

**ON THE IMPACT OF LAND USE & LANDCOVER CHANGE ON GHG EMISSIONS
USING ADVANCED REMOTE SENSING TECHNOLOGY**

IMA-OBONG S. ITUEN

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN EARTH & SPACE SCIENCE & ENGINEERING
YORK UNIVERSITY
TORONTO, ONTARIO

July, 2024

© Ima Ituen, 2024

ABSTRACT

This doctoral research investigates the problem of insufficient reliable data causing lower accuracy in the simulation results of ecosystem carbon. There is an interesting effect of the changing climate: it has been observed in Northern Ontario that the warming weather has lengthened the growing season. Consequently, there has been increased land use conversion from natural forests to agriculture in Northern Ontario to capitalise on the new economic opportunities resulting from longer growing seasons. However, the long-term effects of land conversion from forests to agriculture are unknown. Yet, past studies indicate that the conversions will affect soil carbon, nutrients, and greenhouse gas emissions (GHGs). Given that over 66% of Ontario's land is forested, and this exemplar of land transformation could be experienced in other regions, it is vital to study the impacts of land conversion.

Remote sensing tools are valuable for mapping and monitoring changes in large geographic areas. This research study leveraged advanced remote sensing technology to detect changes in land cover and land use, an essential input in the simulation and prediction of ecosystem carbon and carbon flux exchange between the land surface and the atmosphere. In particular, a change detection method was developed wherein change analysis results were continuously adapted based on the most realistic data and the most current information. The novel change detection method developed in this research study was shown to be suitable for large-area mapping and to yield high accuracy.

A major problem addressed in this work was the lack of landcover classification data which is current, extends over the entire study area, and is of high spatial and temporal resolution with which to analyse the study region properly. (Furthermore, the existing classification maps do not always agree with each other in Northern Ontario and the Clay Belt area, which is the site of this study.) The lack of such data makes it difficult to accurately determine the effect of converting forested areas to agricultural lands in terms of changes in soil organic carbon, carbon sequestration, and greenhouse gas emissions. This thesis proposed a new change detection framework which uses the best available data for landcover classification and disturbance information, even in the event of scarce input or training data. The methodology combines a unique change detection method, a machine learning technique to classify the land cover types, and a protocol to select suitable carbon models to estimate carbon stock and GHG emissions. This thesis sought to exploit existing remote sensing datasets and the best available knowledge to generate realistic data. This data would serve

as input for simulations to obtain more accurate historical and future predictions of ecosystem carbon stock and emissions.

The landcover products employed in commonly-used carbon models are usually from one data source, e.g. Landsat maps. This study explored using multisource satellite data for landcover classification and change detection information being input to carbon models. The proposed method achieved an overall classification accuracy of 90.97%.

Another significant contribution of this study is a framework for curating the data which would be input into the carbon models. The carbon models require as input the specific land use and landcover type, as well as the disturbances which occur in the area under study. Hence, a unique disturbance-informed landcover change detection approach was created. Thus, an analyst can progressively adapt the available data sources to present a more accurate assessment of the disturbances. More accurate input data for the carbon models was shown in this study to decrease carbon and GHG estimates by between 15% and 32%.

The study concludes with tools that an everyday user can utilize to examine their region. The applications can run for free on their devices, permitting users to discover the landcover changes in their region, and estimate forest loss or gain. The applications could serve as a bridge to discussions concerning the areas that are best suited to concentrate carbon sequestration efforts.

ACKNOWLEDGMENTS

It has been, oh, so many years ago now, but I still remember where I was sitting the Saturday afternoon that I pulled out my credit card and applied for doctoral studies at York. It was a big decision, leaving a comfy job to enter the unknown world of academia. For sure, the experience was more challenging than I'd thought, yet extremely rewarding. I'm glad I embarked on the journey. Realising that I couldn't have reached here myself, I acknowledge some individuals:

So much gratitude for my lab mates who helped pick me up during darker times: Wendy, you're always an inspiration to me. Steven, Rory, Cheryl, Mike, Ben, Yongjie, and Miller, I'm incredibly grateful for the debates, discussions, fun – and hard work!! – in the lab and during the field campaigns. I always appreciate the congeniality you showed as we pursued our degrees. And as Cheryl constantly encouraged, let's keep going!

Mine was the best PhD Supervisory Committee on Planet Earth. I'm immeasurably grateful for your input. Dr. Taylor, I am always grateful that you continually gave me another perspective to consider during our talks. Gently, yet with deep conviction, you presented alternatives that could shape my research. Gunho, "thank you" is not enough for your input in my academic career. Your levity while challenging me to go further, continually calling me up higher, was a gift. I am ever grateful for the amount of time you invested in me. I'll pay it forward some day!! My supervisor, Dr. Baoxin, taught me the push-and-pull during supervising. I am forever grateful for your endless commitment to making me think bigger and write better. My Committee, you all stretched my thinking while navigating appropriate ways to engage the scientific discourse. And I feel incredibly blessed to have worked with such terrific educators. From my heart, thank you.

I'm also extremely grateful to the ESSE Dept which offered me the Enbridge Student Award and the Diane Michelangeli Memorial Scholarship for four years. The financial contribution was of great help to me during my studies.

Last of all, I'm incredibly grateful to my family and friends who kept me going on this journey – even those who went ahead in glory along the way. Thank you for your sacrifices, your encouragement, your endless supply of food and snacks, and most of all your belief in me. God being my help, I won't let you down. *Now on to bigger things! :-)*

Forever grateful to You, my Shepherd. You've been my strength through it all. *May my life be my worship...*

TABLE OF CONTENTS

Table of Contents

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES.....	ix
1. INTRODUCTION	1
1.1 Background	1
1.1.1 History repeated?	2
1.1.2 Changing Landscape	3
1.1.3 Remote sensing technology for landcover change detection	4
1.1.4 Introduction to Land Use and Land Cover Changes	5
1.1.5 Carbon Modelling for Canada: Counting Carbon emissions	8
1.1.6 Remote Sensing Tools for Carbon Modelling	9
1.2 Research Objectives and Contributions	13
1.2.1 Contributions	14
1.3 Thesis Outline.....	15
2. CARBON MODEL SIMULATING	18
2.1 Selecting suitable model from Carbon models in common use	18
2.2 Data input for Carbon Models: Patterns of disturbances in Clay Belt	26
2.3 Examining the Data Problem	30
2.4 Existing Land Use and Land Cover Change Products.....	31
2.4.1 Global Products.....	32
2.4.2 Regional and National Products	33
2.5 Uncertainty in Carbon Modelling	35
2.6 Creating Framework for Predicting Carbon	37
3. A NOVEL APPROACH TO CHANGE DETECTION IN THE CLAY BELT	41
3.1 Remote sensing systems and data	41

3.2 Land change detection	44
3.2.1 Machine-/Deep-Learning Change Detection	45
3.2.2 <i>Other Novel Change Detection Methods</i>	49
3.2.3 Fusing Multiple Image Sources	50
3.3 Statistical Method of Change Detection	51
3.4 Results	55
SUMMARY	60
4. MULTISOURCE LANDCOVER CLASSIFICATION	61
4.1 Satellite Imagery Data Sources.....	62
4.1.1 Data preprocessing.....	62
4.2 Data and Classifier Preparation	64
4.2.1 Data Sources for Land Use/Landcover Classification	64
4.2.2 Selecting Landcover Classes	66
4.2.3 Feature Selection.....	67
4.2.4 Obtaining Training and Testing Areas	68
4.2.5 Preparing the Classifier	69
4.2.6 Acquiring Disturbance Information for Change Detection	70
4.3 Change Assessment from Landcover Classification	73
4.3.1 Comparing Disturbance Maps to Classification Maps.....	75
4.3.2 Object-based landcover classification	76
4.4 Results	78
SUMMARY	88
5. CASE STUDY: CARBON AND GHG UNDER LAND CONVERSION IN THE CLAY BELT	90
5.1 Simulation framework: Model inputs and parametrisation.....	90
5.1.2 Field Survey and Community Consultation	96
5.2 RESULTS.....	101
5.2.1 Carbon Simulating and Predictions	101
5.2.2 Field survey results for validation	107
5.2.3 Sensitivity and Uncertainty Analysis	109

SUMMARY	115
6. CONCLUSIONS AND FUTURE WORK	120
6.1 Conclusions: Considerations from Carbon Accounting in the Great Clay Belt	120
6.1.1 Impacts of a Changing Climate	120
6.2 Contributions	121
6.2.1 On Evaluating and Improving Methods of Landcover Classification.....	122
6.2.2 On Exploring Effective Methods for Land Use and Landcover Change Detection	123
6.2.3 On Characterising the Impact Land Use Change has on GHG Emissions and Soil Carbon Stock.....	124
6.3 Future Research.....	126
REFERENCES	129

LIST OF TABLES

Table 2.1: Summary of some Carbon Models	23
Table 2.2: Common Land cover products used for Landcover and Land Use change analysis.....	35
Table 3.1: Summary table of sensors explored.....	44
Table 3.2: Confusion matrix representing change detection accuracy of NDVI-based test: 2005-2009	57
Table 3.3: Confusion matrix representing change detection accuracy of LST-based test: 2005-2009	57
Table 3.4: Confusion matrix representing change detection accuracy of NDVI-based test: 2010-2016	58
Table 3.5 Accuracy results from NDVI- & LST-based (joint) test: 2010-2016.....	58
Table 4.1: Change Detection Succession Matrix	75
Table 4.2: Overall Accuracy of Classification for the different sensors	83
Table 5.1: Average Carbon amount by transect	92
Table 5.2: Properties of the 10 sites simulated on the DNDC simulator	93
Table 5.3: Input parameters simulating carbon in sites and nearby forests for baseline comparison	95
Table 5.4: Average uncertainty of carbon and GHG (CO ₂) stock estimates on the different agricultural lands	112
Table 5.5: Comparing carbon and GHG amounts simulated under approximate and more accurate inputs.....	113

LIST OF FIGURES

Figure 1.1: Study Area of Ontario's Great Clay Belt	2
Figure 2.1 Historical disturbances from the past two decades in the Clay Belt: (a) Weather, infrastructure, and cuts; (b) Fire; (c) Forest disease damage; and (d) Forest insect damage	30
Figure 2.2: Illustrating lack of data – AAFC land use map of the Great Clay Belt area, and the GEDI Laser data product, available only below 51.6°N in Ontario	31
Figure 2.3: Workflow diagram highlighting the Progressive Adaptive Method (PAM) for carbon simulations	39
Figure 3.1: Example of smoothing done on NDVI data	53
Figure 3.2: Changed areas (coloured) detected from 2001 to 2005 in (a) NDVI; (b) LST overlaid on the map of the study area	55
Figure 3.3: Changed areas (coloured) detected from 2005 to 2009 in (a) NDVI; (b) LST overlaid on the map of the study area	55
Figure 3.4 Changed areas (coloured) detected from 2010 to 2016 in (a) NDVI; (b) LST overlaid on the map of the study area	56
Figure 4.1: Example of forest management plan for forest in Ontario's Clay Belt [91] showing areas delineated for harvest.....	72
Figure 4.2: Corrections made in reverting from changes in water class between (a) 2015 and (b) 2016	74
Figure 4.3: Segmentation perimeters from Simple Non-Iterative Clustering method	77
Figure 4.4: Landcover classification using Landsat 8 only.....	79
Figure 4.5: Variable Importance of the Random Forest model using Landsat 8 data	79
Figure 4.6: Landcover classification using Sentinel-2 only	80
Figure 4.7: Variable Importance of the Random Forest model using Sentinel-2 data	80
Figure 4.8: Landcover classification using Landsat 8 and Sentinel-2 imagery.....	81
Figure 4.9: Variable Importance of the Random Forest model using Landsat 8 and Sentinel-2 data.....	81
Figure 4.10: User's accuracy from Sentinel-2 classification	84
Figure 4.11: Landsat 7-derived Land cover areas in 2002.....	85
Figure 4.12: Landsat 7-derived Land cover areas in 2012.....	85
Figure 4.13: Landsat 8-derived Land cover areas in 2022.....	86
Figure 4.14: Comparing user's accuracy from pixel-based and object-based classification performed with Sentinel-1 and Sentinel-2	87
Figure 5.1: Pasture sites S17, S18, S27.....	92
Figure 5.2: Sample growth and yield curves of (a) white spruce, (b) larch, and (c) black spruce trees	94

Figure 5.3 Examples of various sample sites encountered in field survey: a) to c) is forested. In c), the overstory is trembling aspen, with understory of balsam poplar and raspberries; d) to f) are croplands: barley, cereal, and corn; g) and h) are hay	97
Figure 5.4: GPS base station for acquiring coordinates of plots	98
Figure 5.5: Collecting soil samples using a PVC tube	99
Figure 5.6: Bagging the soil sample collected using an auger	100
Figure 5.7: Ecosystem carbon stock for forest sites with conversion to land uses (predicting future carbon stocks from past disturbances)	101
Figure 5.8: Ecosystem carbon stock predicted for natural forest	102
Figure 5.9: Dead Organic Matter for sites, barring conversion to agricultural land though experiencing disturbances.....	102
Figure 5.10: Comparing the baseline forest carbon to some agricultural sites.....	103
Figure 5.11a: Sources of ecosystem carbon from deciduous forest, by site F1	104
Figure 5.12: Simulated soil organic carbon in (a) winter wheat field, (b) spring wheat field, and (c) CO ₂ emissions on oat farm	107
Figure 5.13: Comparing SOC from field survey to carbon estimates from CBM model	108
Figure 5.14: Comparing Soil Organic Carbon stocks and GHG (CO ₂) emissions – simulated from historical records, and simulated as climate-impacted (CI) conditions – on different sites: a) of F1 corn farm field; b) F5 soybean farm field, and c) P3 beef cattle on pasture; and d) P5 pig farm on pasture.....	111
Figure 5.15: Comparing carbon estimates from (a) Total Ecosystem carbon, and (b) Soil Carbon, using approximate vs. specific input data on forest trees.....	113
Figure 5.16 Comparing carbon estimates from accurate input data vs. estimates when disturbance and land use input data has been modified data on forested areas	114

1. INTRODUCTION

In this introductory section, I describe the background and approach to my study. I describe the Clay Belt area and its significance in Canada. This section also introduces the typical carbon modelling performed in Canada and the remote sensing tools used. Thereafter, the research objectives of the dissertation are provided. The outline of the dissertation is presented, followed by the publications which stemmed from my doctoral study.

1.1 Background

The warming climate is not only forcing populations to adapt and reconsider their normal modes of operation, but the changing climate is also exposing opportunities for certain industries. One such industry is the agricultural industry in the Great Clay Belt of Ontario, Canada. This area is undergoing a transformation to expand its agricultural land since the warmer weather experienced in recent decades has extended the growing season.

The Great Clay Belt extends from Quebec into Ontario; in Ontario, the area spans about 120,000 sq. km, as depicted in Figure 1.1. As its name goes, the region is known for its predominantly clayey soil. Rich in glaciolacustrine and morainal calcareous clays and silts, the soil has typically allowed barley, oat, and wheat crops to be grown successfully. Yet, with the increased temperature in recent years, other crops (e.g., soybeans, canola, corn grain, and silage corn) are now being grown. However, this area has harsh weather: the average temperature in July is about 17°C, with peak mean monthly rainfall of about 95 mm in July, and an average of 557mm of rain annually. Furthermore, the first frost occurs in mid-September [1]. Such weather conditions have discouraged several aspiring farmers and forced them to leave [2].

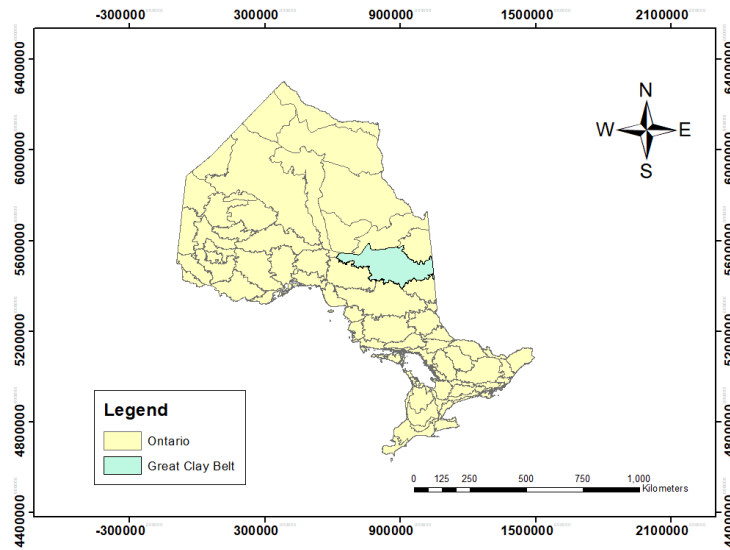


Figure 1.1: Study Area of Ontario's Great Clay Belt

In recent years, annual crop heat units have risen by 25% [3]. This rise has come with an increase in yield per acre and allowed other crops not traditionally in the crop rotation in the Clay Belt to be added. Ontario's government wants to expand agricultural opportunities in Northern Ontario to take advantage of the more favourable growing conditions, as an avenue to address food insecurity [4]. One way the government will increase agricultural activity is by attracting potential farmers for both livestock and crop farming to work in the region.

1.1.1 History repeated?

This is not the first time that the government has driven an initiative to populate the Clay Belt region with farmers. The 1920 Report Commission of Enquiry, Kapuskasing Colony, details the failed effort to maintain farmers living in that area. Soldiers who returned from the First World War were enticed to Ontario's Clay Belt with the promise of free land, twenty million acres of virgin soil, where alfalfa, clover, field roots, barley, and other grains, could be "grown with phenomenal success" [2]. The public announcement also stated that the area had no stones, had good drainage, and that the winter "sometimes touches a lower notch than Ottawa or Montreal" [2]. Unfortunately, the conditions were too harsh for many; the government had over-promised on the potential to succeed. Thus, those who had settled there were eventually given refunds on land and animals, and were given free transport to relocate to any other place in Ontario.

The challenges to agriculture in the Clay Belt region still exist. For instance, the clayey soil is known to lack good drainage, and the summer months are still wet, with an average of 14 rain days per month. However, a solution to the poor drainage has been found with tile drainage. And, with long-term warmer weather being expected due to the effects of climate change, agricultural endeavours could become even more possible, and even profitable.

1.1.2 Changing Landscape

Most of the area in which the Ontario government, in its most recent efforts, is initially targeting to establish agricultural lands is currently forested. The reasons for the land conversion are for food security, given that the population on Earth is expected to increase to 9 billion by 2050 [5], as well as the potential for increased land productivity from the longer growing seasons brought on by warmer weather, better drainage infrastructure, new technologies and crop varieties, and the low cost of land. Furthermore, the agricultural expansion could diversify the economy in that part of Ontario for the First Nations and Métis peoples and sustainably expand the livestock sector [4]. Thus, the Northern Livestock Pilot (NLP) Program in Ontario's Clay Belt region was born in 2017 [6], [7].

The motivation for my work is based on my desire to better understand the effects of global warming and to better understand the implications for land use/land cover change. Global warming is projected to push warmer climate more northwards. Several forested areas in the Clay Belt study area are expected to be converted for agricultural uses. This work aims to identify the possible soil carbon changes and greenhouse gas (GHG) emissions that will occur when the current forested land is transformed into agricultural lands. The long-term effects on soil organic carbon following the land conversion from forests to agriculture are not yet known and can be difficult to predict because of the variations in the driving factors such as forest properties, climate, and soil properties [8]–[10]. The quantification of the difference in soil properties such as carbon and nitrogen amounts can be done using carbon models. An essential input to the carbon models is historical and current land use status, as well as disturbances which have occurred in the region. Forest disturbances considered in this work include natural and anthropogenic disturbances. Natural disturbances include, for example, disease, wildfires, and insect infestation. Anthropogenic disturbances include infrastructure (such as constructing roads and buildings) and deforestation

for logging.

1.1.3 Remote sensing technology for landcover change detection

In my study, I investigated and tested several sources of satellite imagery to establish a reliable and operational method for detecting landcover changes in the Clay Belt. Sensors tested include: (a) The Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS satellite data is commonly used for large-scale long-term monitoring of the Earth's environment. In agriculture and forestry, changes detected from such satellite imagery can be used to monitor crop growth, detect soil moisture, and monitor deforestation and forest fires. (b) Landsat sensors, which have a higher resolution than the MODIS. Landsat is the most frequently used source of time series data for landcover mapping and change detection [11]–[13]. (c) The Sentinel-1 and Sentinel-2 sensors are also higher-resolution sensors, and they began acquiring data in 2015. They have been used in recent years to monitor land use and landcover changes. However, since they do not have the long-running history of MODIS and Landsat, the Sentinel data are not used as extensively for long-term landcover change mapping.

This study aimed first to observe the land use and landcover changes which have occurred in the study region over the past two decades to establish a trend of historical land disturbances, e.g., wildfires, harvesting, forest insect damage. Thereafter, predictions were made as to the frequency and extent of land disturbances in the region in the future and the GHG emissions which would result from land conversions and disturbances were simulated.

There are two common approaches to change detection. In the first category, the objective is to detect whether a change has occurred or not. The outcomes are always dualistic: “Change” or “no change.” In the second category, the nature of the change is also determined, and thus the solution provides “from-to” information. For studies focusing on large-scale monitoring, “change/no-change” detection is often carried out. In my study, I found the results from “change/no-change” detection could be used initially to identify areas for further detailed “from-to” change analysis. In this study, both the “change/no-change” and “from-to” detection were conducted. A detailed change detection based on high spatial resolution imagery was carried out for the areas and the periods where/when changes occurred. As discussed in Section 2.6, the “from-

to” detection was integral to the novel method developed for disturbance-informed change detection.

1.1.4 Introduction to Land Use and Land Cover Changes

“Land use” refers to how land is being used, and what functional purpose the land is serving, e.g., if it is agricultural, it is recreational, and so on. “Land cover” refers to the physical characteristics of the land. For instance, if the land cover is water, exposed rock, if the area is forested, etc. “Landcover” is more explicit in defining the vegetation; for instance, where the land use may be defined as agricultural, the landcover can be classified not only as crops, but as buckwheat or millet. Similarly, a recreational land use designation could include grassland (e.g., a park) as well as urban (e.g., community centre or arena) areas, yet the grassland and urban sites are separate landcover classes. In conducting image classification from remotely sensed data, both land use and landcover are commonly referenced, and both are included in the work. However, more emphasis is placed on landcover and the landcover changes that occur throughout the period under investigation.

Conducting landcover and land use classification is vital as they provide important information to policymakers and decisionmakers. The results from the classification are used in urban planning – providing an understanding of how growth is taking place in an area, and the connectivity or fragmentation of features in an area [14]. Furthermore, landcover and land use classification finds application in land monitoring, natural resource management, carbon accounting, and disaster mitigation and management.

The categorisation of a digital image is typically done by unsupervised or supervised classification methods. The supervised method uses training samples to classify the image into the different landcover classes. The classification is done by comparing each pixel to the spectral signature of each landcover class and labelling the pixel with the class to which it is most similar. Unsupervised classification can produce some spectrally-distinct areas that would not have been obvious or seen visibly/ with the eye. However, the unsupervised classification does not make use of training samples. Instead, the algorithm subdivides the image based on similar digital numbers. Clustering algorithms determine the natural groupings in the image data [15]. The input the user could give would be the number of ‘classes’ to create. Some standard unsupervised classification

algorithms include K-means, Self-Organizing Maps, and Iterative Self-Organizing Data Analysis Technique (ISODAT) classifiers [16]–[18].

The unsupervised classification does not permit one to know, for instance, what percentage of the land has changed from one type of landcover to another, whereas this information can be determined from the supervised classification. However, supervised classification has the drawback of requiring several training samples to improve its accuracy [19], [20]. Supervised classification requires a priori knowledge of the landcover types which are present in the image in order to ‘supervise’ the categorisation of the classes, whereas an analyst does not need to be familiar with the geographical area for unsupervised classification to be successful. Some supervised classification classifiers include the minimum distance, maximum likelihood, Support Vector Machine (SVM), Random Forest classifiers, and the artificial neural networks [21], [22]. Landcover classifiers can also be divided into parametric and non-parametric classifiers. Parametric classifiers assume a Gaussian distribution of the image data. The parameters adopted in the classification, such as mean vector and covariance matrix, are often obtained from the training samples [23]. Examples of parametric classifiers include the maximum likelihood, minimum distance, and linear discriminant analysis. The early landcover classification programs were mostly parametric because they grouped the pixel values on a probability distribution basis [21]. The non-parametric classifiers do not use statistical parameters to separate landcover classes, and they do not make assumptions as to the distribution of the image data. The artificial neural network, decision tree, and support vector machine classifiers are some examples of non-parametric classifiers. Advancements in pattern recognition techniques through artificial intelligence and machine learning methods have supported the growth of these classifiers.

Parametric classification algorithms are not deemed suitable for multisource classification; rather, non-parametric classifiers (e.g., neural network, decision tree, or knowledge-based methods) would be used [23]. In addition, non-parametric classifiers have been shown to perform better than the parametric methods in landcover classification over complex landscapes [20].

Another way to look at a classification method is by the level at which it is performed – pixel-based, sub-pixel-based, and object-based approaches. Pixel-based classifiers are the traditional classifiers which usually assign a class to each pixel as determined by its spectral

signature. The supervised and unsupervised classification methods, and most of the parametric and non-parametric classifiers [21], are pixel-based classification methods.

Whereas the pixel-based classification methods assume each pixel is a homogenous landcover type, sub-pixel-based classification methods are driven by the understanding that many pixels may contain more than one landcover type. Mixed pixels are often seen in medium- and coarse-resolution images. Thus, the spectral value of a pixel is assumed to be a linear or non-linear combination of the classes. Examples of sub-pixel classifiers include fuzzy-set classifiers [24], [25], for spectral mixture analysis. However, subpixel classification suffers from the complication in accuracy assessment – it is very difficult to obtain reference data, hence making it hard to evaluate such classification [23].

Object-based image analysis (OBIA) is based on the creation of homogenous geographic objects within the image. Image segmentation techniques merge the pixels into groups of objects with similar characteristics. The object-based methods typically integrate more information than the pixel-based methods, utilising texture, spectral, and spatial information into the landcover classification effort. Although creating the image clusters prior to classification incurs more computational cost, object-based methods have been shown to attain higher accuracy levels than similar pixel-based methods [26]–[28].

Overall, advances in remote sensing have permitted the development of automated or semi-automated change detection methods [29]. The thresholding method of change analysis establishes a threshold beyond which a change is assumed to have occurred. For instance, image differencing uses a threshold for grey levels of the image to note pixels which have undergone land cover changes [30], [31]. The image differencing can be performed through algebraic image differencing, vegetative index differencing, image ratioing, regression, or change vector analysis [32]–[35].

In my work, I aimed to improve on a change detection method which examines pixel-by-pixel differences. The different data sources and software that I explored as I worked to find a suitable change detection method will be discussed in later chapters. A framework that I have dubbed the Progressive Adaptive Method (PAM) was generated to determine the disturbances and landcover changes in a study area using mostly remotely sensed data [36]. The PAM method was relied on to generate higher accuracy inputs for the carbon models. Thus, it contributed to reducing the uncertainty in the final estimates from the simulations.

1.1.5 Carbon Modelling for Canada: Counting Carbon emissions

Canada plays a very important role in fighting global climate change and the effects thereof. It has been noted that Canada is warming at twice the rate of the global average [37]. Furthermore, about 6% of the world's forest carbon is stored in Canadian forests. Thus, the country's contribution to the warming of the global climate is not trivial. Accurate accounting of the country's emissions has proven to be a challenge, which my study aims to address.

There are many initiatives to address carbon emissions at a regional, national, and even global scale. For instance, New Brunswick has had a Climate Change Action Plan for almost two decades aimed at decarbonisation and climate resilience [38]. The United Nations also has a 13-pronged detailed approach for Reduced Emissions from Deforestation and Forest Degradation (REDD+) [39]. And Canada has a vision of having net-zero carbon emissions by 2050 [40]. As such, Canada has participated in various conventions such as the UN Framework Convention on Climate Change (UNFCCC), whose aim is to stabilise GHG concentrations to prevent dangerous anthropogenic interference with the climate system [41], and Canada signed on to numerous agreements such as the 2009 Bali Roadmap and in 2015 to the Paris Agreement. (Canada had ratified the Kyoto Protocol in 2002 but subsequently pulled out in 2011 [42]. These agreements are legally binding, and the federal government continues to progress toward achieving the carbon emissions targets.

Indeed, Canada has gone through several cycles in adopting what can be called a well-defined federal climate change policy. For instance, some of the policies which have been introduced in the past few decades include the 1995 Federal-Provincial National Action Plan, a National Implementation Strategy and Business Plan in 2000, a bilateral federal-provincial agreement, a 2007 Action Plan to Reduce Greenhouse Gases and Air Pollution [43], and more recently the Pan-Canadian Framework on Clean Growth and Climate Change inaugurated in 2016 [44], and the 2022 Emissions Reduction Plan [45]. The Pan-Canadian Framework was formed with the provinces and territories, having input from the Indigenous Peoples. The Emissions Reduction Plan is an ambitious roadmap to have the country reduce emissions by 40-45% below the 2005 levels by the year 2030 [45]. (2005 is normally referred to as the 'base year', being the

base year under the Kyoto Protocol. The year 2005 and 2007 are tied for among the highest GHG emissions per capita (of 23t CO₂ eq/capita) since reporting from the base year [46].)

Apart from the national plans and policies, Canada signing on to global accords usually requires regular reporting of progress toward the stated targets. For instance, there are annual National Inventory Reports to the UNFCCC and regular contributions to the reports of the Intergovernmental Panel on Climate Change (IPCC). The climate change sectors tracked for reporting include waste, agriculture, Industrial Processes and Product Use, Energy (categorised into fugitive sources, transportation, and stationary combustion), and land use, land use change, and forestry (LULUCF). Carbon accounting is a vital part of resource management, and, as noted above, is required in the fulfilment of agreements such as Kyoto Protocol and the Paris Agreement [47]. However, estimates are typically fraught with great uncertainties. Some of the reasons for the uncertainties include insufficient data inputs, inconsistent accounting methods, and enormous cost and effort to obtain field data. For instance, Wulder et al. [48] wrote that on a national scale, the ground plot data being relied on is provincial and inconsistencies exist among provinces in the methods of arriving at their values [48]. Geospatial accounting methods, such as the approach in my study, would give a boost to natural resource monitoring. Furthermore, consistent carbon counting methods across provinces would also instill more confidence in the estimates calculated. My body of work seeks to improve the carbon and GHG prediction by improving the input to the models. Thus, the cycle could be broken where ‘garbage-in’ produces ‘garbage out.’

1.1.6 Remote Sensing Tools for Carbon Modelling

Geospatially explicit mapping and monitoring of carbon stocks face challenges, many times because of the heterogeneity of the vegetation structure, the difficulty in characterising the various means of land cover change (such as in agricultural areas), and even the tools or techniques applied for carbon accounting. However, geospatial mapping tools perform better than traditional field plot estimations and are easier to deploy for large-scale mapping of carbon stocks. Traditionally, carbon stocks have been estimated from field plot calculations: researchers would visit the field and conduct measurements of the vegetation, obtain soil samples, and, through a series of formulae, estimate the carbon in that locale. (By way of definitions, carbon stock is the carbon amount which is stored in the ecosystem’s biomass and soil, having been sequestered from the atmosphere. It is typically measured in tonnes of carbon. The carbon flux is the average carbon

stock change, usually measured on a yearly timescale. For reasons of comparison, carbon density values may be presented for carbon stock and carbon flux, thus yielding measurements in units of tonnes of carbon/ha, and tonnes of carbon/ha/year.) Many field measurements are required for developing the models, for instance in estimating aboveground biomass. These plot measurement methods are usually valid for a relatively small portion of land, so it is difficult to accurately report on the carbon flux in a spatially explicit manner without making broad generalisations across a large area. Thus, geospatial mapping has been an additional tool to more accurately assess the land carbon inventory.

Remote sensing technology is widely used for advancing forest data and monitoring. Remote sensing is useful because it provides a reliable and repeatable method to monitor temporal changes of the earth's surface. This repeatability characteristic is useful for change monitoring and projection of carbon fluxes over an area. Common geospatial remote sensing tools used in carbon accounting efforts include Light Detection and Ranging (LiDAR). Airborne LiDAR has been effective at supplementing the field plot data [49], [50]; it has been used to derive aboveground carbon density with the LiDAR measurements providing the forest tree heights [50]. However, due to the high cost of obtaining LiDAR data for large study areas, it has been recommended that it is used with satellite imagery: integration of LiDAR and optical imagery for estimating vegetation structure is best done “using spaceborne optical and radar sensors imaging at 10-m to 1-ha spatial resolution in order to resolve fine-scale variation in vegetation cover and condition” [50]. Additionally, data from Landsat [50]–[53], MODIS [54], Advanced Very High Resolution Radiometer (AVHRR) [55], and SPOT (Satellite Pour l'Observation de la Terre) sensors [56] have also been used to provide input data into empirical estimations of biomass or above- and below-ground carbon. Together, these all provide a reliable source for landcover and land use mapping.

Remote sensing science has thus contributed significantly to carbon modelling. Remote sensing can yield an increased sampling across space and time compared to traditional field plot methods. Although field surveys should not be eliminated since they are needed for validation data, leveraging remote sensing with the empirical and process carbon models has been shown to improve the accuracy of forest carbon estimates [57]. Some of the parameters remote sensing can accurately provide include land cover, leaf area index (LAI), biomass, phenology, disturbances, and estimates of gross and net primary production. Sensors such as the optical, RADAR, LiDAR,

and hyperspectral sensors are the sources for these parameters. Those parameters would then be inputs to both empirical and process-based forest carbon models [49].

