# PREDICTIVE MAPPING OF WETLAND TYPES AND ASSOCIATED SOILS THROUGH DIGITAL ELEVATION MODEL ANALYSES IN THE CANADIAN PRAIRIE POTHOLE REGION

A Thesis Submitted to the College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Soil Science
University of Saskatchewan
Saskatoon, SK, Canada

By
Jeremy Joel Kiss

## PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use that may be made of any material in my thesis. Requests for permission to copy or to make other uses of materials in this thesis, in whole or part, should be addressed to:

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan
Canada, S7N 5C9

Head, Department of Soil Science
College of Agriculture and Bioresources
University of Saskatchewan
51 Campus Dr.
Saskatoon, SK
Canada, S7N 5A8

## DISCLAIMER

Reference in this thesis to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

## ABSTRACT

Effective management strategies are needed to control phosphorus loading of prairie watersheds that contribute to the eutrophication issues of Lake Winnipeg. Prairie Pothole Region (PPR) wetlands provide many ecosystems services including reducing nutrient mobility. Preferential conservation of PPR wetlands with calcium carbonate ($CaCO_3$)-enriched soils may be a more effective strategy for controlling phosphorus loading, as these soils have greater potential to retain phosphorus from agricultural runoff. The spatial distribution of $CaCO_3$-enriched wetland soils is controlled by hydrologic processes that may be modellable using high-resolution digital elevation models (DEMs). Two modelling approaches were tested to map spatial distributions of wetlands and wetland soils expected to be enriched with $CaCO_3$. The models were trained and tested with wetland salinity and soil profile information collected at three Saskatchewan PPR sites, near Swift Current, St. Denis, and Smith Creek. The first model was developed to approximate landscape-scale hydrologic processes from high-resolution DEMs to predict the distributions of fresh and solute-rich wetlands; the solute-rich wetlands represent wetlands expected to have $CaCO_3$-enriched soils. Spill channel connections between wetlands were modelled to characterize wetlands in terms of the runoff contributions they receive, their potential for contributing runoff downslope, and their relative position within the landscape; solute-richness predictions were based on these characteristics. This model was successful and achieved acceptable predictive accuracies based on external validation tests. Digital soil mapping (DSM) methodologies were tested for predicting the spatial distribution of wetland soil classes within PPR landscapes. Target soil classes were defined by hydropedological units that reflect differences in soil $CaCO_3$ enrichment. Multiple machine-learning techniques were tested, which incorporated many topographic attributes derived from the DEMs as predictor variables, including knowledge-based topographic attributes developed specifically to characterize the PPR's morphology. Certain DSM models achieved acceptable predictive accuracy based on external validation tests and mapped soils in expected distributions, but none predicted the occurrence of wetlands with $CaCO_3$-enriched soils distributed throughout their basins. Both modelling approaches could potentially be used to 1) identify wetlands with $CaCO_3$-enriched soils to target for conservation efforts to maximize phosphorus retention and 2) create upscaled estimates of phosphorus retention across the PPR.

## ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Angela Bedard-Haughn for her guidance and support as my supervisor. She is a caring and thoughtful person who always made the time to help. It is a joy to discuss research with Angela because she has such genuine interest and excitement for it; it is inspiring. I would like to also thank my advisory committee, Dr. Dan Pennock, Dr. Bob Clark, and Dr. Bing Si, who always provided useful insight and feedback. Dan and Bob have played major roles in motivating me to pursue work in the sciences, which I am hugely grateful for. As well, a big thank you to Irena Creed for her contributions as an ideal external examiner.

This project would not have been possible without the help of the Applied Pedology lab group and others in the Soil Science Department. Thanks to Robin Brown, Jay Bauer, Louis Comeau, Lukas Smith, and Jacqui Gelineau for help conducting the field work and thanks to Dr. Renato de Freitas, Robin Brown, Courtney Clarke, Megan Horachek, and Eric Neil for help and guidance in conducting the laboratory analyses. It is a pleasure to be a part of the friendly and collaborative community of the Soil Science Department here at the University of Saskatchewan.

This project received funding from Environment Canada (The Lake Winnipeg Basin Stewardship Fund), the Natural Sciences and Engineering Research Council of Canada, Ducks Unlimited Canada, and the University of Saskatchewan. Without this financial support, this project would not have been possible.

A loving thank you to my parents, Vern and Judy, their partners, Karen and Kim, my sister, Kimberly, and brother-in-law, Devin, for being a loving and supportive family. Finally, I am ever grateful to Yonina for being a part of my life and supporting me along this path.

# TABLE OF CONTENTS

Permission to use ..... i
Disclaimer ..... ii
Abstract ..... iii
Acknowledgements ..... iv
List of Tables ..... viii
List of Figures ..... ix
List of Abbreviations ..... xi
1 Introduction ..... 1
1.1 Context ..... 1
1.2 Research objectives ..... 2
1.3 Organization of thesis ..... 2
2 Literature Review ..... 4
2.1 Eutrophication of Lake Winnipeg ..... 4
2.2 Prairie Pothole Region ..... 5
2.3 PPR wetlands and nutrient retention ..... 6
2.4 PPR wetland hydrology and hydrochemistry ..... 7
2.5 PPR soil distributions ..... 10
2.6 Hydrologic modelling in the PPR using digital elevation models ..... 12
2.7 Digital soil mapping in the PPR ..... 14
3 Predictive mapping of solute-rich wetlands in the Canadian Prairie Pothole Region through high-resolution digital elevation model analyses ..... 18
3.1 Preface ..... 18
3.2 Abstract ..... 19
3.3 Introduction ..... 20
3.4 Materials and methods ..... 22
3.4.1 Study areas ..... 22
3.4.2 Digital elevation models and aerial imagery ..... 24
3.4.3 Road and railroad removal from DEM ..... 25
3.4.4 Wetland boundary delineation ..... 25
3.4.5 Stream channel networks ..... 27
3.4.6 Wetland Strahler order ..... 28
3.4.7 Predicted active spill channel networks ..... 30

3.4.8 Terminal wetlands ..... 33
3.4.9 Model tuning parameters ..... 34
3.4.10 Wetland salinity data collection ..... 36
3.4.11 Wetland solute-richness classification ..... 39
3.5 Results and discussion ..... 40
3.5.1 Model training results ..... 40
3.5.2 External validation results ..... 41
3.5.3 Predictive accuracies per solute-richness class ..... 43
3.5.4 Site-specific characteristics ..... 43
3.5.5 Best-performing model parameters ..... 44
3.5.6 Mapped distributions of solute-rich wetlands ..... 45
3.5.7 Modelling wetland salinity for other applications ..... 46
3.6 Conclusion ..... 48
4 Predictive digital soil mapping of wetland soil types in the Canadian Prairie Pothole Region ..... 50
4.1 Preface ..... 50
4.2 Abstract ..... 51
4.3 Introduction ..... 52
4.3.1 Wetland soil classes and their expected spatial distributions ..... 53
4.4 Materials and methods ..... 55
4.4.1 Study areas ..... 55
4.4.2 Sample design ..... 56
4.4.3 Wetland soil type classification ..... 59
4.4.4 Spatial resolution ..... 64
4.4.5 Predictor variables ..... 66
4.4.6 Machine-learning techniques ..... 70
4.4.7 Model tuning and cross-validation ..... 72
4.4.8 External validation datasets ..... 72
4.4.9 Selection of best-performing models ..... 74
4.5 Results ..... 75
4.5.1 Tuned model parameters and cross-validation results ..... 75
4.5.2 External validation results ..... 77
4.5.2.1 BIOCAP – St. Denis external validation ..... 77

4.5.2.2 Brown – Smith Creek – Undrained external validation ..... 80
4.5.2.3 Brown – Smith Creek – Drained external validation ..... 80
4.5.3 Visual assessment ..... 80
4.5.3.1 Visual assessment of 4-class maps ..... 80
4.5.4.2 Visual assessment of 3-class maps ..... 84
4.6 Discussion ..... 84
4.6.1 4-class vs. 3-class mapping ..... 84
4.6.2 Fully discharge/strongly calcareous wetlands ..... 85
4.6.3 Predictor variable importance ..... 87
4.6.4 DEM resolution and smoothing ..... 89
4.6.5 Principal component analysis ..... 91
4.6.6 Drained wetlands ..... 92
4.7 Conclusion ..... 93
5 Synthesis and Conclusions ..... 95
5.1 Summary of findings and general conclusions ..... 95
5.2 Model application for prioritizing conservation efforts ..... 98
5.3 Suggested model improvements and future research directions ..... 100
6 References ..... 102
Appendix A: GIS methodologies for wetland solute-richness class model ..... 111
Appendix A-1: Remove roads from DEMs ..... 111
Appendix A-2: Delineate depression boundary polygons and determine depression depths . 115
Appendix A-3: Create predicted active spill channel networks and ascribe Strahler order and terminal status to wetland polygons ..... 124
Appendix B: GIS methodologies to generate new topographic attributes used as predictor variables in the DSM study ..... 137
Appendix C: Soil profile information ..... 147

## LIST OF TABLES

Table 3.1 General study area characteristics. ..... 23

Table 3.2 Model tuning parameters. ..... 34

Table 3.3 Number of wetlands sampled per stratification at each site for the model training and external validation datasets. ..... 37

Table 3.4 Wetland solute-richness class criteria and the number of observations per class per testing and external validation datasets. ..... 40

Table 3.5 Prediction accuracies for the best-performing models for predicting the solute-richness classes of the Training Set of observations overall and for each study area ..... 42

Table 3.6 Accuracies for predicting the solute-richness classes of the external validation sample sets using the best-performing models for the Training Set. ..... 42

Table 3.7 Wetland solute-richness class criteria based on salinity risks to agronomic productivity and the number of observations per class per dataset. ..... 47

Table 3.8 Prediction accuracies and parameters for the best-performing models for predicting the solute-richness classes based on criteria that represent agronomic risks. ..... 48

Table 4.1 General study area characteristics. ..... 56

Table 4.2 Depression size classes ..... 57

Table 4.3 Number of depressions sampled per study area per relative position in watershed and size stratifications. ..... 57

Table 4.4 Soil class criteria per 3-class and 4-class mapping objectives ..... 60

Table 4.5 Number of soil class observations per 4-class mapping classification scheme. ..... 65

Table 4.6 Number of soil class observations per 3-class mapping classification scheme. ..... 65

Table 4.7 Descriptions of predictor variables and associated references. ..... 66

Table 4.8 Number of soil class observations according to the 4-class mapping classification scheme per external validation set ..... 73

Table 4.9 Number of soil class observations according to the 3-class mapping classification scheme per external validation set ..... 74

Table 4.10 Model parameters and cross-validation predictive accuracy of the best-performing models per model objective ..... 76

Table 4.11 External validation predictive accuracy metrics for the best-performing 4-class mapping models. ..... 78

Table 4.12 External validation predictive accuracy metrics for the best-performing 3-class mapping models. ..... 79

Table 5.1 Wetland type definitions used within this thesis. ..... 99

## LIST OF FIGURES

Fig. 2.1 Approximate extents of the Prairie Pothole Region and the Lake Winnipeg Watershed Basin. ..... 5

Fig. 3.1 Study area locations and approximate extents of the PPR and Lake Winnipeg Watershed Basin. ..... 23

Fig. 3.2 Wetland polygon boundaries based on nested closed depression boundaries and maximum closed depression boundaries at the SDNWA portion of the St. Denis study area. ..... 27

Fig. 3.3 a) Stream channel network Strahler orders ascribed to intersecting wetland polygons for the SDNWA portion of the St. Denis study area. b) Wetland solute-richness class predictions based on wetland Strahler order threshold. The predictions pictured reflect an example where wetlands with Strahler order ≥ 6 are predicted to be solute-rich. ..... 29

Fig. 3.4 Example schematic of a sequence of wetlands. b) Depressions are filled to create continuous stream channel drainage networks as in traditional stream network methodologies. c) Example schematic demonstrating how certain hydrologic connections between wetlands (via fill-and-spill and shallow groundwater flow) are more likely to be active than others. “Slow” and “Fast” refer to the expected movement of groundwater. Wetland 1 requires substantial water contributions before it can contribute fill-and-spill flow and before the water table is high enough for the groundwater to move down-slope through the zone of higher hydraulic conductivity. . 31

Fig. 3.5 a) WDPM water distribution outline outputs based on a small water input (left) and a large water input (right). b) Stream channel networks clipped by the extents of the WDPM output polygons to create the predicted active spill channel networks. c) Strahler orders of the predicted active spill channel networks ascribed to the wetland polygons. d) Solute-richness classes predictions per wetland (dependent on the specified Strahler order threshold to distinguish fresh vs. solute-rich wetlands). ..... 32

Fig. 3.6 Flowchart of wetland solute-richness prediction model methodology. ..... 35

Fig. 3.7 Boxplots of SDNWA wetland pond water EC observations measured between 1968 and 2013. Wetlands with more than five observations are shown. Adapted from Pennock et al. (2013). ..... 39

Fig. 3.8 Predicted spatial distributions of fresh and solute-rich wetlands based on Model F for portions of the Swift Current (left), St. Denis (center), and Smith Creek (right) study areas. ..... 46

Fig. 4.1 Schematic diagram of soil class distributions, depth to $CaCO_3$, groundwater table, and direction of groundwater and fill and spill flow based on diagrams from Van der Kamp and Hayashi (2009), Pennock (2011), and Pennock et al. (2014). Soil class distributions are indicated by the uppercase letters U = Upland, CW = Calcareous Wetland, T = Transition, and R = Recharge. ..... 54

Fig. 4.2 Sample point transect at a depression within the Swift Current study area. ..... 58

Fig. 4.3 Comparison of B horizon colour for a non-gleyed Calcareous Wetland soil (left) and the adjacent Upland soil (right). ..... 63

Fig. 4.4 3D representation of the topographic predictor variable *Elevation above basin bottom* for a portion of the St. Denis study area. The attribute reflects the meters above the associated depression's bottom..... 69

Fig. 4.5 3D representation of the topographic predictor variable *Wetland Strahler order maximum* for a portion of the St. Denis study area. The stream channel lines are not included in the final variable raster surface, only the values for each depression polygon. ..... 70

Fig. 4.6 3D representations of aerial imagery and 4-class map outputs from select best-performing models for example wetlands from each site..... 82

Fig. 4.7 3D representations of aerial imagery and 3-class map outputs from select best-performing models for example wetlands from each site. Calc. = Calcareous. ..... 83

Fig. 4.8 Predictor variable importance for the 4C_2mNoSm_OV_RF model plotted based on MDG. Predictor variables developed for this study are highlighted in gray. ..... 88

Fig. 4.9 Comparing surficial errors within the 2 m – not smoothed and 2 m – 5x smoothed DEM. a) and b) show 3D representations of the elevation surfaces for a wetland within the St. Denis DEM with a near-flat surface due to water presence during LiDAR collection. c) and d) show the resulting raster surfaces for the Convergence Index variable for the same wetland. ..... 91

Fig. C.1 Soil profile locations at Swift Current study area. Legal land descriptions are formatted: Quarter section – section – township – range – west of the 3rd meridian. ..... 147

Fig. C.2 Soil profile locations at the St. Denis study area. Legal land descriptions are formatted: Quarter section – section – township – range – west of the 3rd meridian. ..... 148

Fig. C.3 Soil profile locations at the Smith Creek study area. Legal land descriptions are formatted: Quarter section – section – township – range – west of the 1st meridian.... 149

## LIST OF ABBREVIATIONS

| | |
|---|---|
| $CaCO_3$ | Calcium carbonate |
| CSSC | The Canadian System of Soil Classification |
| ctree | Classification tree |
| DEM | Digital elevation model |
| DSM | Digital soil mapping |
| EC | Electrical conductivity |
| EM | Electromagnetic conductivity |
| Fe | Iron |
| GIS | Geographic information systems |
| GPS | Global Positioning System |
| HCl | Hydrochloric acid |
| LiDAR | Light Detection and Ranging |
| MDG | Mean decrease in Gini |
| MLR | Multinomial logistic regression |
| OV | Original topographic variables |
| PASCN | Predicted active spill channel network |
| PC | Principal component |
| PCA | Principal component analysis |
| PPR | Prairie Pothole Region |
| RF | Random forest |
| RMSE | Root mean squared error |
| SAGA | System for Automated Geoscientific Analysis |
| SOC | Soil organic carbon |
| S-Rich | Solute-rich |
| treebag | Classification tree with bagging |
| WDPM | Wetland DEM ponding model |
| 3C | 3-class mapping |
| 4C | 4-class mapping |

# 1 INTRODUCTION

## 1.1 Context

Wetlands of the Prairie Pothole Region (PPR) are important landscape features that limit agricultural phosphorus runoff contributions to downstream waterways. Wetland drainage has increased in portions of the PPR to reduce occurrences of flooding and increase arable land for cultivation (Brown et al., 2017a); however, drainage also reduces phosphorus retention potentials in PPR watersheds, which results in increased loading of downstream waterbodies (Blann et al., 2009; Badiou et al., 2018). This has major implications for Lake Winnipeg, which has already been severely impacted by eutrophication from phosphorus loading and is largely fed by PPR watersheds (Environment Canada, 2011).

Wetlands with soils enriched with calcium carbonates ($CaCO_3$) have a greater potential for reducing phosphorus mobility (Zhang et al., 2014; Brown et al., 2017b). Enrichment of $CaCO_3$ within PPR wetland soils is determined by the hydrologic characteristics of the wetland (Pennock et al., 2014). On the landscape scale, wetlands that a) receive greater runoff contributions, b) cannot contribute runoff downslope, and c) have a greater potential to receive groundwater discharge are more likely to be solute-rich and have soils enriched with $CaCO_3$ (Van der Kamp and Hayashi, 2009; Pennock et al., 2014). These hydrologic processes may be approximated through analysis of high-resolution digital elevation models (DEMs) to predict the spatial distributions of solute-rich wetlands. Successful predictions of solute-rich wetlands would allow for identification of wetlands expected to have $CaCO_3$-enriched soils, which could be prioritized for conservation to maximize the phosphorus retention ecosystem service within the PPR.

Calcium carbonates are not evenly distributed within PPR wetland soils. The distributions of calcareous soils (those enriched with $CaCO_3$) within wetlands are controlled by hydrologic processes occurring over landscape scales and within individual wetlands (Pennock et al., 2014). Wetlands can be classified as recharge, discharge, and flow-through, depending on their relationship with the groundwater, which causes different distributions of calcareous soils within each wetland type (Arndt and Richardson, 1988). Relationships between the spatial distribution

of calcareous wetland soils and a few key topographic attributes have been identified (Bedard-Haughn and Pennock, 2002; Pennock et al., 2014). However, wetland soil formation is highly complex (Pennock et al., 2014). Those complexities may be modelled through digital soil mapping methodologies that incorporate machine-learning techniques, which determine relationships between soil distributions and many environmental variables (McBratney et al., 2003). The combination of machine-learning modelling techniques and topographic attributes derived from high-resolution DEMs, that reflect individual wetland-scale and landscape-scale hydrologic characteristics, may be used to predict spatial distributions of $CaCO_3$-enriched wetland soils. Successful predictions of the extents of these soils within PPR landscapes would allow for upscaled estimates of phosphorus retention potentials for the PPR and could also potentially inform conservation efforts to target wetlands with greater potential to reduce phosphorus mobility.

This research builds on the Pennock et al. (2014) study *Application of hydropedology for predictive mapping of wetland soils in the Canadian Prairie Pothole Region*. That study provides the foundation for many of the concepts used in the models of this study and is recommended as supplemental reading.

### 1.2 Research objectives

The first objective of this research was to develop a model to predict spatial distributions of fresh and solute-rich wetlands within PPR watersheds by approximating hydrologic processes using high-resolution DEMs. The second objective of this research was to assess digital soil mapping methodologies for predicting the spatial distributions of wetland soil types in PPR landscapes using high-resolution DEMs.

### 1.3 Organization of thesis

This thesis is presented in the manuscript-style format. The first two chapters provide a general introduction (Ch. 1) and review of relevant literature (Ch. 2); the next two chapters (Ch. 3 and 4) present the main body of the research in manuscript format. The first research objective, focused on wetland type, is addressed in Chapter 3: *Predictive mapping of solute-rich wetlands in the Canadian Prairie Pothole Region through high-resolution digital elevation model analyses.*

Chapter 4 addresses the second objective: *Predictive digital soil mapping of wetland soil types in the Canadian Prairie Pothole Region*. Chapter 5 summarizes and synthesizes the two research chapters, suggests how they can be applied, and proposes how they could be improved upon. Chapter 6 provides a comprehensive reference list for the entire thesis. There are multiple appendices included in this thesis: Appendix A describes the GIS methodologies used in the wetland solute-richness class prediction model proposed in Chapter 3, Appendix B describes GIS methodologies used to develop new knowledge-based topographic attributes used in the digital soil mapping study described in Chapter 4, and Appendix C provides maps of the soil sample locations and directs the reader to the soil profile descriptions.

# 2 LITERATURE REVIEW

## 2.1 Eutrophication of Lake Winnipeg

Lake Winnipeg is one of the largest lakes in the world (roughly 23,750 $km^2$) (Environment Canada, 2011). It is considered a crucial resource for Manitoba's tourism, fishing, and hydroelectric industries and is home to over 23,000 permanent residents, including Métis and First Nation communities (Environment Canada, 2011). Lake Winnipeg experienced a 71% increase in phosphorus inputs and an 18% increase in nitrogen inputs between 1994 and 2007 (Schindler et al., 2012). The nutrient increases have caused large-scale increases in phytoplankton and shifts in phytoplankton populations to the dominance of cyanobacteria (greater than 90%), commonly known as blue-green algae (Kling et al., 2011). The more frequent and intense algal blooms have negative effects on the water quality and ecological health of the lake and have negative impacts on its socio-economic uses (Environment Canada, 2011).

Lake Winnipeg's watershed is Canada's second largest watershed at 1,000,000 $km^2$ in size (Environment Canada, 2011). It spans four provinces and portions of the U.S. (Fig. 2.1). Nutrient loading is derived from both point and non-point sources within the watershed, with agricultural activity providing a significant source of nutrients (Schindler et al., 2012). Increased numbers of livestock operations and increased use of synthetic fertilizers in this region over the last century have contributed excessive phosphorus and nitrogen loads to the waterways within the watershed that eventually feed into Lake Winnipeg (Jones and Armstrong, 2001). Effective management strategies to address the various nutrient sources are needed to control and reduce loading affecting Lake Winnipeg (Environment Canada, 2011). The effects of reducing nitrogen inputs for controlling eutrophication are debateable, but there is no doubt that phosphorus inputs need to be reduced to effectively control eutrophication (Schindler, 2012).

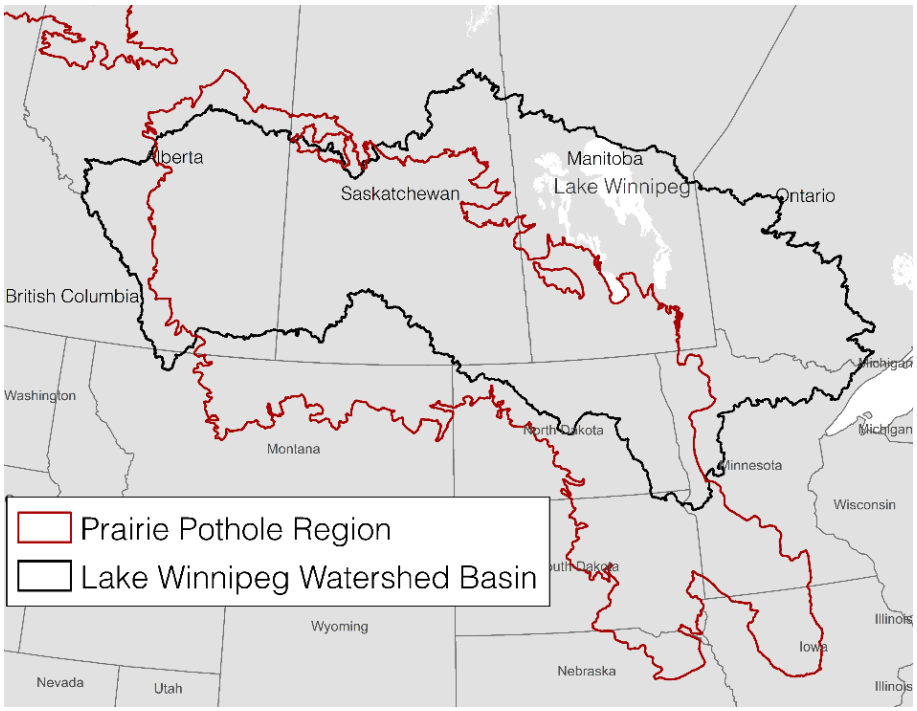



**Fig. 2.1 Approximate extents of the Prairie Pothole Region and the Lake Winnipeg Watershed Basin.**

## 2.2 Prairie Pothole Region

There is substantial overlap between the Lake Winnipeg watershed basin with the Prairie Pothole Region (PPR) (Fig. 2.1). The PPR represents a unique prairie landscape that is characterized by an abundance of shallow wetlands. The undulating to hummocky landscape was formed by uneven depositions of glacial till (Pennock et al., 2010) left behind from the most recent glaciation, the Wisconsin glaciation, which retreated fully from the prairies approximately 11,000 years ago (Klassen, 1989). The till parent materials include finer silts and clays with a low hydraulic conductivity, especially at depths 5 – 10 m below the surface (Miller et al., 1985; Hayashi et al., 1998a); precipitation ponds within low-conductivity depressions to form wetlands (Bedard-Haughn and Pennock, 2002). Since deglaciation, the region has mostly experienced arid and sub-humid climates, which have not provided the quantity of precipitation necessary to generate fluvially-eroded integrated drainage networks (rivers and streams) in much of the landscape (Shook et al., 2013). The wetlands are typically hydrologically isolated, except during wetter periods where water is passed through the wetland systems via shallow groundwater and surficial fill-and-spill flow (Van der Kamp and Hayashi, 2009). Phosphorus mobility is largely through surficial runoff because movement through groundwater is typically limited in the PPR

(Hayashi and Rosenberry, 2002); therefore, under natural conditions, much of the phosphorus does not easily move through these wetland complexes to reach the rivers that eventually feed into Lake Winnipeg. However, wetland drainage converts this system to more closely resemble an integrated drainage system where surface water (and phosphorus in runoff water) can be more readily passed into downstream waterways (Blann et al., 2009).

### 2.3 PPR wetlands and nutrient retention

Wetlands have been found to act as sinks for phosphorus and nitrogen in runoff through physical and chemical mechanisms (Johnston, 1991). On watershed scales, wetlands physically control nutrient runoff by increasing surface storage and reducing effective contributing areas (Blann et al., 2009). At the individual wetland scale, the ponds and associated riparian vegetation physically slow the movement of runoff causing sedimentation of nutrients (Mitsch et al., 1995). Badiou et al. (2018) found intact PPR wetlands to have greater potential for phosphorus retention compared with drained PPR wetlands. Wetland drainage has become more commonly implemented in the PPR, especially in the northern and eastern parts where wetlands have exacerbated flooding issues and reduced arable cropland (Brown et al., 2017a). This has been especially true since 2010 because the PPR has experienced an unusually wet period (Brown et al., 2017a). Further wetland drainage within the PPR would result in increased phosphorus loading of the prairie watersheds and would perpetuate eutrophication issues for Lake Winnipeg.

Certain PPR wetland soils may have a greater potential for phosphorus retention due to their chemical characteristics. Soils enriched with calcium carbonates ($CaCO_3$) react with the mobile phosphorus to form less soluble Ca-phosphates (Zhang et al., 2014). This reduces the mobility of the phosphorus and prevents it from moving into downstream waterbodies. This finding was also observed in the Rahman et al. (2014) study, where phosphorus content within runoff was reduced in soil treated with $CaCO_3$. In a study comparing phosphorus content between calcareous and non-calcareous surficial PPR wetland soils, Brown et al. (2017b) found total phosphorus was not significantly different between the soil types, but available phosphorus was six times greater within the non-calcareous wetland soil. Therefore, the non-calcareous wetland soil had a greater potential to contribute phosphorus downstream.

## 2.4 PPR wetland hydrology and hydrochemistry

The geochemical makeup of the PPR glacial deposits reflects the characteristics of the bedrock and other materials that they were sourced from (Pennock et al., 2011). Across the PPR, the glacial deposits are rich with calcium carbonates (St. Arnaud, 1976), sourced from limestone outcroppings located just to the south of the Canadian Shield in Manitoba and Saskatchewan (Pennock et al., 2011). The glaciers advanced from Hudson Bay and moved south-west across the prairies (Klassen, 1989), redistributing $CaCO_3$ from the limestone outcroppings throughout the region. The glacial deposits closer to the original limestone outcroppings are more enriched with $CaCO_3$ (Pennock et al., 2011) and the soils of these regions have remained enriched with $CaCO_3$ even after thousands of years of weathering (Saskatchewan Soil Survey Staff, 1991). Solutes, including $CaCO_3$, within the parent materials have been redistributed through the various wetland-dominated hydrologic processes occurring in the PPR.

PPR wetlands are found within topographically closed basins and are often hydrologically isolated due to the low hydraulic conductivity of the glacial deposits, especially during dry periods (Van der Kamp and Hayashi, 2009). Snow is redistributed by wind to form snow drifts within the wetland depressions, which increases spring water levels (Fang and Pomeroy, 2008). Wetlands also receive much of the snowmelt as surficial runoff due to the reduced infiltration through frozen soils, although a portion of snowmelt does infiltrate the soils (Woo and Rowsell, 1993). Throughout the growing season, wetlands receive precipitation through direct rainfall and lose water through evaporation. Due to the low hydraulic conductivity of the deeper glacial tills, hydrologic losses and additions through deep groundwater movement are typically very slow and are not considered to significantly affect wetland hydroperiods, unless there is direct interaction with coarse-textured intertill aquifers (Van der Kamp and Hayashi, 2009). The hydraulic conductivity of the glacial tills increases exponentially near the surface due to weathering (Miller et al., 1985; Hayashi et al., 1998b). Water within the wetland soil directly below the pond is conducted laterally outward from the wetland through the near-surface materials due to the higher hydraulic conductivity compared to the deeper tills (Hayashi et al., 1998a; Heagle et al., 2013). This water can then move upwards to the soil surface at the wetland fringe through capillary rise (Knuteson et al., 1989; Hayashi et al., 1998a). Evaporation and evapotranspiration from vegetation in the wetland fringe contribute to this phenomenon (Hayashi et al., 1998a). This

has a significant effect on a wetland's water balance; Millar (1971) determined that the recession rate of a wetland was strongly related to the wetland's shoreline to pond-area ratio due to this effect.

Runoff can move from adjacent uplands to the wetlands and can be transferred between wetlands via shallow groundwater and overland flow (Van der Kamp and Hayashi, 2009). These have been recognized as a threshold-mediated processes that occur once sufficient water levels have been achieved (Spence, 2010). The water table typically sits within the deeper, low-conductivity tills resulting in limited groundwater flow. During wetter periods, the water table will rise and encompass the higher hydraulically conductive near-surface tills which allows the groundwater to move; this is referred to as activation of the effective transmission zone (Brannen et al., 2015). Overland flow can occur when wetlands fill beyond their holding capacity and spill into adjacent wetlands; this is referred to as fill-and-spill flow (Cook and Hauer, 2007). Millar (1976) developed wetland classifications based on the runoff (fill-and-spill and shallow groundwater) contributions they receive: 1) isolated wetlands which do not receive nor contribute runoff to and from adjacent wetlands, 2) overflow wetlands which do not receive runoff contributions but can contribute them downslope, 3) channel wetlands which receive runoff contributions from upslope wetlands and contribute runoff to wetlands downslope, and 4) terminal wetlands which receive runoff contributions but cannot contribute them downslope.

Solutes present in the glacial deposits of the PPR have been redistributed via these hydrologic mechanisms. The movement of $CaCO_3$ is very slow due to its low solubility (Knuteson et al., 1989). Solutes are transferred with the lateral and upwards movement of water to the wetland fringes (Knuteson et al., 1989; Hayashi et al., 1998b; Heagle et al., 2013). Enrichment with $CaCO_3$ in the soils of the wetland fringe is a common feature observed in the PPR and is referred to as the wetland discharge ring (Pennock et al., 2014). This phenomenon was studied in North Dakota on glaciolacustrine parent materials by Knuteson et al. (1989) who determined that upward transfers of $CaCO_3$ equivalent occurred at a rate between 0.1 to 0.2 mol $m^{-2}$ $yr^{-1}$. During periods of intense rainfall, the water table within the hillslopes adjacent to the wetlands can move into the effective transmission zone and flow will reverse back downslope into the wetlands

(Winter and Rosenberry, 1998) causing a limited amount of the solutes to be redistributed to the wetland pond (Berthold et al., 2004).

Solutes are transferred from wetlands in topographically high positions to those in lower positions through fill-and-spill, shallow and deep groundwater flow (Van der Kamp and Hayashi, 2009). Shallow groundwater and fill-and-spill contributions cause increases in solute loads in lower-lying wetlands (Cook and Hauer, 2007; Nachshon et al., 2013). The classifications defined by Millar (1976) for a wetland's potential to receive and contribute runoff are important concepts for wetland solute accumulation as isolated and overflow wetlands are less likely to accumulate solutes and terminal wetlands are most likely to accumulate solutes because they cannot readily be removed via runoff (Nachshon et al., 2013).

Although deep groundwater flow movement is typically very limited, it is important for determining if wetlands are dominantly recharge or discharge wetlands (Van der Kamp and Hayashi, 2009). Recharge wetlands recharge water to groundwater and discharge wetlands receive groundwater discharge (Lissey, 1971). With only minimal deep groundwater movement, over thousands of years, the solute additions and removals via these processes are important controls on wetland chemistry (Van der Kamp and Hayashi, 2009). Solutes in the sediments within recharge wetlands are slowly leached out through this process. Solutes are moved into the discharge wetlands through groundwater movement. These processes occur at much faster rates for wetlands that interact directly with coarse-textured intertill aquifers (Zebarth et al., 1989). Discharge wetlands are found in low-lying landscape positions where the wetland is located beneath or level with the surrounding water table (Lissey, 1971). Recharge wetlands are situated at an elevation above the surrounding water table. The relationships between wetlands and groundwater can change with changes in hydrologic regimes. Flow-through wetlands recharge the groundwater and receive groundwater discharge depending on the water table levels (Arndt and Richardson, 1988). Flow-through wetlands represent a midpoint between recharge and discharge wetlands in terms of solute accumulations (Arndt and Richardson, 1989).

## 2.5 PPR soil distributions

The PPR spans the Brown, Dark Brown, Black, and Dark Gray soil zones, which reflect general climatic conditions; warmer and drier in the southwest (Brown) and cooler and wetter in the northeast (Black and Dark Gray) (Pennock et al., 2011). The colour reflects the organic carbon storage in the surface horizon due to differences in available moisture (Pennock et al., 2011). The soil zones correspond to the great groups of the Chernozemic soil order, which is the dominant soil of the prairies.

A common catenary relationship is observed in the PPR. The upland landscape positions are dominated by Chernozems and Regosols. Thin Chernozems (often with Rego and Calcareous sub-groups) and Regosols form on the eroded shoulder and summit positions of hillslopes (Pennock et al., 2011). This distribution is more prominent in cultivated landscapes where tillage erosion in these positions can be substantial (Pennock, 2003). Further down the hillslope, in the mid-slope position, thicker Orthic Chernozems commonly form (Pennock et al., 2011). Eluviated and Gleyed Chernozems often form in foot-slope positions. Gleysols form in the toe-slope and depressional positions due to the redistribution of moisture to these positions (Bedard-Haughn and Pennock, 2002).

Gleysols in the PPR typically reflect wetland soils. These form in positions where excessive moisture causes anaerobic conditions. The anaerobic conditions allow for oxidation-reduction processes to transform iron and manganese (Bedard-Haughn, 2011). During anaerobic conditions, $Fe^{3+}$ is used as an electron acceptor by bacteria causing it to be reduced to $Fe^{2+}$ (Bedard-Haughn, 2011). The reduced iron is more mobile and can be leached from the profile leading to a dull grey or bluish colour. When the soil dries out, the iron can be oxidized and form pockets of reddish mottles (Bedard-Haughn, 2011). The excessive moisture in wetland soils also reduces microbial activity. This slows the decomposition of organic material and results in thick layers of soil organic carbon in the upper horizons (Pennock et al., 2014). Wetlands of the PPR, however, do not typically form peat deposits due to periodic drought conditions which allow for decomposition of the organic material (Pennock et al., 2014).

Wetland soils differ in terms of their accumulation of $CaCO_3$, which is largely controlled by the soil's relationship with groundwater. As groundwater discharges into discharge wetlands, $CaCO_3$ is deposited throughout the soils of the wetland (Pennock et al., 2014). With the dominant downward movement of water within recharge wetlands, the $CaCO_3$ can be leached out from the soils directly beneath the wetland pond (Pennock et al., 2014). Flow-through wetlands can experience fluxes in the direction of groundwater in their soils. They accumulate $CaCO_3$ with groundwater discharge; however, intermittent periods of downward water movement cause the concentrations of solutes to be lower than in discharge wetlands (Arndt and Richardson, 1989). The relationship between deep groundwater movement and a wetland is an important control on the distribution of $CaCO_3$ within its soil, but wetlands that receive significant solute contributions through fill and spill and shallow groundwater flow can accumulate $CaCO_3$ within the soils throughout the wetland basin regardless of whether the wetland is recharge or discharge (Cook and Hauer, 2007; Pennock et al., 2014).

As discussed in section 2.4, the upward and lateral movement of groundwater to the wetland fringe causes accumulations of the $CaCO_3$ in a discharge ring pattern. This phenomenon is observed in all wetland types: recharge, discharge, and flow-through. These soils are enriched with $CaCO_3$ throughout their full profile. The upward movement of water in these profiles often inhibits the development of a B horizon (Pennock et al., 2014). The discharge ring may or may not be gleyed. Much of the upward and lateral movement of water to the discharge ring occurs through unsaturated flow (Knuteson et al., 1989) which, alone, would not provide the anaerobic conditions necessary to form gleyed soil characteristics. Soil gleying within the discharge ring likely forms during wetter periods when the positions were saturated with wetland water. Gleying in these positions also indicates that the dominant movement of water within these profiles likely changed over time because gleyed soils are expected to develop with downward movement of water to cause iron to be leached from the profile (Bedard-Haughn and Pennock, 2002).

Wetlands are often defined in terms of their vegetation and the permanence of pond surface water (Stewart and Kantrud, 1971; Millar, 1976). However, water levels can vary significantly in this region year to year and within a season (Winter and Rosenberry, 1998). Vegetative

characteristics will shift in response to varying water levels and are difficult to use to reflect wetland extents in cultivated areas because they can be completely removed (Millar, 1976). Wetland soils can be used to indicate the fixed extents of wetlands. Soil gleying is an irreversible process (Bedard-Haughn, 2011) and the additions and removals of $CaCO_3$ have required thousands of years to form (Knuteson et al., 1989); these characteristics reflect a more permanent stamp of the wetland on the landscape. Pennock et al. (2014) found that the boundary between wetland recharge soils and the discharge ring closely matched the maximum water level observed in the last half century. Wetland soil characteristics therefore indicate historic wetland extents.

### 2.6 Hydrologic modelling in the PPR using digital elevation models

High-resolution digital elevation models (DEMs) have allowed for detailed modelling of surficial hydrology in the PPR. A DEM is a grid-based raster surface in which every grid cell has an elevation value. The spatial resolution refers to the width of the individual cells. In a higher resolution DEM, there are more grid cells per area, providing a more detailed representation of the earth's surface. Currently, 1-m DEMs exist only for select areas of Canada (Natural Resources Canada, 2018) and the rest of Canada is covered with 30 – 90 m DEMs (Natural Resources Canada, 2016). The latter are too coarse in resolution to represent the small-scale topographic variation of the PPR, which is characterized by countless wetlands, many of which are only tens of meters wide (Li et al., 2011). Higher-resolution DEMs (< 10 m) provide detailed representations of the earth's surface which can be used to model small-scale hydrologic characteristics of a PPR landscape (Li et al., 2011). High-resolution DEMs can be generated from Light Detection and Ranging (LiDAR) remote sensed imagery which uses a pulsed laser to collect elevation information at densely-spaced points along the earth's surface. Most modern LiDAR methods capture many points per square meter which are interpolated to create DEMs, usually with resolutions less than 5 m.

One issue with high-resolution DEMs is that there can be many errors within the surface, creating artefactual features like depressions (Lindsay and Creed, 2006). For hydrologic modelling within landscapes with more integrated drainage systems, all depressions are assumed to be errors and are removed to allow for flow routing (Lindsay and Creed, 2005). However, this

practice is not appropriate for much of the PPR (Li et al., 2011) because depressions are common landscape features that define the region's hydrology (Van der Kamp and Hayashi, 2009). To remove artefactual depressions from the DEM surface, Li et al. (2011) recommend smoothing a 1-m DEM 10 to 20 times and removing any depressions with areas less than 200 $m^2$ or depths less than 10 cm.

