

Comparing object- and pixel-based random forest landcover classification techniques for better pre-flooding carbon balance estimation of a boreal reservoir

Lia Sophie Hamel

2026

Department of
Earth and Environmental Sciences
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden



Lia Sophie Hamel (2026)

Comparing object- and pixel-based random forest landcover classification techniques for better pre-flooding carbon balance estimation of a boreal reservoir

Bachelor's degree thesis 15 credits in Physical Geography and Ecosystem Science

Department of Earth and Environmental Sciences

Lund University

Level: Bachelor of Science (BSc)

Course duration: March 2026 until June 2026

Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed herein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Comparing object- and pixel-based random forest
landcover classification techniques for better pre-
flooding carbon balance estimation of a boreal reservoir

Lia Sophie Hamel

Bachelor thesis, 15 credits, in Physical Geography and Ecosystem Science

Supervisor

Jonathan Seaquist

Department of Earth and Environmental Sciences

Lund University

Examiner

Anna Terskaia

Department of Earth and Environmental Sciences

Lund University

Acknowledgement

I would like to thank my supervisor Jonathan Seaquist for his invaluable support, constructive feedback and encouraging words throughout the entire project. No matter the struggle I faced, he always found a way to make me feel more positive and I consistently left our weekly Monday morning meetings feeling optimistic.

Abstract

The impoundment of hydroelectric reservoirs is associated with significant emissions of greenhouse gases. In addition, the loss of the pre-flooded landscape is a substantial factor in the long-term carbon footprint of hydroelectric reservoirs. The Paix-des-Braves reservoir in Québec was one of the first reservoirs to be studied in regards of its total carbon footprint, combining pre- and post-impoundment carbon balance estimates. However, carbon balance estimates for the pre-impoundment landscape of the Paix-des-Braves reservoir currently rely on landcover information which has not been assessed for its accuracy or otherwise validated. This leads to uncertainties in the accuracy of the carbon balance estimates. Therefore, this study aimed to improve the landcover classification for the pre-impoundment landscape of the Paix-des-Braves reservoir in Québec by using Landsat 5TM imagery and auxiliary data for object- and pixel-based random forest landcover classification. The study was able to produce a map accurate enough for carbon balance estimates and proposes two approaches for accuracy assessment without field validation. Both the out-of-bag error per class and the correlation matrices using a subset of the training data as validation yielded very high accuracies for both modelling approaches. Based on the updated landcover information, the pre-impoundment PDB reservoir landscape was found to be a sink of CO₂ and a small source of CH₄. Furthermore, differences in landcover classification were found to clearly impact carbon balance results with diffusive flux values for CH₄ being outside the range of previous studies. The methodology proposed in this study can be transferred to other boreal hydroelectric reservoirs to improve our understanding of the carbon footprint of hydroelectric energy.

Keywords: Hydroelectric reservoir, carbon balance, greenhouse gases, random forest, boreal, Québec, PDB, LULC, remote sensing

Contents

1	INTRODUCTION	1
2	BACKGROUND	3
2.1	Carbon balance of hydroelectric reservoirs	3
2.1.1	The flooded landscape	4
2.1.2	The pre-flooded landscape	5
2.2	Landcover classification	5
2.2.1	Interpreting remotely sensed imagery	5
2.2.2	Image classification algorithms	7
2.2.3	Object- vs. pixel-based classification approaches	9
2.2.4	Improving the classification	9
3	DATA & METHODS	10
3.1	Study area	10
3.2	Data	12
3.2.1	Landsat 5 TM imagery	12
3.2.2	Canadian National Fire Database	13
3.2.3	National Topographic Database	14
3.2.4	Digital elevation model	16
3.2.5	LakeAtlas	16
3.3	Methods	16
3.3.1	Map quality and stability	16
3.3.2	Accuracy assessment	25
3.3.3	Carbon balance of the pre-flooded landscape	26
4	RESULTS	27
4.1	Map quality and stability	27
4.1.1	Map stability	29
4.2	Accuracy assessment	30
4.3	Carbon balance of the pre-flooded landscape	30

4.3.1	Sensitivity of carbon balance to landcover classification	35
5	DISCUSSION	36
5.1	Map quality and stability	36
5.1.1	Map stability	36
5.1.2	Possible improvements	38
5.2	Accuracy assessment	39
5.3	Carbon balance	41
5.4	Limitations	44
5.5	Outlook	45
6	CONCLUSION	46
	References	47
	Appendix	i
A	Additional maps of the study area	i
B	Parameters for modeling and model output	vi
C	Carbon balance	vii

Abbreviations

CNFD	Canadian National Fire Database
DEM	Digital elevation model
ER	Ecosystem respiration
FCC	False color composite
GEE	Google Earth Engine
GHG	Greenhouse gases
GPP	Gross primary productivity
LULCC	Land use and land cover change
ML	Machine learning
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
MSAVI	Modified Soil-Adjusted Vegetation Index
NIR	Near-infrared
NTDB	National Topographic Database
OB	Object-based
OM	Organic matter
OOB	Out-of-bag
PCA	Principal component analysis
PB	Pixel-based
PDB	Paix-des-Braves
RF	Random forest
RS	Remote sensing
SWIR	Shortwave-infrared
TM	Thematic mapper
UA	User's accuracy

1. INTRODUCTION

Land use and landcover change (LULCC) has contributed 33% of total global CO₂ emissions over the last 150 years (IPCC, 2000). The primary drivers of LULCC emissions are deforestation for the expansion of agricultural lands (Houghton and Nassikas, 2017), urbanization (Khan et al., 2024), and forest degradation due to wood harvesting practices (Houghton et al., 2012). However, the effect of LULCC on ecosystem carbon dynamics is visible in many different sectors, and LULCC changes related to the impoundment of hydroelectric reservoirs, for example, can also lead to substantial emissions (Tremblay (2005), Demarty et al. (2026), Kumar et al. (2023)). No matter the anthropogenic activity it is related to, LULCC often leads to the transformation of natural carbon sinks to net carbon sources. This can happen through, e.g., the burning of biomass (Chen et al., 2013), loss of soil organic carbon (Houghton, 1995), and the reduction of the area’s carbon density (Houghton et al., 2012). However, uncertainties regarding emissions related to LULCC are still high, with recent estimates ranging in the order of $\pm 0.7 \text{ Gt C yr}^{-1}$ (Friedlingstein et al., 2023), which corresponds to roughly 6% of total anthropogenic emissions (Friedlingstein et al., 2023).

Remote sensing (RS) technology can provide data across large spatial and temporal scales and can thus be very helpful in mapping LULCC (Khorram, 2012). One factor influencing the quality of LULCC assessment, and thus the certainty of carbon balance estimates related to LULCC, is the unreliability of LULCC data (IPCC, 2021). Historically, LULCC mapping often relied on national statistical data, which tended to be non-spatial, aggregating LULCC data from different regions on a national level and thus making the attribution of emissions to specific LULCC activities difficult (Houghton et al., 2012). Furthermore, these historic datasets reported net changes in landcover, which implies that the simultaneous creation and abandonment of two plots of agricultural land, for example, would not be reflected in the data. This can lead to an underestimation of LULCC emissions (Houghton et al. (2012), Houghton and Nassikas (2017)). Moreover, national differences in the type of change monitored and the landcover and land use considered can lead to substantial differences in LULCC carbon balance estimates (IPCC, 2022). For example, failure to accurately track forest management practices can lead to an underestimation of global net LULCC emissions

by 25%–35% (Houghton et al., 2012).

With the advent of globally available, high-resolution satellite imagery, the estimation of global and local LULCC dynamics has improved drastically, and modern LULCC emission estimates are substantially less uncertain (IPCC, 2022). In the US, improved LULCC mapping has reduced uncertainties in carbon sink estimates by roughly 50% (Sleeter et al., 2018). Nonetheless, inconsistencies in mapping technique, methodology, and data sources remain, leading to continued uncertainties in LULCC impact assessment (IPCC, 2022).

One anthropogenic system in which LULCC-related emissions are particularly uncertain is hydroelectric reservoirs (IPCC, 2022). While this issue has received increasing attention over the last decades, full life-cycle impact assessments, taking the loss of the pre-flooded landscape over the entire lifespan of the hydroelectric reservoir into account, are rare (Teodoru et al., 2012). While hydroelectric power is generally considered a low-carbon energy source (IPCC, 2022), it has been established that the physical and biological changes caused by the impoundment of hydroelectric reservoirs lead to significant greenhouse gas (GHG) emissions (Teodoru et al. (2012), Kumar et al. (2023), Demarty et al. (2026)). Additionally, the loss of the pre-impoundment landscape often represents a loss of a net carbon sink, further altering the net carbon footprint of hydroelectric reservoirs. As a result, greenhouse gas emissions from hydroelectric reservoirs are explicitly considered in recent reports by the Intergovernmental Panel on Climate Change (IPCC, 2022).

As mentioned above, there are major challenges related to accurate LULCC assessment. This holds particularly true for hydroelectric reservoirs, which tend to be very remote, making the collection of field data for accurate landcover mapping difficult (Tremblay, 2005). Furthermore, many reservoirs were flooded during the mid-to-late 20th century (Kumar et al. (2023), Tremblay (2005)), and pioneering studies, assessing changes in the local carbon balance, did not have access to modern mapping technologies to accurately assess LULCC. One such case is the Paix-des-Braves (PDB) reservoir (formerly Eastmain-1) in boreal Québec, which was one of the first reservoirs to be extensively studied with regard to its long-term carbon balance (Teodoru et al., 2012). While many of the original findings presented by Teodoru et al. (2012) have since been validated by subsequent studies, the uncertainty in the accuracy of their landcover map, used to assess the carbon balance of the pre-impoundment

landscape, has not been addressed. This leads to large uncertainties in the net carbon footprint estimates of the PDB reservoir.

Hydropower is the largest renewable energy source globally, and an improved understanding of the impact landcover change has on its carbon footprint is paramount for assessing the "greenness" of this energy source. The PDB reservoir provides a unique opportunity for assessing this since numerous emission studies have been conducted here whose quality can be elevated by an improved understanding of the pre-impoundment landcover. Therefore, the aim of this study is to produce a landcover map which can be used for carbon balance estimates of the pre-impoundment PDB reservoir landscape. This study will provide answers to the following research questions:

1. Can a supervised landcover classification approach produce a map of sufficient quality and stability to support reliable carbon balance estimates?
2. How can map accuracy be assessed given the limitations in reference data and the lack of field validation?
3. To what extent does a new landcover classification alter previous carbon balance estimates of the pre-flooded PDB reservoir landscape?

2. BACKGROUND

The following provides some background information on how landcover information can be used to estimate carbon balances and which aspects are particularly important to take into account in the context of hydroelectric reservoirs. The carbon balance of boreal hydroelectric reservoirs is explained and previous findings, regarding the carbon balance of the pre-impoundment PDB reservoir specifically, are presented. Finally, background information regarding the use of remotely sensed imagery for landcover mapping are provided and different classification techniques and strategies are explored.

2.1. Carbon balance of hydroelectric reservoirs

To apply the information derived from landcover classification to carbon balance estimates, the so-called "bookkeeping" model is often used (Houghton et al., 2012). The carbon bal-

ance of each specific landcover type is measured and values are extrapolated over the entire area to arrive at a total carbon balance of the landscape prior to LULCC (Teodoru et al., 2012). Then, landcover data is updated based on LULCC information and an assessment of the carbon balance after the change is made and compared to the pre-change balance. Thus, the net carbon impact of the LULCC can be assessed (Houghton et al., 2012). In the case of hydroelectric reservoirs, an assessment of the long-term carbon balance of the reservoir is of particular relevance to assess the total carbon footprint of the hydroelectric energy produced (Teodoru et al., 2012). Therefore, the impact of the loss of the pre-flooded landscape as well as the carbon balance of the created reservoir must be assessed over the entire lifespan of the reservoir (Teodoru et al., 2012). While this study focuses on the carbon balance of the pre-impoundment landscape, a brief explanation of diffusive carbon fluxes post-impoundment is provided to better contextualize the importance of long-term carbon estimates for hydroelectric reservoirs.

2.1.1. The flooded landscape

In general, GHG emissions from hydroelectric reservoirs are the result of a net imbalance between carbon inputs and production (sources) versus carbon storage and absorption (sinks) (Wang et al., 2024). Most reservoirs are net heterotrophic, meaning the rate of ecosystem respiration (ER) exceeds the rate of gross primary productivity (GPP) (Barbosa et al., 2023). Thus, carbon mineralization is higher than carbon fixation leading to a surplus of carbon which is released into the atmosphere as CO_2 or CH_4 (Barbosa et al., 2023). GHG emissions from hydroelectric reservoirs generally follow three main pathways: Diffusive emissions, ebullition (bubbling) and degassing (Demarty and Tremblay, 2017). For boreal reservoirs, diffusive emissions have been found to be the dominant emission pathway (Demarty et al., 2026).

Diffusive surface emissions are driven primarily by the decomposition of organic matter (OM) in the reservoir (Wang et al. (2024), Barbosa et al. (2023)). Particularly in young reservoirs, recently flooded vegetation and soil lead to a sudden high availability of labile OM which is easily decomposed by micro-bacteria, thus increasing the reservoir’s ER (Barbosa et al., 2023). Diffusive surface emissions from reservoirs tend to be highest within the first

10 years after impoundment (Teodoru et al., 2012) since this is when autochthonous carbon (carbon produced within the reservoir) is most readily available (Wang et al., 2024). This trend in rapidly decreasing diffusive emissions during the first 10 years post-impoundment has been confirmed for the PDB reservoir (Demarty et al., 2026). Furthermore, for the PDB reservoir as well as for the La Romaine reservoir complex in Québec, spatial heterogeneity of diffusive emissions has been related to the pre-impoundment landcover (Teodoru et al. (2012), Rust et al. (2021)).

2.1.2. The pre-flooded landscape

The carbon balance of the pre-impoundment landcover influences the net carbon footprint of hydroelectric reservoirs by representing the loss of a carbon sink or source (Teodoru et al., 2012). As mentioned in the introduction, Teodoru et al. (2012) was the first to include this in a carbon life-cycle assessment of hydroelectric reservoirs. Their carbon balance estimates for the pre-flooded PDB reservoir found the landscape to have been a slight sink of carbon. However, the error margins of their estimates were so high that they assumed the landscape to be roughly neutral in terms of carbon for their total carbon estimates (Teodoru et al., 2012). It is important to note that their error calculations are solely based on uncertainties related to GHG flux estimates while uncertainties related to the landcover of the pre-flooded PDB landscape are not addressed. A main goal of this study is to reassess their carbon balance estimate of the pre-impoundment PDB landscape based on updated landcover data, and to discuss the sensitivity of this balance to uncertainties in landcover classification.

2.2. Landcover classification

2.2.1. Interpreting remotely sensed imagery

The basic premise of landcover mapping based on remotely sensed imagery is the distinct spectral signature of different landcover types (Richards, 2022). Depending on the landcover type, light is absorbed and reflected to different degrees at different wavelengths which makes it possible to tell them apart using remotely sensed imagery (Richards, 2022). A general visualization of spectral signatures of different landcover types taken from Richards (2022)

is shown in Figure I. Considering the landcover classes which are assumed to be prevalent around the PDB reservoir (Teodoru et al., 2012) some distinct spectral patterns are to be expected.

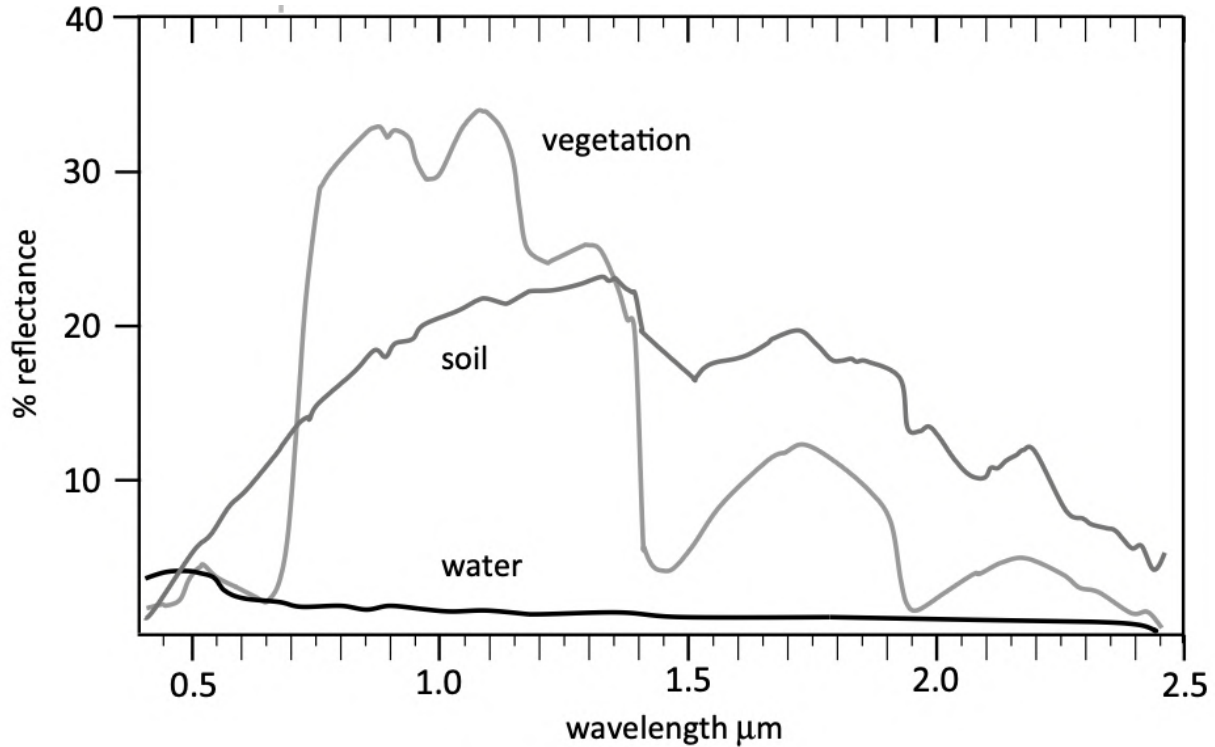


Figure I: Generalized spectral reflectance of water, bare ground and vegetation. Graphic taken from Richards (2022)

Firstly, the reflectance curve for healthy green vegetation shows strong absorption of blue and red light, making plants appear green, and high reflectance in the Near-Infrared (NIR) spectrum (0.7–1.3 μm) (Roy, 1989). The specific reflectance curve of each individual plant depends on plant pigmentation, water content and cell structure (Richards, 2022). While chlorophyll content influences blue and red light absorbance, cell structure is responsible for the high NIR reflectance (Richards, 2022). A key feature of healthy vegetation is the "red edge," a very abrupt increase in reflectance near 0.7 μm (Roy, 1989). This is shown by both deciduous and coniferous trees though coniferous trees tend to have relatively lower NIR reflectance and higher absorbance across the visible spectrum than deciduous trees in a summer scene (Khorram, 2012). Given that the spectral signature of plants is determined by their internal structure, a main factor influencing the spectral signature of green vegetation is

vegetation health. Burned vegetation exhibits much lower NIR reflectance since fire destroys the cell structures usually responsible for high NIR reflectance (Richards, 2022). Fire also drastically reduces the amount of chlorophyll in the plant’s cells, leading to less absorption of red light, making burned vegetation appear more purple on remotely sensed imagery (Richards, 2022).

