moderate to strong CaCO <sub>3</sub> throughout the profile	<ul style="list-style-type: none"> <li>• Gleysols OR</li> <li>• Gleyed soils OR</li> <li>• Rego Chernozems that do not occur in shoulder or summit positions of a hillslope. Adjacent upland profiles along transect must either have a B horizon or have weak or no CaCO<sub>3</sub> in their 0–15 cm increment OR</li> <li>• Calcareous Chernozems that do not occur in shoulder or summit positions of a hillslope. The B horizon must have a chroma of less than 5 and a chroma 1 less than the B horizon of the adjacent upland profile along the transect</li> </ul>	<ul style="list-style-type: none"> <li>• Form in a continuous sequence from the wetland edge, that is, once an upland soil is encountered, no calcareous wetland soils should be found upslope along the transect</li> </ul>
Buried/depositional	Any distribution	<ul style="list-style-type: none"> <li>• Any</li> </ul>	<ul style="list-style-type: none"> <li>• Soil profiles with more than 30 cm of depositional material on top of them</li> </ul>
Upland	Any distribution	<ul style="list-style-type: none"> <li>• Any</li> </ul>	<ul style="list-style-type: none"> <li>• Soils that do not meet criteria of other soil types</li> </ul>

CSSC, Canadian System of Soil Classification.

**Table 3.** Number of observations per soil type in the training dataset.

Soil type	Study area (n)			Total
	Swift Current	St. Denis	Smith Creek	
Noncalcareous wetland	24	39	27	90
Calcareous wetland	6	30	13	49
Upland	37	47	27	111
Total	67	116	67	250

**Table 4.** Number of soil type observations per independent validation set.

Soil type	Independent validation sets (n)		
	BIOCAP–St. Denis	Brown–Smith Creek	
		Undrained	Drained
Noncalcareous wetland	20	10	28
Calcareous wetland	28	0	4
Upland	13	10	31
Total	61	20	63

of this study, except there were six observations that were left unclassified as the data required to distinguish them as either calcareous wetland or upland soil types were not available.

### Spatial resolution

The PPR landscape is characterized by topographic variation at scales where hilltops and depressions can occur within

5–10 m of each other and very different soils develop in these positions (Pennock et al. 1987). These features would be smoothed out and lost in low-resolution DEMs and so the DEMs were kept at resolutions that correspond to the scale of variation for terrain derivative calculation. Separate models were tested using terrain derivatives generated at 2, 5, and 10 m to assess differences in predictive accuracy at varying spatial resolutions. Mapping was tested at 5 m spatial resolution to stay consistent with Bedard-Haughn and Pennock’s (2002) PSM study on the distribution of PPR wetland soil types. Due to the specific sampling design used in our study, at a resolution of 5 m, many adjacent sampled point pairs shared the same pixel. Based on recommendations in Hengl (2006), a spatial resolution of 2 m was also tested because at this resolution, less than 5% of closest point pairs shared pixels. The 2 m DEMs provide a more detailed representation of the land surface and may capture features that would be smoothed out at 5 m resolution. Ten metre DEMs were also tested to assess if similar predictive performance could be achieved at coarser spatial resolution, which is typically easier to obtain and reduces computational workloads.

### DEM processing

LiDAR collection for the three sites occurred either late summer or fall when water levels are lowest; however, at St. Denis and Smith Creek sites, standing water was still present in the largest wetlands. This caused the DEM surface in these positions to reflect the water surface and not the underlying sediment. The water surface causes a noisy DEM in these positions. A method for identifying and reinterpolating water surfaces in the DEM is described in Appendix A, as are the methods used to smooth and resample the DEM to the target spatial resolutions following reinterpolation.



## Predictor variables

As the focus of this study was to determine the effectiveness of using terrain derivatives derived from high-resolution DEMs as predictor variables for predictively mapping wetland soil types and to determine which terrain derivatives were most useful, predictor variables were limited primarily to terrain derivatives. It is likely that the predictive mapping of wetland soil types would be improved by including predictor variables derived from other remote and proximal sensing data, like synthetic aperture radar (SAR), optical remote sensing data, band indices, electromagnetic surveys, etc. but they were not included to enable more straightforward evaluation of terrain derivatives as predictor variables.

A total of 35 predictor variables were calculated at each DEM resolution (2, 5, and 10 m) (Table 5). Most of these were calculated using the System for Automated Geoscientific Analysis (SAGA) (Conrad et al. 2015) or Whitebox Geospatial tools (Lindsay 2014). Measures of Euclidean distance were included based on Behrens et al. (2018) and were calculated using R code published by Saurette (2021). Cartographic depth-to-water (Murphy et al. 2009) was calculated with GRASS GIS (GRASS Development Team 2017) using R code published by Schönauer and Maack (2021).

Several terrain derivatives were specifically developed to better represent the unique morphological characteristics of the PPR that were expected to relate to wetland soil type distributions (Fig. 3). Three of these derivatives were based on a grid cell's position relative to a local depression. These were developed to capture relationships identified by Pennock et al. (2014) between soil type distribution and topographic position relative to depression bottoms and maximum water levels.

The contributing watersheds were determined for each closed topographic depression within the DEM. Grid cells were related to their local depression based on the contributing basin they fell within. Values were calculated for a grid cell's elevation above the associated depression's basin minimum (elevation from depression minimum derivative) and elevation above or below the depression's spillover elevation (elevation from depression spillover derivative). The spillover elevation reflects the maximum elevation of a closed depression before water would spill externally from it. For the depression depth derivative, the depression depth value was ascribed to all pixels within the depression's contributing basin. The measure of elevation from depression minimum per depth was included to potentially capture differences in relationships of soil distribution for wetlands of varying sizes. The watershed max catchment area derivative ascribes the maximum catchment area for a depression's contributing area to all pixels within that contributing area.

Wetland Strahler order minimum and maximum were calculated for each depression following the methodologies described in Kiss and Bedard-Haughn (2021). For these variables, the characteristics were quantified for each closed depression; all pixels within the depression were ascribed the same value and all pixels not included in closed topographic depressions were given values of zero. R scripts for calculating the terrain derivatives developed in this study and parameters used to calculate existing derivatives are available on

GitHub (Kiss 2022). Both the newly developed terrain derivatives and existing derivatives were calculated using a suite of Whitebox Geospatial (Lindsay 2014) and SAGA GIS (Conrad et al. 2015) tools.

Prior to model training, recursive feature elimination was conducted to reduce the number of predictor variables used in the final models to reduce the risk of model overfitting. Feature reduction has been found to improve model performance in PSM studies focused on soil class mapping (Brungard et al. 2015; Hounkpatin et al. 2018). The recursive feature elimination was implemented using the caret package (Kuhn 2008) in R (R Core Team 2020), using a random forest model as the classifier and 5-fold cross validation with five iterations using Kappa score as the evaluation metric to determine the optimal predictor variables to include in the final models. Importance of each predictor variable in the final models was determined using the Gini index (Wright and Ziegler 2017). Overall variable importance was determined by summing the Gini index of each variable within the final random forest models.

## Model training, tuning, and validation

The objective of this study was to assess the performance of various terrain derivatives as predictor variables for mapping wetland soil types and not to determine the optimal machine learning model type for this problem. Random forest is a tree-based machine learning model capable of modelling complex nonlinear relationships between numerous predictor variables and categories (Strobl et al. 2009). In addition to its frequently cited success in the PSM literature (Brungard et al. 2015; Heung et al. 2016; Jeune et al. 2018; Assami and Hamdi-Aissa 2019), random forest models provide an intuitive evaluation of predictor variable importance, which makes it an ideal candidate for this study.

Following the recursive feature elimination step, the caret package (Kuhn 2008) was used again to optimize model parameters and train the final random forest models through K-folds cross-validation. Each model was trained with the training dataset using 5-fold cross-validation with 20 iterations using Kappa score as the evaluation metric. From this, cross-validation accuracy and Kappa scores were determined as well as the optimal mtry parameter setting per model, which is the number of predictor variables available for each split in a random forest model. Each random forest was grown with 1000 trees (ntree).

Model performance was assessed based on overall accuracy, Kappa score, producer's accuracy, and user's accuracy as determined from the cross-validation and independent validation tests. Cross-validation performance was assessed in addition to independent validation tests because independent validation data were not available for the Swift Current study area, so cross-validation performance allowed for a fairer assessment of model performance across all three sites. Overall accuracy refers to the total number of correct classification predictions per the total number of observations. Kappa score corrects the overall accuracy by accounting for the possibility of a correct prediction based on chance alone (Landis and Koch 1977). Producer's accuracy, which is reported per



**Table 5.** List of predictor variables and their source.

Terrain derivative	Reference
Euclidean distance from site corners	Behrens et al. (2018)
SAGA wetness index	Boehner et al. (2002)
Slope height	Boehner and Selige (2006)
Normalized height ( $t = 10$ , $t = 1000$ )	Boehner and Selige (2006)
Valley depth	Boehner and Selige (2006)
Mid-slope position	Boehner and Selige (2006)
Specific dispersal area	Costa-Cabral and Burges (1994)
Multi-resolution ridge top flatness index	Gallant and Dowling (2003)
Multi-resolution valley bottom flatness index	Gallant and Dowling (2003)
Wetland Strahler order minimum	Kiss and Bedard-Haughn (2021)
Wetland Strahler order maximum	Kiss and Bedard-Haughn (2021)
Max elevation deviation (local, meso, and broad)	Lindsay et al. (2015)
Stochastic depression analysis	Lindsay and Creed (2005)
Slope length and steepness factor	Moore et al. (1991)
Cartographic depth-to-water	Murphy et al. (2009)
Catchment area (multiple flow direction)	Quinn et al. (1991)
Specific catchment area	Quinn et al. (1991)
Terrain ruggedness index	Riley et al. (1999)
Aspect (northness)	Zevenbergen and Thorne (1987)
Slope	Zevenbergen and Thorne (1987)
Plan curvature	Zevenbergen and Thorne (1987)
Profile curvature	Zevenbergen and Thorne (1987)
Tangential curvature	Zevenbergen and Thorne (1987)
Elevation	—
Elevation from depression spillover	Kiss (2022)
Elevation from depression minimum	Kiss (2022)
Elevation from depression minimum per depth	Kiss (2022)
Depression depth	Kiss (2022)
Watershed max catchment area	Kiss (2022)

Note: SAGA, System for Automated Geoscientific Analysis.

soil type class (Table 8), refers to the number of correctly predicted soil classifications of a particular soil type per the total number of observations of that soil type (Malone et al. 2017). User's accuracy, also reported per soil type, refers to the number of correctly predicted soil classifications of a soil type per the total number of predictions of that soil type.

### Visual assessment of model performance

The predictive accuracies based on the internal and independent validations provide an excellent performance summary of the predictive models. However, it is crucial to ensure that in addition to achieving acceptable accuracy scores, model predictions of soil distributions make conceptual sense. Maps were generated for each site from each of the models and were visually inspected using expert knowledge to determine if the distribution of soils matched the concepts described in the introduction. This involved inspecting areas at each site where wetland types were known either from previous studies or from the sampling conducted in this study.

For most wetlands, inspection was conducted to determine if noncalcareous wetland soil types were predicted to be distributed throughout the wetland basin floor with calcareous wetland soils formed in a discharge ring pattern in the foot

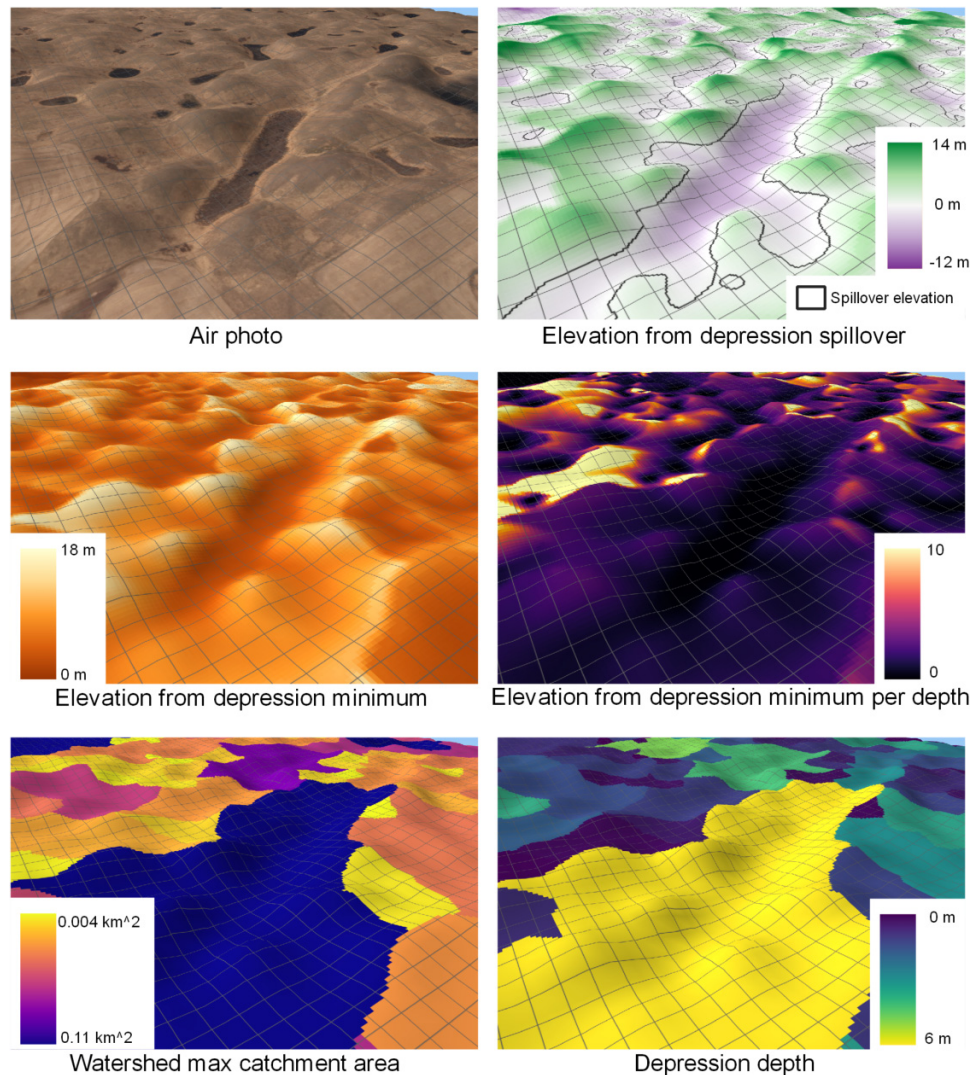
to mid slope positions surrounding the wetland basin. At the St. Denis study area, there were numerous wetlands known to have calcareous wetland soils distributed not only in discharge rings surrounding the basin, but throughout their wetland basins; these were inspected to determine if calcareous wetland soils were predictively mapped throughout these wetland basins. Distributions of upland soils mapped within known wetlands or wetland soils mapped in locations well beyond the boundaries of discernable wetlands were deemed erroneous. Soil types are expected to be distributed continuously and therefore single pixel or small clusters of soil types were typically considered erroneous artefacts, except for the models generated based on predictor variables of 10 m resolution, as wetland depressions may be small enough to be represented by a single 10 m pixel.