Numerous methods have been developed to model surficial hydrology in PPR landscapes using high-resolution DEMs, all of which account for the influence of depressional storage. Wu and Lane (2016) developed a method for delineating wetland depressions and quantifying the hierarchical relationships between nested wetlands using a contour-tree approach. Shaw et al. (2013) developed a method for determining effective contributing areas within PPR watersheds that considers the storage capacity of the wetland depressions. Yang and Chu (2015) developed a similar hydrological model for the PPR that considers the hierarchical relationships of nested wetlands, but also accounts for spatio-temporally varied rainfall and infiltration rates in heterogeneous soil. Shook et al.'s (2014) Wetland DEM Ponding Model (WDPM) was developed to model distributions of runoff within PPR landscapes, which also considers the storage capacity of the wetland depressions. The advantage of the WDPM is that it uses Shapiro and Westervelt's (1992) algorithm for redistributing water on the landscape as oppose to the D8 direction of drainage algorithm commonly used in other models (used in Shaw et al. (2013) and Yang and Chu (2015)). This algorithm allows for water to move in multiple directions from a single cell, whereas the D8 method determines one direction of water movement for a single cell. This iterative algorithm simulates the movement of water on the landscape; if water fills a depressions, additional water is able to spill over the depression (Shook et al., 2014). These models were largely developed for water budget purposes; no methods specifically model wetland solute accumulations influenced via hydrological processes. Successful modelling of solute distributions would allow for identifying wetlands with greater phosphorus retention potential. Wetland solute distribution mapping could serve other purposes as wetland salinity has important influences on biological, ecological, and agronomic functions.

## 2.7 Digital soil mapping in the PPR

The hydrological modelling approaches described typically model processes occurring over the greater landscape and watershed. These approaches are appropriate for determining characteristics for larger hydrological units, like wetlands, but are less appropriate for determining individual soil pedon characteristics, which are influenced by various processes, hydrologic and otherwise, occurring on much smaller scales. Digital soil mapping approaches can consider a multitude of influences, including hydrologic ones, occurring over various scales, which makes them more appropriate for predicting individual soil characteristics.

In the past 30 years, there have been considerable advances in the field of digital soil mapping (DSM) (McBratney et al., 2003). This has resulted from increased access to spatially continuous environmental information and the consistent and rapid advancements in computational capabilities. DSM involves mapping either soil properties or types based on relationships established between the target soil characteristic and the available environmental variables in a region. McBratney et al. (2003) propose the *scorpan* framework for DSM which builds on Jenny's (1941) conceptual model that soil formation is the function of the soil forming factors: climate, organisms, relief, parent material, and time. The *scorpan* model also includes: 1) the soil – which refers to information on a soil at a point either previously measured or for another soil property and 2) space – the geographic position of the soil. The environmental variables used to predict the spatial distributions of soil properties or types can reflect a single or multiple soil forming factors that are most pertinent for predicting the target soil characteristic. The environmental variables often include terrain attributes derived from DEMs, spectral reflectance bands from satellite imagery, and soil survey information. The *scorpan* framework can be applied to predict wetland soil distributions within the PPR. However, there are challenges for doing so because the typical information sources used in DSM, like soil survey and aerial imagery, do not provide much useful information for predicting the distributions of wetland soil types in this region.

Legacy soil surveys of the Canadian PPR do not provide information necessary to map wetland soil types in the PPR. Surveys are typically at resolutions between 1:50,000 to 1:250,000 (Agriculture and Agri-Food Canada, 2017). The soil surveys map large areas with a single

polygon that describes the soil series expected within the polygon: the dominant and commonly-found soil orders and a description of the parent material. Wetlands are not mapped specifically, instead the presence of Gleysols within a polygon will be indicated within the soil series, which includes an estimate of the proportion of the polygon expected to have Gleysols; there is no indication of where within the polygons the Gleysols are found.

Aerial imagery exists for most of the PPR and can provide information on spatial extents of wetland pond water. This approach has been used to create an inventory of wetlands across the PPR (Canadian Wetland Inventory Technical Committee, 2016). Pennock et al. (2014) found that distributions of wetland recharge and discharge ring soils corresponded to the maximum observed water levels, but this information is not available at most locations. Soil distance to wetland water has been found to be a key control on distributions of wetland soils (Bedard-Haughn and Pennock, 2002; Murphy et al., 2009). However, due to variability of wetland pond water extents, it would be difficult to create standardized measures to relate the extents within imagery to the spatial distribution of wetland soil types (Bedard-Haughn and Pennock, 2002).

High-resolution topographic information is likely more useful than legacy soil data or air photos for mapping detailed soil distributions in this region. Currently, high-resolution DEMs exist for portions of the PPR. Clear relationships between topographic attributes derived from high-resolution DEMs and spatial distributions of soils within upland positions in PPR landscapes have been established. Pennock et al. (1987) identified landform elements based on plan and profile curvatures and hillslope gradient that could be used to predict soil A horizon depth and depth to $CaCO_3$. Similarly, Manning et al. (2001) found relationships between landform elements and other soil properties including organic carbon, pH, and solum thickness in an undulating Manitoba landscape. Florinsky et al. (2002) related soil properties to various topographic attributes including aspect, curvatures, specific catchment area, and topographic index. These attributes reflect the soil hydrology at the hillslope scale that largely controls the common catenary relationship observed in much of the region (Thompson et al., 1997).

Depressional positions can be more challenging for characterizing in terms of topographic attributes, for example, depressions are often flat and therefore cannot be defined by their

curvature, which complicates differentiating them from flat hilltop positions (Pennock et al., 1987). Other topographic attributes have been found to relate specifically to distributions of PPR wetland soil types in depressional positions. Bedard-Haughn and Pennock (2002) found that gleyed recharge soils occurred in positions with specific dispersal areas less than 2 $m^2$ $m^{-1}$ and that non-gleyed recharge and discharge soils occurred in positions with specific dispersal areas greater than 2 $m^2$ $m^{-1}$. Pennock et al. (2014) found the distributions of wetland recharge and discharge soils were related to elevation above wetland basin bottom. There have been many other DEM-derived topographic attributes developed which quantify all types of topographic characteristics like wetness indices (Beven and Kirkby, 1979), terrain roughness indices (Riley et al., 1999), curvature derivatives (Zevenbergen and Thorne, 1987), multi-resolution valley bottom flatness index (Gallant and Dowling, 2003), to name a few. These attributes may provide useful information for predicting spatial distributions of PPR wetland soil types.

Spatial distributions of wetland soil types in the PPR may relate to many topographic attributes and so modelling approaches that can account for many variables may be most applicable for this objective. Digital soil mapping methodologies often incorporate machine learning techniques to use as predictive models (McBratney et al., 2003). Machine-learning involves data mining where models will determine patterns and relationships between large numbers of predictor variables and the target variable; these relationships can then be used to make predictions for new data (Witten and Frank, 2005). Soil formation is a complex process involving many factors (Jenny, 1941); machine-learning approaches have proven to be useful in DSM because they allow for the consideration of many variables that can reflect the effects and interactions of the forming factors on soil characteristics (McBratney et al., 2003).

Many machine-learning approaches have been used in DSM for mapping soil classes. The most popular among these are tree-based learners which include classification trees, classification trees with bagging, and random forest models (Heung et al., 2016). These involve the development of one or thousands of decision trees which are based on if-then rules to differentiate soil class observations based on their predictor variable characteristics (Strobl et al., 2009). These model types are useful because they can reflect non-smooth and non-linear relationships between the target classes and predictor variables (Strobl et al., 2009). Due to their

hierarchical structure, these models can reflect effects and interactions between the predictor variables, where the importance of a predictor variable can be dependent on another predictor variable value (Heung et al., 2016). As an example, surface flatness could reflect the bottom of a depression or hill summit within a PPR landscape; these features could be distinguished with a second predictor variable that quantifies their position relative to the midpoint of the hillslope. The tree-based learners would be able to reflect these types of relationships in their decision trees.

Linear regression models are commonly used in DSM to map soil properties (McBratney et al., 2003). A type of generalized linear model, multinomial logistic regression, can be used instead to map soil classes (Heung et al., 2016). Multinomial logistic regression is an extension of binomial logistic regression which models the probability of soil class occurrence between 0 and 1 based on relationships with predictor variables. This model is also frequently used in DSM studies to map soil classes (Kempen et al., 2009; Debella-Gilo and Etzelmüller, 2009; Heung et al., 2016).

As mentioned, the distribution of wetland soils of the PPR have been related to key topographic attributes derived from high-resolution DEMs (Bedard-Haughn and Pennock, 2002; Pennock et al., 2014). However, no studies to date have attempted to predictively map spatial distributions of wetland soil types in the region using machine-learning DSM approaches in combination with topographic attributes derived from high-resolution DEMs as predictor variables. This approach could potentially allow for highly detailed mapping of these soils and improve our understanding of the distributions of them. The need to target and quantify the extents of phosphorus retaining wetland soils provides the impetus to apply these methods to map wetland soil types.

# 3 PREDICTIVE MAPPING OF SOLUTE-RICH WETLANDS IN THE CANADIAN PRAIRIE POTHOLE REGION THROUGH HIGH-RESOLUTION DIGITAL ELEVATION MODEL ANALYSES

## 3.1 Preface

Solute-rich wetlands are expected to have soils enriched with $CaCO_3$, meaning they have a greater potential for phosphorus retention. Solute-rich wetlands develop from the influence of hydrologic processes that redistribute solutes present in the parent material. There have been several studies that used high-resolution DEMs to model hydrologic processes within the PPR, but none have attempted to model solute accumulations that result from those hydrologic processes. The model proposed in this chapter incorporated ideas used in the preceding PPR hydrologic models and applied them to predict binary wetland solute-richness classes. Wetland solute-richness classes were defined by electrical conductance measures of the pond water and wetland soil; wetland soil $CaCO_3$ content measurements were not considered in the training or testing of this model. The model made predictions for individual wetlands; there was no attempt to predict where the solutes accumulate within the wetlands. Wetlands predicted to be solute-rich by the model proposed in this chapter would be recommended to be preferentially conserved due to their greater potential for phosphorus retention.

### 3.2 Abstract

Wetlands of the Prairie Pothole Region (PPR) play an important role in reducing nutrient mobility in prairie watersheds. Wetlands with strongly calcareous soils have a greater potential for phosphorus retention than non-calcareous wetland soils. Accurate predictions of the spatial distributions of wetlands with $CaCO_3$-enriched soils would allow for prioritized wetland conservation efforts to encourage this ecosystem service. Accumulations of $CaCO_3$ and other solutes are largely determined by the wetland's topographic position and its relationship with groundwater. A model for estimating these characteristics was developed and tested using LiDAR-derived digital elevation model analyses and measurements of pond water and soil electrical conductivity (i.e., solute-richness) from three study areas within the PPR, near Swift Current, SK; St. Denis, SK; and Smith Creek, SK. Spill channel connections between wetlands were predicted to characterize wetlands in terms of the potential shallow groundwater flow and fill-and-spill flow contributing to them, their potential for contributing flow downslope, and their relative position within the local landscape and greater watershed. Wetlands were then ascribed Strahler orders and determined to be terminal or not based on the predicted spill channel networks. Spatial distributions of solute-rich wetlands were predicted with a simple decision tree model that predicts wetlands as either fresh or solute-rich based on their Strahler order and terminal status. The model achieved total predictive accuracies between 69 and 82% based on training and external validation tests. The model showed potential to be used to map the distribution of solute-rich wetlands for other applications, i.e. salinity risks to agronomic productivity. Wetland Strahler order and terminal status variables could be incorporated into more complex multi-variate models for the purposes of digital soil mapping or hydrological studies in the Prairie Pothole Region.

### 3.3 Introduction

Agricultural intensification and wetland drainage across the Prairie Pothole Region (PPR) have contributed to phosphorus loading of the watersheds that ultimately feed into Lake Winnipeg, resulting in its eutrophication. Among the many ecosystem services that wetlands of the PPR provide, they reduce phosphorus mobility within the prairie watersheds (Badiou et al., 2018). Certain wetland soils (those enriched with calcium carbonates ($CaCO_3$)) have a greater potential to retain phosphorus. Brown et al. (2017b) found total phosphorus amounts in calcareous and non-calcareous PPR wetland soils to be similar, but available phosphorus was found to be significantly higher in non-calcareous wetland soils compared to calcareous wetland soils. Calcium reacts with phosphorus to form less soluble Ca-phosphates (Zhang et al., 2014) and thereby reduces the available phosphorus within the watershed.

Calcium carbonate accumulation within wetlands is controlled by several hydrologic processes. During drier periods, PPR wetlands are predominantly hydrologically isolated from each other due to the low permeability of the deeper glacial tills, unless they are connected to coarse-textured intertill aquifers (Van der Kamp and Hayashi, 2009). During wetter periods, connectivity between wetlands is increased due to increased fill-and-spill flow and shallow groundwater flow (Cook and Hauer, 2007; Brannen et al., 2015). Fill-and-spill flow occurs once wetlands fill to their maximum water holding capacity and contribute overland flow downslope. The glacial tills nearer to the surface have exponentially greater hydraulic conductivities than the deeper tills due to greater weathering (Miller et al., 1985; Hayashi et al., 1998b). Shallow groundwater flow occurs when the water table is near to the surface and the groundwater can move through the hydraulically conductive near-surface tills (Brannen et al., 2015). Water moving through fill-and-spill and shallow groundwater flow redistributes solutes present in the parent material to wetlands in lower-lying positions. The parent materials of the PPR are rich with $CaCO_3$ (St. Arnaud, 1976), which is redistributed through these processes. Wetlands that receive greater contributions of shallow groundwater and fill-and-spill flow are expected to have greater accumulations of solutes. Certain wetlands may receive substantial solute contributions and, due to their topographic position or morphological characteristics, do not contribute fill-and-spill or shallow groundwater flow further downslope; these are referred to as terminal wetlands

(Millar, 1976). Terminal wetlands have a greater potential to accumulate solutes because solutes are not easily leached from them.

Although deep groundwater movement in glacial till is very slow and has little influence on wetland hydroperiods (Van der Kamp and Hayashi, 2009), the relationship between wetlands and deep groundwater movement is important for determining if a wetland is dominantly recharge or discharge (Lissey, 1971). This relationship determines if the limited movement of deep groundwater has dominantly caused additions or removals of solutes within a wetland over time. Recharge wetlands recharge water to the groundwater, whereby solutes are slowly leached from the wetlands into the groundwater. Groundwater discharges solutes to discharge wetlands. Solutes cannot be readily leached from discharge wetlands, causing them to accumulate. Discharge wetlands occur where wetlands are situated at an elevation below or level with the surrounding water table. They are less common than recharge wetlands; they only occur in low-lying landscape positions (Lissey, 1971) and are often terminal wetlands (Van der Kamp and Hayashi, 2009). A wetland's relationship with the groundwater can change with changes in groundwater levels (Winter and Rosenberry, 1998); flow-through wetlands can recharge the groundwater and receive groundwater discharge depending on groundwater levels. Flow-through wetlands fall on the spectrum between recharge and discharge wetlands in terms of solute accumulation (Arndt and Richardson, 1989).

The solute-rich wetlands that develop from the described hydrologic processes (increased fill-and-spill, shallow groundwater flow, and deep groundwater discharge contributions) are expected to have greater potential for phosphorus retention due to the increased concentrations of $CaCO_3$ in their soils. As these hydrologic mechanisms for wetland solute accumulation are largely determined by the wetland's topographic position, spatial distributions of solute-rich wetlands may be predictable through digital elevation model (DEM) analyses. Pennock et al. (2014) found all solute-rich wetlands (those with electrical conductivity greater than 1000 $\mu S\ cm^{-1}$) were below a certain elevation at the St. Denis National Wildlife Area (SDNWA). These wetlands received greater shallow groundwater and fill-and-spill flow contributions and had greater potential to be discharge wetlands. This information is useful for understanding the wetlands of the SDNWA, but the absolute elevations are not applicable for other areas of the

PPR, or even other portions of the St. Denis watershed. There is a need to develop a methodology that interprets more general topographic characteristics to predict spatial distributions of solute-rich wetlands within PPR landscapes. This study proposes a method for predicting the spatial distribution of fresh and solute-rich wetlands by approximating the hydrological processes influencing wetland solute accumulation through analyses of high-resolution Light Detection and Ranging (LiDAR)-derived DEMs. Successful predictions would enable prioritized wetland conservation and restoration efforts to preserve and encourage phosphorus retention within the PPR.

## 3.4 Materials and methods

### *3.4.1 Study areas*

Wetlands were sampled and tested for water and soil salinity at three study areas across the Saskatchewan portion of the PPR to assess model predictions of solute-richness classes. The study areas included sites near Swift Current, SK; St. Denis, SK; and Smith Creek, SK. The areas were selected because high-resolution LiDAR-derived 1-m DEMs exist for their entire watersheds. The Swift Current study area was located 20 km east of Swift Current, SK; the St. Denis study area was located 40 km east of Saskatoon, SK; and the Smith Creek study area was located 60 km south-east of Yorkton, SK (Fig. 3.1). Multiple sites were sampled within each watershed. The sites were selected to have undulating to hummocky topography and glacial till parent material to be representative of the PPR. The sites included cultivated, pasture, and native land uses. The study areas were also selected based on soil survey salinity data. Sites with salinity ratings of *moderate* to *severe* (effects on productivity) were avoided. The study areas span the climate gradient of the PPR. Swift Current has the warmest and driest climate considering potential evaporation and Smith Creek has the coolest and wettest climate. General site characteristics are described in Table 3.1.

The Swift Current study sites are found on either medium textured glacial till or moderately fine textured glacial till with thin deposits of moderately coarse textured fluvial or fluvio-lacustrine deposits on top (Ayres et al., 1985). The glacial deposits of the Swift Current study area are relatively thin (between 2 to 5 m thick) before bedrock is reached. Intertill aquifers were not mapped for the Swift Current area because the glacial deposits are so thin (Pennock et al., 2014).

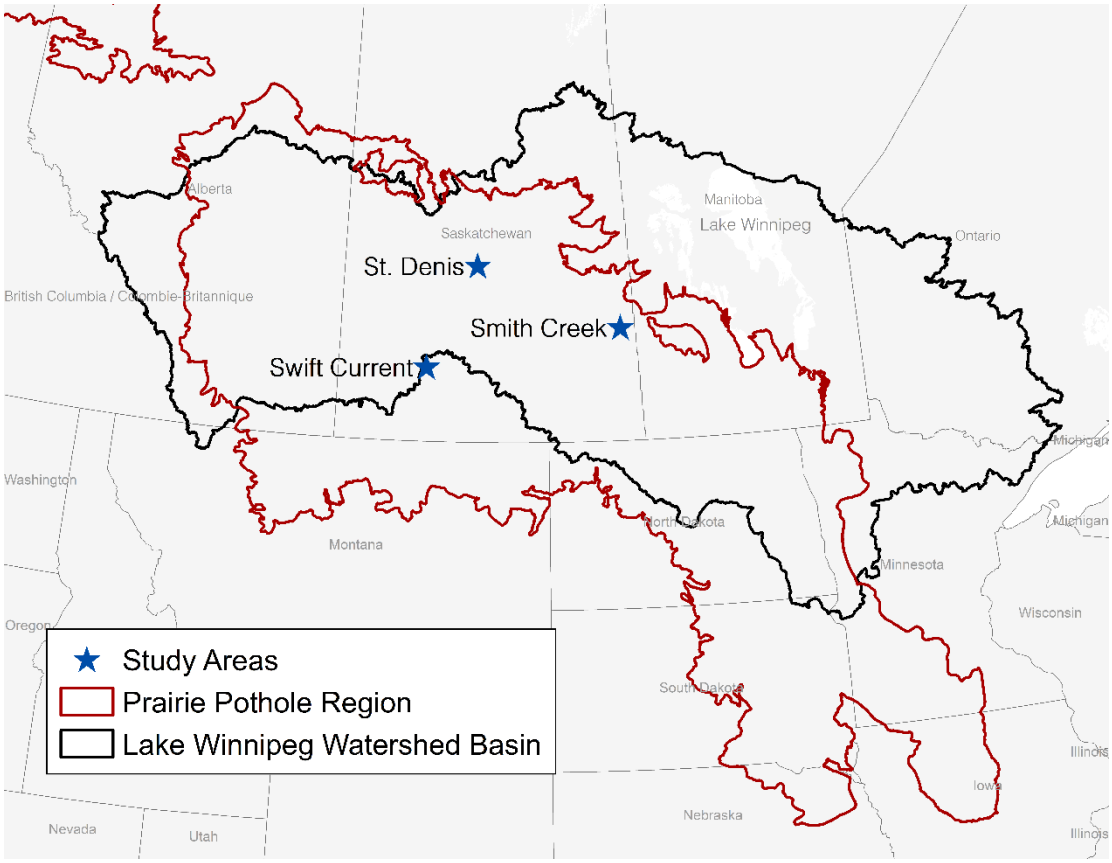



**Fig. 3.1 Study area locations and approximate extents of the PPR and Lake Winnipeg Watershed Basin.**

**Table 3.1 General study area characteristics.**

| Study areas | Area extent ($km^2$) | Landform | Slope (%) | Soil salinity effects on agricultural productivity | Soil zone | Mean annual temperature¶ ($^oC$) | Mean annual precipitation¶ (mm) |
|---|---|---|---|---|---|---|---|
| Swift Current | 13 | Hummocky-dissected, undulating-dissected† | 2 - 30† | None to slight† | Brown | 4.1 | 392.5 |
| St. Denis | 6 | Hummocky‡ | 2 - 30‡ | Very slight‡ | Dark brown | 3.3 | 340.4 |
| Smith Creek | 14 | Hummocky§ | 0 - 15§ | Very slight to slight§ | Black | 1.8 | 463.5 |

†Ayres et al. (1985)
‡Acton and Ellis (1978)
§Saskatchewan Soil Survey Staff (1991)
¶Climate normals for 1981 – 2010 (Government of Canada, 2018)

The landforms at the Swift Current sites were dominantly hummocky-dissected with some hummocky and undulating-dissected areas.

The St. Denis study sites are found on medium textured glacial tills with thin surficial deposits of fine-textured glacio-lacustrine materials in certain areas (Acton and Ellis, 1978). The St. Denis study area contains the SDNWA. The hydrology of the SDNWA has been studied extensively. A thin sand intertill aquifer is located roughly 20 – 30 m below ground level in the SDNWA portion of the watershed (Hayashi et al., 1998b). Pond water depth, salinity, and chemical characteristics have been monitored for the wetlands in the SDNWA since 1968 (Pennock et al., 2013).

The Smith Creek study sites are found on medium textured glacial tills (Saskatchewan Soil Survey Staff, 1991). The area has a hummocky landform, but the relief is gentler than the Swift Current and St. Denis study areas; the slopes ranged from 0 – 15% (Table 3.1). The study area overlies the Empress Group Aquifers (Saskatchewan Soil Survey Staff, 1991). These aquifers are at depths greater than 50 m and so the interaction between the surface waters and these aquifers are considered to be negligible.

### *3.4.2 Digital elevation models and aerial imagery*

LiDAR data were collected for the Swift Current study area between October 16 and 25, 2009 (Agriculture and Agri-Food Canada, 2009). The DEM had a horizontal resolution of 0.2 m and a vertical accuracy of 0.081 m (RMSE) (McElhanney Consulting Services Ltd., 2009). LiDAR data were collected for the St. Denis study area on August 9, 2005. The data had a horizontal resolution of 0.5 m and a vertical accuracy of 0.14 m (RMSE) (Töyrä et al., 2008). Because the data for St. Denis was collected earlier in the season, water was present at the base of the larger wetlands. The elevation surfaces in these positions reflect the water surfaces rather than the underlying sediment. Although this does affect the storage capacity of the depressions, the water levels were low enough within the depressions that it is not expected to significantly affect the proposed methodology. LiDAR data were collected for the Smith Creek study area between October 14 and 16, 2008. The data had a horizontal resolution of 1 m and a vertical accuracy of

0.05 m (RMSE) (Lidar Services International, 2009). To ensure consistency in the methodologies, 1-m resolution versions of all three site DEMs were used.

Aerial imagery was used to inform wetland boundary delineation and identify road networks. Ortho photos from 2008 – 2016 for each site were acquired from FlySask2 (Saskatchewan Geospatial Imagery Collaborative, 2018). Ortho photos that were collected at the same time as the LiDAR collection for the Swift Current area were also used (Agriculture and Agri-Food Canada, 2009).

### *3.4.3 Road and railroad removal from DEM*

Solute accumulations in wetlands are largely the result of redistribution through hydrologic processes occurring since the Wisconsin glaciation. To better reflect the historic surficial hydrologic connections between wetlands, roads and railroads were removed from the 1-m DEM. The road features were identified using the Saskatchewan Road Network Database 2014 shapefile (ISC, 2014). Any missing grid roads and railroads segments that were discernible from the aerial imagery and the DEM were added to the shapefile. Buffer polygons were created for each road segment to cover all portions of the DEM that contained road-related features and the portions of the DEM that fell within the buffer polygons were removed. The removed area was filled through interpolation using inverse distance weighting (ESRI, 2016). The interpolated areas were smoothed and added to the DEM to create the road-removed 1-m DEM. This DEM was used for all aspects of the model. A detailed description of the full road-removal methodology can be found in Appendix A-1.

### *3.4.4 Wetland boundary delineation*

Wetland boundaries were delineated by determining the closed topographic depressions within the DEM. Closed topographic depressions are groups of cells that represent features with internal drainage (Lindsay, 2016). These are common features in the PPR due to its glacial morphology; however, they can also be the result of errors within the DEM. The 1-m DEM was smoothed using a 2 x 2 mean filter and then resampled to 2 m to reduce the number of artefact depressions. The boundaries of the wetlands were defined by the spillover elevation of the closed depressions. The spillover elevation represents the maximum elevation of the depression before water would

spill from it. The methodology for determining the spillover elevations of the closed depressions is described in Appendix A-2. The spillover elevation boundary was used to define depression depth and area. Depressions with depths of less than 10 cm or areas of less than 50 $m^2$ were removed to further reduce the number of artefact depressions in the DEM. The minimum depth size was based on the recommendations of Li et al. (2011) who found wetlands with depths less than 10 cm were often erroneous features. The minimum area size was based on the smallest wetland observed as a part of the SDNWA wetland inventory that had a depth greater than 10 cm.

At the St. Denis and Swift Current sites, some of the closed topographic depressions encompassed very large areas that contained many smaller nested wetlands. For some of these closed depressions, the spillover elevation represented an unrealistic wetland boundary that would only be representative in extremely wet periods that have yet to be recorded (Fig. 3.2). It cannot be assumed that the nested wetlands within the massive closed topographic depressions would have the same solute contents under typical conditions. To separate out the nested wetlands, the closed topographic depressions entirely within the maximum closed topographic depressions were identified. This process was repeated until the closed depressions could not be subdivided into two or more nested wetlands (that met the minimum size criteria outlined above) and the boundaries of the closed depressions more closely matched the minimum wetland pond water and associated vegetation extents observed in the aerial imagery (Fig. 3.2). The methodology for determining the maximum closed depressions and the nested closed depression boundaries is described in Appendix A-2. This procedure was not necessary at the Smith Creek sites because the maximum closed topographic depressions closely matched the observable wetland extents in the aerial imagery. This was due to the subtler relief and increased moisture levels at the Smith Creek study area. The wetland inventories for each study area were based on the nested closed depression boundaries; the sampling design considered each polygon as an individual wetland. The model was tested using both the nested closed depression boundaries and the maximum closed depression boundaries to represent wetland boundaries.

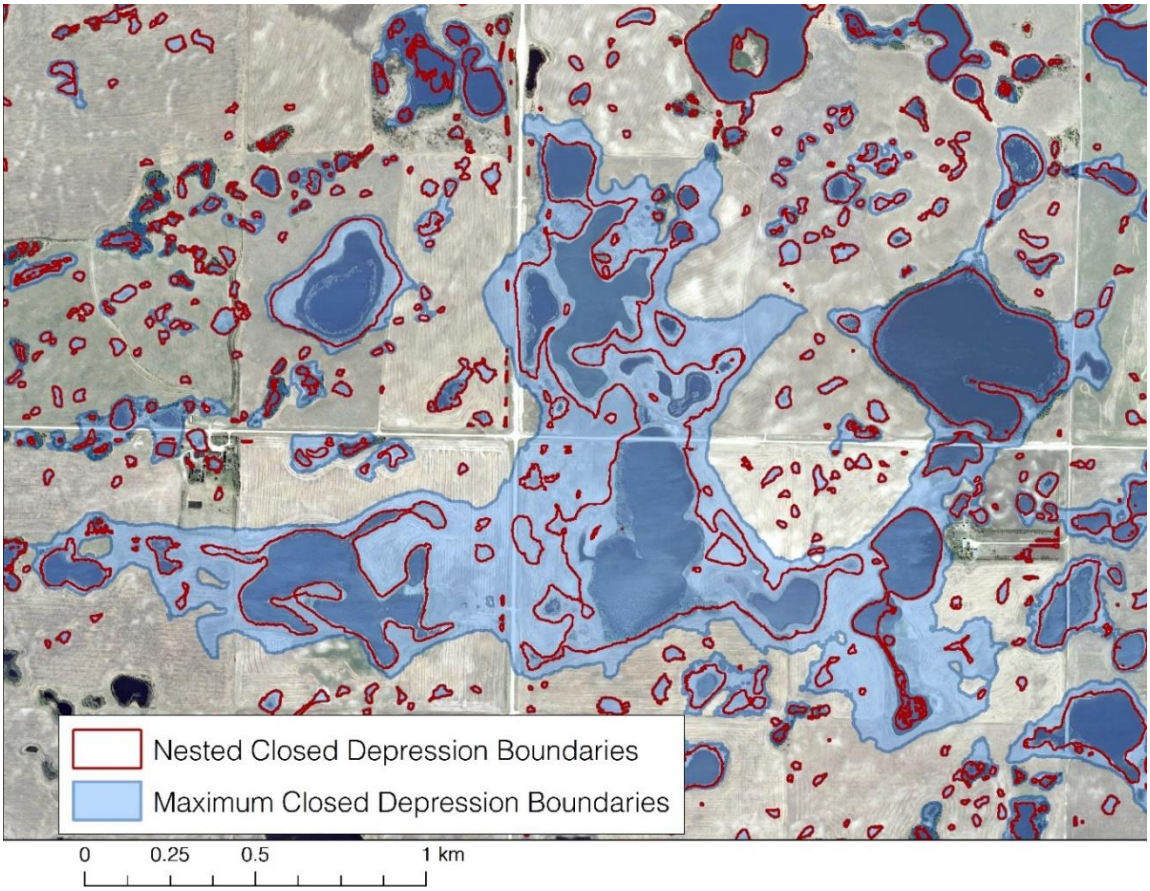



**Fig. 3.2 Wetland polygon boundaries based on nested closed depression boundaries and maximum closed depression boundaries at the SDNWA portion of the St. Denis study area.**

*3.4.5 Stream channel networks*

Surficial hydrologic connections between wetlands were represented using dendritic stream channel networks created from the DEM using the Basic Terrain Analysis module of the System for Automated Geoscientific Analysis (SAGA) (Conrad et al., 2015) (Fig. 3.3a). The road-removed 1-m DEMs were smoothed using a 5 x 5 mean filter and resampled to a resolution of 5 m before running the module. This DEM resolution is detailed enough to capture the important hydrological relationships for modelling at this scale and reduces computational requirements. A study by Shook et al. (2013) on specific wetland storage dynamics in similar regions used DEMs with spatial resolutions ranging from 6 to 15 m to adequately represent landscape-scale hydrologic processes.

Stream channel networks were ascribed Strahler order. Strahler ordering is a method of assigning relative orders to stream channels based on the number of stream channels that contribute flow to them (Strahler, 1952). Stream channels with Strahler orders less than three were not included in

the stream channel network. This was set using the channel density parameter of the Basic Terrain Analysis module. This parameter defines the minimum Strahler order required to begin a channel segment. Strahler orders were reassigned to the stream channel network so that the initial stream segments had an order of one. Stream channel network Strahler order values are central to the prediction model. Their values are affected by the spatial resolution of the DEM and the channel density set in the Basic Terrain Analysis module. To follow the recommendations of this study, it is necessary to use the same specifications for DEM spatial resolution and channel density.

### *3.4.6 Wetland Strahler order*

Wetland polygons were ascribed the maximum Strahler order of the stream segments that they intersected using the Spatial Join tool (ESRI, 2016) (Fig. 3.3a). The wetland Strahler order can be interpreted as a rough estimate of the fill-and-spill and shallow groundwater contributions a wetland is expected to have historically received. A wetland with a higher Strahler order would be expected to have received greater contributions of solutes.

Wetland Strahler order also quantifies the relative position of a wetland in the landscape. The stream channels flow in the direction of the greatest downhill slope gradient. This causes them to flow towards the lowest-lying landscape positions; in more typical drainage structures these would be streams and rivers, but in PPR landscapes these are dominantly wetland depressions. Stream channel networks converge in the depressions causing increases in Strahler orderings. The Strahler order reflects that a depression is the lowest lying landscape feature compared to the surrounding landscape. It also works to reflect a wetland's relative position within the watershed. Wetlands within a localized neighbourhood will have a higher Strahler order than the surrounding landscape features, but the Strahler order value is limited by its contributing area. Whereas, the lowest-lying wetlands within a watershed potentially have flow contributions from the entire watershed and therefore would have much higher Strahler orders. These lowest-lying wetlands with the highest Strahler orders represent wetlands with the greatest potential to be discharge wetlands. Therefore, although deep groundwater flow is not specifically modelled, lower-lying wetlands that have a greater potential to be discharge wetlands are predicted to be solute-rich.

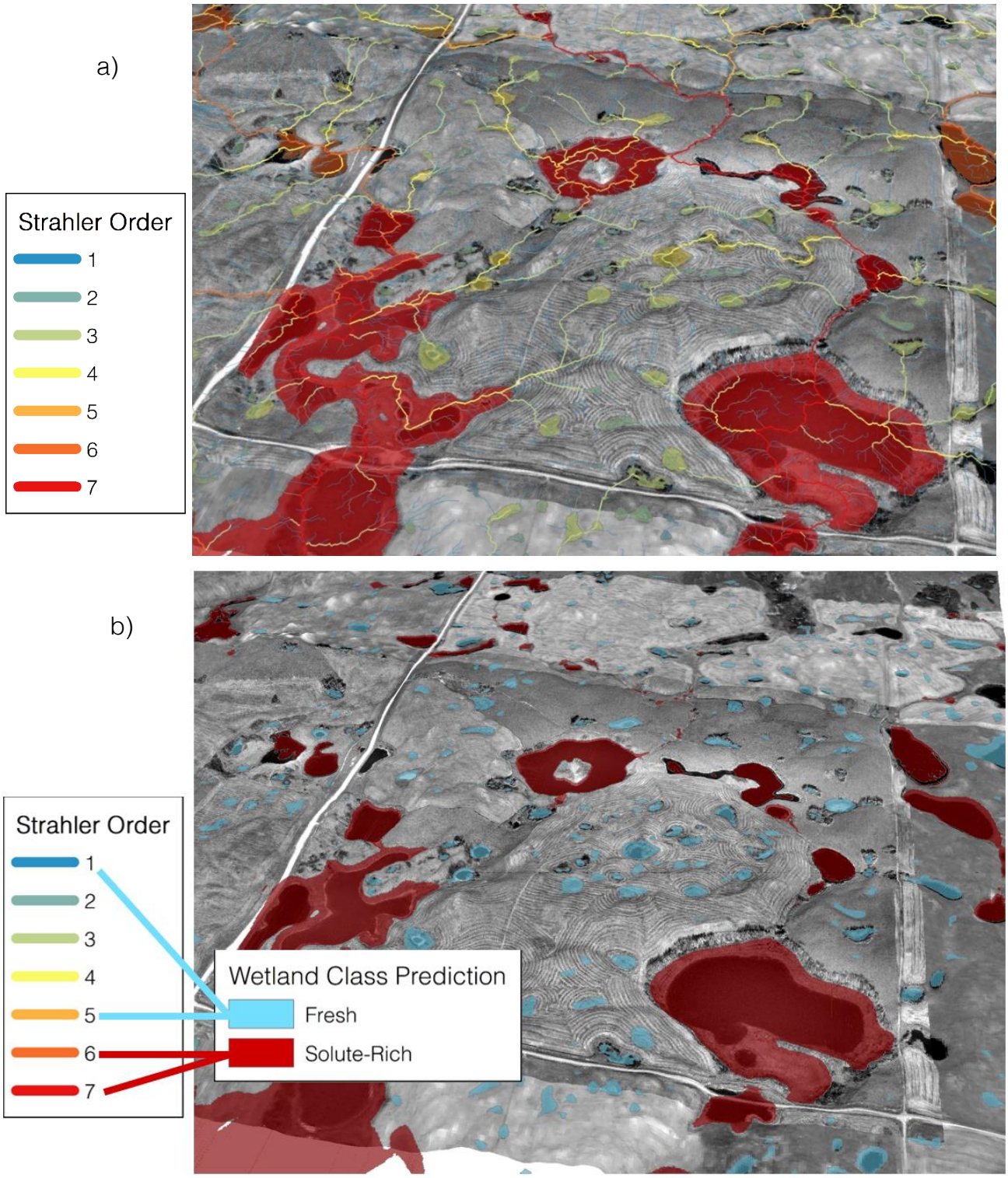



**Fig. 3.3 a) Stream channel network Strahler orders ascribed to intersecting wetland polygons for the SDNWA portion of the St. Denis study area. b) Wetland solute-richness class predictions based on wetland Strahler order threshold. The predictions pictured reflect an example where wetlands with Strahler order ≥ 6 are predicted to be solute-rich.**

Binary predictions of wetland solute-richness class were made with a simple decision tree. Wetland Strahler order was a variable considered in the model decision tree. Wetlands with a Strahler order above a specified threshold were predicted to be solute-rich (Fig. 3.3b). The specified wetland Strahler order threshold used to define wetland solute-richness class predictions was a model parameter that was tuned in the model training.

### *3.4.7 Predicted active spill channel networks*

Traditional stream channel network methodologies have been developed to model flow-routes in landscapes with more typical integrated hydrological drainage systems (rivers and streams). Within these methodologies, sinks and depressions within the DEM are filled to allow for stream flow to continue downstream uninterrupted (Fig. 3.4b). However, these sinks and depressions are a defining feature of the PPR hydrology; they interrupt hydrologic connections by providing intermittent storage basins. Therefore, the stream channel network methodologies developed for more typical hydrologic systems likely overestimate the level of hydrologic connectivity within a PPR landscape (Li et al., 2011).

In a PPR landscape, certain hydrologic connections between wetlands (via both fill-and-spill and shallow groundwater flow) are more likely to be active based on the volume capacity of the closed depressions and the amount of water present in a system (Fig. 3.4c). Certain depressions do not receive sufficient water to fill to their capacity and contribute water further downslope either through fill-and-spill or shallow groundwater flow (e.g. wetland 1 in Fig. 3.4c). Other wetlands may not require substantial water contributions to fill to their capacity (e.g. wetland 2 in Fig. 3.4c) or do require substantial water contributions to fill to their capacity but achieve them via large catchment areas. These wetlands are more likely to contribute fill-and-spill and shallow groundwater flow downslope.