Yet, with all these advances, in my view, advanced remote sensing technology can be more fully exploited as a tool for geospatial carbon accounting. Although there are multiple data products for classified landcover/land use change as listed above, I have found a unique challenge: it is not always possible to obtain landcover classification data which is up to date, of high spatial and temporal resolution, and covers the study area. The currently available sensors, with their different spatiotemporal ranges, can be a solution to the problems of insufficient or inaccurate data input and lack of spatially explicit carbon accounting. A recent study in Canada that explored the opportunities for remote sensing science in the next generation of carbon models called for an integration of multiple data sources and advanced technology to estimate forest biomass and carbon accurately over time [49].

Though there are several important inputs to the carbon models, this my research focuses on improving the accuracy of the carbon and GHG estimates by providing more accurate land use and landcover change information, in addition to disturbance information, to the models. Multi-temporal remote sensing data from satellite images is used in this study to establish if there has been a historical pattern of change in the study area. Remotely sensed images are used instead of ground-based surveying because the images can be generated in a more cost-effective, efficient, and timely manner.

Now, datasets from multiple sources have proven better at discriminating between different landcover types than single-source data [58]. Image fusion integrates digital images from different sources to produce a single or more complete image, with minimal loss or distortion to the original image. Thus, more information is derived from the fused image than could have been derived from the individual images [59]. Investigations into multi-source landcover classification have been widely done [35], [58]. Some of the most used multi-source classification techniques in remote sensing are Dempster–Shafer theory of evidence, neural networks, decision tree classifier and knowledge-based methods.

In this study, I have conducted image classification from multiple sources: the Sentinel-1 SAR and the Sentinel-2 optical images, as well as the Landsat datasets, with a machine learning

classification method. Whereas optical data collection can be hampered by cloud cover, the active SAR sensor is able to gather data regardless of the conditions. Thus, the SAR data can improve the classification accuracy of the optical sensor. Furthermore, the textural features can add information which would be useful for classification. I used the grey-level co-occurrence matrix (GLCM) in this work to add a textural component to the classification. The GLCM is a pixel-level feature; it measures the number of times a combination of grey levels occurs in an image. For instance, it counts how often a pixel of value P appears next to a pixel of value Q, in a particular direction and distance.

Another input to the carbon models explored in this work was the record of disturbances which occur in the study area. In Canada, disturbances such as wildfires are tracked by each province [60], and each province's methods for recording incidents might be different. For instance, the categorisation of the wildfire, the time of burning, and the area affected, may not all be recorded – depending on the requirements of the region. Hence, the lack of complete data can hamper the estimates of dependent variables. A measure to counter the paucity of data could be using general data – such as the expected frequency of a wildfire in an area.

Indeed, such substitutions are sometimes required, if it is impossible to ascertain the actual disturbance information. Yet, advanced remote sensing tools offer a method to improve the knowledge of the historical influence of disturbances on carbon and GHG emissions. My study explores the added benefits of using the best available data for disturbance input to carbon models. Also, it compares the difference in carbon and emissions estimated from the models when more accurate data is input to the models vs. having approximate data as input.

Also important as parameters to the simulations was the identification of land use status (e.g. forested, pastureland, etc.), disturbance information, and tree type or tree specie if forested. The land use status is usually obtained from a landcover classification map of the area, if not previously known. Furthermore, the classification map could define the type of tree present, e.g. softwood, hardwood, or mixed forest. Now, the analysis of change in the landcover is required because the carbon models need the cover type at each time step to be defined. So, if an area has undergone the change of being an oat farm for three years, and is now fallow for a year, the expected emissions and soil carbon would be different. To properly simulate the carbon amounts, the change which occurred on the land is to be clearly defined. Similarly, if a forest stand has

increased from being a stand of poplar trees for the past 20 years to now having trembling aspen and balsam poplar present, the ecosystem carbon of the area is different. In order to properly estimate the carbon stock, the change analysis is required.

Another aspect of change detection related to the carbon models is the reason for the change. If a forested area shows loss of tree coverage between two periods, the analyst will want to know the cause of the deforestation. The carbon amounts simulated by the models would differ if the deforestation was due to fire or due to harvesting of the trees. Thus, the driver of the disturbance should be input to the model as well. Ontario's geospatial database GeoHub is regularly updated and so could also be a source of information regarding disturbance history.

With more accurate land cover information obtained from fused remote sensing data, the accuracy of the carbon and GHG estimates is expected to also increase. Disturbances and landcover change play a role in ecosystem emissions and carbon sequestration. This research demonstrates how multi-sensor landcover classification and change detection using the PAM method developed is a way to improve the accuracy of the carbon models. In turn, it will lead to increased accuracy of the carbon and GHG estimates from proposed land use conversion to provide greater insight for public policy makers.

1.2 Research Objectives and Contributions

Other studies have shown the impact agricultural land use changes have on health, the environment, and local climate [61], e.g., the increases in ammonium when forested land is converted to pasture lands [62]. A large-scale study was performed to estimate the changes in woody carbon stock in the entire cropland of Canada over a decade [63]. A study in Canada estimated the above- and below-ground organic carbon, differentiating the amounts from coniferous and deciduous trees [64]. These studies have been successful and contributed significantly to the knowledge of the differences in the various ecozones of Canada. Yet, to my knowledge, an in-depth look at the effect of climate change in Ontario's Boreal forests or the Clay Belt and the resulting carbon dynamics from land conversions has not often been done [65]–[67]. The challenge in performing carbon modelling using multi-source land classification is gathering the relevant data for classification and land disturbances.

My thesis focusses on more accurate simulations from better inputs of disturbances and land cover changes. As far as I've seen, research in this area is sparse, and it is hoped that this work can kickstart further studies for increasing accuracy of the carbon model outputs. It is worth mentioning that there is a deficiency of historical geoproducts in this area of Northern Ontario. By conducting the present research, I not only contribute to methodology development, but also to closing the knowledge gap. In addition, I demonstrate how multi-sensor landcover classification and change detection can improve the accuracy of the carbon models, and ultimately increase the accuracy of GHG emissions data.

To address the existing challenges and limitations, the goal of this thesis is to develop advanced remote sensing products to improve the assessment of the impact of land cover conversion on carbon emissions. To fulfil this goal, the following three main scientific objectives as investigations were carried out:

- 1) Develop methods to improve land cover and land use classification and change detection
- 2) Explore and determine the most effective carbon models to evaluate the impact of land cover conversion on carbon emission
- 3) Develop an effective framework and methods to characterise the impact land use change has on greenhouse gas emissions and soil carbon stock

1.2.1 Contributions

This study is unique in the position of the methodology, its scope, and application: I use multiple sources of data for land cover and disturbance information, which is not typically done for the application of carbon calculations. Furthermore, the study was conducted on a large-scale, in a northern part of Ontario, Canada where accurate remote sensing products needed for carbon modelling are not always available. A contribution of this study to scientific knowledge is that it provides a framework to analysts on how to best utilise remote sensing and existing geospatial data products for detecting land use/landcover changes, so the change information is an input to the carbon models. Consequently, the better data input for land classification and landcover change

increases the accuracy of the carbon model, and ultimately increases the accuracy of GHG emissions data.

Land cover changes have socio-economic impacts on biodiversity, the livelihoods and rights of Indigenous Peoples, food and water security [68], and environmental impacts such as soil and atmospheric carbon flux. Under the United Nations Framework on Climate Change, Canada is required to report on their GHG emissions and carbon removals annually [69]. Land transformations such as those studied in this work are a component of the input to such a report. Thus, it is vital to estimate the soil carbon and GHG emissions as accurately as possible.

Another important outcome of this study is providing a method to predict the potential GHG emissions and carbon content on a piece of land, particularly where there has been transformation in land use. The tool I developed permits the user to use historical and current land use information to estimate the ecosystem carbon. It is essential that the tool developed is operational so that it is accessible to as many people as possible, without users requiring extensive prior training. With these emission estimates, decisionmakers can be informed about the most sustainable routes in land transformation or development. The knowledge gained can be used and shared in various settings, and potentially has broad personal and industry usage.

1.3 Thesis Outline

The chapters of the dissertation are arranged as follows: This first chapter has introduced the study area and has laid the foundation for an understanding of remote sensing for carbon modelling. In addition, the background and motivation of the study has been discussed and the scientific research objectives enumerated. The second chapter introduces various carbon models which are used worldwide. The chapter also highlights some of the inputs to the models – disturbances and land use change. Thereafter, I discuss the disturbances commonly seen in the Clay Belt. In addition, the existing land use and landcover change products which can be used in current carbon models are discussed. Chapter 3 is based on my publication from 2021, coauthored with my supervisor Dr. Baoxin Hu, in the ISPRS International Journal of Geo-Information [70]. This chapter provides the current state of land cover/land use mapping and classification methods,

specifically highlighting multi-sensor data fusion. The focus of the article is ‘change/no-change’ detection. Thus, an introduction is given to the automatic and operational change detection method developed during this research study. In addition, some of the results from that work are presented.

The fourth chapter is also based on an article submitted to an IEEE journal [71]. This section of the work focussed on “from-to” change detection and the land classification conducted for carbon modelling. The data used for the analysis, the features selected, and the classification methodology are presented. Chapter 4 also displays some of the land use change analysis results in pixel-based and object-based change detection. Also, it addresses various image classification techniques and change detection methods. The fifth chapter presents the carbon estimation from the different carbon models used. Simulations are illustrated in this chapter to demonstrate how better input data can affect the carbon and GHG emissions estimated. There are also projections of the carbon sequestration and emissions that are possible when different land use schemes are embarked on in the areas marked out for conversion in the study area. Finally, Chapter 6 is the synthesis of the project, offering land management practices which consider the principles of sustainability. Suggestions for future work are also proffered in this chapter.

The dissertation research generated the following publications and presentations:

Publications

- ◆ Ima Ituen, Baoxin Hu (2024), *Assessing the Impact of Land Conversion on Carbon Stocks and GHG Emissions*, **Land 13 (8), 1291**
- ◆ Ima Ituen, Baoxin Hu (2024), *A Multisource data approach for Change and Disturbance Mapping of Ontario’s Clay Belt towards more accurate Carbon and Emissions Estimation*, **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing** (Paper submitted)
- ◆ Ima Ituen (2022), *Estimating Landcover Change and Greenhouse Gas Emissions using spatiotemporal MODIS data: A case study in Northern Ontario, Canada*, **Canada Watch**
- ◆ Ima Ituen, Baoxin Hu (2021), *An Automatic and Operational Method to Predict Landcover change using spatiotemporal MODIS data: A case study in Northern Ontario, Canada*, **ISPRS International Journal of Geo-Information, 10 (5), 325**

Conference Proceedings and Presentations

- ◆ Ima Ituen, Baoxin Hu (2023), *Assessment of Disturbances on Greenhouse (GHG) Emissions in Forests in Northern Ontario*, **GIS in Education and Research Conference, Toronto, Canada**
- ◆ Ima Ituen, Baoxin Hu (2023), *Multi-Source Data Approach for Land Use/Landcover Change in the Clay Belt of Ontario, Canada*, **International Geoscience and Remote Sensing Symposium (IGARSS), Pasadena, CA, USA, 2023, pp. 1553-1556, doi: 10.1109/IGARSS52108.2023.10282290**
- ◆ Ima Ituen, Baoxin Hu (2023), *Potential Carbon Emissions and Carbon Sequestration in the Clay Belt following Land Conversions*, **International Geoscience and Remote Sensing Symposium (IGARSS), Pasadena, CA, USA, 2023, pp. 1301-1304, doi: 10.1109/IGARSS52108.2023.10281454**
- ◆ Ima Ituen, Baoxin Hu (2023), *Modelling Soil Carbon & GHG Emissions from Land Use Changes in Ontario's North*, **Canadian Symposium on Remote Sensing, Yellowknife, Canada (Poster)**
- ◆ Ima Ituen, Baoxin Hu (2022), *Estimating Landcover Change and Greenhouse Gas Emissions using spatiotemporal MODIS data: A case study in Northern Ontario, Canada*, (Canada in Conversation: Crisis, Challenge and Change), **Robarts Centre for Canadian Studies Graduate Student Conference**
- ◆ Ima Ituen, Baoxin Hu (2021), *Estimating Greenhouse Gas Emissions from Landcover Changes: A case study in Northern Ontario, Canada*, **Association of Ontario Land Surveyors (Poster)**
- ◆ Ima Ituen, Baoxin Hu (2021), *Estimating carbon and greenhouse gas emissions in remote regions of Canada*, **Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., June 2021**

2. CARBON MODEL SIMULATING

Carbon in its various forms (e.g., carbon monoxide (CO) and carbon dioxide (CO₂)) comes from sources such as anthropogenic activities – burning fossil fuels, cement manufacturing, farming, and land use change – and the natural carbon cycle from respiration and decomposition of matter. On land, carbon stock comes from biomass (living organisms in forest) while dead organic matter is comprised of woody debris, the forest floor, and soil.

The greenhouse gases related to agriculture are nitrous oxide (N₂O), methane (CH₄), and CO₂. N₂O enters soils after fertilizer is applied. However, changes occur to the soil nutrient cycles after a harvest [72]. Where a flux tower provided by field work might be unavailable, the carbon and nitrogen flux within the ecosystem is often calculated using carbon models. The models permit the carbon and GHG emissions to be predicted under various scenarios, such as modelling the growth of different plants, or deforestation and afforestation. Additionally, natural and anthropogenic disturbances can be simulated. Some modifiers to the models may include different climate conditions, different land use proportions, different agricultural land types (e.g., pasture, cropland), and a variety of disturbances experienced.

In this chapter, standard carbon and GHG modellers are discussed, including CANDY, Roth-C, FORCARB, CENTURY, and CBM-CFS3. I discuss some of the inputs to the models generally, and then relate these inputs – disturbances and land use changes – to the models which can be used to study the Clay Belt.

2.1 Selecting suitable model from Carbon models in common use

CANDY

The CARbon and Nitrogen DYnamics (CANDY) model is an agro-ecosystem model [73]. Hence, the model is only for agricultural lands. CANDY is usually used to calculate the short-term dynamics of nitrogen transformation and the accumulation of carbon in cropland. The CANDY model is useful for assessing the carbon and nitrogen uptake from the presence of livestock. It has also been used successfully to estimate carbon flux in the long term. The main focus of CANDY is the soil properties, e.g., the soil water dynamics, the soil structure, and the soil nitrogen dynamics.

Some of the inputs to run the model are soil and plant characteristics, the agricultural management processes, and climate details such as precipitation and mean air temperature. The model will thus evaluate the soil organic properties – the carbon, nitrogen, and water content.

Roth-C

The Rothamsted Carbon (Roth-C) model [74] simulates soil organic carbon dynamics. In [75], an example is provided where Roth-C has been used to analyse the regional carbon dynamics. Soil samples were collected for soil analysis, and the soil organic carbon was determined. The soil organic carbon dynamics were estimated from the “trade” between five pools: decomposable plant material, resistant plant material, microbial biomass, humified organic matter, and inert organic matter. The study used the measured soil organic carbon content as the initial content, and then ran simulations of soil organic carbon dynamics based on a 0-15cm soil depth and a 0-20cm soil depth [75]. The results showed that the Roth-C did model the trend in soil organic carbon change. However, it underestimated the amount of change in both scenarios.

FORECAST

An ecosystem model called FORECAST is a forest carbon model. It has been used to simulate the soil carbon content and storage [76]. The forest carbon storage, C , was calculated as in equation 2.1:

$$C = B_i \times C_c \times A_i \quad (2.1)$$

where

B_i = the biomass of the type of forest

A_i = the distribution area of the forest, and

C_c = the carbon density of this forest.

The model was used to estimate the regional or national forest carbon storage from the forest inventory data. Jie et al. [76] found limitations in that the model did not simulate carbon storage very well, as it did not closely match the observed data.

CBM-CFS3

The Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3) is commonly used for assessing carbon stocks and forecasting. Although developed in British Columbia, Canada, it is widely used in other countries for their carbon accounting and reporting to regulatory bodies such as IPCC. This model simulates dead organic matter dynamics, carbon dynamics of above- and belowground biomass including soils, and can represent both stand- and landscape-level forest dynamics. CBM tracks carbon stocks, transfers between pools, and CO₂, CH₄, and CO emissions. The model is representative of Canadian forests, with data representing 10 of Canada's 12 ecozones that contain forests [77].

Some of the model's inputs include land-use classification changes, disturbances (e.g., fire, insects, deforestation, harvest impacts), and stand age. Among the numerous outputs obtainable from the tool are carbon stocks and fluxes, annual carbon transfers between pools, emissions to the atmosphere, transfers and emissions associated with disturbance types, and annual carbon stock change for the total ecosystem. These outputs are then used in various ways, e.g., for carbon budgeting.

FORCARB

The U.S. FORest CARBon Budget Model (FORCARB) is another model which has been widely used to calculate carbon stocks from forest inventories. It runs as a FORTRAN program. This model integrates many relevant factors: natural disturbances, land use change, land management, and climate. FORCARB considers the carbon storage in four ecosystems: Trees, soil, the forest floor, and understory vegetation. If forested land had reverted from agricultural land, the model assumes the soil carbon accumulated to levels similar to forest land that had never been cultivated.

The outputs from FORCARB simulations include harvested carbon by tree type (e.g., softwoods, hardwoods), total forest carbon by age class, total carbon sequestered and emitted, and net change in carbon – aboveground, soil, floor, understory, etc. This model is an inventory change method, so a base carbon amount needs to be known. Estimating non-CO₂ greenhouse gas emissions such as CH₄ and N₂O is also possible, but the model requires other input data to do so [78].

CO2FIX

The CO2FIX model simulates carbon dynamics in vegetative structure on a yearly timescale. Programmed in C++, it uses empirical yield for growth modelling. In estimating carbon amounts, CO2FIX considers carbon in above- and below-ground, soils, and derived wood products using the relation in equation 2.2.

$$C_T = C_b + C_s + C_p \quad (2.2)$$

where

C_T = Total carbon stored in the forest stand

C_b = Biomass carbon above- and below-ground

C_s = Stored carbon in soil organic matter, and

C_p = Carbon stored in wood products

Within the CO2FIX model, individual trees or species are assumed to exhibit similar growth. It also estimates carbon derived in multi-cohort stands, and the user can modify tree growth by different types of interactions in such stands. Thus, tree growth can be simulated where there is no competition, or competition exists, or the multi-cohort stands have synergistic effects on each other.

Some of the inputs to this tool are the crop/forest rotation length, rate of thinnings (e.g., 20% removed per quarter), competition effect (e.g., understory), mortality rate, initial soil carbon, etc. A study done with this model observed that the simulated biomass carbon calculation was mostly lower than the reported carbon amounts [79]. However, the lower values likely arose because the simulation was performed on a managed forest with regular thinning, keeping average carbon content at a lower level, while the data it was being compared with came from undisturbed systems.

CENTURY

CENTURY is another model for ecosystem analysis. It evaluates the effect of changes in management and climate on ecosystems. It also simulates the amount of carbon, nitrogen, phosphorus, sulphur, and soil water present in an area. CENTURY has been effective in ecosystems such as grassland, cropland, forests, and savannahs [80].

This model permits the simulation of disturbances such as fire, harvest, grazing, and cultivation. Some inputs used in CENTURY include monthly precipitation, soil texture, monthly maximum and minimum air temperature, and the timescale to be used.

DNDC

The DeNitrification-DeComposition (DNDC) model is used to estimate the carbon content on agricultural land. The model simulates carbon and nitrogen biogeochemistry in agro-ecosystems [81]. It can also predict crop growth, soil carbon dynamics, soil temperature and moisture regimes, and emissions of gases such as N₂O, nitric oxide (NO), ammonia (NH₃), CH₄, and CO₂. The DNDC model is a process-based model driven by four main ecological drivers: the climate, soil, vegetation, and management practices (such as crop type and rotation, tillage, fertilization, grazing, and manure application) of the area under study. Thus, the parameters required to run the model include the fertilizer scheme, the tillage applied, climate data, and grazing details of livestock on farmland.

Plants' biomass is calculated from the grain, stem, leaves, and roots. The decomposition is estimated from the soil organic carbon (SOC), soil climate, and soil nitrogen. The crop growth simulated by the DNDC is done on a daily time step. The soil organic matter is set in pools in DNDC, and the daily decomposition rate of the pools is dependent on such parameters as the soil clay content, nitrogen availability, and soil temperature and moisture content. The decomposed SOC is lost as CO₂ or distributed to other SOC pools.

The DNDC model simulates the effect of rainfall as well as NH₃ concentration in the soil, allowing the estimation of nitrification and denitrification which occurs in the soil. As a check for accuracy, the model outputs the daily photosynthesis, carbon allocation, water and nitrogen uptake so that the model results can be verified with the ground truths.

EFISCEN

The European Forest Information SCENario (EFISCEN) model is another forest inventory model which is free to any user. It is suitable for large areas, e.g., a region or country. The minimum area for simulation is 10ha.

EFISCEN depicts a forest as an area distribution over age and volume classes within a matrix. For each forest type, a separate matrix is set up. The model is structured for 5-year time steps, and is suitable for 10-12 timesteps (about 50 to 60 years) [82]. Therefore, EFISCEN is unsuitable for fast-growing tree species with short rotations. In addition, the model should not be used for unmanaged forests or uneven-aged forests because it was developed for even-aged and managed forests [82].

Other models

Estimating soil organic carbon at the regional or national scale is done using various models. One of such is the Global Environment Facility Soil Organic Carbon (GEFSOC) Modelling System. From a graphical user interface, the user can estimate soil organic carbon stock and stock change using the Roth-C model, the CENTURY model, and the IPCC method. Data inputs include climate, historical and current land use, and types of soil. The GEFSOC system predicts the effects of land use change on the soil organic carbon stock [62], [83]. Such information can be used for land use planning at regional or national scales, and for identifying areas for carbon sequestration [62].

The FullCAM model was developed to assist in Australia's carbon accounting [84]. It joins two submodels – CAMFor, the carbon accounting model for forests, and CAMAg, the carbon accounting model for agricultural areas – into a single code. The FullCAM model considers the different cycles that forests or agricultural lands can go through, e.g., deforestation, afforestation, and mixed land systems (agroforestry). CAMAg also considers the carbon output from the various possibilities of farmland; for instance, it counts the carbon for cropped land and grazed land. The FullCAM model incorporates the different disturbances that could affect the land such as fire, cropping, and harvest. It can also use landcover change data from remote sensing sources.

Table 2.1 gives an overview of some carbon models which were examined in this study.

Table 2.1: Summary of some Carbon Models

MODEL	FEATURES	APPLICATIONS
Integrated Model to Assess the Global Environment (IMAGE) v.2.4 [85]	Global model. Model drivers include demographics, economics, land cover and uses, carbon and nitrogen	Estimates Greenhouse Gas Emissions in 14 landcover types, including forested,

		cycles. Incorporates global carbon cycle, global landcover change, global temperature and precipitation patterns	deserts, and arable agricultural land
ROTH-C [74]		Ecosystem Model: Simulates the turnover of organic carbon in non-waterlogged topsoils that allows for the effects of soil type, temperature, moisture content and plant cover on the turnover process. Operates on a monthly timestep. Useful for modelling large areas, as is a regional-level model.	The impact of climate change on soil organic carbon dynamics [86]
CANDY [73]		Model for simulation of soil carbon and nitrogen. Consists of several process modules such as soil water dynamics, soil temperature dynamics, crop development, grassland with livestock, and nitrogen dynamics.	Model is used to calculate the short-term dynamics of nitrogen and carbon on cropland and livestock lands.
CENTURY [80]		Regional-level ecosystem process model. Evaluates changes in management and climate on ecosystems. Uses monthly time steps for simulations, even up to several thousand years.	Simulates amount of carbon, nitrogen, phosphorus, sulphur, and soil water in various ecosystems, such as grassland, cropland, forest, and savannah.
BIOME-BGC (BioGeochemical Cycles model) [87]		Global Biogeochemistry process model: simulates the development of forest carbon and nitrogen	Estimates net primary productivity and soil organic carbon of forested areas. [86]
CBM-CFS v.3 [77]		Forest ecosystem model. Simulates ecosystem carbon and GHG stocks and dynamics on a yearly time step. Also estimates annual carbon transfers between	Used for carbon budgeting of forests. CBM commonly used for assessing and forecasting carbon stocks and emissions to the atmosphere.

	pools. Is a stand-level and regional-level model, and so is used by several countries to model their nationwide carbon and GHG amounts.	
CO2FIX [79]	Stand-level carbon simulator of carbon stocks and fluxes on an annual scale.	Forest carbon model
DNDC [81]	Process-based model which simulates carbon and nitrogen in agricultural areas. Produces results on a daily time step. Sequence of execution of sub-models is soil climate – plant growth – decomposition – nitrification and denitrification – fermentation.	Model predicts crop growth, soil carbon dynamics, soil temperature and moisture regimes, and GHG emissions. Only for agricultural land uses.
EFISCEN [82]	For each forest type, separate matrix is set up consisting of age and volume classes.	Models forest carbon. Suitable for large areas, e.g., region or country. Model developed only for even-aged and managed forests.
FORCARB [78]	Calculates Carbon stocks from forest inventories. Model integrates factors such as natural disturbances, land use change, land management, and climate.	Estimates carbon and GHG emissions in forests.
IBIS [88]	Dynamic model which integrates various ecosystem phenomena. Has 4 modules: land surface, Vegetation phenology, vegetation dynamics, and carbon balance modules.	Determines ecosystem response to land use changes and climate changes in forested regions, grasslands, and deserts.

2.2 Data input for Carbon Models: Patterns of disturbances in Clay Belt

A critical parameter in carbon modelling is the history (or projection) of disturbances. A disturbance can be characterised as an interruption to vegetation. The disturbances which occur in the area could be caused by anthropogenic means and natural causes, such as burns, weather events, forestry harvesting, infrastructure, and pests. In the CBM model, the disturbances can be simulated to occur over the entire study area or can be limited to specific stands. Examples of the disturbances which have occurred in the Clay Belt are wildfires, logging, disease damage, and infrastructure disturbance such as road construction. These disturbances produce corresponding changes in soil carbon [89].

Historical land disturbances (such as logging, fires, road/infrastructure creation, and insect damage) over the period of study were obtained from various governmental datasets [90], [91], as well as some from the farmland owners. Figure 2.1a illustrates some of these disturbances. It is possible to simulate the transitions which occur following a disturbance. For instance, if a stand consisted of mixed species such as trembling aspen, eastern larch, and black spruce, following a harvest disturbance, the stand could now be only eastern larch and black spruce, or even eastern larch, black spruce, and balsam fir. The post-disturbance dynamics of the stand would depend on the growth and yield curves applied in the model. After a stand-replacing disturbance, the CBM model resets the age of a stand to 0. The stand will begin to grow according to the same growth curve if the same species of tree continues growing in that place.

In this work, transitions have been modelled for changes in landcover class, e.g., forest harvest, and for forest management (e.g., pruning and cutting; removing dead trees), wildfires, disease damage, insect damage, and agricultural expansion. Disturbances simulated in this study include:

- i. Land use/landcover changes: Landcover change was tested annually via the spectral angle method. If a change in landcover class was found, the simulation was updated with the new cover type. In modelling for the future, the historical trend was used to forecast the transitions until 2043. Similarly, an assumption was made about the area allocated to cropland, pastureland, and livestock. The agricultural expansion sequence was based on common farming practice in Ontario, as well as the information given by farmers in the

Clay Belt on how they operate their farms. Urbanisation was not considered in this work as it is unlikely that there will be substantial development in the region of interest (an area within the Clay Belt) within the coming years: The region under study is already accessible by local and major roads and is served by existing infrastructure. Thus, this study does not investigate an urbanisation transition.

- ii. Forest/Grass Cutting and Harvest: The DNDC model allows the user to input the yearly farming management information regarding grazing and grass-cutting: for example, the number of hours each day the livestock grazes, the number of days in a month of grazing, and the intensity (in heads/ha) of the specified grazing animal, are some parameters the user can input. Similarly, the number of times the grass is cut per month, as well as the parts of the plant cut should be defined – root, stem, leaf, or grain. The harvest and cutting activities performed in the forested land was also simulated. Some of the events were clear-cut harvesting of 60% of merchantable trees followed by salvage of the snags, or clear-cut harvesting of 30% of merchantable trees followed by burning organic residue (slash). Those events were stand-replacing, and the disturbance transitions were defined as such.

The Ontario government geodatasets provide information on the plans for commercial harvesting on Crown land [91]. These forest management plans are defined in two phases of five years. Following the completion of Phase I, each forest project is examined with an optional refresh of objectives or activities. During this period, a public consultation is conducted. Now, where trees which were allocated for harvest in Phase I were not logged, they can be incorporated into the Phase II plans, if necessary. However, areas designated to be cut in the second phase cannot be harvested before that, i.e. in Phase I. Using the historical forest management plans in concurrence with maps produced for landcover change assessment, the deforestation activities were simulated in the CBM model. Also simulated were the different events that started a farmland, having been converted from forested land: e.g., clearcut with slash-burn, and clearcut harvesting with salvage of dead stem snags from the stand. The Progressive Adaptive Method (PAM) [36] was adopted to establish the disturbances that most likely occurred, based on the forest management plans and the remote sensing data from the Landsat and Sentinel satellites. This feedback loop between what is proposed and what actually occurs regarding forest disturbances by

way of harvest is a key component of the PAM method. Thus, the maps from the provincial forestry unit are closely reviewed.

- iii. Fires: Wildfires are a significant natural disturbance in Canada's forests. They can arise from such sources as brush burns, lightning strikes, campfires, fireworks, railway locomotives, and powerline shorts. Historically, wildfires in Ontario have burned for a variety of periods – from a few hours up to several weeks. On average, wildland fires consume 2.5 million ha. each year in Canada, roughly half the size of the province of Nova Scotia [92]. They may typically burn more of the understory and midstory – though it still depends on how severe the fire is. After the fire, the understory of the forests could be open to light, permitting subsequent growth in the area. The transitions of the forest composition following each wildfire was simulated in the CBM model.

Ontario Wildfire Archive [93] and the provincial government's database Ontario GeoHub [94] contain historical records of the fires in the province since 1960. Thus, over the study period (from two decades ago until two decades ahead), the trend of historical fires in the region were used to simulate future wildfire patterns. Figure 2.1b illustrates the history of fires in the Clay Belt in the past two decades. In addition, a future effect of climate change could be increased extreme weather events and wildfires. Hence, the impact of climate change was introduced to the model in terms of the frequency and severity of wildfires after year 5. The different disturbances have varying impacts on carbon pools and transfers. For instance, wildfires could transfer carbon to the atmosphere (as they consume debris on the forest floor and subsequently release the carbon to the atmosphere), as well as to stands, and to pools like the dead organic matter (DOM) pool. However, forest harvesting transfers the carbon to the forest product sector or final wood products and stands. The annual emissions from Canada's wildfires have been known to equal the annual emissions from burning fossil fuels [95].

Other fires which were simulated were the regular burns performed on cropland. Farmers sometimes burn leftover straw after a harvest as fields are prepared for seeding the following spring. The controlled burning is easier and more cost-effective than tilling the straw. Furthermore, in areas with a high clay content, the soil is more likely to have compaction and draining problems. Thus, it can be difficult to till the straw/residue back into the clayey soil. Therefore, burning would be a choice a farmer could take to manage

the residue [96]. Apart from clearing the land for planting, burning of agricultural fields is also done in Canada to control pests, disease, and weeds [97].

- iv. Disease damage: The provincial geodatabase provides the spatial extent of diseases which have afflicted the forests since the early 2000s. Thus, the past forest disease damage events were simulated in CBM, and predictions for future events were made based on the trend observed. Some of the diseases which have been predominant in the Ontario forests include brown spot needle blight, ink spot of aspen, spruce needle rust, and Septoria leaf spot and canker [98]. Particularly in the Clay Belt, Septoria leaf spot and ink spot of aspen were the diseases that afflicted the forests, with the extent of the damage listed as moderate to severe. (See Figure 2.1c.)
- v. Insect damage: The insect damage within the study region has come from the forest tent caterpillar, the larch casebearer, and the birch skeletonizer during the study period. Again, the effects of the damage were moderate to severe. Figure 2.1d illustrates the insect damage experienced by vegetation in the Clay Belt in the past two decades. In the simulations conducted in CBM, salvage logging was only simulated after significant insect mortality. Significant insect infestation was taken as a stand-replacing disturbance. Partial stand mortality could also be simulated as a lesser effect of insect damage.