## Results

### Tuned model parameters and internal cross-validation results

Table 6 shows the optimized model parameters and cross-validation accuracy per model. Based on the cross-validation

**Fig. 3.** 3D scene for an example area in the Swift Current study site (3× vertical exaggeration based on DEM) showing an air photo and the terrain derivatives developed for this study. [Colour online.]



results, the models developed on predictor variables of 2 m resolution had the highest predictive accuracies, but cross-validation accuracy was generally similar between the models based on different resolutions (ranging from 75% to 82% accurate) (Table 6). According to the cross-validation results, the site-specific models performed slightly better than the generalized models; the average Kappa scores across all three sites (All (mean)) for the site-specific models were slightly higher than the generalized models at all resolutions (Table 6). Overall, there were only minor differences in performance across all models shown in Table 6 based on the cross-validation evaluation. The models vary widely in the total number of predictor variables used (from 7 to all 35). There were no noticeable trends in the difference of performance of models that incorporated many or few predictor variables.

### Independent validation results

The models were tested for predicting the soil types of the independent datasets from the St. Denis and Smith Creek

study areas. Table 7 shows the overall accuracy and Kappa scores based on the independent validation tests for the models. Table 8 shows the producer's and user's accuracy for each soil type classification for the BIOCAP–St. Denis and Brown–Smith Creek–Undrained independent validation tests for the models.

### Independent validation results—BIOCAP–St. Denis

The best-performing model for the BIOCAP–St. Denis dataset was the 2 m site-specific model, which was built using predictor variables with a spatial resolution of 2 m and used only training data from the St. Denis study site. This model achieved an overall accuracy of 76% and a Kappa score of 0.63. The site-specific models consistently performed better than the generalized models for the St. Denis independent validation test.

The models consistently had low producer's accuracy and high user's accuracy for calcareous wetland soil types. This

**Table 6.** Model parameters, cross-validation overall accuracy, and Kappa score per model per site.

Model		Site	Optimized parameters		Cross-validation	
Resolution	Training data		<i>n</i> pred var*	mtry	Acc.† (%)	Kappa
2 m	Generalized	All	35	5	80	0.67
	Site-specific	All (mean)‡	—	—	80	0.68
		Smith Creek	12	2	80	0.68
		St. Denis	33	3	82	0.72
		Swift Current	24	11	79	0.62
5 m	Generalized	All	32	5	78	0.64
	Site-specific	All (mean)	—	—	79	0.66
		Smith Creek	8	2	79	0.67
		St. Denis	30	6	79	0.68
		Swift Current	35	10	78	0.61
10 m	Generalized	All	34	4	75	0.59
	Site-specific	All (mean)	—	—	78	0.64
		Smith Creek	8	4	75	0.61
		St. Denis	32	2	79	0.67
		Swift Current	7	3	80	0.64

\*Number of predictor variables included in the model training.

†Overall accuracy.

‡While "All" for the generalized models includes cross-validation accuracy and Kappa scores for models trained with training data from all sites, "All (mean)" provides mean summaries of the cross-validation accuracy and Kappa scores for the three site-specific models.

**Table 7.** Overall accuracy and Kappa score per model per independent validation test sets.

Model		Independent validation					
		BIOCAP–St. Denis		Brown–Smith Creek			
				Undrained		Drained	
Resolution	Training data	Acc.* (%)	Kappa	Acc. (%)	Kappa	Acc. (%)	Kappa
2 m	Generalized	64	0.48	90	0.80	85	0.70
2 m	Site-specific	76	0.63	100	1.00	80	0.66
5 m	Generalized	63	0.45	100	1.00	93	0.85
5 m	Site-specific	69	0.54	90	0.80	75	0.55
10 m	Generalized	64	0.45	85	0.70	85	0.70
10 m	Site-specific	76	0.62	84	0.67	66	0.40

\*Overall accuracy.

communicates that this soil type is underestimated but where it is predicted, the map user can have confidence they will find that soil type there. The 2 and 10 m site-specific models achieved the most balanced results for the calcareous wetland soil types with producer’s accuracies of 54% and 65% and user’s accuracy of 94% and 90%, respectively. Conversely, the models had very high producer’s accuracy and low user’s accuracy for noncalcareous wetland soil types. This means that all noncalcareous wetland observations were correctly predicted as such, but that they were overpredicted elsewhere.

### Independent validation results—Brown–Smith Creek

The models generally had high accuracies and Kappa scores for predicting the soil types of the Undrained independent validation set; some achieving perfect (100% accurate, Kappa score = 1.0) and near perfect (90% accurate, Kappa

score = 0.80) results (Table 7). The 10 m resolution models had the lowest accuracies for this validation test (ranging from 84% to 85% accurate). Site-specific and generalized models performed relatively similar for the Undrained dataset, but for the Drained test set, the generalized models consistently performed better, with some Kappa scores 0.30 greater. Predictions for the Drained validation set were generally less successful than for the Undrained validation set but still typically had high levels of agreement. This indicates that despite the impact drainage may have on the landscape (and the derived DEM), the distribution of wetland soils may still be predictable in landscapes that have undergone mechanical drainage, meaning these methods could potentially be used in restoration projects in drained landscapes to identify historic wetland extents.

The Brown–Smith Creek–Undrained validation dataset did not contain any calcareous wetland soil type observations



**Table 8.** Producer and user accuracy per soil type classification per model per independent validation test set (Brown–Smith Creek–Drained not included).

Model		Independent validation (%)									
		BIOCAP–St. Denis						Brown–Smith Creek			
		Calc. wetland*		Noncalc. wetland†		Upland		Undrained		Upland	
		Prod.‡	User§	Prod.	User	Prod.	User	Prod.	User	Prod.	User
Resolution	Training data										
2 m	Generalized	25	100	100	53	93	75	80	100	100	84
2 m	Site-specific	54	94	95	66	93	75	100	100	100	100
5 m	Generalized	22	100	100	50	93	80	100	100	100	100
5 m	Site-specific	40	100	100	55	85	85	100	84	80	100
10 m	Generalized	40	79	100	55	62	80	80	89	90	82
10 m	Site-specific	65	90	100	65	62	80	100	73	70	100

\*Calcareous wetland soil type.

†Noncalcareous wetland soil type.

‡Producer's accuracy.

§User's accuracy.

due to the sampling design. Therefore, models that underpredicted the extent of calcareous wetland soil types were likely to perform well for this dataset. The 2 and 5 m resolution-based models performed similarly, sometimes overpredicting noncalcareous wetland soil types and sometimes overpredicting upland soil types (as indicated by high producer's accuracy and lower user's accuracy (Table 8)).

### Predicted soil type maps—visual assessments

Predicted soil maps for the full study areas from each model for St. Denis and Smith Creek are shown in Figs. 6 and 8. The full map of the Swift Current study area is not included as differences in soil distributions between the models are not obvious at coarse scale. Full-sized maps for each study area from each model are provided in the Supplementary file. Figures 4, 5, and 7 show predicted maps for example areas from each site.

Generally, the predicted maps generated by the models were acceptable; wetland soils (calcareous and noncalcareous soil types) were typically mapped within or directly adjacent to depressional positions and upland soils were mapped in upslope positions. Wetland basin centers were often mapped with noncalcareous wetland soils with rings of calcareous wetland soils mapped surrounding the wetlands reflecting the discharge ring phenomenon. Wetland soils were most frequent at the Smith Creek study area and least frequent at the Swift Current study area, which coincides with the general climatic moisture trends for these sites. The St. Denis study area had the most wetlands predicted to have calcareous wetland soils throughout their basins, which is consistent with historic site observations (Pennock et al. 2014). Calcareous wetland soil types were much less frequently predicted within the Swift Current study area, which is consistent with the small amount of calcareous wetland soils encountered in the field data collection (Table 3).

At the Swift Current study area, with exception to the 10 m based models, all models generated relatively consistent maps (Fig. 4). The largest wetlands were mapped to have calcareous wetland soils surrounding the wetland basin, but

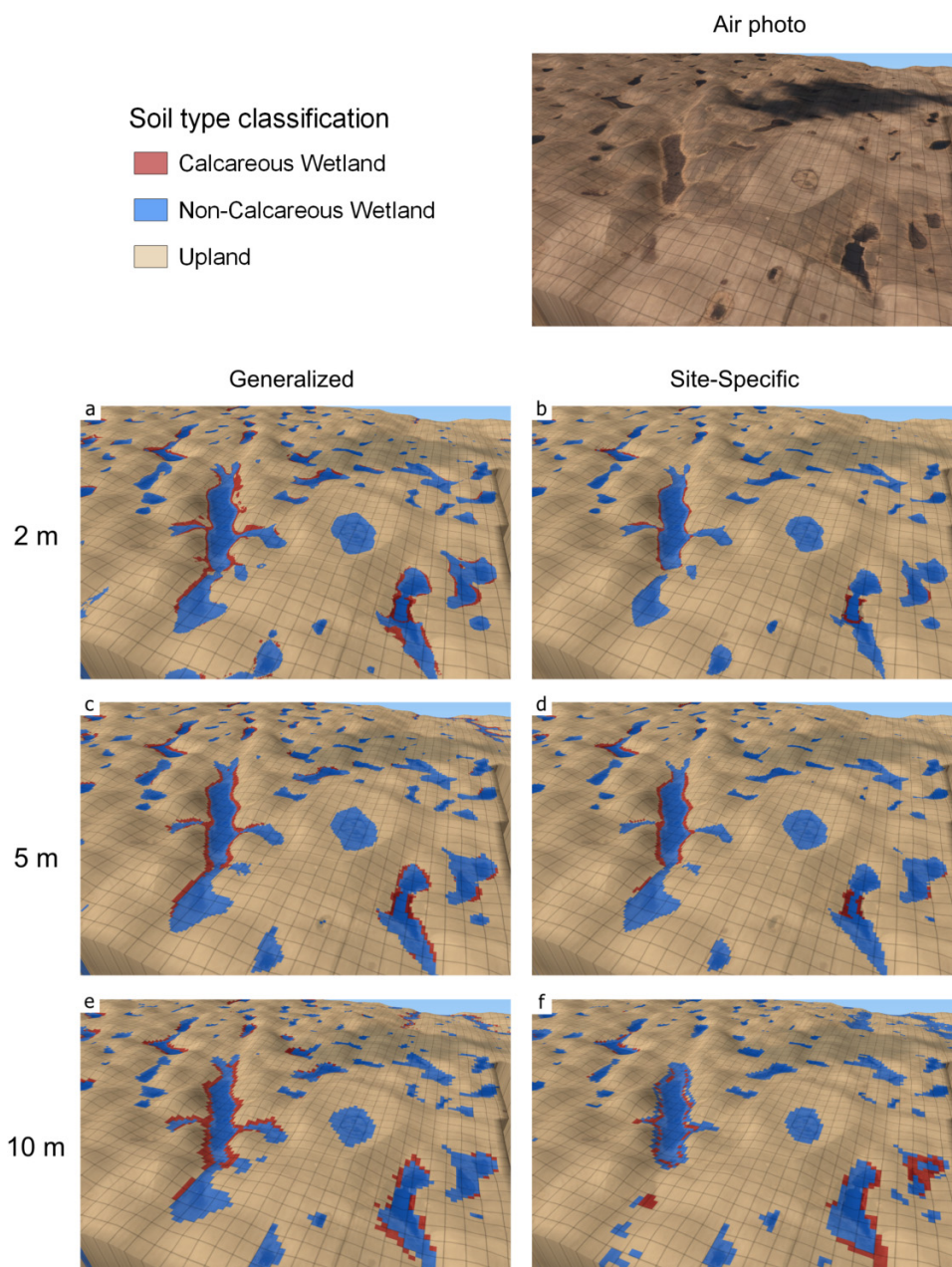
most other wetlands were mapped with only noncalcareous wetland soil types. The generalized models predicted more occurrences of calcareous wetland soil types than the site-specific models. The maps generated by the 10 m based models typically had more single-cell and small cluster distributions of soil types compared to those of the finer-scale resolutions. These scattered soil type predictions were often wetland soil types erroneously mapped in upland positions.