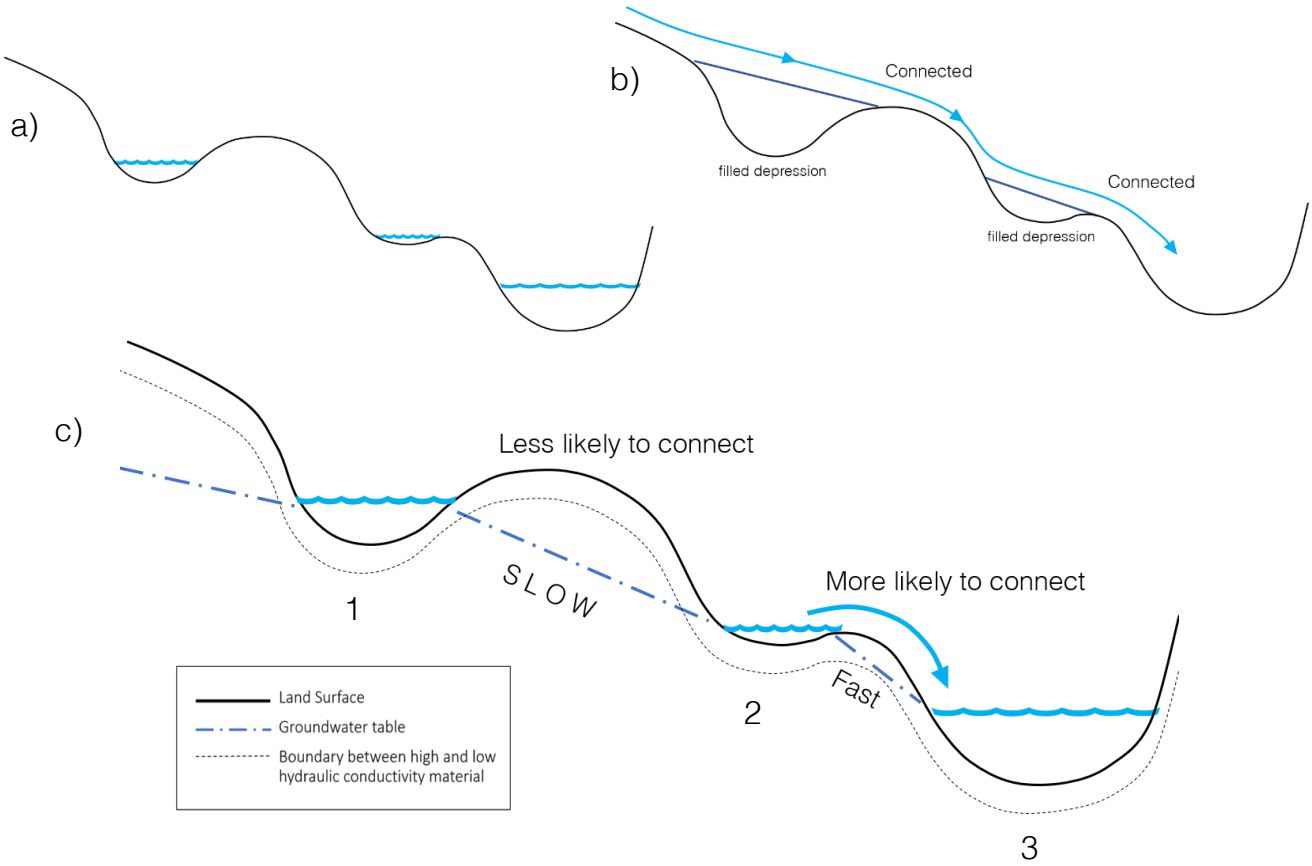



**Fig. 3.4 Example schematic of a sequence of wetlands. b) Depressions are filled to create continuous stream channel drainage networks as in traditional stream network methodologies. c) Example schematic demonstrating how certain hydrologic connections between wetlands (via fill-and-spill and shallow groundwater flow) are more likely to be active than others. "Slow" and "Fast" refer to the expected movement of groundwater. Wetland 1 requires substantial water contributions before it can contribute fill-and-spill flow and before the water table is high enough for the groundwater to move down-slope through the zone of higher hydraulic conductivity.**

The issue of overestimating the hydrologic connectivity within a PPR landscape using traditional stream channel networks was addressed using the Wetland DEM Ponding Model (WDPM) (Shook et al., 2014). The WDPM was developed to model distributions of runoff in the PPR. Within this model, water is applied evenly across the DEM and is redistributed using Shapiro and Westervelt's (1992) iterative algorithm to determine the water's final spatial distribution. Instead of allowing the water to settle to its final distribution, the WDPM model was stopped partway to observe which spill channels were active (Fig. 3.5a). Spill channels were determined to be active if water was distributed throughout the area between wetlands to form a continuous stream channel. The stream channel networks were adapted to reflect only the spill channels that were observed to be active by clipping the stream channel network by the WDPM spill channel

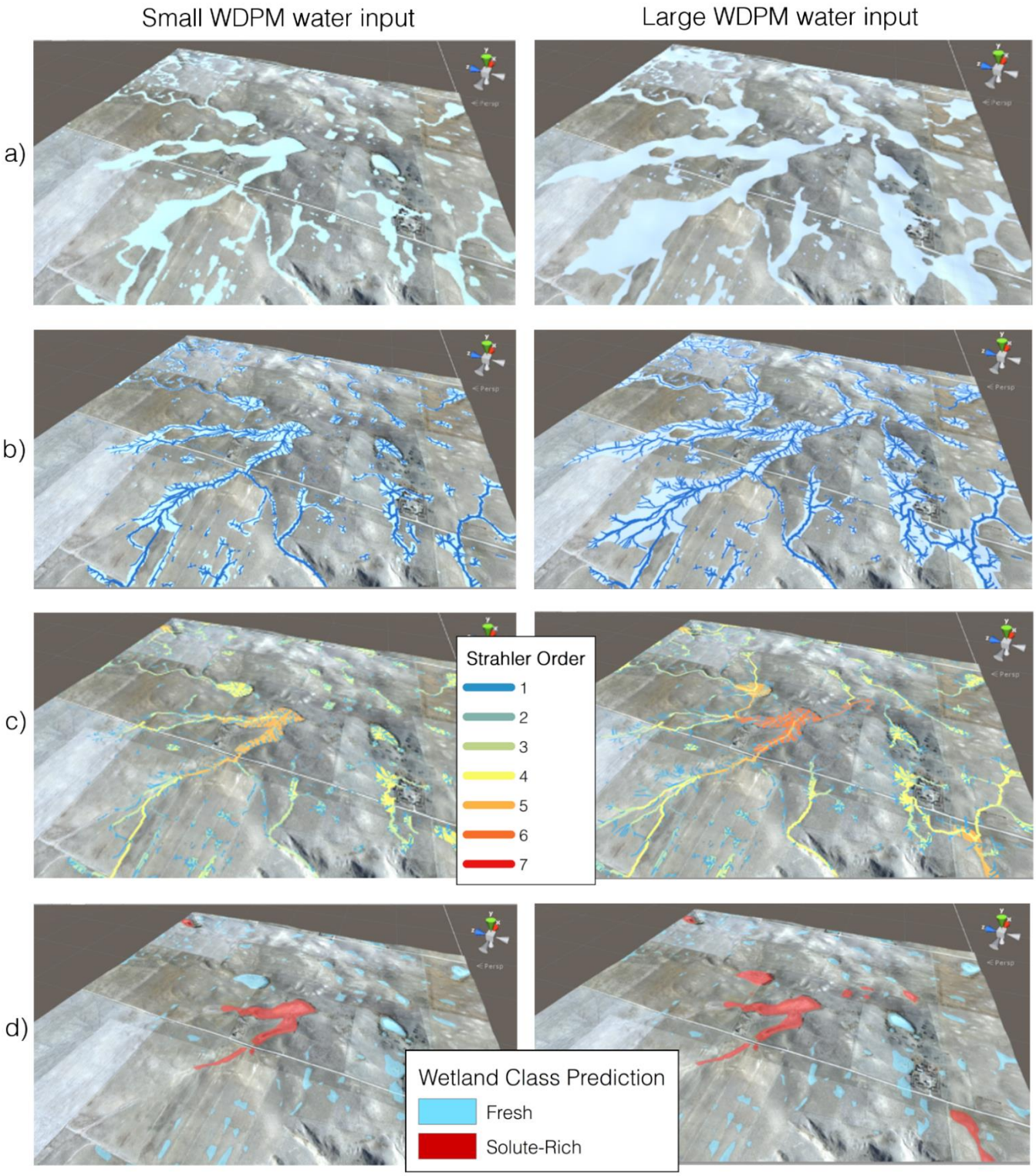



**Fig. 3.5 a) WDPM water distribution outline outputs based on a small water input (left) and a large water input (right). b) Stream channel networks clipped by the extents of the WDPM output polygons to create the predicted active spill channel networks. c) Strahler orders of the predicted active spill channel networks ascribed to the wetland polygons. d) Solute-richness classes predictions per wetland (dependent on the specified Strahler order threshold to distinguish fresh vs. solute-rich wetlands).**

polygon outputs (Fig. 3.5b). These adapted stream channel networks are referred to as the predicted active spill channel networks.

The amount of water distributed on the DEM is a controllable parameter in the WDPM. By adding greater amounts of water (Fig. 3.5, right), more depressions fill beyond their capacity and spill downslope, therefore more spill channels will be observed to be active and the stream channel network will be adapted to reflect a greater number of connections between wetlands. By adding smaller amounts of water (Fig. 3.5, left), there will be a smaller number of active connections between wetlands. The objective of this aspect of the model is to predict which spill channels have historically been most active to cause accumulations of solutes in wetlands lower lying in the landscape. Five versions of the predicted active spill channel networks were generated that reflect different levels of hydrologic connectivity. The predicted active spill channel networks were created by clipping the original stream channel network with the water distribution outputs from the WDPM (Fig. 3.5b). The volumes of water distributed on the DEMs were not meant to reflect realistic rainfall or snowmelt additions of precipitation; they were selected to achieve varying levels of connectivity between wetlands. The full methodology for creating the predicted active spill channel networks is described in Appendix A-3. In the model training, the model is tested using each of the five versions of the predicted active spill channel networks to determine which version results in the most accurate predictions of wetland solute-richness classes.

### *3.4.8 Terminal wetlands*

As described above, the wetland Strahler order reflects the contributions that a wetland receives. To predict the accumulations of solutes within a wetland, it is also important to consider whether the wetland can contribute water (and associated solutes) further downslope. Wetlands that receive water contributions, but cannot contribute water downslope, are terminal wetlands (Millar, 1976). The predicted active spill channel networks allowed for the determination of wetland terminal status. This would not be possible using the traditional methodologies for stream network generation because they assume that all depressions contribute water downslope. Wetlands were determined to be terminal if no segment of a spill channel had flow moving out of the wetland boundary polygon. This methodology is described in Appendix A-3. Terminal

wetlands needed to be distinguished from isolated wetlands. Isolated wetlands also do not contribute water further downslope but differ from terminal wetlands in that they do not receive significant contributions from upslope wetlands (Millar, 1976). Terminal wetlands were required to have a wetland Strahler order of greater than two to ensure that the wetlands would receive substantial upslope contributions. The number of wetlands determined to be terminal was affected by the level of hydrologic connectivity predicted by the predicted active spill channel networks. With more expected connectivity, there would be less wetlands expected to be terminal because they are more likely to contribute water downslope.

### *3.4.9 Model tuning parameters*

The methodologies of the prediction model are summarized in the flowchart in Fig. 3.6. There were four model parameters tuned to refine model predictions (Table 3.2). Two parameters determined the GIS methodologies for ascribing Strahler orders and terminal status to wetlands (Fig. 3.6). The first parameter determined how wetland boundaries were defined. The model was tested using the maximum closed topographic depression boundaries and the nested closed topographic depression boundaries for the wetland observations. Using the maximum closed topographic depression boundaries increases the likelihood of a wetland having a higher Strahler order and being terminal because of the increased wetland boundary size.

**Table 3.2 Model tuning parameters.**

| Model tuning parameters | | | |
|---|---|---|---|
| Parameters affecting wetland Strahler order and terminal status | | Decision tree parameters | |
| Wetland boundary | Spill channel connectivity | Solute-rich Strahler order | Terminal status rule |
| Nested depressions<br>Maximum depressions | Minimum<br>Near minimum<br>Moderate<br>Near maximum<br>Maximum | $\geq 3$<br>$\geq 4$<br>$\geq 5$<br>$\geq 6$ | Not considered<br>Terminal only<br>OR Terminal<br>AND Terminal |

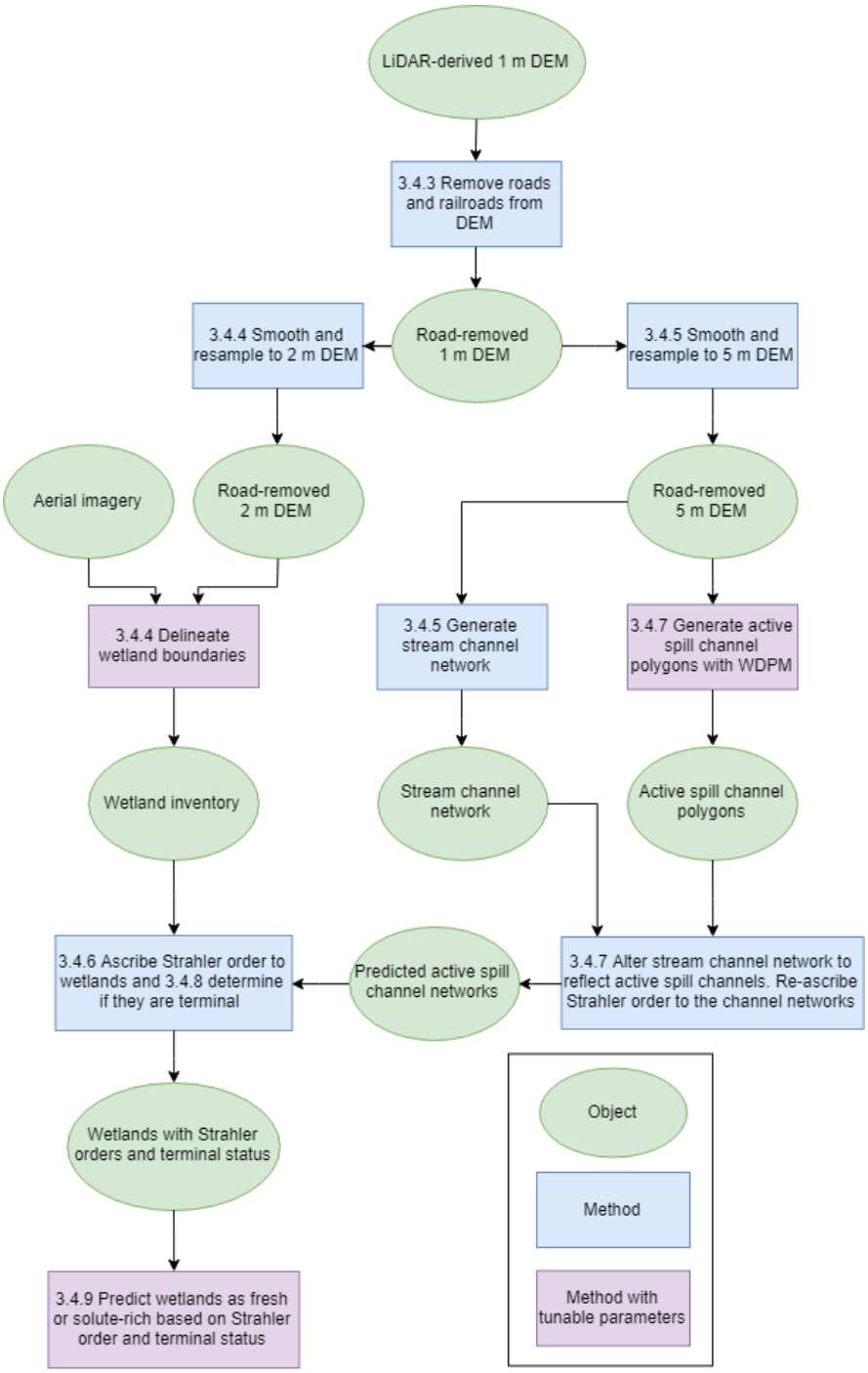



**Fig. 3.6 Flowchart of wetland solute-richness prediction model methodology.**

The second parameter determined the predicted active spill channel network to represent hydrologic connections between wetlands. Maximum was the stream channel network unaltered to reflect spill channels that were more likely to be active. The remaining spill channel networks were altered versions of the stream channel network to reflect spill channels that were more

likely to be active using the water distribution outputs of the WDPM model; the predicted active spill channel networks.

The model was a simple decision tree consisting of either one or two nodes to make binary fresh vs. solute-rich wetland predictions. Two model parameters altered the decision tree rules: 1) the Strahler order threshold between fresh and solute-rich predictions and 2) how wetland terminal status is considered in the decision tree. For an example of the solute-rich Strahler order threshold, at a threshold of ≥ 3, wetlands with a Strahler order greater than or equal to three were predicted to be solute-rich. Four options were tested for the terminal status rule: 1) Not considered – The model did not consider wetland terminal status and wetland solute-richness class predictions were based entirely on the wetland Strahler order. 2) Terminal only – The model does not consider wetland Strahler order and terminal wetlands were predicted as solute-rich and non-terminal wetlands as fresh. 3) OR Terminal – wetlands were predicted to be solute-rich if they either had a Strahler order above the defined threshold OR were terminal. 4) AND Terminal – Wetlands were predicted to be solute-rich if they both had a Strahler order above the defined threshold AND were terminal.

*3.4.10 Wetland salinity data collection*

A stratified random sampling design was used to select wetlands to test for salinity. The nested closed depression boundary polygons were used to delineate individual wetlands. The sampling design was informed by the prediction model, but it did not consider wetland terminal status; that aspect of the model was developed after the field sampling. For the sampling design, wetlands were classified as Expected Solute-Rich or Expected Fresh based on their wetland Strahler order. Wetlands with a Strahler order greater than or equal to five, according to any of the predicted active spill channel networks, were classified as Expected Solute-Rich. Wetlands with a Strahler order less than five, according to all the predicted active spill channel networks, were classified as Expected Fresh. The Expected Solute-Rich wetlands were further stratified according to which predicted active spill channel networks ascribed them a Strahler order greater than or equal to five. However, there were not enough wetlands from each category available at each site to achieve a balanced number of samples per category (Table 3.3). There were many small Expected Fresh wetlands and so they were stratified based on size to ensure larger Expected

Fresh wetlands were sampled (Table 3.3). Wetlands with areas greater than 4047 $m^2$ were classified as large and wetlands with areas less than 4047 $m^2$ were classified as small. Size classes were informed by Millar's (1976) size classifications, which were described in acres. A minimum size of 10 cm depth and 100 $m^2$ area was used to avoid sampling very small depressions that did not exhibit wetland characteristics. The Expected Fresh wetlands were also stratified by relative elevation in watershed. Relative elevation reflected two categories, upper and lower halves of the watershed, which were determined using Jenks natural breaks classification method (Jenks, 1967).

**Table 3.3 Number of wetlands sampled per stratification at each site for the model training and external validation datasets.**

| Sample set | Model prediction | PASCN† causing ≥ 5 Strahler order | Relative elevation | Size | Swift Current (n) | St. Denis (n) | Smith Creek (n) |
|---|---|---|---|---|---|---|---|
| Training Set | Expected Fresh | None | Upper | Small | 3 | 5 | 9 |
| | | | | Large | 5 | 5 | 10 |
| | | | Lower | Small | 5 | 5 | - |
| | | | | Large | 3 | 3 | - |
| | | | | Total | 16 | 18 | 19 |
| | Expected Solute-Rich | Minimum | | | 4 | 9 | 12 |
| | | Near minimum | | | 3 | 9 | 2 |
| | | Moderate | | | 2 | - | 4 |
| | | Near maximum | | | 4 | 2 | - |
| | | Maximum | | | 4 | - | - |
| | | | | Total | 17 | 20 | 18 |
| Soil Sample Test Set | | | | | 16 | 12 | 11 |
| St. Denis Historic Test Set | | | | | - | 115 | - |

†PASCN = Predicted active spill channel network

Water samples were collected where pond water was present and tested for electrical conductivity using a PC 700 conductivity meter (Oakton Instruments, Vernon Hills, Illinois). Electromagnetic conductivity surveys were conducted with an EM38 unit (Geonics Ltd., Mississauga, Ontario) at the edge of the pond water extent for wetlands that were inundated and at the wetland center when standing water was not present. The Swift Current and St. Denis sites were sampled in September and October of 2015 and the Smith Creek sites were sampled in June

2016. Because the Smith Creek sites were sampled earlier in the season, it is likely that the wetlands had greater water levels than they would have in the fall because they would not have experienced the same period for evaporative loss (Euliss et al., 2014). Therefore, the solute concentrations within these wetlands may have been diluted. The wetland solute-richness classifications were based on very broad ranges of salinity values and so variability of measurements between seasons should not have significantly affected classifications. This dataset is referred to as the Training Set. It was used to tune the model parameters (discussed in section 3.4.9) to maximize predictive accuracy.

Drained wetlands were not sampled. The majority of wetlands located in the lower half of the Smith Creek watershed were drained. All sampling in this study area was done in the upper elevation of the watershed. Wetlands sampled directly adjacent to roads were subsequently removed from the Training Set. Several wetlands sampled at Swift Current and a few at Smith Creek were determined to be on parent material other than glacial till and were subsequently removed from the Training Set.

Wetland salinity was tested at an additional 42 wetlands across the three study areas as a part of the *Predictive digital soil mapping of wetland soil types in the Canadian Prairie Pothole Region* study (Ch. 4). A stratified random sampling design was used to select these wetlands; they were not stratified based on wetland Strahler orders, they were stratified based on size and relative elevation in the watershed (methodologies described in section 4.4.2). This is referred to as the Soil Sample Test Set.

Since 1968, pond water electrical conductivity data has been collected for all wetlands within the SDNWA (Waiser, 2006). This is referred to as the St. Denis Historic Test Set. These two datasets, the Soil Sample Test Set and the St. Denis Historic Test Set, were used as external validation tests for the prediction model after it was trained using the Training Set. For the St. Denis Historic Test Set, wetland salinity averages were calculated for wetlands with multi-year data. A few wetlands identified in the SDNWA wetland inventory shared single wetland polygons. Each wetland in the SDNWA inventory had individual salinity measurements and so they were considered individual observations even when they shared a single wetland polygon.

*3.4.11 Wetland solute-richness classification*

The objective of this model was to predict the spatial distribution of solute-rich wetlands for the purpose of identifying wetlands with $CaCO_3$-enriched soils. However, water and soil salinity data can be collected far more readily and efficiently than testing for soil $CaCO_3$ content. It is assumed that, generally, wetlands with greater accumulations of salinity will also have greater accumulations of $CaCO_3$ as landscape-scale solute distribution is controlled by the same hydrological mechanisms (Pennock et al., 2014). Due to the inherent variability of wetland salinity (Euliss et al., 2014), the model aimed to predict wetlands as simple binary classes: solute-rich or fresh. The wetland classification criteria were based on electrical conductivity (EC) of pond water and electromagnetic conductivity (EM) of the wetland soil (Table 3.4). Wetlands with an EC of greater than 1000 µS $cm^{-1}$ were classified as solute-rich because calcium carbonates begin to precipitate out of the pond water and concentrate in the soil at this threshold (Arndt and Richardson, 1989). This value also reflects a natural break in the distribution of wetland salinity observations measured at the SDNWA between 1968 and 2013 (Fig. 3.7) (Pennock et al., 2013). The EM value of 70 mS $m^{-1}$ was chosen because wetlands with EC greater than 1000 µS $cm^{-1}$ typically had soil EM values greater than 70 mS $m^{-1}$.

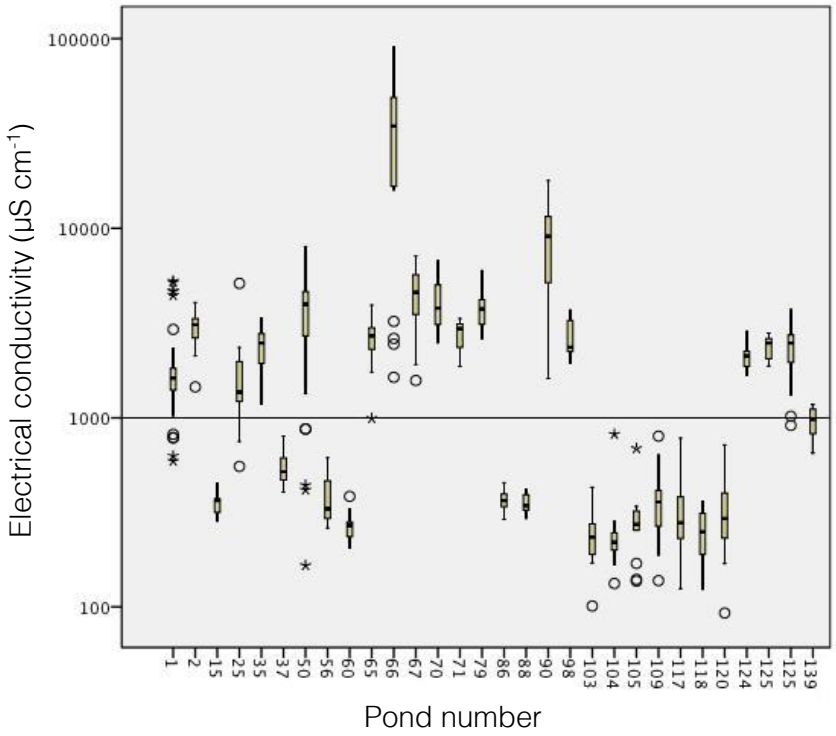



**Fig. 3.7 Boxplots of SDNWA wetland pond water EC observations measured between 1968 and 2013. Wetlands with more than five observations are shown. Adapted from Pennock et al. (2013).**

**Table 3.4 Wetland solute-richness class criteria and the number of observations per class per testing and external validation datasets.**

| | Class criteria | | Training | Soil | St. Denis |
|---|---|---|---|---|---|
| Wetland class | Water EC ($\mu S\ cm^{-1}$) | EM for 0 - 1.5 m ($mS\ m^{-1}$) | Set (n) | Sampled Set (n) | Historic Set (n) |
| Fresh | < 1000 | < 70 | 38 | 35 | 48 |
| Solute-Rich | > 1000 | > 70 | 46 | 4 | 67 |

Wetland classifications were made on available data if an observation did not have values for both measurements. Wetland observations with contradicting pond EC and soil EM values were removed from the training and external validation datasets. It was undetermined which measurement is more representative of wetland $CaCO_3$-enrichment. These observations fall in the middle of the spectrum between fresh and solute-rich, as defined here. The objective of this model is only to predict wetlands as two broad classes of solute-richness. Correct or incorrect predictions for these questionably classified wetlands would not offer useful information to the modelling process.

## 3.5 Results and discussion

### *3.5.1 Model training results*

A wide range of predictions for the spatial distributions of solute-rich wetlands were generated by the models based on the varying combinations of model parameters. The best-performing models for predicting the solute-richness classes of the Training Set of observations are shown in Table 3.5. Total accuracy reflects the percentage of total observations that were correctly predicted. The Kappa score adjusts the overall predictive accuracy according to the likeliness of predictions being correct based on chance alone (Landis and Koch, 1977) and is defined as

$$Kappa = \frac{p_o - p_e}{1 - p_e}$$

where $p_o$ represents the overall accuracy and $p_e$ represents the expected accuracy between observations and predictions. Higher Kappa scores indicate more agreement between predictions and observations. Producer's accuracy, reported per solute-richness class, refers to the number of observations of a soil class that were correctly predicted divided by the number of observations of that soil class (Malone et al., 2017). User's accuracy (discussed later in the results) refers to

the number of observations of a soil class that were correctly predicted divided by the number of predictions of that soil class.

The best-performing models included those that had the highest total accuracy predicting for the entire Training Set, as well as the models that had the highest total accuracy per each study area. Model A had the highest total accuracy and Kappa score. Models C and D were the best-performing models at the Swift Current study area. Models B and G were the best-performing models at the Smith Creek study area. Several models achieved the highest total accuracy for the St. Denis study area. The Smith Creek study area had disproportionately more solute-rich wetlands than fresh wetlands and models that predicted greater frequencies of solute-rich wetlands performed better in this area (e.g. in Models B and G). The Swift Current study area had disproportionately more fresh wetlands than solute-rich wetlands. Model parameters that worked best at the Swift Current study area (e.g. in Models C and D) were not as accurate at the Smith Creek study area.

### *3.5.2 External validation results*

The best-performing models based on the training data listed in Table 3.5 were tested for predicting the wetland solute-richness classes of the external validation sample sets: the Soil Sample Set and the St. Denis Historic Set (Table 3.6). The models all performed well for the Soil Sample Set, however, there were only four solute-rich wetlands in this dataset. Model D had the highest total accuracy but a very low accuracy for predicting the solute-rich wetland observations. It had the lowest Kappa score for the Soil Sample Set. Model E and F were the best-performing models for the St. Denis Historic Set of observations. Their Kappa scores correspond to moderate levels of agreement between predictions and observations (Landis and Koch, 1977). Based on the predictive accuracies for both the training and external validation datasets, Model F represents the best-performing model overall. Model E had a higher total accuracy for the St. Denis Historic Set of observations, but Model F was slightly more successful for correctly predicting the solute-rich observations. Model F also achieved more balanced performance for both the Smith Creek and Swift Current study areas (Table 3.5).

**Table 3.5 Prediction accuracies for the best-performing models for predicting the solute-richness classes of the Training Set of observations overall and for each study area. Total accuracy values are written in bold to indicate they were either the highest overall or highest for a specific site.**

| Model | Model parameters | | | | Overall Training Set | | | | Swift Current | | | St. Denis | | | Smith Creek | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Wetland boundary | Spill channel connectivity | S-Rich† Strahler order | Terminal status rule | Total accuracy (%) | Kappa | Fresh pr. acc.‡ n = 38 (%) | S-Rich pr. acc. n = 46 (%) | Total accuracy (%) | Fresh pr. acc. n = 23 (%) | S-Rich acc. n = 7 (%) | Total accuracy (%) | Fresh pr. acc. n = 9 (%) | S-Rich pr. acc. n = 11 (%) | Total accuracy (%) | Fresh pr. acc. n = 6 (%) | S-Rich pr. acc. n = 26 (%) |
| A | Nested | Near Minimum | ≥ 4 | Not considered | **83** | 0.66 | 82 | 85 | 84 | 83 | 86 | **90** | 89 | 91 | 79 | 67 | 81 |
| B | Nested | Near Minimum | ≥ 4 | OR terminal | 82 | 0.63 | 77 | 87 | 77 | 74 | 86 | **90** | 89 | 91 | **82** | 67 | 85 |
| C | Nested | Minimum | ≥ 4 | Not considered | 82 | 0.64 | 87 | 78 | **90** | 92 | 86 | 85 | 89 | 82 | 72 | 67 | 74 |
| D | Nested | Moderate | ≥ 5 | Not considered | 80 | 0.59 | 93 | 69 | **87** | 92 | 72 | 80 | 89 | 73 | 72 | 100 | 66 |
| E | Maximum | Minimum | ≥ 4 | Not considered | 81 | 0.61 | 79 | 82 | 84 | 79 | 100 | **90** | 89 | 91 | 72 | 67 | 74 |
| F | Maximum | Near Minimum | ≥ 4 | Not considered | 82 | 0.63 | 77 | 87 | 80 | 74 | 100 | **90** | 89 | 91 | 79 | 67 | 81 |
| G | Maximum | Near Minimum | ≥ 4 | OR terminal | 82 | 0.63 | 74 | 89 | 77 | 70 | 100 | **90** | 89 | 91 | **82** | 67 | 85 |

†S-Rich = Solute-rich
‡pr. acc = producer's accuracy



**Table 3.6 Accuracies for predicting the solute-richness classes of the external validation sample sets using the best-performing models for the Training Set. Total accuracies were written in bold to indicate they were the best-performing models for the St. Denis Historic Set.**

| Model | Model parameters | | | | Soil Sample Set | | | | St. Denis Historic Set | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Wetland boundary | Spill channel connectivity | S-Rich† Strahler order | Terminal status rule | Total accuracy (%) | Kappa | Fresh pr. acc.‡ n = 35 (%) | S-Rich pr. acc. n = 4 (%) | Total accuracy (%) | Kappa | Fresh pr. acc. n = 48 (%) | S-Rich pr. acc. n = 67 (%) |
| A | Nested | Near Minimum | ≥ 4 | Not considered | 80 | 0.41 | 78 | 100 | 60 | 0.24 | 88 | 39 |
| B | Nested | Near Minimum | ≥ 4 | OR terminal | 80 | 0.41 | 78 | 100 | 59 | 0.22 | 84 | 41 |
| C | Nested | Minimum | ≥ 4 | Not considered | 80 | 0.41 | 78 | 100 | 62 | 0.30 | 96 | 38 |
| D | Nested | Moderate | ≥ 5 | Not considered | 85 | 0.16 | 92 | 25 | 61 | 0.28 | 96 | 36 |
| E | Maximum | Minimum | ≥ 4 | Not considered | 80 | 0.41 | 78 | 100 | **70** | 0.42 | 90 | 56 |
| F | Maximum | Near Minimum | ≥ 4 | Not considered | 80 | 0.41 | 78 | 100 | **69** | 0.40 | 86 | 57 |
| G | Maximum | Near Minimum | ≥ 4 | OR terminal | 80 | 0.41 | 78 | 100 | 65 | 0.30 | 75 | 57 |

†S-Rich = Solute-rich
‡pr. acc = producer's accuracy

### *3.5.3 Predictive accuracies per solute-richness class*

All models had low producer's accuracy for predicting the solute-rich wetland observations of the St. Denis Historic Set, meaning many of the solute-rich wetland observations were predicted to be fresh wetlands. The St. Denis Historic Set observations were all located in the lower portion of the St. Denis watershed. Some of these wetlands receive groundwater discharge from an intertill aquifer (Hayashi et al., 1998b; Heagle et al., 2013). Deep groundwater movement was not specifically modelled and so the model is unable to predict all wetlands that receive groundwater discharge from intertill aquifers. The user's accuracy for solute-rich wetlands according to Model F is 85%. User's accuracy reflects the likelihood of finding that specific predicted class to be correct. Therefore, although the model underestimates the frequency of solute-rich wetlands in this area, the wetlands that are predicted to be solute-rich are confidently predicted as such.

### *3.5.4 Site-specific characteristics*

It is important to consider specific site characteristics, like geochemistry and climate, that may influence the distributions of fresh and solute-rich wetlands. The model parameters that work best for one area will not necessarily work best for others. In terms of mapping wetlands with $CaCO_3$-enriched soils, the parent materials of the Smith Creek study area are characterized by an abundance of $CaCO_3$ (Saskatchewan Soil Survey Staff, 1991) and therefore, more $CaCO_3$-rich wetlands would be expected in this area. Site-specific information is available in the detailed soil surveys. Soil $CaCO_3$ content is not specifically mapped but site-specific salinity information is available. The Smith Creek study area had more sites mapped as having salinity levels with *slight* affects on productivity than the other areas and therefore, more solute-rich wetlands would be expected in this area. The inclusion of site-specific salinity information could refine the model's predictions to reflect general geochemical characteristics for a target area. The study areas also experience different climates, which can influence hydrologic characteristics due to differences in available moisture. With greater moisture inputs, more hydrologic connectivity within the landscape is expected, which could potentially mean greater amounts of solute-rich wetlands due to increased redistribution of solutes. These site-specific climatic differences may be best reflected by different parameter values used in the model. Model parameters could be

calibrated for a target area using wetland salinity data from that area to improve the model performance.

### *3.5.5 Best-performing model parameters*

The prediction results show that adapting the stream channel networks to reflect spill channels that are more likely to be active is advantageous for this modelling purpose. The best-performing models were based on predicted active spill channel networks that represented either minimum, near minimum, or moderate levels of connectivity between wetlands. The maximum predicted active spill channel network, which was the original stream channel network unaltered to reflect spill channels that are more likely to be active, did not perform as well. The unaltered stream channel network overestimates the hydrologic connectivity between wetlands in a PPR landscape. Hydrologic connectivity within a PPR landscape is controlled largely by the characteristics of the depressional features (Shook et al., 2013) and hydrological models built for this region need to account for this (Li et al., 2011). The method proposed in this study accounts for these features which contributed to its predictive success.

The model accuracy was not improved by separating nested wetlands from the larger closed topographic depressions. The predictive accuracies were similar using the maximum closed topographic depressions for the wetland boundaries. Although the nested wetlands would each have unique hydrologic characteristics causing differences in their solute accumulations, the characteristics of their respective maximum closed topographic depression may be more important for determining their solute-richness. The additions and removals of solutes have been developing for thousands of years (Van der Kamp and Hayashi, 2009). Although some maximum closed topographic depressions seem exceedingly large compared to current wetland extents, they may have functioned as single waterbodies for extensive periods of time. Water records dating back only slightly more than half a century show huge variation in the water levels at the SDNWA (Pennock et al., 2013); wetland extents would have varied considerably more than this since the time of glaciation. Pond water salinity within a single wetland does not vary substantially spatially (Euliss et al., 2014). Therefore, during periods where the maximum wetland boundaries were fully inundated, the nested wetlands within likely accumulated solute depositions at similar rates.

Wetland Strahler order was more important than wetland terminal status for predicting wetland solute-richness classes. Most of the best-performing models did not consider terminal status of wetlands to determine their predictions. The best-performing models predicted wetlands with Strahler orders of four or greater to be solute-rich, except Model D. This Strahler order is relatively low compared to the maximum Strahler orders observed across the entire watersheds. This indicates that wetland solute accumulations (to the level of interest in this study) can occur from the hydrologic contributions generated over relatively small areas. Wetlands expected to be enriched with $CaCO_3$ are not only found in the lowest positions of the watershed but are also present in more localized low-lying landscape positions. This corresponds to the findings of Cook and Hauer (2007) who observed significant solute enrichment in wetlands with expected hydrologic connectivity to an upslope wetland. The connected wetlands did not require contributions from many wetlands; with just one adjacent hydrologically-contributing wetland, solute-enrichment was found to be much greater than that of isolated wetlands.

### *3.5.6 Mapped distributions of solute-rich wetlands*

The maps in Fig. 3.8 show the predicted distributions of solute-rich and fresh wetlands of the three study areas based on Model F, which was the best performing model overall. The maps indicate that the solute-rich wetlands were typically larger wetlands that are expected to receive hydrologic contributions from wetlands upslope. These wetlands would be expected to have soils enriched with $CaCO_3$ and, therefore, these wetlands could be prioritized in terms of conservation to maximize phosphorus retention in the PPR. Many of these wetlands would be considered "gatekeeper" wetlands (Phillips et al., 2011), which control the contributions from numerous upstream wetlands to downstream waterbodies by providing an intermittent storage basin. Hence, in addition to stopping runoff from moving into key waterways downstream, these wetlands are also likely to have highly $CaCO_3$-enriched soil, and therefore greater phosphorus retention capacity, stressing the importance of conserving these wetland types.

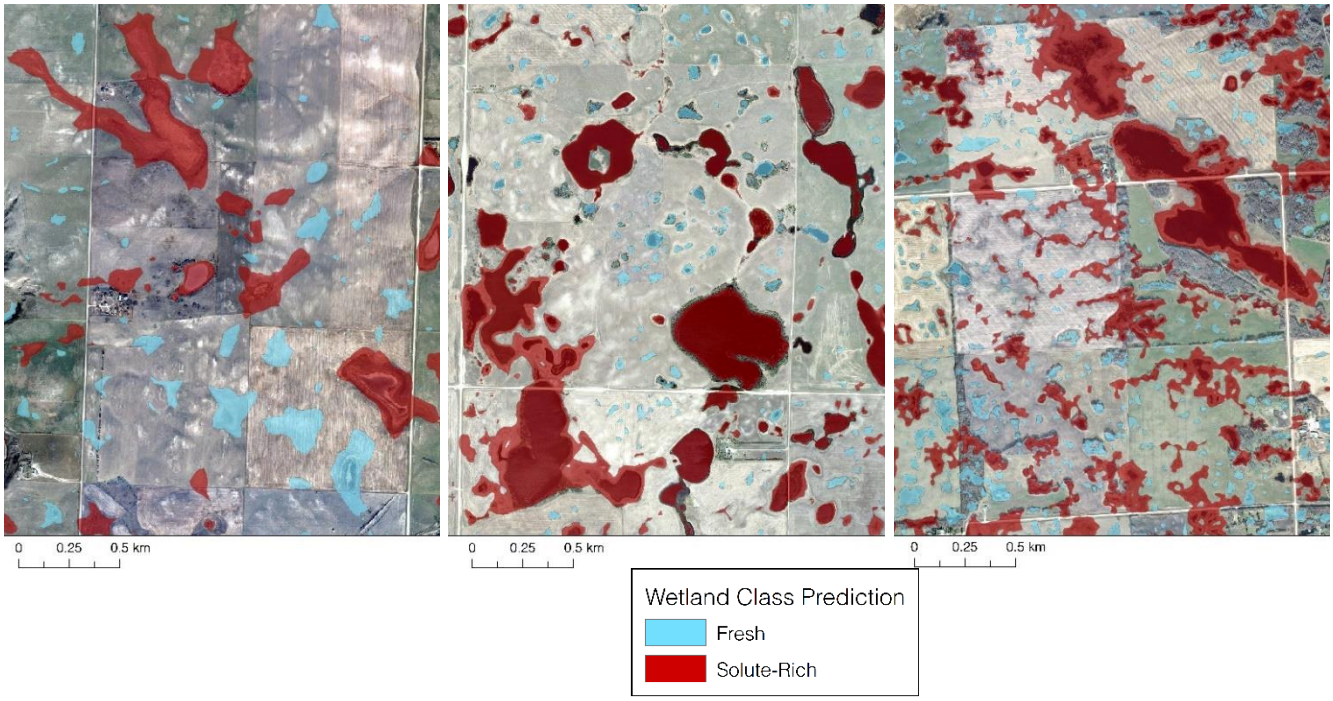



**Fig. 3.8 Predicted spatial distributions of fresh and solute-rich wetlands based on Model F for portions of the Swift Current (left), St. Denis (center), and Smith Creek (right) study areas.**

*3.5.7 Modelling wetland salinity for other applications*

The goal of this study was to predict the spatial distributions of solute-rich wetlands that would be expected to have soils enriched with $CaCO_3$. The salinity criteria of the solute-richness classes were based on thresholds reflecting that characteristic. The model has potential to be used to map wetland salinity for other purposes. To predict the spatial distributions of solute-rich wetlands based on different salinity thresholds, different model parameters may perform better. The model was also tested to predict the spatial distribution of saline wetlands that would pose risks to agronomic productivity. The salinity classes were based on different EC and EM criteria (Table 3.7). Soils with EM values greater than 100 mS $m^{-1}$ are expected to have salinity levels that seriously limit crop productivity (Henry, 2003). The threshold of 2000 µS $cm^{-1}$ was selected as the criteria for wetland pond water EC, as this value corresponded to wetlands with EM values greater than 100 mS $m^{-1}$.