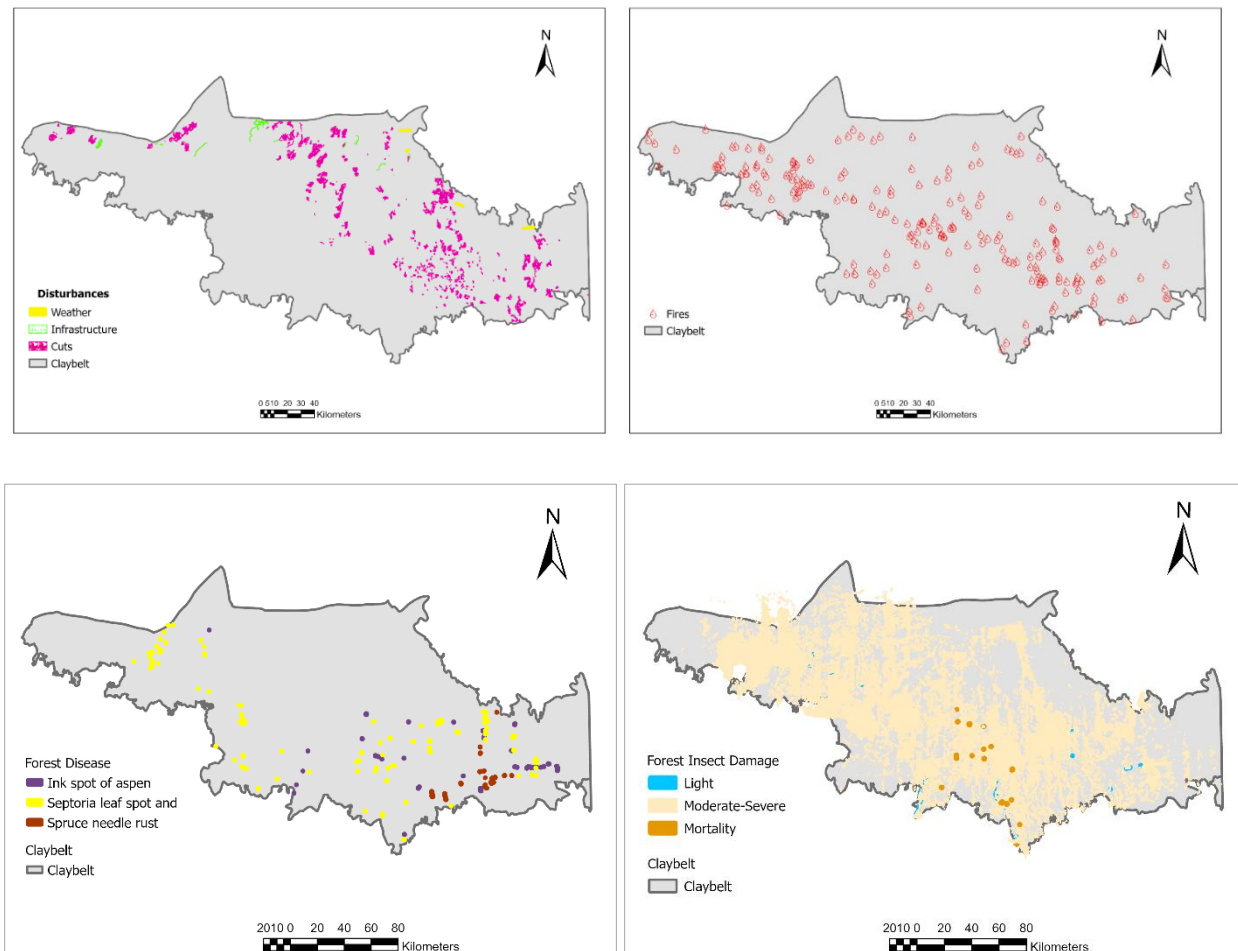


Figure 2.1 Historical disturbances from the past two decades in the Clay Belt: (a) Weather, infrastructure, and cuts; (b) Fire; (c) Forest disease damage; and (d) Forest insect damage

2.3 Examining the Data Problem

Northern Ontario lacks reliable, accessible, and continuous geospatial information – specifically land use information – which has resulted in expensive and protracted solutions, as well as overly broad portrayals of the region. An essential component of carbon modelling is the input of landcover or land use type. Additionally, information on the land transitions, such as afforestation, deforestations, fallowing, etc., is necessary for carbon modelling. However, the lack of continuous landcover maps in the Great Clay Belt region hampers consistent and reliable carbon estimation in the area, since the landcover type is required as a data input. My research is driven by the urgent need to develop suitable methodologies for detecting and mapping landcover dynamics in Northern Ontario for forest and agricultural lands. For example, the forest height data

for Canada is only produced every 10 years [99], and the GEDI canopy height data [100] (at ~25m resolution) is not collected for the entire Northern Ontario (see Figure 2.2 below). Figure 2.2 illustrates the issue of the lack of coverage in the Clay Belt area. These examples – with no data coverage at the topmost portion of the Clay Belt – buttress the issue of a data problem in Northern Ontario. In addition, they highlight the urgent need to develop suitable methodologies for detecting and mapping landcover dynamics in Northern Ontario.

In the absence of detailed maps that indicate when land conversions occurred, I propose to fill the gap with multiscale change detection methods using remotely sensed data. This will, in turn, improve the carbon estimates following land cover changes.

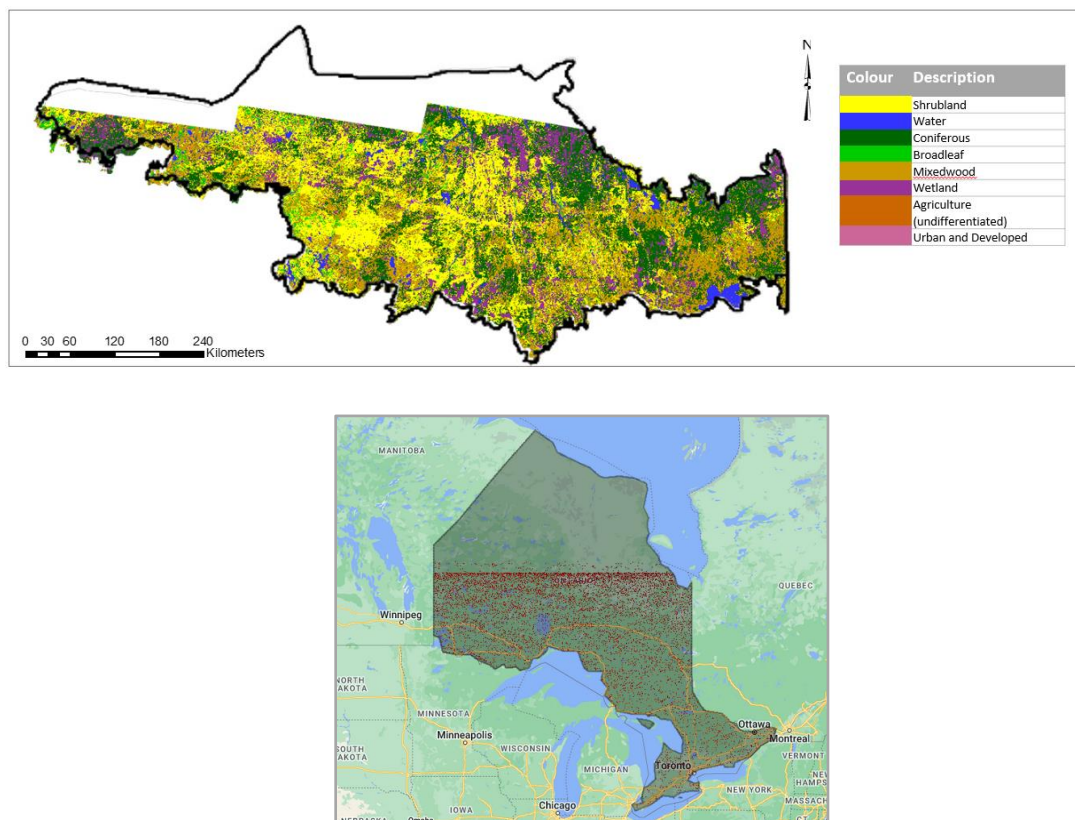


Figure 2.2: Illustrating lack of data – AAFC land use map of the Great Clay Belt area, and the GEDI Laser data product, available only below 51.6°N in Ontario

2.4 Existing Land use and Land cover Change Products

Now, the availability of satellite imagery has led to analysis-ready datasets, at varying spatial and temporal resolutions, which support landcover classification and change detection

globally and regionally. Currently, there are dozens of such mapping initiatives at varying spatial and temporal resolutions. Some of the products will be reviewed here:

2.4.1 Global Products

From 1992 to 1993, the Global Land Cover Characteristics (GLC) dataset was acquired at 1km spatial resolution off the AVHRR sensor. This dataset has 17 landcover classes and an accuracy of about 66.9% [101], [102]. The GLC 2000 project is also a 1km spatial resolution dataset, from the SPOT4 vegetation sensor. It was produced in 2000 with 23 landcover classes, having an accuracy of 68.6%. The dataset is based on the UN's Food and Agriculture Organisation (FAO) Land Cover Classification system. It follows, then, that the focus of the dataset is cover types of forests, grasslands, and croplands [103]. The MODIS Land Cover collection has five classification schemes, the primary one being the International Geosphere-Biosphere Programme (IGBP) [102]. This dataset has 17 landcover classes as well, at 500m spatial resolution. The accuracy is just over 70% [104], and the dataset is available from 2001 to 2018. Another global dataset is the GlobCover series. The GlobCover2005 dataset spans from December 2004 to June 2006. It has 22 landcover classes and an accuracy of 73.1% [102]. The GlobCover2009 dataset covers January to December 2009, with an accuracy of 67.5%. The GlobCover datasets were produced at 300m spatial resolution.

The Geo-Wiki initiative produced landcover maps using a crowdsourcing approach to validation. For these maps, users around the world volunteer geographic information by validating three landcover maps – GLC 2000, the MODIS IGBP, and GlobCover2005 [105], [106]. Two hybrid maps were produced for 2005, yielding an 82.8% accuracy [102].

The European Space Agency produced a Climate Change Initiative Land Cover product at 300m spatial resolution [107]. A baseline landcover map was prepared for 1992-1999, and the first global landcover map was produced for 1992. Another set of 10-year baseline landcover maps was made for the 1999-2013 and 2014-2015 periods. Thus, the dataset provides annual global landcover maps from 1992 to 2015 with 23 landcover classes, each at 300m resolution. Several satellite data sources were used for these landcover maps; they were the SPOT-VGT (1999-2013), AVHRR (1992-1999), MERIS-FR/RR (2003-2012), and PROBA-V (2013-2015). In 2016, another landcover map was planned using Sentinel-3 Ocean and Land Colour Instrument (OLCI) and Sea

and Land Surface Temperature Radiometer [108]. This dataset is widely used for change detection purposes.

The Global Land Cover 250 m from China (GLC250 m_CN) is also widely used for change detection. It is a 250m dataset, produced in 2001 and 2010. The datasets are based on MODIS data and have an accuracy of roughly 75% in both years [109]. The Chinese National Administration of Surveying, Mapping and Geoinformation created the GlobeLand30 landcover product on a decadal timeframe. Thus, the 30m maps are available for 2000, 2010, and 2020 [110]. This dataset has ten landcover classes and an expected accuracy of 78% [110].

2.4.2 Regional and National Products

The Coordination of Information on the Environment (CORINE) Land Cover (CLC) dataset is produced by the European Environment Agency for the European region. The first inventory was from 1990, and the dataset was updated in 2000. After that, an update cycle of 6 years was implemented; thus, there are currently five assets, the most recent landcover asset being from 2018. This CLC dataset has a spatial resolution of 100m, and each asset has 44 classes. Therefore, it is possible to use the dataset for change detection as there will be no ambiguity with the categories of a classification [111].

Another European land use and landcover data source is the Land Use and Cover Area frame Surface (LUCAS) product [112]. It is not sourced from satellite imagery, but from aerial photographs and field data. The changes in landcover have been calculated this way since 2006. The LUCAS product has seven landcover classes, provided at a spatial resolution of 2km [102]. Over in Africa, the AFRICOVER project began in 1995 to map the entire landcover of Africa; however, it has not yet been completed: only East Africa has fulfilled the mandate. Each country is to map their landcover with an aim of 80% accuracy, using 21 FAO classes [102], [113].

In America, the National Land Cover Database (NLCD) is perhaps the most widely used and referenced landcover dataset for the conterminous U.S. The NLCD began providing landcover maps in 1992 with a 21-class map at 30m resolution [102]. However, all the other map products released since have been 16-class datasets. Thus, the time series maps are very useful for analysing

multi-temporal landcover change. So far, there have been nine epochs of the 16-class maps from 2001 to 2021 [114].

Canada has also had a long-running time series product: The LCC85-05 was based on data from the AVHRR sensor and had landcover maps produced from 1985 to 2005. The 12-class dataset had an overall accuracy of 65% [115]. Similarly, the 250m LC Time Series 2000-2011 dataset was a landcover product created from the MODIS sensor, with a spatial resolution of 250m and an accuracy of between 70 and 80% [102]. This product uses 19 FAO landcover classes [116]. Another product created for Canada was the 20m LC product based on SPOT 4/5. It has 16 classes and was mapped for the period 2005-2010. The dataset has an overall accuracy of 71% [117].

Dynamic World is a unique dataset that was recently introduced to the change detection and land use/landcover mapping space. It provides a near real-time land use/landcover classification. Furthermore, it is produced to leverage deep learning at a high resolution – based on the 10m Sentinel-2 imagery. Therefore, landcover classifications can be updated continually with each new Sentinel-2 acquisition. Dynamic World appears to be the first of its kind as a near real-time dataset [118]. This product has been available since the inception of the Sentinel-2 offering in June 2015, and it is being hosted on the Google Earth Engine platform. By way of accuracies, when compared with expert and non-expert annotations, the nine classes of Dynamic World had an agreement of between 73.8% and 77.8% [118]. An interesting feature of the Dynamic World is that it does not attempt to classify an image's pixels; rather, it provides the probability that a pixel belongs to any of the nine classes in the Dynamic World dataset. Thus, it is not a ready-to-use landcover product like the preceding are [119].

Table 2.2 below provides a summary of some existing landcover products which are used in landcover change analysis.

Table 2.2: Common Land cover products used for Landcover and Land Use change analysis

Landcover product	Spatial resolution	Accuracy	Number of classes
Global Land Cover Characteristics	1km	66.9%	17
MODIS IGBP	500m	73.6%	17
GlobCover2005	300m	73.1%	22
GlobCover2009	300m	67.5%	22
Climate Change Initiative Land Cover	300m	71.5%	23
GlobeLand30	30m	75%	10
CLC	100m	78.5%	44
LUCAS	2km	79% (producer's accuracy)	7
NLCD	30m	86.4-90.6%	21, 16
LC Time Series	250m	70.8%	19

A drawback of global datasets is the restriction it could place on an analyst in defining classes. Therefore, collecting training samples from global datasets is not always ideal as the dataset might not have the land cover classes the analyst desires. Furthermore, a global dataset is sometimes less accurate at local scales [120]. This study provides a framework – particularly for regions which do not have regularly updated landcover/land use datasets of high accuracy (>80%) and higher resolution – so analysts can best use remote sensing and existing geospatial data products in determining the landcover/land use changes as inputs for carbon modelling.

2.5 Uncertainty in Carbon Modelling

Canada's forest carbon and GHG accounting is done under the National Forest Carbon Monitoring Accounting and Reporting System (NFCMARS). The NFCMARS relies on the v.3 Carbon Budget Model of the Canadian Forest Sector [77] as its core ecosystem carbon model [121], [122]. The model uses the best available data and scientific understanding of processes related to forest cycles. The CBM-CFS3 model adheres to the Intergovernmental Panel on Climate Change (IPCC) reporting guidelines for carbon and GHG estimates [123].

Inherent in the CBM are sources of uncertainty such as data input to the model by the user, the ecological parameters used in the simulations, processes incorrectly specified or even excluded, and the model's algorithms [77]. Such is not uncommon with carbon models; apart from the CBM, ecosystem carbon models have been known to have sources of errors from the uncertainty in inputs, the incorrect calibration of the models, the size and/or scope of the sample data, and even insufficient field measurements. On the other hand, these factors, such as sufficient field data and correct calibration of the models, could contribute to more robust model estimates [124]. Models are not able to simulate all the ecosystem biogeochemical processes completely. So approximations have to be made, although uncertainties have been said to diminish over large spatial scales as the internal variability averages out [125], [126]. The international climate reports also require countries to assess the uncertainties in their carbon and GHG emissions estimates. Indeed, carbon accounting is a complex endeavour, and Canada has always reported the uncertainties associated with its estimates. However, a criticism of the reports is that the CBM-CFS3 results are published but the input variables are not publicly available [122].

The carbon and GHG flux estimates typically have large uncertainties, especially for large areas since modelling all the physical, chemical, and biological processes entailed could be difficult. For example, the national uncertainty range for GHG fluxes from non-converted forest land was $\pm 39\text{MT}$ of carbon dioxide (CO_2) between 1990 and 2021 (where non- CO_2 emissions do not contribute much to the total uncertainty) [127]; or uncertainties of emissions/removals on cropland was reported as $\pm 19\%$; and the uncertainty in estimates of CH_4 emissions from enteric fermentation in 2020 was between -3Mt and $+5\text{Mt}$ CO_2 eq. [127]. It is not unusual for the results to have uncertainties up to 20% of the mean value from carbon modelling [121], [128]. Due to the various inputs and the complexities of the carbon models, uncertainties are bound to result. The aim of the carbon analyst is to identify sources of error or uncertainty, quantify them, and minimise them while studying the carbon dynamics. This paper examines uncertainties in estimating the potential carbon and GHG emissions, as well as carbon sequestration potential, when forested land is converted to agricultural land.

One of the methods to calculate uncertainty in carbon models is to select several parameters related to the factor one is investigating, such as soil organic carbon. The parameters are perturbed within a range of variation and their effect on the factor can be estimated. White et al. [128] also

used the Gaussian Emulation Machine for Sensitivity Analysis (GEM-SA) software [129] to calculate the uncertainty of dead organic matter. GEM-SA permits the user to build an emulator based on Bayesian statistics. GEM-SA proposes to perform uncertainty and sensitivity analyses with fewer code runs than Monte Carlo-based methods [129]. Metsaranta et al. calculated uncertainty by ratioing individual uncertainties to the total uncertainty: running the model over the entire study area of 2.3×10^6 sq. km five times, while varying one factor each time. Five factors were being assessed for uncertainty in that study [121].

2.6 Creating Framework for Predicting Carbon

Two models were eventually selected to conduct the simulations – DNDC and CBM-CFS3 – as they were suited to the goals of this work. The DNDC model estimated the carbon stock and GHG emissions on crop and pastureland, while the CBM was used to account for the carbon amounts in the forested regions of the study area. The model simulated the carbon and GHG amounts that would have accumulated without land use or landcover changes. Thus, the CBM results were the baseline for the study – to be compared to the results from the DNDC models for areas which had been converted to crop, livestock, or pasture lands.

One of the ways to reduce model uncertainty is to select models which have been well-validated and shown to be robust under a variety of situations, such as the models used in this study. It is crucial that the model correctly simulates the processes in the ecosystem. So, in this study, different models were explored regarding suitability for the required task. The carbon models rely greatly on the inputs given by the user. Thus, as this study demonstrates, the simulations should be performed with the best or most reliable input data.

A limitation of the models used in this study is their spatial response: both the agricultural and forest carbon models are aspatial. As such, the CBM model, for instance, can represent a stand of trees (or a group of stands), but it does not recognise the spatial relationships of stands within the spatial unit (a subdivision of the study area). Therefore, I was constrained to only looking at plots individually. In future work, I will use the new version of the CBM, which is spatially explicit. Then, a map of the emissions and carbon stock can be created. Additionally, CBM cannot

accommodate stands with different age classes. Another limitation of the DNDC and CBM carbon models is that they cannot directly predict the impact of climate on disturbance regimes. Thus, the model user would need to include the changes in disturbance regimes when setting up the simulation, for example, when using the models to predict future events. Although these two models were chosen for the study, other process models could have been selected in their place, e.g. from those listed in Table 2.1. Selecting a model representative of the study is vital to achieve more relevant results.

Another perspective would be to explore the use of process-based models (which are also called mechanistic models) for carbon modelling. They are also useful in predicting future states of ecological systems. Some of the commonly used process-models for forest ecosystems include Roth-C (which is a spatially explicit SOC dynamics model), CENTURY, and BIOME-BGC. CANDY and DNDC are process-models for agro-ecosystems.

Initially, the CANDY model was used to simulate the carbon dynamics in converting forested areas to agricultural land. The model is usually suitable for the scenarios which were used in this study – crops grown interspersed with fallow periods, and livestock raised on pastureland. However, the model failed during some of the simulations. I contacted the model's creators, and after they had tried several times to solve the problem, they were unable to. They reproduced the problem, but finally concluded that I had reached the limit of the model [130]. Therefore, I switched to the DNDC model, which is also capable of simulating the circumstances in the crop and pasture sites. Anyone conducting carbon modelling would similarly perform tests to ensure the model chosen is compatible with the analysis needs the model's capabilities.

Having selected the carbon models which would be used to simulate the ecosystem carbon and GHG emissions over the study area, I examined the inputs to the models. The forest model requires parameters such as transition information following a disturbance, the forest inventory, and the climate conditions. Similarly, the DNDC model uses parameters such as climate conditions, tillage information, and land use changes to generate simulation results. Figure 2.3 highlights some of these model inputs. The elements recognised for improvements to the carbon model results and reduction in uncertainties were the landcover change and disturbance history of the study area. Thus arose the research objective to improve detection of landcover changes.

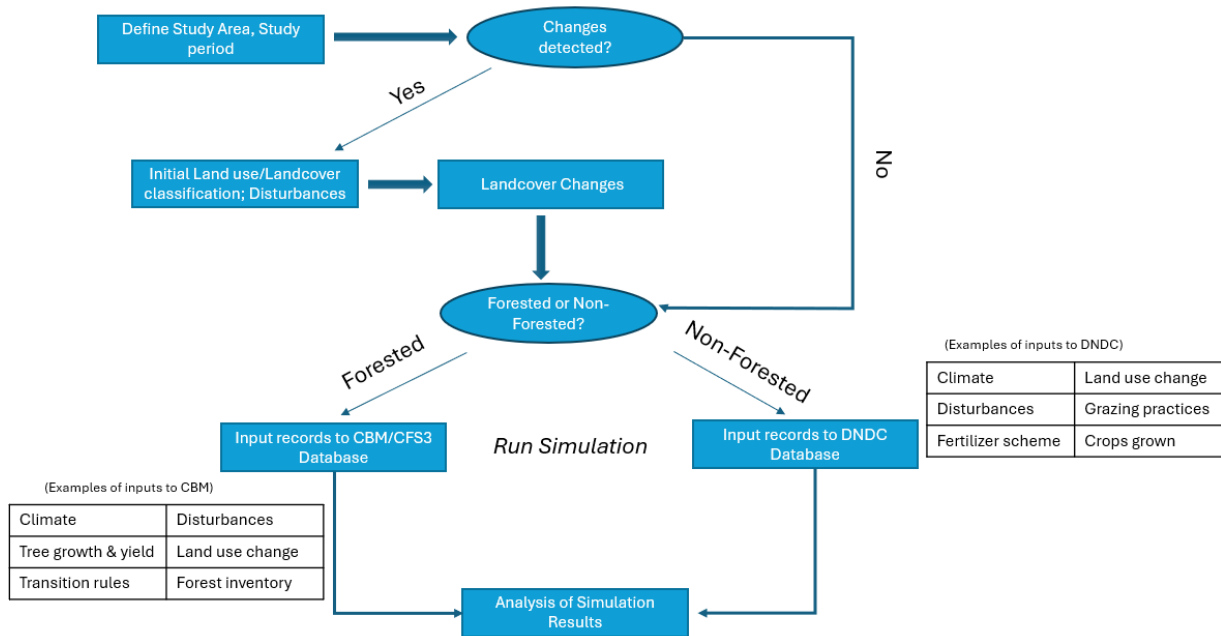


Figure 2.3: Workflow diagram highlighting the Progressive Adaptive Method (PAM) for carbon simulations

Figure 2.3 illustrates the workflow for the PAM method. The first step, having selected a region of interest, is to determine if there has been a change in landcover during the period of study. The determination of change is carried out using remote sensing data. Chapter 3 gives more details of how this first stage was conducted using data derived from satellite imagery. If there has been no change in landcover and the area is vegetated, it is then tested if the area is forested or not. However, if there has been a change in landcover type, the initial landcover type and the changed type are determined. In addition, the disturbances which might have occurred are obtained. It should be noted that following the branch of ‘non-forested’ areas only leads to examining the croplands and pasturelands. Other non-forested lands such as urban areas or water bodies are not considered in this work.

The specific landcover types are determined through a machine learning classifier. The landcover change detection is performed using a “from-to” analysis. Chapter 4 describes the algorithm for this second stage of the workflow. The final step is the simulation of carbon and GHG emissions over the study area. The estimates were only made over vegetated areas. If forested, the CBM model would be used. The DNDC model was proposed for all non-forested areas. Using the input parameters derived in the previous steps, such as land cover type, land use

change, and disturbance history, the simulations are run, and the resulting estimates can be analysed. Chapter 5 goes further in-depth on how the final stage of the workflow was enacted over the forests and agricultural lands in the Clay Belt study area and the types of results which could be generated from the carbon models.

3. A NOVEL APPROACH TO CHANGE DETECTION IN THE CLAY BELT

My objective for land use and landcover change analysis is to examine the impacts of climate change under different land management schemes. Of the two common approaches to change detection – change/no change, and from-to change detection – this chapter presents a proposed method utilising the change/no change approach. First, a background is given on the advanced remote sensing technology considered in this study. Customary change detection methods were then discussed. Thereafter, the automatic change detection process I enacted with my supervisor is presented. This novel method uses a hypothesis test as the basis for determining where land cover change has occurred. To my knowledge, many change analysis methods use thresholding, but a hypothesis-based test is scarcely used for such analyses. The method we developed was validated based on data collected in Ontario's Clay Belt.

A part of the research from this chapter is published in the journal paper:

Ituen, I & Hu, B. (2021) "An Automatic and Operational Method for Land Cover Change Detection Using Spatiotemporal Analysis of MODIS Data: A Northern Ontario (Canada) Case Study," *ISPRS International Journal of Geo-Information*, vol. 10.

3.1 Remote sensing systems and data

The multispectral remote sensors considered for landcover and land use change in the study are:

- i. The Moderate Resolution Imaging Spectroradiometer (MODIS). It captures data in 36 spectral bands and acquires data at three spatial resolutions: 250m, 500m, and 1km. (Bands 1 and 2 have 250m resolution, bands 3-7 have 500m, and bands 8-36 have 1km spatial resolution.) MODIS's revisit time of about twice a day is useful for monitoring dynamic land surface changes. With its moderate spatial resolution, models have been created for the fusion of MODIS with other higher-resolution imagery whose satellites may have a

lower temporal frequency. MODIS satellite data is commonly used for large scale long-term monitoring of the earth's environment. In agriculture and forestry, changes detected from such satellite imagery can be used to monitor crop growth, detect soil moisture, and monitor deforestation and forest fires.

- ii. Landsat: The Landsat 7 Enhanced Thermal Mapper (TM/ETM+) has 7 and 8 bands, while the Landsat 8 Operational Land Imager (OLI) sensor has 11 bands. Landsat data has been used extensively for global change research, landcover mapping, forestry, water and coastal studies, and resource management. The panchromatic bands on these sensors have a 15m resolution, while the visible and infrared bands have a 30m spatial resolution. Both sensors have thermal bands: for the ETM+, the spatial resolution of that band is 60m, while the two thermal bands of the OLI are at a 100m resolution. In this study, Tier 1 images have been used as best as possible, as opposed to Tier 2 images. The Landsat dataset is widely used in ecological applications because it has an extensive historical time-series, and a spatial resolution suitable for investigating land cover and land use change. Furthermore, its spectral coverage is appropriate for studying vegetation properties.
- iii. Sentinel-1: The Sentinel-1 satellite has a C-band Synthetic Aperture Radar (SAR) instrument on board. SAR is a radar (RAdio Detection And Ranging) tool; it is an active sensor and so transmits microwave radiation and receives the signal which is scattered back to the sensor. That is unlike the passive optical sensors – they receive the radiated reflection from the sun in certain parts of the electromagnetic spectrum. SAR is advantageous because it is not affected by cloud cover or the absence of the sun's illumination: it can acquire data at night as well as day. The Sentinel-1 SAR sensor is insensitive to the atmosphere, except for very dense rain cells [131]. Hence, the Sentinel-1 is chosen as a data source which is reliable and repeatable for earth monitoring under just about all weather conditions. This satellite also has a revisit time of 12 days [132]. A drawback of the SAR microwave radiation is that it produces speckle noise. The noise causes a 'salt and pepper' appearance in areas where one would expect a constant backscattering response. Although it can be reduced, speckle is difficult to eliminate [131].
- iv. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): The data from ASTER is used to create detailed maps of land surface temperature, emissivity maps,

elevation models, and change detection. The design of the ASTER has a unique feature in that Band 4 has the same spectral resolution as Band 3, except it is backward-looking, thus giving a second view of the study area. This multispectral radiometer has five thermal bands of 90m spatial resolution, 6 SWIR bands of 30m resolution, and 4 VNIR bands, which are all of 15m spatial resolution.

- v. Sentinel-2: This satellite carries the MultiSpectral Imager (MSI), which is a high-resolution imager. It is useful for landcover maps, land change detection mapping and gathering geophysical variables. The MSI is an optical sensor with 13 spectral bands: 4 in the visible range with a 10m resolution; 6 NIR bands with 20m spatial resolution, and 3 SWIR bands of 60m spatial resolution. Furthermore, three of the sensor's NIR bands are red-edge bands which are useful for vegetation discrimination as they capture the strong reflectance of vegetation. Its high resolution imagery is useful in classification efforts, though its surface reflection data is only available over the study area from 2017. Similar to Sentinel-1, the Sentinel-2 mission has a constellation of two polar-orbiting satellites. Sentinel-2 has a high revisit time over the study area (5 days from the two satellites under cloud-free conditions) [133] and is thus very useful in monitoring vegetation changes, particularly during the growing season.
- vi. PlanetScope: Launched off the Planet satellites, the PlanetScope sensors have a resolution of 3.7m. This is the highest resolution data which is freely available from 2009. The Planet sensor acquired images from four multispectral bands (red, green, blue, and near-infrared) on the RapidEye satellite until the RapidEye constellation was deactivated in March 2020. Thereafter, the Dove satellites carried the PlanetScope sensors and have added four bands – coastal blue, green I, yellow, and red edge [134], [135]. Planet satellites have a daily revisit time. In this study, the data was used to validate how or when an area had changed, e.g., conversion from forested to pasture, etc. For the study area, the ~3m resolution Planet data was not available until 2014. However, the 5m and 6.5m Rapid Eye was available from 2010.
- vii. RADARSAT: RADARSAT is Canada's earth observation satellite. It uses a radar sensor; it is an active remote sensing system that uses microwave energy. A vital advantage of this sensor is that it permits observations in all weather conditions. The RADARSAT mission was initially focussed on maritime security – the mission was established from the need to

monitor the icy waters of Canada [136]. The sensor's purpose has since expanded. RADARSAT-2 was launched in 2007, with a continued view to provide radar image data for coastal and marine surveillance, and security and foreign policy [137]. RADARSAT has a very high revisit time of up to four times a day in Canada's north [136]. Thus, this sensor is very useful for disaster management, monitoring land use changes, monitoring the effects of climate change, and coastal changes. RADARSAT is also widely used to monitor forest and cropland dynamics using textural analysis. The RADARSAT has a very high resolution: the C-Band wavelength is about 5.6cm; the L-Band's is 23.5cm.

A summary of the satellite sensors which were explored in this work is shown in Table 3.1 below.

Table 3.1: Summary table of sensors explored

Sensor Name	Satellite Onboard	Spatial Resolution	Acquisition Period
MODIS	Terra	250m, 500m, 1km	2000 to present
ASTER	Terra	15m to 90m	2000 to present
Sentinel-1 SAR	Sentinel-1	10m	2014 to present
MultiSpectral Imager	Sentinel-2	10m to 60m	2015 to present
Enhanced Thermal Mapper+	Landsat 7	15m to 60m	1999 to 2024
Operational Land Imager	Landsat 8	15m to 100m	2013 to present
RADARSAT SAR	RADARSAT	1m to 100m	1995 to present
PlanetScope	RapidEye, Dove	3.7m	2009 to present

3.2 Land change detection

Change detection is identifying the bitemporal change in an area from the remotely sensed images. The changes in land use and landcover are actively tracked by organisations worldwide because the effects thereof have ramifications in human well-being and environmental areas, such as impacts on biodiversity and water balance [138]–[140]. Change detection is sometimes considered as either a binary ‘change/no-change’ situation, or a “from-to” scenario, where the exact land use or landcover type is initially known as well as its final state. Either of the change detection types is used based on the application. For instance, an analyst might want to know if an

area was affected by a flood or included in an urban development project, so the change/no-change analysis might be conducted. However, if specific information is required on how the area has changed, e.g., if it went from forested to built-up or forested to bare earth, then the from-to analysis is required.

To understand geographical trends and processes that have occurred over a land area, the from-to change analysis can be conducted on areas identified as changed. This form of change analysis requires the landcover to be broken into information classes, i.e. the types of landcover that need to be identified. On a digital image, these information classes have a distinct spectral signature. Thus, an algorithm can distinguish the chosen classes from the signatures they present. The objective would be to associate an information class with the spectral signatures. Several representative areas are selected for each information class. Usually, homogenous groups of pixels selected across the image identify the class; they are termed training samples. An algorithm could then use the samples to predict what cover class is on the entire image.

3.2.1 Machine-/Deep-Learning Change Detection

Machine learning algorithms are often used to detect landcover changes. The change detection is done by classifying landcover change events over two or more dates using remote sensing variables. These machine learning methods can be termed ‘stacked classification’, ‘multidate classification,’ or ‘compound classification’ [141]. A benefit of deep learning is finding non-linear data patterns with little or no prior assumptions. Deep learning is being used for change detection purposes, with one of the methods being the Long Short-Term Memory (LSTM) neural network. It will learn the time series over a long period, being able to predict new data from historical time series [142]. The LSTM is an online detection method: it detects anomalies with minimum delay. Thus, it is a real-time or near-real-time method [142].

The deep learning sequence is to train, predict, and then detect changes. To remove noise or cloud effects from remotely sensed data, a threshold can be set for the maximum cloud cover or noise permitted. Otherwise, if it goes beyond the threshold, the data is discarded as unusable. Setting these parameters sometimes means that the remote sensing data has gaps. An advantage of the LSTM is that it can learn from training time series with missing data [142].

Machine learning techniques have made great strides in detecting landcover changes. However, they generally suffer from some problems due to the unique nature of remote sensing data, as well as the landscape being surveyed. For instance, one of the challenges facing machine learning for landcover change detection is the diverse nature of the earth: going from one place on the map to another, one would see the variation in vegetation types and climate conditions. So, where a savannah forest can be classified in one place, another forest categorised as a savannah forest in another part of the world can behave differently due to the seasonal cycles in that part of the world. This means the heterogenous nature of the remote sensing variables would impact the development and learning of a uniform (for savannah) change detection model for all locations.