Water is characterized by an almost complete absorption in the NIR range and a very low reflectance in the visible spectrum, thus appearing very dark in remotely sensed imagery (Richards, 2022). Bare soil usually shows a linear increase in reflectance across the electromagnetic spectrum. Depending on the soil’s water content, the same absorption bands (1.4, 1.9, 2.7 μm) as for healthy green vegetation can be present (Richards, 2022). Vegetated wetlands, such as bogs, fens, and marshes, which are present in the study area (Teodoru et al., 2012), generally follow the reflectance pattern of healthy green vegetation (Amani et al., 2018), with high NIR reflectance and low red and blue light reflectance (Amani et al., 2019). However, the unique proportions of water, vegetation and soil present in different types of wetland lead to distinct spectral signatures (Amani et al., 2018).

To help with the identification of different landcover types, false color composites (FCC) and vegetation indices can be used (Richards, 2022). FCC are created by matching Landsat image bands to primary display colors (red, green, and blue) in combinations other than the natural true-color sequence which can help to highlight particular landcover characteristics (Khorram, 2012). Vegetation indices are image derivatives, mathematical transformations of the raw image data that are specifically developed to highlight e.g., vegetation or water features (Silva et al., 2020).

2.2.2. Image classification algorithms

The main aim of image classification is to analytically group the spectral information provided in the remotely sensed imagery into easily interpretable classes (Khorram, 2012). Automated image classification approaches can be broadly categorized into unsupervised and supervised algorithms. Unsupervised classification algorithms group pixels into spectral clusters without prior labeling (Gibson and Power, 2000).

In contrast, supervised image classification algorithms require the collection of spectrally

representative training samples to define the spectral signature of each class before analysis. While many different supervised classification algorithms exist, some of the most commonly used ones are Maximum Likelihood Classification (MLC), Support Vector Machine (SVM), Convolutional Neural Networks and Random Forest (RF) (Richards, 2022).

RF classification is a relatively new addition to the supervised classification algorithms available and has become increasingly popular for landcover classification in recent years (Jhonnerie et al. (2015), Rodriguez-Galiano et al. (2012)). Compared to other classification algorithms, it has a few advantages which work well with the premise of this study. Firstly, RF is a non-parametric classifier which allows the analysis of non-normally distributed and multi-modal data (Amani et al. (2019), Zhe Zhu (2016)). Furthermore, the algorithm provides a variable importance metric as well as an unbiased accuracy assessment, the out-of-bag (OOB) error (described below), which does not require any field validation data (Breiman, 2001). Lastly, RF is not as sensitive to noise or small sample sizes as other classifiers (Rodriguez-Galiano et al., 2012) and requires the setting of fewer parameters (Pal, 2005) all of which make it very amenable to the scope of this study.

In essence, RF is an ensemble machine learning (ML) algorithm based on numerous independent decision trees where the final classification of any pixel depends on the majority vote of these trees. Each decision tree in the forest is "grown" using a different subset of data (a process called bagging) and around 1/3 of the data, the so-called out-of-bag (OOB) data, is never used for any individual tree (Breiman, 2001). After classification, OOB data is run down only those trees that did not include them in their training subset. The OOB error then is the proportion of the OOB data classifications that are incorrect compared to the true labels in the training set (Breiman, 2001). Breiman (2001) proposed OOB error as an independent accuracy assessment. To assess how much the model depends on specific input variables, RF randomly shuffles the OOB data of one input variable at a time, destroying the relationship between that variable and the predicted class. The resulting decrease in model accuracy is then used as an indication of the importance of that variable in determining the classification result (Breiman, 2001). While ensemble complexity, unknown tree split rules and randomness introduced in the bagging process limit the model's transparency (Rodriguez-Galiano et al., 2012), the user does have control over the number of trees, maximum tree

depth, minimum leaf size and the number of random variables sampled per tree (Millard and Richardson, 2015). The number of trees is the total number of independent decision trees which are grown in the RF (Rodriguez-Galiano et al., 2012). Maximum tree depth defines the maximum number of splits per tree while minimum leaf size determines the minimum number of samples that have to be present at a terminal node to reach a classification result (Amani et al., 2019). All these parameters can be tuned to increase accuracy and reduce overfitting (Amani et al., 2019).

2.2.3. Object- vs. pixel-based classification approaches

RF image classification can be performed using different minimum units of analysis. Pixel-based (PB) classification uses an individual pixel as the input unit, considering only the variations of variables in this single pixel to classify it. In contrast, the minimum unit of object-based classification (OB) is a range of contiguous pixels (Amani et al., 2018). To define pixels belonging to an object, OB classification first needs image segmentation to be performed, a process in which pixels are grouped into homogeneous areas based on their spectral and spatial similarity. OB classification has been shown to generally yield higher accuracies (Amani et al., 2018).

2.2.4. Improving the classification

Landcover classification algorithms can be improved by using Principal Component Analysis (PCA) (Richards, 2022). In essence, PCA groups the original data on a number of orthogonal vectors, thus reducing its dimensionality. The first of these so-called principal components holds as much of the variation of the data as possible, while the second component holds as much variation as possible without any correlation to the first and so forth (Khorram, 2012). PCA produces a metric called eigenvalue which indicates the amount of variance captured by each of the components and is an indication of which principal components should be included in analysis (Richards, 2022). Thus, PCA can be used to reduce the amount of bands required for analysis which enhances the efficiency and effectiveness in supervised image classification (Richards, 2022). In RF classification, PCA can help to simplify the generation of decision trees by removing correlated variables and reducing noise in the input

data (Richards, 2022). Furthermore, PCA can be particularly useful when performing image segmentation as a composite with the first three principal components as input is often more distinctly separable than a standard RGB composite (Gibson and Power, 2000).

3. DATA & METHODS

All spatial analysis in this study was performed in ArcGIS Pro (Version 3.4) by Esri, and Python (Version 3.12.11). All maps were created using ArcGIS and OpenAI (2026) was used to assist with coding in Python. Prior to analysis, all datasets were projected into Canada Albers Equal Area Conic NAD83 (EPSG: 102001) to avoid distortion of area.

3.1. Study area

The Paix-des-Braves (PDB) reservoir, formerly Eastmain-1, is located in the northwestern part of boreal Canada, roughly 1000 km north of Montréal, Québec (Demarty et al., 2026). Figure II shows the location of the PDB reservoir in Québec. The region experiences a cool continental climate (Amani et al., 2019) with 430 mm of rain and 260 mm of snow per year and an average monthly temperature from -23° to 14°C (Demarty et al., 2026). PDB is connected to one of the largest hydroelectric networks in the world, the La Grande hydroelectric complex, comprising a total of nine reservoirs constructed in the Eastern James Bay region between 1973 and 1996 (Messier, 1998). From 2014 to 2022, the La Romaine reservoir complex was added at the southern coast of Québec (Rust et al., 2021). All reservoirs are operated by Hydro-Québec which has funded extensive environmental impact studies, making the La Grande complex one of the world’s most extensively studied hydroelectric projects (Tremblay, 2005).

The impoundment of the PDB reservoir started in 2005, and in 2009 and 2011 the Rupert diversion bays were flooded to increase water volume in the PDB reservoir (Demarty and Tremblay, 2017). The PDB reservoir has a size of roughly 603 km^2 (Teodoru et al., 2012) and its mean depth varies from 8.5 m to 11 m, depending on the season and water management (Demarty et al., 2026). Water input to the reservoir comes from the Eastmain river which has a watershed of roughly $46,000\text{ km}^2$ and flows into James Bay (Teodoru et al., 2012). The region is dominated by podzolic and peat soils on igneous bedrock (Demarty and Tremblay,

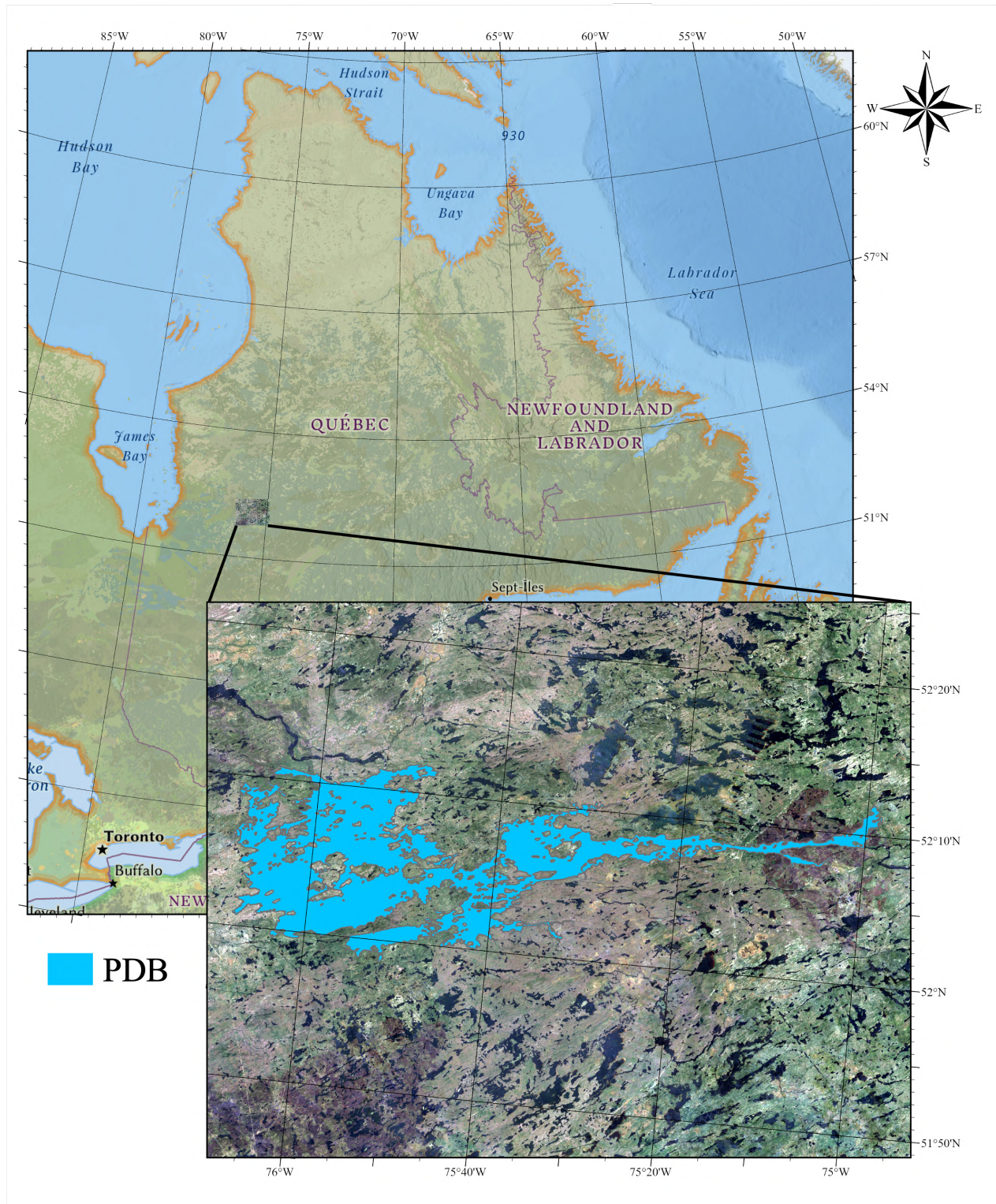


Figure II: Map showing the location of the study area around the PDB reservoir within Québec. Basemap: Esri, Maxar, Earthstar Geographics, and the GIS User Community (World Imagery) (Esri, 2025). Landsat Median Composite (EROS, 2020). Reservoir extent from Lehner et al. (2022).

2017). Mean water residence time is around 2.3 months (Demarty et al., 2026); the surface reservoir is usually frozen from mid-December to mid-May (Demarty and Tremblay, 2017); and the reservoir’s two hydroelectric power stations have an installed capacity of 1275 MW (Hydro-Québec, 2026).

According to Teodoru et al. (2012), the PDB reservoir flooded a heterogeneous landscape, consisting of mature forest (deciduous and coniferous), burned forest, non-forested soil, wetlands, as well as lakes, rivers and streams. To be able to collect enough training samples for landcover mapping, the study area was defined 10 times bigger than the reservoir extent, at a total size of 6300 km^2 . The extent of the study area in relation to the reservoir can be seen in Figure II.

3.2. Data

To train the RF model for land cover classification of the PDB, remotely sensed Landsat 5TM imagery as well as several auxiliary datasets were used. RF is an algorithm which works well with data from multiple sources (Rodriguez-Galiano et al., 2012) and while Landsat 5TM data allows for easy transferability of the methods employed in this study and has a high temporal resolution (making it very suitable for LULCC mapping), local auxiliary datasets were found to provide helpful contextual information for certain landcover classes during the classification process. An overview of the data used in this study is found in Table I. A more detailed description of the individual datasets ensues below.

3.2.1. Landsat 5 TM imagery

Landsat 5 Thematic Mapper (TM) data has successfully been used for historical landcover mapping before (Krivoguz et al., 2023) and its temporal scope encompasses all but the first impoundment of hydroelectric reservoirs in Québec. Furthermore, previous studies have used Landsat imagery as a basis for accuracy assessment without field validation (Millard and Richardson (2015), Amani et al. (2019), Huang et al. (2017)). Thus, Landsat 5TM imagery was chosen as a basis for landcover mapping in this study.

Landsat 5 TM surface reflectance imagery (Level 2) (EROS, 2020), was accessed through Google Earth Engine (GEE). The TM sensor includes six spectral bands (Table II) with a

Table I: Overview of datasets used in this study

Dataset	Dataset period	Study period	Spatial resolution / scale	Source
Landsat 5 TM Surface Reflectance (Level 2)	1984–2012	1 June–30 Sep. 2002	30 m	EROS (2020), accessed via GEE
Canadian National Fire Database (CNFD)	1973–present	1973–2002	Varies	Canadian Forest Service (2022)
National Topographic Database (NTDB)	1944–2005	1955–1990	1:50,000	Natural Resources Canada (2005)
NASA Digital Elevation Model (DEM)	11–22 Feb. 2000	11–22 Feb. 2000	30 m	NASA JPL (2020)
LakeAtlas	~2016	~2016	15 arc-seconds (≈ 500 m)	Lehner et al. (2022)

spatial resolution of 30 m. Based on this spatial resolution, minimum mapping unit (MMU) for this project was set at 9 pixels or 900 m^2 . Each Landsat scene is $170\text{km} \times 183\text{km}$ in size and to cover the extent of the study area, five different Landsat tiles (016/024, 017/023, 017/024, 018/023, 018/024) had to be mosaicked. The dataset has a temporal resolution of 16 days and is available between March 1984 and May 2012. The Level 2 surface reflectance product is already atmospherically and geometrically corrected, but the bands had to be scaled to convert raw digital numbers (DN) to surface reflectance (SR) values, according to Equation 1:

$$SR = DN \times 0.0000275 - 0.2 \quad (1)$$

3.2.2. Canadian National Fire Database

The area around the PDB reservoir had an estimated fire cycle between 132 and 153 years during the last century and fire is thus an important metric to take into consideration when

Table II: Spectral bands and wavelength ranges for Landsat 5 Thematic Mapper sensor

Band	Spectral region	Wavelength range (μm)
Band 1 (Blue)	Visible Blue	0.45 – 0.52
Band 2 (Green)	Visible Green	0.52 – 0.60
Band 3 (Red)	Visible Red	0.63 – 0.69
Band 4 (NIR)	Near-infrared	0.76 – 0.90
Band 5 (SWIR1)	Shortwave-infrared	1.55 – 1.75
Band 7 (SWIR2)	Shortwave-infrared	2.08 – 2.35

calculating long-term carbon balance based on landcover (Dessureault et al., 2015). Historical forest fire data was obtained from the Canadian Wildland Fire Information System (Canadian Forest Service, 2022). Their Canadian National Fire Database (CNFD) contains polygons designating forest fire perimeter. For the province of Québec, the dataset starts in 1973. No accuracy metrics are provided.

3.2.3. *National Topographic Database*

The Canadian National Topographic Database (NTDB) has been previously used for land-cover estimates of the PDB reservoir region (Teodoru et al., 2012). The historic LULC product has a temporal scope from 1944 - 2005 and a scale of 1 : 50,000 (Natural Resources Canada, 2005). Mean planimetric accuracy is 25 m for rural areas though this varies by tile and ranges from 18 to 27 m for the study area which encompasses 12 tiles in the NTDB. While the NTDB has a higher resolution than the Landsat imagery, the NTDB only includes water, vegetation and wetlands as landcover categories for the study area and no accuracy metric per class is provided. In addition, temporal scope for the data varies by tile and ranges from 1955 to 1990 for the study area which makes accurate landcover mapping of the PDB landscape directly prior to impoundment difficult. However, in combination with the Landsat data, the NTDB was deemed very useful for this study as a reference dataset which clearly delineates wetland and water. Particularly for the wetlands, the NTDB was considered to be a helpful aid in learning how to identify them on the Landsat scene. A map derived from the NTDB and CNFD polygon layers is shown in Figure III.