Using these criteria, there were limited numbers of solute-rich wetland observations in each sample set. A new set of observations (referred to as the Agronomic Risk Sample Set) was created to assess the model's applicability for predicting solute-richness classes based on these

different criteria. The Agronomic Risk Sample Set included all 68 solute-rich wetlands from the datasets, and 69 fresh wetlands were randomly sampled from the Training Set: 23 from each study area. The models were not tested with external validation observations due to the limited number of solute-rich wetlands corresponding to these new class criteria.

**Table 3.7 Wetland solute-richness class criteria based on salinity risks to agronomic productivity and the number of observations per class per dataset.**

| | Class criteria | | | Training | Soil | St. Denis | Agronomic Risk |
|---|---|---|---|---|---|---|---|
| Wetland class | Water EC ($\mu S\ cm^{-1}$) | EM for 0 - 1.5 m ($mS\ m^{-1}$) | Exceptions | Set (n) | Sampled Set (n) | Historic Set (n) | Sample Set (n) |
| Fresh | < 2000 | < 100 | | 92 | 40 | 71 | 69 |
| Solute-Rich | > 2000 | > 100 | OR EM > 150 regardless of EC | 16 | 2 | 44 | 68 |

The best-performing model parameters for predicting the distributions of solute-rich wetlands according to these class criteria (Table 3.8) varied considerably from the best-performing model parameters discussed earlier (Table 3.5), mostly in terms of the Strahler order threshold at which the model predicts wetlands to be solute-rich. Wetlands with a Strahler order greater than or equal to four were predicted to have solute accumulations reflective of $CaCO_3$-enriched soils, according to the previous models. Wetlands with a Strahler order greater than or equal to five or six were predicted here to have salinity levels that pose agronomic risks. This indicates that the salinity accumulations that reflect agronomic risks require greater hydrologic contributions to develop. This makes conceptual sense as wetlands are often discussed in terms of their position along a salinity/solute-richness spectrum; their position in that spectrum is influenced by the hydrologic contributions they receive (Euliss et al., 2014).

The models for predicting the distribution of wetlands that pose agronomic salinity risks performed well but had lower accuracies (70 – 75% total accuracy, Table 3.8) as compared to the previous models' internal validation predictive accuracies (80 – 83% total accuracy, Table 3.5). The wetlands that had salinity levels representative of $CaCO_3$-enrichment but not agronomic risks (i.e. wetlands with EC values between 1000 – 2000 $\mu S\ cm^{-1}$ and EM values between 70 –

**Table 3.8 Prediction accuracies and parameters for the best-performing models for predicting the solute-richness classes based on criteria that represent agronomic risks.**

| Model | Model parameters | | | | Agronomic Risk Sample Set | | | |
|---|---|---|---|---|---|---|---|---|
| | Wetland boundary | Spill channel connectivity | S-Rich† Strahler order | Terminal status rule | Total accuracy (%) | Kappa | Fresh pr. acc.‡ n = 69 (%) | S-Rich pr. acc. n = 68 (%) |
| H | Maximum | Minimum | ≥ 6 | Not considered | 75 | 0.48 | 96 | 52 |
| I | Maximum | Moderate | ≥ 6 | Not considered | 71 | 0.40 | 79 | 62 |
| J | Maximum | Minumum | ≥ 5 | Not considered | 71 | 0.42 | 76 | 67 |
| K | Maximum | Maximum | ≥ 6 | Not considered | 70 | 0.39 | 69 | 71 |

†S-Rich = Solute-rich
‡pr. acc = producer's accuracy

100 mS $m^{-1}$), will be referred to as moderately saline wetlands. The lower predictive accuracy for these models indicates that the models had more difficulty distinguishing the moderately saline and fresh wetlands from the very saline wetlands. The models performed better distinguishing the moderately and very saline wetlands from the fresh wetlands. This is likely because the moderately saline wetlands had Strahler orders more similar to the very saline wetlands and the models could not differentiate between them. Wetland Strahler order may work best to differentiate fresh from moderately saline wetlands. Wetlands with Strahler orders of three and four seem to exhibit real differences in terms of solute-richness whereas differences between higher wetland Strahler orders (i.e. six and seven) seem more arbitrary. The differences between moderately saline and very saline wetlands may result from influences that are not fully reflected by wetland Strahler order. The models still performed well for his purpose, but these implications should be considered if attempting to differentiate between saline and very saline wetlands within models.

### 3.6 Conclusion

The proposed model used high-resolution DEMs to 1) approximate the shallow groundwater and fill-and-spill contributions wetlands receive, 2) determine if whether wetlands are terminal, and 3) quantify the relative position of wetlands within the landscape, in order to predict wetlands as either fresh or solute-rich. The model was relatively successful in its predictions. For the purposes of predicting wetlands expected to have soils enriched with $CaCO_3$, the best-

performing model achieved total predictive accuracies between 69 and 82% based on training and external validation data sets. The model parameters of Model F are recommended for mapping $CaCO_3$-enriched wetlands in PPR landscapes, as it was the best performing model overall; although, calibrating model parameters for a target area using wetland salinity data collected in that area would result in the best model performance. Mapped distributions of solute-rich wetlands generated by this model could be used to inform conservation efforts by identifying wetlands that are expected to have soils enriched with $CaCO_3$, which have a greater potential for phosphorus retention.

The model performance was improved by adapting the stream channel networks to reflect spill channels that are more likely to be active. This allowed for better representation of expected hydrologic connectivity between wetlands to estimate which wetlands have historically received the greatest contributions of solutes. Wetland terminal status was not as valuable as wetland Strahler order for predicting wetland solute-richness classes. There was little to no improvement to the model performance by separating out nested wetlands from the maximum closed topographic depressions. Based on the validation results for predicting wetland classes for the St. Denis Historic Set of observations, the model potentially underestimates the distribution of solute-rich wetlands. Therefore, if the model were used to create an upscaled estimate of wetlands with greater potential for phosphorus retention, it would represent a conservative estimate. The models indicate that the larger "gatekeeper" wetlands are expected to have $CaCO_3$-enriched soils which emphasizes the importance of their conservation for reducing phosphorus mobility within PPR watersheds. The model showed potential to be used to map wetland salinity for other purposes, although the model parameters should be adapted for the specific purpose. Modelling could be improved by incorporating site-specific salinity or climatic information to potentially address differences in site geochemistry and hydrology. The model made predictions based entirely on wetland Strahler order and terminal status. These variables could be incorporated into more complex multi-variate models to improve model performance or potentially map specific salinity values rather than binary solute-richness classes.

# 4 PREDICTIVE DIGITAL SOIL MAPPING OF WETLAND SOIL TYPES IN THE CANADIAN PRAIRIE POTHOLE REGION

## 4.1 Preface

The study presented in this chapter assessed the capability of digital soil mapping methodologies, which incorporated high-resolution DEMs to map PPR wetland soil types that differ in terms of their $CaCO_3$ content. A few key topographic attributes have been related to the distribution of PPR wetland soil types (Bedard-Haughn and Pennock, 2002; Pennock et al., 2014), but no studies have attempted to model spatial distributions of these soils using a variety of high-resolution DEM-derived topographic attributes in combination with machine-learning modelling techniques. The model proposed in Chapter 3 makes predictions of solute-richness for individual wetlands, whereas these models attempt to predict the spatial distribution of the calcareous soils *within* the individual wetlands and overall landscape. Wetland types, other than solute-rich and fresh, are discussed in this study that refer specifically to the distribution of soils within the wetlands; these are recharge, flow-through, and discharge/strongly calcareous wetlands. Although there are clear relationships between the soil distribution types and wetland solute-richness types, the terms are not used interchangeably, i.e. a solute-rich wetland is not necessarily a discharge/strongly calcareous wetland. The models in this study incorporated topographic attributes developed for the model proposed in Chapter 3, that reflect landscape-scale hydrologic processes and characteristics. Other attributes that reflect topographic variation within individual wetlands were included as predictor variables in this study. These models do not attempt to predict the concentration of $CaCO_3$ within the soil; instead, they predict the occurrence of soil classes that reflect differences in $CaCO_3$ enrichment. The resulting soil map outputs from successful models would provide area-per-hectare estimates of calcareous wetland soils which could be used to upscale estimates of phosphorus retention potentials for the PPR. Although the modelling approach used in this study differs substantially from the model proposed in Chapter 3, the resulting soil maps generated by the models could be used for the same purpose: to identify wetlands with a greater potential for phosphorus retention. These models could achieve this goal by indicating which wetlands have the greatest extents of calcareous soils.

## 4.2 Abstract

Wetlands of the Prairie Pothole Region (PPR) reduce phosphorus mobility within prairie watersheds. Wetland soils that are strongly calcareous at their surface have a greater potential for retaining phosphorus from wetland pond water. Recent studies have indicated that the distribution of calcareous wetland soils within PPR landscapes may be predictable through analyses of high-resolution digital elevation models (DEMs). Digital soil mapping methodologies were assessed for modelling the spatial distribution of wetland soil types in the PPR. Soil profiles were sampled at three Saskatchewan PPR sites: Swift Current, St. Denis, and Smith Creek. Soils were classified and mapped according to two classification schemes: 1) Calcareous Wetland, Recharge, Transition, Upland and 2) Calcareous Wetland, Non-Calcareous Wetland, Upland. Four machine-learning model types were tested including classification trees, classification trees with bagging, random forest, and multinomial logistic regression. Model predictor variables were generated from high-resolution LiDAR-derived DEMs. Knowledge-based topographic variables were developed to reflect the unique characteristics of the PPR's morphology. Various spatial resolutions and levels of smoothing were tested for the base DEMs. Models were tested using the original topographic variables and principal components as predictor variables. The models were trained through five-fold cross-validation repeated 20 times and were tested through external validation on datasets from previous studies in the St. Denis and Smith Creek areas. The best-performing models had acceptable validation predictive accuracies. The classification trees with bagging and random forest models using predictor variables derived from the 2-m DEMs with no smoothing produced maps that reflected expected soil distributions. Visual assessment identified unexpected distributions of soils in the maps generated from models based on the DEMs with greater degrees of smoothing and coarsening; this was due to issues caused by near-flat water surfaces within the DEM. No models successfully predicted the occurrence of wetlands with Calcareous Wetland soils throughout the wetland basins because samples could not be collected from more permanent wetland basin centers due to deep water inundation. However, wetlands were mapped with Transition soils throughout their basin floors which may be used to indicate potential discharge/strongly calcareous wetlands.

## 4.3 Introduction

Lake Winnipeg and other prairie waterbodies have been negatively impacted through eutrophication. Over-application of phosphorus fertilizers for agricultural production is a nonpoint source for nutrient loading of the prairie watersheds, many of which ultimately contribute to Lake Winnipeg. Wetlands of the Prairie Pothole Region (PPR) work as filters to reduce phosphorus mobility within these watersheds (Badiou et al., 2018). Wetland soils that are enriched with calcium carbonates ($CaCO_3$) are especially effective at retaining phosphorus. Although calcareous wetland soils were found to have similar total phosphorus as non-calcareous wetland soils, they were found to have six times less available phosphorus (Brown et al., 2017b). Phosphorus movement in calcareous soils is limited due to the formation of insoluble Ca-phosphates (Zhang et al., 2014). Phosphorus is retained within the calcareous wetland soils and stopped from moving into downstream waterways. This ecosystem service is likely lost with wetland drainage (Badiou et al., 2018). Understanding spatial distributions of calcareous wetland soils can inform conservation and restoration efforts to maintain this service.

A model for predicting the spatial distributions of wetlands that are expected to have the greatest accumulations of $CaCO_3$, and therefore the greatest potential for phosphorus retention, is discussed in Chapter 3. That model makes predictions based on individual wetlands but does not model the distribution of $CaCO_3$ within the wetland soils. Calcareous soils are not evenly distributed within PPR wetlands (Pennock et al., 2014). The objective of this study was to assess the suitability of digital soil mapping (DSM) methodologies to predict the spatial distributions of calcareous wetland soils in PPR landscapes. Successful predictions could provide extent estimates of calcareous wetland soils within PPR landscapes. If phosphorus storage within calcareous soils were quantified, the extent estimates could be used to establish upscaled estimates of wetland phosphorus retention potential across the PPR.

The hydropedologic influences on the distributions of wetland soils are highly complex; they include hydrologic processes occurring over landscape-scales as well as within individual wetlands. These characteristics could potentially be modelled from topographic attributes determined from high-resolution LiDAR-derived digital elevation models (DEMs). Methodologies for quantifying many topographic attributes from DEMs have been developed in

the last 30 years; these have proven to be fundamental to DSM studies for predicting spatial distributions of soil types and properties (McBratney et al., 2003). Pennock et al. (2014) and Bedard-Haughn and Pennock (2002) suggest a few topographic attributes that relate specifically to the distribution of PPR wetland soil types, such as specific dispersal area, elevation from wetland basin bottom, and position relative to maximum observed wetland water levels. The spatial distributions of wetland soil types are likely related to many topographic attributes. Machine-learning modelling techniques are well suited to determine relationships between soil classes and large numbers of variables. The relationships defined by the machine-learning models reflect complex effects and interactions of the topographic attributes as soil forming factors.

### *4.3.1 Wetland soil classes and their expected spatial distributions*

Calcium carbonates are abundant in the glacial deposits of the PPR (St. Arnaud, 1976). The distribution of $CaCO_3$ within the wetland soils is largely determined by the wetland's relationship with groundwater. Wetlands can be classified as recharge, discharge, or flow-through (Arndt and Richardson, 1988). The pond water level in recharge wetlands is at an elevation higher than the surrounding water table; the dominant movement of water in the soils directly beneath the pond is downward (Pennock et al., 2014), and $CaCO_3$ can be leached out of these soils with the downward movement (Fig. 4.1). The pond water elevation within discharge wetlands is dominantly below or level with the surrounding water table elevation. Discharge wetlands are only found in low-lying landscape positions (Lissey, 1971). Groundwater discharges into the wetland. Groundwater carries dissolved solutes, including $CaCO_3$, and distributes them throughout the discharge wetland's soil (Fig. 4.1) (Pennock et al., 2014). Wetland soils that receive solute depositions throughout their profile from the upward and lateral movement of groundwater are commonly referred to as discharge soils and are included in the Calcareous Wetland soil class within this study. Flow-through wetlands represent a transitional wetland type. Flow-through wetlands both recharge the groundwater and receive groundwater discharge depending on water table levels (Winter and Rosenberry, 1998). The soils of flow-through wetlands are characterized by having reduced solute accumulations in the upper depths of their profile compared to Calcareous Wetland soils due to leaching during periods of dominantly downward movement of water (Arndt and Richardson, 1989). These are referred to

here as Transition soils. Wetlands can also receive significant inputs of solutes from fill and spill and shallow groundwater flow from upslope wetlands (Fig. 4.1) (Cook and Hauer, 2007; Nachshon et al., 2013). This can result in $CaCO_3$-enriched soils throughout the wetland basin regardless of the wetland's relationship with groundwater (Cook and Hauer, 2007; Pennock et al., 2014). These wetlands are referred to as strongly calcareous wetlands and their soils are also included in the Calcareous Wetland soil class. The hydrologic characteristics of wetlands are complex and can change with changes in hydrologic regimes (Winter and Rosenberry, 1998). However, the accumulations and removals of $CaCO_3$ require long periods of time to occur due to their low solubility (Knuteson et al., 1989) and, therefore, the distribution of $CaCO_3$ can be used to interpret the dominant hydrologic characteristics of a wetland.

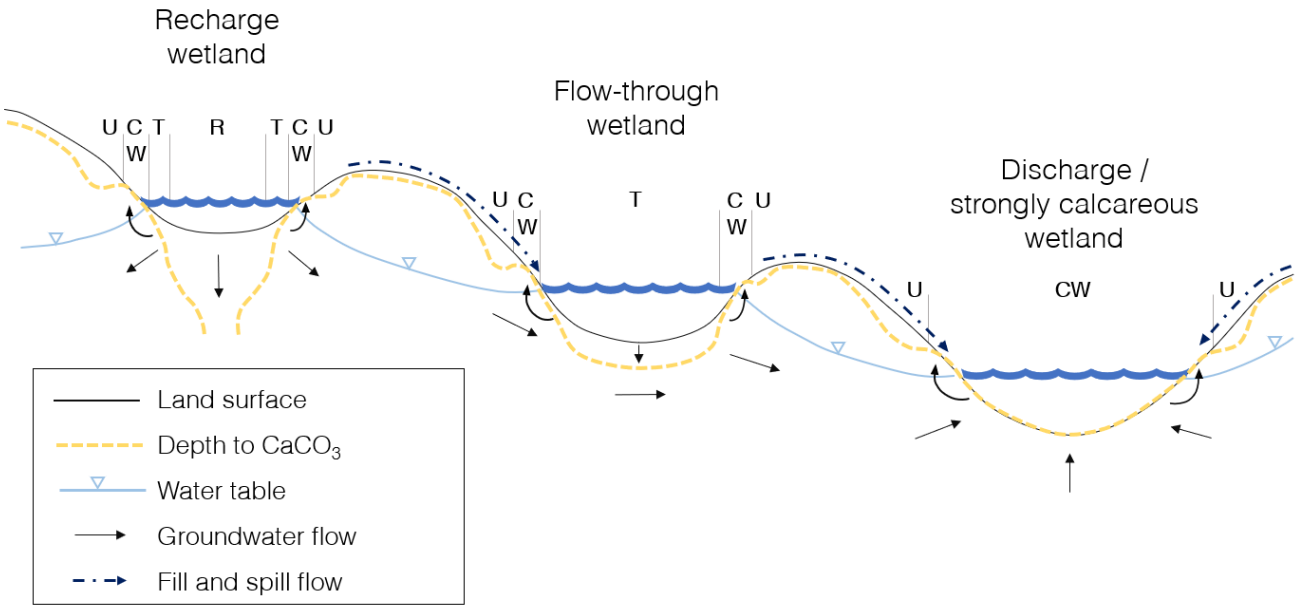



**Fig. 4.1 Schematic diagram of soil class distributions, depth to $CaCO_3$, groundwater table, and direction of groundwater and fill and spill flow based on diagrams from Van der Kamp and Hayashi (2009), Pennock (2011), and Pennock et al. (2014). Soil class distributions are indicated by the uppercase letters U = Upland, CW = Calcareous Wetland, T = Transition, and R = Recharge.**

Regardless of wetland type, a ring of calcareous discharge soils forms at the wetland fringe (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. 4.1) (Knuteson et al., 1989; Hayashi et al., 1998b; Heagle et al., 2013). Evaporation and evapotranspiration from plants in the wetland fringe

contribute to this phenomenon (Hayashi et al., 1998b). The upward movement of water to the wetland fringe redistributes $CaCO_3$ to these positions causing the development of a ring of Calcareous Wetland soils surrounding the wetlands (Fig. 4.1) (Pennock et al., 2014). In addition to describing the soils of flow-through wetlands, Transition soils are used in this study to describe the soils in the positions between recharging wetland basin floors and the discharge ring. These soils are affected by upward and lateral movement of water through most of their profile, but downward movement of water in the upper portion of their profile causes $CaCO_3$ to be leached out of the upper horizons. These are specified as a separate class here because they are not enriched with $CaCO_3$ in their upper depths of their profile to the extent of Calcareous Wetland soils, nor do they match the commonly-used concept of a Recharge soil.

Movement of dissolved phosphorus through the soil matrix is limited (Hayashi and Rosenberry, 2002; King et al., 2015) and so soil calcium within the surficial horizon has much greater potential to interact with phosphorus within wetland pond water. Calcareous Wetland soils, either within wetland fringes or within the basins of fully discharge/strongly calcareous wetlands, are characterized by accumulations of $CaCO_3$ throughout their full profile. Recharge and Transition soils lack $CaCO_3$ in the surficial depths of their profile and therefore are not expected to have the same potential for phosphorus retention. In this study, models were generated to predict the distribution of Calcareous Wetland, Recharge, Transition, and Upland soil classes. This modelling objective is referred to as 4-class mapping. Models were also generated to predict the distribution of Calcareous Wetland, Non-Calcareous Wetland, and Upland soil classes where Recharge and Transition soils were grouped into the single class, Non-Calcareous Wetland, as they are expected to have the same effect on phosphorus retention. This modelling objective is referred to as 3-class mapping.

## 4.4 Materials and methods

### *4.4.1 Study areas*

Soil samples for the DSM model training and testing were collected from study areas near Swift Current, SK; St. Denis, SK; and Smith Creek, SK. The study areas are summarized in Table 4.1. More detailed descriptions and a map of the study areas within the PPR are found in section 3.4.1. 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 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 Gleysols and eluviated and gleyed Chernozems form in the depressional positions (Pennock et al., 2011). Soils in the Smith Creek area are commonly strongly enriched with $CaCO_3$ due to the limestone-rich parent materials (Saskatchewan Soil Survey Staff, 1991). LiDAR-derived DEMs were available for all three sites (section 3.4.2).

**Table 4.1 General study area characteristics.**

| Study areas | Area extent ($km^2$) | Landform | Slope (%) | Soil salinity effects on agricultural productivity | Soil zone | Mean annual temperature¶ ($^oC$) | Mean annual precipitation¶ (mm) |
|---|---|---|---|---|---|---|---|
| Swift Current | 13 | Hummocky-dissected, undulating-dissected† | 2 - 30† | None to slight† | Brown | 4.1 | 392.5 |
| St. Denis | 6 | Hummocky‡ | 2 - 30‡ | Very slight‡ | Dark brown | 3.3 | 340.4 |
| Smith Creek | 14 | Hummocky§ | 0 - 15§ | Very slight to slight§ | Black | 1.8 | 463.5 |

†Ayres et al. (1985)
‡Acton and Ellis (1978)
§Saskatchewan Soil Survey Staff (1991)
¶Climate normals for 1981 – 2010 (Government of Canada, 2018)

### *4.4.2 Sample 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. Soil sample locations were based on individual depressions to focus the sampling in depressional areas. 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 sample 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 sample 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. Depressions were selected for sampling based on a stratified random

sample design. Individual depression boundaries were determined by the maximum closed topographic depressions within the DEM (described in section 3.4.4). Depressions were stratified based on size (Table 4.2) and relative elevation in the watershed (upper and lower) (Table 4.3). Size classes were based on Millar's (1976) size classifications for wetlands in the Canadian PPR. The smallest depressions (≤ 4047 $m^2$) were further stratified based on depth to capture very small depressions in the sampling design. These could be sampled quickly and could provide information on minimal depths required to form wetland soil characteristics. 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).

**Table 4.2 Depression size classes**

| Size class | Area† ($m^2$) | Depth (m) |
|---|---|---|
| 1 | 0 - 4047 | 0.1 - 0.3 |
| 2 | 0 - 4047 | > 0.3 |
| 3 | 4047 - 40470 | > 0.1 |
| 4 | > 40470 | > 0.1 |

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

**Table 4.3 Number of depressions sampled per study area per relative position in watershed and size stratifications.**

| Relative elevation in watershed | Size class | Study area | | |
|---|---|---|---|---|
| | | Swift Current (n) | St. Denis (n) | Smith Creek (n) |
| Upper | 1 | 3 | 3 | 3 |
| | 2 | 2 | 2 | 2 |
| | 3 | 2 | 2 | 3 |
| | 4 | 2 | 1 | 3 |
| Lower | 1 | 3 | 3 | - |
| | 2 | 2 | 2 | - |
| | 3 | 2 | 2 | - |
| | 4 | 2 | 2 | - |

Sampled depressions were located in sites with a range of land uses including cultivated, pasture, and native. Some of the smaller depressions within the cultivated sites were tilled through. The wetland soil characteristics of gleying and discharge rings are expected to persist regardless of land use. Tillage erosion can result in redistribution of $CaCO_3$ to wetland positions, but this effect can be differentiated from natural, hydrologically-influenced $CaCO_3$ distributions with field inspections. Wetland vegetation has been found to contribute to the formation of the discharge ring (Hayashi et al., 1998b). 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 the process. 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 4.3).

Randomly placed sample points focused within depressional areas are likely to end up in inundated positions making them inaccessible for sampling. The wetland water levels can change dramatically year to year and even within a season and therefore, aerial imagery could not be reliably used to inform sample point placement to avoid inundated positions. 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 (example: Fig. 4.2). 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. Transect points were sampled with a truck-mounted hydraulic corer (Giddings Machine Company Ltd., Windsor, CO) if it was accessible to do so. Otherwise, transect points were sampled by hand auger. Samples were taken upslope along the transect until soils that were certain to be Upland soils were observed.

**Fig. 4.2 Sample point transect at a depression within the Swift Current study area.**

In a study on PPR wetlands in similar regions, Pennock et al. (2014) found Recharge soils distributed throughout the 0 – 0.95 m elevation above the wetland basin bottom and discharge (Calcareous Wetland) soils distributed throughout the 0.95 – 2 m elevation above the wetland basin bottom. 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 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. Spacing between sample points depended on the depth of the depression because the height of the depression depth marked the spillover elevation. For depressions with depth 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 depth between 0.3 to 0.5 m, sample points were spaced at elevation increments of 0.25 m. For depressions with depth 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 RMSE of 0.02 m. The number of sample points per depression varied, with an average of five sample points per depression. Some of the smallest depressions only had one to three sample points, whereas some of the larger depressions had greater than ten.

Sample point locations were collected with a Trimble GeoExplorer 2005 Series GeoXT GPS (Trimble, California, U.S.A.). At each sample point, sample cores were taken to 90 cm when possible. The profiles at each sample point were described and classified according to the Canadian System of Soil Classification (CSSC) (Soil Classification Working Group, 1998). Soil samples were collected 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 wetlands were at their lowest. The sampling 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.

*4.4.3 Wetland soil type classification*

Each soil profile was classified according to the two classification schemes:

1) 4-class mapping: Calcareous Wetland (CW), Recharge (R), Transition (T), Upland (U), or buried/depositional
2) 3-class mapping: Calcareous Wetland (CW), Non-Calcareous Wetland (NCW), Upland (U), or buried/depositional

Soil classes were determined from CSSC classifications and profile descriptions (Table 4.4). The 3-class mapping classifications used the same classification criteria as the 4-class mapping except Recharge and Transition soils were grouped into the Non-Calcareous Wetland soil class.

The distribution of $CaCO_3$ within soil profiles was determined through application of 10% hydrochloric (HCl) acid during the field assessment. Soils were considered to have moderate to

**Table 4.4 Soil class criteria per 3-class and 4-class mapping objectives.**

| Soil Class | | $CaCO_3$ Distribution | Other Criteria |
|---|---|---|---|
| **3-Class** | **4-Class** | | |
| Non-Calcareous Wetland | Recharge | • Weak or no $CaCO_3$ observed in top 70 cm of profile | • Gleysols<br>• Gleyed soils |
| | Transition | • Moderate to strong $CaCO_3$ within top 70 cm of profile but weak or no carbonates in top 15 cm of profile | • Gleysols<br>• Gleyed soils |
| Calcareous Wetland | Calcareous Wetland | • Moderate to strong $CaCO_3$ throughout the profile | • Gleysols<br>• Gleyed soils<br>• Rego Chernozems that do not occur in shoulder or summit positions of a hillslope. Adjacent upland profiles must either have a B horizon or have weak or no $CaCO_3$ in their 0 - 15 cm increment<br>• 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<br>• Form in a continuous sequence from the wetland edge, i.e. once an upland soil is encountered, no Calcareous Wetland soils should be found upslope along the transect |
| Upland | Upland | • Any distribution | • Soils that do not meet criteria of other classifications |
| Buried / depositional | Buried / depositional | • Any distribution | • Soil profiles with more than 30 cm of depositional material on top of them |

strong presence of $CaCO_3$ when moderate to strong effervescence was observed with HCl application. Moderate to strong effervescence is observed when bubbles form either thick or low foam (Watson and Pennock, 2016).

The discharge ring surrounding wetlands is expected to have a gradational effect where some soils accumulate substantial contributions of $CaCO_3$ and soils in adjacent positions will accumulate $CaCO_3$ through the same processes, but not to the same degree. The Calcareous Wetland soil class was restricted to profiles with moderate to strong $CaCO_3$ throughout their profile. Recharge and Transition soils are expected to have the same potential for phosphorus retention as they both lack moderate to strong $CaCO_3$ in the surficial depth (0 – 15 cm) increment hence their grouping as Non-Calcareous Wetland soil class for the 3-class mapping modelling objective.

Gleyed soils are the most characteristic feature of wetland soils. Excessive moisture can lead to anaerobic conditions within the soil which causes iron to be reduced from $Fe^{3+}$ to $Fe^{2+}$ (Bedard-Haughn, 2011). The reduced iron is more mobile and is redistributed within the soil profile. It can be leached out of the profile completely, causing grey or even blue soil colours. It can also form red pockets of oxidized iron (mottles) within soil matrices that experience both anaerobic and aerobic conditions to allow for the reduction and oxidation of iron. Gleysols and gleyed soils were determined based on the criteria outlined in the CSSC (Soil Classification Working Group, 1998). Some of the soils in the Swift Current area had very dark colours 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 colour because the dark parent material had colours 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 were analyzed using a Bartington MS-2D meter (Bartington, Oxfordshire, U.K.) following the protocol described in de Jong et al. (2000). Profiles with horizons having potentially gleyed soil colour and magnetic susceptibility less than $150 \times 10^{-9}$ $m^3$ $kg^{-1}$ 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 $CaCO_3$ 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 ring 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 gradational effect of the discharge ring and the ubiquity of $CaCO_3$ in PPR soils, it was difficult to develop class criteria to distinguish non-gleyed Calcareous Wetland soils from adjacent Upland soils that had moderate to strong $CaCO_3$ throughout their profile. Upland soils may contain $CaCO_3$ in the surface horizon of their profile but would not be in landscape positions where they would interact with phosphorus within wetland pond water. Therefore, several soil characteristics were explored to distinguish non-gleyed Calcareous Wetland soils from Upland soils. 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 colour of the B horizon, when present. Upland soils have greater colour chroma due to the oxidation of iron in the soils, this gives the soil a more reddish colour (Fig. 4.3) (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 colours or mottling. The non-gleyed Calcareous Wetland soils would not have had the moisture conditions to form those qualities, but enough moisture to cause a reduced soil colour compared to the adjacent oxidized Upland soils. The following colour criteria for B horizons was incorporated into the classification: for non-gleyed Calcareous Chernozems to be classified as a Calcareous Wetland soil, the B horizon must have a chroma $< 5$ and a chroma at least 1 chroma less than the adjacent Upland soil B horizon.

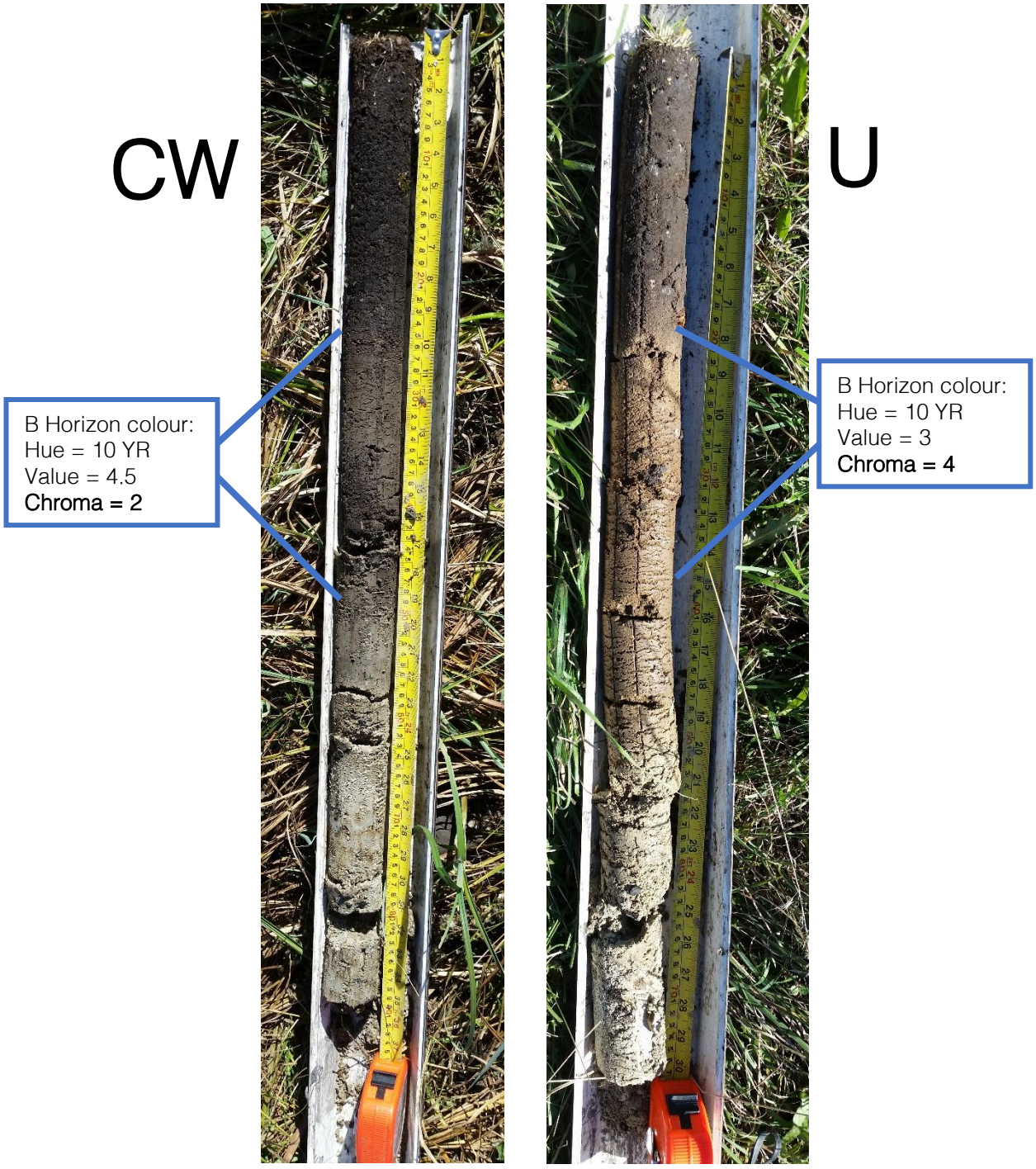



**Fig. 4.3 Comparison of B horizon colour for a non-gleyed Calcareous Wetland soil (left) and the adjacent Upland soil (right).**

Soils with greater than 30 cm of depositional material at the top of their profile were classified as buried/depositional. Buried/depositional soils were not considered in the soil class prediction models because it would not be possible to accurately predict their distribution based on the current elevation surface (Bedard-Haughn and Pennock, 2002). 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 8 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).

Tables 4.5 and 4.6 show the number of observations per 4-class and 3-class classification schemes. These observations were used to train the predictive models. There were proportionally more Upland observations than other soil classes. There were 118 wetland soil class observations total (Recharge, Transition, Calcareous Wetland) which represents a balanced number of wetland vs. upland soil class observations. There were fewer Calcareous Wetland soil class observations as compared to Recharge and Non-Calcareous Wetland soil classes, which is reasonable as they are expected to be less abundant within a typical landscape.

### *4.4.4 Spatial resolution*

The PPR landscape is characterized by small scale topographic variation. Hilltops and depressions of the hummocky landscapes can occur within five to ten meters of each other. 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 high resolutions for the soil modelling. The spatial resolution of the DEMs used to generate the predictor variables define the spatial resolution of the resulting maps. Soil mapping was tested at 2 m and

**Table 4.5 Number of soil class observations per 4-class mapping classification scheme.**

| Soil Class | External validation sets | | |
|---|---|---|---|
| | BIOCAP - St. Denis | Brown - Smith Creek | |
| | | Undrained | Drained |
| | (n) | (n) | (n) |
| Recharge | 20 | 8 | 25 |
| Transition | 0 | 2 | 3 |
| Calcareous Wetland | 28 | 0 | 4 |
| Upland | 13 | 10 | 31 |
| Total | 61 | 20 | 64 |

**Table 4.6 Number of soil class observations per 3-class mapping classification scheme.**

| Soil Class | Study Area | | | |
|---|---|---|---|---|
| | Swift Current | St. Denis | Smith Creek | Total |
| | (n) | (n) | (n) | (n) |
| Recharge | 19 | 20 | 19 | 58 |
| Transition | 5 | 8 | 8 | 21 |
| Calcareous Wetland | 6 | 20 | 13 | 39 |
| Upland | 37 | 47 | 27 | 111 |
| Total | 67 | 95 | 67 | 229 |

5 m to assess differences in predictive accuracy at varying spatial resolutions. The original LiDAR-derived 1-m DEMs were resampled using block averaging to the specified resolutions. Mapping was tested at 5-m spatial resolution, staying consistent with the Bedard-Haughn and Pennock (2002) DSM study on the distribution of PPR wetland soil types. Due to the specific sampling design used in this study to capture the extent of the discharge rings, at a resolution of 5 m, many closest sample 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. However, down-sampling reduces the numbers of errors that may be present in the elevation surface at the

higher resolutions (Lindsay, 2016), which means that the 2-m DEM likely has more features that are the result of errors than the 5-m DEMs. To correct for this, mapping was tested using 2-m DEMs without smoothing after resampling and 2-m DEMs smoothed five times using a 3 x 3 mean filter after resampling. This is based on the findings of Li et al. (2011) who recommend smoothing LiDAR-derived 1-m DEMs 10 to 20 times for depression identification within PPR landscapes to reduce erroneous features. Smoothing the 2-m DEM five times uses the same smoothing window as smoothing the 1-m DEM ten times.

### *4.4.5 Predictor variables*

A total of 32 predictor variables of topographic attributes were used in the predictive models. Predictor variable information exists for all locations of the areas to be sampled and mapped. All were derived from the DEM. The predictor variables were generated for each DEM resolution and smoothing: 2-m – not smoothed, 2-m – 5x smoothed, and 5-m – not smoothed. The predictor variables included many attributes commonly used in DSM methodologies (Table 4.7). These were calculated using the System for Automated Geoscientific Analysis (SAGA) (Conrad et al., 2015).