The predicted maps generated for the St. Denis study area were relatively consistent across all models, except that the site-specific models predicted more frequent occurrences of wetland soils generally and more frequent occurrences of calcareous wetland soils (Fig. 6). This contributed to its better performance when predicting for the BIOCAP validation test set. When mapping at 10 m, the rings of calcareous wetland soils surrounding the wetlands become blocky and less detailed as a result of the pixel size. As discussed, larger depressions were more frequently predicted to have calcareous wetland soils within their basin floors at the St. Denis study area. All models predicted Pond 90 (the largest wetland in the top right of the example area scene (Fig. 5)) to be predominantly covered with calcareous wetland soils, which is appropriate as the largest low-lying wetlands in this portion of the St. Denis watershed are known to have calcareous wetland soils throughout their basins (Pennock et al. 2014). However, resulting maps often include irregular distributions of noncalcareous soils within these basins (as seen in ponds adjacent to Pond 90 in Fig. 5), which is not expected to occur. Additionally, not all wetlands at St. Denis known to have calcareous wetland soils throughout their basins were consistently predicted as such.

Compared to the other sites, the resulting maps for the Smith Creek study area differed substantially between some models, especially in the eastern portion of the study area where training data were lacking (Fig. 8). The site-specific models predicted greater occurrences of calcareous wetland soil types and wetland soils generally (Figs. 7 and 8). Noncalcareous wetland soils appear to extend beyond the wetland boundaries observable in the air photo in the maps generated



**Fig. 4.** 3D scene ( $3\times$  vertical exaggeration based on DEM) of predicted soil type maps and air photo for an example area at the Swift Current study site based on the generalized and site-specific models built at each spatial resolution. Satellite RGB imagery from *Agriculture and Agri-Food Canada (2009)*. [Colour online.]



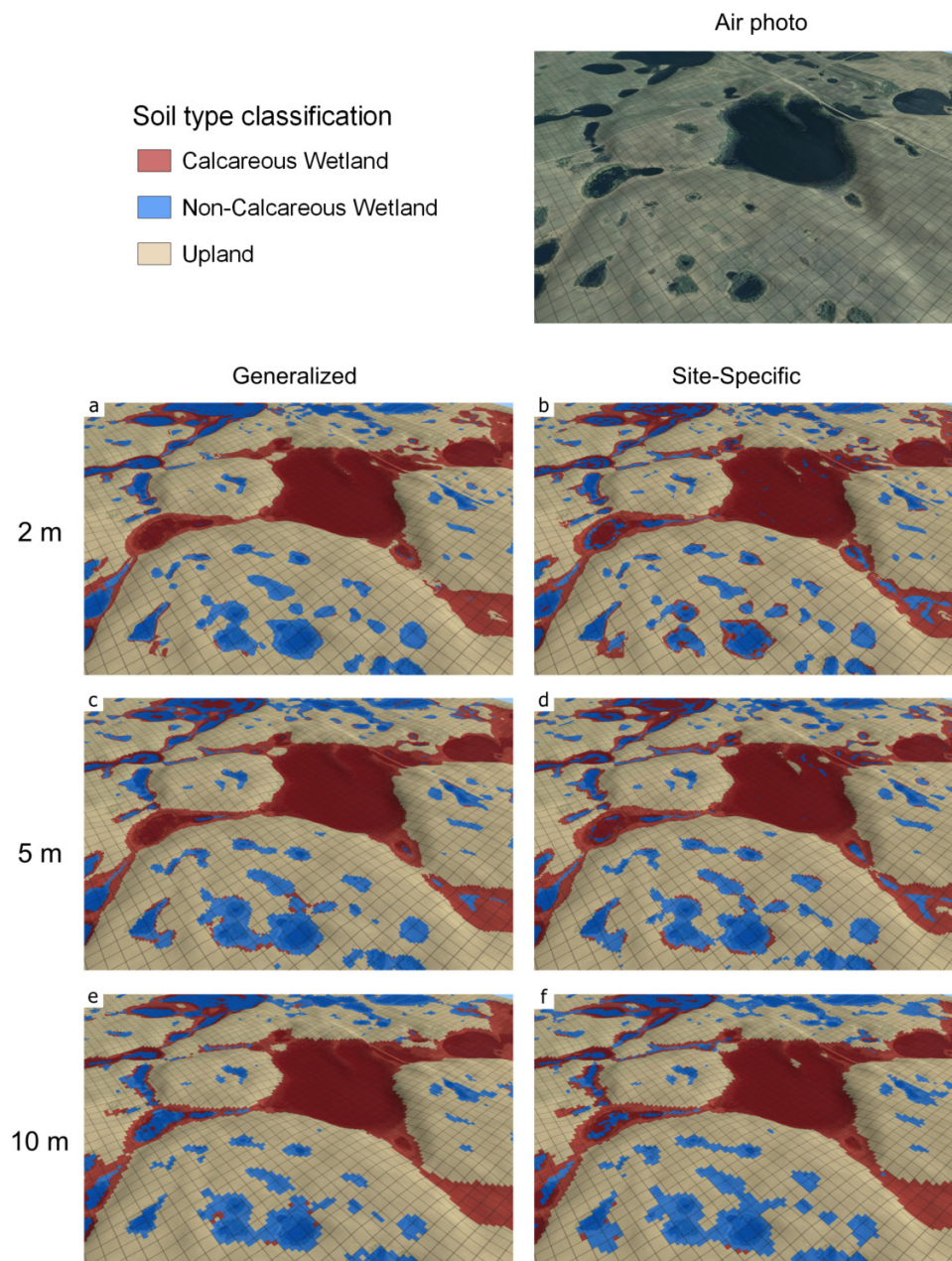
by the site-specific models (Fig. 7) and may be overestimated. It is possible that this is appropriate in an area as dense with wetlands as shown in the example area in Fig. 7; however, in the eastern portion of the study area where wetland drainage is more common, it is likely that wetland soils are being overestimated. This explains the poorer performance of the site-specific models for predicting the classes for the Drained validation test set.

The general high accuracies of each model based on the independent validation tests despite the apparent differences in resulting maps is likely due to the nature of the validation test sets from the area. Point observations in these datasets

were collected from distinctly different slope positions and were likely easier to predict than if the points spanned the gradient between these positions. The questionable results generated by the site-specific models for the eastern portion of the study area became apparent only through the visual assessment, as they achieved acceptable predictive accuracies based on the validation datasets.

The map generated by the 10 m based generalized model for Smith Creek contains many instances of single pixel and small clusters of the soil types which results in a speckled pixelated-looking output as seen in Fig. 7e. The map generated by the 10 m based site-specific model follows generally

**Fig. 5.** 3D scene ( $3\times$  vertical exaggeration based on DEM) of predicted soil type maps and air photo for an example area at the St. Denis study site based on the generalized and site-specific models built at each spatial resolution. Satellite RGB imagery from *Saskatchewan Geospatial Imagery Collaborative (2010)*. [Colour online.]



the same trends mapped by the 2 m site-specific model but because of its coarseness, the mapped soil type distributions do not follow the contours of the wetlands and depressions with the same level of precision (Figs. 7b and 7f).

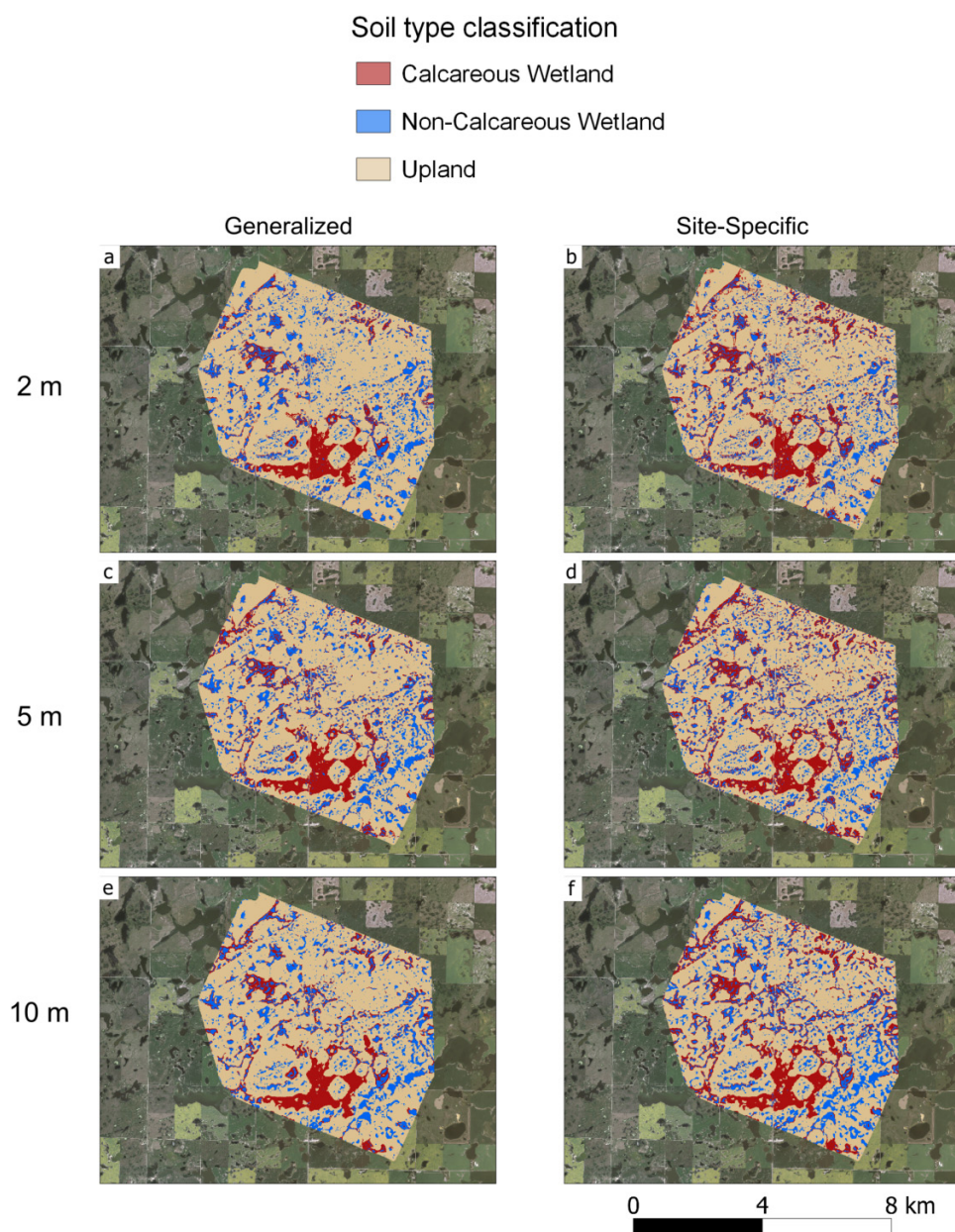
### Variable importance

Figure 9 shows a plot of the overall importance of every predictor variable within the random forest models. Of the top 10 most important variables, three were developed specifically for this study (elevation from depression minimum per depth, elevation from depression spillover, and elevation from depression minimum) and two were devel-

oped for the related study focused on predicting wetland solute-richness classes in PPR landscapes (Wetland Strahler order minimum and maximum) (Kiss and Bedard-Haughn 2021). Elevation from depression minimum per depth and elevation from depression spillover were the most important predictor variables overall. Four of the top 10 predictor variables were calculated using Whitebox Geospatial tools, including stochastic depression analysis (Lindsay and Creed 2005) and maximum elevation deviation (local, meso, and broad scales) (Lindsay et al. 2015). SAGA-calculated terrain derivatives slope height and normalized height ( $t = 1000$ ) (Böhner and Selige 2006) were the seventh and 11th most



**Fig. 6.** Predicted soil type maps for the full St. Denis study site based on the generalized and site-specific models built at each spatial resolution. Full-sized versions of each map are provided in supp1a. Satellite RGB imagery from [Saskatchewan Geospatial Imagery Collaborative \(2016\)](#). [Colour online.]



important predictor variables, respectively. Euclidean distance from NE site corner, Watershed max catchment area, and Aspect (northness) had the lowest Gini index scores, indicating that they were the least important predictor variables.