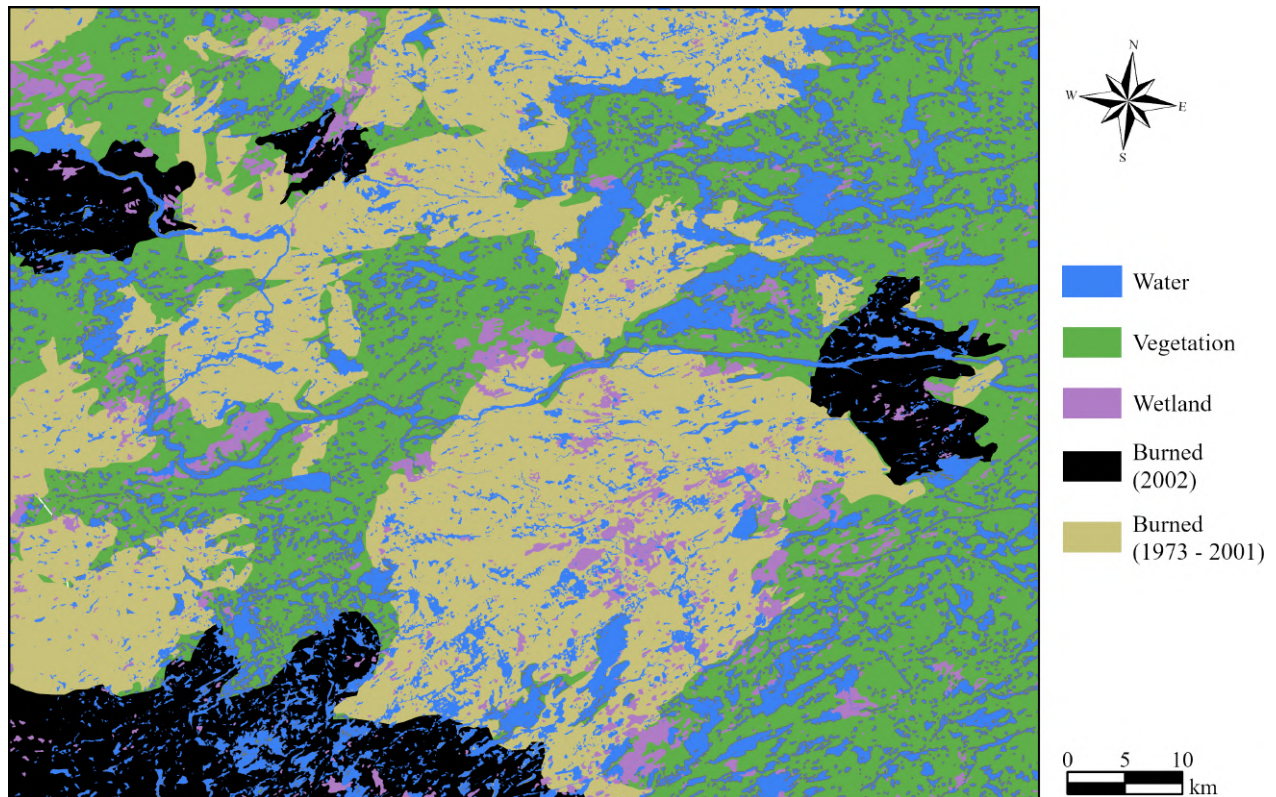


Figure III: Landcover map based on water, vegetation and wetland polygons from the NTDB (Natural Resources Canada, 2005). Burned area polygons from the CNFD (Canadian Forest Service, 2022).

3.2.4. Digital elevation model

Elevation data has been found to be an important variable in landcover classification using RF (Millard and Richardson, 2015). Thus, historic digital elevation model (DEM) data for the study area was obtained from the global NASA 30 m digital elevation model (NASA JPL, 2020). The product is a reanalysis of the shuttle radar topography mission which collected topographic data for 80% of the world from 11 February to 22 February, 2000. The DEM was accessed through GEE and the DEM of the study area can be found in the Appendix A4b.

3.2.5. LakeAtlas

For delineation of the PDB reservoir extent, shapefile data provided through the LakeAtlas database (Lehner et al., 2022) was used. The database has a resolution of 15 arc seconds and for Canada it has been assembled based on information provided through CanVec (Natural Resources Canada, nd).

3.3. Methods

The methods section is structured according to the order of the research questions. First, the mapping process itself will be described as well as the steps undertaken to assess map stability. Then, the accuracy assessment process will be described followed by a description of the carbon balance calculations. A visual overview of the methodology used for the landcover classification is provided in Figure IV.

3.3.1. Map quality and stability

This section will explain methods related to the first research question: "Can a supervised landcover classification approach produce a map of sufficient quality and stability to support reliable carbon balance estimates?"

Pre-processing of the data

The chosen Landsat 5 TM data (EROS, 2020) was pre-processed in Google Earth Engine (GEE). Firstly, to reduce data volume, a median composite image was produced. Median

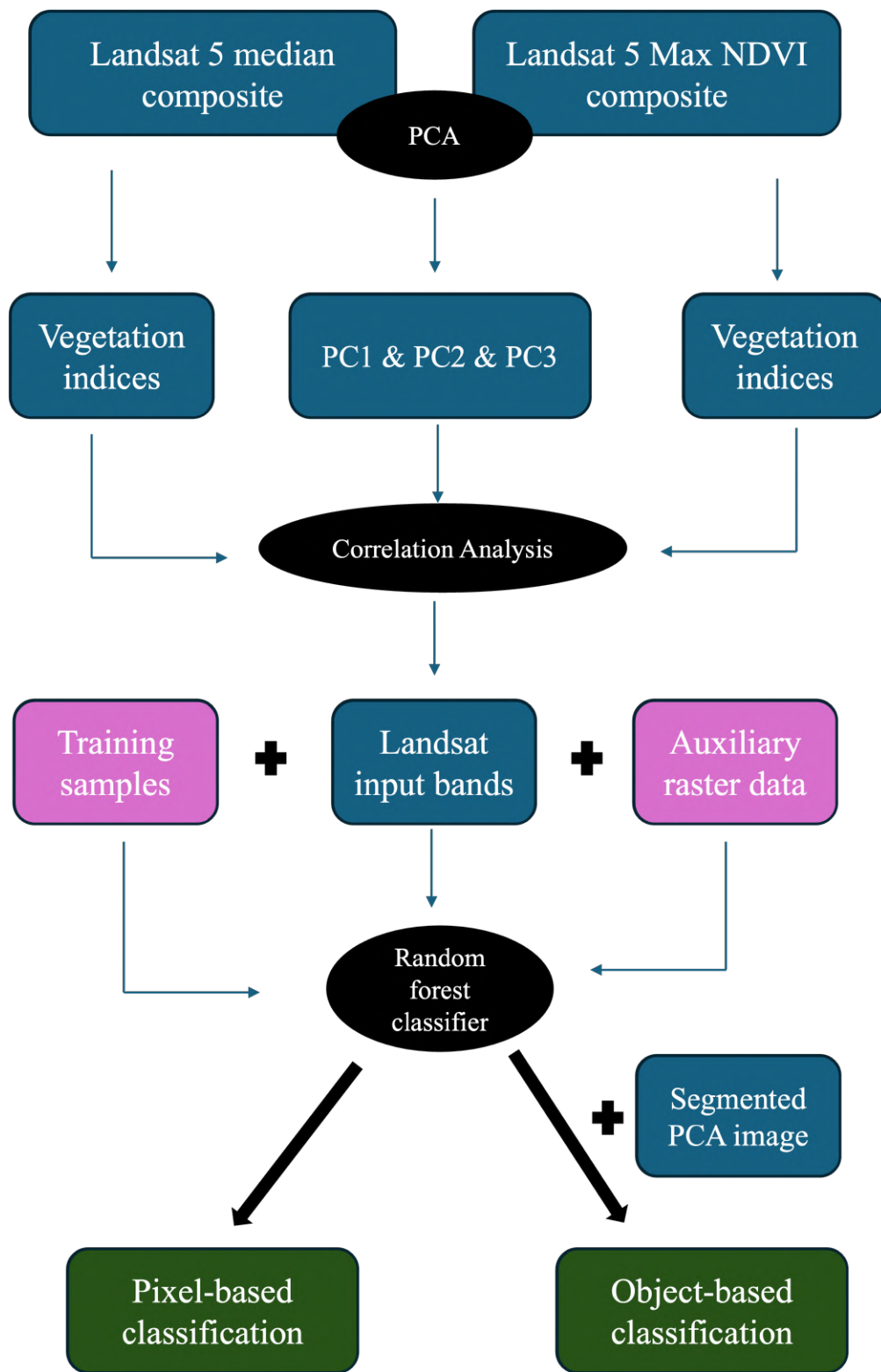


Figure IV: Flowchart broadly visualizing the methods used for pixel- and object-based classification of the study area.

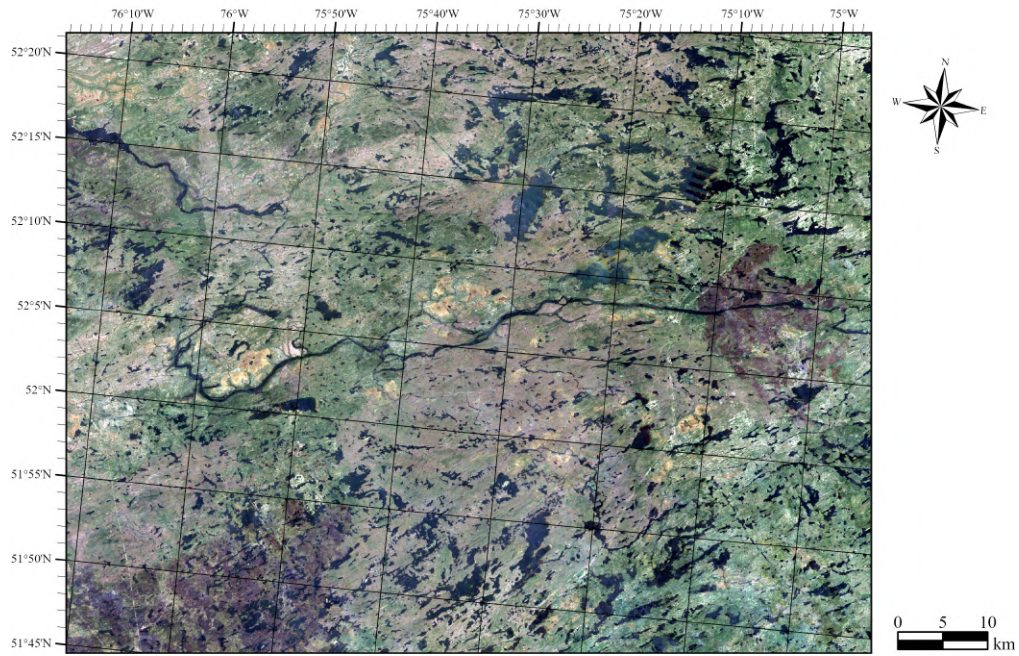
composites produce good results in RF landcover classification (Phan et al., 2020) and the method is widely used in landcover classification studies (Gebreegziabher et al. (2025), Li et al. (2022), Nasiri et al. (2022)). Phan et al. (2020) have shown that median composites produce best RF classification results when only images from one season of one year are used. To avoid snow cover and capture vegetation at full bloom, the summer season (01 June - 30 September) was chosen for analysis. Since the reservoir was flooded in 2005 (Demarty et al., 2026), Landsat scenes during the summer months of 2001, 2002, 2003 and 2004 were assessed for quality and suitability for mapping. Average cloud cover during the summer across all scenes and images was 58% in 2004, 59% in 2003, 40% in 2002 and 53% in 2001. Based on this, images between 1 June 2002 and 30 September 2002 were downloaded for analysis.

In order to get the best composite result, filled pixels, dilated clouds, clouds, cloud shadows and snow were masked out based on the quality assessment band provided through the CFMASK algorithm as part of the Landsat 5 TM Level 2 product. An assessment of the performance of the CFMASK algorithm for Landsat 5 TM imagery can be found in Foga et al. (2017). The images were not pre-filtered for cloud cover since this has been found to reduce accuracy (Phan et al., 2020). Thus, images with high cloud cover were included though cloudy pixels were of course not included in the final composite due to the applied CFMASK mentioned above.

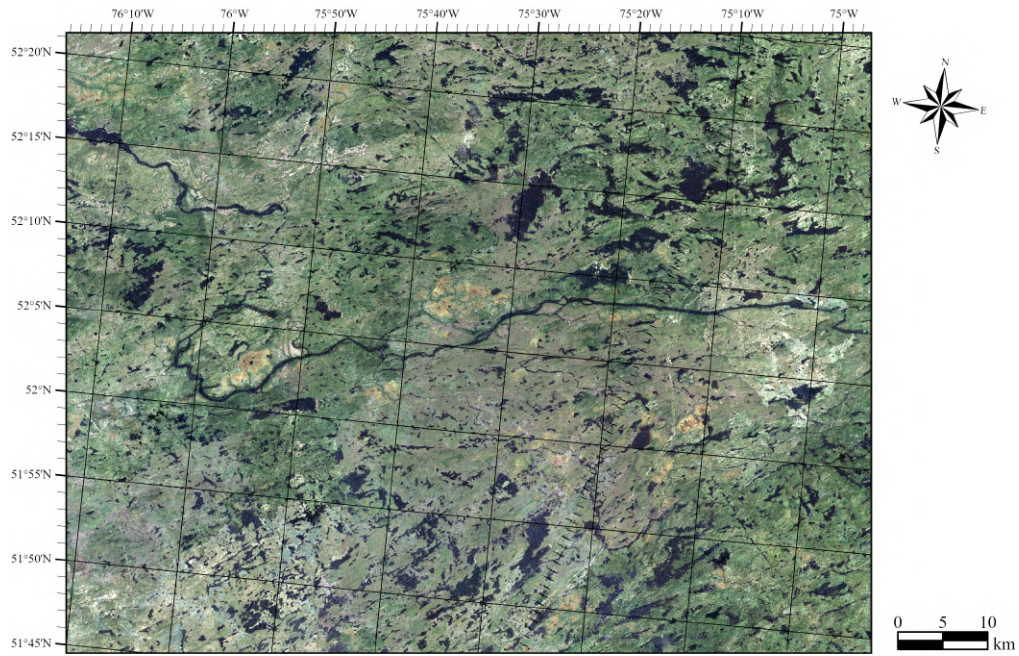
Although median composites have been shown to perform well for RF classification (Phan et al., 2020), in certain contexts, maximum value composites can outperform median composites (Xu et al., 2025). Therefore, two Landsat images were computed for analysis: One median composite (Landsat Median, see Figure Va) and one NDVI maximum composite (Landsat Max, see Figure Vb). NDVI maximum composites select the date with the maximum NDVI to be used for all bands of a given pixel.

Principal component analysis

As mentioned previously, PCA can lower data volume and thus computing time in classification while at the same time reducing noise and correlation between input parameters (Gibson and Power, 2000). Given the overall lower quality of the Landsat Max composite, a correlation matrix was computed between all bands of both images, to assess whether including the Landsat Max composite in PCA would provide additional information and thus offset



(a) Landsat Median composite



(b) Landsat Max composite using maximum NDVI

Figure V: Landsat Median and Landsat Max composite used as a basis for random forest classification. Data source EROS (2020), accessed and processed in Google Earth Engine. Temporal scope of the composites is 1 June - 30 September 2002.

the chance of the lower quality potentially worsening results (Appendix A6). The SWIR1 and SWIR2 bands of the Landsat Median image showed a correlation > 0.9 with their Landsat Max image counterparts. However, all other bands showed correlations < 0.9 and were therefore included in the PCA analysis. Thus, PCA was run with Bands 1-5 and 7 from the Landsat Median image and Bands 1-4 from the Landsat Maximum image.

Landcover classification

After an initial inspection of the imagery, six landcover classes were selected for mapping. The goal was to stay as close to landcover classes used in previous studies to be able to compare results, while at the same time only including classes that could be sufficiently sampled and which seemed distinguishable with the provided data. Table III summarizes the landcover classes used in this study.

Table III: Landcover classes used in this study

Landcover class	Description
Water	Including lakes, rivers, and streams
Deciduous	Forest dense enough to be clearly visible within a 9-pixel area to comply with MMU
Coniferous	
Wetland	Including marshes, swamps, fens, and bogs
Burned	Area burned in 2002
Bare	Non-forested soil as well as areas with very sparse vegetation cover (older burned area)

Delineation of landcover classes

To help with the identification of the different landcover types for training sample collection, several vegetation indices were calculated based on the Landsat composites. Calculations were performed in GEE according to Equations (2)–(5). While other indices were visually assessed in GEE, these four seemed the most useful in delineating the different landcover types in the study area.

The Normalized Difference Vegetation Index (NDVI), the most widely used vegetation index, was calculated according to Equation 2 (Khorram, 2012). It produces a linear scale where values near 1 indicate dense, healthy vegetation, while negative or near-zero values represent water, snow, urban areas, or bare soil (Zaidi et al., 2017). An NDVI-derived map

for the study area can be found in Appendix A3a.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

The Normalized Difference Water Index (NDWI) is sensitive to the water content in vegetation and soil and useful to monitor changes in vegetation biomass and water stress (Silva et al., 2020). Here, it was deemed useful to clearly distinguish recently and older burned areas. It is calculated according to Equation 3 (Gao, 1996).

$$NDWI = \frac{NIR - SWIR1}{NIR + SWIR1} \quad (3)$$

An adaptation of the NDWI, the modified Normalized Difference Water Index (mNDWI), uses the green band instead of the NIR band and has been proven particularly useful to delineate water bodies. It was calculated according to Equation 4 (Xu, 2005). See an example for the study area in Appendix A4a.

$$mNDWI = \frac{Green - SWIR1}{Green + SWIR1} \quad (4)$$

Finally, the second Modified Soil-Adjusted Vegetation Index (MSAVI2) was calculated according to Equation 5 (Qi et al. (1994), see Appendix A3b). The MSAVI2 is particularly useful to distinguish sparse vegetation from bare soil (Qi et al., 1994).

$$MSAVI2 = \frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - Red)}}{2} \quad (5)$$

Apart from the original Landsat images vegetation indices, PCA outputs as well as FCCs were used to recognize different landcover classes during sample collection. PC1 (Appendix A1) clearly showed the difference between land and water areas which was also nicely reflected in the mNDWI (Appendix A4a). PC3 proved to be particularly useful for delineation of recently burned area (Appendix A2b) while PC2 showed a more general division between healthy vegetation and bare/burned ground (Appendix A2a). A color-infrared composite proved particularly useful to distinguish between deciduous (bright pink) and coniferous (purple) which were also quite easy to differentiate when stretching the NIR band of the

Landsat Max composite. The band combination NIR, Red, SWIR1 has been successfully used in wetland mapping before (Kulawardhana et al., 2007) and proved to be very useful for wetland delineation in the study area as well, making them appear bright yellow (see Appendix A5b). Finally, stretching of the SWIR1 bands helped to differentiate between bare ground and forest due to the sensitivity of the SWIR1 band to moisture content (Amani et al., 2018).