**Table 4.7 Descriptions of predictor variables and associated references.**

| Predictor variable | Description | Reference |
|---|---|---|
| Topographic wetness index | An index of expected moisture accumulation that considers catchment area and slope angle | Beven and Kirkby (1979) |
| SAGA wetness index | Similar to the topographic wetness index but considers a modified catchment area | Boehner et al. (2002) |
| Slope height | Elevation above the nearest stream channel determined in the DEM | Boehner and Selige (2006) |
| Normalized height | A measure of a grid cell's relative position in the local landscape which considers its vertical offset and catchment area | Boehner and Selige (2006) |
| Standardized height | The normalized height multiplied by the absolute elevation | Boehner and Selige (2006) |
| Valley depth | Elevation below the nearest ridge | Boehner and Selige (2006) |
| Mid slope position | Elevation above or below the mid-slope position of a local hill-slope | Boehner and Selige (2006) |
| Specific dispersal area | The total area of land that a grid cell contributes flow towards per unit contour. It is calculated based on inverted catchment area | Costa-Cabral and Burges (1994) |

**Table 4.7 -** continued

| | | |
|---|---|---|
| Multi-resolution valley bottom flatness index | A calculation that identifies flat valley bottoms based on elevation and slope values across a range of spatial scales | Gallant and Dowling (2003) |
| Multi-resolution ridge top flatness index | A complementary calculation to the valley bottom flatness index that identifies flat hill tops using a similar approach | Gallant and Dowling (2003) |
| Convergence index | An index reflecting if the slopes of adjacent grid cells face the target grid cell | Koethe and Lehmeier (1996) Kiss (2004) |
| Relative hydrologic slope position | Similar to the normalized height measure but a more simplified measure of a grid cell's relative position in the local landscape. It is calculated as: catchment area / (inverted catchment area + catchment area) | MacMillan (2005) |
| Slope length and steepness factor | An index which considers slope length and slope gradient | Moore et al. (1991) |
| Catchment area | The total area of land that contributes flow to a grid cell. The catchment area was calculated using a multiple flow direction algorithm which considers that water flows more than one direction from a grid cell | Quinn et al. (1991) |
| Specific catchment area | The total area of land that contributes flow to a grid cell per unit contour. It is calculated based on the catchment area | Quinn et al. (1991) |
| Terrain ruggedness index | An index that quantifies topographic heterogeneity based on the total change in elevation of a grid cell compared to its adjacent cells | Riley et al. (1999) |
| Aspect | Direction of the slope face | Zevenbergen and Thorne (1987) |
| Slope | Angle of inclination relative to the horizontal plane | Zevenbergen and Thorne (1987) |
| General curvature | A summary of curvature of the entire surface | Zevenbergen and Thorne (1987) |
| Plan curvature | The curvature along the horizontal plane. This is often referred to as the contour curvature as it reflects the curvature along a hypothetical contour line | Zevenbergen and Thorne (1987) |
| Profile curvature | The curvature in the direction of the steepest slope | Zevenbergen and Thorne (1987) |
| Tangential curvature | The curvature perpendicular to the steepest slope gradient | Zevenbergen and Thorne (1987) |
| Total curvature | A summary of curvature of the entire surface, calculated differently than general curvature | Zevenbergen and Thorne (1987) |
| Elevation | Meters above sea level | - |

**Table 4.7 -** continued

| | | |
|---|---|---|
| Elevation percentile | A grid cell's elevation percentile in relation to the entire study area watershed | This study |
| Elevation from depression spillover | Elevation above or below the associated depression's spillover elevation | This study |
| Elevation above basin bottom | Elevation above the associated depression bottom elevation | This study |
| Elevation above basin bottom / depth | Elevation above the associated depression's bottom elevation divided by depression depth | This study |
| Depression depth | Elevation difference between a depression's spillover elevation and bottom | This study |
| Depression max catchment area | The maximum catchment area of a depression ascribed to all cell's within a depression | This study |
| Wetland Strahler order minimum | A measure of a wetland's hydrologic position in the watershed, considering minimum hydrologic connectivity | This study |
| Wetland Strahler order maximum | A measure of a wetland's hydrologic position in the watershed, considering maximum hydrologic connectivity | This study |

Several topographic attribute measures were developed to better represent the unique morphological characteristics of the PPR that were expected to relate to wetland soil type distributions. Detailed descriptions on the GIS methodologies for calculating each new topographic attribute are included in Appendix B. *Elevation percentile* determines each grid cell's elevation percentile in relation to the entire study area watershed. Fully discharge wetlands are expected in lower positions within a watershed. This variable may help to capture these features. Three of the variables were based on a grid cell's position relative to a local depression. These were developed to potentially capture relationships identified by Pennock et al. (2014) between soil class distribution and topographic position in relation to depression bottoms and maximum water levels. The contributing basins were determined for each closed topographic depression with depths greater than 10 cm. 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 bottom (Fig. 4.4) and elevation above or below the depression's spillover elevation. The *Depression depth* variable ascribes a depression's depth value to the depression's contributing basin. The measure of *Elevation above depression bottom divided by depression depth* was included to potentially capture differences in relationships of soil distribution for wetlands of varying sizes.

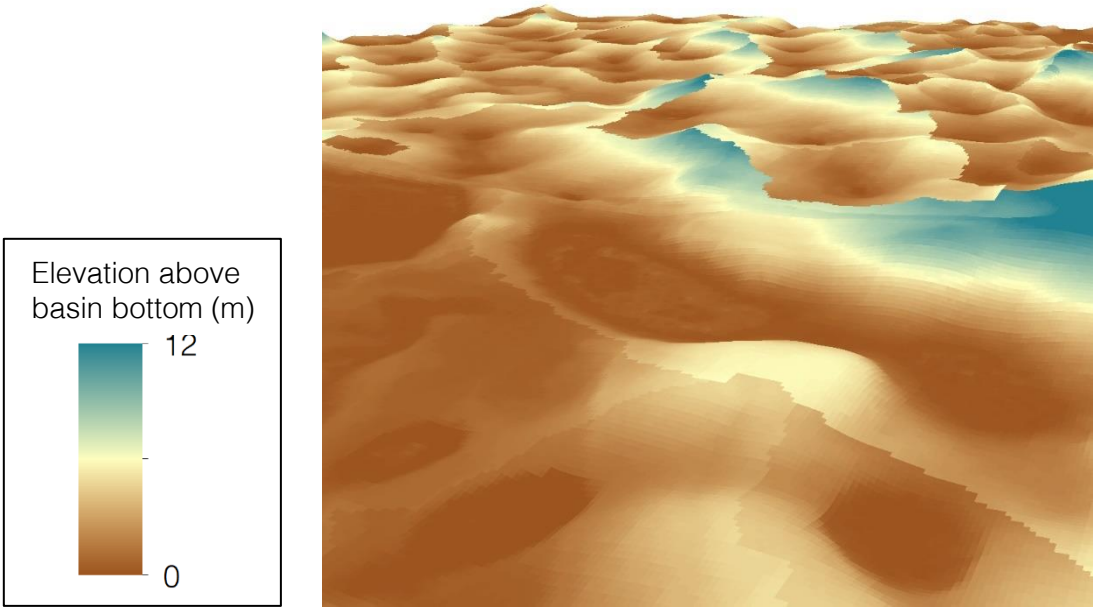



**Fig. 4.4 3D representation of the topographic predictor variable *Elevation above basin bottom* for a portion of the St. Denis study area. The attribute reflects the meters above the associated depression's bottom.**

Several variables were developed to reflect the characteristics of individual depressions including *Depression max catchment area, Wetland Strahler order minimum,* and *Wetland Strahler order maximum* (Fig. 4.5). These characteristics were quantified for each closed topographic depression. All cells within a closed topographic depression were given the same value according to the characteristics of that depression. Cells outside of the closed topographic depressions were given values of zero. Wetland Strahler orders were determined following the methodologies described in section 3.4. The two measures vary in the expected connectivity between wetlands.

Because the predictor variables were all derived from the DEM, it is expected that many of them were collinearly related. Principal component analysis (PCA) can improve model performance by reducing the number of collinear variables. There have been mixed results in terms of the effect of using principal components in random forest models (Svetnik et al., 2004; Xiong et al., 2012). Non-penalized multinomial logistic regression models (as used in this study) are sensitive to overfitting with too many available predictor variables (Hastie et al., 2009). Reducing the number of predictor variables through PCA may reduce the likelihood of the multinomial logistic regression models overfitting. PCA also helps to reduce computational requirements. Models

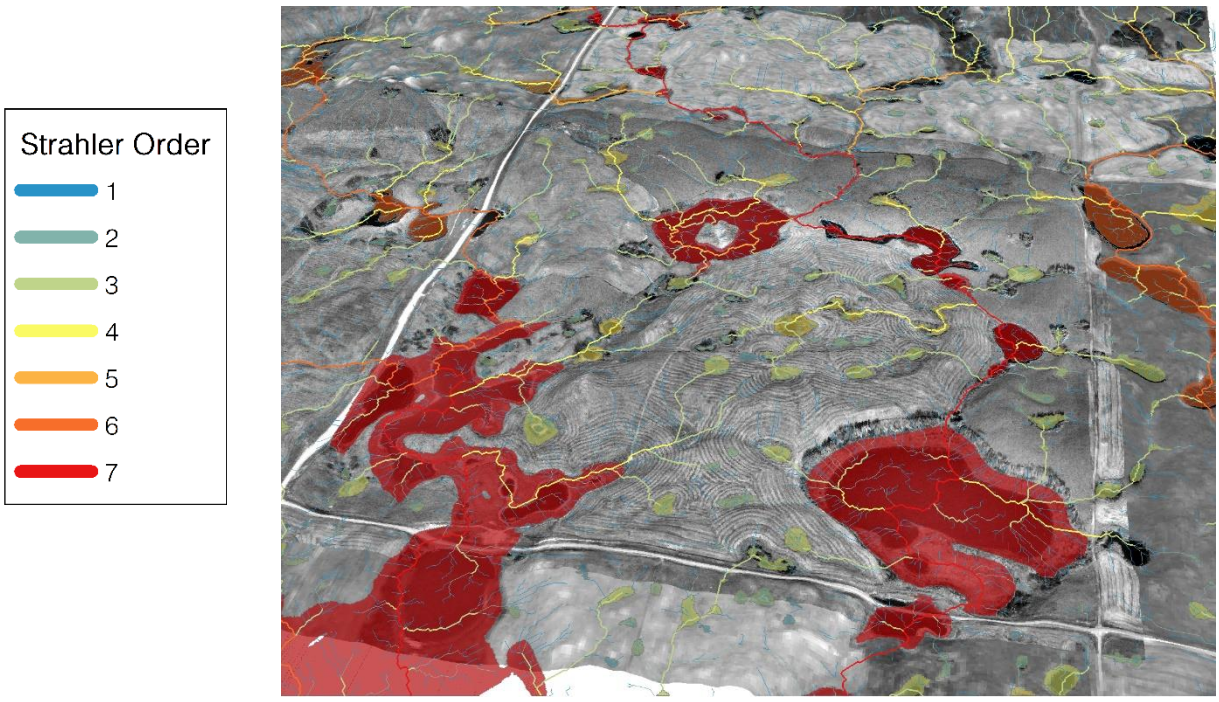



**Fig. 4.5 3D representation of the topographic predictor variable *Wetland Strahler order maximum* for a portion of the St. Denis study area. The stream channel lines are not included in the final variable raster surface, only the values for each depression polygon.**

were tested using the original 32 topographic variables as predictor variables and using principal components of the topographic variables as predictor variables. The number of principal components used accounted for 95% of the total variance within the original topographic variables. For the 2-m – not smoothed DEM, the original 32 topographic attributes were reduced to 20 principal components and for the 2-m – 5x smoothed and 5-m – not smoothed DEMs, they were reduced to 19 principal components.

*4.4.6 Machine-learning techniques*

Four machine-learning techniques were tested to model the distributions of soil classes. These included classification trees, classification trees with bagging, random forest, and multinomial logistic regression.

A classification tree is a decision tree of if-then rules based on the inputted predictor variables to predict soil classes. The rules are generated based on characteristics between predictor variables and soil classes as seen in the training observations. At each step in the decision tree, the soil observations are split per the predictor variable value that results in the greatest impurity

reduction between groups of observations (Strobl et al., 2009). After each split, the groups of observations become more homogeneous. Finally, the terminal nodes or leaves (where no further splitting occurs) represent a mostly uniform group of observations. The model is applied to new observations with the same predictor variable information to predict the classes of those new observations. Classification trees can develop enough splits to perfectly classify the training data, which would result in an overfitted model that would not perform well for predicting classes of new observations. The number of splits is limited to ensure the model is not overfitted (Venables and Ripley, 2002).

Classification trees with bagging is an ensemble classification tree model that involves the creation of many classification trees based on bootstrap sample sets of the training data. In this study, for each bagged classification tree model, 1000 trees were created. Each tree is based on a different 70% of the training data. After the creation of the ensemble of trees, the model makes predictions for new observations by running them through each tree. Each tree casts its "vote" on what class it predicts the new observation to be and the class that receives the most votes is selected as the prediction. Random forest is another form of an ensemble classification tree model. More diverse trees are grown by limiting the number of variables available to make each individual splitting rule (Strobl et al., 2009).

Multinomial logistic regression is a generalized linear model used to predict the occurrence of soil classes. It is an extension of binomial logistic regression which models the relationship between predictor variables and the probability of soil class occurrence between 0 and 1 using the logit link function (Kempen et al., 2009):

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \eta_i$$

where $p_i$ is the probability of occurrence of soil class $i$ and $\eta_i$ is the vector of predictor variables and associated coefficients that define the linear regression relationship for soil class $i$ to logit($p_i$). In the multinomial case, where n is the number of soil classes, $\eta$ is determined for each soil class and $p_i$ is determined by:

$$p_i = \frac{\exp(\eta_i)}{\exp(\eta_1) + \exp(\eta_2) + \ \ldots + \exp(\eta_n)}$$

The soil class with the highest probability of occurrence is selected as the predicted class.

*4.4.7 Model tuning and cross-validation*

Some of the machine-learning techniques have parameters that can be tuned to improve predictive accuracy. The R-package *caret* (Kuhn, 2008) was used to optimize model parameters. *caret* uses k-folds cross-validation which splits the training data into k number of folds (five folds were used in this study). The model is trained using four of the five folds and then tested on the 5th fold. The process is repeated so that each fold is used as the testing fold. This was repeated 20 times where the data was split into different sets of five folds and the accuracies from each cross-validation were averaged. This whole process is repeated for each tuning parameter value to determine the tuning parameter value with the highest cross-validation accuracy. This process not only optimizes the tuneable model parameters, but it also gives a cross-validated estimate of the model accuracy.

As mentioned, the classification trees could potentially create enough splits to classify every observation of the training dataset so that there are no impurities among the observations in each terminal node. This model would be overfitted and would not perform well predicting for new data. The tuneable model parameter of the classification tree models was the complexity parameter cp, which controls the number of splits made by the classification tree. The number of splits per trees is not controlled for the classification trees with bagging or random forest models. Each random forest and bagged classification model consisted of 1000 trees. There were no optimizable parameters for the classification trees with bagging or multinomial logistic regression models, but their cross-validation predictive accuracy was still determined through the k-folds cross-validation. The number of predictor variables available for each split (mtry) in the random forest models was optimized through the cross-validation.

*4.4.8 External validation datasets*

Two datasets of soil observations from previous studies were used as external validation sets to test the predictive accuracy of the models on observations that were not used to train the models. 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 2 transects, each with 10 – 12 points per transect. The transects extended from the upland, through the wetland center, to the upland on the other side of the wetland. 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. The Brown dataset consisted of samples from 42 wetlands, however, 32 of those wetlands had been drained at some point. The data from the undrained and drained wetlands were separated into two validation sets. The set of drained wetland soil observations allows for assessment of whether the models are applicable in these types of landscapes. Each of the 42 wetlands were 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 external validation dataset per modelling objective are shown in Table 4.8 and Table 4.9.

**Table 4.8 Number of soil class observations according to the 4-class mapping classification scheme per external validation set.**

| Soil Class | External validation sets | | |
|---|---|---|---|
| | BIOCAP - St. Denis | Brown - Smith Creek | |
| | | Undrained | Drained |
| | (n) | (n) | (n) |
| Recharge | 20 | 8 | 25 |
| Transition | 0 | 2 | 3 |
| Calcareous Wetland | 28 | 0 | 4 |
| Upland | 13 | 10 | 31 |
| Total | 61 | 20 | 64 |

**Table 4.9 Number of soil class observations according to the 3-class mapping classification scheme per external validation set.**

| Soil Class | External validation sets | | |
|---|---|---|---|
| | BIOCAP - St. Denis | Brown - Smith Creek | |
| | | Undrained | Drained |
| | (n) | (n) | (n) |
| Non-Calcareous Wetland | 20 | 10 | 28 |
| Calcareous Wetland | 28 | 0 | 4 |
| Upland | 13 | 10 | 31 |
| Total | 61 | 20 | 63 |

There were six observations in the external validation datasets that were difficult to classify as either Calcareous Wetland or Upland soils. The soil colour data required to distinguish them as Upland or non-gleyed Calcareous Wetland soils did not exist, so these observations were removed from the external validation datasets.

Two methods of external validation testing were used. The mapping was done at a high resolution to maintain a high level of detail in the DEM and not because it is the objective of the study to map the soil class distributions so specifically. To allow for some spatial uncertainties between the GPS positions of the observations and predictor variables, r = 0 cell and r = 1 cell methods of validation were used for the external validation tests (Heung et al., 2014; Nauman and Thompson, 2014; Vincent et al., 2018). r = 0 cell validation determines if the observed soil class matches the prediction at a single grid cell location. In r = 1 validation, an observation is considered correctly predicted for if its soil class matches the predictions at the single grid cell location or any of the eight adjacent grid cell locations.

### *4.4.9 Selection of best-performing models*

Models were generated for the two modelling objectives: 4-class and 3-class mapping. The best-performing models per modelling objective were selected based on their predictive accuracies for the cross-validation and the r = 0 cell external validation. These include the four models with the highest accuracy based on the cross-validation, the four models with the highest accuracy based on the BIOCAP external validation, the four models with the highest accuracy based on the

Brown – Undrained external validation, and the four models with the highest average predictive accuracy of all three validations. The selection of the best-performing models did not consider the predictive accuracy based on the Brown – Drained external validation.

The predictive accuracies based on the cross and external validations provide a summary of the best-performing models among the many combinations of model parameters. However, before deciding which models performed best overall, it is crucial to ensure that the outputted maps of the soil distributions make conceptual sense. This is especially true for this study, where the distribution of the wetland soil types is conceptually well understood but has been difficult to quantify. Maps were generated for each site from each of the best-performing models. The maps were visually inspected to determine if the distribution of soils matched the concepts described in section 4.3.1.

## 4.5 Results

### *4.5.1 Tuned model parameters and cross-validation results*

Table 4.10 shows the model parameters and the predictive accuracies based on the 5-fold cross-validation repeated 20 times for the best-performing models for the two modelling objectives. The model parameters include the spatial resolution and level of smoothing used for the underlying DEM that all predictor variables were derived from, whether original topographic variables or principal components were used as predictor variables, the machine-learning technique, and the machine-learning parameter value optimized through the cross-validation. The best-performing models included a range of model parameters. Overall predictive accuracy and Kappa scores are reported. Cross-validation predictive accuracy of the best-performing models for 4-class mapping ranged from 60 to 72% (Table 4.10). The Kappa scores ranged between 0.33 to 0.57 which correspond to fair to moderate levels of agreement between predictions and observations (Landis and Koch, 1977). The cross-validation predictive accuracy for the best-performing models for 3-class mapping were higher and ranged from 64 to 77% accurate. These Kappa scores ranged from 0.4 to 0.63 which correspond to moderate to substantial levels of agreement between predictions and observations. The lower accuracy for 4-class mapping is largely a result of error in differentiating the Transition and Recharge soil classes that were grouped as the single Non-Calcareous Wetland class for the 3-class mapping objective.

**Table 4.10 Model parameters and cross-validation predictive accuracy of the best-performing models per model objective.**

| Modelling Objective | Model code† | Model parameters | | | | Training data cross-validation | |
|---|---|---|---|---|---|---|---|
| | | Base DEM | Predictor variables‡ | Machine-learning technique§ | Optimized tuning parameter | Accuracy (%) | Kappa |
| 4-class | 4C_2mNoSm_OV_treebag | 2m no smooth | Original | Bagged class tree | - | 71 | 0.56 |
| | 4C_2mNoSm_OV_RF | | Original | Random forest | mtry = 21 | 70 | 0.55 |
| | 4C_2mNoSm_PC_ctree | | Prin. comp. | Class tree | cp = 0.200623 | 60 | 0.33 |
| | 4C_2mNoSm_PC_RF | | Prin. comp. | Random Forest | mtry = 4 | 62 | 0.39 |
| | 4C_2m5xSm_OV_ctree | 2m 5x smooth | Original | Class tree | cp = 0.035195 | 68 | 0.50 |
| | 4C_2m5xSm_OV_treebag | | Original | Bagged class tree | - | 72 | 0.57 |
| | 4C_2m5xSm_OV_RF | | Original | Random forest | mtry = 28 | 71 | 0.57 |
| | 4C_2m5xSm_PC_ctree | | Prin. comp. | Class tree | cp = 0.058803 | 63 | 0.41 |
| | 4C_5mNoSm_OV_ctree | 5m no smooth | Original | Class tree | cp = 0.012798 | 63 | 0.43 |
| | 4C_5mNoSm_OV_treebag | | Original | Bagged class tree | - | 71 | 0.56 |
| | 4C_5mNoSm_PC_RF | | Prin. comp. | Random forest | mtry = 5 | 68 | 0.50 |
| 3-class | 3C_2mNoSm_OV_treebag | 2m no smooth | Original | Bagged class tree | - | 76 | 0.60 |
| | 3C_2mNoSm_OV_RF | | Original | Random forest | mtry = 18 | 76 | 0.61 |
| | 3C_2mNoSm_PC_ctree | | Prin. comp. | Class tree | cp = 0.03459 | 64 | 0.40 |
| | 3C_2mNoSm_PC_RF | | Prin. comp. | Random forest | mtry = 6 | 71 | 0.51 |
| | 3C_2m5xSm_OV_treebag | 2m 5x smooth | Original | Bagged class tree | - | 77 | 0.62 |
| | 3C_2m5xSm_OV_RF | | Original | Random forest | mtry = 12 | 77 | 0.63 |
| | 3C_2m5xSm_PC_MLR | | Prin. comp. | Multinom. log. reg. | - | 73 | 0.57 |
| | 3C_5mNoSm_OV_treebag | 5m no smooth | Original | Bagged class tree | - | 75 | 0.59 |
| | 3C_5mNoSm_OV_RF | | Original | Random forest | mtry = 22 | 75 | 0.59 |
| | 3C_5mNoSm_PC_MLR | | Prin. comp. | Multinom. log. reg. | - | 69 | 0.50 |

†Models are referred to by their model code in the following results and discussion. It reflects the model objective: 4C = 4-class mapping, 3C = 3-class mapping; and model parameters: base DEM: 2mNoSm = 2-m – not smoothed, 2m5xSm = 2-m – 5x smoothed, 5mNoSm = 5-m – not smoothed; predictor variables: OV = original topographic variables, PC = principal components; and machine-learning technique: ctree = classification tree, treebag = classification tree with bagging, RF = random forest, MLR = multinomial logistic regression
‡Original = original topographic variables, Prin. comp. = principal components
§Multinom. log. reg. = multinomial logistic regression

### *4.5.2 External validation results*

The predictive accuracies for the BIOCAP – St. Denis, Brown – Smith Creek – Undrained Wetlands, and Brown – Smith Creek – Drained Wetlands external validations for the best-performing models are shown in Tables 4.11 and 4.12. Overall predictive accuracy and Kappa scores are reported for r = 0 cell and r = 1 cell validation methods. The producer's and user's accuracy are reported for the Calcareous Wetland soil class observations only for the BIOCAP – St. Denis external validation dataset because there were few observations (n = 4) within the Brown – Smith Creek datasets.

#### *4.5.2.1 BIOCAP – St. Denis external validation*

The predictive accuracies for the BIOCAP – St. Denis external validation were lower than the predictive accuracies for the cross-validation and the other external validation tests. The best-performing models for 4-class mapping had predictive accuracies between 48 to 73% and Kappa scores between 0.28 to 0.57 according to the r = 0 cell validation method. As expected, the predictive accuracies and Kappa scores were higher according to the r = 1 cell validation method and ranged between 55 to 87% accurate with Kappa scores of 0.37 to 0.79. The producer's accuracy for Calcareous Wetland observations were low for each model and ranged from 0 to 47% accurate, whereas the user's accuracy was high and ranged from 69 to 100% accurate (excluding the 4C_2mNoSm_PC_ctree model). This indicates that the models underestimate the amount of Calcareous Wetland soils but, when they were predicted, they were confidently predicted as such. The producer's and user's accuracy for the Calcareous Wetland observations for the 3-class mapping models were similar with the exception of the 3C_5mNoSm_PC_MLR model which had both high producer's accuracy (75%) and user's accuracy (81%). The overall predictive accuracies for the BIOCAP – St. Denis external validation were similar between the two modelling objectives (Table 4.11 and 4.12). This is because there were no Transition observations in this dataset and none of the 4-class mapping models predicted there to be Transition soil classes within this dataset.

**Table 4.11 External validation predictive accuracy metrics for the best-performing 4-class mapping models.**

| Model code | BIOCAP - St. Denis | | | | | | Brown - Smith Creek | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Undrained | | | | Drained | | | |
| | r = 0 cell | | | | r = 1 cell | | r = 0 cell | | r = 1 cell | | r = 0 cell | | r = 1 cell | |
| | Accuracy | Kappa | Calcareous W. obs.† | | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa |
| | | | Prod. acc.‡ | User acc.§ | | | | | | | | | | |
| | (%) | | (%) | (%) | (%) | | (%) | | (%) | | (%) | | (%) | |
| 4C_2mNoSm_OV_treebag | 60 | 0.46 | 22 | 75 | 69 | 0.58 | 75 | 0.60 | 80 | 0.66 | 70 | 0.54 | 78 | 0.65 |
| 4C_2mNoSm_OV_RF | 60 | 0.47 | 22 | 86 | 69 | 0.58 | 70 | 0.53 | 85 | 0.74 | 75 | 0.61 | 78 | 0.65 |
| 4C_2mNoSm_PC_ctree | 48 | 0.28 | 0 | 0 | 55 | 0.37 | 75 | 0.56 | 85 | 0.73 | 77 | 0.58 | 86 | 0.74 |
| 4C_2mNoSm_PC_RF | 58 | 0.42 | 18 | 100 | 66 | 0.52 | 80 | 0.64 | 90 | 0.82 | 89 | 0.81 | 93 | 0.86 |
| 4C_2m5xSm_OV_ctree | 73 | 0.57 | 61 | 74 | 87 | 0.79 | 50 | 0.26 | 60 | 0.37 | 64 | 0.37 | 69 | 0.44 |
| 4C_2m5xSm_OV_treebag | 66 | 0.51 | 47 | 69 | 73 | 0.61 | 56 | 0.26 | 70 | 0.48 | 70 | 0.46 | 74 | 0.52 |
| 4C_2m5xSm_OV_RF | 64 | 0.50 | 40 | 69 | 78 | 0.67 | 60 | 0.35 | 80 | 0.65 | 69 | 0.44 | 75 | 0.55 |
| 4C_2m5xSm_PC_ctree | 61 | 0.44 | 25 | 100 | 68 | 0.53 | 75 | 0.56 | 75 | 0.56 | 86 | 0.74 | 86 | 0.74 |
| 4C_5mNoSm_OV_ctree | 64 | 0.50 | 40 | 69 | 71 | 0.59 | 50 | 0.19 | 70 | 0.45 | 59 | 0.27 | 66 | 0.38 |
| 4C_5mNoSm_OV_treebag | 63 | 0.50 | 33 | 82 | 73 | 0.61 | 75 | 0.59 | 80 | 0.66 | 72 | 0.55 | 83 | 0.72 |
| 4C_5mNoSm_PC_RF | 60 | 0.47 | 33 | 100 | 71 | 0.60 | 75 | 0.54 | 85 | 0.73 | 77 | 0.59 | 81 | 0.66 |

†Calcareous W. obs. = Calcareous Wetland observations
‡Prod. acc. = producer's accuracy
§User acc. = user's accuracy

**Table 4.12 External validation predictive accuracy metrics for the best-performing 3-class mapping models.**

| Model code | BIOCAP - St. Denis | | | | | | Brown - Smith Creek | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Undrained | | | | Drained | | | |
| | r = 0 cell | | | | r = 1 cell | | r = 0 cell | | r = 1 cell | | r = 0 cell | | r = 1 cell | |
| | Accuracy | Kappa | Calcareous W. obs.† | | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa |
| | | | Prod. acc.‡ | User acc.§ | | | | | | | | | | |
| | (%) | | (%) | (%) | (%) | | (%) | | (%) | | (%) | | (%) | |
| 3C_2mNoSm_OV_treebag | 61 | 0.43 | 18 | 84 | 71 | 0.57 | 80 | 0.67 | 90 | 0.82 | 75 | 0.59 | 85 | 0.73 |
| 3C_2mNoSm_OV_RF | 60 | 0.41 | 18 | 84 | 71 | 0.57 | 80 | 0.67 | 95 | 0.90 | 83 | 0.71 | 88 | 0.78 |
| 3C_2mNoSm_PC_ctree | 64 | 0.46 | 40 | 79 | 81 | 0.70 | 85 | 0.70 | 95 | 0.90 | 81 | 0.65 | 91 | 0.82 |
| 3C_2mNoSm_PC_RF | 63 | 0.46 | 18 | 100 | 66 | 0.50 | 90 | 0.80 | 100 | 1.00 | 93 | 0.85 | 94 | 0.88 |
| 3C_2m5xSm_OV_treebag | 61 | 0.40 | 36 | 67 | 73 | 0.58 | 80 | 0.60 | 85 | 0.70 | 81 | 0.64 | 81 | 0.64 |
| 3C_2m5xSm_OV_RF | 64 | 0.47 | 25 | 100 | 73 | 0.59 | 95 | 0.90 | 95 | 0.90 | 88 | 0.76 | 91 | 0.82 |
| 3C_2m5xSm_PC_MLR | 64 | 0.46 | 36 | 77 | 76 | 0.63 | 90 | 0.82 | 100 | 1.00 | 77 | 0.62 | 86 | 0.75 |
| 3C_5mNoSm_OV_treebag | 66 | 0.49 | 36 | 84 | 76 | 0.63 | 85 | 0.73 | 95 | 0.90 | 78 | 0.63 | 88 | 0.77 |
| 3C_5mNoSm_OV_RF | 60 | 0.41 | 22 | 86 | 76 | 0.64 | 90 | 0.82 | 95 | 0.90 | 81 | 0.67 | 88 | 0.77 |
| 3C_5mNoSm_PC_MLR | 78 | 0.65 | 75 | 81 | 87 | 0.80 | 70 | 0.48 | 75 | 0.55 | 83 | 0.69 | 91 | 0.83 |

†Calcareous W. obs. = Calcareous Wetland observations
‡Prod. acc. = producer's accuracy
§User acc. = user's accuracy

#### *4.5.2.2 Brown – Smith Creek – Undrained external validation*

The best-performing models for 4-class mapping had predictive accuracies between 50 to 80% and Kappa scores between 0.19 to 0.64 according to the r = 0 cell validation method for the Brown – Smith Creek – Undrained external validation. The predictive accuracies and Kappa scores according to the r = 1 cell validation method ranged between 60 to 90% and 0.37 to 0.82. The models for 3-class mapping had higher predictive accuracy which ranged from 70 to 95% with Kappa scores between 0.55 to 1.0 according to the r = 0 cell validation method. Most of these models were able to successfully predict the soil classes of this external validation dataset, however, there were no Calcareous Wetland observations within this dataset. This external validation test provided an assessment of whether the models overestimated the extent of Calcareous Wetland soil classes and the results indicate that they did not.

#### *4.5.2.3 Brown – Smith Creek – Drained external validation*

The models were similarly successful in their predictions for the Brown – Smith Creek – Drained external validation set. The 4-class mapping models had predictive accuracies between 61 to 89% according to the r = 0 cell validation method and the 3-class mapping models predictive accuracies between 70 to 93%. This indicates that the distribution of wetland soils may still be predictable within mechanically drained wetlands.

### *4.5.3 Visual assessment*

Fig. 4.6 and 4.7 show three example wetland series from each study site and the predicted soil maps from a few of the best-performing models. The predicted extents of wetland soils vary from site to site: Swift Current had the smallest extents of wetland soils and Smith Creek had the greatest extents of wetland soils, which is consistent with what was observed in the field sampling.

#### *4.5.3.1 Visual assessment of 4-class maps*

The 4C_2mNoSm_OV_RF (Fig. 4.6d-f) and 4C_2mNoSm_OV_treebag (not shown) models generated similar, acceptable-looking maps. Calcareous Wetland soil rings were consistently mapped surrounding the wetland basins. Transition soils were often mapped between the Recharge basins and the Calcareous Wetland ring as expected (most apparent in Fig. 4.6f).

Upland soils were restricted to upland landscape positions. The wetland basin floors were mapped as either Recharge or Transition soils. There was some speckling of Recharge and Transition soils within wetland basin floors, which is not expected. Other than the transition from Recharge to Transition or Calcareous Wetland soils at the edge of the basin, wetland basin floors are expected to have uniform distributions of soils throughout them (Pennock et al., 2014). This speckling issue was far less common in the maps generated from these models (4C_2mNoSm_OV_RF and 4C_2mNoSm_OV_treebag) compared to most of the other 4-class mapping model outputs. Another issue was that no models mapped wetlands as having Calcareous Wetland throughout their basins. Field observations in previous studies confirm that there are several known discharge wetlands in the lower portion of the St. Denis watershed (Pennock et al., 2014). These same wetlands were consistently mapped as having Transition soils throughout the basin floors which would indicate that these are flow-through wetlands. Some recharge wetlands were mapped as having Transition soils throughout their basin floor, as is seen in the largest wetland pictured in the Swift Current map (Fig. 4.6), but most recharge wetlands were mapped with Recharge soils throughout their basin floor. Aside from the minor speckling issue and mapping the discharge wetland basins at St. Denis with Transition soils, the maps from these models matched the conceptual understanding of the distribution of soil types in these landscapes.

The random forest and bagged classification tree models based on the 2-m – 5x smoothed and the 5-m – not smoothed DEMs using the original variables generated similar-looking maps (4C_2m5xSm_OV_treebag, shown Fig. 4.6g-i) as those discussed above. However, for many of the wetlands at St. Denis, Calcareous Wetland soils were speckled throughout the wetland basins (seen in Fig. 4.6h). This issue was far more common in these maps than in the maps generated by the 4C_2mNoSm_OV_RF and 4C_2mNoSm_OV_treebag models discussed above.

The 4C_5mNoSm_OV_ctree model did map some wetlands with Calcareous Wetland soils throughout their basins (Fig. 4.6k) but it was not consistent in its mapping of discharge/strongly calcareous wetlands. There were several other issues associated with these outputted maps, including Upland soils being mapped in wetland landscape positions. The maps generated from models built using principal components consistently had soil classes in more scattered

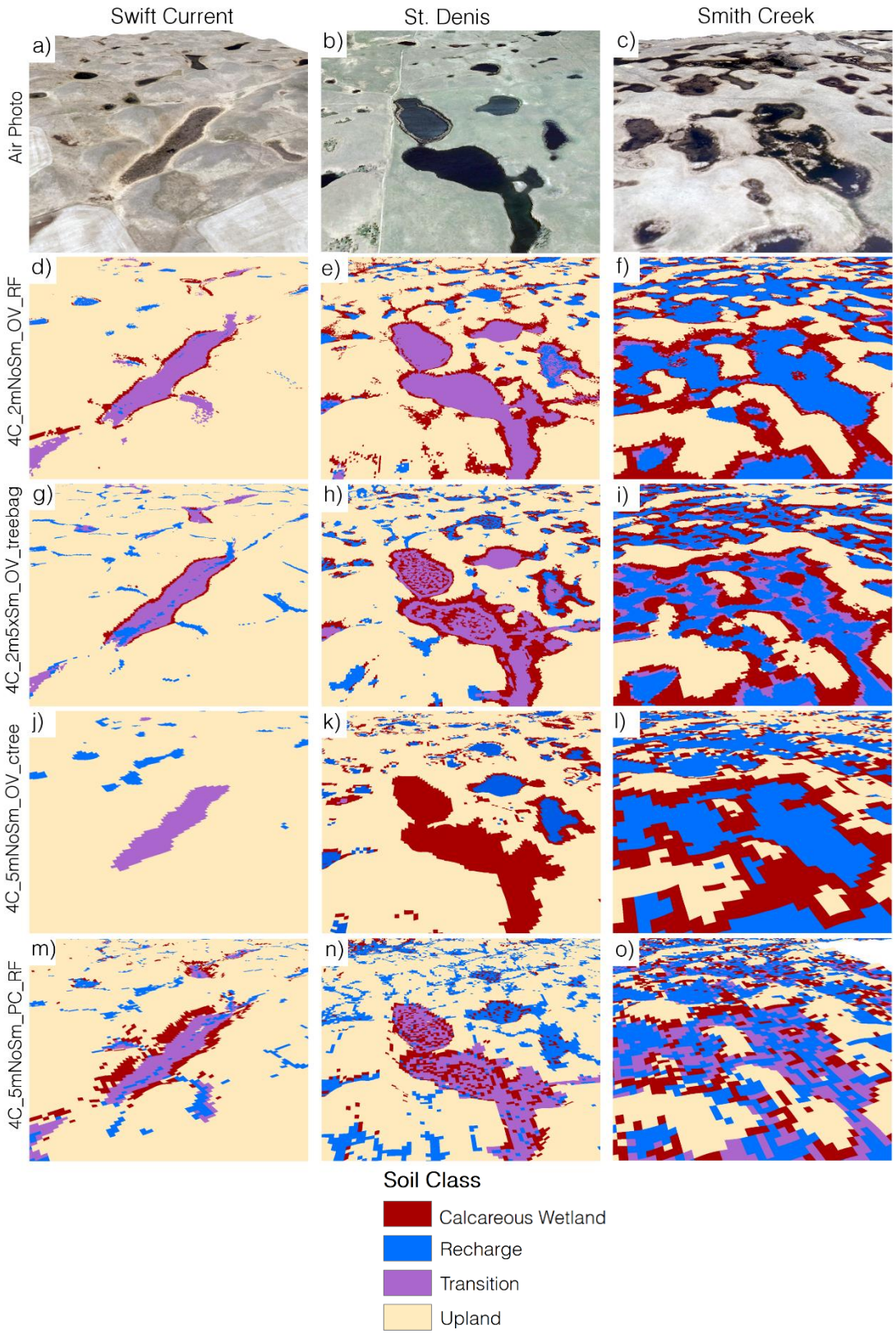



**Fig. 4.6 3D representations of aerial imagery and 4-class map outputs from select best-performing models for example wetlands from each site.**

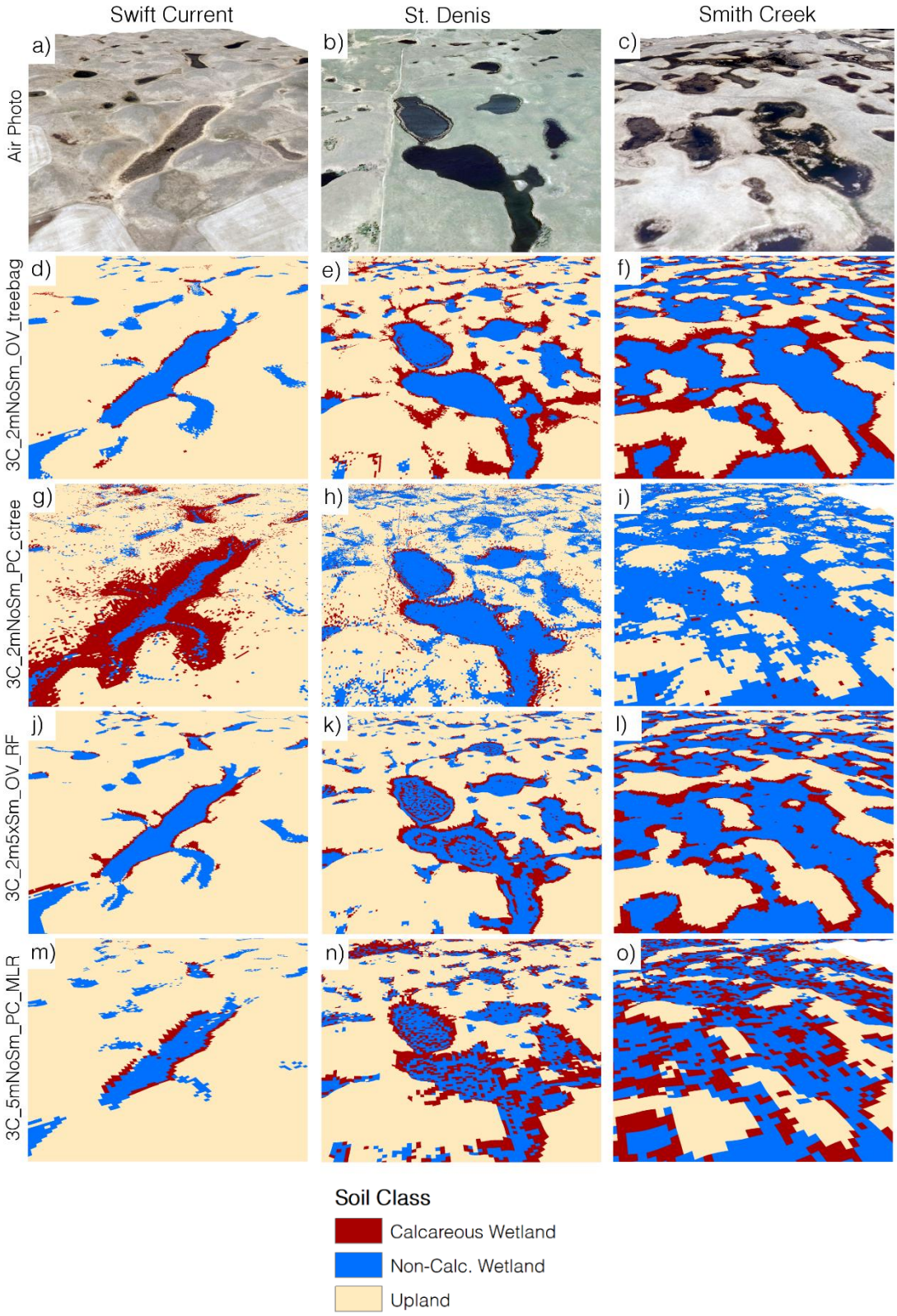



**Fig. 4.7 3D representations of aerial imagery and 3-class map outputs from select best-performing models for example wetlands from each site. Calc. = Calcareous.**

distributions including wetland soil classes mapped in Upland positions (4C_5mNoSm_PC_RF shown Fig. 4.6m-o.).