Another way to view how heterogeneity affects machine learning modelling is the varied nature of the remote sensing image pixels: [141] examined the response of mapping MODIS 500m data from a forest fire in California. Within each 500m were a variety of vegetation types. Since vegetation behaves differently having been burned, a spatial heterogeneity was observed. Thus, training a classifier to detect burned or unburned regions becomes difficult. In that study, the classifier had to be trained with widely varying properties across the entire area being studied. Therefore, different models need to be generated for each time and region whenever data shows a different characteristic [141].

Another difficulty encountered by machine learning in change detection is the use of datasets from various sources. It is not unusual to use data from multiple sensors to generate a holistic view of landcover change. This allows the analysis of the area's change at different temporal and spatial scales. The problem of having multiple data sources at different resolutions is that the machine learning algorithms are designed to function with datasets at a single scale of observation [141].

A hindrance faced by not only machine learning methods is the scarcity of training data. When there is insufficient ground-truth data, the supervised learning method requires the detailed learning of the classifier, and to guard against overfitting. Some solutions that have been proffered for machine learning when there is little training data are to have semi-supervised learning and active learning approaches [141].

Advances in machine learning are set to address such problems as highlighted above. E.g., Multitask Learning can address the heterogeneity and scarcity of training data; Multi-Instance

Learning could also address the scarcity of training data; and Multiview Learning would counter the challenges associated with having data coming from multiple sources [141].

Deep learning is a subfield of machine learning. Deep learning automates the ability to perform representation from raw data, whereas machine learning algorithms would need humans to extract features with which to work. Deep learning is attractive as it removes the costly and sometimes inefficient element of humans from the process. The automated process enables the analysis of large amounts of data. Specific to landcover/land use classification, deep learning seeks to advance beyond machine learning as more information is received; deep learning algorithms build artificial neural networks which learn and make classification decisions on their own [143]. These algorithms analyse data and draw mapping conclusions like a human would.

The most common deep learning algorithms used in landcover and land use mapping are the convolutional neural networks and the recurrent neural networks. Convolutional neural networks are able to extract low-level features like angles and outlines of objects on a landscape, regardless of its shape or size. Similarly, recurrent neural networks' ability to "remember" past data points permits the networks to understand the present and even predict future events. As such, these networks are proficient at handling time series data and are useful for change analysis [144]. Thus, these networks' algorithms find great applicability in landcover mapping and change detection.

Whereas feature extraction and classification are performed as two separate steps and by the human in traditional machine learning, in deep learning, the feature extraction and classification are performed as an integrated process. The classification algorithm predicts each class's accuracy via the neural network, and that without human help [144]. Deep learning has been shown to surpass machine learning in landcover classification [145]. Some of the overall accuracies these models have been able to attain include 99.04% with the EuroSAT dataset [146], 95.95% [147] and 99.81% [148], both using the University of California Merced dataset, 71% from Sentinel-2 data [145], and between 92.46 and 99.64 using various architectures [149].

The process proposed in this study was designed to be repeatable and easily deployable by a wide range of users. For instance, the algorithm for this study is provided upon request as a freely available code; it can be used immediately as long as a study area is defined, the period of interest, and training/validation points are selected. The user can also choose which features to use for

classification, e.g., removing some of the spectral bands proposed in the code or adding more indices. As an example of its functionality, the code executes within a few minutes on the extensive study area of about 120,000 sq. km. Comparing this machine learning method to deep learning methods, the latter require an enormous amount of training data. Many times, labelled datasets (such as ESRI, UCM, EuroSAT, and LandSAT) are used because of the extensive manual effort required to extract the training datasets and generate the image tile required as input to the deep learning models. As an example, EuroSAT is based on cities in over 30 European countries [150], and the dataset might not properly reflect the landscape of Congo or the deserts in Egypt for land classification. When labelled data specific to a region is unavailable, it takes a lot of manual work to generate.

To properly conduct landcover/land use mapping, the deep learning methods require much more data than the traditional machine learning methods. The deep learning's multi-layer structure necessitates the large training dataset "to eliminate fluctuations and make high-quality class interpretations" [144]. If a large training dataset is not used, deep learning could have the same or even worse performance than a Support Vector Machine method [151]. In addition, the computing power demanded by deep learning methods is significantly more than that required by traditional machine learning methods. It is not unusual for neural networks to need a GPU (which has thousands of processor cores, compared to CPUs) to be trained. Furthermore, deep neural networks can take several hours to train, as an increased number of epochs usually increases accuracy; for instance, running 100 epochs can take 10 hours to train the model.

The deep learning user also needs to define a loss function which measures the cost incurred from incorrect predictions by the neural network. It is very difficult to train neural networks, and loss functions can be challenging to optimize. Another parameter the designer of the network sets is the learning rate. There are significant consequences in setting the learning rate; if it is set too low, the learning is slow to converge – or maybe might not converge at all (e.g., being stuck at a false minima instead of the global minimum). Also, if the learning rate is set too high, the process might start to diverge from the solution. Adaptive learning can help with setting this learning rate rather than the iterative process of trying several rates until an optimal value is found. An adaptive learning algorithm would adapt to the network and find a suitable learning rate.

It is because of the ease of use that this work's methodology is being proposed, using a

Random Forest classifier. Since the purpose is to generate better information as input to carbon models, I aimed to make this carbon simulation intermediate step less demanding and far less complex than would be encountered in deep learning while maintaining a high level of accuracy in landcover/land use classification. A Random Decision Forest is a method that operates by constructing multiple decision trees during training phase. The decision of the majority of the trees is chosen by the algorithm as the final decision. All the trees together make up what is called a Random Forest. So, from the forest, a vote is taken and the majority decision is what will represent the forest. Random Forest uses classifiers like bagging, which averages many low-bias and high-variance predictors, and so can have high accuracy without overfitting. the benefits of a Random Forest classifier are it uses multiple trees to prevent overfitting of data.

In the first part of the change detection work, as well as in the second part (described in Chapter 4), the commitment to an easily operational method was kept. For example, a Random Forest change detection method was followed as the processing takes a much shorter time and provides very high accuracy. So comparatively for ease of use, the Random Forest method was chosen. If set up in a deep learning environment, it might dissuade many potential users from ever attempting to use my methodology which could decrease the uncertainty associated with the estimation of carbon in the soil or atmosphere and GHG emissions. It is crucial that the method developed is operational and accessible to as many people as possible without requiring extensive training prior.

3.2.2 Other Novel Change Detection Methods

The Continuous Change Detection and Classification (CCDC) algorithm processes a set of images from different dates together to identify landcover change. The algorithm starts with three consecutive images free of snow, cloud, or cloud shadow (cloud cover <80%) to ascertain if a pixel has changed or not [12]. When a pixel has changed over multiple images, it is likely landcover change has occurred. Of a set of three images, if pixels show change for one or two consecutive times, it is labelled as a “possible change.” If the pixel also changes in the third consecutive image, it is labelled and assigned to the “change” class.

It is important to note that this algorithm tests the change across all 7 of the Landsat bands. A threshold for determining change is also set as $3 \times \text{RMSE}$ of the surface reflectance for each band. If the observed data point is outside this threshold ($-3 \times$ to $+3 \times \text{RMSE}$), it is taken that a landcover change has occurred.

The COntinuous monitoring of Land Disturbance (COLD) algorithm is a method built off the CCDC algorithm [152]. In this method, change occurs when six consecutive images have anomaly observations. It has been noted that COLD will miss grassland fires or floods which do not present for six consecutive images [152]. This method uses normalized change vector magnitude to determine the change in landcover/land use.

3.2.3 Fusing Multiple Image Sources

Sometimes, remote sensing images are fused to provide more reliable results or a better interpretation of the results than from a single sensor. The fused image combines two or more images of the same scene into a single image.

Image fusion has been done with various sensors, e.g., multispectral sensors, multispectral sensors with panchromatic (images acquired in almost the same spectral range), and even panchromatic and hyperspectral sensors. However, the spectral range of a hyperspectral imager is much wider than a corresponding panchromatic imager [153]. An example of the benefit of fusing image data from multiple sensors can be seen with Landsat and MODIS:

Landsat sensors offer higher resolution imagery than the MODIS and so the data is very useful for landcover mapping. However, Landsat data is only acquired every 16 days, while MODIS acquires data on a much higher temporal scale. Thus, MODIS would be used at broader spatial scales to observe dynamic changes in land surfaces, while Landsat can be used to see finer details of the land surface when the image is available. Many models have been created to fuse such data as Landsat, which has a low temporal frequency but higher spatial resolution, with MODIS, the higher temporal frequency sensor, which has a moderate resolution. Many of the models exist for fusing data from different sensors, e.g., for MODIS and SAR data [154], MODIS and Sentinel-2 [155], and Landsat and ASTER data for vegetative indices' comparison [156].

3.3 Statistical Method of Change Detection

This section describes the development of the statistical change detection method we proposed. As illustrated in the methodology diagram in Figure 2.3, an initial determination is made of whether an area has experienced change or no change. The remotely sensed data (Landsat Normalised Difference Vegetation Index (NDVI) and MODIS Land Surface Temperature (LST) images) employed in this first part of the study were taken from Google Earth Engine. The Google Earth Engine is a cloud-based platform for geospatial algorithm development and processing from which a user can access hundreds of public earth science datasets. The publicly available satellite imagery from all around the globe can be manipulated for a variety of purposes. The datasets on the platform include raw images such as at-sensor radiance and top-of-atmosphere reflectance images, as well as derived datasets such as Enhanced Vegetation Index (EVI), Normalized Burn Ratio (NBR), and the Normalized Difference Water Index (NDWI) images. Google Earth Engine's vast cloud computing resources permit large datasets to be analysed rapidly.

The MODIS Terra data products used for the change analysis were the 16-day 250m resolution product for NDVI (MCD13Q1) [157], the 500m resolution surface reflectance (MOD09A1.006), and the 8-day 1km resolution product for LST (MOD11A2) [158]. Linear interpolation was performed for missing data, interpolating both temporally and spatially. An Image Collection with a total of 23 images was used for the months' NDVI maps, and for the LST, an Image Collection of 4 images was used for each month. The data in Google Earth Engine has typically undergone preprocessing. The analysis was performed with the Google Earth Engine code editor, ArcMap 10.7, and MATLAB.

Normalised Difference Vegetation Index (NDVI) is a well-used vegetation index because of its success at detecting vegetation, the ease with which it is calculated and interpreted, and its accessibility from various satellite data. The NDVI index is known to be the primary vegetation index for global operational applications [159]. NDVI was used to monitor change on an intra-annual basis and an interannual basis. As to the other feature considered in this work (the LST), studies have shown the impact landcover and land use changes have on LST [160]–[162]. Depending on the type of landcover present, the surface temperature in an area could vary [163]. Similarly, when the land undergoes a change in landcover or land use type, it could experience a simultaneous change in LST. Majumder et al. [164] noted that LST could decrease in an area which

increases in agriculture. Zhang and Liang [165] used China's land use/cover datasets (CLUDs) to study the impact of landcover change on surface temperature. The database contains images of 1km raster size having 25 landcover classes; these classes are further grouped into cropland, grassland, woodland, unused land, water, and urban areas. The Zhang and Liang [165] study showed that landcover change from cropland to built-up areas increased the LST, while a change from forest to cropland reduced surface temperature.

It is acknowledged that many other features could have been used for the change analysis. At a large scale and for change in broad land cover/land use, the seasonal changes in the cover types are the key for operational change detection. NDVI was chosen because of its historical success as a vegetation monitoring measure. The developed method should work for other products, such as Enhanced Vegetation Index (EVI), Leaf Area Index (LAI), and Fraction of Photosynthetically Active Radiation (FPAR), which will be explored in future work.

Image preprocessing was carried out to remove unwanted artefacts, such as those created by atmospheric interferences. The MODIS Surface Reflectance product at 500m resolution was estimated to be the surface reflectance at ground level. It was corrected for atmospheric conditions such as gasses, aerosols, and Rayleigh scattering. Similarly, the MODIS NDVI data from Google Earth Engine consists of atmospherically corrected bi-directional surface reflectance, which has been masked for water, clouds, heavy aerosols, and cloud shadows. The LST Level 3 products used in the simulations had cloud-contaminated LSTs removed already. Another element of image preprocessing was the smoothing of input data to remove noise. In this work, I used a Savitzky-Golay filter as the smoothing technique. Other smoothing methods that were considered were the Best Index Slope Extraction, logistic and double logistic smoothing, asymmetric Gaussian, and mean- and median-moving average. The premiere method did not change or smooth any of the NDVI data it was applied to, so was not used further. Of the latter two methods, the moving median gave a better representation of my data than did the moving mean. The rest of the filtering methods tested also produced a reasonable degree of smoothing to the input data. Yet, of all these methods, the Savitzky-Golay gave the best result – using the greatest number of data points while maintaining the general trend of the plots. A window size of 5 was chosen as the proposed approach obtains its optimal performance with that window size. Figure 3.1 is an example of the effect of

smoothing on the NDVI data in 2010. The rest of this section describes the approach taken to create my landcover change detection method and the procedure used to assess its accuracy.

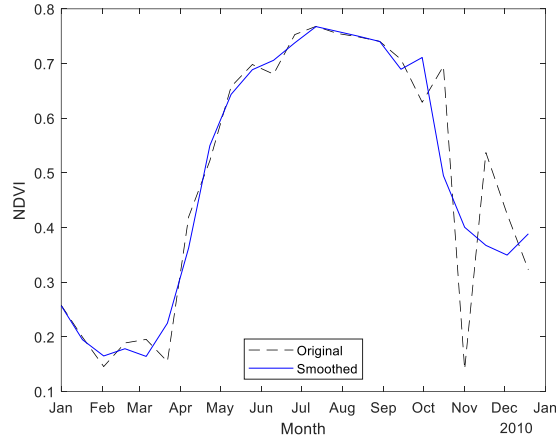


Figure 3.1: Example of smoothing done on NDVI data

A Principal Component Analysis (PCA) [166] was performed on a 2001 image of the surface reflectance with the first 7 MODIS bands. The PCA showed that the first three components explained more than 98% of the data. The rest of the computations were then conducted on the first three principal components of MODIS imagery covering the optical bands. The statistical test was performed to determine if significant change had occurred, utilising the NDVI and LST maps as input.

Hotelling's T^2 test [167] was selected in this method as the T^2 test can compare multivariate data. So, where univariate data can be compared using a t-test, multivariate data can be compared from the T^2 test. The T^2 test, like the t-test, compares samples using means. In Hotelling's case, the multivariate mean is used because it compares the means of two or more samples. Thus, the T^2 statistic can be thought of as comparing a point in space defined by the means of all the individual variables.

When $X_1, \dots, X_N$ are k -component vectors, and the normal distribution has a vector mean μ and a covariance matrix Σ with the sample mean defined as

$$\bar{x} = \left(\frac{1}{N}\right) \sum_{i=1}^N X_i \quad (3.1)$$

and the sample covariance matrix as

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})(X_i - \bar{x})' \quad (3.2)$$

then the generalised T^2 statistic is given as

$$T^2 = (\bar{x} - \mu_0)' \left(\frac{S}{n} \right)^{-1} (\bar{x} - \mu_0) \quad (3.3)$$

where μ_0 is a known vector, and there are $n-1$ degrees of freedom,

and the paired Hotelling's T^2 test statistic is given by the expression:

$$T^2 = n\bar{x}'S^{-1}\bar{x} \quad (3.4)$$

The T^2 test is more advantageous than a t-test because the Type I error rate is better controlled in the former than in the latter. Another benefit of the T^2 test is that it considers the relationship between the different variables; thereby, the differences between groups can be obtained. The assumptions for this statistic are that the samples have normal distributions, they are independent, and if there are only two samples, then they have equal variance-covariance matrices.

The test was done to confirm if change had occurred in an area. Thus, the Null Hypothesis statement is: *No landcover change has occurred in these areas in a five-year period.* The hypothesis was tested by using the difference in mean values of the parameters under study within the five years. The null hypothesis would be rejected for the alternative hypothesis if the sample data used for the study showed the null hypothesis to be false. That is, if change had occurred, the null hypothesis would be rejected. The null hypothesis implies that the mean LST and the mean NDVI remained the same. That is, their population mean vectors are equal.

An extension of the proposed change detection method was explored whereby the NDVI and LST data of each year were combined into one matrix and their landcover change was predicted using the algorithm described here. That entailed normalising the NDVI and LST data before making the comparison between the years. (The method is hereafter referred to as the joint method.)

3.4 Results

Following the statistical tests, the program outputs an image of the study area, with 0 representing a no-change area and 1 representing a changed superpixel. The areas where there is a zero value were made transparent, and a changed area has a grey colour. The results in the form of binary maps are shown in Figures 3.2-3.4 below.



Figure 3.2: Changed areas (coloured) detected from 2001 to 2005 in (a) NDVI; (b) LST overlaid on the map of the study area

Figure 3.2 illustrates the differences in areas where the LST test detected change and where the NDVI test detected change between 2001 and 2005. Overall, the LST detected more areas of significant change than the NDVI test. The most noticeable differences are observed in the north central part of Hearst and in the mid-eastern part: The NDVI test returned less significant change in landcover than the LST in those areas. The parts which were more similar are in the western and central part of the town.



Figure 3.3: Changed areas (coloured) detected from 2005 to 2009 in (a) NDVI; (b) LST overlaid on the map of the study area

In Figure 3.3, the landcover change detected between the LST and NDVI tests is shown as more similar in the west and central portions of the town. The change detection results were dissimilar mainly in the north-central part of the town, where the NDVI test predicted mostly no significant change in that area, whereas the LST predicted significant change had occurred in that area between 2005 and 2009.

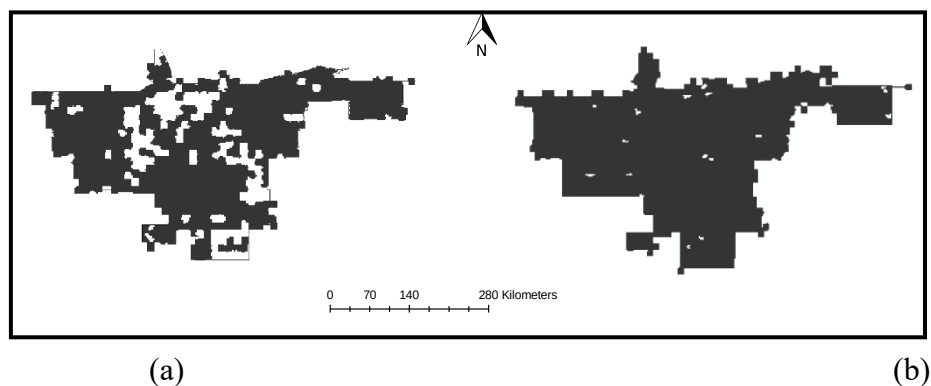


Figure 3.4 Changed areas (coloured) detected from 2010 to 2016 in (a) NDVI; (b) LST overlaid on the map of the study area

There was good agreement in the change detection results between the NDVI and LST for the period 2010 to 2016 (Figure 3.4). Apart from the north central part of the town, and an area in the central west part of the town, the change detection maps produced from NDVI and the LST indicate similar landcover change occurring. In the areas of dissimilarity noted, the NDVI indicated no significant change, while the LST had predicted a significant change in landcover had occurred.

An accuracy assessment was conducted on each of the NDVI and LST change maps produced, comparing areas of change/no-change with the high-resolution maps – maps of the area at 30m resolution, which is the Landsat-level resolution – for the years under study. Tables 3.2-3.4 below give examples of the results achieved. The accuracies obtained varied from 74% to 88%. For instance, the overall accuracy of the change predicted by NDVI between 2001 and 2005 was 78%, while the LST had a 74% accuracy. Similarly, the accuracy of the LST change detected was 84% for the period 2005-2009 (see Table 3.3). The average overall accuracy of the change detection process was 81%.

Table 3.2: Confusion matrix representing change detection accuracy of NDVI-based test: 2005-2009

	<i>Reference Data</i>			
<i>Classified Data</i>	Changed	Unchanged	Totals	User's Accuracy (%)
Changed	21	4	25	84
Unchanged	8	17	25	68
Total	29	21		
Producer's Accuracy (%)	72	81	Overall Accuracy: 76%	

The overall accuracy of the NDVI change map was 76%, with a *kappa coefficient* of 0.52 in the period 2005 to 2009. The user's accuracy for changed areas was 84%, while the user's accuracy for the unchanged areas was 68%. Similarly, the producer's accuracy was 72% for the changed areas because eight areas which had undergone landcover change were incorrectly classified as being unchanged. The producer's accuracy was better for the unchanged areas as only 4 of the unchanged areas were predicted by the statistical test to be changed (see Table 3.2).

Table 3.3 below provides the accuracy results from the same timeframe, when LST was the feature on which the change assessment was made.

Table 3.3: Confusion matrix representing change detection accuracy of LST-based test: 2005-2009

	<i>Reference Data</i>			
<i>Classified Data</i>	Changed	Unchanged	Totals	User's Accuracy (%)
Changed	24	3	27	89
Unchanged	5	18	23	78
Total	29	21		
Producer's Accuracy (%)	83	86	Overall Accuracy: 84%	

The LST change map for the change between 2005 and 2009 yielded an overall accuracy of 84%. Like the NDVI test for that period, more errors of omission were found in the changed areas than errors of commission. However, for the unchanged areas, most of those chosen for accuracy assessment were correctly classified as unchanged.

Table 3.4 displays another example of the results from the change detection test. When NDVI was used to predict landcover change, the overall accuracy was 88% with a *kappa coefficient* of 0.76.

Table 3.4: Confusion matrix representing change detection accuracy of NDVI-based test: 2010-2016

	<i>Reference Data</i>			
	Changed	Unchanged	Totals	User's Accuracy (%)
Changed	21	5	26	85
Unchanged	1	23	24	96
Total	22	28		
Producer's Accuracy (%)	96	85	Overall Accuracy: 88%	

In the detection period 2010-2016, there was less confusion in the NDVI change maps pertaining to the changed areas. Only one area was incorrectly classified as unchanged (see Table 3.4). This resulted in a high producer's accuracy for the changed areas.

When the joint method of NDVI and LST together were used to predict landcover change, the accuracies were in the same range as the standard (i.e. proposed) method using individual parameters. The average overall accuracy using this approach was 80%. Table 3.5 below illustrates an example of the landcover change prediction results from the joint method.

Table 3.5 Accuracy results from NDVI- & LST-based (joint) test: 2010-2016

	<i>Reference Data</i>			
<i>Classified Data</i>	Changed	Unchanged	Totals	User's Accuracy (%)
Changed	20	2	22	91
Unchanged	5	23	28	82
Total	25	25		
Producer's Accuracy (%)	80	92	Overall Accuracy: 86%	

The resulting map from the joint method had a lower producer's accuracy than user's accuracy of the changed areas, signifying a higher omission error than commission error (Table 3.5). The lower producer's accuracy was evident in the changed category because changed superpixels were identified as unchanged, while only two superpixels which were unchanged were incorrectly classified as changed.

Comparing the joint method's results to the NDVI change for the same period, the joint method misclassified 2 of the changed areas and 5 of the unchanged areas, while the NDVI method misclassified 5 and 1 of the changed and unchanged areas respectively. In this case, the joint

method had a lower commission error. The overall accuracy from this joint method was 86% for the period 2010-2016.

Uncertainties in fusion were not considered in this work which operated on feature-level fusion. In addition, a drawback of feature-level fusion is that because all the features are usually treated equally, the features that might need more weight in identifying a cover class are not made prominent and so do not contribute to the classification optimally [168]. A higher processing level is the decision-level fusion. In decision fusion, information is merged from the different data sources (the individual classifiers) to make a final decision based on the decision rules. In the future, I will explore if decision-level fusion is a better approach to employ.

As a quick check to confirm the changes in vegetated areas in Hearst, the Hansen et al. [169] method of detecting forest loss and forest gain was used to estimate the yearly forest area changes over the study period. It was seen that there was approximately 18,000,000m² loss between 2006 and 2007, and also between 2008 and 2009 there was 8,000,000m² loss; in addition, a 26,000,000m² forest loss occurred between 2010 and 2011, and approximately 12,000,000m² forest gain was estimated in the town between 2014 and 2015. These figures again confirm the expectation of observing significant change from the statistical test method. Furthermore, I created an application which draws on the Hansen et al. dataset [169] that shows the maps of forest loss and forest gains, as well as has charts to view the changes in forest volumes over the Clay Belt, Hearst, and Cochrane from 2001 to 2023. I called this app the [Forest Change App](#), found at this link: [Forest Change App](#)

Another application I created is called [The Landcover App](#). Users can choose any year from 2000 to 2023 to observe their region of interest. The area is classified based on the MODIS IGBP data of 500m resolution, and contains 17 landcover classes. The scope of the map is Canada, the 291 regions are as defined by the administrative units by the UN FAO. Again, it can be accessed from here: [Landcover App](#)

SUMMARY

The method proposed for determining the landcover change that took place in a specified time period uses an algorithm I developed. It applies the Hotelling T^2 statistical test to determine when a significant difference between two sets of data has occurred. An analyst only needs to define the required significance level of the T^2 test. This is easier for a user and would provide more consistent results than existing methods of change detection, which necessitate choosing a threshold for determining change, where that choice is made by trial-and-error. Image processing techniques of Principal Component Analysis and image segmentation were performed on the input data.

Using the NDVI and LST parameters of MODIS data, I determined where significant landcover change had occurred in the study areas. The contribution of this study was to find an operational yet automatic method to determine where change in landcover has occurred.

Additionally, by utilising NDVI and LST, which are readily available data from satellite imagery, this method can automatically detect landcover change without much other influence or input from the user. The platform used for processing the data, Google Earth Engine, is also easily accessible. Thus, the proposed method is highly operational and can find wide applicability in the monitoring of landcover change.

This study is also significant in that NDVI and LST are used together to predict landcover changes. To the best of my knowledge, such experiments are not widely seen in literature, though the features have been used individually for that purpose. The result from this study indicates that this approach being proposed is a valuable area to continue developing since its outcomes are similar to existing and established change detection methods.

4. MULTISOURCE LANDCOVER CLASSIFICATION

The previous chapter described a method developed during my doctoral research whereby an analyst could easily identify areas of change between two time periods. It is an automatic method to detect binary land cover change from a statistical test using the NDVI and LST parameters. Building on that ‘change/no-change’ method, this chapter introduces the second part of the work: in areas that have experienced landcover change, what was the area initially and what cover type was it changed to over the period of investigation?

The carbon models used in the study require the landcover and land use type to be defined, as well as their disturbance history. Therefore, rigorous “from-to” landcover classification and change detection needs to be done (see Figure 2.3). Hence, this chapter presents the remote sensing data used for the classification effort, as well as the steps taken for preprocessing. In striving for a more accurate estimation of carbon stock and GHG emissions, the entire suite of data sources examined will be described. Thereafter, the Random Forest classifier setup is presented, and the change detection results are also shown. The PAM method of change detection developed in this study is also explained.

Another aspect of this work is the section on segmentation: I explored using the GLCM for the landcover classification and compared the results of the object-based classification to the pixel-based classification. As mentioned earlier, this study is unique in that the data of multiple sensors was fused to generate landcover information for carbon models. Few studies consider multi-sensor image classification for estimating soil carbon flux [47], which could increase the accuracy of carbon estimates – since multi-sensor classification has been shown to increase the overall accuracy of classification [58], [170]. Most studies that use satellite imagery to classify land use prior to estimating soil carbon only use Landsat data [171]–[176]. This research was motivated by exploring if multi-sensor landcover classification and change detection can improve the carbon models’ accuracy over large geographical areas.

The change detection results derived from the PAM technique are displayed at the end of this chapter. Part of the research presented in this chapter was submitted in Feb. 2024 to the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing under the title *A multisource data approach for change and disturbance mapping of Ontario’s Clay Belt towards more accurate carbon and emissions estimation*. The article is currently under review.

4.1 Satellite Imagery Data Sources

The land classification and change detection data sources are the Landsat-7 and -8, and Sentinel-1 and -2 sensors. The optical data were corrected for atmospheric conditions such as gasses, clouds, heavy aerosols, cloud shadows, and Rayleigh scattering. If images are from different sensors (meaning they have different calibrations) and are not cross-calibrated, the results obtained will not be correct. For example, the Landsat-7 and -8 sensors are different (have slightly different signatures), so all the Landsat collections would need to be harmonised. Such harmonisation was required for Landsat Collection 1. In this study, Landsat Collection 2 was used for both Landsat 7 and Landsat 8. Hence the harmonisation step is not required: the sensor data has already been intercalibrated.

Similarly, for the Sentinel-1 data, Google Earth Engine (GEE) performs the preprocessing to derive the backscatter coefficient (in dB) in each pixel. In this work, an additional boxcar filter was added – the Refined Lee filter. The additional border noise reduction implemented here was a masking procedure to remove border artifacts.

4.1.1 Data preprocessing

The optical Landsat-8 Collection 2, Level 2 imagery has been orthorectified and atmospherically corrected. The cloud mask method was applied to generate cloud-free pixels during the period of interest (the summer months). The *qa_pixel* attribute is a pixel quality parameter from the C of Mask (CMASK) algorithm. The values were set to mask cloud and cloud shadows, while the *qa_radsat* band was set for no saturation of any of its data.

The optical Sentinel-2 imagery has been orthorectified and atmospherically corrected. The Sentinel-1 image collection in Google Earth Engine consists of Level-1 Ground Range Detected (GRD) scenes processed to backscatter coefficient. The preprocessing already conducted on the dataset includes radiometric calibration and terrain correction (orthorectification). Its preprocessing goes through these steps:

1. Orbit file application: Updates the orbit metadata with a restituted orbit file
2. GRD border noise removal: As of January 2018, this step removes low-intensity noise as well as invalid data on the scene edges

3. Thermal noise removal: This step “removes additive noise in sub-swaths to help reduce discontinuities between sub-swaths for scenes in multi-swath acquisition modes” [177]. The removal of thermal noise can only be applied to images from July 2015 onward.
4. Radiometric calibration: This computes the sensor backscatter intensity from the calibration parameters
5. Terrain correction (orthorectification): This correction converts the ground range geometry to backscatter coefficients for high latitudes. Terrain correction is required because the ground range geometry does not account for terrain, so the correction is applied for high latitudes (above 60° or below -60°). Products such as the ASTER DEM are used for this correction.

The Sentinel-1 imagery was also filtered according to the parameters desired. First, the acquisition mode chosen was the Interferometric Wide mode, which is for the primary acquisitions over land. The subsequent filter employed was used to filter according to polarisation – the VV or VH. In this study, both parameters were used: the single co-polarisation vertical transmission-vertical receiving (VV), as well as single co-polarisation vertical transmission and horizontal receiving (VH). In addition, the choice can also be made to select the ascending or descending pass of the satellite. Although the initial choice was for the descending pass, no results were rendered in the study region with that setting. Thus, the setting was changed to the ascending pass. It is also possible to select the metadata’s resolution – either 10, 25, or 40m. This study used the 10m setting. To reduce speckle on the images, the Refined Lee speckle filter was applied with a 3x3 square filter. (Other speckle filters such as the Lee filter, Gamma Maximum A-posterior, and Lee sigma filter were also explored. The additional border noise reduction implemented here was a masking procedure to remove border artifacts.)

Sometimes, after the data preprocessing, clouds were still present. Thus, a percentile composite was sought that would be closest to the representation of the area while eliminating the cloudy pixels. The percentile composite works in this fashion: given all the values of a band in an Image Collection for a pixel and sorting them, the 20th percentile, for instance, would be 20% along the cumulative density function of values for that pixel. I created the composite image of each tenth percentile, from 0 to 80%. This study found that the 30th percentile provided the best results, so that image was used going forward in the analysis. This composite proved better than

the others because it had the fewest cloudy pixels, whereas the higher percentile images have higher reflectances, like cloudy pixels which have higher reflectance values. Similarly, water, as well as the hills and cloud shadows, have lower reflectance and could still appear in the composite images on the lower percentile range [120]. With the preprocessing of the remote sensing data completed, the next steps for landcover classification can be conducted.

4.2 Data and Classifier Preparation

4.2.1 Data Sources for Land Use/Landcover Classification

A vital component of estimating carbon is identifying the landcover/land use type present. For forested regions, existing forest inventory information can be used as the source for the cover classification.

Ontario's T1 Forest Resource Inventory (FRI) for 2008-2017 provides the inventory for all the forests in the province [178]. A Random Forest (RF) classifier requires both training and testing samples. The FRI dataset was used to select training samples for forested areas. Likewise, the testing samples were taken from both the FRI dataset and the Land cover product of Natural Resources Canada (NRCan). NRCan has produced landcover maps every five years since 2010; however, the 2020 product used a different method than the previous ones. It has been recommended to avoid comparing it for change detection [179]. For this work, the 2020 map based on the Landsat-8 sensor was used to collect sample areas. Subsequently, LiDAR data (only available adjacent to the major highway of the Clay Belt) was used to refine some forest training areas since it has a higher spatial resolution. That was also the standard used if there was a discrepancy in cover type identified in an area between two maps: if there was no field survey data there, I would select the classification type which appeared most correct when compared to the higher resolution map. For instance, comparing the 282 training points obtained from the FRI data with the Agriculture and Agri-Food Canada landcover product, it was observed that 8 of the data points had a discrepancy. With both datasets being of 30m resolution, the LiDAR dataset could be referenced to ascertain the area's 'most likely' landcover type.

Now, the FRI contains information on tree species for the forests located in the Clay Belt. With this information, areas could be identified as coniferous-, deciduous-, or mixed-forests. The

threshold for any point or area was the percentage of a tree type within a pixel: if 80% or more had coniferous trees, it was categorised as part of the coniferous forest. If a pixel had mixed types of trees, and the total tree type was 60% or less, it would be categorised as a mixed forest. For example, Point #002 of the test points was 60% black spruce, 30% larch, and 10% white cedar; it was counted as coniferous. Point #065 was 60% white birch, 20% trembling aspen, 10% black spruce, and 10% balsam fir; it was classed as deciduous. And Point #214 was 40% balsam poplar, 20% trembling aspen, 20% black spruce, and 20% balsam fir; it was classed in the mixed forest pixels. The precise locations of the training points were obtained from ArcMap 10.6.