Training sample collection

For training sample collection, stratified sampling was used to ensure that training points would cover the entire extent of the Landsat image and that each class would be sufficiently represented in the sample collection. Given that the image is mosaicked from different scenes, this was deemed important to capture the spectral signature of each landcover class over the entire extent of the image (Gibson and Power, 2000). As a basis for the stratified sampling, auxiliary data was used to define polygons in which the presence of certain landcover classes was to be expected. This approach was recommended by McCoy (2005). Wetland and water polygons were based on the NTDB, while burned area polygons were based on CNFD data from 2002. Compared to areas that had never burned, areas that had burned since 1973 seemed to visually show much lower vegetation density. To sufficiently sample the bare class, burned area prior to 2002 served as the basis of another polygon. Now, 200 random points were created within each of the five polygon classes (wetland, water, burned in 2002, burned prior to 2002, rest) at a minimum distance of 500 m.

Training polygons of 9 pixels, surrounding the created sample point, were classified until the minimum number of pixels per class had been collected. Zhe Zhu (2016) tested the sensitivity of the RF algorithm to sample size and found 20,000 pixels to be the optimal number for one Landsat scene. However, the extent of the study area is 6300 km^2 , or about 20% of one Landsat scene. For two Landsat scenes, Zhe Zhu (2016) found that the optimal training sample size increased linearly. Assuming a similar linear trend when downsizing the study area, 4000 pixels should be sufficient to classify the area using RF. Following the approach of 9 pixels per sample, this translates into roughly 450 samples. This approach aligns with the recommendations by McCoy (2005) to collect a maximum of 10 pixels per sample and to collect 10 times as many samples as input bands per class (here: $6 \text{ bands} \times 10 =$

60 samples per class). Gibson and Power (2000) recommend that training samples should be as representative of a class as possible and sample points which fell into very heterogeneous areas were therefore skipped to reduce data noise. To assess the spectral separability of the classes, a spectral signature plot was computed. Based on a preliminary spectral signature plot (not shown), the water classes (river and lake) were merged since they did not show enough spectral separability during an initial test computation of the spectral signature plot. Furthermore, during sample collection it became apparent that a differentiation between different wetland classes would not be possible due to insufficient knowledge and the small overall number of certain wetland classes to be expected based on estimates from (Teodoru et al., 2012).

Samples were collected until each class had a minimum of 60 samples, since this would leave enough data to be excluded for accuracy assessment. Despite stratified sampling, certain classes are more prevalent in the scene than others and thus more likely to fall on a sample point. Therefore, the total number of pixels (5500) collected was somewhat unevenly distributed between the classes (Table IV). This, however, is a desirable result as Millard and Richardson (2015) have shown that a representative as opposed to an equal number of samples per class leads to better RF classification results.

Table IV: Number of pixels as well as approximate number of samples within each training sample class. Sample number is only approximate since around nine pixels form one training sample polygon but this is not entirely consistent and deviates from 7 to 10, leading to slightly different values.

Class	Number of Pixels	Number of Samples
Water	1303	144
Deciduous	811	90
Coniferous	1224	136
Wetland	569	63
Burned	671	74
Bare	922	102

Random forest model inputs

Millard and Richardson (2015) recommend removing highly correlated variables (≥ 0.9) from the input layers of RF as the model tends to treat correlated variables with preference

which leads to biased variable importance metrics while simultaneously adding noise and computational complexity to the model without improving accuracy. Therefore, in order to assess which additional layers would be useful to add to the model, a correlation matrix was computed between the computed vegetation indices from both Landsat images and PCA layer one, two and three respectively (see: Appendix A6). Indices which had a correlation of ≥ 0.9 with either PCA layer or one of the other indices from either image were removed. Thus, apart from the three PCA layers, the NDVI and mNDWI from Landsat Max, as well as the MSAVI2 from Landsat Median and the NDWI from both images were included in the modeling. In addition, the DEM layer as well as the NTDB layer (with the three classes wetland, vegetation, and water) were included. Since data accuracy of the CNFD varies by province and based on the mapping techniques used, this dataset was used merely to better identify burned areas during training sample collection and was not included in the RF model itself.

Random forest classification

Random forest classification was performed using the ArcGIS Pro adaptation of the RF algorithm developed by Leo Breiman (Breiman, 2001). Both object-based (OB) and pixel-based (PB) classification were performed in order to test their performance and benefit from their slightly different in- and output options. In addition, a majority (modal) filter was applied to the final classification outputs to reduce pixel-level noise and improve spatial coherence of the resulting maps.

The RF parameters "number of trees" and "maximum tree depth" were optimized for overall accuracy using grid search optimization and were kept the same for both OB and PB classification. Both models were run once using 100% of the training data and a summary of all model inputs and parameter settings as well as variable importance metrics can be found in Appendix A1. For OB classification, the PCA image, using PC1, PC2, and PC3 as RGB respectively, was segmented. The minimum segment size was set to 9 in order to match the minimum sample size and the minimum mapping unit (MMU) achievable with Landsat imagery at 30 m resolution. OB classification was run with the input layers mentioned above, as well as with all training samples and the segmented image as an additional input. In contrast to OB classification, PB classification in ArcGIS Pro provides OOB metrics.

According to Breiman (2001) the OOB error is an independent accuracy assessment in its own right. However, Millard and Richardson (2015) compared OOB and independent accuracy assessment and found that OOB tended to overestimate the model's accuracy. This was particularly true when training samples were spatially autocorrelated as is the case with the individual pixels that make up the sample training polygons in this study. Since this study aims to explore the possibilities of accuracy assessment without ground truth data, training samples for the PB classification were reduced to one central pixel. This avoids spatial autocorrelation and allows for a more accurate OOB.

Model stability

To assess whether the model is sufficiently stable to be used for carbon balance estimates, a confidence raster was computed for the 20 runs of OB and PB respectively. This was done by dividing the number of times the most frequently predicted class in each cell was predicted by the total number of iterations. For example, a pixel that was classified as "Coniferous" 15 out of 20 times is the most frequently classified class for that pixel and would be assigned a value of $15/20$, or 0.75 in the confidence raster. This approach for model stability validation has previously been used by Millard and Richardson (2015). To assess model stability in more practical terms in regard to carbon balance estimates, total area predictions for each landcover class were computed for all 20 model runs. Then, the resulting variations in landcover area across model runs were visualized in a boxplot.

3.3.2. Accuracy assessment

This section will explain methods related to the answer of the second research question: "How can map quality be assessed given the limitations in reference data and the lack of field validation?" It is important to note that the nature of the research question refers to the possibilities of an accuracy assessment without any field validation. Thus, all references made in the following text to "validation samples" apply to samples that have not been collected using ground validation but were instead collected based on Landsat imagery and the auxiliary data mentioned above.

Both OB and PB classification were run 20 times, each time with a different random subset of 80% of the training data (once again using training sample polygons for OB classi-

fication and single pixels for PB classification). The remaining 20% of the training samples were used for point-validation and the computation of an error matrix for both models. The training and validation data used here were created class-stratified, meaning that 80/20 splits were randomly made for each class individually in order to keep their ratio within the total dataset constant. Note that the OB classification was run with the training sample polygons as input while the PB classification was run with the single pixel samples, as per the clarification above. This allowed for the computation of an average OOB error per class across the 20 PB classification runs.

3.3.3. Carbon balance of the pre-flooded landscape

This final methods section will describe the steps undertaken to answer the third research question: "To what extent does a new landcover classification alter previous carbon balance estimates of the pre-flooded PDB reservoir landscape?"

To calculate the carbon balance of the pre-flooded landscape, the two model outputs were both clipped to the reservoir extent, using the reservoir shape provided by the LakeAtlas database (Lehner et al., 2022). For carbon balance calculations of the pre-flooded landscape, diffusive flux measurements per landcover class were taken from Teodoru et al. (2012). This was done in order to be able to compare the carbon balance derived in this study with estimates made by Teodoru et al. (2012) and to thus be able to assess the impact of differing landcover classification on carbon balance estimates. An overview of their carbon balance measurements for the different landcover classes can be found in Appendix A7. Since Teodoru et al. (2012) does not provide any flux measurements for bare land, this class was merged with the burned class for carbon balance calculations. To arrive at a total carbon balance for the reservoir, each landcover area was multiplied by its mean diffusive flux (based on A7), and then all values were summed up. Finally, average flux per m^2 was calculated in order to be able to compare values to those used by Teodoru et al. (2012) since they assume a larger total reservoir extent.

To assess how sensitive the total carbon balance of the reservoir is to different landcover classifications, the carbon balance calculations were repeated for each of the 40 model runs (combining both PB and OB approaches). Thus, an average carbon balance as well as

minimum and maximum values could be computed, reflecting the sensitivity of the total carbon balance to changes in landcover classification.

4. RESULTS

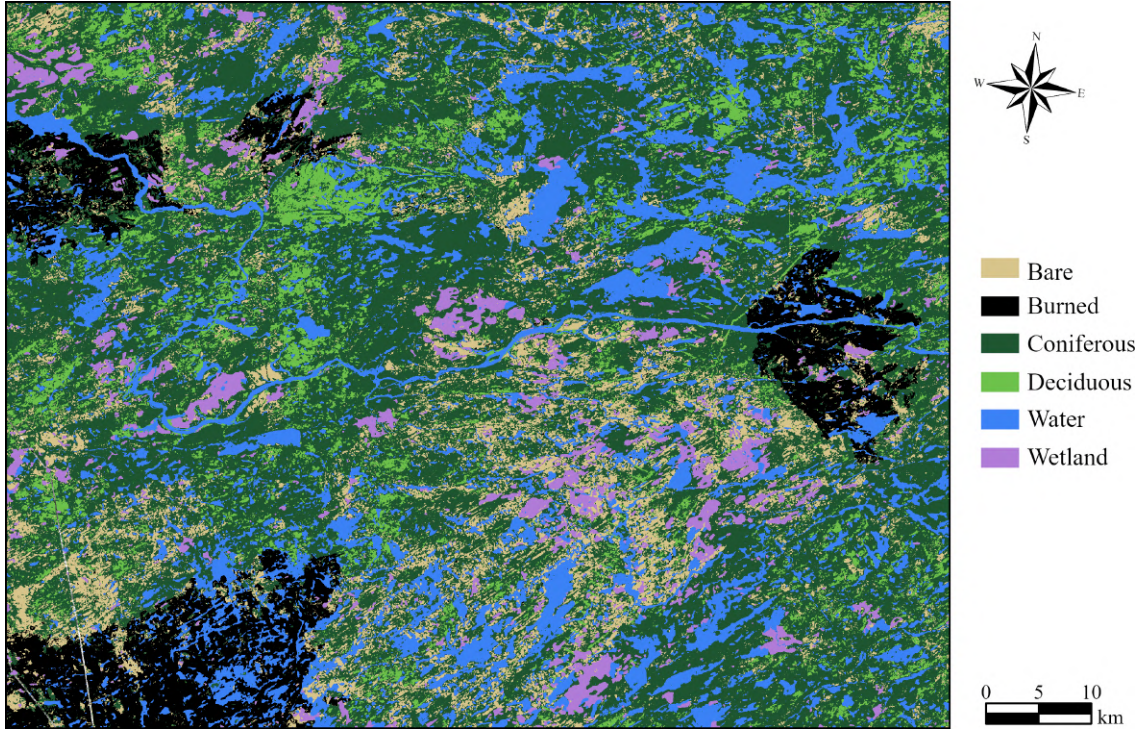
The following will present the results in the order in which they address the three research questions of this study. Section 4.1 presents results related to the first research question by first exploring map quality and then presenting model stability metrics. Subsequently, section 4.2 explores the results relating to the second research question and presents the accuracy metrics derived from the two different accuracy assessment approaches. Finally, section 4.3 addresses the third research question by presenting new carbon balance estimates for the pre-flooded PDB landscape and a sensitivity analysis of carbon balance values to landcover changes.

4.1. Map quality and stability

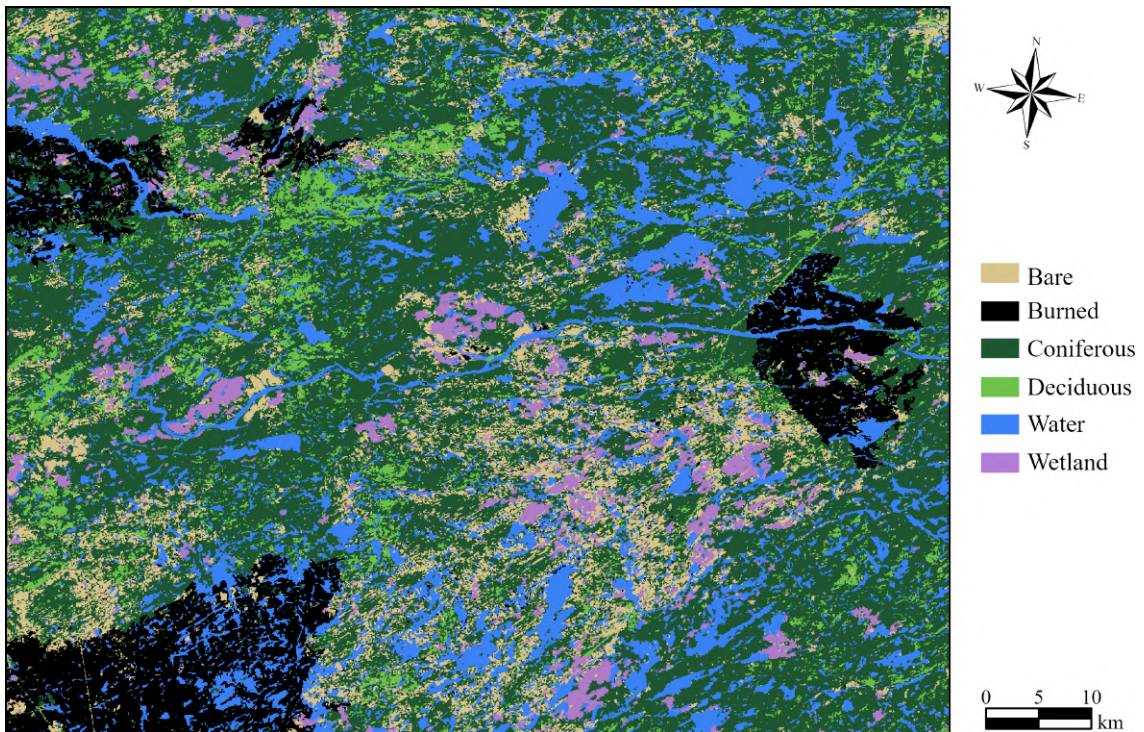
Both models predict a heterogeneous landscape, dominated by coniferous forest and lakes, as well as the Eastmain river flowing through the center of the scene. The final maps from both PB and OB classification can be found in Figure VI. Burned areas are clearly delineated in the southwest, northwest and east of the scene. Deciduous forest is more common in the west while bare ground is more common in the east. Wetlands are present throughout the entire area though they seem to be much rarer in the northeastern corner of the scene. While minor differences between the two maps, such as a higher prediction of burned area in the center of the map by the object-based model, are visible, they generally look very similar.

Spectral Signatures of Training Samples

The spectral signatures of the training samples are quite similar in the visible spectrum but increasingly distinct in the NIR and SWIR bands. The mean spectral signature of the six different landcover classes in relation to the Landsat Median Composite is shown in Figure VII. Water and burned forest have much lower reflectance in the NIR band compared to all other classes. Bare ground has the highest SWIR1 reflectance while having a lower reflectance in the NIR band compared to wetlands and the two forest classes. Deciduous



(a) Pixel-based classification result



(b) Object-based classification result

Figure VI: Results from the random forest models, using 100% of the training data for both object- and pixel-based classification. Input data sources: EROS (2020), Natural Resources Canada (2005), NASA JPL (2020)

trees and wetland samples show a lot of spectral similarities, though deciduous trees have slightly higher SWIR1 and NIR reflectance. Coniferous trees are seemingly distinct from deciduous trees with much lower NIR reflectance. Apart from the differences in the mean reflectance curves, the standard deviation around each mean curve shows that there is clear overlap between the classes as well. This is especially apparent for the deciduous and wetland classes, where one standard deviation of the deciduous class completely encompasses the mean NIR surface reflectance for the wetland class.

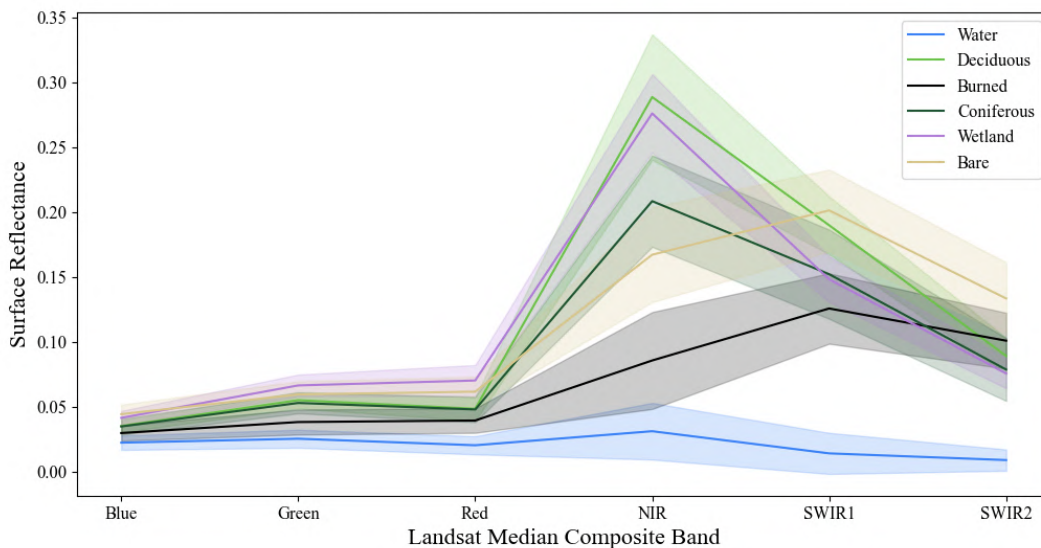


Figure VII: Mean spectral signatures of polygon training samples for the Landsat Median Composite. Bands around the mean lines indicate one standard deviation.