*4.5.4.2 Visual assessment of 3-class maps*

The 3C_2mNoSm_OV_RF (Fig. 4.7d-f) and 3C_2mNoSm_OV_treebag outputs were similar to the 4C_2mNoSm_OV_treebag (Fig. 4.6d-f) and 4C_2mNoSm_OV_RF discussed above, except that the Recharge and Transition soils were predicted as Non-Calcareous Wetland soils. All wetland basin floors were mapped with Non-Calcareous Wetland soils. There was no issue of soil class speckling within the wetland basins in the maps generated by these two particular models. Again, no wetlands were mapped with Calcareous Wetland soils throughout the wetland basin. Similar to their 4-class mapping equivalent models, apart from missing these features, the maps generated from these models were consistent with the conceptual understanding of the distribution of wetland soil types.

Similar issues of soil class scattering throughout the St. Denis wetland basins were seen in the maps generated from the models based on the 2m – 5x smoothed and 5m – not smoothed DEMs. The maps generated from the models using principal components again had issues with wetland soil classes predicted in upland positions. This is most apparent from the 3C_2mNoSm_PC_ctree model maps (Fig. 4.7g-i) and less apparent with the 3C_5mNoSm_PC_MLR model maps (Fig. 4.7m-o).

## 4.6 Discussion

*4.6.1 4-class vs. 3-class mapping*

The purpose of the modelling was to predict the extents of wetland soils that offer greater potential for phosphorus retention due to their enrichment with $CaCO_3$. The 3-class mapping was designed to be more appropriate for this purpose as it classified the wetland soil based on whether it is enriched with $CaCO_3$ at the surface of its profile where soil calcium could interact with the mobile phosphorus within the wetland pond. The classifications of Calcareous Wetland, Recharge, and Upland were also explored because they are important hydropedological concepts in this region and successful mapping of them could serve additional purposes. The Transition class was incorporated because Calcareous Wetland and Recharge soils represent very specific

concepts and many soil profile observations did not fit into those classes. The extent of the Transition soils between the recharging wetland center and the discharge ring is not necessarily important to distinguish for the purposes of phosphorus retention potential estimates. However, based on the map outputs, the 4-class maps may provide more information in terms of phosphorus retention potentials than the 3-class maps. This is because the predicted extents of Transition soils provide more information than they were intended to, as they seem to indicate wetlands with the potential to be fully discharge/strongly calcareous wetlands; this is discussed further in section 4.6.2. Otherwise, the 4-class and 3-class 2mNoSm_OV_RF and 2mNoSm_OV_treebag models predict similar extents for the Calcareous Wetland soils. The only difference is that Non-Calcareous Wetland soils were further specified as either Transition and Recharge soil classes in the 4-class maps. The predictive accuracies were generally lower for the 4-class maps, but this is due to the higher predictive error in differentiating between the Recharge and Transition soil classes. The predictive accuracies of the 4C_2mNoSm_OV_RF and 4C_2mNoSm_OV_treebag models were still within an acceptable range. There was also some speckling of Recharge and Transition soil classes within some wetland basins in the 4-class maps generated by these models, but this was not as prominent as was seen in other model outputs.

### *4.6.2 Fully discharge/strongly calcareous wetlands*

None of the models predicted wetlands with Calcareous Wetland soils throughout their basins. This contributed to the low predictive accuracies for the BIOCAP external validation tests. One of the three sampled wetlands in this dataset had Calcareous Wetland soils throughout its basin; the 3C_5mNoSm_PC_MLR model was the only model that predicted Calcareous Wetland soils within that specific wetland basin. However, it did not consistently predict Calcareous Wetland soils throughout the known discharge wetlands in the St. Denis study area.

Within the dataset collected for this study (the model training data), there were only six wetlands sampled that were likely to be fully discharge/strongly calcareous wetlands: five in the lower portion of the St. Denis study area and one in the lower portion of the Swift Current study area. Even though they were sampled, it was not possible to confirm if they were fully discharge/strongly calcareous wetlands because sampling could not be conducted at the basin floor due to deep water inundation. This was an issue with the sample design that was likely to

have caused the lack of these wetland types being predicted. Generally, discharge wetlands are less likely to dry out annually (Van der Kamp and Hayashi, 2009). Recharge wetlands can also remain inundated throughout the year, but smaller, less permanent wetlands are typically recharge wetlands. The pattern of soil distribution observed at most recharge wetlands was a ring of Calcareous Wetland soils surrounding the Recharge soils within the basin floor (potentially with a ring of Transition soils between them). Because this pattern was so common in the training data, the models likely assumed that within the potentially discharge/strongly calcareous wetlands, at the positions at lower than the observed Calcareous Wetland soil profiles, the same pattern of Transition and Recharge soils would be encountered.

The low producer's accuracy and high user's accuracy for the Calcareous Wetland soil class indicates that the models underestimated the frequency of these soils, but where they were mapped they were confidently mapped as such. There were fewer Calcareous Wetland observations within the training data than Recharge, Non-Calcareous Wetland, and Upland soil class observations (Tables 4.5 and 4.6). Therefore, models that predict greater frequencies of these other soil classes will be preferentially selected because they are more likely to have greater overall predictive accuracy. One method to address this issue is to influence the model to predict more Calcareous Wetland soil classes by creating synthetic observations of Calcareous Wetland soil classes in the training data using the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002). This involves generating new predictor variable information for each synthetic observation based on the predictor variable information of the existing observations. This was tested with the 3C_2mNoSm_OV_RF model where 40 synthetic observations of Calcareous Wetland soils were incorporated into the training data to match the number of Non-Calcareous Wetland soils (n = 79). However, the resulting model had similar producer's and user's accuracies for Calcareous Wetland observations and similar overall predictive accuracy for the BIOCAP external validation test. Due to the water inundation during the time of sample collection at potential discharge/strongly calcareous wetlands, Calcareous Wetland soils were only observed in the fringe surrounding wetlands and not within wetland basin floors. Therefore, these soil class observations occupied a relatively small feature space in terms of their predictor variable values. The newly generated synthetic observations would be

ascribed predictor variable information similar to the existing observations and would not necessarily encourage the mapping of these soil classes in the wetland basin centers.

The wetlands that were mapped as having Transition soils throughout their basins (flow-through wetlands) by the 4C_2mNoSm_OV_RF and 4C_2mNoSm_OV_treebag models have characteristics that would indicate the potential to be fully discharge/strongly calcareous wetlands; they were typically large and found in lower positions within the landscape. This included most of the known discharge wetlands in the St. Denis study area. This is likely due to the influence of a sampled flow-through wetland in this area that had Transition soils throughout its basin center. Wetlands mapped as flow-through wetlands could be interpreted as wetlands with greater potential to be discharge/strongly calcareous wetlands. Certain wetlands found to be recharge through the field sampling were also mapped as being flow-through wetlands, as seen in the largest wetland shown in the Swift Current output map (Fig. 4.6). But beyond this example, known recharge wetlands were not commonly mapped as flow-through wetlands.

### *4.6.3 Predictor variable importance*

Random forest models do not allow for easy interpretation of their decision-making, but they do provide a measure of the predictor variable importance to the decision-making. This is reported as mean decrease in Gini coefficient (MDG) which reflects how effective a variable is at creating homogeneous splits in the nodes across the entire forest (Strobl et al., 2009). Fig. 4.8 shows the MDG of each predictor variable for the 4C_2mNoSm_OV_RF model. Many of the knowledge-based predictor variables developed for this study were found to be important according to this and the other best-performing random forest models (Fig. 4.8).

According to the 4C_2mNoSm_OV_RF model, the *Elevation from depression spillover* variable was one of the most important predictor variables (Fig. 4.8). This is consistent with the findings of Pennock et al. (2014) that found the distribution of Recharge and discharge (Calcareous Wetland) soils was predictable based on maximum recorded water levels. Their finding was determined at sites where pond water levels have been recorded annually since at least 1968. This information does not exist for most locations but *Elevation from depression spillover* works as a proxy for this information, as for many wetlands, the depression spillover elevation will

match the maximum water level. This will not necessarily be true for the largest wetlands. This highlights the advantage of more complex model types, like the tree-based models, that determine instances where certain variables are most applicable for determining class predictions.

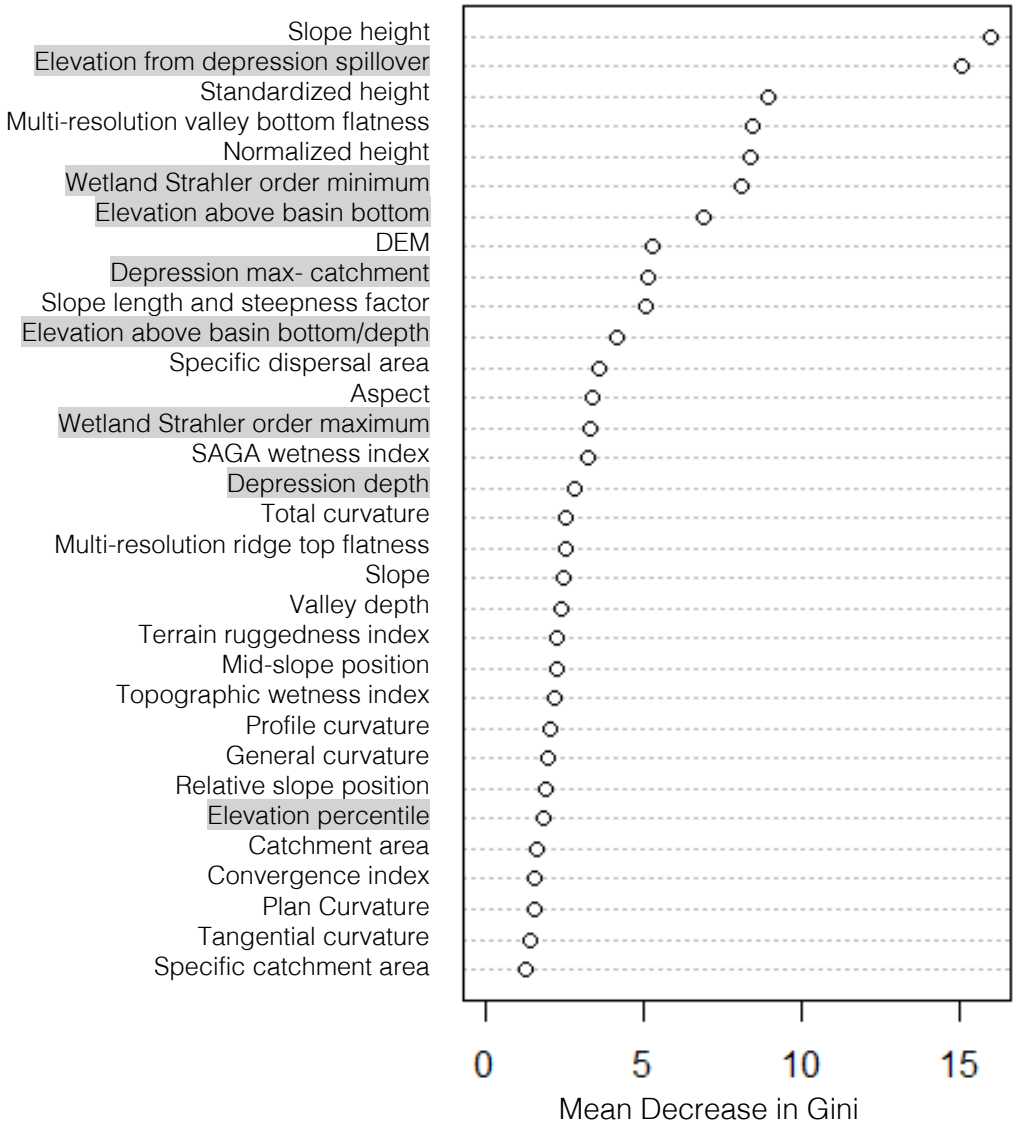



**Fig. 4.8 Predictor variable importance for the 4C_2mNoSm_OV_RF model plotted based on MDG. Predictor variables developed for this study are highlighted in gray.**

The *Elevation above basin bottom* attribute was another useful knowledge-based predictor variable developed for this study. Pennock et al. (2014) found that 95% of Recharge soils were found within the 0 - 0.95 m elevation above the wetland basin bottom and 90% of discharge

(Calcareous Wetland) soils occurred between 0.95 to 2 m elevation above the wetland basin bottom. The relationships between *Elevation above basin bottom* and the soil profiles sampled in this study (which included a broader range of wetland sizes) were analyzed. It was found that all Recharge soils were within 0 to 1.2 m and all Transition soils were within 0 to 2.1 m elevation above the basin bottom. All Calcareous Wetland soils were found between 0.5 to 3 m elevation above the basin bottom except for five profiles that were adjacent to the two massive, potentially lake-sized wetlands at St. Denis. This attribute is important for predicting the distributions of soil classes. It does not provide all necessary information, but again, it provides useful information to more complex model types to make predictions.

### *4.6.4 DEM resolution and smoothing*

The models based on the 2-m – 5x smoothed and the 5-m – not smoothed DEMs performed well in the cross-validation and external validation tests. The issue with these models was identified in the visual assessment. Most commonly seen in the St. Denis study area, the wetland basins were often speckled with varying soil types. The DEM smoothing and coarsening was performed to reduce the erroneous sinks and artefacts present in the DEM and therefore, it was expected that class prediction errors would have been more likely to be found in the maps generated based on the 2-m – not smoothed DEM.

The class speckling within the basins was commonly associated with wetlands where low levels of water were present during the time of LiDAR collection. This results in flat surfaces in the DEM where the water was located (Li et al., 2011). Upon close inspection of these positions within the 2-m – not smoothed DEM, it was observed that these flat surfaces were not perfectly flat; there were small differences in elevation from cell to cell (less than 30 cm) (Fig. 4.9a). These differences were errors within the DEM surface. The small degree of variation in cell to cell elevations within a relatively flat surface can cause drastically different values between the cells for certain topographic attributes. For example, the *Convergence Index* algorithm considers the direction of slope of the adjacent cells compared to the target cell, it does not consider the degree of slope (Kiss, 2004). The variation caused by the error in the flat surface causes drastically different Convergence Index values between cells which results in a very noisy raster surface (Fig. 4.9c). This was the case for several topographic attribute raster surfaces generated

from the 2-m – not smoothed DEM. The machine-learning models would be unable to determine useful relationships between the soil classes and the noise of the topographic attributes in these positions and would not use them to inform predictions. Conversely, when the flat surfaces were smoothed and coarsened to create the 2-m – 5x smooth and 5-m – not smoothed DEMs, the cell to cell variation was reduced but erroneous features still existed (Fig. 4.9b.). The topographic attributes of these features look less like random noise (Fig. 4.9d) and may end up providing information to the machine-learning models causing the models to predict different soil classes at different positions along these features.

The artefactual features within the water-caused near-flat surfaces could be corrected by removing these surfaces from the DEM and re-interpolating them, as was done by Li et al. (2011). These positions would still have flat surfaces, but they would either be perfectly flat or have the same generalized slope across them so that artefactual features would not be interpreted from them. Li et al. (2011) provide a simple method for determining these surfaces by using the LiDAR intensity data to determine where water was present during the time of LiDAR collection. However, the LiDAR intensity data was not available for these study areas. Detection of these water-caused near-flat surfaces could not be automated and so re-interpolation of these surfaces was not conducted for this study.

Aside from this class speckling within the wetland basins, most of the classification trees with bagging and random forest models generated using the 2-m – 5x smoothed and the 5 m – not smoothed DEMs created acceptable map outputs. The models built from the 5-m DEM did not capture features in as much detail but captured the same major distributions of wetland soils as the 2-m resolution maps. To map PPR wetland soils on a much larger scale, it would be especially beneficial to work with 5-m DEMs to reduce computational memory and processing time requirements.

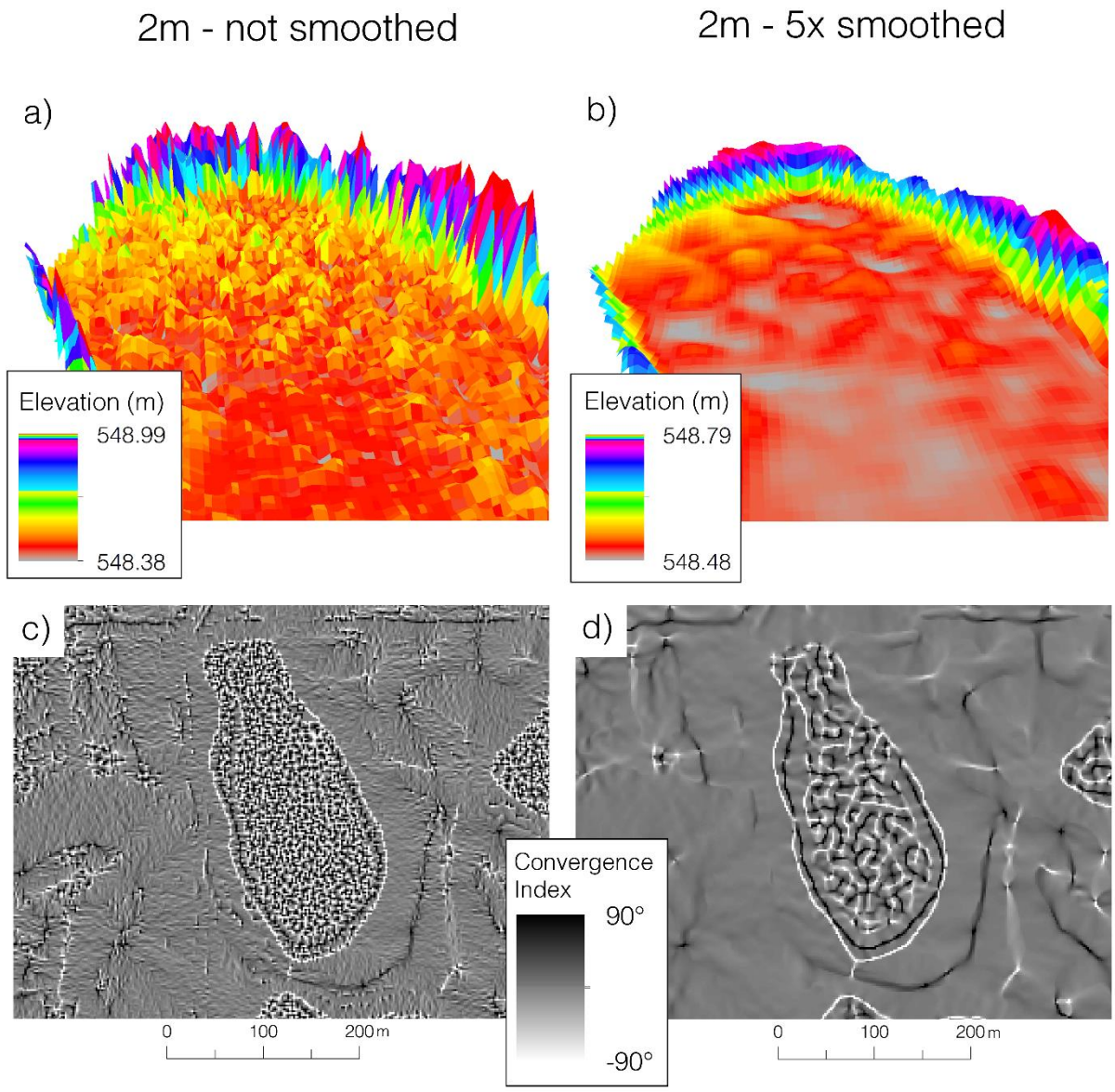



**Fig. 4.9 Comparing surficial errors within the 2 m – not smoothed and 2 m – 5x smoothed DEM. a) and b) show 3D representations of the elevation surfaces for a wetland within the St. Denis DEM with a near-flat surface due to water presence during LiDAR collection. c) and d) show the resulting raster surfaces for the Convergence Index variable for the same wetland.**

*4.6.5 Principal component analysis*

There were fewer principal component-based models amongst the best-performing models (only 4 of 21). Based on the validation accuracy metrics, it is unclear if using the principal components had any effect on the performance of the tree-based machine-learning models. In terms of visual assessment, the principal component-based models often generated maps with speckling of unexpected soil classes in both the wetland and upland landscape positions. The use of principal

components improved the performance of the multinomial logistic regression models as the only MLR models included as best-performing models were principal component-based. Non-penalized multinomial logistic regression models are susceptible to overfitting when too many predictor variables are available (Hastie et al., 2009). Because principal component analysis reduced the number of predictor variables, the principal component-based MLR models performed better than the original variable-based MLR models.

One issue with PCA for this study is that the variables were simplified in terms of their overall variance and not strictly in terms of their variance in relation to the target soil classes. Not all features of the landscape were as important as others to this modelling exercise; large expanses of upland areas would be expected to have the same soil type, whereas three to four soil types may be found in the mid-slope, foot-slope, and toe-slope of a depression within a few meters of each other. The topographic attributes of these more important features may be simplified through PCA causing useful modelling information to be lost.

### *4.6.6 Drained wetlands*

Wetlands that had been mechanically drained were avoided in the sample design because it was unknown how the distribution of soil types would have been affected. It was also unclear how the drainage would affect the topographic attributes of the wetlands. Many of the topographic attributes used as predictor variables in this study quantify water movement potential in the landscape. As well, many of the topographic attributes developed in this study were based on the closed topographic depressions in the DEM. Mechanical drainage alters the elevation surface and so these attributes would be affected.

The Brown – Smith Creek – Drained dataset allowed for the assessment of the performance of the prediction models in areas that had undergone mechanical drainage. The models were still able to successfully predict the distribution of soil types for this dataset. There were only four Calcareous Wetland soil observations within this dataset and they were predicted with varying accuracy (0 – 75 % producer's accuracy). But the high overall predictive accuracy still indicates that the models were at least able to successfully differentiate between wetland and upland soils.

Therefore, modelling could potentially be used for restoration projects in these drained landscapes to identify previous wetland extents.

### 4.7 Conclusion

The combination of high-resolution LiDAR-derived DEMs and machine-learning model techniques were assessed for modelling the spatial distribution of wetland and upland soil types within PPR landscapes. Two modelling objectives were tested: 1) 4-class mapping where soils were mapped to reflect hydropedological concepts of Recharge, Calcareous Wetland, Transition, and Upland soils and 2) 3-class mapping where soils were mapped in terms of their potential for phosphorus retention within wetlands.

The 4C_2mNoSm_OV_RF and 4C_2mNoSm_OV_treebag models were the best-performing models for 4-class mapping. The 3C_2mNoSm_OV_RF and 3C_2mNoSm_OV_treebag models were the best-performing models for 3-class mapping. These models had similar cross-validation and external validation accuracies as the other best-performing models but the maps they generated were more consistent with the conceptual understanding of the distribution of wetland soil types within the PPR. There were issues of soil class speckling in unexpected landscape positions within the maps generated by the models based on the 2 m – 5x smoothed and 5 m – not smoothed DEMs as well as in the maps generated by the principal component-based models. No models consistently predicted wetlands with Calcareous Wetland soils throughout their basin. This is due to being unable to sample the basin centers of potentially fully discharge/strongly calcareous wetlands because they were inundated at the time of sampling.

The 4-class and 3-class 2mNoSm_OV_RF and 2mNoSm_OV_treebag models successfully mapped the distribution of soil types within recharge wetlands. The distribution of Calcareous Wetland soils in the wetland fringe is intuitively understandable when observed in the field; however, it is difficult to quantify this relationship with a single or even several topographic attributes. The more complex machine-learning techniques, like the ensemble tree learners (classification trees with bagging and random forest models), were able to successfully model this soil distribution for wetlands of varying sizes and characteristics. The inclusion of knowledge-based predictor variables based on the relationships established by Pennock et al.

(2014) improved these predictive models. The map outputs from these models convey the intuitively-understood distribution of these soil types.

Map users can be confident to find Calcareous Wetland soils where they were mapped by these models. However, the models underestimate the extent of these soil types because no models consistently predicted the occurrence of wetlands with Calcareous Wetland soils throughout their basins. This must be considered if the models were used to upscale estimates of these soil types across the PPR. The 4-class mapping models consistently mapped wetlands with greater potential to be discharge/strongly calcareous wetlands as flow-through wetlands (which have Transition soils throughout their basin floors). For this reason, the 4-class mapping models provide more information than the 3-class mapping models for the purpose of identifying wetlands with greater potential for phosphorus retention because the mapped flow-through wetlands could be interpreted as potentially discharge/strongly calcareous wetlands. These wetlands could be given greater priority in terms of conservation for the purposes of maintaining phosphorus retention potentials in the PPR.

Overall, the results indicate that DSM methodologies are well suited to mapping wetland soil types in the PPR. However, the presence of water surfaces within the DEMs and the challenge of sampling more permanent wetlands with inundated basin centers must be considered. Also, the landscapes are defined by their small-scale topographic variability and so mapping of soil distributions should be done at high-resolutions. It is likely that at much coarser resolutions, wetland depressional features would be smoothed over and predictive capacities would be reduced.

# 5 SYNTHESIS AND CONCLUSIONS

Phosphorus loading of prairie watersheds has led to degradation of water quality in major waterbodies within the region, including Lake Winnipeg (Environment Canada, 2011). Drainage of PPR wetlands reduces the natural ability of these wetlands to retain mobile phosphorus from agricultural runoff (Blann et al., 2009; Badiou et al., 2018). Efforts to maintain phosphorus retention potentials in the PPR may be more effective by prioritizing the conservation of wetlands with soils enriched with $CaCO_3$, as these soils have greater potential to retain mobile phosphorus (Zhang et al., 2014; Brown et al., 2017b). The spatial distributions of wetland solute and $CaCO_3$ enrichment are controlled by hydrologic processes that occur over landscape scales and on individual wetland scales (Van der Kamp and Hayashi, 2009; Pennock et al., 2014). The studies presented in this thesis attempted to predict the spatial distributions of solute-rich wetlands and calcareous soils through analysis of high-resolution LiDAR-derived DEMs. Two modelling approaches were tested, each with slightly differing objectives. The model discussed in Chapter 3 proposes a method to estimate hydrologic characteristics for individual wetlands and predict them to be either fresh or solute-rich. Digital soil mapping methodologies were tested in Chapter 4 to predict the spatial distribution of wetland soil types through use of machine-learning modelling techniques and many topographic attributes derived from the DEM. The models predictively mapped soil classes based on hydropedologic units: Recharge, Calcareous Wetland, Transition, and Upland soils as well as more generalized soil classes that related more specifically to $CaCO_3$ enrichment: Calcareous Wetland, Non-Calcareous Wetland, and Upland soils.

## 5.1 Summary of findings and general conclusions

The modelling studies of this thesis indicate that it is possible to predictively map the spatial distributions of solute-rich wetlands and calcareous wetland soils within PPR landscapes. Both modelling approaches were dependent on high-resolution DEMs. The small-scale topographic variation of the PPR defines its hydrologic and hydropedologic processes which control the distributions of $CaCO_3$ (Van der Kamp and Hayashi, 2009; Pennock et al., 2014). The high-resolution DEMs allowed for more realistic approximations of these processes and characteristics. At too coarse of resolutions, depressional features can be smoothed out and important modelling information is lost (Li et al., 2011).

Deep groundwater movement can be an important control on the distribution of solutes in PPR wetlands (Van der Kamp and Hayashi, 2009). The inability of the models to specifically account for deep groundwater movement using only surface elevation models likely resulted in some model error for both studies. The model proposed in Chapter 3 underestimated the number of solute-rich wetlands in the SDNWA which may have been because it could not account for the intertill aquifer known to contribute solutes to select wetlands in that area (Hayashi et al., 1998b; Heagle et al., 2013). The models discussed in Chapter 4 had difficulty predicting the occurrence of discharge wetlands which dominantly receive deep groundwater discharge. It would require enormous resources to develop spatially continuous maps of deep groundwater characteristics on wide-scales. Although high-resolution DEMs can be expensive to generate (i.e. via LiDAR) and cannot specifically account for deep groundwater characteristics, they are relatively accessible, and they provided adequate information to allow for the modelling of the hydrologic and hydropedologic processes occurring at this scale, which would not have been possible otherwise. Furthermore, unmanned aerial vehicles are increasingly being used to collect data to generate high-resolution DEMs; this approach requires comparatively minimal costs.

The model proposed in Chapter 3 was successful in predicting the spatial distributions of fresh and solute-rich wetlands within PPR landscapes. The best-performing model had acceptable predictive accuracies based on the training and external validation tests (between 69 and 82% accurate). The methodologies proposed for this model, where stream channel networks were adapted to reflect spill channel connections that were more likely to occur, proved to be advantageous for this modelling purpose. This methodology accounts for the influence of depressional features on hydrologic connectivity within PPR landscapes. When this methodology was not incorporated into the model, the model did not perform as well. These characteristics need to be considered in any attempt to model hydrology in the PPR (Li et al., 2011; Shook et al., 2013).

The model also showed potential to be used for predicting wetland solute-richness for other purposes and could, therefore, potentially inform studies in biology, ecology, or agrology. It was tested to predict the distributions of wetland solute-richness classes based on salinity thresholds that reflected risks to agronomic productivity. Different model parameters were used for this

objective. These models were successful in their predictions, although they had slightly lower predictive accuracies than the models generated to predict wetland solute-richness classes that reflected $CaCO_3$ enrichment. This may indicate that the model is less effective at differentiating between wetlands that are at the higher end of the solute-richness spectrum, i.e. differentiating solute-rich wetlands from very solute-rich wetlands.

There was varying success for predicting the spatial distributions of wetland soil types depending on the DSM methodologies used. The classification trees with bagging and random forest models that used predictor variables derived from the 2-m DEM with no smoothing were the best-performing models. These models had reasonable predictive accuracies based on the cross-validation and external validation tests and mapped the soils in distributions that matched the conceptual understandings established in the literature (Arndt and Richardson, 1988; Pennock et al., 2014). The classification trees with bagging and random forest models are complex modelling techniques that can capture effects and interactions between many predictor variables. This likely contributed to their success in predicting the distribution of wetland soil types, which result from complex interactions of soil forming factors. The topographic attributes developed to specifically reflect characteristics of the PPR morphology contributed to the overall success of the models. A few of these were based on the relationships between topographic attributes and PPR soil distributions determined by Pennock et al. (2014) and represent knowledge-based predictor variables. The incorporation of knowledge-based predictor variables into DSM models has been shown to improve model performance (MacMillan et al., 2005).

Many of the DSM models had reasonable predictive accuracies but produced maps with unexpected distributions of soils. This was largely due to the effect of near-flat surfaces within the DEM in the positions where water was present during the time of LiDAR collection. If not for these specific issues, the map outputs from these models would be acceptable. The models that were generated using the principal component-based predictor variables produced maps with soil classes scattered in unexpected positions. Conducting principal component analysis for the predictor variables did not prove to be beneficial for this modelling objective.

None of the DSM models successfully predicted the occurrence of fully discharge/strongly calcareous wetlands, which have Calcareous Wetland soils throughout their basins. This was the result of an issue with the sample design where the basin floors of the larger, more permanent wetlands could not be sampled due to deep water inundation and, as a result, there were minimal observations of Calcareous Wetland soil classes in wetland basin floor positions. The lack of these observations in the training dataset resulted in no models predicting the occurrence of discharge/strongly calcareous wetlands. This highlights the importance of the training data to DSM. The training data must include observations that are representative of every landscape feature that is meant to be mapped. Some of the 4-class mapping models predicted flow-through wetlands (wetlands with Transition soils throughout their basin) occurring where known discharge wetlands are found at St. Denis. These could be interpreted to indicate wetlands with *potential* to be discharge/strongly calcareous wetlands.

The models in both studies incorporated topographic attributes that reflected landscape and watershed-scale hydrologic processes. Distributions of $CaCO_3$ are controlled by processes occurring over this scale. It is uncertain how the models would perform using DEMs for only a portion of a watershed. The topographic characteristics derived from the DEM would not be able to account for the hydrologic processes occurring on the larger scales. Soil distributions within individual wetlands could still likely be predicted, but the models would have difficulty determining wetlands as either fresh or solute-rich or recharge, discharge/strongly calcareous, or flow-through, depending on the characteristics and size of the area to be modelled.

### 5.2 Model application for prioritizing conservation efforts

A number of different wetland types were discussed throughout this thesis (Table 5.1). There are clear relationships between the wetland types but, for the purposes of this study, the terms are not used interchangeably.

**Table 5.1 Wetland type definitions used within this thesis.**

| Chapter | Wetland Type | Definition |
|---|---|---|
| 3 | Fresh | EC < 1000 µS $cm^{-1}$<br>EM < 70 mS $m^{-1}$ |
| | Solute-rich | EC > 1000 µS $cm^{-1}$<br>EM > 70 mS $m^{-1}$ |
| 4 | Recharge | Recharge/Non-Calcareous Wetland soils throughout wetland basin floor |
| | Flow-through | Transition soils throughout wetland basin floor |
| | Discharge / strongly calcareous | Calcareous Wetland soils throughout wetland basin floor |

The models discussed in this thesis were developed using very different approaches but they both could be used for the same objective: to identify wetlands with soil enriched with $CaCO_3$ that have a greater potential for phosphorus retention. Both approaches have underlying uncertainties associated with them. The model proposed in Chapter 3 maps solute-rich wetlands, which can include both recharge, flow-through, and discharge/strongly calcareous wetland types. Recharge wetlands can have strongly calcareous soils within their discharge rings, but discharge/strongly calcareous wetlands have calcareous soils throughout their basin and therefore would likely have a greater potential for retaining mobile phosphorus because there is more soil-surface area to interact with the wetland pond-water phosphorus.

Although the DSM models did not predictively map discharge/strongly calcareous wetlands, the wetlands mapped as flow-through wetlands could be indicative of wetlands with greater potential to be discharge/strongly calcareous. There were much fewer flow-through wetlands mapped than solute-rich wetlands, and those that were mapped as flow-through wetlands were often known discharge wetlands. Therefore, the wetlands that were mapped as flow-through wetlands were the most likely to be discharge/strongly calcareous wetlands and would be considered to have the

greatest potential for phosphorus retention. Other wetlands predicted to be solute-rich by the model described in Chapter 3 would still be expected to be enriched with $CaCO_3$ but not to the extent of those most likely to be discharge/strongly calcareous wetlands. Wetlands predicted to be fresh are expected to have the lowest potential for phosphorus retention. Wetlands could be prioritized for conservation based on their expected phosphorus retention potential from most to least important: 1) discharge/strongly calcareous/flow-through, 2) other solute-rich wetlands, and 3) fresh wetlands. This is not to indicate that fresh wetlands do not contribute to managing phosphorus runoff; all wetlands provide increased water storage capacities and reduce conveyance within PPR watersheds (Blann et al., 2009). Other factors should also to be considered in conservation decision-making, including a wetland's potential to contribute runoff to rivers and streams. Certain wetlands would not be expected to contribute phosphorus runoff to downstream waterways regardless of their chemical characteristics due to their topographic position or morphology. Where as "gate-keeping" wetlands would be most important to conserve as they control connectivity to downstream waterways.

### 5.3 Suggested model improvements and future research directions

There are several improvements that could be made to each of the modelling approaches. The model proposed in Chapter 3 based its predictions on only one or two variables (Strahler order and terminal status), and the performance of the model varied at the different study areas. The study areas varied in terms of the $CaCO_3$ content of their parent materials and their climatic characteristics, which affect wetland area extents and general hydrologic connectivity within the landscape. These characteristics would have implications for the distributions of solute-rich wetlands. The model could be improved upon by incorporating site specific information that reflect these characteristics, such as soil survey salinity maps. This issue could also be addressed by calibrating the model parameters for a target mapping area by using wetland salinity data from that area.

The main issues encountered in the DSM study (Chapter 4) were due to water within the wetlands, which limited sampling access and caused error-prone near-flat surfaces within the DEMs. Capturing wetland soil observations in the centers of larger, more permanent wetlands are essential for accurate predictive mapping in these positions. The soil sampling was conducted

during a particularly wet period. Wetland basin centers for larger, more permanent wetlands would be more accessible during drier periods. Other sampling methods could also be explored, such as coring through the ice in winter. The issue of the near-flat surfaces within the DEMs could be corrected by identifying these features through an automated detection procedure and re-interpolating them.

Testing wetlands for salinity and field-testing wetland soil cores through application of hydrochloric acid were both very quick measures of solute-richness and $CaCO_3$ content. The classification criteria used for the solute-rich wetland class reflects a very broad range of wetland salinity and the wetlands within this class would have had a wide range of $CaCO_3$ content within their soils. Similarly, the soils that were determined to have moderate to strong $CaCO_3$ within their surface horizons to be classified as Calcareous Wetland soil may have contained widely varying actual $CaCO_3$ content. It would be beneficial in a future modelling study to analyze and quantify soil surface $CaCO_3$ content and attempt to map it. There has been extensive work in the field of DSM to map continuous variables like $CaCO_3$ content. However, the model would need to also account for the potential for the $CaCO_3$ to interact with mobile phosphorus within runoff. There are many landscape positions with $CaCO_3$-enriched soils that would not have the same potential to interact with phosphorus runoff as wetland positions do. This is a strength of the class-based approach, as the wetland soil classes reflect soils that are most likely to interact with runoff.

Before any further modelling is conducted, it is necessary to further quantify the relationships between phosphorus retention and calcareous wetland soils. Brown et al. (2017b) compared available and total phosphorus content within calcareous and non-calcareous wetland soils but had a limited sample set. It would be beneficial for future studies to analyze the relationships between soil type and specific phosphorus fractions, as they have different loss potentials (Brown et al., 2017b). It would also be useful to compare phosphorus retention of discharge ring soils and calcareous basin floor soils, as this would have implications for prioritizing wetland conservation. With better relationships established between calcareous wetland soils and phosphorus retention, the map outputs from the DSM models could be used to estimate the stored phosphorus within wetland soils across the PPR.

# 6 REFERENCES


Acton, D.F., and Ellis, J.G. 1978. The soils of the Saskatoon map area 73-B Saskatchewan. Saskatchewan Institute of Pedology Publication, S4.

Agriculture and Agri-Food Canada. 2009. ISO 19131 Swift Current LiDAR Project 2009 - Data Product Specifications. Available: https://open.canada.ca/data/en/dataset/72003a79-c799-49de-b135-de5af833f029 (verified 22 July 2018).

Agriculture and Agri-Food Canada. 2017. Detailed soil survey (DSS) compilations. Available: http://sis.agr.gc.ca/cansis/nsdb/dss/v3/index.html (verified 25 July 2018).

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.

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(3): 848–855.

Ayres, K.W., Acton, D.F., and Ellis, J.G. 1985. The soils of the Swift Current map area 72 J Saskatchewan. Saskatchewan Institute of Pedology Publication, S6.

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.

Bedard-Haughn, A. 2011. Gleysolic soils of Canada: Genesis, distribution, and classification. Can. J. Soil Sci. 91: 763–779.

Bedard-Haughn, A.K., and Pennock, D.J. 2002. Terrain controls on depressional soil distribution in a hummocky morainal landscape. Geoderma 110: 169–190.

Berthold, S., Bentley, L.R., and Hayashi, M. 2004. Integrated hydrogeological and geophysical study of depression-focused groundwater recharge in the Canadian prairies. Water Resour. Res. 40, W06505.

Beven, K.J., and Kirkby, M.J. 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24: 43–69.

Blann, K.L., Anderson, J.L., Sands, G.R., and Vondracek, B. 2009. Effects of agricultural drainage on aquatic ecosystems: a review. Crit. Rev. Environ. Sci. Technol. 39: 909–1001.

Boehner, J., Koethe, R. Conrad, O., Gross, J., Ringeler, A., Selige, T. 2002. Soil regionalisation by means of terrain pnalysis and Process parameterisation. Pages 213 – 222 *in* Micheli, E., Nachtergaele, F., Montanarella, L., ed., Soil Classification 2001. European Soil Bureau, Research Report No. 7, Luxembourg.

Boehner, J. and Selige, T. 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. Pages 13 – 27 *in* Boehner, J., McCloy, K.R., and Strobl, J., ed. SAGA – Analyses and Modelling Applications. Gottinger Geog. Abh. 115.

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.

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.

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.

Canadian Wetland Inventory Technical Committee. 2016. Canadian Wetland Inventory (data model) version 7.0. Available: http://www.ducks.ca/assets/2017/01/CWIDMv7_01_E.pdf (verified 27 July 2018).

Chawla, N. V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. 2002. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 16: 321–357.

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 depressional wetland landscape. Wetlands 27: 719–738.

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.

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(1): 135–142.

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(2): 233–244.

Debella-Gilo, M., and Etzelmüller, B. 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. Catena 77: 8–18.

Environment Canada. 2011. State of Lake Winnipeg: 1999 to 2007. Manitoba Water

Stewardship. Available: https://www.gov.mb.ca/waterstewardship/water_quality/state_lk_winnipeg_report/pdf/state_of_lake_winnipeg_rpt_technical_low_resolution.pdf (verified 27 July 2018)

ESRI. 2016. ArcGIS Desktop: Release 10.4. Redlands,CA.