Agriculture and Agri-Food Canada (AAFC) produces annual crop inventory maps, so along with field survey points, this crop inventory data could be used to generate sample regions for the landcover classification. The AAFC crop inventory dataset was produced from Landsat and Sentinel-2 imagery and is provided at 30m resolution [180]. The dataset was used to select both training samples and testing sample areas. Now, this dataset does not cover the entire extent of the Clay Belt region; hence, another data source was required as a supplement.

Since AAFC's Land Use dataset is semi-decadal [181], sample areas could be taken from maps from 2015 or 2020. Again, the 2020 map was selected as a reference for this study. Although ground sampling was not completed in some provinces due to pandemic restrictions that year, Ontario was not affected, so the dataset did not offer a "general agriculture" class, as was done for British Columbia and some maritime provinces that year, but maintained the distinction in classes – such as types of crops, pasture, etc. Thus, the 2020 map serves the purpose of this study. Both UTM Zones 16 and 17 Land Use maps were needed to cover the entire Clay Belt area. The AAFC Land Use dataset has four cropland classes, two of which were used for this study – Annual Cropland and Cropland (annual and perennial cropland) classes. The others are the Land Converted to Cropland (i.e. land which was not cropland ten years prior) and Land Converted to Annual Cropland (i.e. land which did not appear to be annual cropland ten years prior) – [182]. The latter two were seen as being a subset of the previous two classes; hence, testing samples were not extracted from those classes.

AAFC Land Use has a class called "Grassland Managed", which I define as natural grass and shrubs used for cattle grazing. Thus, apart from the field survey pasture areas which formed part of the study's ground truth, the AAFC land use dataset from 2020 was also used to gather

testing regions for the pastureland classes.

Water samples came from NRCan's Dynamic Surface Water dataset [183]. These maps were generated from Landsat satellite data. Inundation frequency is the percentage of valid observations when water is detected at the surface [184]. The maps in this dataset display from 0% inundation frequency to 100% inundation frequency, i.e. land regions that never have water to areas that are perpetually covered by water. This study chose to select only areas with at least 95% inundation frequency annually to be used in the classification effort. (Another study had used 90% as the inundation frequency threshold for permanent water during validation of their classification [184], while another labelled a pixel as permanent water when its surface water prevalence was between 85-100%, and as ephemeral water when the surface water prevalence was between 15-85% [185].)

Dynamic surface water maps are annual masks representing the maximum annual water extent [184]. And the inundation frequency maps are done over the entire time series; for instance, the one used in this study was from 1984-2019. (Each year, the time series is updated to 1984-2020, 1984-2021, etc.) Thus, the inundation frequency is taken from the time series map. The validation samples were taken by randomly selecting areas from the 2019 surface water mask map – and then confirming the regions are historically of 95% or more inundation frequency. Any sample region seen to have a frequency of less than 95% was removed from the validation sample set.

These datasets are all freely available to those conducting landcover classification and change detection. As is central to the PAM method, the input data is easily accessible for the study and monitoring of landcover change. Given the various sources of land cover/land use data, the proposed method compares the classification of training and testing sample points from exiting data products. If the data maps disagree at any point, the remote sensing data can be used to confirm the cover type.

4.2.2 Selecting Landcover Classes

The land cover classes chosen for the supervised classification in this study were urban areas, forested, pasture, water, cropland, and wetland. The urban class comprises the built-up areas

such as roads, buildings, and other human-made structures, such as railroads. Initially, the roads and buildings land uses were separated, but there was no need in this study to have them analysed, e.g., comparing roads to building class, and so it was sufficient to group them. Furthermore, all their spectral signatures are similar, which lends to their being in one class. The forested areas were divided into mixed forests, coniferous forests, and deciduous forests. Prior knowledge of the area informs that the Clay Belt has mostly mixed and coniferous forests, with deciduous forests being more prevalent in the western part. Given that the study aims to track the conversion of forests to agricultural land – regardless of the type of agriculture, such as livestock, cropland, etc. – the pasture and cropland separation was deemed necessary. The pasturelands could form areas for grazing livestock (significant when estimating carbon and GHGs from manure in the area) or could grow hay which is harvested to feed animals.

Wetland is a significant land cover type in Canada: [186] refer to studies that estimate that 13% of Canada’s terrestrial area is covered by wetlands [187], and 25% of the global wetlands are in Canada [188]. One study’s land classification estimated Canada’s landcover is 36% wetland [186]. Wetlands are dynamic, even so with the different seasons, and so the discrepancy in classifying this cover type is not surprising. Nonetheless, the selection of samples for this land cover through the land use datasets is adequate.

4.2.3 Feature Selection

From the Landsat and Sentinel optical sensors, features input to the classifier for the supervised classification were surface reflectances from the visible, near-infrared, and shortwave infrared bands. Also included were indices: Normalized Difference Vegetation Index (NDVI), which is sensitive to vegetation characteristics; the Normalized Difference Built-up Index (NDBI) for delineating built-up areas; and the Modified Normalized Difference Water Index (MNDWI), which is sensitive to the water content in vegetation. The Sentinel-1 SAR provided features of the VV single co-polarization band and the dual- band cross-polarization VH. These features were input to the Random Forest classifier, with parameters such as the number of trees, bag fraction, and number of variables per split being chosen based on hyperparameter tuning.

The area’s elevation and slope were also included in the classification. These topographic

parameters were obtained from the Canadian Digital Elevation Model [189], a dataset provided by NRCan. The elevation affects a region's precipitation level and temperature; thus, it is essential to include such features in the classification. The slope is calculated within GEE using the *ee.Terrain.slope* operation. Given the elevation image in metres, the local gradient is computed from the 4-connected neighbours of each pixel. All the features were normalised between 0 and 1.

The final composite images were of the growing season each year, from July to October. For the period under study, the median composite of the images was used for the analysis. This choice was made over the mean composite of the optical data so as not to introduce possible errors of high reflectance values from a cloud if the mean was selected. Since the classification was performed as a multisource task, the images and features from both the Landsat and Sentinel sources were merged. Similarly, when the SAR data was included for analysis, the resulting median image was concatenated with the optical data after individually filtering the VV and VH SAR datasets.

4.2.4 Obtaining Training and Testing Areas

Field surveys of the study area were conducted in 2018, 2019, and 2021. Thereby, I was able to record several ground truth coordinates for the landcover types before the COVID-19 pandemic prevented my taking further trips during my studies. However, due to the limitations of field testing and since the Clay Belt is so large at over 120,000 sq. km, more sample regions were required for the training and testing data of the classifier. Using the datasets such as discussed in Sec. 4.2.1 above, polygons of each landcover type were primarily created in ArcMap 10.6 to be used as training and validation areas. This selection could have been performed in Google Earth Engine; however, some datasets are not available on that platform; for instance, the AAFC annual crop inventory maps are available, but not the AAFC Land Use maps.

Although training areas were taken as polygons, Google Earth Engine handles them as individual pixels. Therefore, the size of each class's training polygons (in terms of pixels) was counted, and then the numbers were pared down to be relative to the class's areal ratio. For instance, 700 polygons were created as training samples for the coniferous class; this amounted to 35,955 pixels. Furthermore, the total number of training pixels was less than 55,000. Therefore,

the coniferous forest class would be overwhelmingly favoured in classification. Given that conifer forest cover type is usually about 30-35% of the study area, the training samples were screened in an attempt to make the fraction of samples consistent with the fractional distribution of the classes within the study area. Hence, of the 35,955 pixels of the coniferous forest class, less than 50% of that class's sample pixels were used for training the classifier. (The division was achieved using the *randomColumn* function in Google Earth Engine, filtering to only select a random number of less than 50% of the pixels in that class.)

Having gathered all the input parameters to the RF classifier, the values of the parameters are normalised. Looking at some of the values (for instance, the reflectances could be numbers in the thousands while the indices are fractions), it becomes apparent why normalisation is necessary prior to performing the classification. Machine learning algorithms give more weight to higher numbers as being more important. Therefore, the training and testing feature values were normalised to be between 0 and 1 so that no feature was weighted heavier than others.

A design was explored in which the training sites were taken from Hearst and its environs (towards the west of the Clay Belt), while the test sites were mainly from the far west, as well as Kapuskasing and Cochrane and their environs (the central to eastern parts of the Clay Belt). Performing this region-based classification might reduce bias and give a better assessment of the classifier.

4.2.5 Preparing the Classifier

The Google Earth Engine platform ran the Random Forest classifier for the land classification and change detection process. The number of trees used (set by the function *numberOfTrees*) affects how quickly the model will load. For instance, if the number of trees is 3000, the classifier will take much longer to run than if the parameter had been set to 30. However, having more trees generates a more statistically robust RF model. For this initial classification, two values between 10 and 300 trees were experimented with. For the results presented here, the RF classification was done with 80 trees. Another parameter which was specified was the number of variables per split. Again, this parameter was investigated to see its effects on the classification results. It was initially set at 5, 7, and then given the default value. (If unspecified, the classifier uses the square root of the number of variables as its value).

Hyperparameter tuning was done as part of the classification process: After initially generating an accuracy report from the Random Forest classifier, hyperparameter tuning was done to ascertain, for example, the least number of trees necessary to give the highest accuracy. Usually, more trees mean more accuracy, but that also means more computation capacity. Thus, the tuning becomes necessary.

Two other parameters of the Random Forest (Bag Fraction and Number of variables per split) were included so that all three parameters could be optimised simultaneously. The Bag Fraction was tested between 0.1 and 0.9 at an interval of 0.1, while the number of variables was tested based on the variables specified for each sensor's classification. Thus, the ideal parameters were selected, whereby accuracy was highest.

The classifier was then run separately on the training and validation datasets. The classification was performed individually on the optical and SAR images, then their parameters were merged to classify both images simultaneously.

4.2.6 Acquiring Disturbance Information for Change Detection

The disturbances in the study area are crucial to understanding the changes which occur within the investigation periods. They are categorised broadly into natural and anthropogenic occurrences. Some drivers of natural disturbances include fires, floods, hurricanes, drought, insect damage, disease, and blowdown of trees by wind. Similarly, some human-caused disturbances to the natural environment include road-building, laying of powerlines or pipelines, oil and gas sites, and harvest activities. The questions an analyst might seek to answer regarding a disturbance include the type of disturbance, when it occurred, what caused it, where it took place, and the duration of the event – in other words, the “what, when, why, where, and how long?” questions. Remote sensing data can assist in answering these questions. For this work, disturbance information was primarily sourced from public datasets provided by the government. If analysts cannot access such information, other sources, such as remotely sensed data, can be used to acquire disturbance information.

The first areas which were assessed for disturbances were the cropland and pasture areas. The assumption is that there was initially continuity between the nearby natural forests and any

area developed for agriculture. Thus, the lands would have similar properties and would have gone through similar disturbances, e.g., flooding, drought, etc. Furthermore, the agricultural areas might have been generated from deforestation or cutting down of trees. Historical maps were studied to determine when the land conversion to agricultural land occurred. (PlanetScope and Google Earth maps were used for this purpose.) In addition, a landowner could provide the information during a field visit, e.g., when the land was cleared, if the vegetation there was burned, and if there had been prior land use change. Remote sensing images could also verify these accounts.

Ontario's forest management plans capture other deforestation activities [60]. The plans are specific for the province's Crown forests, with the planning areas defined as forest management units. Projects are set out in decadal timeframes but in two phases of five years. Phase I is created with its objectives, and areas delineated for harvest, for instance. Yet, any allocation for harvest within Phase II cannot be cut during Phase I. Furthermore, if some of the logging allocation of the first five years is not required, those trees need not be cut: during Phase II, any of the harvest areas can be cut. Phase II gives the optional refresh on the project if changes are required after Phase I activities. (Figure 4.1 shows an example of a forest management plan for a forest in the Clay Belt, for 2020 to 2030.)

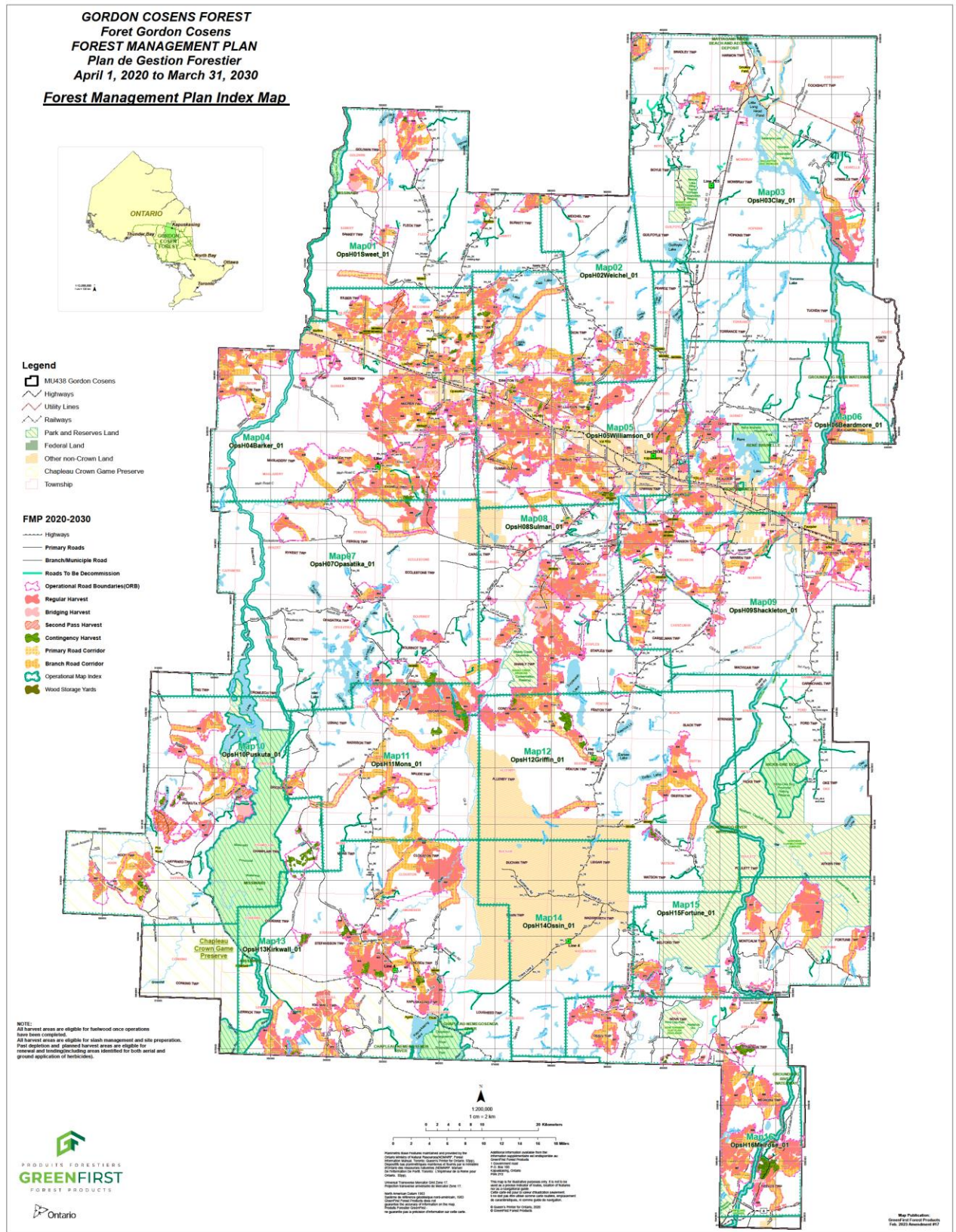


Figure 4.1: Example of forest management plan for forest in Ontario's Clay Belt [91] showing areas delineated for harvest

Natural disturbances which the Ontario governmental database, Ontario GeoHub, provides include insect infestation, disease damage, and wildfires in the forests [94]. Insects which have been detected in Ontario forests in recent years include:

- Jack pine budworm
- Spruce budworm,
- Forest tent caterpillar,
- Spongy moth (previously known as gypsy moth),
- Emerald ash borer.

The brown spot needle blight, beech leaf disease, beech bark disease, and Septoria leaf spot and canker were predominant diseases in Ontario's forest over the past two decades. Specific to the Clay Belt, the Septoria leaf spot and ink spot of aspen were the diseases that afflicted the area's forests. They mostly inflicted moderate to severe damage during the period under study. Another significant disturbance which affects forests is fires. The Ontario GeoHub is a reliable data source for historical fires, and this dataset was used in the study to identify the fires that occurred during the period under study. The dataset includes the start and end date of the fires, the location burned, and the cause of the fires, e.g., industrial, resident, lightning, or recreation.

4.3 Change Assessment from Landcover Classification

From visual observations of the study area maps, the area did not appear to change significantly each year. Landcover classification could be done by taking samples of each cover type on the digital maps for every year being classified. However, with the assumption that the region's textural properties have remained the same, the training samples can be migrated from one year to another, with some preconditions. This helps to address a limitation of the machine-learning classification process: obtaining adequate and accurate training samples.

Instead of finding new samples each year, those pixels or clusters which have not changed class can be reused. For example, if there are 900 random training sample areas in a class one year, those training areas would be sampled the following year. Then, the spectral angle between the two sample areas is determined. If a sample region's spectral angle is high, then it would be discarded as a training sample. If not (or it is less than a chosen threshold), then the sample area would be kept and

used for the following year's classification. Thus, in conducting classification in multiple years, the training points or polygons in one class can be migrated to use as classifier input the following year.

Another essential feature of the Progressive Adaptive Method (PAM) is verifying that the 'change' of the landcover class is a real change. Therefore, post-classification analysis needs to be done. Post-classification corrections were performed in this wise: a comparison of the land cover classification from one year to the next was applied but with some stipulations. For instance, if a pixel is classified as urban one year, it could never change to water the following year. Such corrections were applied, changing the pixels which went from urban to water one year back to urban. This was the forward correction. Similarly, a backward correction was done such that if a map section is classified as water this year, it should not have been classified as urban the previous year. Doing corrections such as these produced a change in the area sometimes; for instance, in one class, it went from over 294 sq. km. to 1731 sq. km (see Figure 4.2 which illustrates the method).

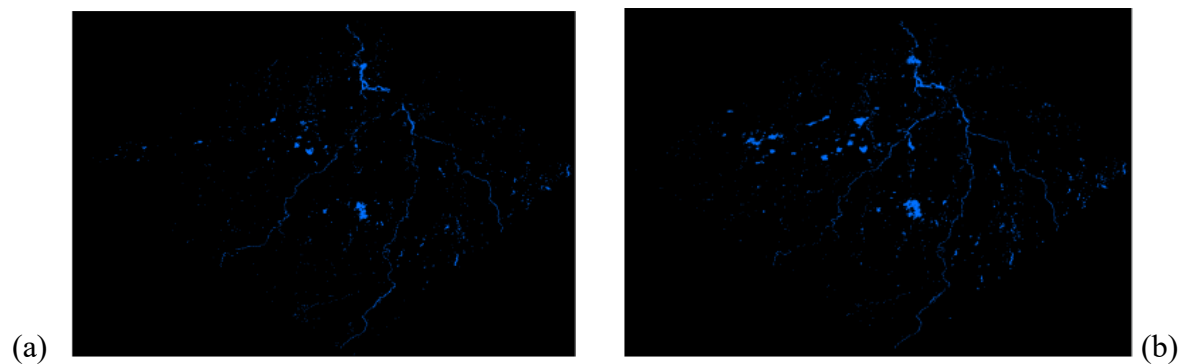


Figure 4.2: Corrections made in reverting from changes in water class between (a) 2015 and (b) 2016

Now, the classes used for the study area were urban, water, forested, pasture, wetland, and cropland. In determining which classes would be constrained in the forward and backward correction, I decided on only those which were highly improbable, like water and urban classes being converted. The matrix in Table 4.1 below shows the logic used to determine the possible land cover class changes. Finally, the changes between two years were assessed by the image differencing method [190] of the classified maps.

Table 4.1: Change Detection Succession Matrix

LAND COVER CLASSES	Urban	Deciduous Forest	Mixed Forest	Coniferous Forest	Pasture	Water	Cropland	Wetland
Urban	3	0	0	0	1	0	1	0
Deciduous Forest	0	3	2	1	2	0	1	1
Mixed Forest	0	2	3	2	2	0	1	1
Coniferous Forest	0	1	2	3	2	0	1	1
Pasture	0	1	1	1	3	1	2	1
Water	0	0	0	0	0	3	0	1
Cropland	0	0	0	0	2	0	3	1
Wetland	0	1	1	1	1	1	2	3

Changes: 0 = Not likely 1 = Possible 2 = Probable 3 = Likely

4.3.1 Comparing Disturbance Maps to Classification Maps

The disturbance data obtained from the GeoHub was filtered to the area under study as well as to the period under investigation. The map of the resulting disturbances – of the various kinds, both natural and anthropogenic – was then created. It then became apparent for each year’s map where a difference in landcover type or landcover area might be expected. Thus, the “when,” the “why,” and the “where” of possible landcover change is known.

The classification maps generated by the Random Forest modeller were differenced to ascertain the areas which had shown a change in landcover class at each time step. That is, the initial and final images in each time step were classified and any region whose resulting

classification value did not equal 0 following the differencing was identified as a possible change area. If an area was still identified as being changed following post-classification corrections, the transitions in class were then ascertained. The class identifications in the second image were remapped by a constant value, e.g., multiplied by 100, and then the class identifications of the first image were added to it. Any part of the merged image which did not have a ‘mirror-value’ as its new class would show what the area had transitioned to. For instance, a mixed forest pixel had class identification 2. If it was unchanged, multiplying its classification value in the second timestep by 100 and adding the initial value of 2 would yield a mirror-value of 202. However, when that area transitioned to deciduous forest (whose class identification was 3), following the merging of images, its value would be 203 – indicating an initial value of mixed forest class 2 and a final value of deciduous forest class 3. The changed areas could then be mapped by individual class transitions, displaying all regions which did not have mirror-values in the classification map.

Having individual classification maps and the disturbance maps permitted the comparison – whether the maps generated by remotely sensed images showed the changes expected by the disturbance maps; in a sense, crosschecking each other. In some cases, the maps could be correlated, for example, if a woodlot had been cleared and converted to pastureland. But if a moderate insect infestation had occurred and there was no noticeable change in forest inventory in that region, the input of an insect disturbance was still fed to the carbon model. In addition, if the Phase I map of a forest management unit suggests a swath of forest would be logged but the land use change maps do not indicate loss of forest inventory, it is likely the Phase I proposal was not followed through. (Perhaps the harvest will be performed in Phase II.) In this case, the disturbance information given to the carbon model is from the remote sensing maps. Thus, the best available data was integrated to determine landcover/land use change or disturbances in the study area to enhance such input to the carbon model.

4.3.2 Object-based landcover classification

Another comparison done in this study was the segmentation analysis. Segmentation was performed using the Simple Non-Iterative Clustering (SNIC) algorithm in GEE. The SNIC is an improvement of the well-known Simple Linear Iterative Clustering (SLIC) segmentation algorithm.

The SNIC algorithm is reported to yield the best land cover classification results because it can be controlled by user parameters [27]. It is also less memory-intensive than the other clustering algorithms [27]. The SNIC was run to simplify the image into small clusters of connected pixels called superpixels.

The superpixels were created based on prior work [70]. From that work using the SLIC algorithm, the optimal superpixel spacing was about 40 pixels in the study area. The image segments created have properties, e.g., perimeter size, standard deviation, mean, cluster width, and cluster area. These properties are used to classify the segments. For this study, the means of the spectral features were used for classification. In this way, over 637 million image objects were generated; Figure 4.3 illustrates some of the segments' perimeters.



Figure 4.3: Segmentation perimeters from Simple Non-Iterative Clustering method

The SNIC algorithm computes the average values of all the objects and their spectral features. These average reflectance values within the object were classified using the RF classifier (from the training samples). Now, with the superpixel spacing at 40 pixels, the standard deviation derived was slightly less than 0.4.

Previous works reflected an increased accuracy when using SNIC on the Google Earth platform to perform Random Forest classifications: [26] showed between 0.1 and 4% increase in

overall accuracy. Reference [28] used a fusion of Sentinel-1 and -2 data, along with the high resolution PlanetScope data, to compare the image classification results when using pixel-based and object-based methods. The results showed an increase of between 2 and 15% in accuracy, depending on the sensor data used, when the object-based classification was performed. (The study also examined classification outcomes from only Sentinel-1 and -2 fusion, or PlanetScope alone or only Sentinel-2 images.) Another study's classification accuracy increased from 94.2% to 96% when a random forest model was used, respectively, for a pixel-based and object-based classification [27]. The SNIC algorithm was used in that study for the object-based classification. The classifier was trained on spectral features, spectral indices, topographic features, soil features, and distance to water bodies.

In this work, the object-based classification was performed on the Sentinel-1 and Sentinel-2 fused images as in [27]. The features from the Sentinel-1 dataset were the filtered VV and VH bands and the texture metrics from the Gray Level Co-occurrence Matrix (GLCM). Five of the possible fourteen GLCM texture metrics were implemented in this classification task, on both the VV and VH: the variance (which measures how dispersed the distribution of gray-level in an image is), entropy (which is a measure of the randomness of gray-level distribution in the image), contrast (that measures the local contrast of image), angular second moment (that measures the number of repeated pairs), and the correlation (which is a measure of the correlation between pixels). Finally, a Principal Component Analysis was performed over the five GLCM metrics to obtain a principal component band containing most of the textural information [26]. It is known that most of the information is contained in the first principal component, while the first three principal components will contain over 90% of the information. In this study, the first principal component derived of each cluster was added to each segment's mean as another band on which to train the Random Forest classifier.

4.4 Results

Following the classification done using Landsat and Sentinel-2 images, both separately and then as fused mages, the results of the Random Forest classification were assessed. Figures 4.4-

4.9 show the results of classifying from the pixel-based method. The same training and validation data were used to classify the landcover maps for the different sensors. The overall accuracy of the classification was calculated in Google Earth Engine. Figures 4.4 and 4.5 display the classification from Landsat data.

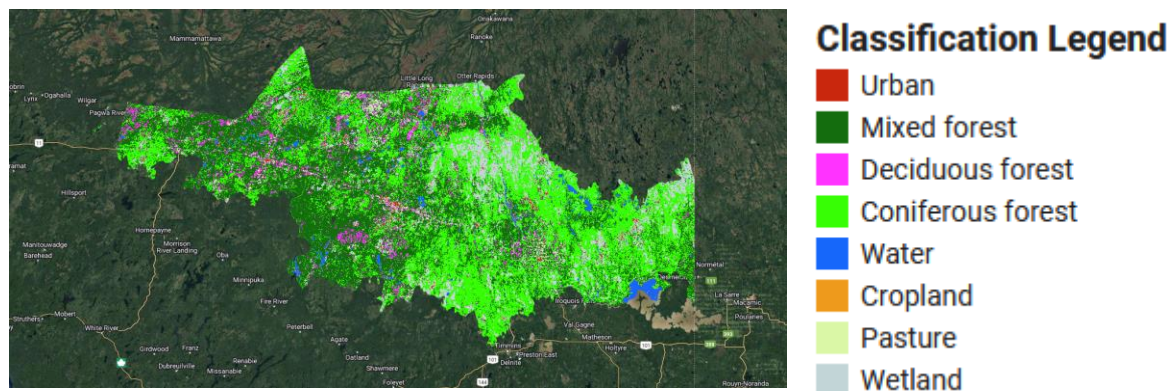


Figure 4.4: Landcover classification using Landsat 8 only

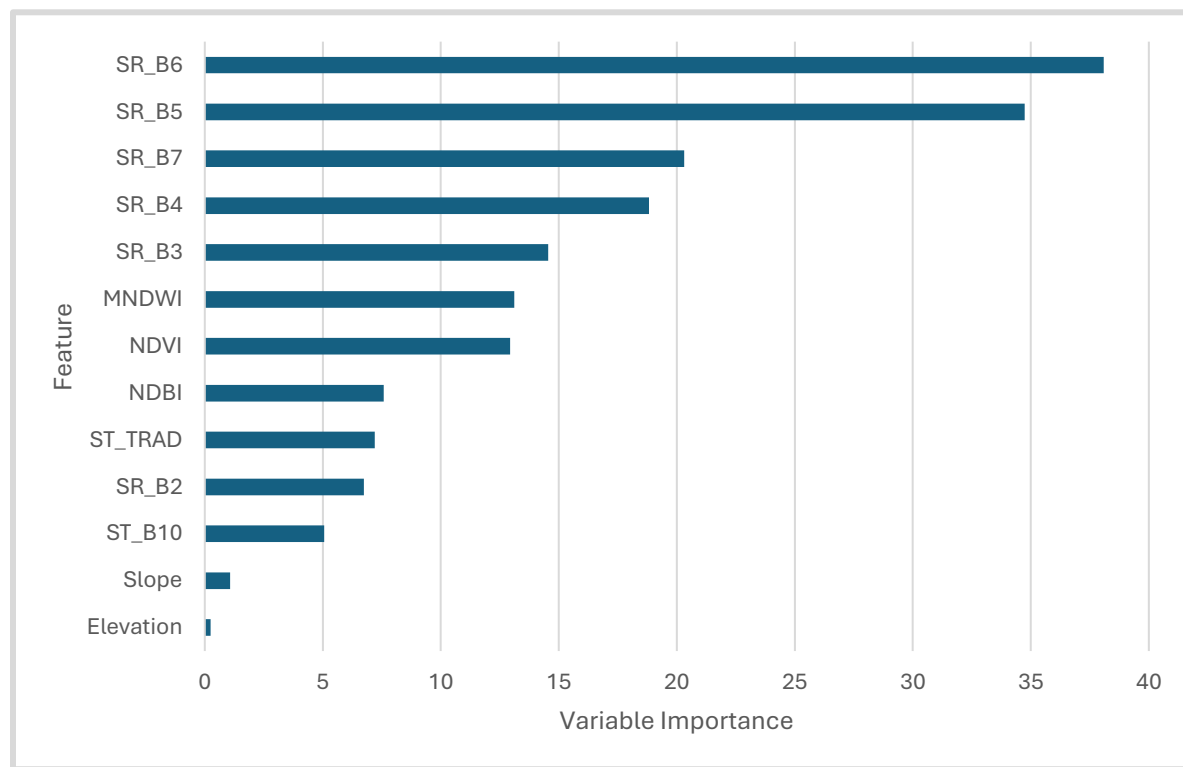


Figure 4.5: Variable Importance of the Random Forest model using Landsat 8 data

Similarly, the classification results using the Sentinel-2 data are presented in Figure 4.6 and Figure 4.7

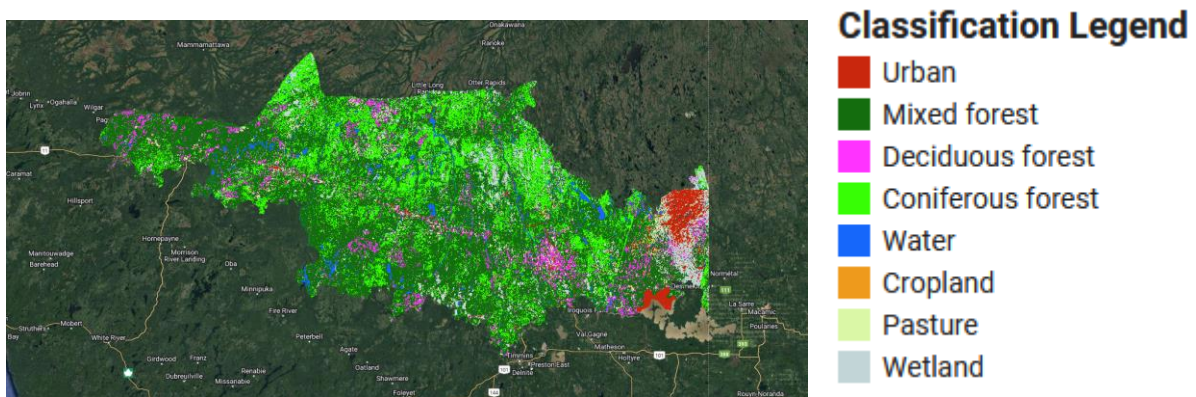


Figure 4.6: Landcover classification using Sentinel-2 only

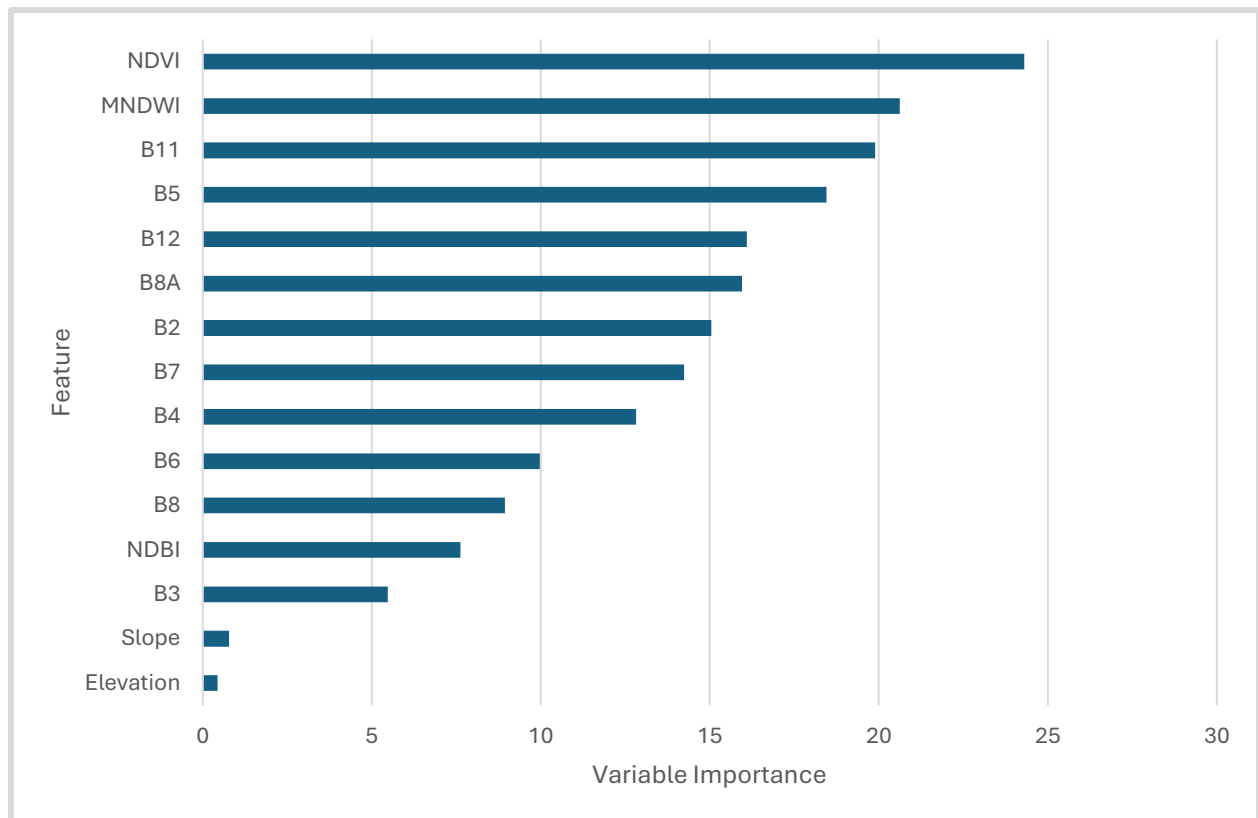


Figure 4.7: Variable Importance of the Random Forest model using Sentinel-2 data

Finally, Figure 4.8 and Figure 4.9 display the classification results from the fusion of Landsat 8 and Sentinel-2 imagery.