4.1.1. Map stability

Both models show overall good stability, with the PB approach showing slightly higher overall agreement across model runs (Figure VIII) while the OB approach shows smaller area ranges per class (Figure IX). The PB approach also has a higher minimum agreement value (31%) compared to the OB approach (24%, see Figure VIII). Both approaches show the lowest agreement in the northwestern part of the scene around the edges of the Eastmain river. Apart from that, no clear patterns are discernible. Considering the ranges of area predictions across all 20 model runs shown in Figure IX it becomes apparent that the coniferous class has

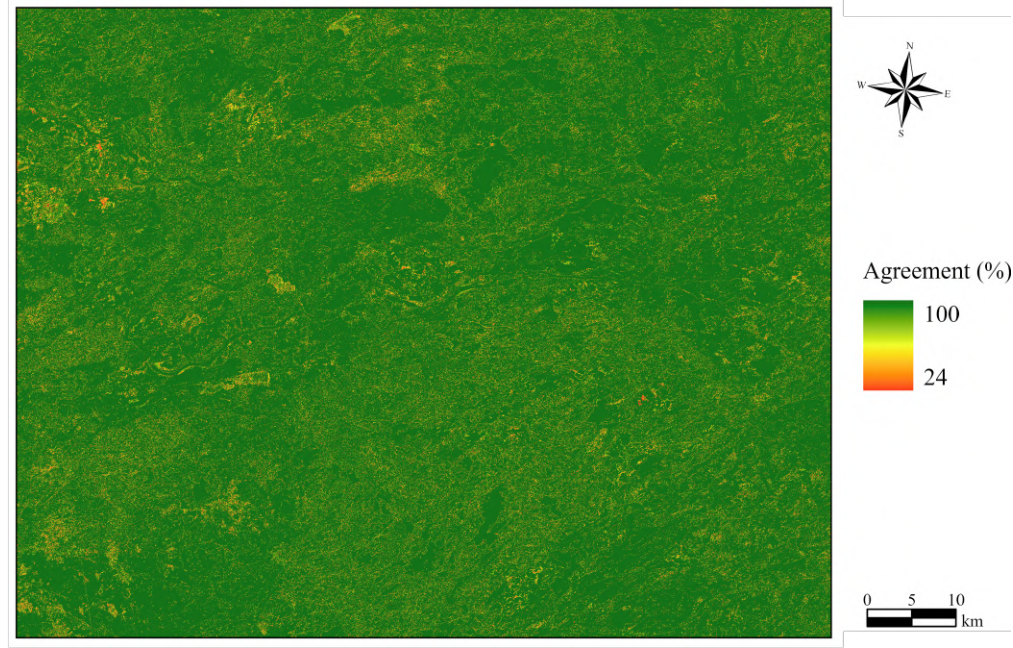
the largest range for both classification approaches, followed by the bare and the deciduous class. These three classes, as well as the burned class, all have smaller ranges in the OB approach compared to the PB approach. Water and wetland have very small ranges and little variation in either approach.

4.2. Accuracy assessment

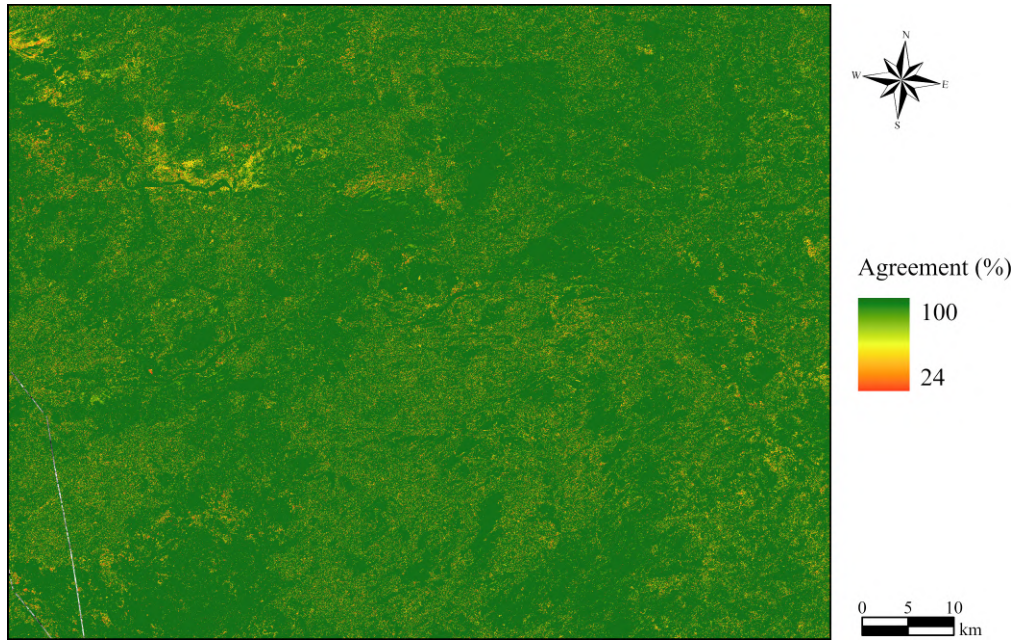
Accuracies for both models are high, with overall accuracy for the OB classification at 90% with a Kappa value of 0.87, and overall accuracy for the PB classification at 87.5% with a Kappa value of 0.85. OOB error metrics (only PB classification) show the smallest error for the wetland class, followed by water and burned. OOB error is highest for the coniferous class at 7.32% (see Table VI). From the confusion matrices (see Table Va for OB results and Table Vb) for PB results) it is clear that water has the highest user’s accuracy (UA) and producer’s accuracy (PA) for both modeling approaches, while the coniferous class has the lowest. Wetland, burned, and bare accuracies are higher for the OB classification (both UA and PA), while the deciduous class has an equal UA for both and has a slightly higher PA in PB classification.

4.3. Carbon balance of the pre-flooded landscape

Both models used in this study predict the pre-flooded PDB landscape to have been acting as a sink of CO₂ and a small source of CH₄ (see Table VIII). The PB approach predicts a larger CO₂ sink than the OB approach. Since these values were calculated proportionately to area per landcover class within the PDB reservoir extent, Figure X shows an example of classified reservoir extent using the PB approach and Table VII summarizes landcover percentages derived from both OB and PB classification within the PDB reservoir extent. PB and OB classification approaches predict relatively similar landcover proportions within the PDB reservoir. The OB approach predicts slightly more water, burned and bare area while the PB approach predicts a slightly larger proportion of deciduous trees. Both approaches predict coniferous to be the most prevalent class, occupying 48% of the area.



(a) Confidence raster across 20 object-based random forest model outputs



(b) Confidence raster across 20 pixel-based random forest model outputs

Figure VIII: Confidence rasters for both modeling approaches, showing agreement for land-cover classification across 20 model runs, using 80% of the training data in each run. Note that the minimum agreement for the pixel-based approach was 31%; scaling was adjusted for better comparability.

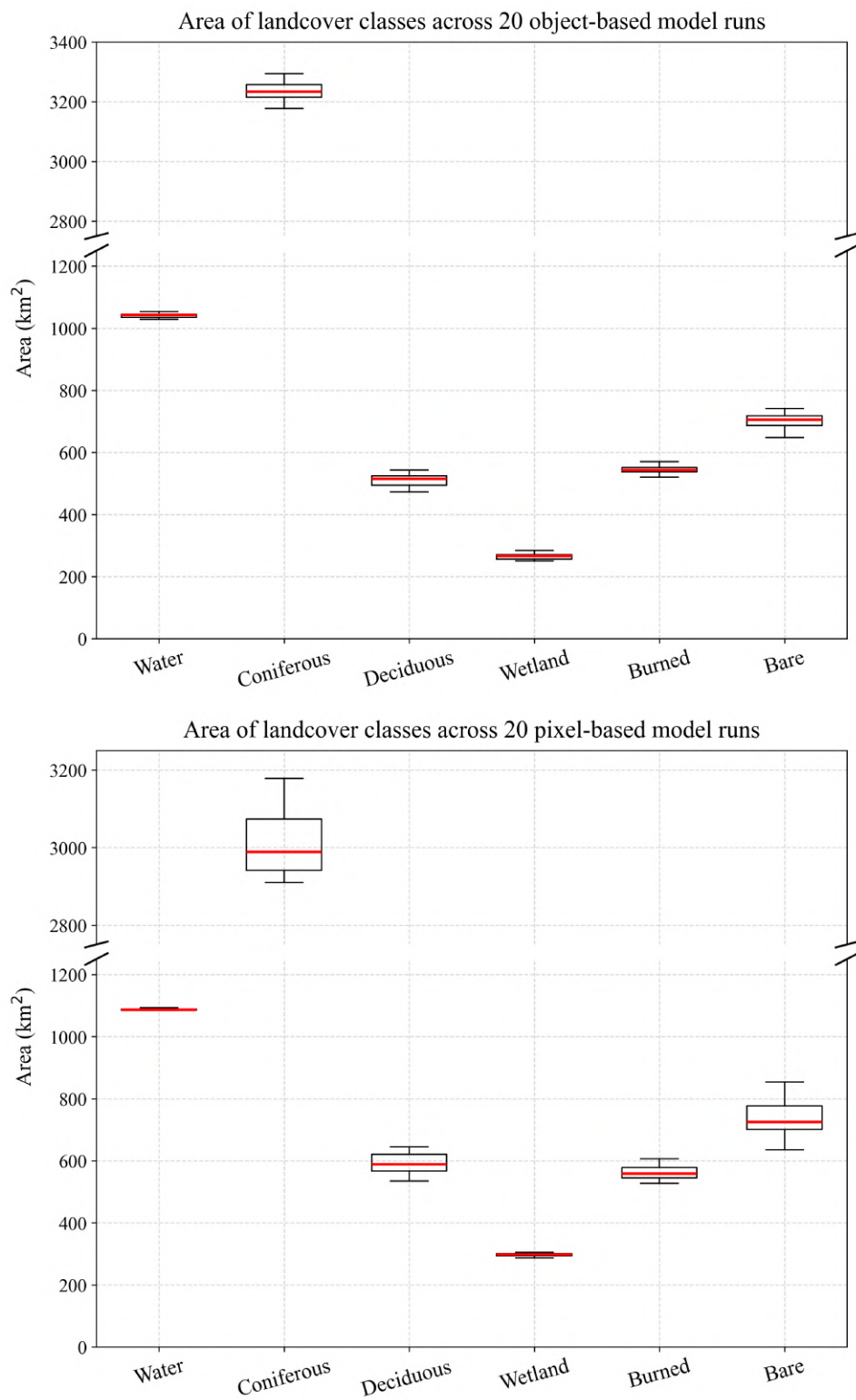


Figure IX: Boxplots showing the area range (black whiskers), median (green line), and one standard deviation (blue box) for each landcover class for both modeling approaches.

(a) Object-based classification

	Water	Deciduous	Coniferous	Wetland	Burned	Bare	PA
Water	623	0	7	0	0	0	98.9%
Deciduous	0	385	56	9	3	9	83.3%
Coniferous	2	63	524	2	14	25	83.2%
Wetland	0	1	6	287	0	0	97.6%
Burned	0	0	0	0	312	24	92.9%
Bare	0	4	39	3	16	421	87.2%
UA	99.7%	85.0%	82.9%	95.3%	90.4%	87.9%	

(b) Pixel-based classification

	Water	Deciduous	Coniferous	Wetland	Burned	Bare	PA
Water	584	0	8	2	0	0	99.3%
Deciduous	0	346	67	13	3	5	85.0%
Coniferous	3	57	466	2	11	54	78.5%
Wetland	0	1	5	267	0	0	93.4%
Burned	1	1	0	0	279	32	90.3%
Bare	0	2	48	2	16	385	80.9%
UA	99.3%	85.0%	78.5%	93.4%	90.3%	80.9%	

Table V: Confusion matrices for the object-based and pixel-based land-cover classifications. PA = Producer’s Accuracy and UA = User’s Accuracy.

Table VI: Random forest mean OOB error by landcover class across 20 pixel-based model runs

Class	OOB (%)
Water	0.86
Deciduous	1.98
Coniferous	7.32
Wetland	0.07
Burned	0.59
Bare	1.93

Table VII: Comparison of classification results using object-based classification and pixel-based classifications. All numbers are percentages of total PDB area.

Class	Percent of PDB Area	
	OB	PB
Water	23	22
Deciduous	6	8
Coniferous	48	48
Wetland	11	11
Burned	3	2
Bare	11	9

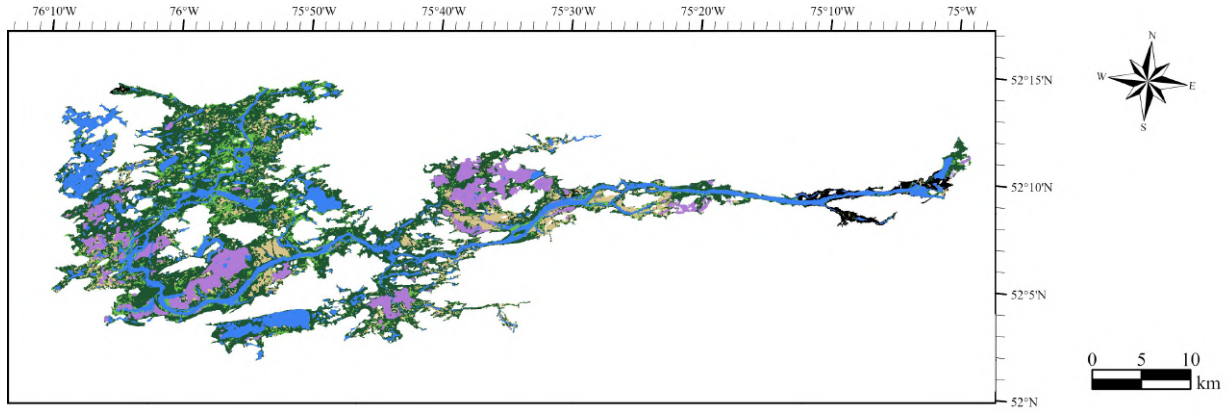


Figure X: Pixel-based classification result clipped to the PDB reservoir extent. Used as basis for Table VII. Data sources for modeling: EROS (2020), NASA JPL (2020), Natural Resources Canada (2005). Reservoir extent: Lehner et al. (2022)

Table VIII: Estimated CO_2 and CH_4 fluxes for the pre-flooded PDB landscape for both the object-based and pixel-based RF classification approach.

Classification	Diffusive Flux ($\text{mg C m}^{-2} \text{ d}^{-1}$)	
	CO_2	CH_4
Object-based RF	-37.3	4.46
Pixel-based RF	-53.9	4.54

4.3.1. Sensitivity of carbon balance to landcover classification

Across all 40 model runs, total carbon balance of the PDB landscape pre-impoundment ranges between -49 to -62 $\text{mg C m}^{-2} \text{ d}^{-1}$ for CO_2 and between 3.7 to 4.2 $\text{mg C m}^{-2} \text{ d}^{-1}$ for CH_4 (see Figure XI). The median of the CO_2 flux is -56.7 $\text{mg C m}^{-2} \text{ d}^{-1}$ while the median for the CH_4 flux is at about 3.95 $\text{mg C m}^{-2} \text{ d}^{-1}$.

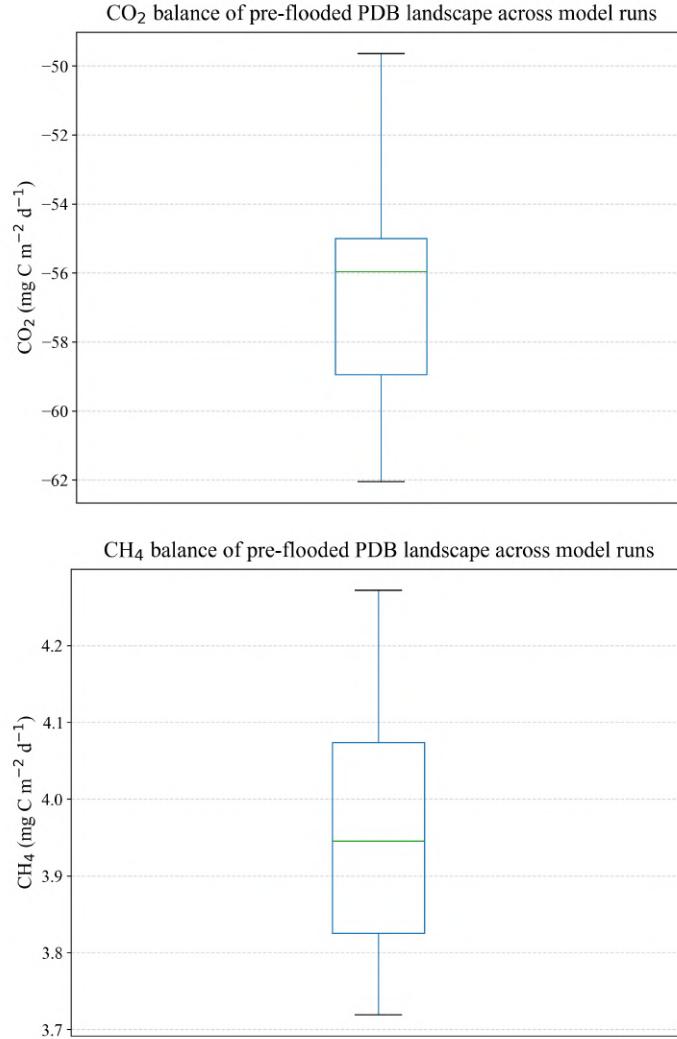


Figure XI: Boxplots showing total CO_2 and CH_4 balance of the PDB pre-flooded landscape per day. The boxplot visualises the range of values across all 40 model runs, as well as median value and one standard deviation.

5. DISCUSSION

The following will discuss the results of this study in the context of the three research questions. First, the question whether or not a supervised landcover classification approach can produce a map of sufficient quality and stability to support reliable carbon balance estimates will be addressed in section 5.1. Subsequently, section 5.2 discusses how map quality was assessed given the limitations in reference data and the lack of field validation. Then, the question to what extent the new landcover classification altered previous carbon balance estimates of the pre-flooded PDB reservoir landscape will be explored in section 5.3. Finally, limitations of this study (section 5.4) as well as future research possibilities (section 5.5) will be brought up.