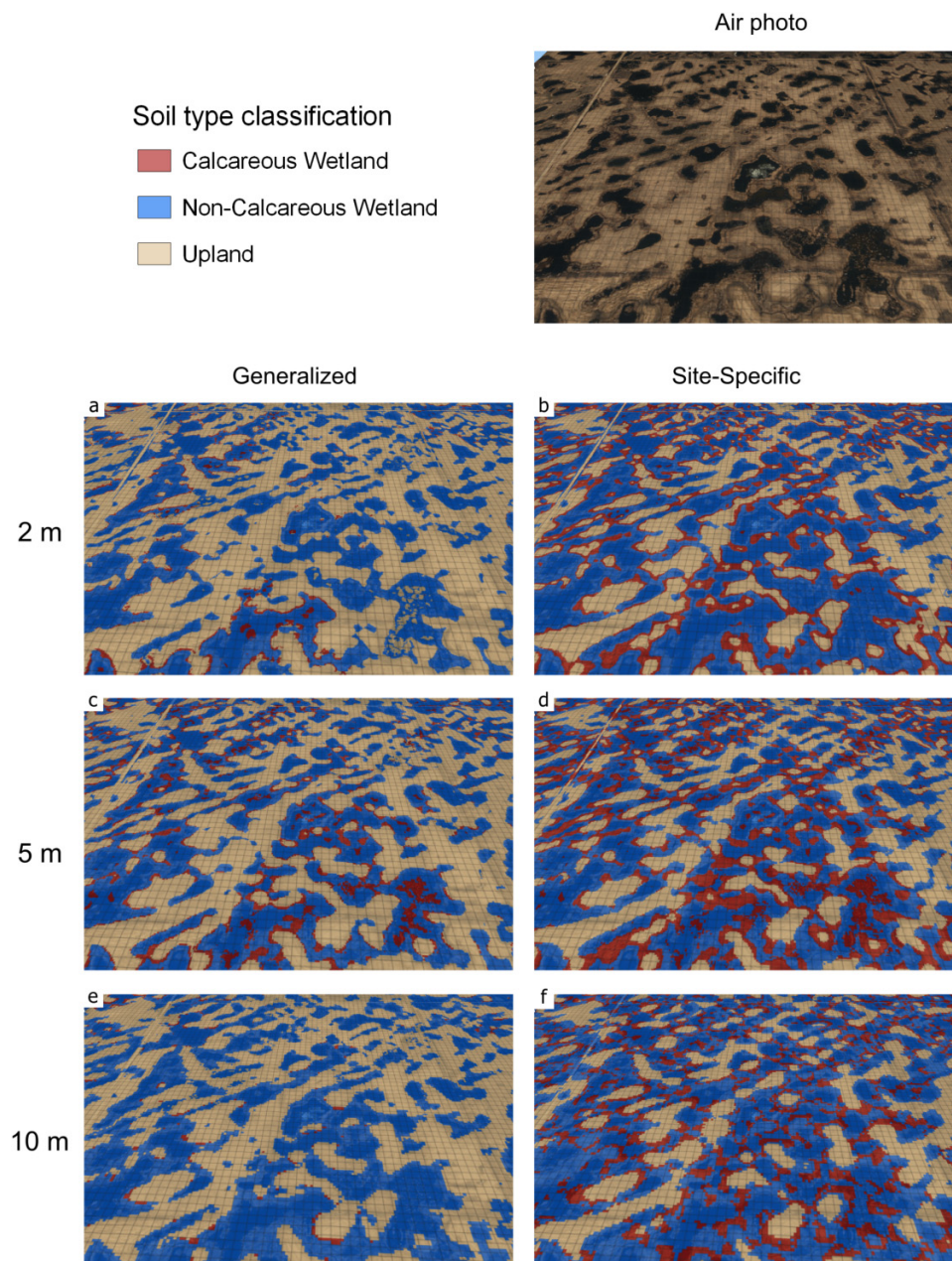
## Discussion

### Overall model performance

The relatively high overall accuracies and Kappa scores according to the cross-validation and independent validation tests, and additionally the acceptable map outputs as determined through visual assessment for many of the mod-

els indicate their success in predictively mapping wetland soil types at the PPR study sites. Based on all model evaluation criteria, the best-performing model overall was the 2 m site-specific model, which achieved overall accuracies and Kappa scores ranging from 80%–82% and 0.62–0.72, respectively, based on cross-validation (Table 6) and 76%–100% and 0.63–1.0, respectively, based on independent validation tests (Table 7). The related study focused on predictively mapping wetland solute-richness classes at the same PPR study sites achieved similar accuracies: independent validation accuracy of 69%–80% and Kappa scores of 0.40–0.41 (Kiss and Bedard-Haughn 2021). The model developed for that study was simpler and based on only a few terrain-derived predictor variables and its objective was only to predict a single

**Fig. 7.** 3D scene ( $3\times$  vertical exaggeration based on DEM) of predicted soil type maps and air photo for an example area at the Smith Creek study site based on the generalized and site-specific models built at each spatial resolution. Satellite RGB imagery from *Saskatchewan Geospatial Imagery Collaborative (2014)*. [Colour online.]



solute-richness class for each wetland, whereas in this study, detailed mapping of soil types within and surrounding individual wetlands was accomplished.

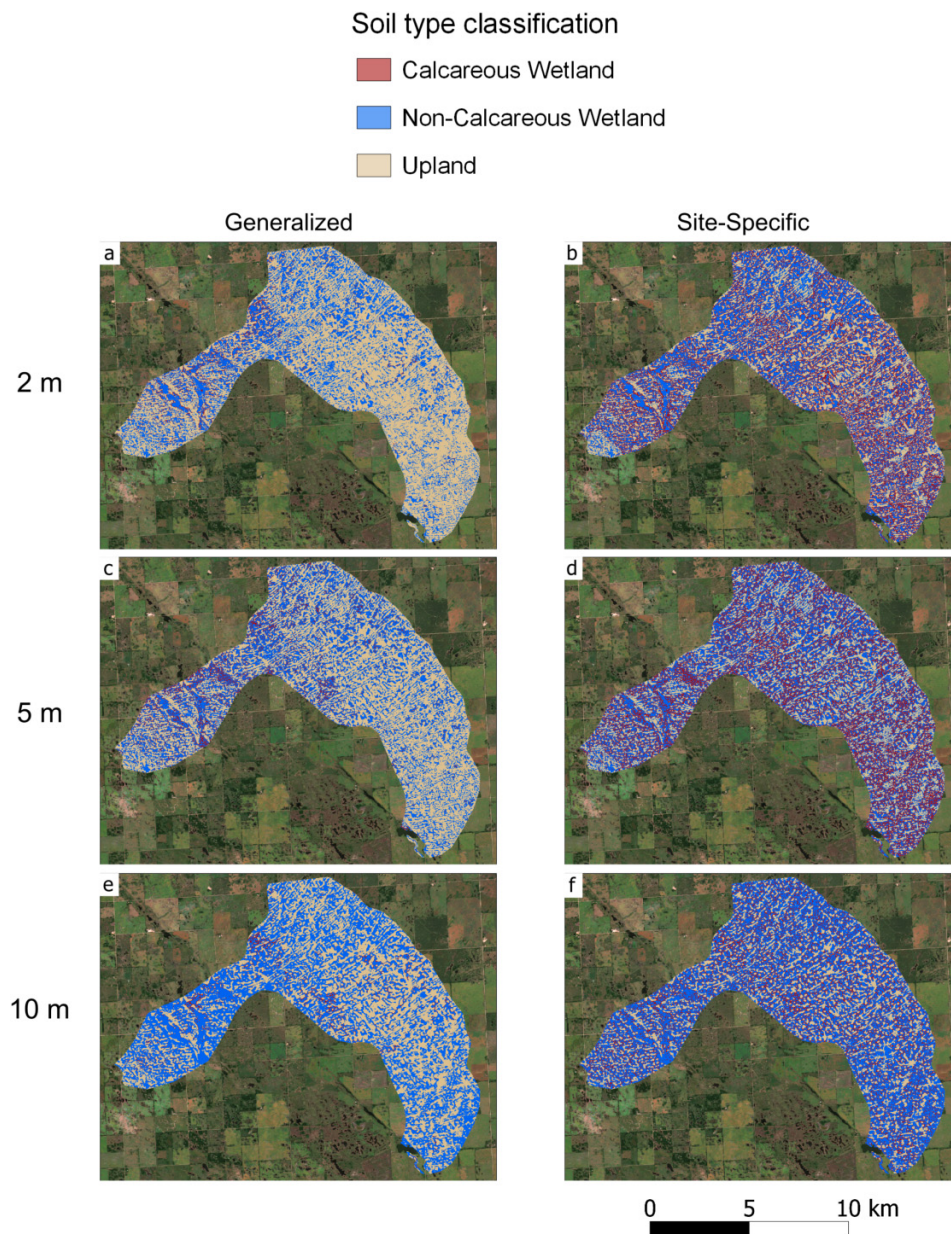
Recent studies focused on predictively mapping wetland classes achieved similar or higher overall accuracies based on independent validation tests: *Mahdianpari et al. (2017)* achieved 94% overall accuracy predictively mapping land cover types (primarily wetland classes) in the Avalon Peninsula of Newfoundland using remote sensed multispectral imagery; *Jahncke et al. (2018)* achieved overall accuracy of 89.2% mapping wetland classes in an area south of Halifax, Nova Scotia by using various remote sensed data, including LiDAR

data; and *Millard and Richardson (2014)* achieved overall accuracy of 72.8% predictively mapping land cover classes in a peatland region near Ottawa, Ontario using SAR and LiDAR data. These studies focused on mapping wetland classes that were typically defined by their vegetative characteristics, which can be closely correlated to spectral reflectance bands and vegetative indices, making them more easily predicted. Achieving similar overall accuracies when predictively mapping soil classes, particularly in agricultural landscapes, is often more difficult.

Other studies focused on soil class mapping in North America achieved similar or poorer overall accuracies based on in-



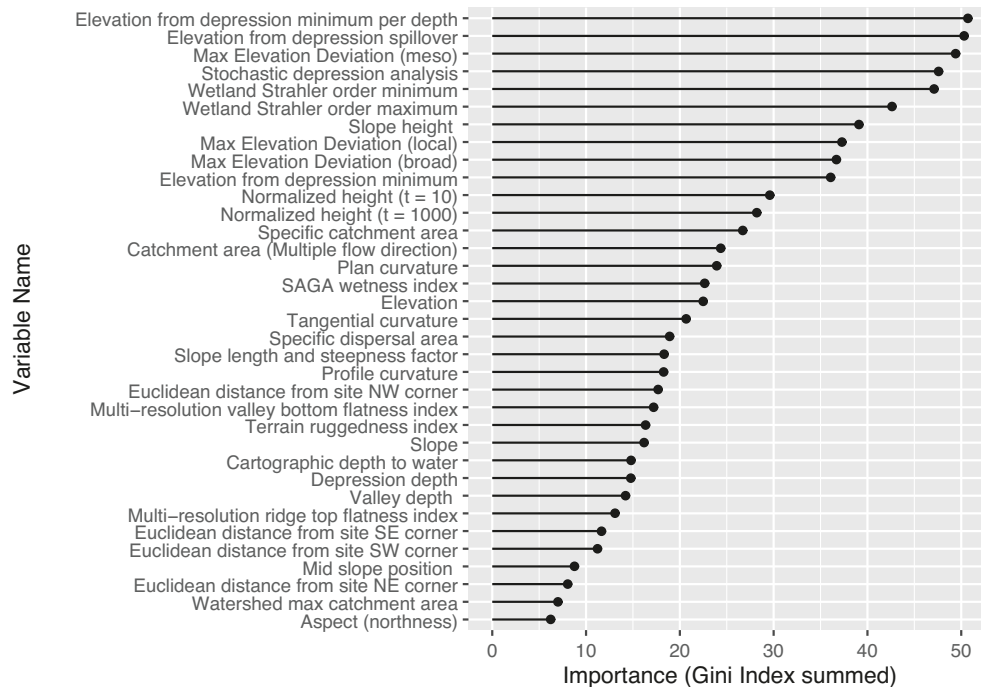
**Fig. 8.** Predicted soil type maps for the full Smith Creek study site based on the generalized and site-specific models built at each spatial resolution. Full-sized versions of each map are provided in supp1a. Satellite RGB imagery from Copernicus (2017). [Colour online.]



dependent validation tests to those achieved in this study: Goldman et al. (2020) achieved overall accuracies of 77.1% and 70.6% mapping soil drainage groups and texture groups, respectively, in a low-relief wetland-rich landscape in the Choptank River Watershed in Maryland and Delaware, USA; Heung et al. (2016) achieved overall accuracies of 70% and 58% mapping soil taxonomic great groups and soil orders, respectively, in the Lower Fraser Valley, British Columbia; and Brungard et al. (2015) achieved Kappa scores of 0.53, 0.32, and 0.19 mapping soil taxonomic subgroups at study sites in Wyoming, New Mexico, and Utah, USA, respectively.

The success of the predictive models of this study indicates that it is possible to accurately predict the spatial distribution of the targeted wetland soil types using primarily terrain

derivatives. This corresponds to the findings of Pennock et al. (2014) who were able to relate key hydropedological features to terrain and elevation-based characteristics due to the substantial influences of topography and hydrology on wetland soil formation in this region. This is not to discount the importance of the SCORPAN model (soil as a function of soil, climate, organisms, relief, parent material, and spatial position) (McBratney et al. 2003), which has proven to be a guiding principle for predictive digital soil mapping throughout the world. The predictive models built in this study would likely improve by incorporating predictor variables that more directly reflect the other elements of SCORPAN like remote sensing multispectral imagery, climate data, and parent material information. The success of the discussed wetland class

**Fig. 9.** Predictor variable importance based on Gini index summed across random forest models.

studies that included remote sensed data beyond terrain data highlights their importance for mapping wetland resources; studying the effects of incorporating these types of features as predictor variables is a crucial next step for research on mapping wetland soils and wetlands more generally in the Canadian PPR.