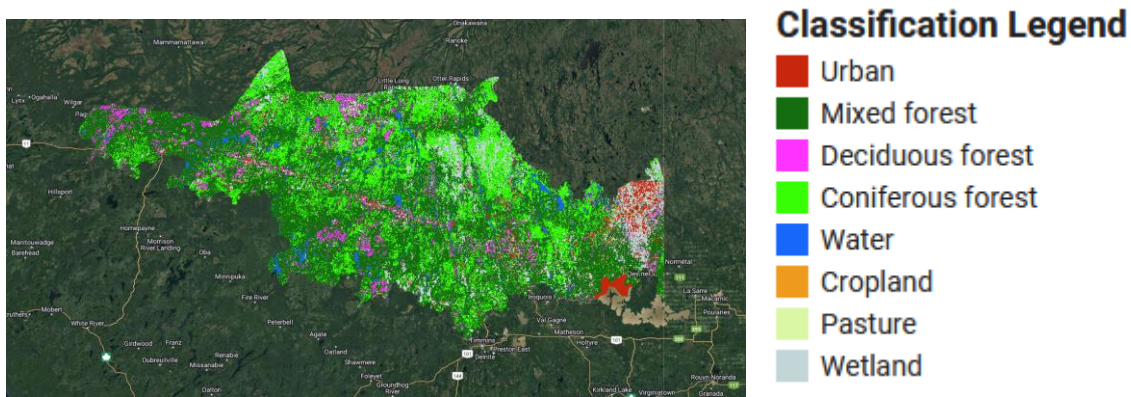


Figure 4.8: Landcover classification using Landsat 8 and Sentinel-2 imagery

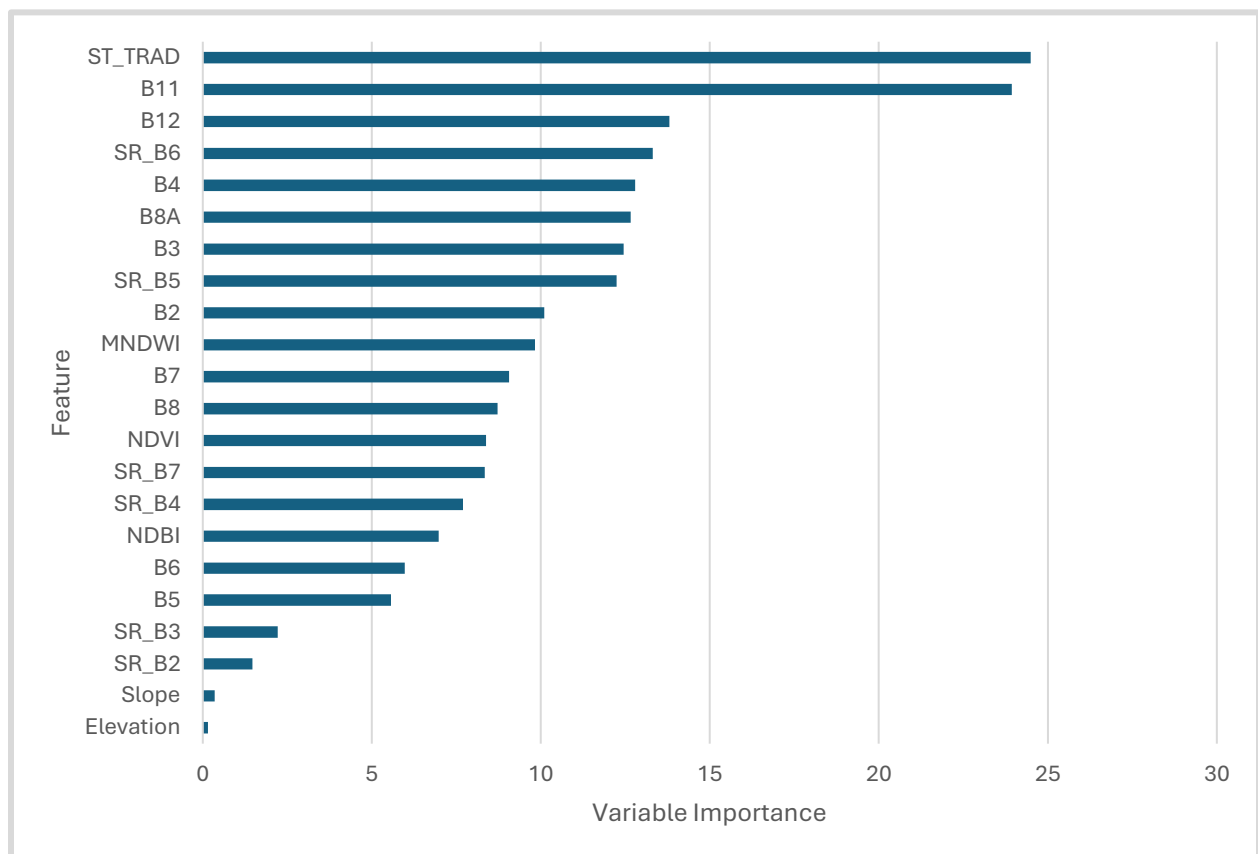


Figure 4.9: Variable Importance of the Random Forest model using Landsat 8 and Sentinel-2 data

The Random Forest classification model was measured using the Statistical Machine Intelligence and Learning Engine (SMILE). The SMILE feature importance measurement method

can be run directly in GEE by calling *.explain* on the classifier. This will yield the importance of each feature in the classifier, determining the final class assigned to an area. It should be noted that the values calculated are relative; the absolute variable importance can be estimated by normalising the importance values. The bands of the Sentinel-2 sensor are denoted as the B* features, while the Landsat 7 and Landsat 8 reflectance bands are denoted as SR_B* in Figs 4.7 and 4.5 respectively. Looking at the charts of variable importance, it is observed that elevation and slope rank as the least important feature in the classification (see Figs 4.5 and 4.7). This is not unexpected since the relief of the Clay Belt is not significantly changing. The slope and elevation features might better serve other regions in distinguishing a landcover type. End-users of the GEE code can include or exclude them as deemed necessary. The shortwave- and near- infrared reflectance bands were ranked highest for the Landsat-8 classification. In the Sentinel-2 classification, water and vegetation indices as well as the red-edge (B5) and shortwave infrared (B11) bands were ranked as being the four most important to the random forest classification. Interestingly, with the fusion of Landsat and Sentinel sensor data, the thermal reflectance band from the Landsat data was most important (see Figure 4.9), followed by the shortwave infrared band (B11) from the Sentinel-2 sensor, along with the shortwave-infrared bands of the Sentinel and Landsat sensors (B12 and SR_B6).

Areas of healthy green vegetation have high reflectance in the red edge and near-infrared bands. Hence, these bands were very important in distinguishing the vegetation in the study area. In addition, the shortwave infrared bands are useful in identifying water content in the area. Discriminating land from water is usually in the near- to mid-infrared region of the electromagnetic spectrum. The ST_TRAD band values are produced from the thermal band reflectance. As noted before, the land surface temperature is a feature that can be used to distinguish cover types. Overall, the features selected assist in defining the land cover classes due to the spectral reflectance characteristics of cover types. Moreover, a continuous evaluation can be performed with different combinations of features to improve the classification accuracy.

The results of the classification accuracy assessment are displayed in Table 4.2.

Table 4.2: Overall Accuracy of Classification for the different sensors

Data source	Number of spectral + auxiliary bands	Overall accuracy
Landsat 8	13	89.91
Sentinel-2	15	90.49
Landsat 8 + Sentinel-2	22	90.97

In the Sentinel-2 classification, the classes which were most confused were the mixed forest being confused with deciduous and coniferous forest classes, and the wetland. The deciduous class mainly was confused with the mixed forest class. However, the coniferous forest class had high user's accuracy (98.7%) and had displayed only relatively minimal confusion with the mixed forest class (see Figure 4.10 below). The wetland class was also confused with the cropland class and vice versa: the wetland class had a 64.2% user's accuracy while cropland's was 48.7%. The pasture class had a high users' accuracy (of 96.4%) and was sometimes confused with the wetland class. The pattern of confusion was similar for the Landsat classification, except with the pasture class, where that landcover class was confused only with the urban class, yielding a user's accuracy of 98.5%. In both the Landsat and Sentinel-2 classifications, the water and urban classes were hardly confused with the other cover classes (having 93% and 99.5% user's accuracy). Similar to the Sentinel-2 results, the wetland class mainly was confused with the croplands, yielding a user's accuracy of just over 15%. Further splitting of the wetland class into more categories might give better discrimination for this cover type, and thus improve its user's accuracy.

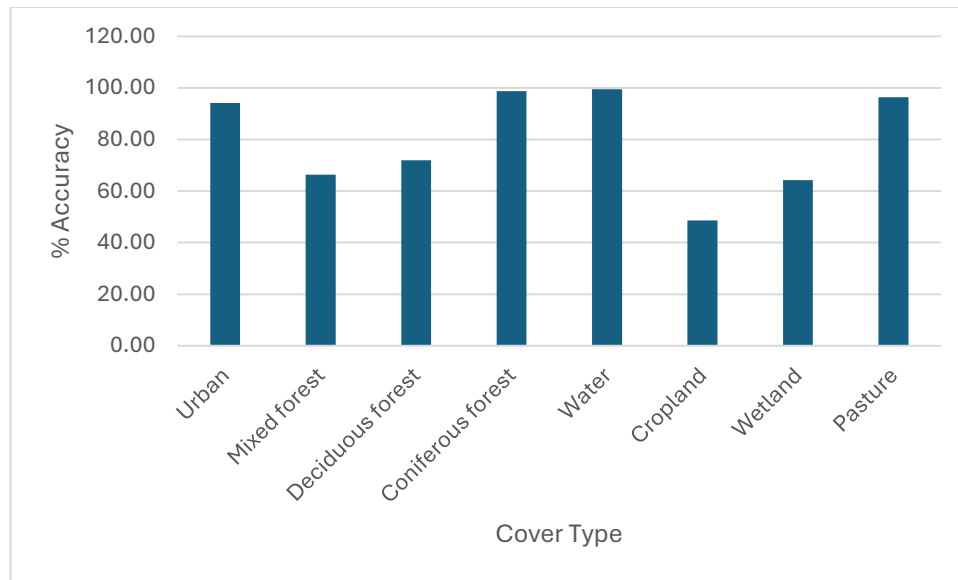


Figure 4.10: User's accuracy from Sentinel-2 classification

The cover types of the Clay Belt were examined over the two decades from 2002 to 2022. The changes in areas were observed, noting particularly classes of landcover which appeared to be part of a landcover conversion. The wetland class sees a marked increase, with 39% of the landcover being classified as wetlands. Meanwhile, pastureland decreased in area from 2002 (from 1421sq. km to 581sq. km in 2022), or as a percentage falling from 3% of the total area to 1%. Now, part of the wetland class can also be considered forested because wetland consists of shrubs and trees, and in the Clay Belt, some of the wetland areas are found within forests.

The urban and permanent water classes remained relatively stable over the study period, while the forest classes of wetland, coniferous- and mixed-forest seemed to increase at the expense of deciduous forest and cropland as seen in Figures 4.11 to 4.13 below.

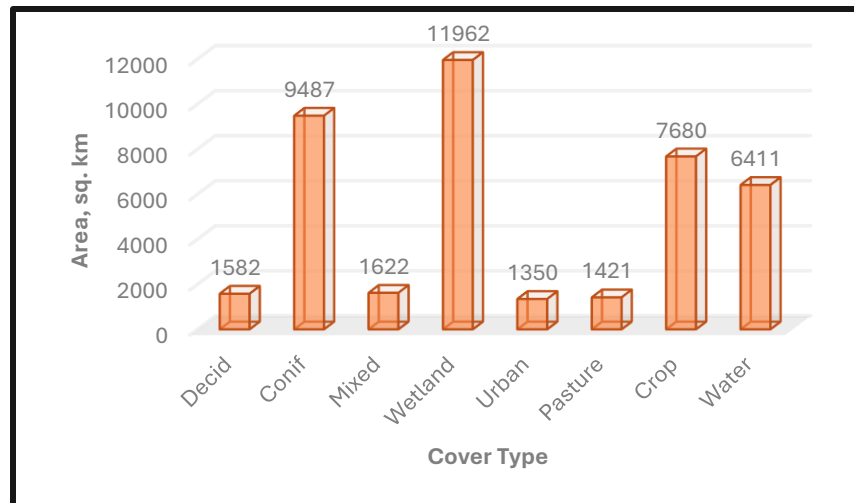


Figure 4.11: Landsat 7-derived Land cover areas in 2002

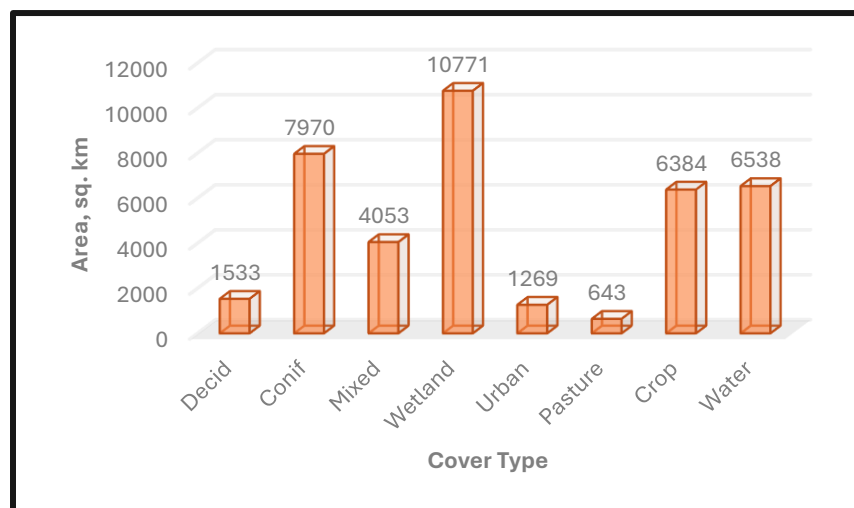


Figure 4.12: Landsat 7-derived Land cover areas in 2012

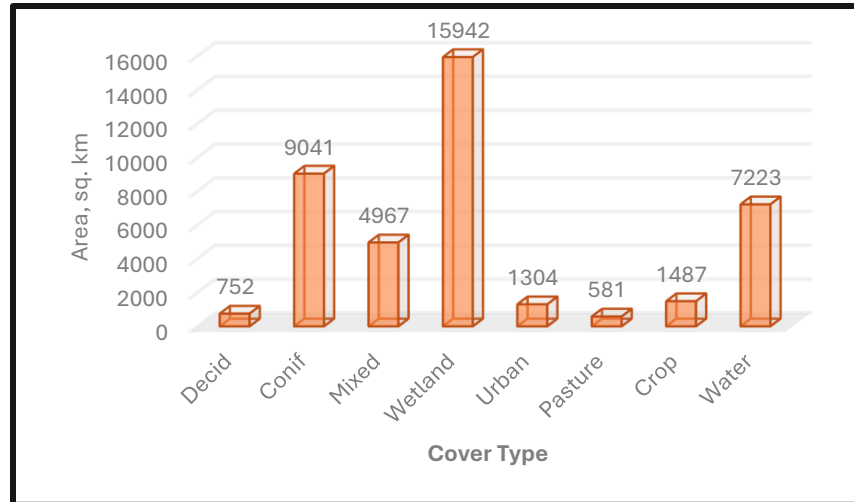


Figure 4.13: Landsat 8-derived Land cover areas in 2022

Segmentation Results

Similar to this work, the study area in [27] was a mixture of landcover types – vegetated, water, and built-up areas. Thus, a similar increase in overall accuracy was expected to result from the object-based classification. However, when the object-based Random Forest model was run on different parts of the Clay Belt, such as the Gordon River Forest, the overall accuracy decreased by 2.56% on the original classification done using the Sentinel-2 image data. When the object-based classification was performed over the entire Clay Belt, there was an increase of 0.67% from the pixel-based classification accuracy. The trend in user's accuracy was similar to that observed after the pixel-based classification: the water and urban classes had very high user's accuracy (~99.6% and 99.7% respectively), the lowest user's accuracy was seen in the cropland class, and the greatest confusion between classes was seen in mixed forest class where it was confused with several other cover types – mostly the deciduous, coniferous, and the pastureland (see Figure 4.14). The cropland was mostly confused with the urban and wetland classes. There was no significant difference in the various classes' user's accuracy between the object-based method and the pixel-based method except in the cropland, mixed forest, and the wetland class: the cropland user's accuracy decreased, while that of the mixed forest and wetland classes increased by 0.06%. Cropland is known to be confused with urban areas because of the large variance of spectral characteristics in both cover types [191].

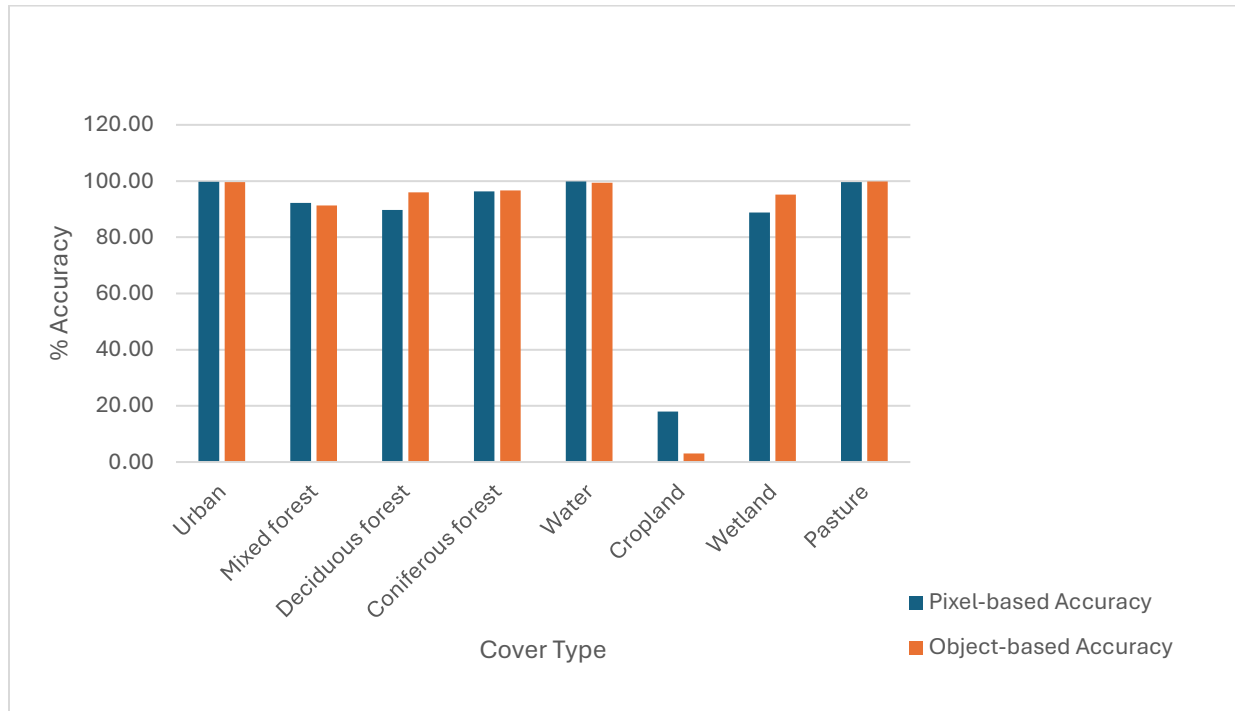


Figure 4.14: Comparing user's accuracy from pixel-based and object-based classification performed with Sentinel-1 and Sentinel-2

Another study which compared pixel-based classification to object-based classification on Google Earth Engine also experienced a drop in overall accuracy from the object-based method [192], and another study saw a 4.7% reduction in accuracy when the object-based method was used, compared to the pixel-based classification [26]. As in this study, there could be several reasons for the drop in overall accuracy, such as the spatial resolution of the training data used to read the spectral textures and classify the area correctly. Perhaps using a training dataset which was collected at 10m spatial resolution (i.e. a Sentinel-2 training dataset) would improve the classification result. Also, the amount of training data was possibly insufficient now that the image was segmented into millions of parts. In addition, other features could be added to the data being classified, such as the soil features and distance to water bodies, as done in [27], whose study saw an increase in classification accuracy when performing object-based classification. The object-based image analysis method should still be investigated further to confirm if this method is viable for generating land use/land cover information for carbon models. The GLCM information would usually improve the classification of complex land use and landcover classes, e.g., identifying vegetation with a regular texture [26].

SUMMARY

This study has introduced an application of spatial data analysis to support the tracking of ecosystem carbon and GHG emissions. It demonstrates how geospatial data can be harnessed to monitor land use and landcover change, providing policymakers with crucial information to make decisions that may foster a more equitable and resilient environment. The framework I designed utilises the best existing products and knowledge to detect change, particularly in carbon and GHG analysis.

In this study, I used multisource landcover classification techniques, endeavouring to use the best available data for the training and subsequent land classification of the area. Permanent water body maps were employed to avoid the effects of seasonality in taking training or validation samples from the water bodies. Also addressed was performing annual change analysis from a base year's land classification, without having to gather new training samples, when classified data from existing regional or national datasets are unavailable.

Highlights of the method used in this study for the classification of the landcover area include:

(a) Using the percentile method of finding cloud-free data for image processing. The percentile method is especially useful in areas subject to rainy or inclement weather.

(b) The fusion of multiple satellite datasets from various sensors. Spectral and textural features from the optical and SAR sensors were inputs to the classifier.

(c) Increased accuracy of carbon model from better data input of land classification and land cover change.

(d) Conducting object-based classification rather than pixel-based classification. The object-based classification method has been shown to increase the accuracy of the Random Forest classification by up to 4% [27], [192]. Results from this study, however, indicated there was not a significant increase in accuracy.

(e) Samples of training and validation for the classification being taken from different parts of the Clay Belt to prevent spatial correlation among them.

It is hoped that this work shows the importance of performing further study into improving the input to carbon models by producing better land cover and land use information. Remote sensing plays a vital role in this aspect, as does disturbance mapping, to identify and interpret the changes that take place in the areas under study.

5. CASE STUDY: CARBON AND GHG UNDER LAND CONVERSION IN THE CLAY BELT

The third objective of this research was to develop useful methodologies to characterise the impact land use change has on greenhouse gas emissions and soil carbon stock. As such, a study was undertaken to estimate the soil carbon and GHG emissions by 2043 (simulating four decades of study) if some of the forested lands in the Clay Belt were converted for agricultural purposes. The CBM model finds application in forests, and the DNDC model is designed for agricultural land. For the study, these two models were used accordingly to derive the estimate of soil organic carbon stock and GHG emissions under different land management schemes. A portion of the Clay Belt, which the Ontario government has set aside to pioneer transformation, was used as the region of interest.

5.1 Simulation framework: Model inputs and parametrisation

The DNDC model was run in the Site mode rather than Regional mode; the input parameters were entered individually for each site – five pasture sites and five cropland sites. The crops chosen for the simulation were corn, wheat, oats, and soybean, i.e. crops typically grown in the Clay Belt. The crop rotation was such that the spring wheat was planted early in the year and harvested in the summer, while the winter wheat was planted late in the year and harvested the following year. A fallow year was also implemented every five years. Fertilizer application was simulated, along with the tillage routine. Inputs to the model included:

- Latitude
- Climate information
- Nitrogen concentration in rainfall
- The clay fraction in the soil
- The soil bulk density
- Soil pH
- Initial nitrogen concentration at surface soil

As per the analysis of samples from field studies, the soil nitrogen and carbon amounts were known.

The lab analysis of the soil samples gathered during the field survey provided both the pH(H₂O) and the pH(CaCl₂). However, for this study, the pH(CaCl₂) was used as it is taken to be more accurate than the pH based on the water scale [193]. Both pH measurements were taken by mixing an air-dried soil sample with four times its weight of the fluid – water or calcium chloride – prior to measuring the pH value. The differences in pH(w) vs. pH(CaCl₂) were not always large, for instance in site F5, the pH values were 5.865 vs. 5.593 respectively, and at the P1 site, the pH(w) was 8.070 while the pH(CaCl₂) was 7.617. Although the differences are not always large in value, the pH(CaCl₂) test is held to be more accurate as it reflects what a plant experiences in the soil [193].

The DNDC model required the SOC at the surface, defined as 0-10cm. However, the soil sampling during our field survey was done at depths of 0-5cm and 5-15cm for each part of the transects. Thus, the 0-15cm SOC value was used in the model. The mean carbon amount was estimated from all sample depths in the transect. Table 5.1 and Figure 5.1 below illustrate carbon amounts in all the transects 1-3 from the lab analysis of the soil samples. (Site S17 is an unmanaged pasture area, while S18 is an unmanaged pasture near a swamp.)

In the DNDC model, the soil textural class was required. The different classes encountered in the sites were loam, clay, sandy loam, clay loam, and sandy loam. The CBM model required the specific type of soil present in a site following conversion from forest. The Canadian soil map [194] was used to determine soil types. The simulated areas were either of the Gleysolic, Luvisolic, or Mesiosol soil type. However, in the absence of Mesiosol as an option in the simulator, Gleysolic soil was used.

Most carbon is stored in soil, compared with that in the atmosphere or vegetation [172]. Proper forest management has long been a tool to ensure carbon sequestration and storage. In addition, harvested wood products can be another tool for the Clay Belt region to increase carbon storage. In particular, forest managers can be intentional about transforming the harvested biomass into long-lived wood products. One reason there is much interest in determining the amount of carbon in soil is that increasing the soil carbon storage is effective at mitigating an increase in atmospheric carbon [172]. Another benefit of increasing the soil organic carbon is to increase the

soil quality, and thereby its productivity. In this study, soil carbon is estimated to learn the climatic impacts on it; additionally, carbon sequestration potential within an ecosystem is assessed.



Figure 5.1: Pasture sites S17, S18, S27

Table 5.1: Average Carbon amount by transect

Transect	Carbon, t/ha		
	S17	S18	27
#1	3.21	3.69	4.73
#2	6.30	3.27	7.54
#3	3.38	2.27	6.83

When simulating the pastureland, perennial grasses (like timothy grass) and alfalfa were selected. The study area underwent grass-cutting once a year to maintain uniform pasture and promote productivity of the grasses. The pasture was also grazed from the beginning of May until the end of October each year, the timing of the pasture in accordance with what usually obtains in Ontario [195].

Another situation proposed was the forest being converted to a livestock farm. Beef farming is already widely carried out in the Clay Belt [7]. An initiative for “Beef Cattle Farming Training Program for Indigenous people” was conducted in 2017, shortly after the Ontario provincial government announced the decision to convert part of the Clay Belt to agricultural land [196]. For this aspect of the simulation, it was proposed that there would be growth in the livestock, e.g., five head of cattle initially on the farm, and after five years, there would be five more head of cattle. Then, by the culmination of the twenty years of the simulation, there would be 20 head of cattle. It was simulated that the cattle would graze in the area for 10 hours a day from May 1 to October 30 each year, and the grass would be cut once a year. Cattle grazing (dairy and beef cattle) is usually done in the Clay Belt [197]. Goats (simulated as sheep in the DNDC program) and pigs

were simulated as the livestock in certain sites, as shown in Table 5.2. There is provision in the model to have external manure applications simulated. However, on the pastureland, I only simulated the excreta from the grazing animals on the fields, and this manure is spread intermittently to prevent damage to the pasture. (If not done, the grass underneath the dung could deteriorate or be killed by the high nutrient amounts.) Furthermore, the hooves of the animals can assist in breaking up and spreading the dung across the pasture. The speed of the dung disintegrating also depends on climate [198]. For this scenario, there was an addition of farmyard manure, simulated as consisting of about 86% carbon for each kg/ha. However, an external manure application was modelled for the cropland areas.

Another important parameter to model was the irrigation. Advances in technology, such as tiled irrigation, have played an important role in making the Clay Belt region more economically viable for crop production. An irrigation index was specified as 100% such that every water deficit is met by irrigation water [81]. The water budget in DNDC has input fluxes of precipitation, irrigation, and groundwater supply.

Table 5.2 below lists the crops, with their attendant parameters, that were simulated as being grown when the forest is converted to farmland.

Table 5.2: Properties of the 10 sites simulated on the DNDC simulator

Name of Site	Land Use	Plant	Planting // Harvest Dates	Texture	Clay %	pH	SOC at surface (0-15cm)
P1	Pasture: Dairy cattle farm	Alfalfa	Perennial	Clay	59.70	5.6	0.318811
P2	Pasture: Goat farm	Timothy grass	Perennial	Clay	47.27	6.7	0.846974
P3	Pasture: Beef cattle farm	Alfalfa	Perennial	Clay loam	35.6	6.6	0.54909
P4	Pasture: Goat farm	Bromegrass	Perennial	Clay loam	54.6	6.2	1.147014
P5	Pasture: Pig farm	Bromegrass	Perennial	Clay	47.16	6.8	1.129642
F1	Crop Farm field	Corn	May // October to November	Clay	49.58	7.0	0.424732
F2	Crop Farm field	Spring wheat	Early April to mid-May // August	Sandy clay loam	13.93	5.9	0.110473
F3	Crop Farm field	Oat	Late April to early May // August	Sandy loam	16.91	6.5	0.457353
F4	Crop Farm field	Winter Wheat	Mid-September to mid-October // Late July or early August	Sandy clay loam	24.54	5.2	0.851421

F5	Crop Farm field	Soybean	Mid-May to early June // October	Clay loam	35.82	7.1	0.876171
----	-----------------	---------	-------------------------------------	-----------	-------	-----	----------

The yield tables input by the user are used to create the growth and yield curves in the model. The growth and yield tables permit the calculation of other parameters such as merchantable wood volume of each species as well as the aboveground biomass components like stemwood, branches, and foliage. These tree growth curves were mainly obtained from provincial sources and some from the Plonski Yield tables, as the growth curves vary by province. For example, the growth curve for red pine in Alberta, a province in western Canada, could be different from that in Ontario, which is in central Canada. A few of the growth curves used were generic to the tree group, e.g., deciduous or coniferous. Figure 5.2 below shows examples of the growth and yield curves I created in CBM, having input the volume of the tree in m^3/ha for each growth stage. The CBM model has volume-to-biomass conversion factors within it [77] which are used to estimate the resulting biomass carbon.

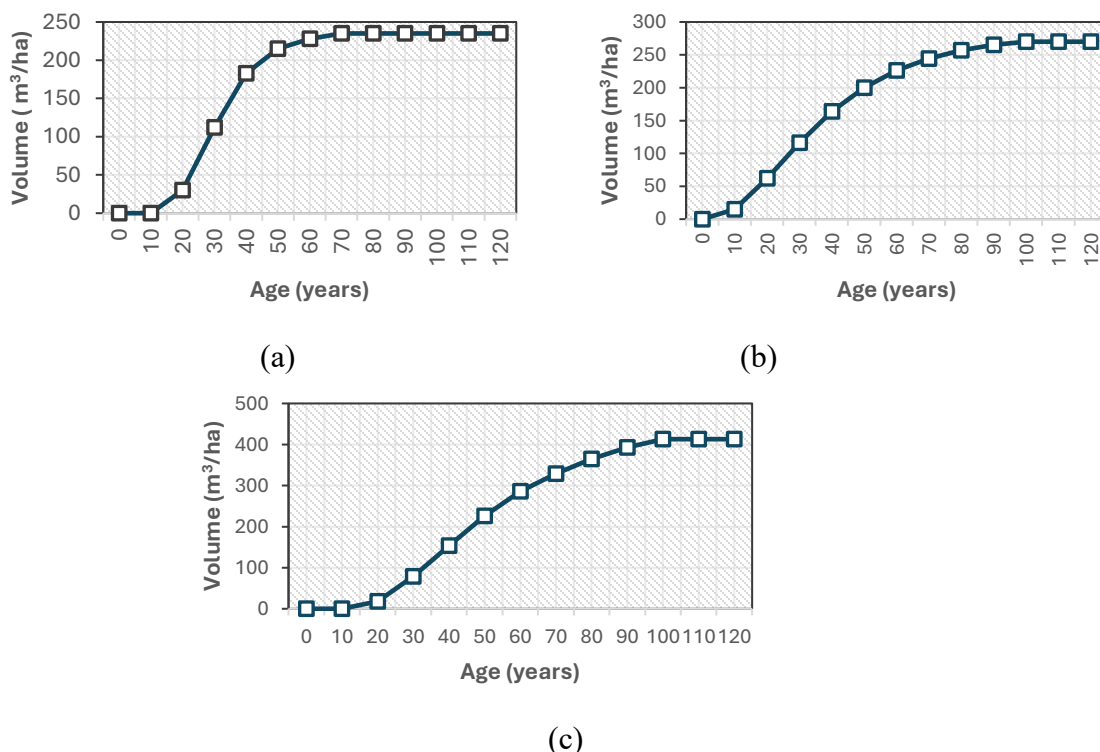


Figure 5.2: Sample growth and yield curves of (a) white spruce, (b) larch, and (c) black spruce trees

In a previous work, I simulated competition in the carbon models by adding a growth retardation factor in areas of mixed stands [89]. Thus, the effect of having different species in the stand as opposed to a uniform stand could be examined.