Euliss, N.H., Mushet, D.M., Newton, W.E., Otto, C.R. V, Nelson, R.D., LaBaugh, J.W., Scherff, E.J., and Rosenberry, D.O. 2014. Placing prairie pothole wetlands along spatial and temporal continua to improve integration of wetland function in ecological investigations. J. Hydrol. 513: 490–503.

Fang, X., and Pomeroy, J.W. 2008. Drought impacts on Canadian prairie wetland snow hydrology. Hydrol. Process. 22(15): 2858–2873.

Florinsky, I.V., Eilers, R.G., Manning, G.R., and Fuller, L.G. 2002. Prediction of soil properties by digital terrain modelling. Environ. Model. Softw. 17: 295–311.

Gallant, J.C., and Dowling, T.I. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39: 1347–1359.

Government of Canada. 2018. Canadian Climate Normals 1981 - 2010 Station Data. Available: http://climate.weather.gc.ca/climate_normals/ (verified 28 July 2018).

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The elements of statistical learning. Springer, New York, NY.

Hayashi, M., Van der Kamp, G., and Rudolph, D.L. 1998a. Water and solute transfer between a prairie wetland and adjacent uplands, 2. Chloride cycle. J. Hydrol. 207: 56–67.

Hayashi, M., Van der Kamp, G., and Rudolph, D.L. 1998b. Water and solute transport between a prairie wetland and adjacent upland, 1. Water balance. J. Hydrol. 207: 42–55.

Hayashi, M., and Rosenberry, D.O. 2002. Effects of ground water exchange on the hydrology and ecology of surface water. Ground Water 40(3): 309–314

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.

Hengl, T. 2006. Finding the right pixel size. Comput. Geosci. 32: 1283–1298.

Henry, J.L. 2003. Henry's handbook of soil and water. Henry Perspectives, Saskatoon, SK.

Heung, B., Bulmer, C.E., and Schmidt, M.G. 2014. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. Geoderma 214–215: 141–154.

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.

ISC. 2014. Saskatchewan Road Network Database 2014. Available: https://www.isc.ca/MapsandPhotos/GISData/Pages/SaskatchewanRoadNetworkDatabase.aspx (verified 19 May 2018).

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

Jenny, H. 1941. Factors of soil formation. Soil Sci. 52(5): 415.

Johnston, C.A. 1991. Sediment and nutrient retention by freshwater wetlands: Effects on surface water quality. Crit. Rev. Environ. Control 21: 491–565.

Jones, G., and Armstrong, N. 2001. Long-term trends in total nitrogen and total phosphorus concentrations in Manitoba streams. Manitoba Conservation Report No. 2001-07.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., and Stoorvogel, J.J. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151: 311–326.

King, K.W., Williams, M.R., Macrae, M.L., Fausey, N.R., Frankenberger, J., Smith, D.R., Kleinman, P.J.A., and Brown, L.C. 2015. Phosphorus transport in agricultural subsurface drainage: a review. J. Environ. Qual. 44: 467.

Kiss, R. 2004. Determination of drainage network in digital elevation models, utilities and limitations. J. Hungarian Geomathematics 2: 16–29.

Klassen, R.W. 1989. Quaternary geology of the southern Canadian Interior Plains. Pages 138–173 *in* R. J. Fulton, ed. Quaternary geology of Canada and Greenland. Geological Survey of Canada, Ottawa, ON.

Kling, H.J., Watson, S.B., McCullough, G.K., and Stainton, M.P. 2011. Bloom development and phytoplankton succession in Lake Winnipeg: a comparison of historical records with recent data. Aquat. Ecosyst. Health Manag. 14(2): 219–224.

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.

Koethe, R., Lehmeier, F., 1996. SARA-Systeme zur automatischen relief-analyse, Benutzerhandbuch, 2. Goettingen University.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28: 1–26.

Landis, J.R., and Koch, G.G. 1977. The measurement of observer agreement of categorical data. Biometrics 33: 159–174.

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.

Lidar Services International. 2009. Manitoba Water Stewardship and Saskatchewan Water Authority October 2008 LiDAR Survey Report.

Lindsay, J.B. 2016. Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. Hydrol. Process. 30: 846–857.

Lindsay, J.B., and Creed, I.F. 2005. Removal of artifact depressions from digital elevation models: towards a minimum impact approach. Hydrol. Process. 19: 3113–3126.

Lindsay, J.B., and Creed, I.F. 2006. Distinguishing actual and artefact depressions in digital elevation data. Comput. Geosci. 32: 1192–1204.

Lissey, A. 1971. Depression-focused transient groundwater flow patterns in Manitoba. Geol. Assoc. Canada - Spec. Pap. 9: 333–341.

MacMillan, R. A., Pettapiece, W.W., and Brierley, J. A. 2005. An expert system for allocating soils to landforms through the application of soil survey tacit knowledge. Can. J. Soil Sci. 85: 103–112.

Maidment, D.R. 2002. Arc hydro: GIS for water resources. ESRI Press, Redlands, CA.

Malone, B.P., Minasny, B., and McBratney, A.B. 2017. Using R for digital soil mapping. Progress in Soil Science. Springer, Switzerland.

Manning, G., Fuller, L.G., Eilers, R.G., and Florinsky, I. 2001. Topographic influence on the variability of soil properties within an undulating Manitoba landscape. Can. J. Soil Sci. 81(3): 439–447.

McBratney, A.B., Mendonça-Santos, M.L., and Minasny, B. 2003. On digital soil mapping. Geoderma 117: 3–52.

McElhanney Consulting Services. Ltd. 2009. Duncairn Dam Study LiDAR Mapping Data, Topography, and Imagery. Available: http://www.agr.gc.ca/atlas/supportdocument_documentdesupport/aafcSkSwiftCurrent2009/en/LiDAR_Mapping_for_Swift_current_Final_Report.pdf (verified 27 July 2018).

Millar, J.B. 1971. Shoreline-area ratio as a factor in rate of water loss from small sloughs. J. Hydrol. 14: 259–284.

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.

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.

Mitsch, W.J., Cronk, J.K., Wu, X., Nairn, R.W., and Hey, D.L. 1995. Phosphorus retention in constructed freshwater riparian marshes. Ecol. Appl. 5(3): 830–845.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5, 3–30.

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.

Nachshon, U., Ireson, A., van der Kamp, G., Davies, S.R., and Wheater, H.S. 2013. Impacts of climate variability on wetland salinization in the North American Prairies. Hydrol. Earth Syst. Sci. Discuss. 10: 13475–13503.

Natural Resources Canada. 2016. Canadian digital elevation model product specifications. Sherbrooke, QC. Available: https://open.canada.ca/data/en/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333 (verified 27 July 2018).

Natural Resources Canada. 2018. High resolution digital elevation model (HRDEM) - CanElevation Series - product specifications. Available: https://open.canada.ca/data/en/dataset/957782bf-847c-4644-a757-e383c0057995 (verified 27 July 2018).

Nauman, T.W., and Thompson, J.A. 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213: 385–399.

Pennock, D.J. 2003. Terrain attributes, landform segmentation, and soil redistribution. Soil Tillage Res. 69: 15–26.

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.

Pennock, D., Yates, T., Bedard-Haughn, A., Phipps, K., Farrell, R., and McDougal, R. 2010. Landscape controls on N2O and CH4 emissions from freshwater mineral soil wetlands of the Canadian Prairie Pothole region. Geoderma 155: 308–319.

Pennock, D., Bedard-Haughn, A., and Viaud, V. 2011. Chernozemic soils of Canada: Genesis, distribution, and classification. Can. J. Soil Sci. 91: 719–747.

Pennock, D.J., Henderson, D., Ireson, A., Naschon, U., Spence, C., Van der Kamp, G., Waiser, M., Wilson, H., and Bedard-Haughn, A. 2013. Where it all comes together: 45 years of research on hydrology-ecosystem interactions at the St. Denis National Wildlife Area. Saskatoon, SK.

Pennock, D., Bedard-Haughn, A., Kiss, J., and Van der Kamp, G. 2014. Application of hydropedology to predictive mapping of wetland soils in the Canadian Prairie Pothole Region. Geoderma 235–236: 199–211.

Phillips, R.W., Spence, C., and Pomeroy, J.W. 2011. Connectivity and runoff dynamics in heterogeneous basins. Hydrol. Process. 25: 3061–3075.

Quinn, P., Beven, K., Chevallier, P., and Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models. Hydrol. Proc. 5: 59–79.

hydrological modeling using digital terrain models. Hydrol. Proc. 5, 59–79.

Rahman, A., Rahman, S., and Cihacek, L. 2014. Influence of soil pH in vegetative filter strips for reducing soluble nutrient transport. Environ. Technol. 35(14): 1744–1752.

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. 2018. FlySask2. [Online] Available: https://www.flysask2.ca/?redirect_to=/imagery/ [2018 Apr. 5].

Saskatchewan Soil Survey Staff. 1991. The soils of Langenburg, Fertile Belt, Churchbridge, Saltcoats rural municipalities Saskatchewan. Saskatchewan Institute of Pedology, S208.

Schindler, D.W. 2012. The dilemma of controlling cultural eutrophication of lakes. Proc. R. Soc. B 279: 4322–4333.

Schindler, D.W., Hecky, R.E., and Mccullough, G.K. 2012. The rapid eutrophication of Lake Winnipeg: Greening under global change. J. Great Lakes Res. 38: 6–13.

Shapiro, M., and Westervelt, J. 1992. r.mapcalc: an algebra for gis and image processing. U.S. Army Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Illinois, 422-425.

Shaw, D.A., Pietroniro, A., and Martz, L.W. 2013. Topographic analysis for the prairie pothole region of Western Canada. Hydrol. Process. 27: 3105–3114.

Shook, K., Pomeroy, J.W., Spence, C., and Boychuk, L. 2013. Storage dynamics simulations in prairie wetland hydrology models: evaluation and parameterization. Hydrol. Process. 27(13): 1875–1889.

Shook, K.R., Armstrong, R., Sharomi, O., Spiteri, R., and Pomeroy, J.W. 2014. The WDPM user's guide. Available: https://www.usask.ca/hydrology/WDPM.php (verified 27 July 2018).

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.

Soil Classification Working Group. 1998. The Canadian system of soil classification, 3rd edition. Research Branch, Agriculture and Agrifood Canada, Ottawa, ON. Publ. 1646.

Spence, C. 2010. A paradigm shift in hydrology: storage thresholds across scales influence catchment runoff generation. Geogr. Compass 4(7): 819–833.

St. Arnaud, R.J. 1976. Pedological aspects of glacial till. Pages 133–155 *in* R.F. Legget, ed. Glacial Till. Royal Society of Canada, Ottawa, ON.

Stewart, R.E., and Kantrud, H.A. 1971. Classification of natural ponds and lakes in the glaciated prairie region. U.S. Bureau of Sport Fisheries and Wildlife Resource Publications No. 92.

Strahler, A.N. 1952. Hypsometric (area-altitude) analysis of erosional topography. GSA Bull. 63(11): 1117–1142.

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(4): 323–348.

Svetnik, V., Liaw, A., Tong, C., and Wang, T. 2004. Application of Breiman's Random Forest to modeling structure-activity relationships of pharmaceutical molecules. Pages 334–343 *in* F. Roli, J. Kittler, and T. Windeatt, eds. Multiple Classifier Systems, Fifth International Workshop. Springer, Berlin.

Thompson, J.A., Bell, J.C., and Butler, C.A. 1997. Quantitative soil-landscape modeling for estimating the areal extent of hydromorphic soils. Soil Sci. Soc. Am. 61(3): 971–980.

Töyrä, J., Pietroniro, A., Craymer, M., and Veronneau, M. 2008. Evaluation of LiDAR-derived ground surface digital elevation model (DEM) in low-relief regions: Case study on the Canadian Prairies. Page *in* Hydroscan: airborne laser mapping of hydrological features and resources. Environment Canada and the Canadian water resources association, Saskatoon, SK.

Trimble. 2005. Getting started guide: GeoExplorer 2005 series. Westminister, Colorado.

University of Saskatchewan. 2018. Saskatchewan Soil Information System. Available: sksis.usask.ca (verified 16 August 2018).

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.

Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S-Plus, 4th edition. Springer, New York, NY.

Vincent, S., Lemercier, B., Berthier, L., and Walter, C. 2018. Spatial disaggregation of complex

Soil Map Units at the regional scale based on soil-landscape relationships. Geoderma 311: 130–142.

Waiser, M.J. 2006. Relationship between hydrological characteristics and dissolved organic carbon concentration and mass in northern prairie wetlands using a conservative tracer approach. J. Geophys. Res. 111.

Watson, K., and Pennock, D.J. 2016. Section 3: Soil profile description. Pages 1–40 *in* D.J. Pennock, K. Watson, and P. Sanborn, eds. Field handbook for soils of Western Canada. Pedology Subcommittee, Canadian Society of Soil Science.

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.

Witten, I.H., and Frank, E. 2005. Data mining: practical machine learning tools and techniques, 2nd edition. Elsevier, San Francisco, CA.

Woo, M.-K., and Rowsell, R.D. 1993. Hydrology of a prairie slough. J. Hydrol. 146: 175–207.

Wu, Q., and Lane, C.R. 2016. Delineation and quantification of wetland depressions in the Prairie Pothole Region of North Dakota. Wetlands 36: 215–227.

Xiong, X., Grunwald, S., Myers, D.B., Kim, J., Harris, W.G., and Comerford, N.B. 2012. Which covariates are needed for soil carbon models in Florida. Pages 109–113 *in* Digital Soil Assessment and Beyond. CRC Press.

Yang, J., and Chu, X. 2015. A new modeling approach for simulating microtopography-dominated, discontinuous overland flow on infiltrating surfaces. Adv. Water Resour. 78: 80–93. Elsevier Ltd.

Zebarth, B.J., De Jong, E., and Henry, J.L. 1989. Water flow in a hummocky landscape in central Saskatchewan, Canada, II. Saturated flow and groundwater recharge. J. Hydrol. 110: 181–198.

Zevenbergen, L.W., and Thorne, C.R. 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landforms 12: 47–56.

Zhang, M., Li, C., Li, Y.C., and Harris, W.G. 2014. Phosphate minerals and solubility in native and agricultural calcareous soils. Geoderma 232–234: 164–171.

# APPENDIX A: GIS METHODOLOGIES FOR WETLAND SOLUTE-RICHNESS CLASS MODEL

**Note: the output files generated throughout the procedures (shapefiles, rasters, etc.) are given names in this document to clarify their origin and purpose, some names may be longer than ArcGIS allows for and abbreviations should be used.**

## Appendix A-1: Remove roads from DEMs

1. Acquire 1-m resolution DEM for the target study area and import into ArcMap. DEMs of other resolutions could also be used but the procedures that are affected by cell size should be adjusted accordingly
2. Acquire or create road network polyline shapefile and import into ArcMap
    a. The Saskatchewan Road Network Database 2014 shapefile provided a polyline shapefile for the roads within each study area
    b. Road segments that were visible in the aerial imagery or in the DEM that were not reflected in the original road network shapefile were added as line segments
3. Use the ArcGIS *Buffer* tool – this step will create polygon buffers of the road network shapefile
    a. Input Features: Road network shapefile
    b. Output Feature Class: Road_Buffer_Polygons
    c. Distance: Linear unit 28 m – this will buffer the roads 28 m on either side (56 m total). 56 m was wide enough to cover the widest road-related features in the DEMs of the three sites. Different values could be used for other sites to either capture larger road-related features or minimize the effect of the procedure on the DEM
    d. Side Type: FULL
    e. End type: ROUND
    f. Method: PLANAR
    g. Dissolve type: NONE
4. Use the ArcGIS *Polygons to Raster* tool
    a. Input Features: Road_Buffer_Polygons
    b. Value field: any input
    c. Output Raster Dataset: Road_Buffer_Raster

d. Cell assignment type: CELL_CENTER
e. Priority field: NONE
f. Cellsize: after setting the cell size in Environments (in the next step), this value should correspond to the DEM resolution
g. Within Environments:
    i. Processing Extent:
        1. Extent: DEM
        2. Snap Raster: DEM – this step ensures that the grid of the created raster matches the original DEM
    ii. Raster Analysis:
        1. Cell Size: DEM

5. Use the ArcGIS *Is Null* tool – this tool will create a new raster of the road buffer. Cells with no road buffer have a value of 1 and cells with a road buffer have a value of 0
    a. Input raster: Road_Buffer_Raster
    b. Output raster: Road_Buffer_Raster_Null
    c. Within Environments…
        i. Raster Analysis:
            1. Mask: original DEM raster file
6. Use the ArcGIS *Set Null* tool
    a. Input Conditional Raster: Road_Buffer_Raster_Null
    b. Expression: Value = 0 – this will set cells that were 0 to NoData, these cells would be located where the road buffers were
    c. Input false raster or constant value: original DEM – for any cell where the value is = 1, it will be given the value of the original DEM
    d. Output raster: DEM_Without_Buffer
7. Use the ArcGIS *Buffer* tool – this 2$^{nd}$ buffer will be used to generate points that will be used to interpolate the elevations within the 1$^{st}$ road buffer area that was removed from the DEM
    a. Input Features: Road_Buffer_Polygons
    b. Output Feature Class: Road_Edge_Buffer_Polygons
    c. Linear unit: 32 m

d. Side Type: **OUTSIDE**
e. End type: ROUND
f. Method: PLANAR
g. Dissolve type: NONE

8. Use the ArcGIS *Polygon to Raster* tool
    a. Input Features: Road_Edge_Buffer_Polygons
    b. Value field: any input
    c. Output Raster Dataset: Road_Edge_Buffer_Raster
    d. Cell assignment type: CELL_CENTER
    e. Priority field: NONE
    f. Cellsize: after setting the cell size in Environments, this value should correspond to the DEM resolution
    g. Within Environments…
        i. In Processing Extent, set the Snap Raster to the original DEM raster file, **do not** set the extent to the original DEM raster file
        ii. In Raster Analysis, set the Cell Size to the original DEM raster file
9. Use the ArcGIS *Raster to Point* tool – this will generate points for each cell of the Road_Edge_Buffer_Raster grid. Generating points can take long processing time depending on the size of the study area
    a. Input Raster: Road_Edge_Buffer_Raster
    b. Output point features: Road_Edge_Buffer_Points
10. Use the ArcGIS *Extract Values to Points* tool – in this step, the points will gain the elevation value for the cell that they overlie
    a. Input Features: Road_Edge_Buffer_Points
    b. Input raster: original DEM
    c. **Uncheck/disable** the following:
        i. Interpolate values at the point locations
        ii. Append all the input raster attributes to the output point features
    d. Output point features: Road_Edge_Buffer_Points_Elevations
11. Use the ArcGIS Spatial Analyst *IDW* tool (Inverse Distance Weighting) – this step will interpolate the elevations of the areas between the road edge using inverse distance

weighting. This tool was chosen because a maximum distance can be set and so areas that do not need to be re-interpolated are not, which reduces computing time

a. Input Point Features: Road_Edge_Buffer_Points_Elevations
b. Z value field: RASTERVALU – this field will have been created in step 10 and reflects the elevations extracted from the DEM
c. Output raster: Interpolated_Road_Elevations
d. Power: 0.5
e. Search radius: Variable
f. Search Radius Settings: Number of points: 120
g. Search Radius Settings: Maximum distance: 250 m
h. Within Environments…
    i. Processing Extent set
        1. Snap raster: original DEM,
        2. **Do not** set Extent to the original DEM
    ii. In Raster Analysis:
        1. Cell size: original DEM.

12. Use the ArcGIS *Focal Statistics* tool – this step will smooth the Interpolated_Road_Elevations. The IDW interpolation process causes a sharp difference along the center of the interpolated road areas which need to be smoothed
    a. Input raster: Interpolated_Road_Elevations
    b. Output raster: Smooth_Interpolated_Road_Elevations
    c. Neighborhood: Rectangle
        i. Height: 30 m
        ii. Width: 30 m
    d. Statistics type: Mean
    e. Check/enable Ignore NoData in calculations
    f. For best results, this process should be repeated multiple times, although it was only done once for this study
13. Use the ArcGIS *Raster Calculator* tool – this step will bind the Smooth_Interpolated_Road_Elevations to the DEM_Without_Buffer

a. Enter the Expression: Con(IsNull("DEM_Without_Buffer"), "Smooth_Interpolated_Road_Elevations", " DEM_Without_Buffer")
    i. This conditional statement states that if a raster cell in the DEM_Without_Buffer raster is NoData, it will be given the value of the Smooth_Interpolated_Road_Elevations raster output, if it has a value it will use the value of the DEM_Without_Buffer raster
b. Output raster: DEM_Road_Removed
c. Within Environments…
    i. Processing Extent:
        1. Extent: original DEM raster
        2. Snap Raster: original DEM raster
    ii. Raster Analysis:
        1. Cell Size: original DEM
        2. Mask: original DEM

**Appendix A-2: Delineate depression boundary polygons and determine depression depths**

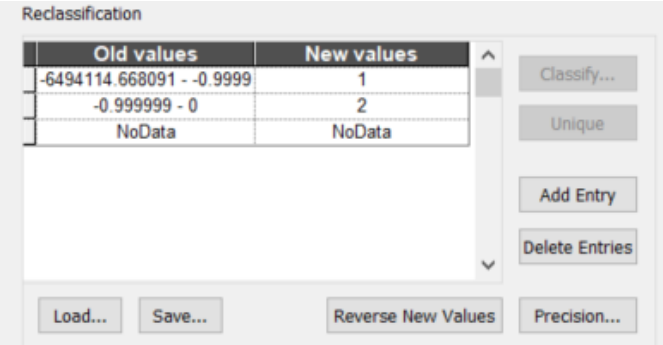
Two methods were used to delineate depression boundary polygons; both methods involved determining the closed topographic depressions within the DEM. The first method determined the maximum closed topographic depressions. This method was tested for delineating depression boundaries in the model proposed in Chapter 3. This method was also used for the DSM study described in Chapter 4 to a) identify depressions for soil sampling and b) generate depression-based predictor variables. The second method determined the nested closed topographic depressions found completely within the maximum closed topographic depressions. This method was tested for delineating depression boundaries in the model proposed in Chapter 3 and was used to identify wetlands for salinity testing for that study. For use in the model proposed in Chapter 3, the closed topographic depressions were determined from the road-removed 2-m DEM. For the depression boundary delineation used in the DSM study described in Chapter 4, the closed topographic depressions were determined from the original (roads not removed) 2-m DEM. For both studies, sampled depressions were selected to be away from roads and so the difference between boundaries for the sampled depressions determined from either the road-removed or original DEMs were minor. The sample design for soil sampling was established

before the road-removed DEM was generated, and so the original DEM was used for all aspects of the DSM study to stay consistent with the sample design. The road-removed DEM was used for the study described in Chapter 3 to stay consistent with the methodologies used in that study.

Delineate the maximum closed topographic depression boundaries

1. Import the 2-m DEM (either road-removed or original) into ArcMap
2. Use the ArcGIS *Fill* tool – this tool will fill all the closed topographic depressions within the DEM
    a. Input surface raster: 2-m DEM
    b. Output surface raster: DEM_filled
    c. Z limit: blank
3. Use the ArcGIS *Fill* tool – this time, only closed topographic depressions that have depths less than 10 cm will be filled. This will allow for the delineation of closed topographic depressions that have depths greater than 10 cm
    a. Input surface raster: 2-m DEM
    b. Output surface raster: DEM_10cm_filled
    c. Z limit: 0.1
4. Use the ArcGIS *Cut Fill* tool – this tool will delineate the closed topographic depressions that were filled in step 2 that were not filled in step 3
    a. Input before raster surface: DEM_10cm_filled
    b. Input after raster surface: DEM_filled
    c. Output raster: DEM_CutFill
    d. Z factor: 1
5. Use the ArcGIS *Reclassify* tool
    a. Input raster: DEM_CutFill
    b. Reclass field: VOLUME

c. Reclassification:




d. Output raster: DEM_CutFill_reclass

e. Uncheck/disable Change missing values to NoData

6. Use the ArcGIS *Raster to Polygon* tool
    a. Input raster: DEM_CutFill_reclass
    b. Field: VALUE
    c. Output polygon features: Max_Depression_Poly
    d. Uncheck/disable Simplify polygons
7. Right-click the Max_Depression_Poly, select Edit Features and Start Editing
8. Open the Max_Depression_Poly Attributes Table
    a. Select by Attributes
        i. gridcode = 2 and press Apply – this will select any area that is outside of a closed topographic depression
    b. Within the attribute table, right-click the block to the left of the observations and select Delete Features – this will delete any polygon that does not represent a closed topographic depression

<u>Delineate the nested closed topographic depression boundaries</u>

9. Use the ArcGIS Data Management *Clip* tool for raster clipping – this step will clip the DEM by the maximum closed topographic depression boundaries, so that the nested closed topographic depressions within those depressions can be determined
    a. Input raster: 2-m DEM
    b. Output Extent: Max_Depression_Poly
    c. Check/enable Use Input Features for Clipping Geometry
    d. Output Raster Dataset: Nested_DEM_1

e. NoData Value: Blank
f. Uncheck/disable Maintain Clipping Extent

10. Follow steps 2 – 8 using the Nested_DEM_1 raster as the input instead of the 2-m DEM and name the output Nested_Depression_Poly_1 – this step will determine the nested closed topographic depressions within the maximum closed topographic depressions.

Some visual inspection was required to determine if the nested closed topographic depression polygons performed better for delineating the wetland boundaries than the maximum closed topographic depression polygons. Aerial imagery from 2008 – 2016 was retrieved from https://www.flysask2.ca/. The aerial imagery with the smallest wetland extents were compared to the nested closed topographic depressions and maximum closed topographic depressions. Nested closed topographic depression polygons were used to delineate wetland boundaries if through visual inspection, they were determined to better represent the wetland boundaries. However, the nested closed topographic depression polygons were only used if there were determined to be multiple closed topographic depressions within a maximum closed topographic depression polygon. Otherwise, if, through the procedure, the maximum closed topographic depression polygon was reduced to a single smaller closed topographic depression polygon, then the polygon boundary would not represent the spillover elevation of that depression and therefore, the maximum closed topographic depression boundary polygon should be used. The following steps determine if a maximum closed topographic depression contained only one nested closed topographic depression:

11. Open the Nested_Depression_Poly_1 Attribute Table
    a. Within Table Options, Add Field
        i. Name: Count_1
        ii. Type: Short Integer
12. Right-click on the Count_1 column heading, and select Field Calculator
    a. Within the blank box under Count_1 =
        i. Enter: 1
        ii. This step will give every polygon within the Nested_Depression_Poly_1 file a count of 1

13. Use the ArcGIS *Spatial Join* tool – this step will add a Count_1 column to the Max_Depression_Poly shapefile and will determine how many nested depressions were within the maximum closed topographic depressions
    a. Target Features: Max_Depression_Poly
    b. Join Features: Nested_Depression_Poly_1
    c. Output Feature Class: Max_Depression_Poly_w_Count
    d. Join Operation: JOIN_ONE_TO_MANY
    e. Check/enable Keep All Target Features
    f. Field Map of Join Features: Remove all Features except Count_1, Right-click the Count_2 Feature, Select Merge Rule, **Select Sum**
    g. Match Option: INTERSECT
    h. Search Radius: blank
    i. Distance Field Name: blank
14. Any maximum closed topographic depression polygon with a Count_1 of only 1 or NULL, is used in the final nested closed topographic depression polygon shapefile. The single nested closed topographic depression polygon within the maximum closed topographic depression polygon need to be deleted:
    a. Open the Attribute Table of the Max_Depression_Poly, use Select By Attributes:
        i. Count_1 = 1 OR Count_1 Is Null
    b. Within the Selection option in ArcMap, use Select By Location:
        i. Selected method: Select features from
        ii. Target layer(s): Nested_Depression_Poly_1
        iii. Source layer: Max_Depression_Poly
        iv. Spatial selection method for target layer feature(s): intersect the source layer feature
        v. Uncheck/disable Apply a search distance
    c. Start Editing the Nested_Depression_Poly_1 and Open its Attribute Table
    d. Right-click a block beside the observations and select Delete Selected
    e. Stop editing and Save Edits
15. The maximum closed topographic depression polygons that did contain multiple nested closed topographic depressions need to be deleted because the multiple nested closed

topographic depression polygons will be included in the final nested closed topographic depression polygon shapefile:

a. Open the Attribute Table of the Max_Depression_Poly, use Select By Attributes:
    i. Count_1 = 1 OR Count_1 Is Null
b. Within the Attribute Table, select Switch Selection
c. Start Editing the Max_Depression_Poly and Open its Attribute Table
d. Right-click a block beside the observations and select Delete Selected
e. Stop editing and Save Edits

16. Use the ArcGIS *Merge* tool
    a. Input Datasets:
        i. Max_Depression_Poly
        ii. Nested_Depression_Poly_1
    b. Output Dataset: Nested_Depression_Poly – This polygon shapefile now represents the nested closed topographic depression boundary polygons which includes the boundary polygon of the maximum closed topographic depressions that did not have nested depressions within
    c. Field Map: all Fields can be removed
17. Option: If the Nested_Depression_Poly polygons are still too large to represent realistic wetland boundaries, then the steps from 9 – 16 can be repeated, using the Nested_Depression_Poly where the Max_Depression_Poly Input was used. These steps were repeated once to determine the nested closed topographic depressions used in the Chapter 3 study.

<u>Determine the depression depths</u>

18. Use the ArcGIS Spatial Analyst *Zonal Statistics* tool
    a. Input raster or feature zone data: specify the depression polygon file, either Nested_Depression_Poly or Max_Depression_Poly. For this tutorial, we will use Nested_Depression_Poly
    b. Zone field: Object_ID
    c. Input value raster: 2-m DEM raster file
    d. Output raster: Depression_Minimum

e. Statistics type: MINIMUM
f. Check/enable Ignore NoData in calculations

19. Use the ArcGIS *Polygon to Line* tool – this step will convert the depression boundary polygons into lines
    a. Input Features: Nested_Depression_Poly
    b. Output Feature Class: Nested_Depression_Line
    c. Check/enable Identify and store polygon neighboring information
20. Use the ArcGIS Spatial Analyst *Zonal Statistics* tool – in this step, the minimum elevation of the polygon **boundary** will be determined, this represents the spillover elevation of the depression
    a. Input raster or feature zone data: Nested_Depression_Line
    b. Zone field: RIGHT_FID
    c. Input value raster: 2-m DEM raster file
    d. Output raster: Boundary_Minimum
    e. Statistics type: MINIMUM
    f. Check/enable Ignore NoData in calculations
21. Use the ArcGIS *Feature To Point* tool
    a. Input Features: Nested_Depression_Poly
    b. Output Feature Class: Nested_Depression_Point
    c. **Check/enable** Inside
22. Use the ArcGIS *Extract Value to Point* tool – this step will ascribe the depression minimum elevation value to the depression point
    a. Input Point Features: Nested_Depression_Point
    b. Input raster: Depression_Minimum
    c. Output point features: Nested_Depression_Point_with_MIN
    d. Uncheck/disable Interpolate values at the point locations
    e. Uncheck/disable Append all the input raster attributes to the output point features
23. Open the Attribute Table of the Nested_Depression_Point_with_MIN shapefile and Add Field
    a. Name: DeprAbsoMIN (abbreviation for depression absolute elevation minimum)
    b. Type: Double

24. Right-click on the new Field and use Field Calculator
    *a.* Within Fields: double click the RASTERVALU to add it to the window below *DeprAbsoMIN* =
    b. Click OK
25. Within the Attribute Table, right-click the RASTERVALU field and Delete Field
26. Use the ArcGIS *Spatial Join* tool to ascribe depression absolute minimum elevation values to the depression polygons
    a. Target Features: Nested_Depression_Poly
    b. Join Features: Nested_Depression_Point_with_MIN
    c. Output Feature Class: Nested_Depression_Poly_with_MIN
    d. Join Operation: JOIN_ONE_TO_ONE
    e. Check/enable Keep All Target Features
    f. Field Map of Join Features: leave as is
    g. Match Option: INTERSECT
    h. Search Radius: blank
    i. Distance Field Name: blank
27. Use the ArcGIS *Feature To Point* tool – this step will create the center point for the depression boundary lines
    a. Input Features: Nested_Depression_Line
    b. Output Feature Class: Nested_Depression_Line_Point
    c. **Check/enable** Inside
28. Use the ArcGIS *Extract Value to Point* tool – this step will ascribe the boundary elevation minimum value to the Nested_Depression_Line_Point
    a. Input Point Features: Nested_Depression_Line_Point
    b. Input raster: Boundary_Minimum
    c. Output point features: Nested_Depression_Line_Point_w_BoundMIN
    d. Uncheck/disable Interpolate values at the point locations
    e. Uncheck/disable Append all the input raster attributes to the output point features
29. Open the Attribute Table of the Nested_Depression_Line_Point_w_BoundMIN shapefile and Add Field
    a. Name: BoundMIN (abbreviation for boundary elevation minimum)

b. Type: Double

30. Right-click on the new Field and use Field Calculator
    a. Within Fields: double click the RASTERVALU to add it to the window below BoundMIN =
    b. Click OK
31. Within the Attribute Table, right-click the RASTERVALU field and Delete Field
32. Use the ArcGIS *Spatial Join* tool to ascribe depression boundary minimum elevation values to the depression polygons
    a. Target Features: Nested_Depression_Poly
    b. Join Features: Nested_Depression_Line_Point_w_BoundMIN
    c. Output Feature Class: Nested_Depression_Poly_w_Abso_and_BoundMIN
    d. Join Operation: JOIN_ONE_TO_ONE
    e. Check/enable Keep All Target Features
    f. Field Map of Join Features: leave as is
    g. Match Option: CLOSEST
    h. Search Radius: 1 m
    i. Distance Field Name: blank
33. Within the Nested_Depression_Poly_w_Abso_and_BoundMIN Attribute Table, in the Table Options, Add Field
    a. Name: Depth
    b. Type: Double
34. Right-click on the Depth Field and select Field Calculator
    a. Enter Within the window below *Depth* = BoundMIN - DeprAbsoMIN
    b. This step will calculate the depression depth by subtracting the depression absolute minimum elevation from the depression boundary minimum elevation
35. Any unnecessary fields can be deleted from the Nested_Depression_Poly_w_Abso_and_BoundMIN shapefile to clean it and the file can be renamed Nested_Depression_Poly which now has the depression depth within a field in its attribute table

## Appendix A-3: Create predicted active spill channel networks and ascribe Strahler order and terminal status to wetland polygons

Five versions of the predicted active spill channel networks (PASCNs) were tested for use in the model for predicting the distributions of solute-rich versus fresh wetlands. The PASCNs differ in terms of the level of hydrologic connectivity expected within the landscape. The PASCNs can be used to determine Strahler orders and terminal statuses of the wetland polygons, which are variables considered in the fresh vs. solute-rich prediction model proposed in Chapter 3. The level of expected hydrologic connectivity of the PASCNs affects these variables. There were 5 PASCNs tested in the model proposed in Chapter 3: Minimum, Near Minimum, Moderate, Near Maximum, and Maximum. There were 2 PASCNs used as predictor variables in the DSM models described in Chapter 4: Minimum and Maximum. For the PASCNs to properly characterize a wetland in terms of its Strahler order and terminal status, it must be generated from the wetland's entire watershed, which includes potential runoff from upslope wetlands. Therefore, the DEM used in the methodology to generate the PASCNs must span the target area's greater watershed.

<u>Resample the road-removed 1-m DEM to 5 m.</u>

1. Import the road-removed 1-m DEM raster (described in Appendix A-1) into ArcGIS
2. Use the ArcGIS *Block Statistics* tool on the DEM. – this step averages the elevation values of the 1-m DEM for a 5 x 5 group of cells, the output will be used in the resampling tool to create a 5-m resolutions DEM
    a. Input raster: road-removed DEM
    b. Output raster: Block_5m_Mean
    c. Neighborhood: Rectangle
        i. Height: 5 m
        ii. Width: 5 m
    d. Statistics type: MEAN
    e. Check/enable Ignore NoData in calculations
3. Resample the Block_5m_Mean raster to create a DEM with a 5-m resolution using the ArcGIS *Resampling* tool
    a. Input raster: Block_5m_Mean

b. Output raster: DEM_5m_NoRoad
c. Output cell size of X and Y as 5 m
d. Resampling Technique: NEAREST – this will use the averaged values provided in the block statistics output

Create the Stream channel network polyline shapefile

This is done using SAGA GIS software. SAGA GIS Software download and installation guides are available at http://www.saga-gis.org/en/index.html

4. Export the DEM_5m_NoRoad as a geotiff from ArcGIS
5. Import the DEM_5m_NoRoad to SAGA GIS
    a. Within the SAGA GIS Tool Libraries:
        i. Import/Export: *GDAL/OGR*
            1. Use the Import Raster tool to import the DEM_5m_NoRoad tif
6. Run the Basic Terrain Analysis on the DEM_5m_NoRoad raster
    a. Within the SAGA GIS Tool Libraries:
        i. Terrain Analysis: *Basic Terrain Analysis*
            1. Grid system: that of the DEM_5m_NoRoad file
            2. >> Elevation: DEM_5m_NoRoad
            3. Options: Channel Density: 3
7. Export the Channel Network as a shapefile
    a. Within the Data Tree, right-click the Channel Network and select save as, save as ESRI shapefile with the name Full_Stream_Channel_Network

Generate the Wetland DEM Ponding Model (WDPM) runoff distribution output polygons

As mentioned, 5 PASCNs were tested for the model proposed in Chapter 3. These were created by altering the Full_Stream_Channel_Network to reflect the runoff distribution outputs from the WDPM. However, the Maximum PASCN is the Full_Stream_Channel_Network unaltered by a WDPM output. Therefore, 4 PASCNs were created using the WDPM output. The WDPM output is an estimate of the runoff distribution on the landscape which can be used to determine connectivity between wetlands. A spill channel connection is predicted to be active if there is continuous runoff distributed along the spill channel. 2 runoff distribution outputs from the

WDPM (1st half and 2nd half) were used to create each version of the 4 PASCNs. Therefore, the model was run 8 times per study area.