### Predictor variable importance

Terrain derivatives generated for this study were some of the most important predictor variables for accurately mapping wetland soil types, specifically the elevation from depression minimum per depth and elevation from depression spillover derivatives, which were the first and second most important predictor variables, respectively. Pennock et al. (2014) found that the boundary between the distribution of recharge (noncalcareous) wetland soils within wetland centers and the surrounding discharge ring of calcareous wetland soils closely matched the extent of maximum recorded water levels. Their finding was determined at St. Denis where pond water levels have been recorded annually since 1968. This information does not exist for most locations but elevation from depression spillover works as a proxy for this information because the depression spillover elevation reflects the maximum possible water level for a wetland. The elevation from depression minimum per depth derivative modified the elevation from depression minimum derivative to reflect different relationships between depression size and soil distributions. A gleysolic soil may form on the edge of a large wetland at an elevation of 1 m above the depression bottom because a large wetland is likely to store enough water to inundate that position, whereas at a position of 1 m along a hill-

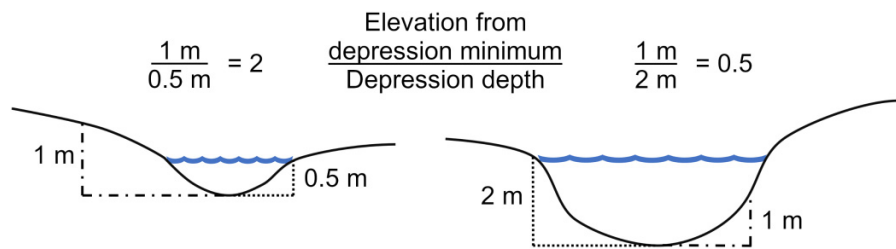
slope above a small depression, it is less likely that a gleysolic or wetland-related soil will form (Fig. 10).

Wetland Strahler order calculations were developed and successfully used to map the distributions of solute-rich wetlands in the same regions of the Canadian PPR (Kiss and Bedard-Haughn 2021). Solute-rich wetlands typically correspond to those with calcareous wetland soils throughout their basin, so these derivatives were likely important in the random forest models for differentiating between wetlands that would and would not form such soil distributions.

The maximum elevation deviation derivatives and the normalized height and slope height derivatives aim to quantify a grid cell's positions in relation to the local and (or) broader landscape. These can reflect similar characteristics to the depression-focused terrain derivatives developed for this study like the elevation from depression minimum derivative but do not limit the calculating window to the watershed of each depression, which helps to better reflect nuanced pedological relationships at the watershed edges. Additionally, hydro-pedological forces occur over a range of scales (Pennock et al. 2014) and features that capture characteristics reflective of varying scales, like these, are clearly important for hydro-pedological mapping.

Other studies focused on mapping wetland soils and wetland classes incorporated similar terrain derivatives that quantify a grid cell's position in relation to its surrounding landscape and found them to be important predictor variables. Goldman et al. (2020) incorporated the deviation from mean elevation derivative, which calculates the difference in a grid cell's elevation compared to the mean elevation from its 200 m radius divided by standard deviation. This was consistently one of the top two most important predictor vari-

**Fig. 10.** Schematic diagram showing varying values for the elevation from depression minimum per depth given a point that is 1 m above the depression minimum for a shallow depression (left) and a deep depression (right). Different soils are expected to form in these two positions. The elevation from depression minimum per depth derivative modifies the elevation from depression minimum derivative by dividing by the depression's overall depth, which helps to quantify the relationships between these two characteristics for depressions of varying sizes. [Colour online.]



ables in their study on mapping soil drainage groups and texture classes in a wetland-rich landscape in the Atlantic Coastal Plain of the USA. This derivative is similar to the maximum elevation deviation derivatives used in this study, except that it only considers the deviation compared to the single spatial scale of 200 m rather than across multiple spatial scales. Jahncke et al. (2018) found that, in addition to slope, topographic position index was an important predictor variable for mapping wetland classes in Nova Scotia. Topographic position index assigns an index to a grid cell according to its relative position compared to local valleys and hilltops. The normalized height and slope height derivatives similarly quantify a grid cell's position compared to local valleys and hilltops but can account for that relationship across various neighbourhood scales.

Whitebox Geospatial's stochastic depression analysis (Lindsay and Creed 2005) derivative (fourth most important variable) estimates the likelihood that a position belongs to a closed topographic depression. Depressions were accurately captured in the outputs generated by this tool, which were likely useful for in distinguishing the distributions of wetland and upland soils in the random forest models. This terrain derivative is recommended for depression or wetland landscape studies in this region.

### Spatial resolution

Predicted maps generated based on predictor variables derived at the highest tested resolution (2 m) generally had the highest predictive accuracies based on cross and independent validation. The resulting soil maps generated at this resolution were acceptable and matched the established understanding of the distribution of these soil types without excessive errors or artefacts. Calculating terrain derivatives and applying predictive models to stacks of predictor variable rasters at this resolution is computationally intensive, which is an important consideration. Predictive mapping at 5 and 10 m resolution is exponentially less computationally intensive. Model performance at these resolutions was only slightly poorer than mapping at 2 m resolution. Predictive accuracy for cross and independent validation tests were similar between the 5 and 10 m based models, but visual inspection determined that the maps generated at 10 m resolution typically lacked the detail achieved at higher resolutions and

soil types were often distributed in blocky and discontinuous patterns. Maps generated at 10 m resolution were still considered successful but working at higher resolutions produced better outputs generally.

### Generalized versus site-specific models

The site-specific models performed slightly better overall compared to the generalized models. They had higher cross-validation Kappa scores on average at all resolutions and generally achieved higher or similar accuracy based on independent validation tests, with the exception for the Brown-Smith Creek-Drained independent validation test, in which the generalized models consistently had higher accuracy.

A notable outcome of the generalized models is that the resulting maps were more consistent across sites than the site-specific models. In this case, this effect may be not desirable because the distribution and frequency of wetland soils at each study area varies substantially. Wetland soils and calcareous wetland soils occur less frequently at the Swift Current study site, which causes the generalized models to underpredict these soil types at the other study sites. The generalized models did, however, achieve higher predictive accuracy according to the Smith Creek-Drained validation test and produced output maps considered more appropriate in the eastern portion of the Smith Creek study area because they predicted fewer occurrences of wetland soils there. The poorer performance of the site-specific models in this instance is likely due to the lack of training data in that region of the study area, as training data were only collected in the western and northern portions (Fig. 2). With better distribution of training data across the entire study area, it is expected that the site-specific models would perform better. This highlights an advantage of the generalized models: because they were built based on more generalizable patterns, they were more generally applicable and are less variable based on local training data.

Most of the terrain derivatives used as predictor variables are relative measures and cannot be used to differentiate between sites in the generalized models. The exception to this is the elevation variable, as each site falls within nearly entirely separate bands of elevation values (Swift Current: 706–832 m; St. Denis: 540–600 m; Smith Creek: 520–545 m). Elevation was a relatively useful predictor variable (Fig. 9), which



was likely due to its ability to split the training data based on study area within the random forest models. Incorporating more predictor variables that can be used to distinguish sites may improve generalized model performance. Incorporating climatic data would provide a variable that directly relates to a key difference between these sites: their moisture regimes.

In the case where predictive mapping models can be trained using either only local or local and global training data, it is recommended to test both and determine on a case-by-case basis which performs better. Extrapolating predictive models to areas where no training data were collected should be done so cautiously. In the case of this study, a model trained using only data from Swift Current is very likely to underpredict the extent of wetland soils at Smith Creek and vice versa. Although St. Denis represents something of a midpoint in terms of frequency of wetland soils, models trained at the other two sites would not be capable of predicting the fully calcareous wetlands at this site.

## Conclusion

Based on the results of this study, mapping of calcareous and noncalcareous wetland soil types in the Saskatchewan PPR is possible using terrain derivatives without additional remote sensing data. This success is partly due to the inclusion of terrain derivatives that were specifically designed to reflect unique morphological characteristics of the PPR that are key to distinguishing the distributions of wetland soil types. Optimal results were obtained using 2 m resolution DEMs. However, acceptable results were still obtained using 5 and 10 m resolution DEMs, which require significantly less computational resources. The techniques developed in this study could help with (i) targeting sampling for studies focused on testing relationships between wetland soil types and ecosystem functions, (ii) guiding wetland conservation and restoration efforts to maximize wetland ecosystem services, and (iii) establishing broad-scale wetland ecosystem function estimates. Future work is needed to evaluate the value of additional covariates beyond the terrain derivatives used in this study.

## Acknowledgements

Financial support was provided for this project by the Lake Winnipeg Basin Stewardship Fund, the Natural Sciences and Engineering Research Council of Canada's (NSERC) Discovery Grant Program, Ducks Unlimited Canada, and the University of Saskatchewan. Predictive models were built and applied using cloud computing resources granted from the Digital Research Alliance of Canada. The soil samples collected from St. Denis between 2000 and 2002 were provided by Dr. Garth van der Kamp from the Global Institute for Water Security and Randy Schmidt from Environment Canada. Soil sampling and profile descriptions for the Brown validation datasets were conducted by Robin Brown as a part of her Master's program in the Department of Soil Science at the University of Saskatchewan. Soil sampling and profile descriptions for the BIOCAP validation datasets were coordinated by Dr. Darwin Anderson, Professor Emeritus at the Department of Soil Sci-

ence at the University of Saskatchewan. Thanks to Dr. Dan Pennock, Professor Emeritus at the Department of Soil Science at the University of Saskatchewan, for guidance in determining profile description criteria for wetland soil type classifications. Thanks to Daniel Saurette, Land Resource Specialist at Ontario Ministry of Agriculture, Food and Rural Affairs, for recommendations for terrain derivatives to use as predictor variables. Thanks to Dr. Louis-Pierre Comeau, Jay Bauer, Lukas Smith, and Robin Brown for field assistance.

## Article information

### History dates

Received: 27 February 2022

Accepted: 12 June 2022

Accepted manuscript online: 23 November 2022

Version of record online: 23 November 2022

### Notes

This paper is part of a Collection entitled "Advances in Soil Survey & Classification in Canada".

### Copyright

© 2022 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

### Data availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

## Author information

### Author ORCIDs

Jeremy Kiss <https://orcid.org/0000-0001-8195-1518>

Angela Bedard-Haughn <https://orcid.org/0000-0002-3971-8509>

0000-0002-3971-8509

Preston Sorenson <https://orcid.org/0000-0002-2958-1246>

### Author notes

Angela Bedard-Haughn served as an Associate Editor at the time of manuscript review and acceptance; peer review and editorial decisions regarding this manuscript were handled by Chuck Bulmer.

### Author contributions

Jeremy Kiss: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, writing—original draft, and writing—review and editing; Angela Bedard-Haughn: conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, and writing—review and editing; Preston Sorenson: investigation, methodology, software, validation, visualization, and writing—review and editing.



## Competing interests

The authors declare there are no competing interests.

## Funding information

Financial support was provided for this project by the Lake Winnipeg Basin Stewardship Fund, the Natural Sciences and Engineering Research Council of Canada's Discovery Grant Program, Ducks Unlimited Canada, and the University of Saskatchewan.

## Supplementary material

Supplementary data are available with the article at <https://doi.org/10.1139/cjss-2022-0034>.

## References

Agriculture and Agri-Food Canada. 2009. ISO 19131 Swift Current LiDAR Project 2009—Data Product Specifications. Available from: <https://open.canada.ca/data/en/dataset/72003a79-c799-49de-b135-de5af833f029>.

Arndt, J.L., and Richardson, J.L. 1988. Hydrology, salinity, and hydric soil development in some North Dakota prairie-pothole wetlands. *Wetlands* **8**: 93–108. doi:10.1007/BF03160595.

Arndt, J.L., and Richardson, J.L. 1989. Geochemistry of hydric soil salinity in a recharge-throughflow-discharge prairie-pothole wetland system. *Soil Sci. Soc. Am.* **53**: 848–855. doi:10.2136/SSAJ1989.03615995005300030037X.

Assami, T., and Hamdi-Aïssa, B. 2019. Digital mapping of soil classes in Algeria – a comparison of methods. *Geoderma. Reg.* **16**: e00215. Elsevier. doi:10.1016/j.GEODRS.2019.E00215.

Badiou, P., Page, B., and Akinremi, W. 2018. Phosphorus retention in intact and drained prairie wetland basins: implications for nutrient export. *J. Environ. Qual.* **47**: 902–913. doi:10.2134/jeq2017.08.0336. PMID: 30025042.

Bedard-Haughn, A.K., and Pennock, D.J. 2002. Terrain controls on depressional soil distribution in a hummocky morainal landscape. *Geoderma* **110**: 169–190. doi:10.1016/s0016-7061(02)00229-x.

Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., and MacMillan, R.A. 2018. Spatial modelling with Euclidean distance fields and machine learning. *Eur. J. Soil Sci.* **69**: 757–770. John Wiley & Sons, Ltd. doi:10.1111/EJSS.12687.

Berhane, T.M., Lane, C.R., Wu, Q., Autrey, B.C., Anenkhonov, O.A., Chepinoga, V. V., and Liu, H. 2018. Decision-Tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. *Remote Sens.* **10**: 580. Multidisciplinary Digital Publishing Institute. doi:10.3390/RS10040580.

Böhner, J., Köthe, R., Conrad, O., Gross, J., Ringeler, A., and Selige, T. 2002. Soil regionalisation by means of terrain analysis and process parameterisation. *Eur. Soil Bur.* **7**: 213–222. Available from [https://esdac.jrc.ec.europa.eu/ESDB\\_Archive/eusoiils\\_docs/esb\\_r/n07\\_ESBResRep07/601Bohner.pdf](https://esdac.jrc.ec.europa.eu/ESDB_Archive/eusoiils_docs/esb_r/n07_ESBResRep07/601Bohner.pdf).