The transition rules specify what occurs when there is a land disturbance. For instance, if a stand had 100% black spruce trees before a wildfire event, it could be modelled such that its composition becomes 20% fir, 30% spruce, and 50% aspen after the fire. Similarly, when a deforestation event occurred, the type of soil was also altered, e.g., from a forest only soil to a non-forest Gleysolic soil. At conversion, it was assumed there was always salvage of tree stems and snags. The disturbance category chosen in the model for that land conversion event was the ‘deforestation - agriculture - with salvage, uprooting, and burning’ type. The forested land was simulated in CBM using inputs such as the number of stands, the specie types in each stand, forest age, and the disturbance events. Table 5.3 below highlights some of the input parameters which were used.

Finally, the option of the forest not being converted was simulated, and the resulting carbon and GHG stocks were analysed from the CBM tool. For these areas, the specie of landcover in the NLP area is used as the input to the model. If the area is not treed, the nearby forest species as taken from the historical landcover maps are used as input to the model. The pattern of insects and disease were simulated as earlier discussed. Table 5.3 contains the properties of the forested sites which were simulated in CBM. As Tables 5.2 and 5.3 describe the attributes of the 10 agricultural sites as well as the 10 forest sites used as a comparison for the baseline carbon, the historical disturbances used in the simulations are as indicated in Table 5.3. In addition, the prediction of carbon stock and GHG emissions up to 2043 was made using the disturbance trends of the region. For instance, wildfires simulated every 30 years with no salvage logging, and disease damage of moderate severity every 15 years with 30% mortality.

Table 5.3: Input parameters simulating carbon in sites and nearby forests for baseline comparison

Site Name	Nearby Forest	Tree Specie	Disturbance History	Conversion Year	Harvest in Past Two Decades?
P1	Coniferous	White spruce	Forest tent caterpillar infestation in 2016	No change since 2002	No
P2	Mixed	Trembling aspen, Larch, Black spruce	Forest tent caterpillar infestation in 2016	2009-2010	Yes

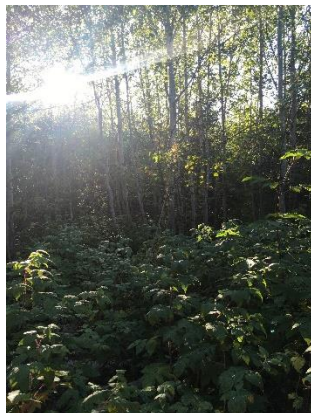
P3	Mixed	Larch, Black spruce	Fire in 2003, forest tent caterpillar infestation in 2015	No change since 2002	No
P4	Mixed	Balsam poplar, Eastern larch, White spruce	Forest tent caterpillar infestation in 2016	Landcover change between 2002 and 2009	Yes
P5	Deciduous	Larch	Fire in 2003, forest tent caterpillar infestation in 2016	Same land use since 2002, only gotten smaller in area	Yes
F1	Deciduous	Trembling aspen	Fire in 2005, forest tent caterpillar infestation in 2016	2017	No
F2	Coniferous	Black spruce	Forest tent caterpillar infestation in 2015. Deforestation, then left fallow; mulched in 2017	2017	Yes
F3	Deciduous	Trembling aspen	Fire in 2003, forest tent caterpillar infestation in 2016. Burnt at deforestation in 2017. Left fallow; not mulched.	2017	Yes
F4	Mixed	Trembling aspen, White spruce, Balsam fir	Forest tent caterpillar infestation in 2016	No change since 2002	No
F5	Mixed	Trembling aspen, Balsam poplar, White birch	Spruce budworm insect infestation in 2021	No change since 2002	Yes

5.1.2 Field Survey and Community Consultation

My research group and I embarked on fieldwork in 2018, 2019, and 2021 within the Clay Belt to collect soil and vegetation data. Thereby, I was able to record several ground truth coordinates for the landcover types before the COVID-19 pandemic prevented my taking further trips during my studies. The field survey was conducted to support the research of digital soil mapping and to assess the impact of land conversion on soil properties and GHG emission. Sample sites were selected at various locations within the Clay Belt in order to have a mix of landcover types – e.g., farmlands, woodlots, forested, abandoned pasture, etc. The understory and overstory observed in the forest areas were always recorded for the plots where sampling was conducted. Figure 5.3 is an example of some of the sites visited for field study.



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)

Figure 5.3 Examples of various sample sites encountered in field survey: a) to c) is forested. In c), the overstory is trembling aspen, with understory of balsam poplar and raspberries; d) to f) are croplands: barley, cereal, and corn; g) and h) are hay

In addition, the discussions with the farmers and landowners where we took some of our soil samples gave us insight into the land history. For instance, they would tell us which part of the land had been unmanaged and for how many years, or their typical rotation of crops, how often or what proportion of farmland they allow to go fallow each year. This information is useful in creating the disturbance history of the area, and also assisted in the design of the simulations which would be performed to predict future land usage and carbon stock.

The coordinates of the locations sampled were recorded using a high accuracy Leica GPS receiver with a Total Station. In order to determine the locations of the plots accurately, the base station was set up (wherever and whenever possible) in an open area near the access points and allowed to run for about 1.5 to 2 hours, during which time the soil sampling was completed. Figure 5.4 shows the base station at one of the plots. At each plot center, a second GPS receiver was also set up.

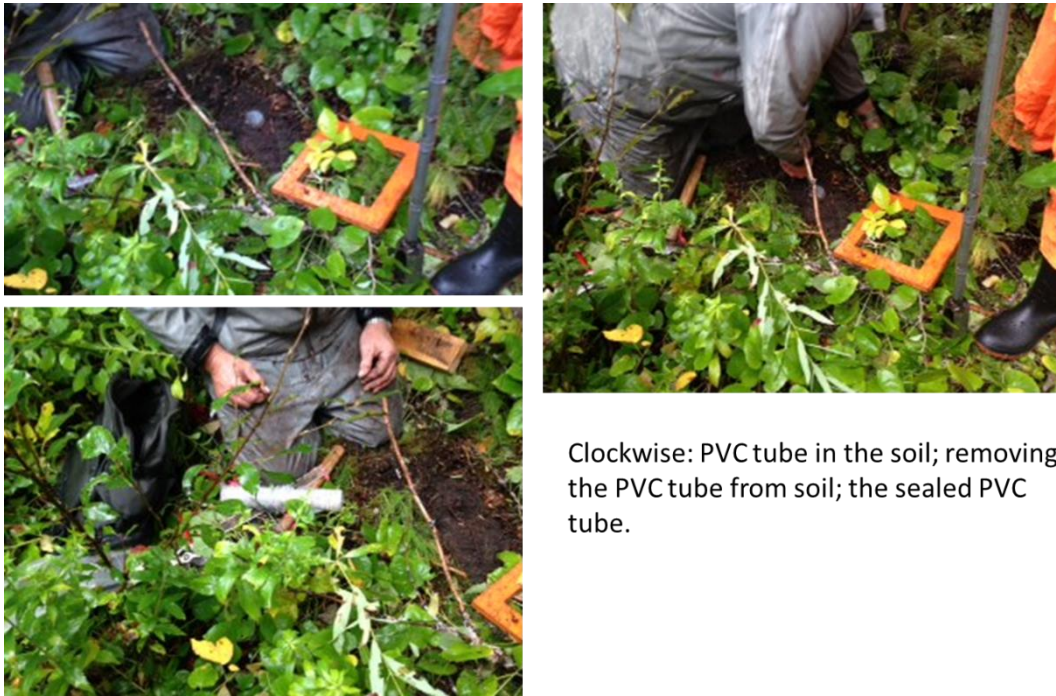


Figure 5.4: GPS base station for acquiring coordinates of plots

The protocol established for the field survey was to set up three 5 m x 5m subplots in each plot. For each subplot, the LFH (Litter, Fermentation, Humus) layer and soil sample core were collected at three different locations. To collect the LFH, a 20x20cm wooden square which would contain the divot was placed firmly at a suitable location in the subplot. The living plants and green leaves in the square area were clipped away and discarded. Then, the loose litter within the divot square was collected into a labelled bag. The interface between the LFH layer and soil layer was identified; then, the LFH layer was removed by using a knife to cut through horizontally along the interface. Having removed the soil attached to the LFH layer, the LFH layer was also placed into

the labelled bag. The depth of the LFH samples varied by the vegetative cover encountered in the area, but it typically varied from 0 – 10cm.

Once the LFH layer was collected, a soil sample for that transect was collected using a PVC tube of about 15cm in length. (The PVC tube was hammered into the soil to gather the sample.) Upon retrieving the tube, both ends were sealed using a specialized plastic bag and rubber band. Figures 5.5 and 5.6 illustrate the procedure for obtaining and storing the soil samples.



Clockwise: PVC tube in the soil; removing the PVC tube from soil; the sealed PVC tube.

Figure 5.5: Collecting soil samples using a PVC tube

For each subplot, soil samples at a depth of 30-45 cm and 90-100 cm were collected.



Figure 5.6: Bagging the soil sample collected using an auger

At almost every sample site, very dense clay was encountered. All the soil samples collected were sent to a lab for analysis. The data was provided in the categories provided, such as 0-5cm, 5-15cm, 30-45cm, and 90-105cm layers for each transect. The lab results provided the mineral content of the soil, e.g. the potassium, magnesium, sulphur, sodium, zinc, and iron. In addition, the soil carbon and nitrogen contents were derived, as well as the bulk density.

Given the bulk density (in g/cm^3) and percentage of carbon in each transect from the soil analysis, the carbon amount was calculated from equation (5.1). Within a specific volume of soil collected:

$$\text{Carbon amount} = \text{Percentage of carbon} \times \text{Mass of soil}$$

or

$$\text{Mass of carbon/area} = \text{Organic carbon \%} \times \text{Bulk density} \times \text{horizon thickness} \quad (5.1)$$

Similarly, the nitrogen amount was estimated, with the percentage of nitrogen in each sample provided by the lab analysis results. Thus, the field data served as validation to the simulated carbon and nitrogen content, where possible. The aim was to get as many useful soil samples as possible, with the emphasis on qualitative samples rather than obtaining a large quantity of samples. Unfortunately, some of the soil samples were not used at the lab because the cores we obtained were not enough for analysis. We obtained samples from 143 sites. Overall, the distribution of sites surveyed over the three field campaigns was 43% forested, 40% pastureland (which included hay fields and shrublands), 13% crop farms, and 4% woodlots.

5.2 RESULTS

5.2.1 Carbon Simulating and Predictions

This section presents the ecosystem carbon and GHG emissions derived from the different carbon models. Whereas the DNDC model tracks flux of GHGs such as CO₂, CH₄, NH₃, NO, and N₂O, the CBM model captures the flux of only CO, CO₂, and CH₄ emissions. Thus, the comparisons and analyses of GHG emissions being performed in this work will only focus on the GHGs which overlap between the two models.

First, the simulation result of the historical ecosystem carbon stocks and GHG emissions for the sites is presented. All the sites had undergone conversion by 2017. Figure 5.7 shows the future prediction for about two decades, with the disturbances expected based on historical trends.

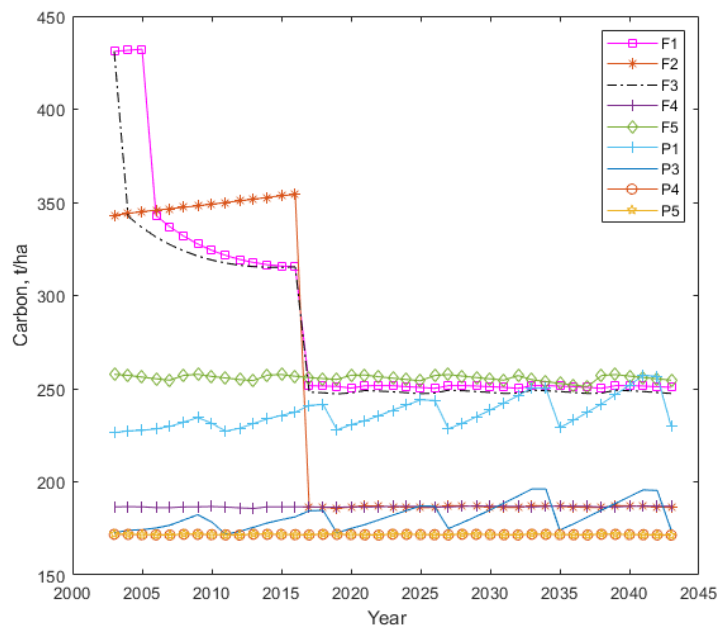


Figure 5.7: Ecosystem carbon stock for forest sites with conversion to land uses (predicting future carbon stocks from past disturbances)

If no land conversion had taken place, the pattern of ecosystem carbon and dead organic matter accumulated over the four decades of simulation would be as shown in Figures 5.8 and 5.9 below. For instance, the total ecosystem carbon was 21,769t/ha in F4, 24,264t/ha in P4 and 7656t/ha in the site adjacent to P5.

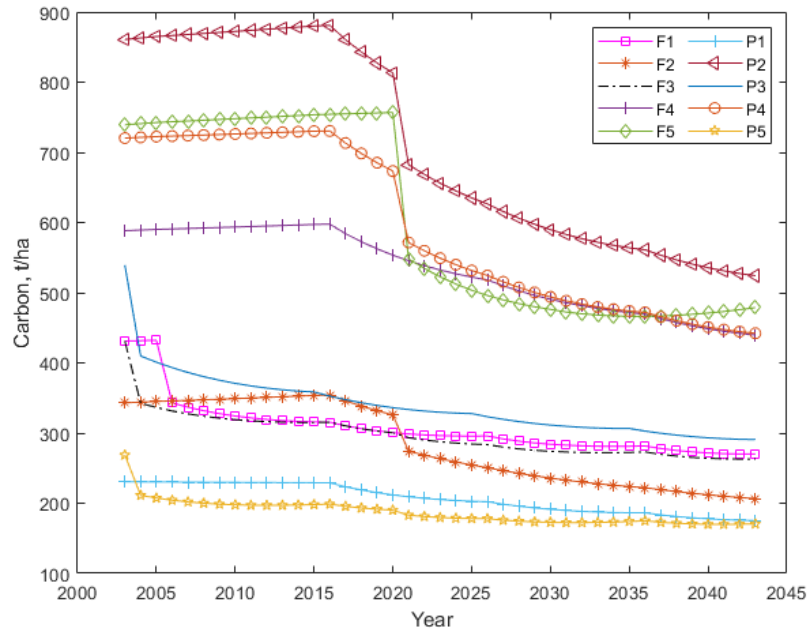


Figure 5.8: Ecosystem carbon stock predicted for natural forest

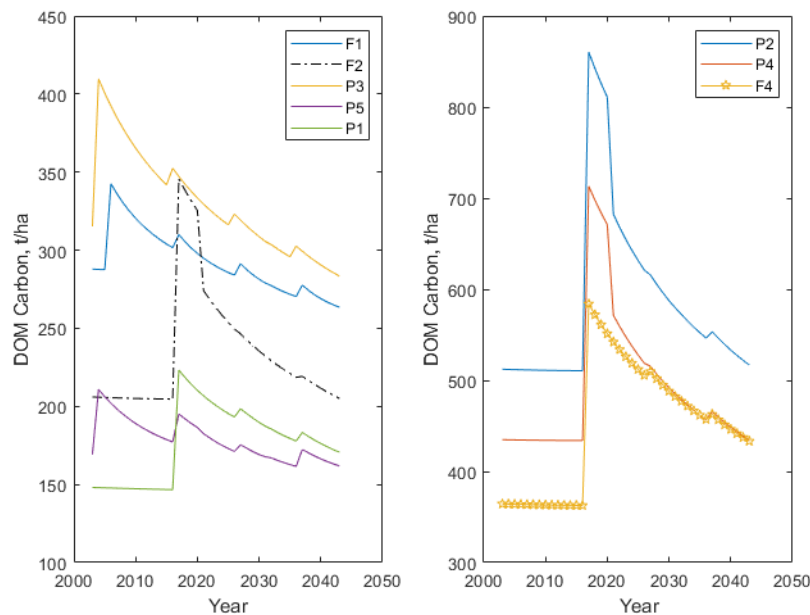


Figure 5.9: Dead Organic Matter for sites, barring conversion to agricultural land though experiencing disturbances

The dead organic matter always shows a change following a disturbance. For instance, in the forest annexing site P2, when there was a wildfire in 2003, there was an increase in the dead organic matter carbon from 315 to 400tC/ha. In addition, after insect damage, there is an increase in DOM carbon from 342 to 352tC/ha. Also noted is the impact of different disturbances on the forested areas. For example, forest sites near P2, P4, and F2 experienced harvest and insect

damage, whereas sites by F4 and P1 experienced only insect damage. The slopes of DOM carbon change for P2 and P4 were -11.5 and -9.3 respectively, while the latter group was -5.4 and -1.8 following the initial disturbances. Similarly, when there is fire and insect damage as in sites F1 and P3, the slope of carbon change was between -4.0 and -6.1.

Comparing the baseline ecosystem (the nearby natural forest) to the total ecosystem carbon in areas which are being used for agricultural purposes, Figure 5.10 below illustrates the difference in amounts. It shows that, for instance, on site P3, there is a -46.5% difference between the baseline and the pastureland, and that on F1, the difference is -7.6% between the baseline forest and the farmland, and at site F4 it is -64.9%. Similarly, the difference in GHG emissions for these scenarios were -74.8%, -44.9%, and -97.7% respectively in CO₂ amounts, as well as -99.9%, +0.05%, and +100% in CH₄ amounts. The GHG values on field P3 are very high because there was a fire in that area in 2003. The CH₄ amount simulated in all the years after is negligible. Site F1 was converted to a farm field in 2017, and the total CH₄ emissions thereafter was 0.02tC/ha, and 41.42tC/ha of CO₂ in the next 27 years under investigation. However, F4 was a wheat field the entire duration of the simulation, compared to a mixedwood forest which never underwent fire nor harvest. Thus, the CO₂ emitted from the forest is ~98% higher than what was produced on the farm field, while the CH₄ emissions are negligible in both scenarios (farm or forest), but the farm has a total of 0.03tC/ha over the four decades to the forest's roughly nil CH₄ emissions. The carbon sequestered by the forest is also ~65% more than that on the farm.

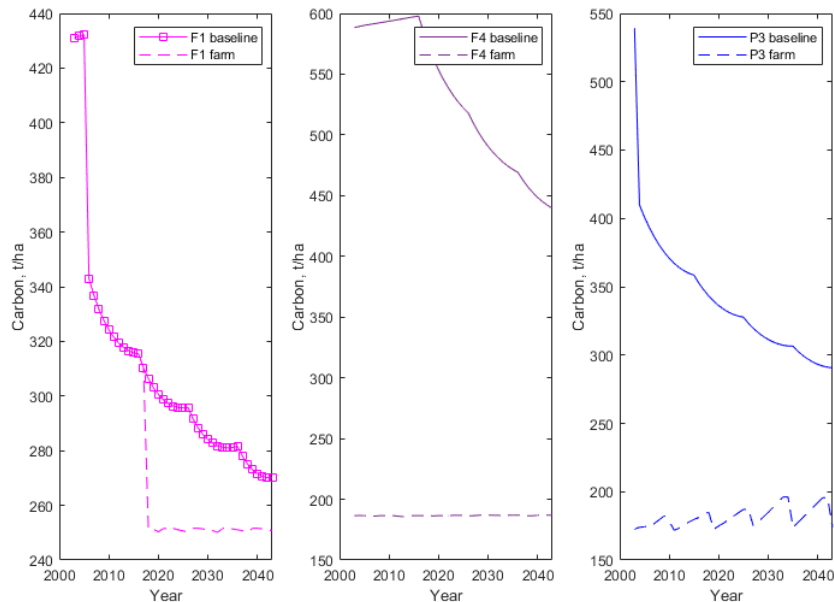


Figure 5.10: Comparing the baseline forest carbon to some agricultural sites

It has been demonstrated that the carbon cycle on land would be considerably modified by land use and land cover changes [199]. In particular, forest ecosystems are vital as carbon sinks. Therefore, if land use change occurs, e.g. from natural forest to managed pastureland or to farmland, there might be a loss of sink capacity in the ecosystem [199], [200]. Due to the various interactions within the carbon cycle, solely the emissions from the soil cannot determine if an area would be a net carbon sink or carbon source. A forest ecosystem would be deemed a carbon sink if it absorbs more carbon from the parts of the carbon cycle than it emits. In Figure 5.11 below, I illustrate the carbon sources in the forest stands. (The figure shows the amount of carbon simulated when the forest remains unconverted.) Prior to disturbances, the carbon stocks are more in the biomass than the DOM. Following disturbances, the biggest contributors to the carbon stocks are the litter and soil carbon in the mixed and deciduous forests. For the coniferous forest, the dead organic matter and soil carbon were the largest sources of ecosystem carbon following the disturbances. Three sites are shown in Figure 5.11 below, with a conifer forest, a deciduous forest, and a mixed forest.

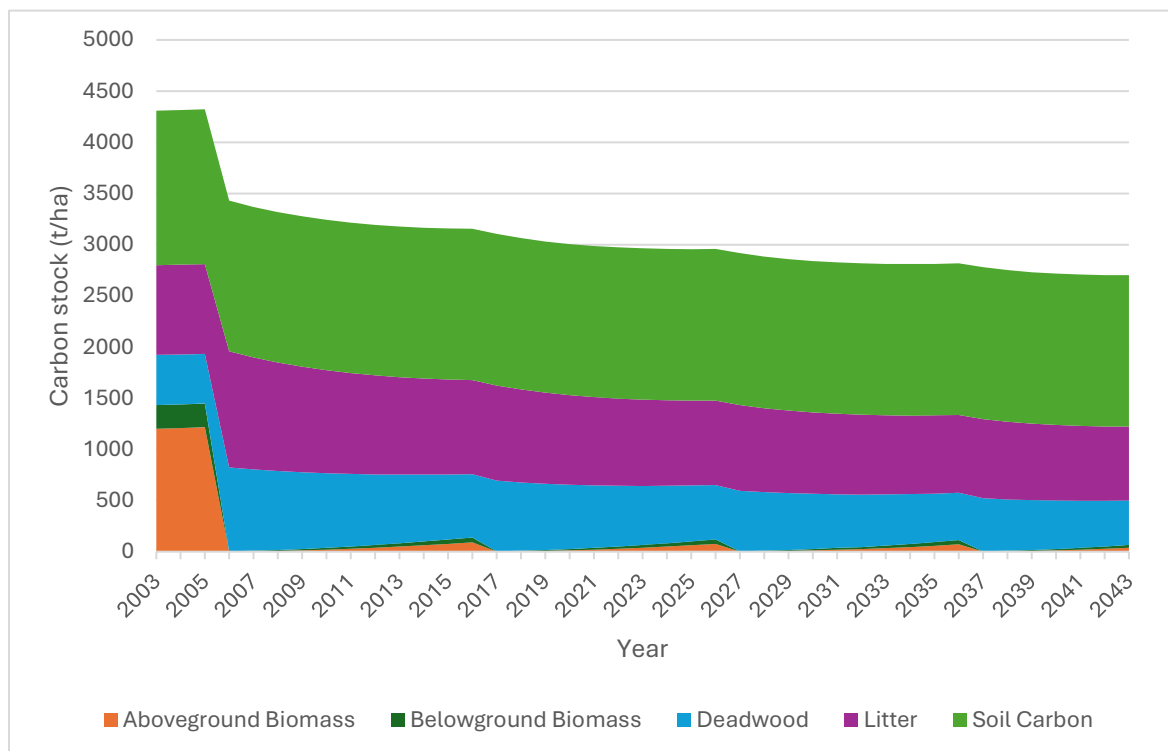


Figure 5.11a: Sources of ecosystem carbon from deciduous forest, by site F1

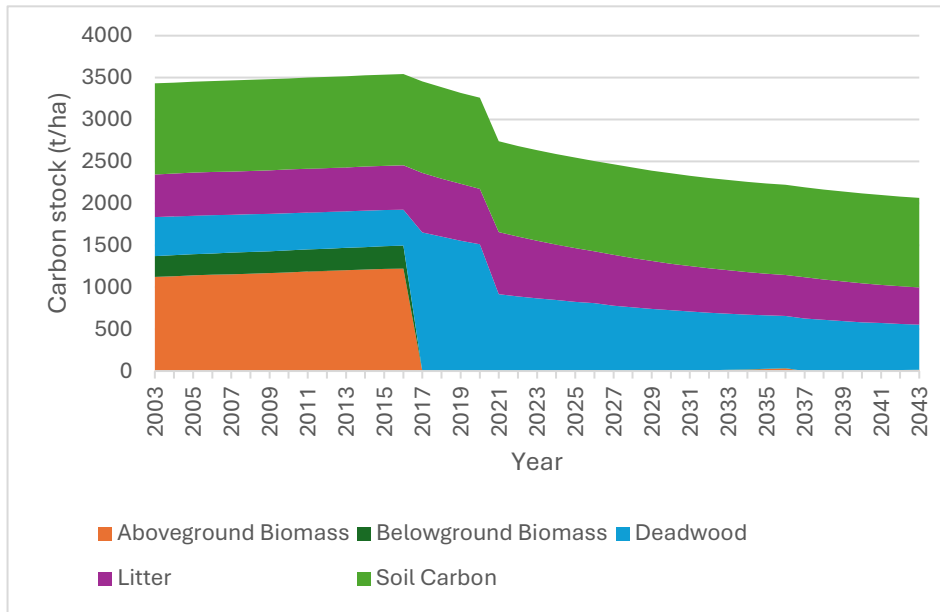


Figure 5.11b: Sources of ecosystem carbon from coniferous forest, by site F2

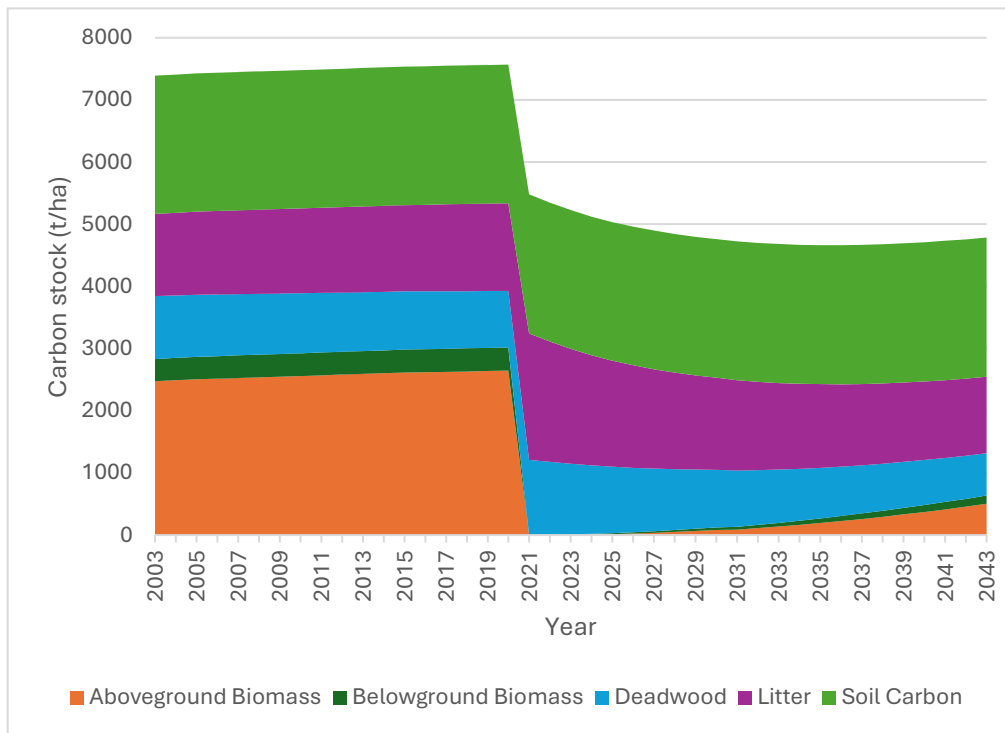
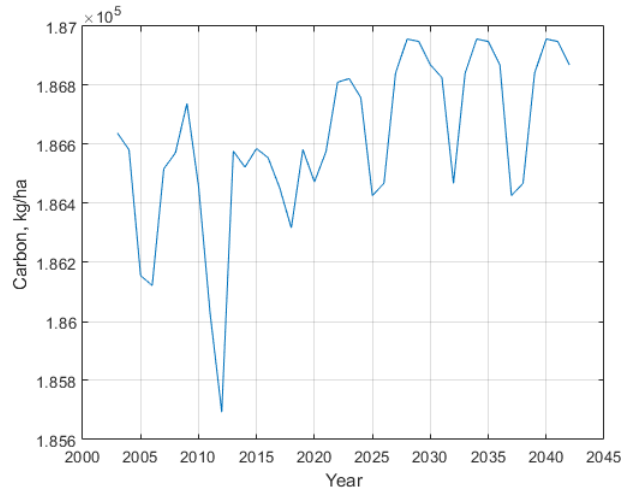


Figure 5.11c: Sources of ecosystem carbon from mixed forest, by site F5

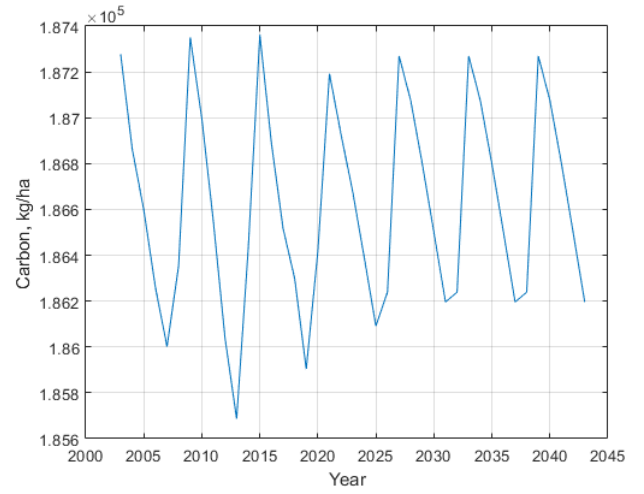
Considering the agricultural lands, simulations were performed to determine the sequestered soil organic carbon and the GHG emissions if the land usage was persistently agricultural for the four decades. Figure 5.12 below illustrates how the soil carbon and CO₂ amounts varied over the years. It is observed that there was always a drop in soil carbon and GHG

emission values during the sixth year when the croplands were left fallow. Furthermore, the sites P2 and P4 had the same animals grazing over the simulation period, with the only difference being the type of grass planted on the sites. Their difference in soil organic carbon was, on average, a difference of 3.85%/year, i.e. a total of 27,098kgC/ha over the four decades simulated. Similarly, P1 and P3 both had cattle being grazed each year, with identical growth patterns of the herds. The grass type was alfalfa in both areas, and the difference in the sites was that the animals being raised were dairy cattle on site P1 and beef cattle on site P3. The GHG emissions total in the former site was 633,652kgC/ha of CO₂ and 0.00985kgC/ha of CH₄, while the beef cattle's total emissions over the simulation period was 620,348kgC/ha of CO₂ and 0.01298kgC/ha of CH₄. (It should also be noted that the soil properties were different at the sites, e.g., different pH, clay percentage, and SOC amount at the surface.)

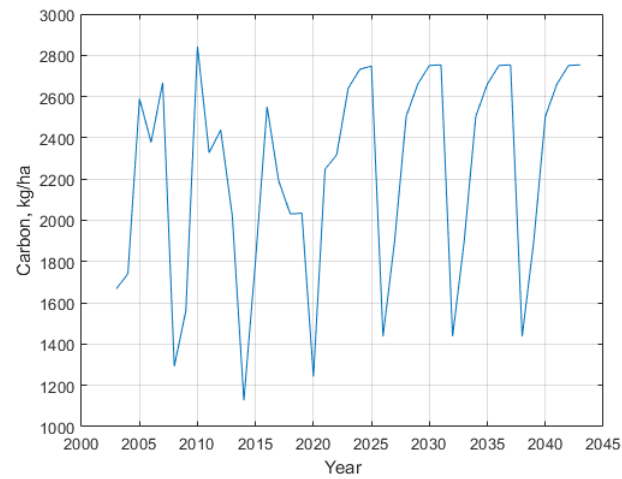
Assessing the baseline forest simulations, the mixedwood forests have considerably more ecosystem carbon than the forested areas consisting of single-species: whereas the uniform-specie areas have between 170 and 431 tonnes of carbon/ha, the mixedwood forests range in ecosystem carbon from 291 to 881t/ha of carbon. In fact, throughout the period simulated, apart from the forest by Site P3, the lowest amount of ecosystem carbon is a little less than 450t/ha of carbon, whereas the highest amount of ecosystem carbon in the single-specie sites was 431t/ha during the period under study.