5.1. Map quality and stability

In answer to the first research question, both classification approaches produced maps of sufficient quality and stability for them to be used for carbon balance estimates of the pre-impoundment PDB reservoir landscape. They visually align with a partial map of the study area produced by Brothers et al. (2012) using similar landcover categories. Furthermore, the calculated carbon balance (Table VIII) is in the order of magnitude of previous studies (Teodoru et al., 2012). However, the very high accuracy values across all landcover classes raise concerns about potential overestimation of classification performance since not all classes would be expected to yield such high accuracies (Amani et al. (2019), Millard and Richardson (2015)). Nonetheless, given the points mentioned above speaking in favor of map quality, as well as the aspects explored in section 5.2, which point towards a faulty accuracy assessment rather than a low-quality map, the map is still considered to be of sufficient quality for carbon balance estimates. The following will discuss map stability as well as consider possible improvements to the mapping methodology used in this study.

5.1.1. Map stability

In general, map stability seems to be relatively high for both modeling approaches. Figure IX shows narrow ranges for most landcover categories, and overall agreement according to Figure

VIII seems high for both maps. However, map stability should ideally be lowest at class-boundaries, suggesting an overall good model stability with some difficulties in heterogeneous areas (Millard and Richardson, 2015). The confidence rasters in Figure VIII do not clearly show this pattern. While the area with the largest disagreement for both models is located at the banks of the Eastmain river in the northwest of the scene (possibly due to an erroneous DEM at this location, see Appendix A4b), agreement across the rest of the study area does not seem to follow any clear pattern. There is lower agreement across the central part of the lower half of both maps and this area is roughly in concordance with a burned (1973 - 2001) area polygon derived from the CNFD (Figure III). Comparing this now to the classified maps (Figure VI), it becomes apparent that big portions of this area were classified as bare. The bare class, together with the coniferous and deciduous class has the lowest PA and UA and also a wider range across model runs (Figure IX). This suggests stability issues regarding the differentiation between the coniferous, deciduous and bare classes which will be further discussed in the later section on accuracy assessment. Coming back to Figure IX, it is noteworthy that water and wetland have the smallest area ranges. This is not surprising given that the water and wetland classes were supported by the NTDB dataset during classification. Burned area also has a lower range than the forest and bare classes and while it does not figure in the NTDB, its delineation is clearly visible in the third principal component (Figure A2b) which most likely improved the stability of classifications in this class. It has to be noted that both OB and PB classification was only run 20 times due to time constraints. While Millard and Richardson (2015) test stability on 25 model runs, additional model runs would certainly improve the confidence in the stability results. Interestingly, the OB classification has the pixels with the lowest agreement (24%) and visually seems to have lower agreement overall. This is in contrast to findings from the literature which suggest that OB classification is generally more stable than PB classification (Amani et al., 2018). This deviation might be due to the slightly different input parameters used for both approaches (Table A1). However, given that RF classification is not transparent in its decision making process (Rodriguez-Galiano et al., 2012), this theory cannot be verified.

5.1.2. Possible improvements

An improvement of map quality would likely have been possible with certain modifications of the methodology. Firstly, both Lück and van Niekerk (2016) and Flood (2015) have developed new compositing techniques which outperform classic maximum NDVI composites. While maximum NDVI composites can help to reduce the impact of clouds, haze and snow on lowering NDVI values (Chen et al., 2003), they tend to still produce more pixelated results overall, despite pre-filtering the images (Ruefenacht, 2016). This phenomenon proved to be true for this study and while the Landsat Max image helped to identify sparse vegetation, a less pixelated NDVI composite, using the methods developed by Lück and van Niekerk (2016) and Flood (2015), might have led to better classification results. This holds particularly true for the differentiation of the coniferous and deciduous classes whose identification heavily relied on the Max NDVI composite and derived indices during training sample collection. This becomes particularly relevant when one considers that the spectral signatures of the coniferous and deciduous classes (Figure VII) are not clearly distinct from one another or other classes and have quite big ranges. According to the confusion matrix, deciduous and coniferous areas are often wrongly classified as the other and a better NDVI composite therefore might have improved classification results. Secondly, a more detailed differentiation of different wetland types would have been desirable for more accurate CH₄ estimates. Due to the limited amount of training samples which would have been available in each wetland class, this detailed wetland mapping was deemed not possible in this study area. However, previous studies have proposed techniques to be able to map different types of wetlands accurately in boreal regions (Amani et al. (2018), Amani et al. (2019), Bubier et al. (1997)). Thus, a bigger study area for training sample collection might have made the differentiation between different wetland classes possible, which would have improved the quality of the produced maps for carbon balance estimates. Lastly, the choice of landcover classes in general is a point of concern. As mentioned previously, the landcover classes used in this study were based on the previous class delineations presented by Teodoru et al. (2012). This was justified through the fact that this approach would make results comparable to Teodoru et al. (2012) and allow the application of their flux measurement values for carbon balance estimates. Furthermore,

apart from the NTDB which only contains wetland, water and vegetation as classes, this was the only available resource for possible landcover classes in the region. Nonetheless a more broad approach to possible landcover classes would have certainly led to more interesting and new results. Similarly to the splitting of certain classes into smaller classes, this was deemed not possible due to a lack of previous knowledge about the region and limited knowledge in the identification of landcover types.

5.2. Accuracy assessment

While this study showed that map accuracy can in theory be assessed given limited reference data and no field validation, the reliability of the accuracy assessment results is highly questionable. This discussion reveals substantial methodological shortcomings for both OOB error and the confusion matrices, and reliable accuracy assessment approaches as an answer to the second research question can thus not be suggested. This section will discuss the OOB error and the confusion matrix in turn, before turning to some more general considerations with regard to accuracy assessment without field validation.

OOB error is below 2% for all classes except for the coniferous class and thus seems to assign a very high accuracy to the entire map. Breiman (2001) states that the OOB error is an independent accuracy assessment. However, as mentioned previously, Millard and Richardson (2015) have found that OOB tends to overestimate accuracies. This was particularly true when samples were autocorrelated or when model input data was highly-dimensional. Both of these points were tried to be addressed in this study by decreasing training sample size to one pixel for PB classification and by removing correlated input variables from the classification. Nonetheless, OOB error might still be inflated due to the fact that, in this study, sample size per landcover class was not exactly proportionate to its actual prevalence in the scene. Millard and Richardson (2015) have found that when a class is overly represented in the training sample set compared to its actual proportion in the landscape, OOB error tends to go towards 0. This is very likely what happened with the OOB error of the wetland class which, at 0.07%, is incredibly low for a class which tends to be difficult to predict (Zhe Zhu, 2016). 10% of all training pixels fell within the wetland class and this is certainly visually disproportionate to the amount of area wetlands

actually cover in the study area (Figure VI). While OOB values for the other classes seem more reasonable compared to previous studies (Millard and Richardson, 2015), it should be noted that OOB error is a model performance metric as opposed to a map accuracy metric and does not establish a true relationship between the map and an independently produced reference sample (Tyukavina et al., 2025). Therefore, while the OOB certainly provides some insights into how well the model reproduces training data, it does not replace an independent accuracy assessment.

The second approach used in this study, using a subset of the training data for validation, certainly gets closer to being an independent accuracy assessment than the OOB error does. Both modeling approaches used in this study yielded very high producer's and user's accuracies overall. While this result is of course desirable, there are some major points of concern regarding the reliability of these values. Millard and Richardson (2015) used the same method as this study to "independently" assess their map's accuracy. While their study regards this accuracy assessment as a reliable good metric, here PA and OA are likely inflated due to the methodology employed in collecting the training samples. Firstly, as mentioned in the methods section, random sample points which fell on spectrally heterogeneous areas were skipped in order to reduce noise in the training sample dataset. While the reduction of noise is arguably a desirable outcome for a training dataset (Rodriguez-Galiano et al., 2012) it also implies that the validation points, derived from the training dataset, all lie within spectrally relatively homogeneous areas. This probably led to inflated accuracies since more heterogeneous areas were not included in the accuracy assessment. Including heterogeneous areas in the validation samples would thus have been a better practice (Congalton and Green, 2008). Secondly, the stratified sampling approach employed used NTDB landcover polygons to ensure sufficient sampling in the wetland class and also used them as a basis for sampling in the water class. Consequently, all validation points within these two classes also fell within NTDB polygons. Considering this in combination with the fact that the NTDB was also used as a model input, an inflation of accuracies in the wetland and water classes is likely. This hypothesis is further supported by the fact that coniferous, deciduous and bare classes had the lowest PA and UA. These are the only three classes which were not supported by the use of auxiliary data for training sample collection or as model input and thus solely relied on

the spectral difference of training samples for classification. Given the large overlap between the standard deviations of the spectral signatures between many classes (Figure VII), it is not surprising that classes without any auxiliary data for classification support yielded lower accuracies.

Finally, some points of concern which apply to both the validation dataset as well as the OOB metrics. Firstly, due to the limited amount of training samples available, the finetuning of the model parameters used for RF was done only by withholding 10% of the training data which naturally were not equal to the 20% which was withheld during each validation run. This introduces a form of circularity where data is used to optimize the model while also being used for validation (Congalton and Green, 2008). Secondly, validation was performed at the pixel level for both classification approaches. This was done in order to produce unambiguous results and was based on the general recommendation of Tyukavina et al. (2025) as well as Millard and Richardson (2015). However, what was not considered is the fact that the validation unit should never be smaller than the MMU (García-Álvarez et al., 2022). This was set to nine pixels for training sample collection in accordance with the resolution of the Landsat 5 imagery. Given the medium resolution of Landsat 5 imagery, it is impossible to derive useful accuracy metric from a pixel-based assessment methodology (Congalton and Green, 2008).

To conclude, while the study has provided two approaches to accuracy assessment, each of them comes with limitations and accuracy results should be interpreted with care.

5.3. *Carbon balance*

In answer to the third research question, it can be clearly said that landcover classification influences the carbon balance of the pre-impoundment PDB landscape (Figure XI). While the new landcover information does not alter the CO₂ balance beyond the upper and lower limits presented in previous studies (Teodoru et al., 2012), both classifications do lead to the prediction of a larger CO₂ sink and a smaller CH₄ source than previously assumed (Teodoru et al., 2012). The following will discuss differences between this study and the main previous study which has addressed the carbon balance of the pre-impoundment PDB reservoir landscape (Teodoru et al., 2012). Furthermore, the impact differing landcover classifications

have on carbon balance estimates will be evaluated.

Given that diffusive flux values for the different landcover types were based on measurements and literature values used by Teodoru et al. (2012), all differences in the total carbon balance of the reservoir are due to differences in landcover classification. Table IX summarizes the area accorded to each class by both models compared to the area it was accorded to by Teodoru et al. (2012) as well as the diffusive flux values applied to each class.

Table IX: Landcover area percentage from object-based and pixel-based classification as well as values derived from Teodoru et al. (2012) (here: TD). Flux values for CO₂ and CH₄, adapted from Teodoru et al. (2012), are shown and were used to calculate the carbon balance of the pre-flooded landscape in this study.

	Water	Deciduous	Coniferous	Wetland	Burned	Bare	Unit
OB	23	6	48	11	3	11	% of Area
PB	22	8	48	11	2	9	% of Area
TD	25	3	28	18	19	8	% of Area
CO ₂	335	-471	-159	-102	-9	-9	mg C m ⁻² d ⁻¹
CH ₄	1.80	-0.08	-0.24	39.40	-0.24	-0.24	mg C m ⁻² d ⁻¹

Considering this, some landcover differences particularly impacting the final carbon balance become apparent. For CO₂ predictions, it is noteworthy that Teodoru et al. (2012) predicts much less forest, both deciduous and coniferous. Instead, they predict more than five times the amount of burned area compared to this study. This naturally impacts the final carbon balance as diffusive flux values for deciduous and coniferous forest are 50 (deciduous) and 15 (coniferous) times more negative than the diffusive flux of burned area. The reason for these very different area values for this study compared to Teodoru et al. (2012) is likely different delineations of the burned landcover class. In their study, Teodoru et al. (2012) include areas burned in 2002 as well as older burned forest in their burned class. In contrast, this study only classified forest burned in 2002 as burned while older burned areas were classified as bare, coniferous, and deciduous, depending on the tree-stand visible from the Landsat imagery. Burke et al. (1997) found that soil respiration rates in forests in boreal Canada recover to pre-burned levels after around 7 years and that CO₂ emission rates are lowest after 2 years post-burn. It is therefore possible to assume that a map used for carbon balance estimates would profit from a differentiation between recently burned and

older burned sites. This assumption is further supported by the fact that the Landsat Max composite which was used in this study clearly shows substantial regrowth of forest in the old burned areas (Figure Vb). Thus, a differentiation between recently and older burned areas is likely preferable to improve the quality of a map in terms of carbon balance estimates. While it would have been ideal to differentiate between coniferous, deciduous and bare areas which have burned pre-2002 and those which have never burned, the maps presented in this study already improve landcover data by Teodoru et al. (2012) by accurately mapping the extent of post-fire regrowth which is assumed to affect the landscape's carbon balance (Burke et al., 1997). In this context it is noteworthy that PB classification predicts a stronger sink than OB classification due to OB classification predicting less deciduous forest (Table VII). While it is impossible to say which classification is more accurate, deciduous is interestingly the only class where OB classification has a lower producer's accuracy than PB classification.

Turning now to methane flux predictions, the higher values found by Teodoru et al. (2012) are due to the higher prediction of wetlands by Teodoru et al. (2012) compared to this study. Wetlands are often comparatively large sources of CH_4 (Bubier et al., 1997) and the CH_4 flux values from Teodoru et al. (2012) for wetlands are between 20 to 500 times higher than for the other landcover classes. Thus, accurate wetland prediction plays an important role in CH_4 balance assessment. Both Teodoru et al. (2012) and this study used landcover data from the NTDB (Natural Resources Canada, 2005) as a basis for landcover classifications. In contrast to other landcover classes such as water and forest, wetlands can be more difficult to distinguish (Amani et al., 2019), and this study therefore heavily relied on NTDB data to find suitable training sample locations for wetlands. In addition, the NTDB was included as an additional raster layer into the RF model which visually clearly increased the accuracy of wetland predictions during first try-out model runs. Unfortunately, it is unclear how Teodoru et al. (2012) used the NTDB to map wetlands. The results in this study follow the wetland predictions of the NTDB pretty accurately and, if anything, seem to still predict slightly more wetlands than the NTDB accounted for. Thus, it is unclear how Teodoru et al. (2012) arrived at their much higher wetland area values. Adding on to that the limitations in accuracy assessment in this study, particularly for the wetland class, it is difficult to argue which values are more accurate in the context of methane balance.

Nonetheless, a comparison of carbon balance values in this study and those made by Teodoru et al. (2012) clearly show that changes in landcover classification impact the carbon balance. This is further highlighted by the sensitivity plot (see Figure XI), which clearly shows the ranges in CO₂ and CH₄ predictions across different model runs. Due to only 80% of the training data being used in each run, each model output had a different landcover composition (see Figure IX) and thus a different carbon balance. While the landscape was always a sink of CO₂ and a source of CH₄, the magnitude varies. Interestingly, CO₂ values across all model runs have a range from -40 to -62 mg C m⁻² d⁻¹ which is much lower than the final OB classification at -37.3 mg C m⁻² d⁻¹ and also on the lower end compared to the final PB classification result at -53.9 mg C m⁻² d⁻¹. This is most likely due to the reduction of training sample size to 80% affecting classification results. McCoy (2005) recommends a minimum of 60 samples per class and a reduction of the amount of training samples per class by 20% leads to the wetland and burned class falling below this threshold. Given that the coniferous class was one of the largest in training sample collection (Table IV) it is likely that this class was less affected by a reduced sample size, leading to an over-prediction of coniferous trees during the validation runs and thus lower total CO₂ balance values.

5.4. *Limitations*

While this study fulfilled its aim and presented a landcover map of the PDB pre-impoundment which can be used for carbon balance estimates, the results should be interpreted with care as per the discussion above. Mainly, uncertainties in the reliability of the accuracy assessment using no field validation data should be considered while the choice and delineation of landcover classes is another major point of concern. Thus, uncertainty in the reliability of the maps produced in this study affects the confidence in the presented carbon balance estimates. While this uncertainty of carbon balance estimates related to landcover accuracy has not been addressed in previous studies, uncertainties related to diffusive flux measurements have been addressed and they do of course also apply to this study. The flux values per landcover class, taken from Teodoru et al. (2012), are not all direct measurement values and mean flux values per class might not be a good representation of fluxes within certain patches of that landcover class in general. Furthermore, in terms of using these carbon balance values

to quantify the loss of the pre-flooded landscape across the lifespan of the reservoir, the fire cycle of the region should be taken into account. Dessureault et al. (2015) looked into this for the PDB reservoir and found that, by taking into account forest fires, the impact the loss of forest has on the reservoir’s carbon footprint would be significantly reduced.

Therefore, carbon balance estimates based on this study should be considered to be highly uncertain and interpreted with care when used as part of life cycle carbon assessments of the hydroelectric PDB reservoir.

5.5. Outlook

Although the results for the carbon balance of the pre-impoundment PDB reservoir landscape presented in this study are highly uncertain, it has definitely been shown that uncertainty in landcover estimates should be considered in carbon balance estimates. This aspect has not received a lot of attention in the James Bay hydroelectric region in general and future studies could focus on an accurate pre-impoundment landcover mapping of other hydroelectric reservoirs in the region. Hydro-Québec is still planning to extend its current hydroelectric network (Hydro-Québec, 2025) and a methodologically clear landcover mapping approach, ideally verified through field accuracy assessment, would enhance the discussion surrounding the carbon footprint of hydroelectric reservoirs. Pre-impoundment landcover mapping should ideally be based on multidecadal imagery in order to account for the potential of forest fires and carbon fluxes of forests in different stages of post-fire regrowth. Furthermore, accurate landcover maps of pre-impoundment landscapes could be useful for the study of other environmental impacts of the reservoirs as well. For example, findings on the relation between pre-flooded landscape and the spatial heterogeneity of diffusive fluxes from the reservoir post-impoundment (Demarty et al. (2026), Rust et al. (2021)) would be more meaningful if the accuracy of the assumed pre-flooded landcover was known. Boreal reservoirs also influence local radiation balance (Ben Aba et al., 2025) and evaporation (Strachan et al., 2016) as well as leading to elevated mercury levels in nearby rivers, affecting the local population (Astrade, 1998). All of these environmental impacts could be more clearly assessed and put into context with an improved and validated map.