Böhner, J., and Selige, T. 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. *Göttinger Geographische Abhandlungen. Vol.* **115**: 13–28.

Brown, R., Zhang, Z., Comeau, L.-P., and Bedard-Haughn, A. 2017a. Effects of drainage duration on mineral wetland soils in a prairie pothole agroecosystem. *Soil. Tillage Res.* **168**: 187–197. doi:10.1016/j.still.2016.12.015.

Brown, R.L., Van der Kamp, G., Zhang, Z., and Bedard-Haughn, A. 2017b. Evaluation of phosphorus available in two prairie wetlands: discharge vs recharge soils. *Can. J. Soil Sci.* **97**: 789–792. doi:10.1139/cjss-2017-0018.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., and Edwards, T.C. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* **239**: 68–83. Elsevier B.V. doi:10.1016/j.geoderma.2014.09.019.

Conrad, O., Bechtel, B., Bock, B., Dietrich, H., Fischer, E. Gerlitz, L., et al. 2015. System for Automated Geoscientific Analyses (SAGA). *Geosci. Model Dev.* **8**: 1991–2007. doi:10.5194/gmd-8-1991-2015.

Cook, B.J., and Hauer, F.R. 2007. Effects of hydrologic connectivity on water chemistry, soils, and vegetation structure and function in an intermontane depressional wetland landscape. *Wetlands* **27**: 719–738. doi:10.1672/0277-5212(2007)27[719:EOHCOW]2.0.CO;2.

Costa-Cabral, M.C., and Burges, S.J. 1994. Digital Elevation Model Networks (DEMON): a model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resour. Res.* **30**: 1681–1692. doi:10.1029/93WR03512.

Copernicus. 2017. Sentinel-2 RGB imagery. European Space Agency.

Ducks Unlimited. 2005. Ducks Unlimited's International Conservation Plan. Available from: <https://www.ducks.org/conservation/international-conservation-plan>.

Gallant, J.C., and Dowling, T.I. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* **39**(12): 1–13. doi:10.1029/2002WR001426.

Goldman, M.A., Needelman, B.A., Rabenhorst, M.C., Lang, M.W., McCarty, G.W., and King, P. 2020. Digital soil mapping in a low-relief landscape to support wetland restoration decisions. *Geoderma*, **373**: 1–16. doi:10.1016/j.GEODERMA.2020.114420.

Government of Canada. 2008. Water Survey of Canada Sub-Sub-Drainage Areas. Available from: <https://databasin.org/datasets/0f0b8d144ea94fdd915ec40b33410ecf>.

GRASS Development Team. 2017. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. Available from: <http://grass.osgeo.org/>.

Hayashi, M., Van der Kamp, G., and Rudolph, D.L. 1998. Water and solute transport between a prairie wetland and adjacent upland, 1. Water balance. *J. Hydrol.* **207**: 42–55. doi:10.1016/S0022-1694(98)00098-5.

Heagle, D., Hayashi, M., and Van der Kamp, G. 2013. Surface-subsurface salinity distribution and exchange in a closed-basin prairie wetland. *J. Hydrol.* **478**: 1–14. doi:10.1016/j.jhydrol.2012.05.054.

Hengl, T. 2006. Finding the right pixel size. *Comput. Geosci.* **32**: 1283–1298. doi:10.1016/j.cageo.2005.11.008.

Hengl, T., and Reuter, H.I. 2008. *Geomorphometry: Concepts, Software, Applications*. Elsevier.

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., and Schmidt, M.G. 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, **265**: 62–77. doi:10.1016/j.geoderma.2015.11.014.

Hounkpatin, K.O.L., Schmidt, K., Stumpf, F., Forkuor, G., Behrens, T., Scholten, T., et al. 2018. Predicting reference soil groups using legacy data: a data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). *Sci. Rep.* **8**: 9959. Nature Publishing Group. doi:10.1038/s41598-018-28244-w.

Jahncke, R., Leblon, B., Bush, P., and LaRocque, A. 2018. Mapping wetlands in Nova Scotia with multi-beam RADARSAT-2 polarimetric SAR, optical satellite imagery, and Lidar data. *Int. J. Appl. Earth Obs. Geoinf.* **68**: 139–156. Elsevier B.V. doi:10.1016/j.JAG.2018.01.012.

Jenks, G.F. 1967. The data model concept in statistical mapping. *Int. Yearb. Cartogr.* **7**: 186–190.

Jeune, W., Francelino, M.R., De Souza, E., Fernandes Filho, E.I., and Rocha, G.C. 2018. Multinomial logistic regression and random forest classifiers in digital mapping of soil classes in western haiti. *Rev. Bras. Ciência do Solo* **42**: 170133. Sociedade Brasileira de Ciência do Solo. doi:10.1590/18069657RBSC20170133.

Kiss, J. 2022. PPR wetland soil type mapping scripts. Available from: [https://github.com/jeremy-kiss/PPR\\_wetland\\_soil\\_type\\_mapping](https://github.com/jeremy-kiss/PPR_wetland_soil_type_mapping) [accessed 14 June 2022].

Kiss, J., and Bedard-Haughn, A. 2021. Predictive mapping of solute-rich wetlands in the Canadian Prairie Pothole Region through high-resolution digital elevation model analyses. *Wetlands*, **41**: 1–17. doi:10.1007/S13157-021-01436-3.

Knuteson, J.A., Richardson, J.L., Patterson, D.D., and Prunty, L. 1989. Pedogenic carbonates in a Calciaquoll associated with a recharge wetland. *Soil Sci. Soc. Am.* **53**: 495–499. doi:10.2136/sssaj1989.03615995005300020032x.

Kuhn, M. 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**: 1–26. doi:10.18637/jss.v028.i05. PMID: 27774042.

Landis, J.R., and Koch, G.G. 1977. The measurement of observer agreement of categorical data. *Biometrics* **33**: 159–174. doi:10.2307/2529310. PMID: 843571.

- Lindsay, J.B. 2014. The Whitebox Geospatial Analysis Tools project and open-access GIS. In Proceedings of the GIS Research UK 22nd Annual Conference. The University of Glasgow, 16–18 April 2014. doi:10.13140/RG.2.1.1010.8962.
- Lindsay, J.B. 2016. Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrol. Process.* **30**: 846–857. doi:10.1002/hyp.10648.
- Lindsay, J.B., Cockburn, J.M.H., and Russell, H.A.J. 2015. An integral image approach to performing multi-scale topographic position analysis. *Geomorphology* **245**: 51–61. Elsevier Ltd. doi:10.1016/j.geomorph.2015.05.025.
- Lindsay, J.B., and Creed, I.F. 2005. Sensitivity of digital landscapes to artifact depressions in remotely-sensed DEMs. *Photogramm. Eng. Remote Sens.* **71**: 1029–1036. American Society for Photogrammetry and Remote Sensing. doi:10.14358/PERS.71.9.1029.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., and Motagh, M. 2017. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. *ISPRS J. Photogramm. Remote Sens.* **130**: 13–31. Elsevier. doi:10.1016/j.isprsjprs.2017.05.010.
- Malone, B.P., Minasny, B., and McBratney, A.B. 2017. *Using R for Digital Soil Mapping. Progress in Soil Science.* Springer, Switzerland. doi:10.1007/978-3-319-44327-0.
- McBratney, A.B., Mendonça-Santos, M.L., and Minasny, B. 2003. On digital soil mapping. *Geoderma* **117**: 3–52. doi:10.1016/S0016-7061(03)00223-4.
- McFarlan, L.S. 2021. *Where's the P in Prairie Potholes? Identifying Patterns of Phosphorus Accumulation in Canadian Prairie Wetlands.* University of Saskatchewan. Available from: <https://hdl.handle.net/10388/13524>.
- Millar, J.B. 1976. Wetland classification in western Canada: a guide to marshes and shallow open water wetlands in the grasslands and parklands of the Prairie Provinces. Canadian Wildlife Service Report Series 37, Ottawa, ON.
- Millard, K., and Richardson, M. 2014. Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier. *Can. J. Remote Sens.* **39**: 290–307. Taylor & Francis. doi:10.5589/M13-038.
- Moore, I.D., Grayson, R.B., and Ladson, A.R. 1991. Digital terrain modeling: a review of hydrological geomorphological and biological applications. *Hydrol. Process.* **5**: 3–30. doi:10.1002/hyp.3360050103.
- Murphy, P.N.C., Ogilvie, J., and Arp, P. 2009. Topographic modelling of soil moisture conditions: a comparison and verification of two models. *Eur. J. Soil Sci.* **60**: 94–109. doi:10.1111/j.1365-2389.2008.01094.x.
- Nachshon, U., Ireson, A., van der Kamp, G., and Wheeler, H. 2013. Sulfate salt dynamics in the glaciated plains of North America. *J. Hydrol.* **499**: 188–199. Elsevier B.V. doi:10.1016/j.jhydrol.2013.07.001.
- Pennock, D., Bedard-Haughn, A., Kiss, J., and van der Kamp, G. 2014. Application of hydrogeology to predictive mapping of wetland soils in the Canadian Prairie Pothole region. *Geoderma*, **235–236**: 199–211. Elsevier B.V. doi:10.1016/j.geoderma.2014.07.008.
- Pennock, D., Bedard-Haughn, A., and Viaud, V. 2011. Chernozemic soils of Canada: genesis, distribution, and classification. *Can. J. Soil Sci.* **91**: 719–747. doi:10.4141/cjss10022.
- Pennock, D.J., Zebarth, B.J., and De Jong, E. 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. *Geoderma* **40**: 297–315. doi:10.1016/0016-7061(87)90040-1.
- Quinn, P., Beven, K.J., Chevallier, P., and Planchon, O. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrol. Process.* **5**: 59–79. doi:10.1002/HYP.3360050106.
- R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <https://www.r-project.org/>.
- Reschke, J., and Hüttich, C. 2014. Continuous field mapping of Mediterranean wetlands using sub-pixel spectral signatures and multi-temporal Landsat data. *Int. J. Appl. Earth Obs. Geoinf.* **28**: 220–229. Elsevier. doi:10.1016/j.jag.2013.12.014.
- Riley, S.J., DeGloria, S.D., and Elliot, R. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermt. J. Sci.* **5**: 23–27.
- Saskatchewan Geospatial Imagery Collaborative. 2010. FlySask2. Available from: <https://www.flysask2.ca/> [accessed 5 January 2022].
- Saskatchewan Geospatial Imagery Collaborative. 2014. FlySask2. Available from: <https://www.flysask2.ca/> [accessed 5 January 2022].
- Saskatchewan Geospatial Imagery Collaborative. 2016. FlySask2. Available from: <https://www.flysask2.ca/> [accessed 5 January 2022].
- Saskatchewan Land Resource Unit. 2009. SKSIDv4, Digital Soil Resource Information for Agricultural Saskatchewan, 1:100,000 scale. Saskatoon, SK.
- Saskatchewan Soil Survey Staff. 1991. The soils of Langenburg, Fertile Belt, Churchbridge, Saltcoats rural municipalities Saskatchewan. Saskatoon, SK.
- Saurette, D. 2021. ON Soil Survey - Euclidean Distance Field. Available from: <https://github.com/newdale/onsoilsurvey/blob/master/R/oss.edf.R> [accessed 13 June 2022].
- Schönauer, M., and Maack, J. 2021. R-code for calculating depth-to-water (DTW) maps using GRASS GIS. doi:10.5281/ZENODO.5718133.
- Segers, R. 1998. Methane production and methane consumption—a review of processes underlying wetland methane fluxes [Review]. *Biogeochemistry* **41**: 23–51. doi:10.1023/A:1005929032764.
- Soil Classification Working Group. 1998. *The Canadian System of Soil Classification.* 3rd ed.
- Strobl, C., Malley, J., and Tutz, G. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods.* **14**: 323–348. doi:10.1037/a0016973. PMID: 19968396.
- Van der Kamp, G., and Hayashi, M. 2009. Groundwater-wetland ecosystem interaction in the semiarid glaciated plains of North America. *Hydrogeol. J.* **17**: 203–214. doi:10.1007/s10040-008-0367-1.
- Winter, T.C., and Rosenberry, D.O. 1998. Hydrology of Prairie Pothole wetlands during drought and deluge: a 17-year study of the Cottonwood Lake wetland complex in North Dakota in the perspective of longer term measured and proxy hydrological records. *Clim. Change.* **40**: 189–209. doi:10.1023/A:1005448416571.
- Wright, M.N., and Ziegler, A. 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**(1): 1–17. doi:10.18637/jss.v077.i01. PMID: 28649186.
- Zevenbergen, L.W., and Thorne, C.R. 1987. Quantitative analysis of land surface topography. *Earth Surf. Proc. Land.* **12**: 47–56. doi:10.1002/esp.3290120107.