8. Use ArcGIS Spatial Analyst *Raster to ASCII* tool to export the road-removed 5-m DEM as an ascii file
    a. Input raster: DEM_5m_NoRoad
    b. Output ASCII raster file: specify the output ASCII file, ensure the file extension is .asc and **not** .txt
    c. The output is referred to as the DEM_5m_NoRoad_ASCII
9. To download the WDPM, send a download request here: https://www.usask.ca/hydrology/WDPM.php
10. Installation guidelines and general information are found within the manual at the website listed above
11. Run the WDPM.exe
    a. Specify the Working Directory the program will save the water distribution ascii files to
    b. Specify *add* as the Method, this will add water to the DEM
    c. DEM File: DEM_5m_NoRoad_ASCII
    d. Water File: NULL

To create the Minimum PASCN, the WDPM needs to be run twice to create the *1st half* and *2nd half* runoff distribution, which will be combined later:

12. Add components:
    a. Output File: Minimum_1st_half
    b. Depth of water (mm): 50 – this parameter specifies the depth of water added per cell to be re-distributed over the DEM
    c. Water runoff fraction: 1
    d. Elevation tolerance (mm): 2000 – this parameter essentially controls the number of iterations or steps that the runoff distribution is moved over. If the maximum change in water depth, at any position within the DEM in an iteration, is smaller than this value, then the model will stop, and the output of the runoff distribution is created. At the tolerance of 2000, the water is mostly redistributed to the

channel and depression positions. At this tolerance, not enough iterations will have been run for the runoff to be distributed throughout the spill channels. This creates the *1st half* WDPM runoff output for the Minimum predicted active spill channel

e. Do not need to set any Subtract or Drain settings, as only the Add method is used.
f. Computation Settings: The WDPM can take very long run-times to create the *2nd half* WDPM runoff outputs (discussed below) for large watersheds. This is a substantial drawback of using this method. Using OpenCL for parallel processing can significantly reduce runtimes. Using OpenCL with GPU reduced the runtimes of the WDPM for this study from days to hours, it is recommended to use this method.
g. Press *Start*

13. Follow the same steps for 11 and 12 to create the 2nd half WDPM runoff output for the Minimum PASCN, except:
    a. Output File: Minimum_2nd_half
    b. Elevation tolerance (mm): 10 – this will stop the model only after the runoff has settled over many iterations. The runoff has run enough iterations to complete the spill channel connections. This step can take very long, and so OpenCL is highly recommended for this step.
14. Repeat steps 12. and 13. 3 times to create the WDPM runoff distribution outputs to create the remaining 3 PASCNs. All steps will be the same except the Output Files will need to be named according to the network to be created and the following Depth of Water (mm) values need to be used per level of connectivity:
    a. Near Minimum: 100 mm
    b. Moderate: 150 mm
    c. Near Maximum: 300 mm
15. In ArcCatalog, specify the projection of each WDPM runoff distribution output raster
16. Import the Minimum_1st_half and Minimum_2nd_half WDPM output rasters into ArcMap
17. Within the properties of each raster, use *Classified* Symbology
    a. There should be one value field for the raster, which reflects the presence of water

b. Specify 2 classes

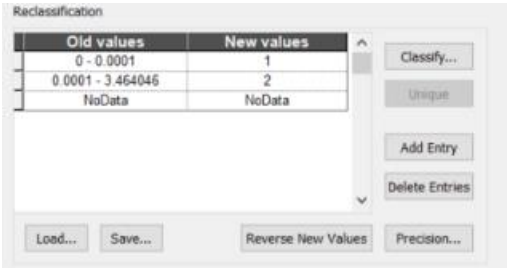
c. Specify the ranges as 0 – 0.0001 as *no water* and 0.0001 – max value as *water*

18. Use the ArcGIS Spatial Analyst *Reclassify* tool

a. Input raster: Minimum_1st_half

b. Reclass field: VALUE

c. Reclassification: the tool should use the class parameters that were specified in the Symbology:




d. Output raster: Min_1st_half_reclass

e. Repeat steps for Minimum_2nd_half raster but save the output raster as: Min_2nd_half_reclass

19. Use the ArcGIS *Raster to Polygon* tool

a. Input raster: Min_1st_half_reclass

b. Field: Value

c. Output polygon features: Min_1st_half_poly

d. Uncheck/disable Simplify polygons option

e. Repeat steps for Min_2nd_half_reclass but name the output Min_2nd_half_poly

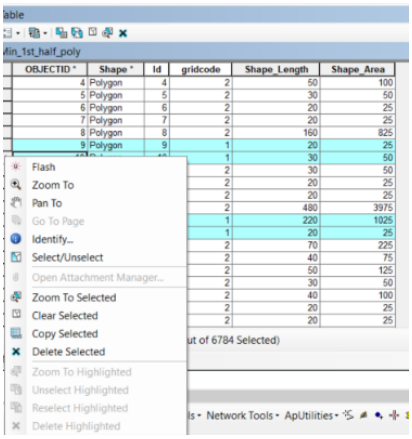
20. Need to delete the polygons that represent *no water*

a. Right-click the Min_1st_half_poly, select Edit Features, and Start Editing

b. Open the Min_1st_half_poly attributes

i. Select by attributes: gridcode = 2 – this highlights all the polygons that represent *water*

ii. Select Switch Selection, now all polygons representing *no water* are highlighted



iii. Right-click a block to the left of a highlighted observation, and select Delete Selected

c. In the Editor toolbar, select Stop Editing and Save Edits

d. Repeat steps for Min_2nd_half_poly

21. Combine the Min_1st_half_poly and Min_2nd_half_poly into one polygon file

    a. Use the ArcGIS Data Management *Merge* tool

        i. Input Datasets: Min_1st_half_poly & Min_2nd_half_poly

        ii. Output Dataset: Minimum_WDPM_Poly

22. Buffer the Minimum_WDPM_Poly by a small amount – this step was done because in some places, some paths of the stream channel networks that will be clipped using the WDPM runoff poly deviated slightly from the path outline by the WDPM runoff polygon

    a. Use the ArcGIS *Buffer* tool

        i. Input features: Minimum_WDPM_Poly

        ii. Output Feature Class: Min_WDPM_Poly_Buff

        iii. Distance: Linear unit: 2.5 m

        iv. Side Type: FULL

        v. End Type: ROUND

        vi. Method: PLANAR

        **vii. Dissolve Type: ALL**

23. Repeat steps 16 – 22 for the Near Minimum, Moderate, and Near Maximum WDPM runoff outputs

<u>Clip the Full_Stream_Channel_Network polyline files by the WDPM runoff outputs</u>

24. Import the Full_Stream_Channel_Network and each of the 4 WDPM runoff outputs into ArcMap

25. Use the ArcGIS *Clip* tool

    a. Input Features: Full_Stream_Channel_Network

    b. Clip Features: Min_WDPM_Poly_Buff

    c. Output Feature Class: Min_Active_Spill_Channel_Network

The Strahler Orders of the Min_Active_Spill_Channel_Network need to be reassigned to reflect relationships of the clipped stream channel network

26. Use ArcGIS *Delete Fields* tool
    a. Input Table: Min_Active_Spill_Channel_Network
    b. Drop Field:
        i. NODE_A
        ii. NODE_B
        iii. BASIN
        iv. ORDER
        v. ORDER_CELL

*27.* Use the ArcGIS *Multipart To Singlepart* tool
    a. Input Feature: Min_Active_Spill_Channel_Network
    b. Output Feature Class: Min_Active_Spill_Channel_Network_Single

28. To re-ascribe Strahler order to the PASCNs, Arc Hydro Tools (Maidment, 2002) will need to be installed. Download and installation information can be found here: http://downloads.esri.com/archydro/archydro/

29. Enable the Arc Hydro Tools toolbar in ArcMap

30. In the Arc Hydro Tools Attribute Tools use *Assign HydroID*:
    a. Dataframe: Layers
    b. Workspace: Specify where to save output
    c. Layers: Select the Min_Active_Spill_Channel_Network_Single
    d. Overwrite Existing HydroID: there should be none, so either
    e. Apply to: All Features

31. In the Arc Hydro Tools Attribute Tools use *Generate From/To Node for Lines:*
    a. Line: Min_Active_Spill_Channel_Network_Single with HydroID

32. In the Arc Hydro Tools Attribute Tools use *Find Next Downstream Line:*
    a. Line: Min_Active_Spill_Channel_Network_Single with HydroID and From/To Nodes

33. In the Arc Hydro Tools Attribute Tools use *Find Next Downstream Line:*
    a. Line: Min_Active_Spill_Channel_Network_Single with HydroID and From/To Nodes

34. Open the Min_Active_Spill_Channel_Network_Single Attribute Table and in the Table Options, Add Field:

a. Name: StrahlerOrder
b. Type: Short Integer

Not all the Arc Hydro Tools, like Assign River Order, are available in the toolbar. Need to access the Toolbox:

35. Right-click in the ArcToolbox and Add Toolbox
    a. Locate where the Arc Hydro Tools were installed to
    b. Select Arc Hydro Tools
36. Within the ArcToolbox, find the Arc Hydro Tools, Select Attribute Tools and use the *Assign River Order* Tool:
    a. Input Feature Class or Table: Min_Active_Spill_Channel_Network_Single
    b. Input River Order Field: StrahlerOrder
    c. River Order Type: **Strahler**
    d. Input Flow Split Table: None
37. Rename the shapefile the Min_Active_Spill_Channel_Network
38. Repeat steps 24 – 36 using the Near Minimum, Moderate, Near Maximum WDPM outputs created in Step 23 to generate the remaining PASCNs
39. The Maximum PASCN was already created as it is the original stream channel network unaltered to reflect WDPM runoff distribution polygon outputs. However, the Strahler order for this network is located within the ORDER field, and **not** the ORDER_CELL field. To keep field name consistent with the other PASCNs:
    a. Open the Full_Stream_Channel_Network Attribute Table
    b. In the Table Options, Add Field
        i. Name: StrahlerOrder
        ii. Type: Short Integer
    c. Right-click on the Field Calculator and enter the ORDER field into the window below StrahlerOrder =
40. Rename the Full_Stream_Channel_Network as Max_Active_Spill_Channel_Network
41. Import the wetland boundary polygons, based on either nested or maximum closed topographic depressions described in Appendix A-2, into ArcMap. For this tutorial, the Nested_Depression_Poly shapefile will be used

Ascribe depression/wetland polygons Strahler order

The Strahler order information from each PASCN will be joined to the wetland boundary polygons

42. Use the ArcGIS *Spatial Join* tool to ascribe the maximum Strahler order value of the intersecting PASCNs to the wetland polygons. This tutorial will use the Min_Active_Spill_Channel_Network as an example
    a. Target Features: Nested_Depression_Poly
    b. Join Features: Min_Active_Spill_Channel_Network
    c. Output Feature Class: Nested_Depression_Poly_Join2Min
    d. Join Operation: JOIN_ONE_TO_ONE
    e. Check/enable Keep All Target Features
    f. Field Map of Join Features: can remove all except StrahlerOrder
        i. Right-click StrahlerOrder and select Merge Rule = Max – the wetland polygons will only be ascribed the maximum Strahler order that they intersect with
    g. Match option: INTERSECT
    h. Search Radius: blank
43. Within the Nested_Depression_Poly, in the Table Options, Add Field
    a. Name: StrahlOrder_Min – this specifies the Strahler order from the join is specific to the Min_Active_Spill_Channel_Network
    b. Type: Short Integer
44. Right-click on the StrahlOrder_Min field and use the Field Calculator, specify the StrahlerOrder field in the window below StrahlOrder_Min =
45. Delete the StrahlerOrder field
46. Repeat steps 42 – 45 using the different PASCNs in the *Spatial Join* tool and specify different StrahlOrder_ fields for each PASCNs. The final Nested_Depression_Poly should have 5 StrahlOrd_ fields, each with the Strahler orders that reflect the maximum value that the polygons intersected with per PASCN

Determine wetland polygon terminal status

Terminal status refers to whether a wetland is terminal or not. A terminal wetland receives runoff contributions from upslope wetlands and does not contribute runoff further downslope. The algorithms for creating the traditional stream channel networks fill all depressions before determining the networks to allow the stream channels to continue beyond any depressions. Therefore, they cannot be used to determine if the wetlands are terminal or not. The PASCNs allow for the wetland polygons to be characterized in terms of their potential for contributing runoff downslope. Here the wetland polygons are determined to be terminal or not based on the 4 PASCNs*:* Minimum, Near Minimum, Moderate, Near Maximum. The methodology outlined here determines for each wetland polygon if any section of the PASCN exits out of the polygon, representing runoff moving out of the wetland. The Maximum PASCN cannot be used to determine if wetlands are terminal or not because it was generated using a traditional approach for determining stream channel networks. According to the Maximum PASCN, no wetland polygons would be terminal.

47. Import the Min_Active_Spill_Channel_Network and the Minimum_WDPM_Poly created in step 21 – this version of the WDPM output is not buffered
48. Use the ArcGIS *Clip* tool
    a. Input Features: Min_Active_Spill_Channel_Network
    b. Clip Features: Minimum_WDPM_Poly
    c. Output Feature Class: Min_Active_Spill_Channel_Network_Clip
49. Use the ArcGIS *Multipart To Singlepart* tool
    a. Input Feature: Min_Active_Spill_Channel_Network_Clip
    b. Output Feature Class: Min_Active_Spill_Channel_Network_Clip_Single
50. Use the ArcGIS *Buffer* tool – the wetland polygons are buffered a small amount to ensure that the channel networks actually move out of the wetland polygon boundary
    a. Input Features: Nested_Depression_Poly
    b. Output Feature Class: Nested_Depression_Poly_Buffer
    c. Distance: Linear Unit: 2.5 m
    d. Side Type: FULL
    e. End Type: ROUND

f. Method: PLANAR
g. Dissolve Type: **NONE**

51. Within ArcGIS Selection, use Select By Location
    a. Selection method: select features from
    b. Target layers: Min_Active_Spill_Channel_Network_Clip_Single
    c. Source layer: Nested_Depression_Poly_Buffer
    d. Spatial selection method for target layer feature(s): **are crossed by the outline of the source layer feature** – this will select any part of the channel network that either enters or leaves the wetland polygons
52. Right-click the Min_Active_Spill_Channel_Network_Clip_Single within the Table Of Contents, select Selection, and Create Layer From Selected Features
53. Use the ArcGIS *Feature vertices to Points* tool – this step will create points at the **starts** of the line segments that either leave or enter the wetland polygons
    a. Input Features: Min_Active_Spill_Channel_Network_Clip_Single_Selection – this was created in step 52
    b. Output Feature Class: Min_Active_Spill_Channel_Network_Start_Points
    c. Point Type: START
54. Use the ArcGIS *Feature vertices to Points* tool – this step will create points at the **ends** of the line segments that either leave or enter the wetland polygons
    a. Input Features: Min_Active_Spill_Channel_Network_Clip_Single_Selection
    b. Output Feature Class: Min_Active_Spill_Channel_Network_End_Points
    c. Point Type: END
55. Use Select Attributes By Location – this step will select any wetland polygon that has a stream segment exiting its boundaries
    a. Selection method: select features from
    b. Target layer(s): Nested_Depression_Poly **not** the Nested_Depression_Poly_Buffer
    c. Source layer: Min_Active_Spill_Channel_Network_**Start_Points**
    d. Spatial selection method for target layer feature(s): **Contain the source layer feature**

56. Open the Nested_Depression_Poly Attribute Table and select *Switch Selection* – now the selection shows the depression polygons that do not have stream segments exiting them. However, these do not necessarily reflect only terminal wetlands. Many of these depressions may not have stream channels entering them, and the stream channels that enter them may be of small Strahler order, and therefore, these depressions are not expected to receive substantial runoff contributions, as is required to be terminal.
57. To select only depression polygons with stream channels entering them:
    a. Again, use Select By Location
        i. Selection method: **select from the currently selected features in**
        ii. Target layer(s): Nested_Depression_Poly
        iii. Source layer: Min_Active_Spill_Channel_Network_**End_Points**
        iv. Spatial selection method for target layer feature(s): **Contain the source layer feature**
58. To ensure that these selected wetlands receive substantial upslope runoff contributions, the depressions polygons with Strahler orders less than three were removed from the selection. In this tutorial example, the terminal status is being determined from the Minimum PASCN, and so the depression polygon Strahler Order from the same PASCN must be used.
    a. Open the Nested_Depression_Poly Attribute Table, use Select By Attributes
        i. Method: **Select from current selection**
        ii. SELECT*FROM Nested_Depression_Poly WHERE: StrahOrder_Min > 2
59. Now all the terminal wetlands based on the Minimum PASCN are selected for the Nested_Depression_Poly. Within the Nested_Depression_Poly Attribute Table, open Table Options, Add Field
    a. Name: Terminal_Min
    b. Type: Text
        i. Field Properties: Length: 20
60. At the bottom of the Attribute Table, select *Show selected records*
61. Right-click the Terminal_Min field and select Field Calculator
    a. Terminal_Min = Terminal
    b. Click OK

62. In the Attribute Table, press Switch Selection
63. Make sure *Show selected records* is still selected and select Field Calculator
    a. Terminal_Min = Non_Terminal
    b. Click OK
64. The steps 47 – 63 are repeated for each PASCN except the Maximum. Within each step, inputs associated with the particular PASCN are used and the fields should be named accordingly
65. In the end, the wetland boundary polygon file should have 9 fields representing the Strahler orders and Terminal status based on the 5 PASCN
    a. StrahlOrder_Min
    b. StrahlOrder_NearMin
    c. StrahlOrder_Moderate
    d. StrahlOrder_NearMax
    e. StrahlOrder_Max
    f. Terminal_Min
    g. Terminal_NearMin
    h. Terminal_Moderate
    i. Terminal_NearMax

# APPENDIX B: GIS METHODOLOGIES TO GENERATE NEW TOPOGRAPHIC ATTRIBUTES USED AS PREDICTOR VARIABLES IN THE DSM STUDY

**Note: the output files generated throughout the procedures (shapefiles, rasters, etc.) are given names in this document to clarify their origin and purpose, some names may be longer than ArcGIS allows for and abbreviations should be used.**

Calculate *Elevation percentile* topographic attribute raster

This attribute calculates each grid cell's elevation percentile in relation to the entire study area watershed from the DEM, and therefore the DEM must be representative of the study area's entire watershed.

1. Import the study area watershed DEM into ArcMap
2. Open the DEM Symbology, select Classified
    a. Classify:
        i. Method: Natural Breaks (Jenks)
        ii. Classes: 100
3. Use the ArcGIS Spatial Analyst *Reclassify* tool
    a. Input raster: DEM
    b. Reclass field: VALUE
    c. Reclassification: After setting the Classification settings in Symbology, the tool should identify the same classes here where New value = 1 should correspond to the smallest elevation class and New value = 100 should correspond to the highest elevation class
    d. Output raster: Elevation_Percentile

Calculate *Elevation from depression spillover* topographic attribute raster

4. Re-create the Max_Depression_Poly described in Appendix A-2, except use the non-road-removed DEM and import into ArcMap
5. Import the non-road-removed DEM into ArcMap (with the resolution of 2 m or 5 m, depending on the target resolution for the predictor variable.
6. Use the ArcGIS *Polygon to Raster* tool
    a. Input Features: Max_Depression_Poly
    b. Value field: Object_ID

c. Output Raster Dataset: Max_Depression_Raster
d. Cell assignment type: CELL_CENTER
e. Priority field: NONE
f. Within Environments:
    i. Processing Extent:
        1. Extent: DEM
        2. Snap Raster: DEM
    ii. Raster Analysis:
        1. Cell Size: DEM

7. Use the ArcGIS *Fill* tool – this will remove any depressions with depths less than 10 cm from the DEM
    a. Input surface raster: DEM
    b. Output surface raster: DEM_10cmFill
    c. Z limit: 0.1
8. Use the ArcGIS Spatial Analyst *Flow Direction* tool
    a. Input surface raster: DEM_10cmFill
    b. Output flow direction raster: Flow_Direction
    c. Uncheck/disable Force all edge cells to flow outward
    d. Output drop raster: blank
9. Use the ArcGIS Spatial Analyst *Watershed* tool – this step will create a raster output which defines the individual contributing areas for each depression. Some areas along the edge of the DEM will not be assigned to a watershed because their associated depression is not within the DEM.
    a. Input flow direction raster: Flow_Direction
    b. Input raster or feature pout point data: Max_Depression_Raster
    c. Pour point field: blank
    d. Output raster: Depression_Watersheds_Raster
10. Use the ArcGIS *Raster to Polygon* tool
    a. Input raster: Depression_Watersheds_Raster
    b. Output polygon features: Depression_Watersheds_Poly
    c. **Uncheck/disable** Simplify polygons

11. Use the ArcGIS *Polygon to Line* tool
    a. Input Features: Max_Depression_Poly
    b. Output Feature Class: Max_Depression_Line
    c. Check/enable Identify and store polygon neighboring information
12. Use the ArcGIS *Generate Points Along Lines* tool
    a. Input Features: Max_Depression_Line
    b. Output Feature Class: Depression_Line_Points
    c. Point Placement: DISTANCE
    d. Distance: 2 meters
    e. Uncheck/disable End Points
13. Use the ArcGIS *Zonal Statistics* tool to determine the minimum elevation along the boundaries of the wetland depression polygons. This represents the spillover elevation of each wetland polygon
    a. Input raster or feature zone data: Max_Depression_Line
    b. Zone field: RIGHT_FID
    c. Input value raster: DEM
    d. Output raster: Depression_Line_Min_Elevation
    e. Statistics type: MINIMUM
    f. Check/enable Ignore NoData
14. Use the ArcGIS *Extract Values to Points* tool
    a. Input point features: Depression_Line_Points
    b. Input raster: Depression_Line_Min_Elevation
    c. Output point features: Depression_Line_Points_with_Spill_Elevation
    d. Uncheck/disable: *Interpolate…* and *Append all…*
15. Open the Depression_Line_Points_with_Spill_Elevation Attribute Table and Add Field in the Table Options
    a. Name: Spillover_Elevation
    b. Type: Double
16. Right-click the Spillover_Elevation field and use Field Calculator
    a. Enter the RASTERVALU field in the window below Spillover_Elevation =
17. Delete the RASTERVALU field

18. Use the ArcGIS *Spatial Join* tool – this step is done to ascribe the spillover elevation values for the depressions to their watershed polygon
    a. Target Features: Depression_Watersheds_Poly
    b. Join Features: Depression_Line_Points_with_Spill_Elevation
    c. Output Feature Class: Depression_Watersheds_Poly_with_Spill_Elevation
    d. Join Operation: JOIN_ONE_TO_ONE
    e. Check/enable Keep All Target Features
    f. Field Map of Join Features: Right-click Spillover_Elevation, select Merge Rule, select Mode
    g. Match Option: Completely Contains
    h. Search Radius: blank
    i. Distance Field Name: blank
19. Use the ArcGIS *Polygon to Raster* tool
    a. Input Features: Depression_Watersheds_Poly_with_Spill_Elevation
    b. Value field: Spillover_Elevation
    c. Output Raster Dataset: Depression_Watersheds**_Raster**_with_Spill_Elevation
    d. Cell assignment type: CELL_CENTER
    e. Priority field: NONE
    f. Within Environments…
        i. Processing Extent:
            1. Extent: DEM
            2. Snap Raster: DEM
        ii. Raster Analysis:
            1. Cell Size: DEM
20. Use the ArcGIS *Raster Calculator* tool – this step will create a raster that reflects each grid cell's elevation above or below the spillover elevation of its associated depression. Grid cells are associated to depressions by the depression contributing area that they fall within
    a. Enter the expression: "%DEM%" – "%Depression_Watersheds_Raster_with_Spill_Elevation%"
        i. The names need to match the raster files exactly

ii. If this expression does not work, enter the Layers manually by double-clicking them in the Layers and variables window

b. Output raster: Elevation_from_depression_spillover

This output will have no data for watersheds on the edge of the DEM that were missing depressions. The raster surface cannot have no data cells to be used as a predictor variable. The areas with missing data were interpolated from the surrounding areas. This step was repeated for all the predictor variables that were based on the depression watersheds

21. Use the ArcGIS *Focal Statistics* tool
    a. Input raster: Elevation_from_depression_spillover
    b. Output raster: Elevation_from_depression_spillover_interpolated
    c. Neighborhood: Rectangle
    d. Neighborhood Settings: - these may need to be adjusted if the no data areas on the edge of the DEMs are significant in size
        i. Height: 350
        ii. Width: 350
    e. Statistics type: MEAN
    f. Check/enable Ignore NoData in calculations
22. Use the ArcGIS *Raster Calculator* tool – this step creates a raster where the values of Elevation_from_depression_spillover are used unless no data exists for a grid cell, if no data exists, then values from Elevation_from_depression_spillover_interpolated are used
    a. Enter the Expression: Con(IsNull("%Elevation_from_depression_spillover %"),"%Elevation_from_depression_spillover_interpolated%","%Elevation_from_depression_spillover%")
    b. Output raster: Elevation_from_depression_spillover_filled
23. Use the ArcGIS *Extract by Mask* tool
    a. Input raster: Elevation_from_depression_spillover_filled
    b. Input raster or feature mask data: DEM
    c. Output raster: Elevation_from_depression_spillover_final

Create the *Elevation above basin bottom* topographic attribute raster

24. Use the ArcGIS *Zonal Statistics* tool – this step creates a raster with the minimum elevation values per depression watershed
    a. Input raster or feature zone data: Depression_Watersheds_Poly_with_Spill_Elevation
    b. Zone field: Id
    c. Input value raster: DEM
    d. Output raster: Watershed_Minimum_Elevation
    e. Statistics type: MINIMUM
    f. Check/enable Ignore NoData
    g. Within Environments…
        i. Processing Extent:
            1. Extent: DEM
            2. Snap Raster: DEM
        ii. Raster Analysis:
            1. Cell Size: DEM
25. Use the ArcGIS *Raster Calculator* tool – this step creates a raster surface of the elevation for each grid cell above its associated depression minimum elevation
    a. Enter the expression: "%DEM%" – "%Watershed_Minimum_Elevation%"
        i. The names need to match the raster files exactly
        ii. If this expression does not work, enter the Layers manually by double-clicking them in the Layers and variables window
    b. Output raster: Elevation_above_basin_bottom
26. Repeat steps 21 – 23 but use the Elevation_above_basin_bottom raster and specify the final output as Elevation_above_basin_bottom_final

Create the *Depression Depth* topographic attribute raster

27. Use the ArcGIS *Raster Calculator* tool – this creates a raster surface where every grid cell within a contributing area will be given the depression depth value of the associated depression

a. Enter the Expression: "%Depression_Watersheds_Raster_with_Spill_Elevation%" - "%Watershed_Minimum_Elevation%"
b. Output raster: Depression_Depth_1

28. Use the ArcGIS *Raster Calculator* tool – for any cell with a depth less than 0.01, this step will set to 0.01
    a. Enter the Expression: Con("%Depression_Depth_1%" <= 0.01, 0.01, "% Depression_Depth_1%")
    b. Output raster: Depression_Depth
29. Repeat steps 21 – 23 but use the Depression_Depth raster and specify the final output as Depression_Depth_final

Create the *Elevation above basin bottom / depth* topographic attribute raster

30. Use the ArcGIS *Raster Calculator* tool
    a. Enter the Expression: "%Elevation_above_basin_bottom%" / "%Depression_Depth%"
    b. Output raster: Elevation_above_basin_bottom_per_depth
31. This output was smoothed using the ArcGIS *Focal Statistics* tool
    a. Input raster: Elevation_above_basin_bottom_per_depth
    b. Output raster: Elevation_above_basin_bottom_per_depth
    c. Neighbourhood: Rectangle
    d. Height: 3
    e. Width: 3
    f. Statistics type: MEAN
    g. Check/enable Ignore NoData in calculations
32. Repeat steps 21 – 23 but use the Elevation_above_basin_bottom_per_depth raster and specify the final output as Elevation_above_basin_bottom_per_depth_final

Create the *Depression max catchment area* topographic attribute

33. Export the DEM from ArcGIS as a .tif
    a. Right-click on the DEM in the Table of Contents and select Data and Export Data

i. Specify a folder, not a geodatabase to save it to
ii. Format: TIFF

34. Import the DEM.tif into SAGA GIS
    a. Within the Tool Libraries
        i. Import/Export: GDAL/OGR:
            1. Import Raster
35. Generate the catchment area raster for the DEM
    a. Within the Tool Libraries
        i. Garden
            1. Introducing Tool Programming
                a. 08: Extended neighbourhoods – catchment areas (parallel)
                    i. Grid systems: specify the grid system of the DEM
                    ii. >> Elevation grid: DEM
                    iii. << Catchment are: <create>
                    iv. Method: **MFD**, not D8
                    v. Press Okay
36. Within the Tool Libraries
    a. Import/Export: GDAL/OGR:
        i. Export GeoTIFF
            1. Grid system: specify the grid system of the DEM and Catchment area rasters
            2. >> Grid(s): Move Catchment area to the right-window, press Okay
            3. File: Specify where to save the tif and name Catchment_area.tif
37. Import the Catchment_area.tif into ArcMap
38. Use the ArcGIS *Zonal Statistics* tool – this step will create a raster surface where every grid cell within a depression polygon will be given the maximum catchment area of the entire polygon
    a. Input raster or feature zone data: Max_Depression_Poly
    b. Zone field: OBJECTID
    c. Input value raster: Catchment_area.tif
    d. Output raster: Depression_max_catchment

e. Statistics type: MAXIMUM
f. Within Environments…
    i. Processing Extent:
        1. Extent: DEM
        2. Snap Raster: DEM
    ii. Raster Analysis:
        1. Cell Size: DEM

39. Use the ArcGIS *Raster Calculator* tool – this step will assign 0 to any area outside of the depression polygon boundaries
    a. Enter the expression: Con(IsNull("%Depression_max_catchment %"),0,"% Depression_max_catchment %")
    b. Output raster: Depression_max_catchment_with_0
    c. Within Environments…
        i. Processing Extent:
            1. Extent: DEM
            2. Snap Raster: DEM
        ii. Raster Analysis:
            1. Cell Size: DEM
            2. Mask: DEM
    d. Use the ArcGIS *Extract by Mask* tool – this step is taken because sometimes defining the mask within the environments does not work
        i. Input raster: Depression_max_catchment_with_0
        ii. Input raster or feature mask data: DEM
        iii. Output raster: Depression_max_catchment_final

<u>Create the Wetland Strahler order minimum and maximum topographic attribute rasters</u>

These topographic attributes were generated using the methodology developed for the model proposed in Chapter 3 described in Appendix A-3. These attributes characterize wetlands in terms of hydrologic contributions they potentially receive from their entire watershed, including potential contributions from upslope wetlands, which are not captured by the watershed calculation done in step 10. Therefore, the methodologies require a DEM input that spans the

study area's entire watershed. Only the wetland Strahler orders determined from the Minimum and Maximum Predicted Active Spill Channel Networks (PASCNs) were incorporated into the DSM because the wetland Strahler orders determined from the other PASCNs do not differ substantially and would provide too many colinear predictor variables for the modelling. To use the wetland Strahler orders as predictor variables in the DSM models, they must be in raster format.

40. Import the Max_Depression_Poly with the Strahler orders ascribed from the PASCNs through the methodology described in Appendix A-3
41. Use the ArcGIS *Polygon to Raster* tool
    a. Input Features: Max_Depression_Poly
    b. Value field: StrahlOrd_Min
    c. Output Raster Dataset: Max_Depression_Strahl_Ord_Min_Raster
    d. Cell assignment type: CELL_CENTER
    e. Priority field: NONE
    f. Within Environments…
        i. Processing Extent:
            1. Extent: DEM
            2. Snap Raster: DEM
        ii. Raster Analysis:
            1. Cell Size: DEM
42. Use the ArcGIS *Raster Calculator* tool – this step will change any grid cell not within a depression polygon boundary to have a value of 0
    a. Enter the Expression: Con(IsNull("Max_Depression_Strahl_Ord_Min_Raster "),0," Max_Depression_Strahl_Ord_Min_Raster ")
    b. Output raster: Wetland_Strahler_order_Min
43. Repeat steps 41 – 42 using the StahlOrd_Max value for the Max_Depression_Poly within step 41 to create the Wetland_Strahler_order_Max

Each topographic variable raster surface can be exported from ArcMap as a geotiff file:

44. Right-click on the raster in the Table of Contents and select Data and Export Data
    a. Specify a folder, not a geodatabase to save it to, Format: TIFF

# APPENDIX C: SOIL PROFILE INFORMATION

Soil profile descriptions were uploaded to the Saskatchewan Soil Information System (SKSIS) at sksis.usask.ca (University of Saskatchewan, 2018). The soil pit descriptions are georeferenced and information for each pit can be accessed by selecting the point within the map. Future versions of SKSIS will allow for querying soil point data by uploader and project name, at which time the points would be under Jeremy Kiss *Predictive digital soil mapping of wetland soil types in the Canadian Prairie Pothole Region*. Until then, the maps below indicate the sample locations for the soil profiles taken for the DSM study described in Chapter 4.

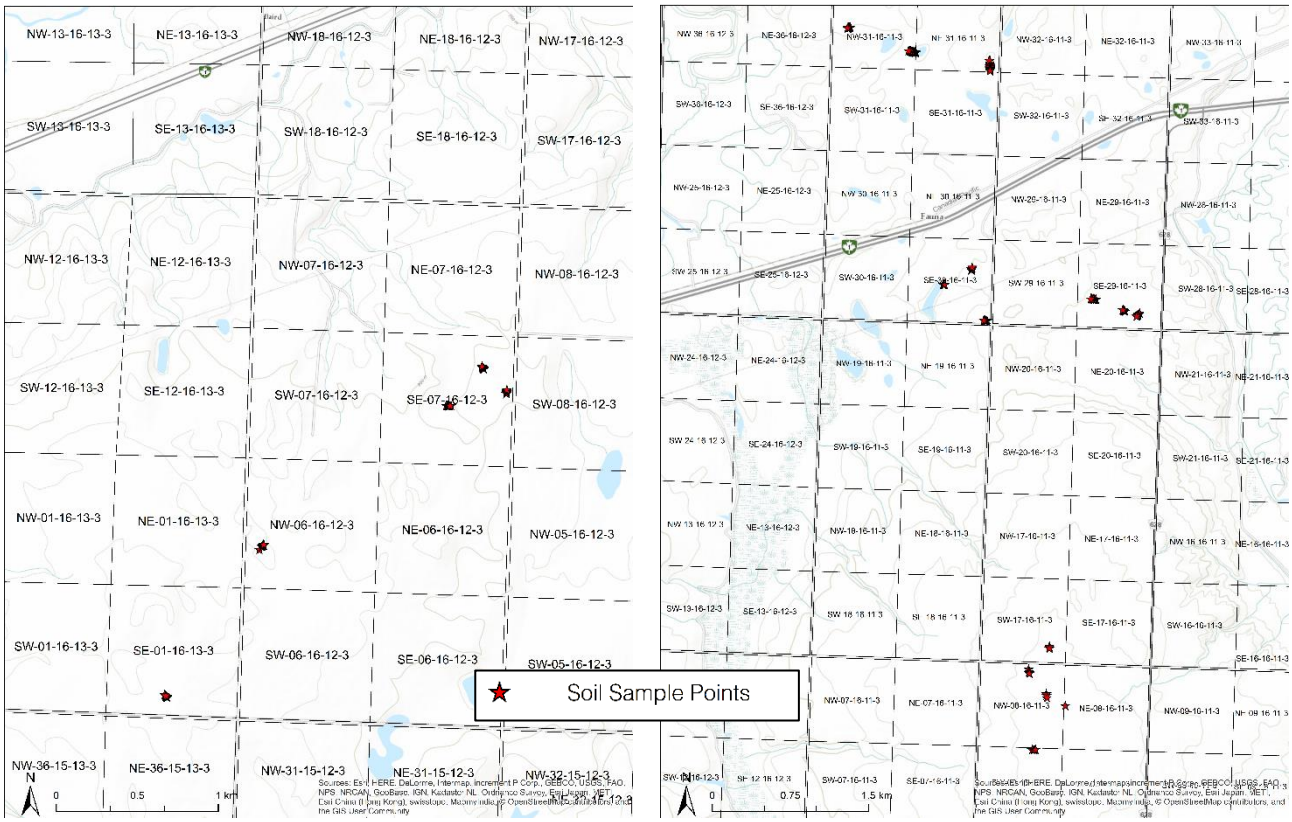



**Fig. C.1 Soil profile locations at Swift Current study area. Legal land descriptions are formatted: Quarter section – section – township – range – west of the 3rd meridian.**

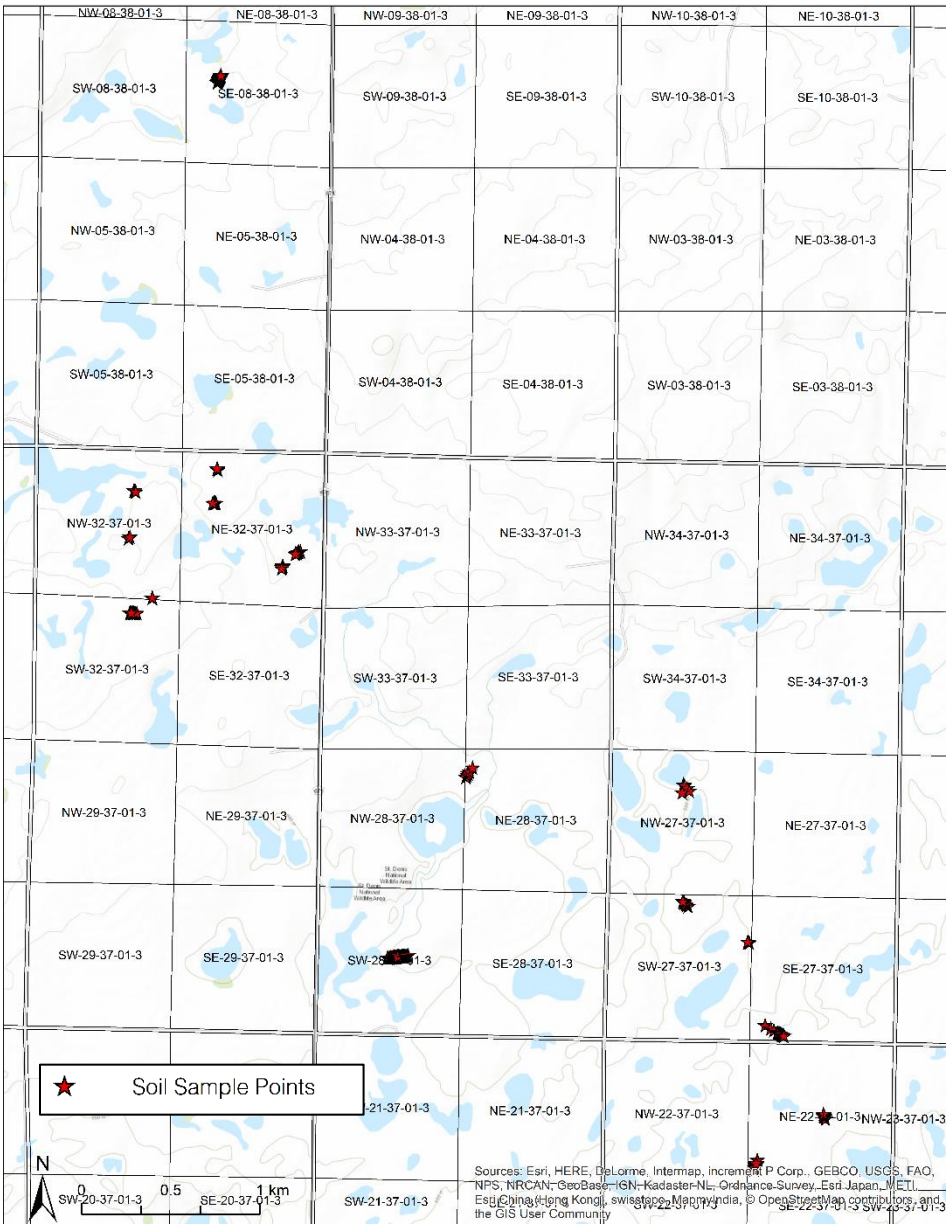



**Fig. C.2 Soil profile locations at the St. Denis study area. Legal land descriptions are formatted: Quarter section – section – township – range – west of the 3rd meridian.**

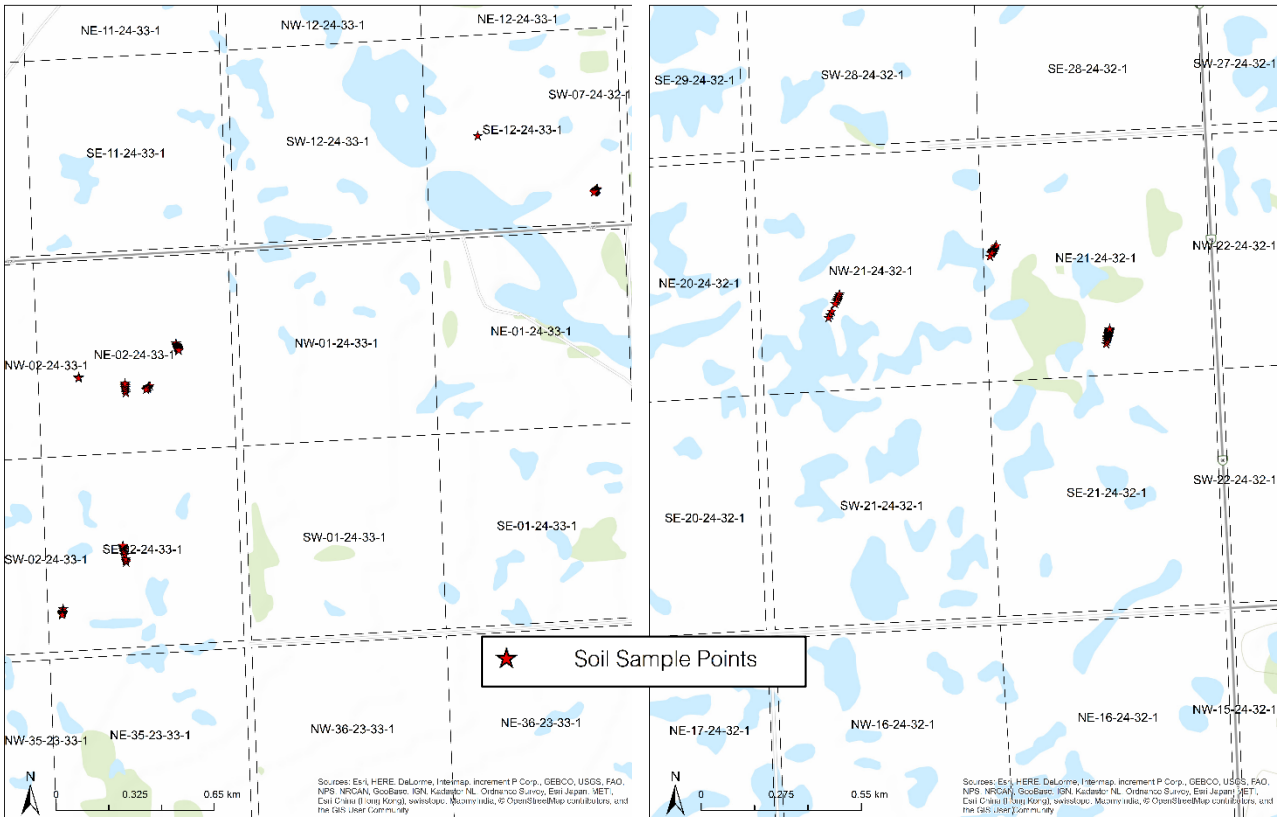



**Fig. C.3 Soil profile locations at the Smith Creek study area. Legal land descriptions are formatted: Quarter section – section – township – range – west of the 1st meridian.**