6. CONCLUSION

This study aimed to investigate the effect landcover classification has on the carbon balance of the pre-impoundment Paix-des-Braves hydroelectric reservoir landscape in Québec. In answer to the first research question whether a supervised land cover classification approach could produce a map of sufficient quality and stability to support reliable carbon balance estimates, two different random forest classification approaches were used which both produced maps suitable for carbon balance estimation. In light of the second research question aiming at exploring the possibilities of map accuracy assessment without field validation, two different accuracy assessments were performed which both found high accuracies across all classes. However, it was discussed that these results should be interpreted with caution as they likely overestimate true accuracies. Finally, the third research question, to what extent the new landcover classification alters the carbon balance of the pre-impoundment PDB reservoir landscape, was addressed. This study found that the pre-impoundment PDB landscape acted as a sink of CO_2 and as a small source of CH_4 and that the magnitude of the fluxes depended on the landcover classification results. Future studies could use a similar methodology to assess the carbon balance of hydroelectric reservoirs pre-impoundment to improve carbon footprint estimates and make results from different areas more comparable. The maps produced in this study could also be used as a basis for further environmental impact assessment of the impoundment of the PDB reservoir.

REFERENCES

- Amani, M., Mahdavi, S., Afshar, M., Brisco, B., Huang, W., Mirzadeh, S. M. J., White, L., Banks, S., Montgomery, J., & Hopkinson, C. (2019). Canadian wetland inventory using Google Earth Engine: The first map and preliminary results. *Remote Sensing*, 11(7), 842. <https://doi.org/10.3390/rs11070842>
- Amani, M., Salehi, B., Mahdavi, S., & Brisco, B. (2018). Spectral analysis of wetlands using multi-source optical satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144, 119–136. <https://doi.org/10.1016/j.isprsjprs.2018.07.005>
- Astrade, L. (1998). La gestion des barrages-réservoirs au Québec: exemples d'enjeux environnementaux. *Annales de Géographie*, 107(604), 590–609. <https://doi.org/10.3406/geo.1998.20878>
- Barbosa, P. M., Bodmer, P., Stadler, M., Rust, F., Tremblay, A., & del Giorgio, P. A. (2023). Ecosystem metabolism is the dominant source of carbon dioxide in three young boreal cascade-reservoirs (La Romaine complex, Québec). *Journal of Geophysical Research: Biogeosciences*, 128(4). <https://doi.org/10.1029/2022JG007253>
- Ben Aba, I., Nadeau, D. F., Sylvain, J.-D., Drolet, G., & Anctil, F. (2025). Estimating the impact of impounding an 85-km² hydropower reservoir in a subarctic environment on the local radiation balance. *Journal of Geophysical Research: Atmospheres*, 130(20). <https://doi.org/10.1029/2025JD044311>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brothers, S. M., del Giorgio, P. A., Teodoru, C. R., Prairie, Y. T., & Smith, R. (2012). Landscape heterogeneity influences carbon dioxide production in a young boreal reservoir. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(3), 447–456. <https://doi.org/10.1139/f2011-174>
- Bubier, J. L., Rock, B. N., & Crill, P. (1997). Spectral reflectance measurements of boreal wetland and forest mosses. *Journal of Geophysical Research: Atmospheres*, 102(D24), 29483–29494. <https://doi.org/10.1029/97JD01246>
- Burke, R. A., Zepp, R. G., Tarr, M. A., Miller, W. L., & Stocks, B. J. (1997). Effect of fire on soil-atmosphere exchange of methane and carbon dioxide in Canadian boreal forest sites. *Journal of Geophysical Research*, 102(D24), 29289–29300. <https://doi.org/10.1029/97JD01131>
- Canadian Forest Service. (2022). Canadian wildland fire information system (CWFIS) – Fire history. <https://cwfis.cfs.nrcan.gc.ca/en/fire-history> (Accessed 10 April 2026)
- Chen, G., Tian, H., Huang, C., Prior, S. A., & Pan, S. (2013). Integrating a process-based ecosystem model with Landsat imagery to assess impacts of forest disturbance on terrestrial carbon dynamics: Case studies in Alabama and Mississippi. *Journal of*

- Geophysical Research: Biogeosciences, 118, 1208–1224. <https://doi.org/10.1002/jgrg.20098>
- Chen, P.-Y., Srinivasan, R., Fedosejevs, G., & Kiniry, J. R. (2003). Evaluating different NDVI composite techniques using NOAA-14 AVHRR data. *International Journal of Remote Sensing*, 24(17), 3403–3412. <https://doi.org/10.1080/0143116021000021279>
- Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: Principles and practices* (2nd ed.). CRC Press.
- Demarty, M., del Giorgio, P. A., Deblois, C. P., Bilodeau, F., & Tremblay, A. (2026). The long-term C footprint trajectory and emissions attribution of a large boreal reservoir (Paix des Braves (Eastmain-1), Québec). *Water Resources Research*, 62(2), e2025WR041366. <https://doi.org/10.1029/2025WR041366>
- Demarty, M., & Tremblay, A. (2017). Long term follow-up of pCO₂, pCH₄ and emissions from Eastmain 1 boreal reservoir, and the Rupert diversion bays, Canada. *Ecohydrology & Hydrobiology*, 19, 529–540. <https://doi.org/10.1016/j.ecohyd.2017.09.001>
- Dessureault, P.-L., Boucher, J.-F., Tremblay, P., Bouchard, S., & Villeneuve, C. (2015). Uncovering the minor contribution of land-cover change in upland forests to the net carbon footprint of a boreal hydroelectric reservoir. *Journal of Environmental Quality*, 44(4), 1111–1118. <https://doi.org/10.2134/jeq2015.02.0071>
- Esri. (2025). World imagery basemap. <https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9> (Accessed 23 May 2026)
- Flood, N. (2015). Seasonal composite Landsat TM/ETM+ images using the medoid (a multi-dimensional median). *Remote Sensing*, 7(12), 6481–6500. <https://doi.org/10.3390/rs5126481>
- Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley, R. D., Jr., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Hughes, M. J., & Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, 194, 379–390. <https://doi.org/10.1016/j.rse.2017.03.026>
- Friedlingstein, P., O’Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., . . . Zheng, B. (2023). Global carbon budget 2023. *Earth System Science Data*. <https://doi.org/10.5194/essd-2023-global-carbon-budget>
- Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- García-Álvarez, D., Camacho Olmedo, M. T., Paegelow, M., & Mas, J. F. (2022). Land use cover datasets and validation tools: Validation practices with QGIS. Springer. <https://doi.org/10.1007/978-3-030-90998-7>

- Gebreegziabher, G. A., Degefa, S., Furi, W., & Abebe, G. (2025). Exploring the nexus between land use and land cover change and environmental parameters using Google Earth Engine. *Environmental Earth Sciences*, 84, 432. <https://doi.org/10.1007/s12665-025-12417-8>
- Gibson, P. J., & Power, C. H. (2000). *Introductory remote sensing: Digital image processing and applications*. Routledge.
- Houghton, R. A. (1995). Land-use change and the carbon cycle. *Global Change Biology*, 1(4), 275–287. <https://doi.org/10.1111/j.1365-2486.1995.tb00014.x>
- Houghton, R. A., House, J. I., Pongratz, J., van der Werf, G. R., DeFries, R. S., Hansen, M. C., Le Quéré, C., & Ramankutty, N. (2012). Carbon emissions from land use and land-cover change. *Biogeosciences*, 9, 5125–5142. <https://doi.org/10.5194/bg-9-5125-2012>
- Houghton, R. A., & Nassikas, A. A. (2017). Global and regional fluxes of carbon from land use and land cover change 1850–2015. *Global Biogeochemical Cycles*, 31(3), 456–472. <https://doi.org/10.1002/2016GB005546>
- Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., Gong, P., Yang, J., Bai, Y., Zheng, Y., & Zhu, Z. (2017). Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sensing of Environment*, 198, 57–72. <https://doi.org/10.1016/j.rse.2017.02.021>
- Hydro-Québec. (2025). Plan d’action 2035: Un plan pour sécuriser notre autonomie énergétique. <https://www.hydroquebec.com/a-propos/publications-rapports/plan-action-2035.html> (Accessed 20 May 2026)
- Hydro-Québec. (2026). Eastmain–Sarcelle–Rupert complex. <https://www.hydroquebec.com/eastmain-sarcelle-rupert-complex/> (Accessed 20 May 2026)
- Intergovernmental Panel on Climate Change. (2000). *Land use, land-use change, and forestry*. Cambridge University Press. <https://www.ipcc.ch/report/land-use-land-use-change-and-forestry/>
- Intergovernmental Panel on Climate Change. (2021). *Climate change 2021: The physical science basis*. Cambridge University Press. <https://www.ipcc.ch/report/ar6/wg1/>
- Intergovernmental Panel on Climate Change. (2022). *Climate change 2022: Mitigation of climate change*. Cambridge University Press. <https://www.ipcc.ch/report/ar6/wg3/>
- Jhonnerie, R., Siregar, V. P., Nababan, B., Prasetyo, L. B., & Wouthuyzen, S. (2015). Random forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries. *Procedia Environmental Sciences*, 215–221. <https://doi.org/10.1016/j.proenv.2015.03.028>
- Khan, Z., Khalid, W., Ali, S. A., Shamim, S. K., & Ahmad, A. (2024). Analysing the potential impacts of land use land cover (LULC) transformation on present and future carbon sequestration capabilities in the central Himalayas. *Discover Geoscience*, 2, 92. <https://doi.org/10.1007/s44288-024-00097-z>

- Khorram, S., van der Wiele, F. C., & Nelson, S. A. C. (2012). Remote sensing. Springer.
- Krivoguz, D., Chernyi, S. G., Zinchenko, E., Silkin, A., & Zinchenko, A. (2023). Using Landsat-5 for accurate historical LULC classification: A comparison of machine learning models. *Data*, 8(9), 138. <https://doi.org/10.3390/data8090138>
- Kulawardhana, R. W., Thenkabail, P. S., Vithanage, J., Biradar, C., Islam, M. A., Gunasinghe, S., & Alankara, R. (2007). Evaluation of wetland mapping methods using Landsat ETM+ and SRTM data. *Journal of Spatial Hydrology*, 7(2).
- Kumar, A., Kumar, A., Chaturvedi, A. K., Joshi, N., Mondal, R., & Malyan, S. K. (2023). Greenhouse gas emissions from hydroelectric reservoirs: Mechanistic understanding of influencing factors and future prospects. *Environmental Science and Pollution Research*, 32(12), 7149–7166. <https://doi.org/10.1007/s11356-023-25717-y>
- Lehner, B., Messenger, M. L., Korver, M. C., & Linke, S. (2022). Global hydro-environmental lake characteristics at high spatial resolution. *Scientific Data*. <https://doi.org/10.1038/s41597-022-01425-z> (Accessed 10 April 2026)
- Li, R., Fang, P., Xu, W., Wang, L., Ou, G., Zhang, W., & Huang, X. (2022). Classifying forest types over a mountainous area in Southwest China with Landsat data composites and multiple environmental factors. *Forests*, 13(1), 135. <https://doi.org/10.3390/f13010135>
- Lück, W., & van Niekerk, A. (2016). Evaluation of a rule-based compositing technique for Landsat-5 TM and Landsat-7 ETM+ images. *International Journal of Applied Earth Observation and Geoinformation*, 47, 1–14. <https://doi.org/10.1016/j.jag.2015.11.019>
- McCoy, R. M. (2005). Field methods in remote sensing. Guilford Press.
- Messier, D. (1998). The James Bay hydroelectric complex: Development and environmental context. *Environmental Monitoring and Assessment*, 49, 1–15.
- Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing*, 7(7), 8489–8515. <https://doi.org/10.3390/rs70708489>
- NASA JPL. (2020). NASADEM merged DEM global 1 arc second V001. https://doi.org/10.5067/MEASURES/NASADEM/NASADEM_HGT.001 (Accessed 2026-04-18)
- Nasiri, V., Deljouei, A., Moradi, F., Sadeghi, S. M. M., & Borz, S. A. (2022). Land use and land cover mapping using Sentinel-2, Landsat-8 satellite images, and Google Earth Engine: A comparison of two composition methods. *Remote Sensing*, 14(9), 1977. <https://doi.org/10.3390/rs14091977>
- Natural Resources Canada. (n.d.). CanVec series. <https://open.canada.ca/data/en/dataset/4c8a8b1e-3c0c-4f9c-9b1b-0f2d7b6a6f1c>
- Natural Resources Canada. (2005). National topographic database (NTDB). <https://open.canada.ca/data/en/dataset/1f5c05ff-311f-4271-8d21-4c96c725c2af/resource/86f7a1cb-8cbf-41b5-8146-2184265e9903> (Accessed 10 April 2026)

- OpenAI. (2026). ChatGPT (GPT-5.5) [Large language model]. <https://chatgpt.com> (Accessed May 2026)
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. <https://doi.org/10.1080/01431160412331269698>
- Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land cover classification using Google Earth Engine and random forest classifier—the role of image composition. *Remote Sensing*, 12(15), 2411. <https://doi.org/10.3390/rs12152411>
- Qi, J., Chehbouni, A., Huete, A. R., & Kerr, Y. H. (1994). Modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119–126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- Richards, J. A. (2022). *Remote sensing digital image analysis*. Springer.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Roy, P. S. (1989). Spectral reflectance characteristics of vegetation and their use in estimating productive potential. *Proceedings of the Indian Academy of Sciences – Plant Sciences*, 99(1), 59–81.
- Ruefenacht, B. (2016). Comparison of three Landsat TM compositing methods: A case study using modeled tree canopy cover. *Photogrammetric Engineering Remote Sensing*, 82(3), 199–211. <https://doi.org/10.14358/PERS.82.3.199>
- Rust, F., Bodmer, P., & del Giorgio, P. A. (2021). Modeling the spatial and temporal variability in surface water CO and CH concentrations in a newly created complex of boreal hydroelectric reservoirs. *Science of the Total Environment*, 815, 152459. <https://doi.org/10.1016/j.scitotenv.2021.152459>
- Silva, V. S. da, Salami, G., Silva, M. I. O. da, Silva, E. A., Monteiro Junior, J. J., & Alba, E. (2020). Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. *Geology, Ecology, and Landscapes*, 4(2), 159–169. <https://doi.org/10.1080/24749508.2019.1608409>
- Sleeter, B. M., Liu, J., Daniel, C. J., Rayfield, B., Sherba, J. T., Hawbaker, T. J., Zhu, Z., Selmants, P. C., & Loveland, T. R. (2018). Effects of contemporary land-use and land-cover change on the carbon balance of terrestrial ecosystems in the United States to 2050. *Environmental Research Letters*, 13(4), 045006. <https://doi.org/10.1088/1748-9326/aab540>
- Strachan, I. B., Tremblay, A., Pelletier, L., Tardif, S., Turpin, C., & Nugent, K. A. (2016). Does the creation of a boreal hydroelectric reservoir result in a net change in evaporation? *Journal of Hydrology*, 540, 886–899. <https://doi.org/10.1016/j.jhydrol.2016.06.067>

- Teodoru, C. R., Bastien, J., Bonneville, B., del Giorgio, P. A., Demarty, M., Garneau, M., Pelletier, L., Prairie, Y. T., Roulet, N. T., Strachan, I. B., & Tremblay, A. (2012). The net carbon footprint of a newly created boreal hydroelectric reservoir. *Global Biogeochemical Cycles*, 26(2). <https://doi.org/10.1029/2011GB004187>
- Tremblay, A., Varfalvy, L., Roehm, C., & Garneau, M. (2005). *Greenhouse gas emissions: Fluxes and processes*. Springer.
- Tyukavina, A., Stehman, S. V., Foody, G. M., et al. (2025). Land cover and change map accuracy assessment and area estimation good practices protocol. <https://doi.org/10.5067/doc/ceoswgcv/lpv/lc.001>
- Wang, Z., Chan, F. K. S., Feng, M.-L., & Johnson, M. F. (2024). Greenhouse gas emissions from hydropower reservoirs: Emission processes and management approaches. *Environmental Research Letters*, 19(7), 073002. <https://doi.org/10.1088/1748-9326/ad560c>
- Xu, H. (2005). A study on information extraction of water body with the modified normalized difference water index (MNDWI). *Journal of Remote Sensing*, 9(5), 589–595.
- Xu, H., Su, G., Li, C., & Deng, W. (2025). The differences, advantages, and disadvantages of various image compositing methods on Google Earth Engine. *Geo-spatial Information Science*, 1–28. <https://doi.org/10.1080/10095020.2025.2542534>
- Zaidi, S. M., Akbari, A., Samah, A. A., Kong, N. S., & Gisen, J. I. A. (2017). Landsat-5 time series analysis for land use/land cover change detection using NDVI and semi-supervised classification techniques. *Polish Journal of Environmental Studies*, 26(6), 2833–2840. <https://doi.org/10.15244/pjoes/68878>
- Zhu, Z., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., Jin, S., Dahal, D., Yang, L., Auch, R. F., & Gallant, A. L. (2016). Optimizing selection of training and auxiliary data for operational land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 206–221. <https://doi.org/10.1016/j.isprsjprs.2016.11.004>

APPENDIX

A. Additional maps of the study area

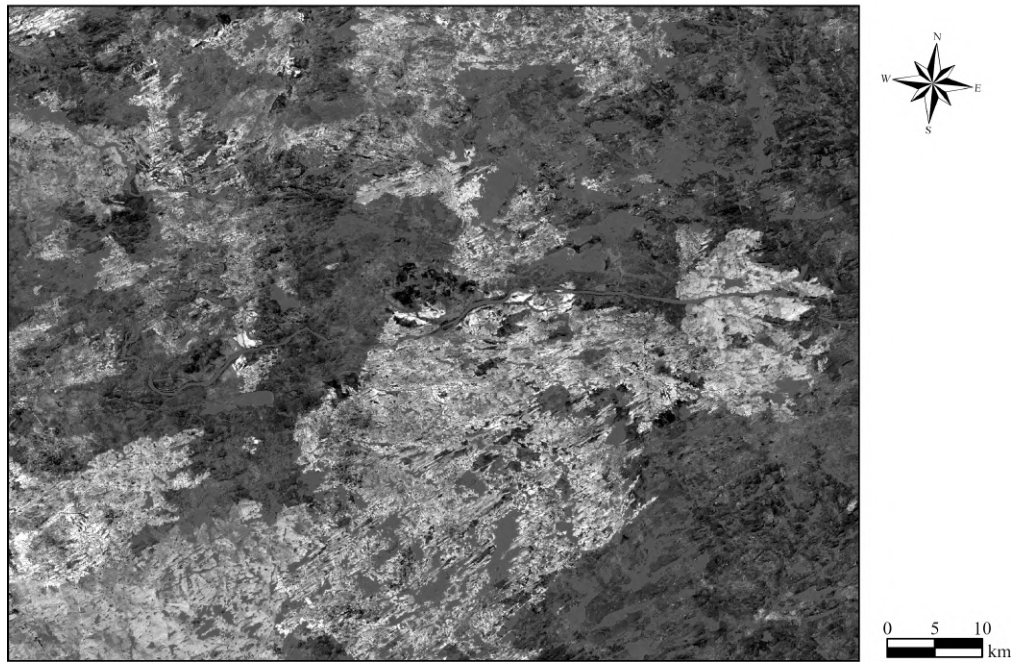


Figure A1: First principal component derived from both Landsat images. Data source: EROS (2020)

The first three principal components explained 98.3% of the variance in the input data. The PC1 held 84.6% of the Eigenvalues, PC2 8.4% and PC3 5.2%. Besides the rather small contribution of PC3, it is visually apparent that the component centers around the variation between burned versus non-burned area. The visualisations of the three first principal components can be found in A1 and A2.