## Appendix A

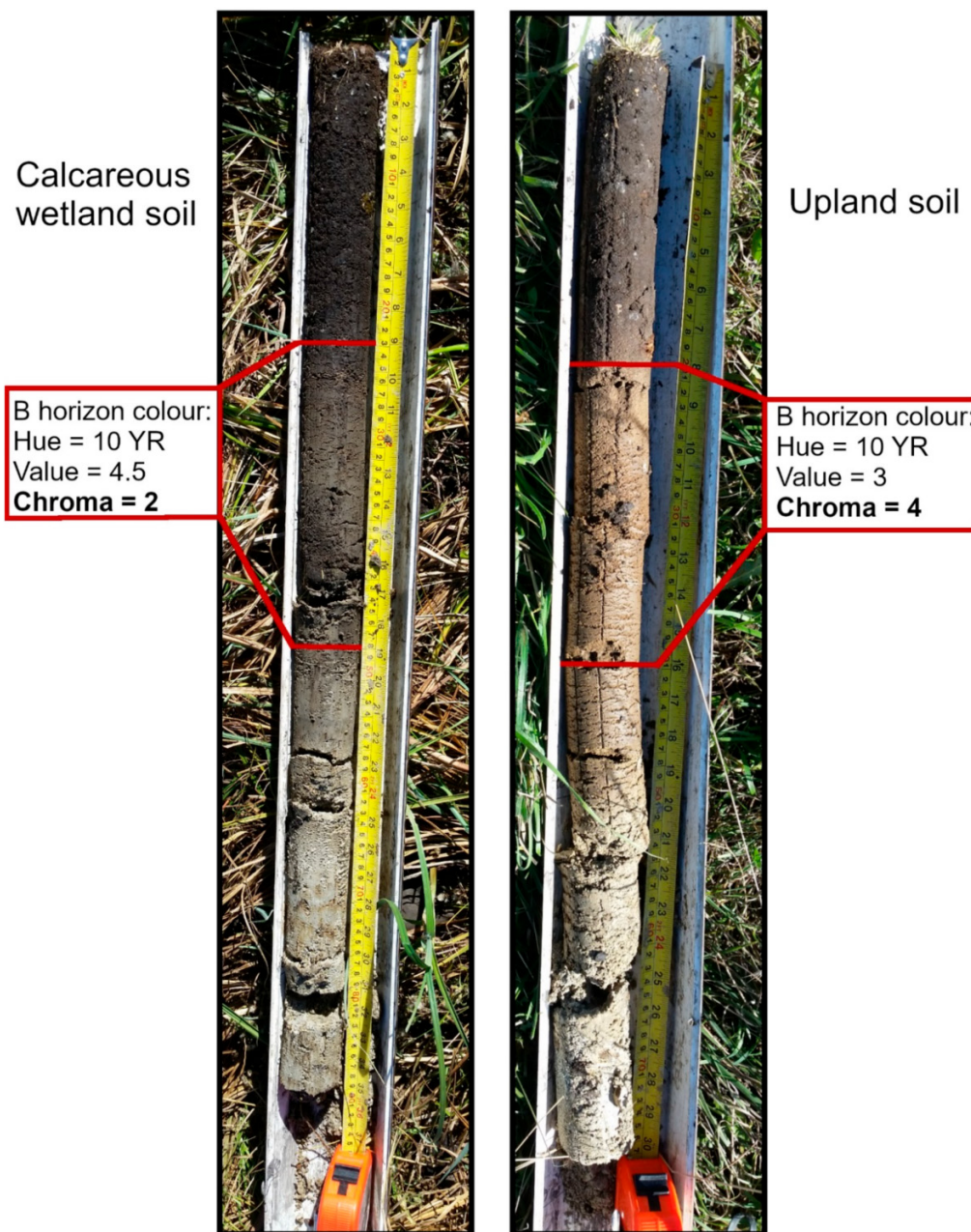
### Formation of fully calcareous wetlands via fill and spill and shallow groundwater flow

The importance of fill and spill flow and shallow groundwater flow on wetland hydrology within Prairie Pothole Region (PPR) landscapes has been recognized over the past decade (Nachshon et al. 2013; Brannen et al. 2015). Wetlands can receive significant inputs of solutes from shallow groundwater flow and fill and spill from upslope wetlands (Cook and Hauer, 2007; Nachshon et al. 2013), which can result in CaCO<sub>3</sub>-enriched soils throughout the wetland basin regardless of the wetland's relationship with groundwater.

Miller et al. (1985), inter-alia, provide a detailed soil profile description for a soil located within wetland 125S at the St. Denis National Wildlife Area (referred to as wetland 14 within that study). This soil profile had high total carbonates and electrical conductivity throughout its profile. They observed a shallow groundwater table beneath this wetland that experienced both recharge and discharge movement of groundwater. Wetland 125S is notably located within a sequence of wetlands where fill and spill flow has been observed during wetter periods (Shaw et al. 2012). The dynamic water levels of the wetlands within this sequence can only be attributed to surface hydrology as deep groundwater movement is too slow to cause such dynamic changes. The high accumulation of solutes within this wetland is likely the result of the same fill and spill processes that cause its large variation in wa-



**Fig. A1.** Comparison photos of B horizon color for a nongleyed calcareous wetland soil (left) and the upland soil adjacent to it along the sampling transect (right). [Colour online.]



ter levels rather than the slow recharge and discharge of its groundwater movement. Wetlands that do not necessarily receive groundwater discharge but do receive substantial solute contributions through fill and spill and shallow groundwater flow can also develop strongly calcareous soils throughout their basins.

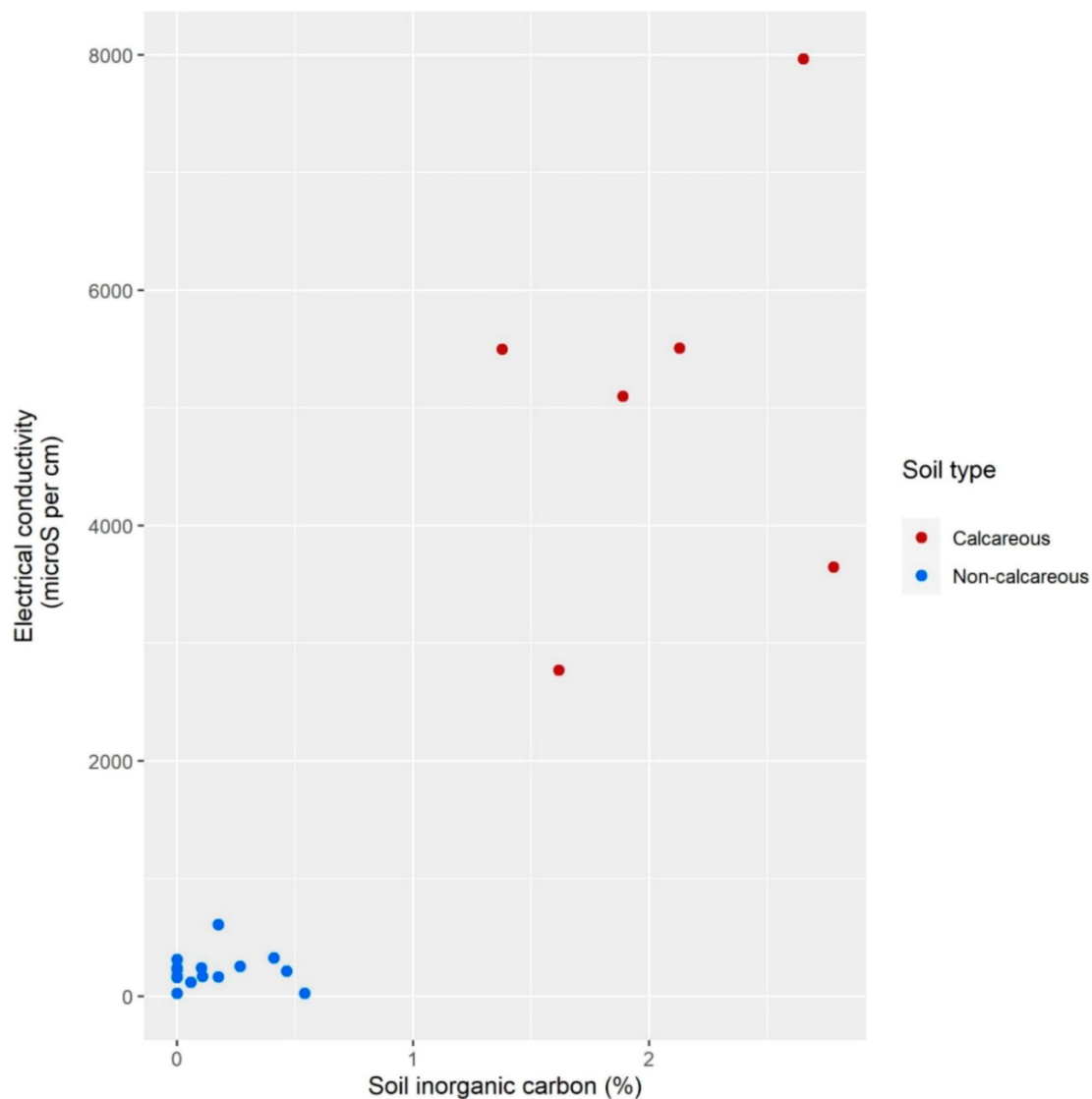
### Transect sampling design

Sample points were placed along a single transect per depression and transect placement was determined in the field. Transects were placed away from potential spill channels and ran in straight lines with the origins at the basin centers. The first transect point was placed in the wetland center if the

wetland was not inundated and at the water's edge if the wetland was inundated. Samples were taken upslope along the transect until soils that were certain to be classified as upland soils were observed.

In a study on PPR wetlands in similar regions, [Pennock et al. \(2014\)](#) found recharge (noncalcareous) soils distributed throughout the 0–0.95 m elevation above the wetland basin minimum and discharge (calcareous wetland) soils distributed throughout the 0.95–2 m elevation above the wetland basin minimum. They also found that the boundary between recharge and discharge ring soils within PPR wetlands roughly matched the maximum observed water level. For smaller wetlands, the maximum water level is expected to

**Fig. A2.** Scatterplot of EC and SIC values for the surficial sample depths (either 0–10 or 0–15 cm) for the profiles sampled from the wetland basin centers at the SDNWA between 2000 and 2002. EC, electrical conductivity; SDNWA, St. Denis National Wildlife Area; SIC, soil inorganic carbon. [Colour online.]



match the depression spillover elevation. The sample point spacing was designed to ensure the points would capture any changes in soil types that may occur over these boundaries.

Sample point placement along the transect was based on changes in elevation and spacing between sample points depended on the depth of the depression. For depressions with depths less than 0.3 m, sample points were spaced at elevation increments of 0.1 m until 0.5 m and in 0.25 m increments after 0.5 m. For depressions with depths between 0.3 to 0.5 m, sample points were spaced at elevation increments of 0.25 m. For depressions with depths greater than 0.5 m, sample points were spaced at elevation increments of 0.5 m. Changes in elevation were determined using a Suunto PM-5 clinometer (Suunto, Vantaa, Finland). This method was tested at a transect with seven sample points using a Sokkisha Set5 Total Station (Sokkia, Kanagawa Prefecture, Japan) and was found to have a root mean squared error of 0.02 m.

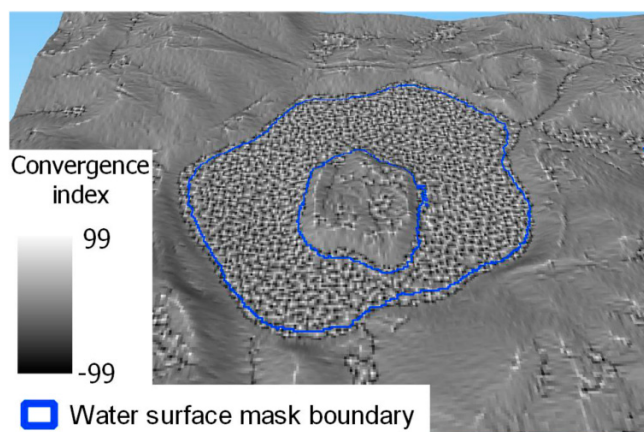
### Wetland soil type classification

The discharge ring surrounding wetlands is expected to have a gradational effect where some soils accumulate substantial contributions of  $\text{CaCO}_3$  and soils in adjacent positions will accumulate  $\text{CaCO}_3$  through the same processes, but not to the same degree. The calcareous wetland soil class was restricted to profiles with moderate to strong  $\text{CaCO}_3$  throughout their profile. The distribution of  $\text{CaCO}_3$  within soil profiles was determined through application of 10% hydrochloric (HCl) acid during the field assessment. Soils were considered to have moderate to strong presence of  $\text{CaCO}_3$  when moderate to strong effervescence was observed with HCl application (bubbles form either thick or low foam (Watson and Pennock, 2016)).

Gleyed soils are the most characteristic feature of wetland soils. They are characterized by a reduction of iron in the soil through anaerobic conditions due to excessive moisture caus-



**Fig. A3.** 3D scene for an example area in the St. Denis study site showing the boundary of the flat-water surface present in the LiDAR-derived DEM as estimated by using the standard deviation of convergence index. LiDAR, light detection and ranging. [Colour online.]



ing the iron to be redistributed either into oxidized pockets, which have a reddish color, or out of the soil causing grey or even blue soil colors (Bedard-Haughn, 2011). Gleysols and gleyed soils were determined based on the criteria outlined in the Canadian System of Soil Classification (CSSC) (Soil Classification Working Group, 1998). Some of the soils in the Swift Current area had very dark colors due to the incorporation of shale within the glacial till parent material (Ayres et al. 1985). This made it difficult to classify soil gleying based on color because the dark parent material had colors with chromas of one according to the Munsell Soil Color Chart (Munsell, Michigan, USA), which can indicate a gleyed horizon (Soil Classification Working Group, 1998). The magnetic susceptibility for the samples from these profiles was analyzed using a Bartington MS-2D meter (Bartington, Oxfordshire, UK) following the protocol described in de Jong et al. (2000). Profiles with horizons having potentially gleyed soil color and magnetic susceptibility less than  $150 \times 10^{-9} \text{ m}^3 \text{ kg}^{-1}$  were determined to be Gleysols according to the findings of de Jong et al. (2005).

The dominant movement of water within gleyed soil profiles is expected to be downward to cause the leaching of iron out of the profile (Bedard-Haughn and Pennock, 2002). Therefore, discharge ring soils with both gleyed profiles and accumulations of  $\text{CaCO}_3$  throughout their profile likely reflect changes in the dominant direction of water movement within the soil over time. The capillary rise of water to the discharge ring is through unsaturated flow (Knuteson et al. 1989), which is unlikely to cause the development of gleyed soils. Therefore, discharge rings and, by extension, calcareous wetland soils are not always gleyed to the extent required to be classified as such according to the CSSC. Due to the gradual effect of the discharge ring and the ubiquity of  $\text{CaCO}_3$  in PPR soils, it was difficult to develop class criteria to distinguish nongleyed calcareous wetland soils from adjacent upland soils that had moderate to strong  $\text{CaCO}_3$  throughout their profile. Several soil criteria were explored to distinguish

nongleyed calcareous wetland soils from upland soils to differentiate where  $\text{CaCO}_3$  accumulations were likely the result of wetland hydrogeological forces.

The upward movement of water to the discharge ring positions often results in a lack of B horizon development, which can be used to distinguish calcareous wetland soils from adjacent upland soils with B horizons (Pennock et al. 2014). However, this characteristic is not always present. Soil organic carbon (SOC) content and A horizon depth thresholds were explored to distinguish the soil classes. Discharge ring positions would be expected to have greater moisture causing greater accumulations of SOC. However, SOC and A horizon depth can reflect many other influences and can be substantially altered with tillage erosion and deposition. Instead, criteria were established to distinguish the soil classes based on soil color of the B horizon, when present. Upland soils have greater color chroma due to the oxidation of iron in the soils, this gives the soil a more reddish color (Fig. A1) (Smith et al. 2011). The presence of moisture within the wetland soils causes the iron to be reduced. In profiles with excessive moisture, the soil forms the gleyed characteristics of grey colors or mottling. The nongleyed calcareous wetland soils would not have had the moisture conditions to form those qualities, but enough moisture to cause a reduced soil color compared to the adjacent oxidized upland soils. The following color criteria for B horizons was incorporated into the classification: for nongleyed Calcareous Chernozems to be classified as a calcareous wetland soil, the B horizon must have a chroma  $< 5$  and a chroma at least one chroma less than an adjacent upland soil B horizon found in the same sampling transect.

Soils with greater than 30 cm of depositional material at the top of their profile were classified as buried/depositional. For soil profiles with less than 30 cm of depositional material, the difference between the soil surface represented by the elevation model and the historic soil surface under which the soil developed is not expected to be substantially different. The depth of 30 cm was selected because it was less than the maximum vertical error considering both the vertical error of the DEM and the potential vertical error caused by the positional error of the GPS. The maximum vertical error of the DEMs used in this study was 22 cm for wetland areas in the St. Denis DEM (Töyrä et al. 2008). There was a 95% probability that the sampled points were within 2 m of the GPS point (Trimble, 2005). Based on the 2 m DEM, the average RMSE of the elevation of the adjacent eight cells for each soil-sampled cell was 13.8 cm. Therefore, the discrepancy between the historic soil surface and the current soil surface with less than 30 cm of depositional material is within the possible range of error between the soil surface and the elevation model (35.8 cm).

### Supplemental training dataset from St. Denis

Due to the lack of training data points from wetland basin centers, a dataset collected from the St. Denis National Wildlife Area (SDNWA) between 2000 and 2002 was used as a supplemental training dataset. This dataset consists of core samples taken from multiple large wetlands at the SDNWA that had dried out during a drought period. A single point was

sampled from the center of the dried-out basin of each wetland and analyzed for SOC, soil inorganic carbon (SIC), electrical conductivity (EC), and pH. Point location coordinates were not recorded for the samples, only the wetland identifier, so for use in this study, sample locations were estimated. The Modified Normalized Difference Water Index (Xu, 2007) was calculated for the study area using Landsat-7 imagery from August 2001 with Google Earth Engine (Gorelick et al. 2017). Pixels with a value greater than 0 were classified as standing water and sample locations were placed as close to the wetland basin center as possible outside of the location classified as standing water.

There were no profile descriptions or classifications for this dataset, only soil samples were collected. Points were sampled at increments following one of two designs: (i) 0–10, 50–60, and 100–110 cm (and occasionally deeper) or (ii) 0–15, 15–30, and in 30 cm increments until at least 120 cm was reached. Each sample was tested for SOC and SIC using a LECO C632 Carbon analyzer (LECO Corp, St. Joseph, MI) and measured for EC and pH using a PC 700 conductivity meter (Oakton Instruments, Vernon Hills, Illinois, USA) and a 1:20 soil to water extract. Because all soil samples were taken from wetland basins, it is assumed that they are not upland soil types. Measured values of soil inorganic carbon and EC were used to determine whether soils were calcareous or noncalcareous wetland soil types. If the surface horizon (0–10 cm depth increment) had a SIC content greater than 1% and an EC greater than 2000  $\mu\text{s cm}$ , the profile was classified as calcareous wetland and otherwise was classified as noncalcareous wetland. Figure A2 shows that these values reflect a noticeable split in the distributions of EC and SIC of these soils.

## DEM processing—water surface identification and reinterpolation

A method similar to that used in Li et al. (2011) to reinterpolate the water surfaces within DEMs was used to reduce noise in the positions and better reflect the shape of the underlying sediment. Li et al. (2011) describe a simple method for determining these surfaces by using the light detection and ranging (LiDAR) intensity data to determine where water was present during the time of collection. However, LiDAR intensity data were not available for these study areas, so water surfaces were estimated from the DEM based on closed topographic depressions and standard deviation of the convergence index terrain derivative (Conrad et al. 2015). Due to the noisy DEM surface in water surface locations, the convergence index exhibited visible patterns of high localized variation in these positions (Fig. A3). The original LiDAR-derived 1 m DEMs were resampled to 2 m using bilinear resampling with GDAL (GDAL/OGR contributors, 2011) and a standard deviation filter with a window of  $19 \times 19$  cells was calculated. Pixels with a standard deviation of convergence index greater than 30 that were within closed topographic depressions were classified as water surfaces, as this threshold best corresponded to the water surface boundary as determined through visual assessment (Fig. A3).

The water surface was masked out of the 1 m DEM, which was then resampled to 5 m, and filled using System for Au-

tomated Geoscientific Analysis's Close Gaps with Spline tool (Conrad et al. 2015). The filled 5 m DEM was then resampled back to 1 m, smoothed using a  $9 \times 9$  mean filter, and used to fill in the water surface areas masked out of the original 1 m DEM. The R script for this method has been made available on GitHub (Kiss, 2022).

The 1 m DEM with water surfaces reinterpolated was resampled to either 2, or 10 m using GDAL's bilinear resampling. Noise present in the DEM was smoothed over three iterations using Whitebox Geospatial's Feature Preserving Smoothing tool (Lindsay et al. 2019) with a filter size of 11 and maximum difference in vectors of 15 degrees, followed by a  $3 \times 3$  mean filter, and filling of single-cell pits.

## References

- Ayres, K.W., Acton, D.F., and Ellis, J.G. 1985. *The Soils of the Swift Current Map Area 72J Saskatchewan*. Saskatoon, SK.
- Bedard-Haughn, A. 2011. Gleysolic soils of Canada: genesis, distribution, and classification. *Can. J. Soil Sci.* **91**: 763–779. doi:10.4141/cjss10030.
- Bedard-Haughn, A.K., and Pennock, D.J. 2002. Terrain controls on depression soil distribution in a hummocky morainal landscape. *Geoderma* **110**: 169–190. doi:10.1016/s0016-7061(02)00229-x.
- Brannen, R., Spence, C., and Ireson, A. 2015. Influence of shallow groundwater-surface water interactions on the hydrological connectivity and water budget of a wetland complex. *Hydrol. Process.* **29**: 3862–3877. doi:10.1002/hyp.10563.
- Conrad, O., Bechtel, B., Bock, B., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Boehner, J. 2015. System for Automated Geoscientific Analyses (SAGA). *Geosci. Model Dev.* **8**: 1991–2007.
- Cook, B.J., and Hauer, F.R. 2007. Effects of hydrologic connectivity on water chemistry, soils, and vegetation structure and function in an intermontane depression wetland landscape. *Wetlands* **27**: 719–738. doi:10.1672/0277-5212(2007)27[719:EOHCOW]2.0.CO;2.
- GDAL/OGR contributors. 2011. *GDAL/OGR Geospatial Data Abstraction Software Library*. Open Source Geospatial Foundation.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**: 18–27. Elsevier. doi:10.1016/j.rse.2017.06.031.
- de Jong, E., Heck, R.J., and Ponomarenko, E.V. 2005. Magnetic susceptibility of soil separates of gleysolic and chernozemic soils. *Can. J. Soil Sci.* **85**: 233–244. doi:10.4141/S04-034.
- de Jong, E., Kozak, L.M., and Rostad, H.P.W. 2000. Effects of parent material and climate on the magnetic susceptibility of Saskatchewan soils. *Can. J. Soil Sci.* **80**: 135–142. doi:10.4141/S99-051.
- Kiss, J. 2022. PPR wetland soil type mapping scripts. [Online] Available from: [https://github.com/jeremy-kiss/PPR\\_wetland\\_soil\\_type\\_mapping](https://github.com/jeremy-kiss/PPR_wetland_soil_type_mapping) [accessed 14 June 2022].
- Knuteson, J.A., Richardson, J.L., Patterson, D.D., and Prunty, L. 1989. Pedogenic carbonates in a Calciaquoll associated with a recharge wetland. *Soil Sci. Soc. Am.* **53**: 495–499. doi:10.2136/sssaj1989.03615995005300020032x.
- Li, S., MacMillan, R.A., Lobb, D.A., McConkey, B.G., Moulin, A., and Fraser, W.R. 2011. Lidar DEM error analyses and topographic depression identification in a hummocky landscape in the prairie region of Canada. *Geomorphology* **129**: 263–275. doi:10.1016/j.geomorph.2011.02.020.
- Lindsay, J.B., Francioni, A., and Cockburn, J.M.H. 2019. LiDAR DEM smoothing and the preservation of drainage features. *Remote Sens.* **11**: 1926. Multidisciplinary Digital Publishing Institute. doi:10.3390/RS11161926.
- Miller, J.J., Acton, D.F., and St. Arnaud, R.J. 1985. The effect of groundwater on soil formation in a morainal landscape in Saskatchewan. *Can. J. Soil Sci.* **65**: 293–307. doi:10.4141/cjss85-033.
- Nachshon, U., Ireson, A., van der Kamp, G., and Wheeler, H. 2013. Sulfate salt dynamics in the glaciated plains of North America. *J. Hydrol.* **499**: 188–199. Elsevier B.V. doi:10.1016/j.jhydrol.2013.07.001.

- Pennock, D., Bedard-Haughn, A., Kiss, J., and van der Kamp, G. 2014. Application of hydopedology to predictive mapping of wetland soils in the Canadian Prairie Pothole Region. *Geoderma* **235–236**: 199–211. Elsevier B.V. doi:[10.1016/j.geoderma.2014.07.008](https://doi.org/10.1016/j.geoderma.2014.07.008).
- Shaw, D.A., Vanderkamp, G., Conly, F.M., Pietroniro, A., and Martz, L. 2012. The fill-spill hydrology of prairie wetland complexes during drought and deluge. *Hydrol. Process.* **26**: 3147–3156. doi:[10.1002/hyp.8390](https://doi.org/10.1002/hyp.8390).
- Smith, C.A.S., Webb, K.T., Kenney, E., Anderson, A., and Kroetsch, D. 2011. Brunisolic soils of Canada: Genesis, distribution, and classification. *Can. J. Soil Sci.* **91**: 695–717. doi:[10.4141/cjss10058](https://doi.org/10.4141/cjss10058).
- Soil Classification Working Group. 1998. *The Canadian System of Soil Classification*, 3rd ed.
- Töyrä, J., Pietroniro, A., Craymer, M., and Véronneau, M. 2008. Evaluation of LiDAR-derived ground surface digital elevation model (DEM) in low-relief regions: case study on the Canadian Prairies. *HYDROSCAN: Airborne laser mapping of hydrological features and resources*. In: Hopkinson, Pietroniro, and Pomeroyeds. pp. 301–326.
- Trimble. 2005. *Getting Started Guide: GeoExplorer 2005 Series*. [Online.] Trimble Navigation Limited, Westminster, CO, USA. Available from: [https://arf.berkeley.edu/files/webfiles/all/arf/equipment/field/gps/trimble/geoXH/GeoExpl2005\\_Getting\\_started.pdf](https://arf.berkeley.edu/files/webfiles/all/arf/equipment/field/gps/trimble/geoXH/GeoExpl2005_Getting_started.pdf).
- Watson, K., and Pennock, D.J. 2016. Section 3: soil profile description. In D.J. Pennock, K. Watson and P. Sanborn, eds. *Field Handbook for Soils of Western Canada*. Pedology Subcommittee, Canadian Society of Soil Science. pp. 1–40.
- Xu, H. 2007. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **27**: 3025–3033. Taylor & Francis. doi:[10.1080/01431160600589179](https://doi.org/10.1080/01431160600589179).