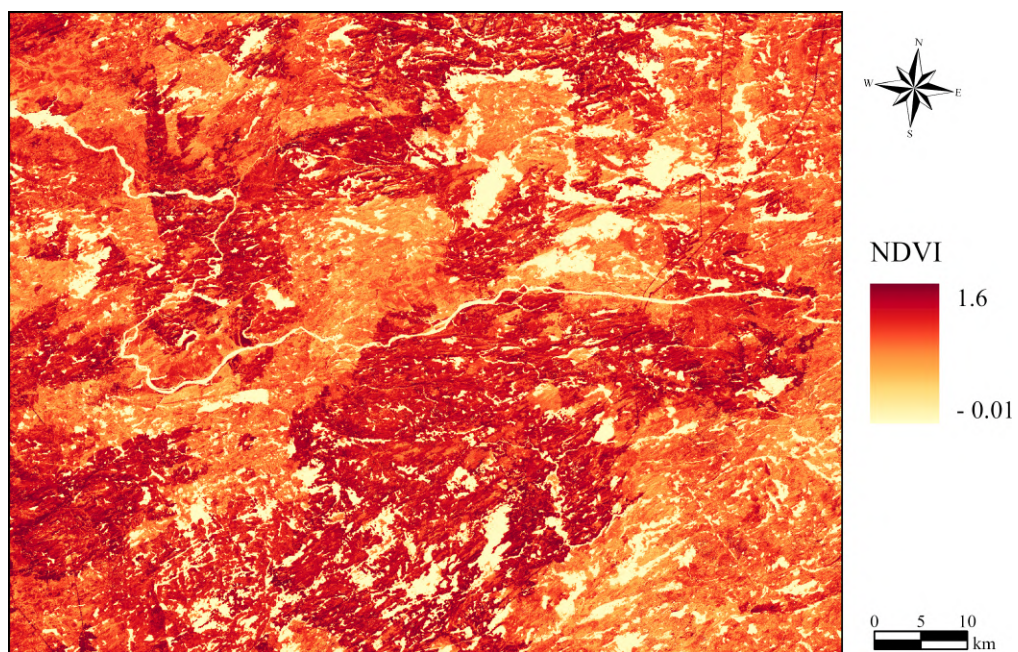


(a) Second principal component

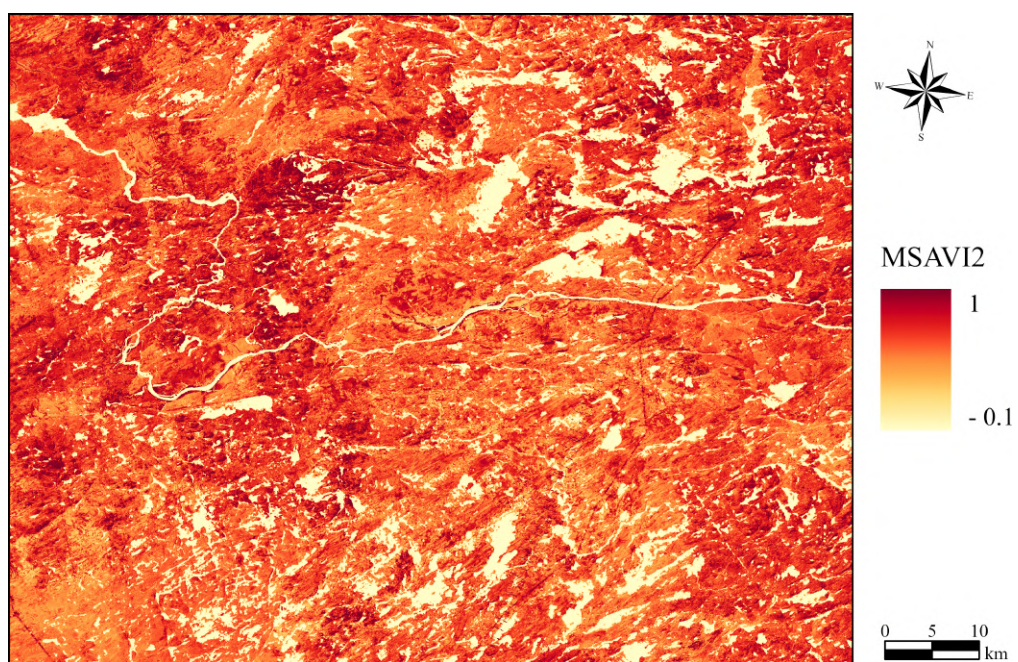


(b) Third principal component

Figure A2: Second and third principal components derived from both Landsat images. Data source: EROS (2020)

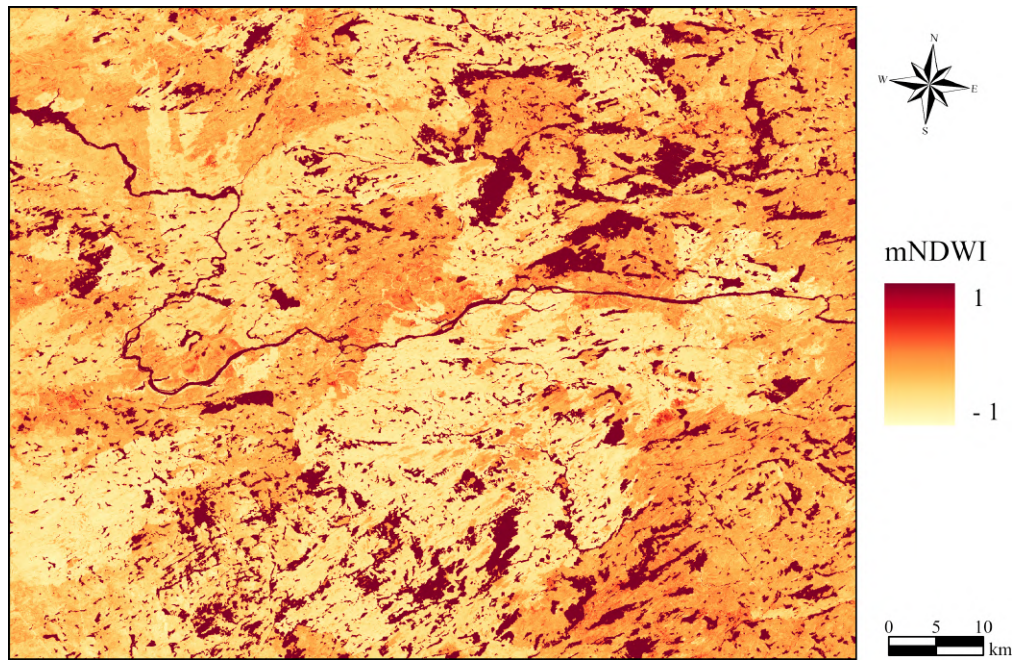


(a) Normalized difference vegetation index based on Landat Max composite

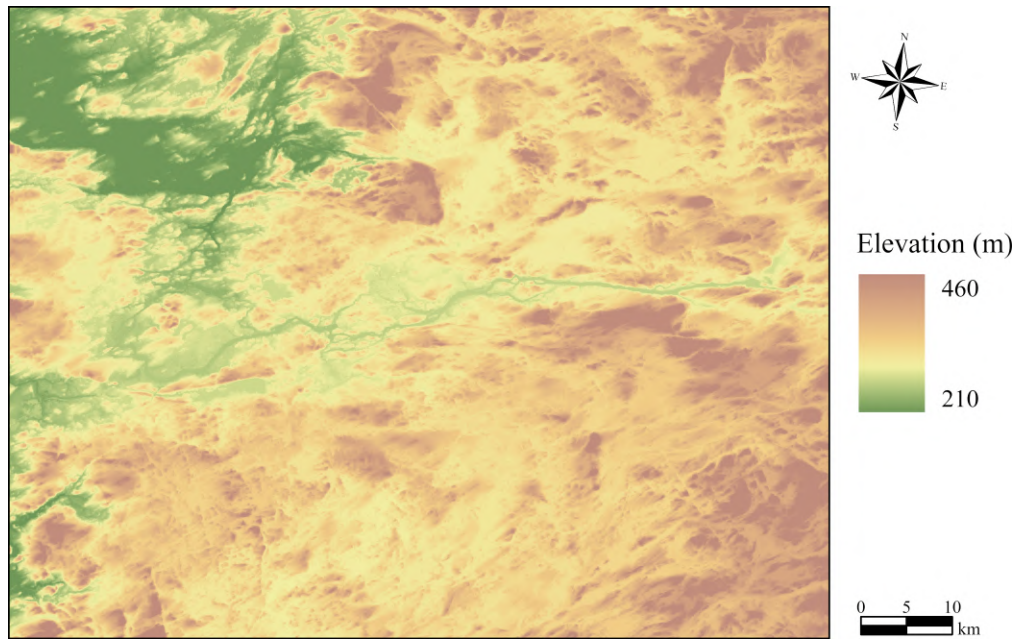


(b) Modified adjusted vegetation index based on the Landsat Median composite

Figure A3: Data source: EROS (2020)

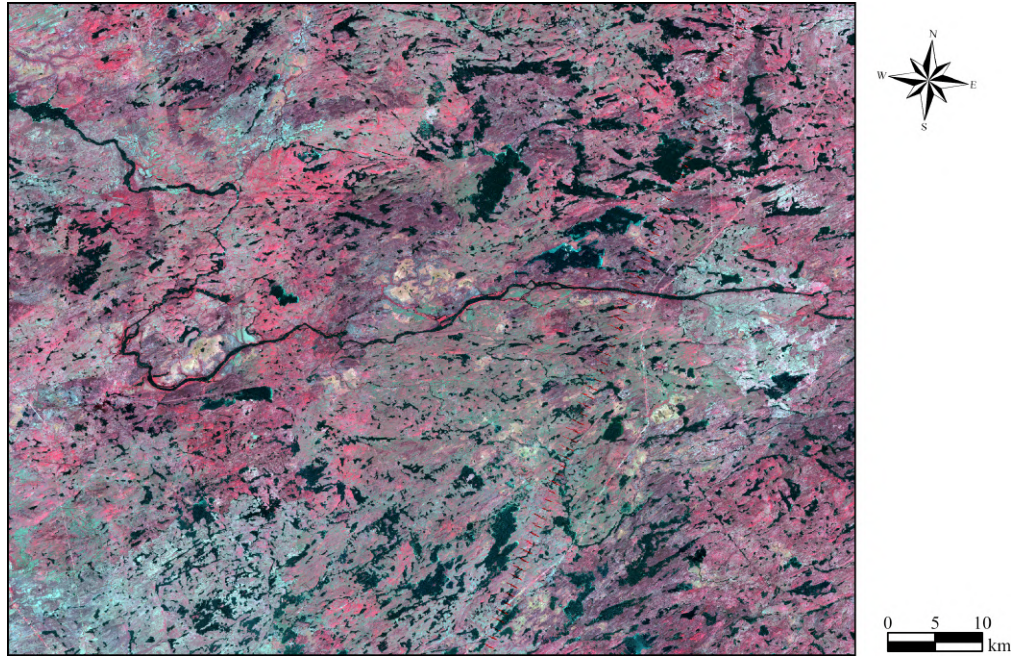


(a) Modified normalized difference water index based on Landsat Median Image

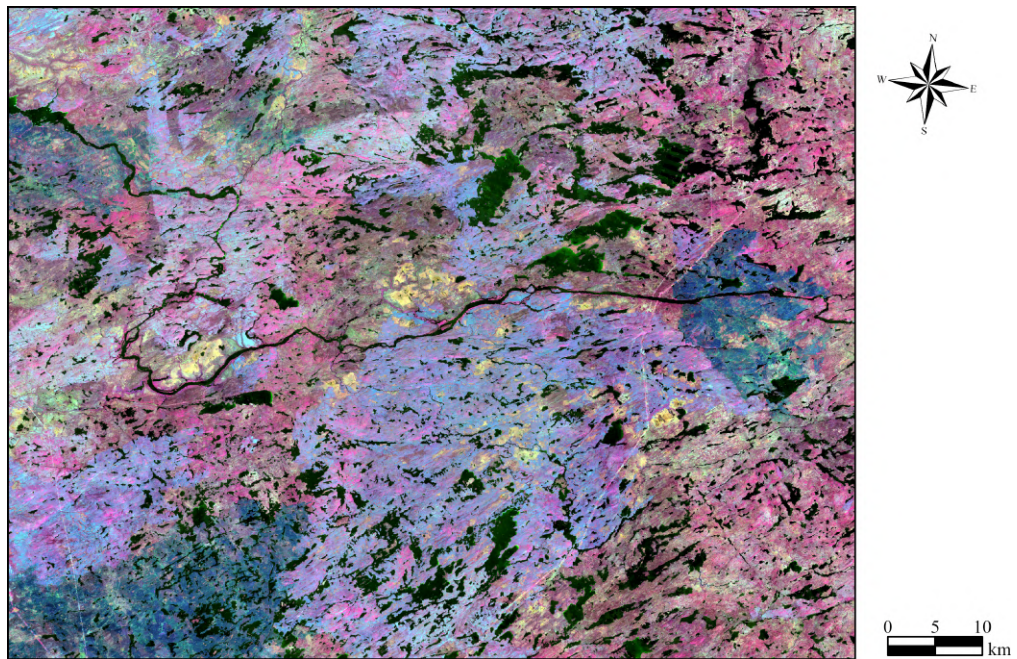


(b) Digital elevation model of the study area

Figure A4: Data sources: EROS (2020) for mNDWI and NASA JPL (2020) for the DEM.



(a) False color composite with bands NIR, Red, Green



(b) False colour composite with bands NIR, Red, SWIR1

Figure A5: False color composites used to differentiate deciduous from coniferous forests (Figure A5a) and wetlands (Figure A5b). Both are based on Landsat Median image (EROS, 2020)

B. Parameters for modeling and model output

Table A1: Random Forest Model Parameters

Model parameter	Input value
Number of trees	100
Minimum leaf size	3
Maximum depth	50
Random variables	3
Sample size	100%
Compensate for sparse categories (pixel-based only)	True
Maximum number of samples per class (object-based only)	1000
Segment attributes (object-based only)	Active chromaticity colour, mean digital number, compactness, rectangularity

	PC1	PC2	PC3	Landsat Median				Landsat Max			
				NDVI	NDMI	MSAVI2	mNDWI	NDVI	NDMI	MSAVI2	mNDWI
PC1	1.00										
PC2	0.00	1.00									
PC3	0.00	0.00	1.00								
NDVI	0.79	-0.29	-0.20	1.00							
NDMI	-0.33	-0.76	-0.27	-0.07	1.00						
MSAVI2	0.88	-0.35	-0.24	0.91	0.03	1.00					
mNDWI	-0.83	-0.23	-0.01	-0.79	0.64	-0.69	1.00				
NDVI	0.73	-0.24	-0.05	0.89	-0.17	0.78	-0.78	1.00			
NDMI	-0.42	-0.62	0.05	-0.26	0.75	-0.18	0.66	-0.23	1.00		
MSAVI2	0.94	-0.20	0.18	0.81	-0.23	0.87	-0.78	0.81	-0.26	1.00	
mNDWI	-0.80	-0.18	0.00	-0.76	0.57	-0.67	0.94	-0.76	0.74	-0.76	1.00

Figure A6: Correlation matrix between PCA components and spectral indices derived from the Landsat Median and Landsat Max composite images

C. Carbon balance

Ecosystem		Area (km ²)	Diffusive Flux (mg C m ⁻² d ⁻¹)		C Storage (mg C m ⁻² d ⁻¹)
Class	Subclass		CO ₂	CH ₄	
Natural aquatic	river	82.0	398 (279/445)	1.0 (0.8/1.1)	ND
	lake	66.9	239 (229/247)	2.8 (0.9/4.2)	-27 (-10/-52)
	stream	1.3	2751 (1115/6019)	ND	ND
Total aquatic		150.2	335 (241/399)	1.8 (0.9/2.5)	-27 (-10/-52)
Forest	coniferous	166.6	-159 (-376/-1)	-0.24 (-0.31/-0.17)	ND
	deciduous	15.6	-471 (-729/-214)	-0.08 (-0.09/-0.07)	ND
	burned	113.8	-9 (-88/71)	-0.24 (-0.30/-0.18)	ND
Total forest		296.0	-117 (-250/15)	-0.22 (-0.29/-0.17)	ND
Non-forest soil		45.9	ND	ND	ND
Wetland	peatland: fen	1.1	1278 (954/1603)	35 (30/40.5)	ND
	peatland: bog	85.4	-120 (-233/-6)	39 (33/45.5)	ND
	swamp/marsh	24.4	-102 (-218/14)	39 (33/45.5)	ND
Total wetland		110.9	-102 (-218/14)	39.3 (33.1/45.5)	ND
Total terrestrial (including non-forest soil)		452.8	-102 (-217/14)	9.5 (7.9/11.0)	ND
Total pre-flood ecosystem		602.9	7 (-103/110)	7.6 (6.2/8.9)	-27 (-10/-52)

^aNegative values represent C sink. Values between parentheses correspond to lower and upper limit estimates. ND denotes not determined.

Figure A7: Carbon fluxes per landcover class from Teodoru et al. (2012). Mean values per class were used for the estimates in this study, upper and lower bounds were not taken into account. Note that subclasses were only used in the forest class and that the measured values for burned forest were applied to bare ground, since no measurements for non-forested soil or any other bare class exist in Teodoru et al. (2012)