(a)



(b)



(c)

Figure 5.12: Simulated soil organic carbon in (a) winter wheat field, (b) spring wheat field, and (c) CO₂ emissions on oat farm

5.2.2 Field survey results for validation

As mentioned earlier, field survey data was acquired in 2018, 2019, and 2021. This data from various ecozones also served as validation for the simulated carbon from both models.

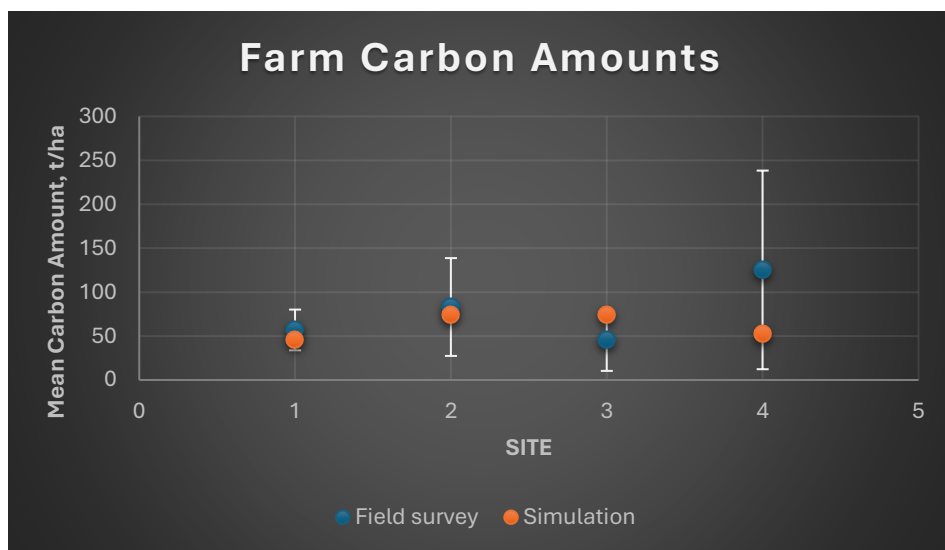
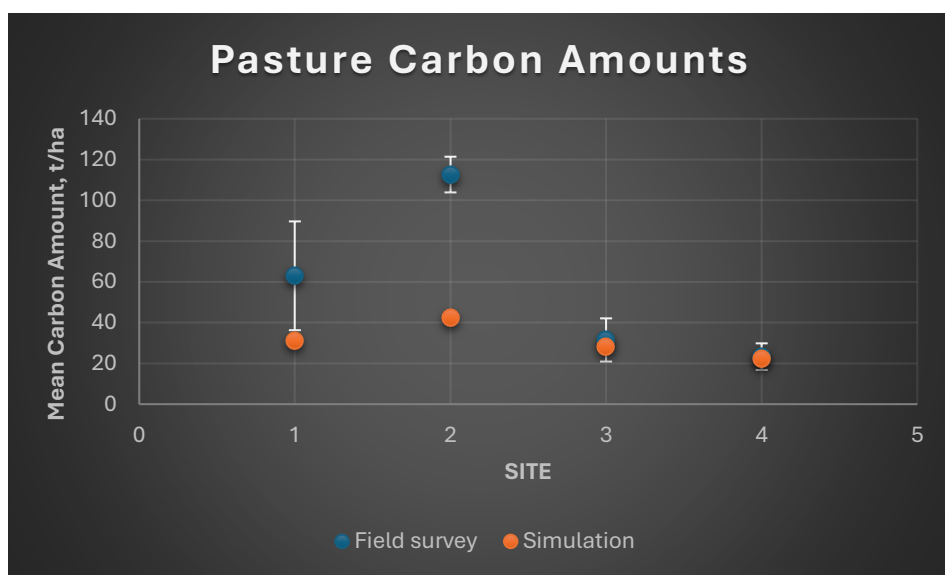


Figure 5.13: Comparing SOC from field survey to carbon estimates from CBM model

I highlight the comparisons in the carbon estimated from some of the sites in Figure 5.13, taking the pasture and farmlands as examples. The farm labelled #3 had its conversion in 2017, according to the high resolution satellite imagery examined. That corresponded with what we observed when we visited the site: it was a clear-cut field, ready to seed. The pasture site #4 was an oat field when we visited for our field campaign. Obtaining some local knowledge to combine with our remote sensing techniques – having the area’s history – allows us to better simulate the

processes which occurred, leading up to the amount of carbon stock we estimate at the current time.

During our field visit, the pasture site #2 was a mix of different grasses. It was a livestock pasture, and we observed about 30 heads of livestock grazing in the area. Performing the comparison of the calculated amount of soil carbon from our field studies to the simulations done, in all the cases, the results were comparable within standard deviation except at pasture site #2. The discrepancy might have arisen because, at that location, we did not ask the farmer how often he did his grass cutting nor confirm the livestock information. The difference in soil carbon estimated at that site from the simulation vs. the amount determined by soil analysis underscores the importance of obtaining sufficient and correct data when performing simulations. Furthermore, having some local knowledge to combine with our remote sensing techniques (to learn about the history of the study area) allows us to simulate better the processes leading up to the current carbon and GHG stock estimated. This, again, is a principle of our PAM method – to use the best available data in generating a depiction of the past and a prediction for future processes.

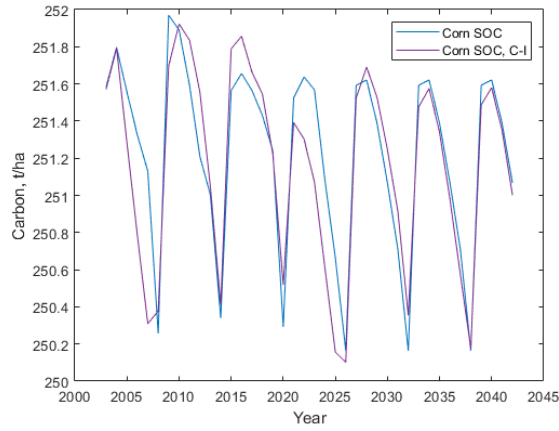
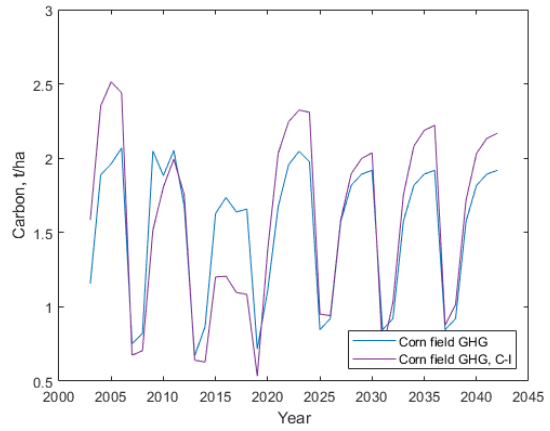
5.2.3 Sensitivity and Uncertainty Analysis

The uncertainty analysis in this study also served to examine the effects of climate change on the soil organic carbon or GHG emissions in the various land uses. For instance, in the DNDC model, the climate parameters of temperature and precipitation were modified in the Monte Carlo uncertainty run. The Monte Carlo test can be performed on different categories of variables like Crop parameters (such as maximum yield, water requirements, plant carbon/nitrogen ratio), Soil parameters (e.g., pH, porosity, bulk density), and Grazing management (such as grazing days and grass cutting amount). With the climate variables chosen to quantify the uncertainty of the analysis, a range was specified for the variables – for air temperature, it was 1.1 and for atmospheric pressure, it was 0.9. Thus, for a 5-year cycle, the Monte Carlo test was run 4000 times for each time step while varying both parameters. For example, the pressure was run having multiplied the original values by a number ranging from -0.9 to 0.9, and the temperatures multiplied by values which ranged from -1.1 to 1.1. For instance, when the temperature was 0.56 times the current air

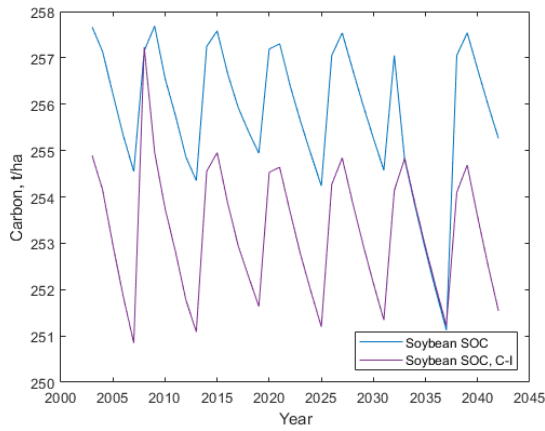
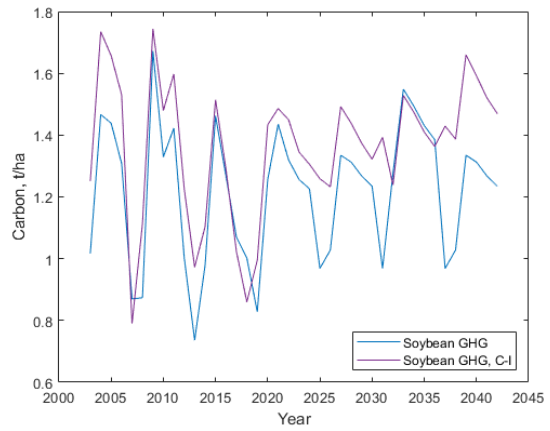
temperature specified in the input file, the pressure was 0.39 times. Or when the temperature was at the -1.01 factor, the pressure was at -0.83 of the original pressure value. The outputs of soil organic carbon and CO₂ emissions were then compared to the original values. Now, although the DNDC models terms the Monte Carlo simulations as an uncertainty analysis, it can also be looked at as being a sensitivity analysis. This is because by perturbing the parameters, the results indicate how sensitive they are to the soil carbon sequestered.

The impact of the changing climate was also examined. This was conducted by introducing a multiplication factor on some climate variables in the DNDC model. The carbon sequestered on the corn and soybean farm fields is shown in Figure 5.14 (a) and (b) (where the stocks simulated under the changed climate are represented as the climate-impacted part of the figures). The soybean soil carbon stock was mostly lower under the changed climate (average of -1.0%), while the soil carbon in the corn fields is -2.0% different on average. However, the CO₂ content on the soybean farm is mostly greater when the climate is changed (an average of +10.65% increase), while the corn fields differed in that respect (decrease of -0.55% CO₂ on average). Figure 5.14 illustrates the impact of the changed climate compared with the current climate scenarios.

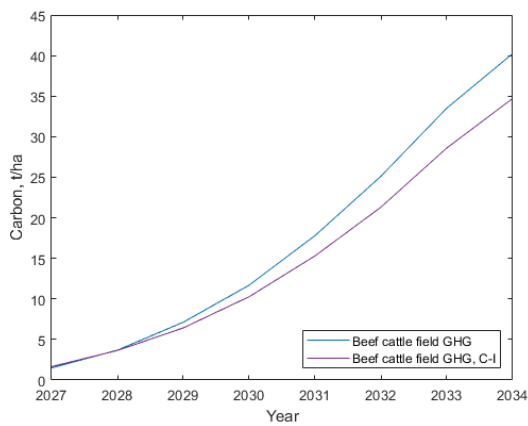
The pasture lands also saw differences in soil organic carbon and GHG emissions under climate change. Site P5, which has livestock and where farmyard manure application was simulated, showed significantly less soil carbon under climate change during the simulation period. Yet for the site P3 where beef cattle grazed, the SOC would be higher under the new climate regimes. Overall, the difference on average is +0.33%. The CO₂ emissions however are on average less for P3 under the changed climate conditions and more for P5, as illustrated in Figure 5.14 (c) and (d). For the 41-year study period, the difference in the GHG emissions was found to be less on site P3 under climate change by an average of -25.6%.



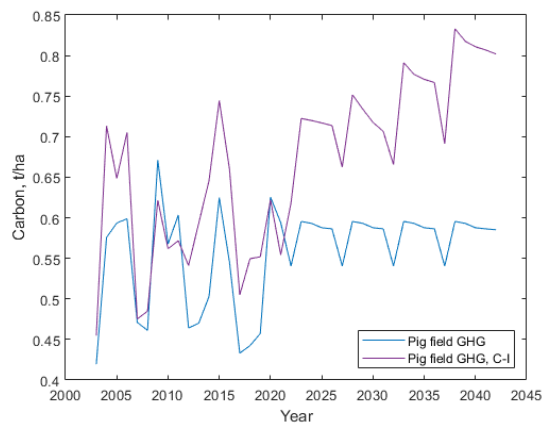
(a)



(b)



(c)



(d)

Figure 5.14: Comparing Soil Organic Carbon stocks and GHG (CO_2) emissions – simulated from historical records, and simulated as climate-impacted (CI) conditions – on different sites: a) of F1 corn farm field; b) F5 soybean farm field, and c) P3 beef cattle on pasture; and d) P5 pig farm on pasture

The average uncertainties of carbon stock estimated on the agricultural lands are not very high, as seen in Table 5.4, while that of the CO₂ emissions were much higher. Additionally, the ranges of CO₂ emissions uncertainties reveal great fluctuations that are not seen in the carbon stock uncertainties. The disparity in average uncertainty might arise because the GHG amounts are smaller than the carbon stock values, so there is a greater impact of differences when the comparisons are made between the original and the result from the uncertainty analysis. That might also play a role in the ranges of uncertainty; for instance, the spring wheat field's CO₂ emission's uncertainty ranged from -17.57 to +22.52, and the pig field's CO₂ emission's uncertainty ranged from -7.99 to +28.51. Table 5.4 lists the outcome of the uncertainty analysis in estimating SOC stock and CO₂ emissions.

Table 5.4: Average uncertainty of carbon and GHG (CO₂) stock estimates on the different agricultural lands

Crop	Soil Organic Carbon uncertainty	GHG uncertainty	Livestock grazed	Soil Organic Carbon uncertainty	GHG uncertainty
Corn	0.022%	0.553%	Beef cattle	0.332%	25.559%
Spring wheat	0.0073%	5.357%	Goats	1.549%	16.157%
Oat	0.129%	0.489%	Goats	0.860%	11.011%
Winter wheat	0.0169%	0.489%	Pigs	1.073%	15.506%
Soybean	0.102%	10.651%	Dairy cattle	0.478%	42.481%

To continue demonstrating that the methodology is easily accessible to an everyday user, the computation times of the simulations were recorded. Processing times were compared for the activity which took the longest time – the Monte Carlo uncertainty runs. The simulations for uncertainty analysis were performed on three different desktop computers. PCs 1 and 2 have an i7-6700 Intel processor, with CPU @3.40GHz; the only difference between them is PC1 has an installed RAM of 16GB while PC2 has 8GB of RAM installed. PC3 has an i7-8700K Intel Core processor, whose frequency is 3.70GHz. The installed RAM is also 16GB. All the systems have a 64-bit operating system. I did several Monte Carlo runs on each computer and noted their

processing times. The average time for completing the run on a 5-year cycle was 5.88hrs on PC1, 5.90hrs on PC2, and 5.05hrs on PC3. For the 1-year fallow runs, the runtimes were similar on the three computers, with processing time ranging from 57 minutes to 1hr and 33 minutes to complete. (A combined 23 fallow runs were tested across the three machines.) Thus, I observe that a standard computer can perform the simulations in this work, though higher-powered computers give a slight difference in processing time.

A test was performed using a carbon modeller to determine if significant differences would be observed from using approximated data inputs to the carbon modeller vs. more accurate data as obtained from the remote sensing tools, as well as model values for the growth and yield volume. The carbon stock was evaluated in the carbon modeller, simulating a period of 100 years. Figure 5.15 shows some outputs obtained: the approximated data always showed a higher estimate of soil carbon, total ecosystem carbon (Figure 5.15a), and the soil carbon (Figure 5.15b).

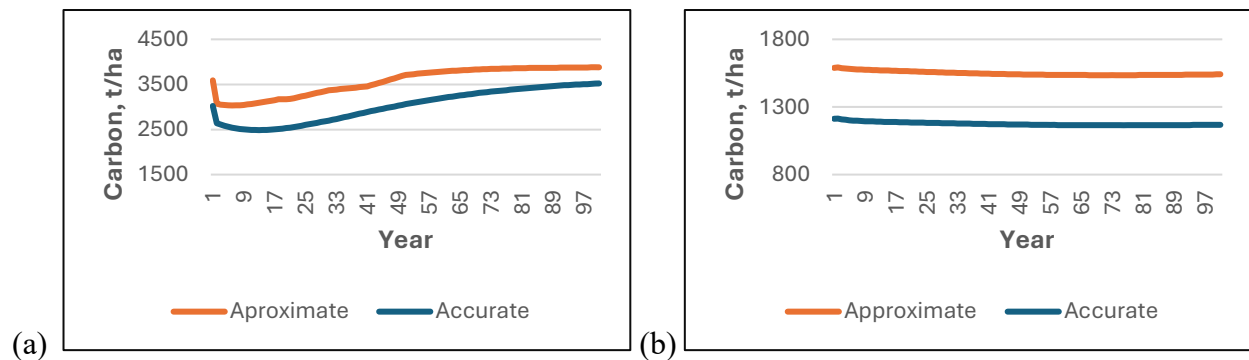


Figure 5.15: Comparing carbon estimates from (a) Total Ecosystem carbon, and (b) Soil Carbon, using approximate vs. specific input data on forest trees

Table 5.5 below also summarises the results obtained when approximate values from generic data are used as input to the carbon model instead of the more accurate data from classification maps obtained from remotely sensed data. The model overestimated the carbon and GHG emissions when approximate data was used.

Table 5.5: Comparing carbon and GHG amounts simulated under approximate and more accurate inputs

Stock	Approximate (t/ha)	Accurate (t/ha)
Total Ecosystem Carbon	360,068	304,919
Soil Carbon	156,413	118,696
GHG (CO, CO ₂ , CH ₄)	6,228	4,245

More specifically, we also used four sites to demonstrate the effects of accurate input data by comparing the SOC and total ecosystem carbon amounts when some of the disturbance data was omitted:

- On site by F1, simulation performed without disturbances
- On site by P5, simulation performed without disturbances
- On site by F4, only one species (white spruce) was simulated as being present
- On site by P4, simulation performed without land use conversion (whereas it should have been simulated as occurring in Year 80)

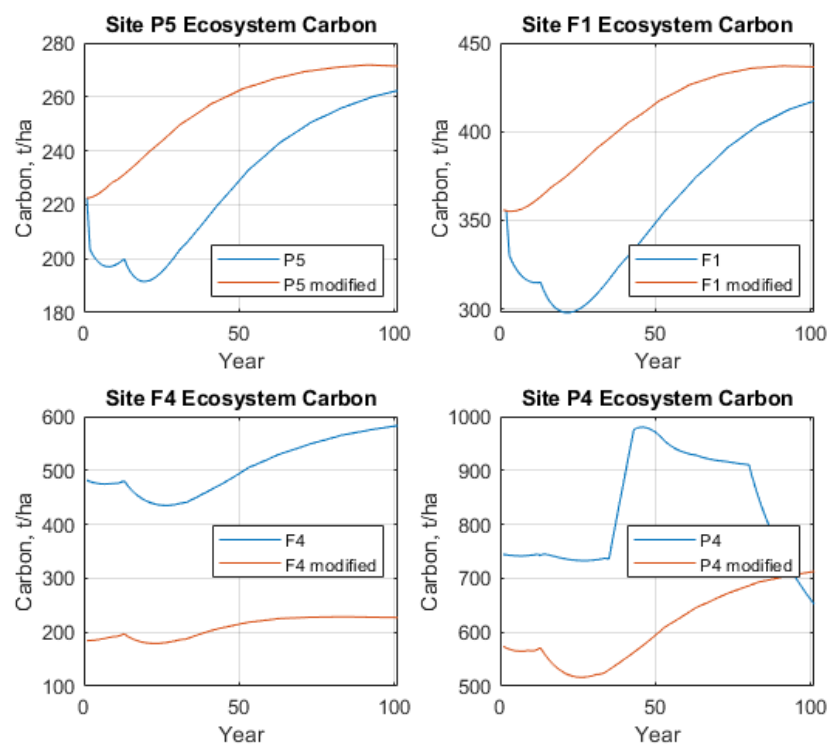


Figure 5.16 Comparing carbon estimates from accurate input data vs. estimates when disturbance and land use input data has been modified data on forested areas

In Figure 5.16, the total ecosystem carbon from the forests by the farm and pasture sites F1, F4, P4, and P5 are plotted vs. the estimated ecosystem carbon without the disturbances or correct landcover information. (The simulations done with incorrect information are labelled as the 'modified' sites in Figure 5.16.) When the disturbances were not included, the estimated

ecosystem carbon amounts were higher than the actual values – by 12.5% by site P5 and by 14.7% by site F1. This is expected, since the disturbances such as insect damage and wildfires create more dead organic matter, thus influencing the total ecosystem carbon amounts. On the other hand, having the forest by site F4 labelled as coniferous with only one tree species, rather than a mixed forest, there was a 58.6% reduction in the total ecosystem carbon over the period simulated. Similarly, when the deforestation by site F4 was not included in the simulation, the estimated carbon amount was incorrect by about 26%. The SOC comparisons showed a similar trend, with sites by P5 and F1 overestimating the SOC by 2.7% and 2.5% respectively. Such results from the uncertainty/sensitivity analysis underscore the importance of having the right input for the models. It is vital to provide realistic data, from the best available knowledge sources, as input for simulation to obtain accurate ecosystem carbon stock and emissions estimates.

SUMMARY

This work aims to contribute to the efforts of reducing uncertainty in estimation of the emissions by increasing the accuracy of input data, particularly the input of landcover or land uses. A study period of four decades was selected to determine the long-term carbon effect of the decision to convert natural forests in the Great Clay Belt of northern Ontario, specifically in the region where the NLP program is being implemented. For this study, the land uses following conversion were stipulated to be cropland and pastureland with livestock.

Future climate conditions were simulated to determine the potential emissions and ecosystem carbon stock under various conditions. (The possible disturbances which could occur, based on the disturbance history in the area, were also simulated.) Using the SSP1.9 pathway from IPCC which predicts the temperatures to be elevated above the most recent decade [201], future temperatures were increased by a factor to account for the effects of the changing climate. Similarly, the precipitation was simulated to be different in the future based on historical trends. Climate change is projected to cause changes in precipitation [202]. The simulations in this body of work were conducted knowing that unprecedented events could occur and that they cannot be properly modelled, however. For instance, in 2023, Canada's eastern province of Nova Scotia experienced three months' worth of rain in a day [203].

This study took the baseline case as a situation where the nearby natural forests are not converted to agricultural lands. Comparing the sequestered carbon simulated in those scenarios shows the forests to have resulted in 1.96 times more sequestered carbon in total. Furthermore, the forests continually act as a carbon sink, not a source, over the four decades simulated, as demonstrated in Figure 5.7. In reviewing the boreal forest carbon estimates around the globe (e.g. in Russia, Canada, Scandinavia and Alaska), the forest carbon ranges from 40-128 t/ha, while the soil carbon stocks range from about 125-339t/ha. That figure does not include the permafrost soils, where the soil carbon is up to 890t/ha [204]. Specifically for Canada, the soil carbon ranged from 139.9-175.8t/ha [204].

Now, it was observed that when the forests were converted to cropland, the average GHG emissions were 60.14tC/ha CO₂ and 0.781kgC/ha CH₄. Furthermore, the average ecosystem carbon sequestered in these scenarios was 9,252.9tC/ha. (The average ecosystem carbon simulated in the forested areas was 16,616tC/ha.) Amongst the crops examined in this study, soybean showed the most ecosystem carbon sequestered while winter wheat had the least. Other studies showed that conversion to cultivated land from forests similarly reduced the soil organic carbon [62], [172], and in pasturelands used for grazing [171].

In addressing the GHG emissions the country will experience in the coming decades, it is essential to examine the input from methane. A result in this work was that the beef cattle produced 1.32 more CH₄ emissions than the dairy cattle. Other studies have shown beef cattle produced 6.9 times the direct CH₄ emissions of dairy cattle [205]; beef cattle produced 2.9 times the dairy cattle CH₄ emissions [206]; and in 2012 in the U.S., beef cattle produced 71% of the total enteric emissions (mainly CH₄), while dairy cattle produced about 25% of the country's enteric emissions [207]. Now, dairy cattle emit less methane than beef cattle because the lactating cows intake primarily for milk production, while the beef cattle consume feed entirely for meat [208].

Pastureland was also simulated, with it being grazed for 10 hours a day from May to October. Under these conditions, the average GHG emission was 303.71tC/ha CO₂ and 0.334kgC/ha CH₄, with the sequestered carbon, being on average 7,708.7tC/ha. The pasture field with the highest emissions was alfalfa pasture with dairy cattle grazing.

Although continuous grazing needs less management and incurs minimal costs, rotational grazing requires more management and costs for watering and fencing certain parts of the fields.

However, the rotational grazing practice is more beneficial in terms of the quality of grasses in the pasture. Disadvantages of rotational grazing include the loss of an area for production each cycle and more work in managing the rotation system; however, the benefits of allowing the plants to rest, not losing forage from trampling, and having more efficient animal distribution and forage use would outweigh the costs. Intensive managed pasture grazing is also known to produce more meat or milk per unit of land [198] and control disease and pest disturbances to the livestock since they are constantly being moved away from fresh waste [209].

As a way to combat climate change due to atmospheric carbon, the Paris Agreement recommends maintaining or increasing carbon sequestration and storage in forests. However, under a changing climate, there could be challenges to maintaining forests as carbon sinks. Minor temperature increases (for instance, between 1°C and 2°C) would increase the rate of forest growth and productivity. Hence, there would be more carbon sequestration in the forests. However, temperature increases of more than 2°C can trigger a decline in growth due to higher mortality [210]. Therefore, the rising or reducing temperature could be used to examine the effect of climate change on carbon sequestration.

Since climate variables were chosen for the uncertainty analysis, the runs were also used to project how the land would respond to an increase in temperature (at 1.1 times the historical temperature) or a reduction in pressure (a fraction of 0.9 the historical or projected future values). Thus, these runs simulated the impact of climate change with a rising air temperature.

Several factors affect the simulation result. Typically, carbon stock predictions generate uncertainty from the carbon models and from the scenarios being simulated. Uncertainty from the carbon models has to do with its response and the model input. Therefore, it is important to use carbon models which have proven robust and well-tested. In addition, the carbon model user should seek accurate input data for a model's parameters and algorithms. In this study, remote sensing data was used to get realistic model input. The estimation of uncertainty within and between carbon models is still a complex issue; for instance, there is no consistent method for calculating confidence intervals in ecosystem models [121]. Yet, it is essential to make uncertainty estimates as it will enable policymakers to know where to direct resources to reduce uncertainties in carbon estimates. Furthermore, the results of carbon models can affect the adaptations made by

the policymakers under the changing climate. Hence, there is potentially a sizable economic value in reducing the models' uncertainty while predicting future carbon and emissions.

In this study, the estimated average uncertainty in emissions was 0.46% in carbon and 12.83% in CO₂ emissions from the agricultural lands. Other studies also vary in their uncertainties, e.g., $\pm 15\%$ of carbon [121] and $\pm 19\%$ uncertainty of emissions on cropland [201]. The conclusions drawn from this study are specific to the type of land use and properties of the agricultural rotations simulated. In addition, the disturbance-related processes introduced in the simulations were based on historical trends in the region. Thus, if different parameters are introduced, the uncertainties calculated would likewise differ.

Comparing the carbon stock and GHG amounts estimated by the carbon models when approximate values from generic data are used as input to the carbon model as opposed to the more accurate data from classification maps obtained from remotely sensed data, a pattern was revealed: the more accurate data input yielded lower amounts. I.e. the total ecosystem carbon was overestimated by 15.32%, the SOC by 24.11%, and the GHG emissions were overestimated by 31.84%. The carbon models used in this study were used in accordance with the landcover type they were designed for (forested or agricultural). The CBM model in particular is used for Canada's carbon accounting [121][122] to fulfil the reporting requirements of the IPCC. This work adjusted two climate parameters to observe their effect on the simulations. Yet, further work should be undertaken to investigate the sensitivity of the carbon models to other parameters.

Some typical limits in the carbon simulating would be the input information, with landcover being one of them. This limitation is also an opportunity: to improve the prediction of carbon and GHG by improving landcover input. Uncertainties still exist; for instance, the DNDC model requires the exact specification of the application of fertilizer, or the irrigation scheme being used. There could also be input data uncertainties to the models from disturbance events, forest inventory, or forest transitions which occurred. However, these uncertainties should be minimised by more accurate data input where possible.

As noted in the study, gathering information from the best available sources yields more accurate results from the carbon simulators. Thus, it would be beneficial to adopt inclusive processes involving local knowledge as well as Indigenous Knowledge in seeking land use history and disturbance history. Furthermore, the local and Indigenous knowledge sources could

contribute towards recommending better land management practices for enhanced carbon sequestration.

It was observed that the area under study acts as a carbon sink, and carbon sequestration increases under the effects of a warming climate. Thus, carefully planned land cover transitions could also be an effective strategy for carbon sequestration. This information is important for reporting agencies, the agricultural industry, policy makers, residents, and public development in this area and beyond as Canada aims to meet sustainability goals. Furthermore, such analyses can be utilized in other regions which are considering or are undergoing similar land transitions.

6. CONCLUSIONS AND FUTURE WORK

This work interprets the initial landcover type and the land history from advanced remotely sensed data – the satellite imagery from Landsat and Sentinel sensors. Additionally, it predicts the potential emissions and carbon amounts under various land management schemes that could occur in the Clay Belt of Ontario, Canada, as the region undergoes a transformation to expand its agricultural land. The study examined how the carbon and GHG emissions would differ from the ‘baseline’ – the area having not undergone any transformation. A progressive and adaptive framework was created, which can be adopted in other regions. This operational framework I developed integrates the best existing geospatial products and knowledge to assist in detecting landcover and land use change as an input to carbon and GHG models.

The study initially focuses on large-area mapping. However, the method generated in this work can be implemented in other areas of diverse landcover/land use types. The two carbon models leveraged in this study (the Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3) and the DeNitrification-DeComposition (DNDC) model) provided insight into the carbon and GHG amounts expected under different land management schemes if a forested area in the Clay Belt is converted to agricultural usage.

6.1 Conclusions: Considerations from Carbon Accounting in the Great Clay Belt

6.1.1 Impacts of a Changing Climate

Globally, the year 2023 was one for the history books: New temperature records were set with it being the hottest year in recorded human history [211][212] and having the most atmospheric GHG concentrations; the El-Nino produced exceptional ocean heat waves [212]; there was widespread flooding, e.g., the atmospheric rivers causing billions of dollars worth of damage in California [213], torrential rainfall intensified by El-Nino killing hundreds and displacing millions in East Africa [214]; and the global air temperatures were so unusually different from previous years that for the first time on record, each day of 2023, the temperature was at least 1°C above the 1800-1950 pre-industrial level. The boreal summer was the warmest season on record,

and temperatures in Europe were above average for 11 months of 2023 [211]. Canada also became a nuisance to its neighbours in the south when its wildfires billowed smoke to various cities in at least 15 states, causing havoc such as travel disruptions, facilities' closures, and event/entertainment cancellations. Furthermore, there was widespread exposure to multiple air pollutants [215]. The Canadian 2023 wildfire season was also record-setting: the 45.7 million acres burned was roughly equal to the total area burned in the previous seven years in the country [216]. All told, roughly 2,200,000,000t of CO₂ was released into the atmosphere [217].

Against this backdrop, one can review the 2030 target Canada has set of reducing emissions by 40-45% below the 2005 levels [45] and Canada becoming a net-zero carbon emitter by 2050 [40]. (For context, 2005 and 2007 are tied for Canada's highest GHG emissions per capita (23t CO₂ eq/capita).) All levels of government will need to make a concerted effort if the 2050 goal is to be achieved. Furthermore, drastic decisions might have to be established to ensure that the requisite policies towards reducing emissions are initiated. Constant assessment and reporting will reveal if Canada is on track to reach the set targets.

Canada's greenhouse gas emissions have hovered above 700 Mt of carbon dioxide equivalent in the past two decades [46]. The pandemic in 2020 influenced the severe reduction in GHG emissions that year, causing a net decrease of 7.4 Mt from electricity/heating production and 27 Mt CO₂ equivalent from transportation [46]. Yet, two of the five most significant sources of Canada's GHG inventory come from Agriculture and Land-Use Change and Forestry [46]. Given the recent thrust to convert forests in the Clay Belt to agricultural land, there is a vital need to assess the attendant effects on carbon and GHG emissions under these scenarios. This estimation can be made with computer carbon modellers.

6.2 Contributions

This dissertation focussed on how the forest air and soil properties transform over time from various land disturbances, and how subsequent management schemes affect the environmental properties such as GHG emissions and soil carbon stock. Prominent in the research was the systematic investigation of methods to use the best available geospatial data to detect changes in the land covers more precisely, given the region's limitations – Northern Ontario often lacks reliable, accessible, and continuous geospatial information.

It is widely understood that the fight for the environment includes decreasing the presence of carbon or carbon dioxide in the atmosphere. Through the results of this study, a better and more accurate understanding of the sources of carbon in a forested area was generated. Thus, it is shown which forest type contributes the most to carbon stocks. Furthermore, the results indicate where more effort could be placed in mitigating the effects of climate change by supporting more carbon sequestration. In the face of many contradicting voices on the subject of the changing climate, its effect on human lives, and how individuals or nations should address it, it is imperative that further efforts must be made to quantify the outcomes of climate change in accurate terms. This work is a contribution to the body of knowledge which seeks to improve the accuracy of carbon models, which in turn will assist policy makers and individuals make decisions grounded in better information.

The main contributions of my research are summarised as follows:

6.2.1 On Evaluating and Improving Methods of Landcover Classification

In this study, I proposed a novel approach to detecting change in land use and/or landcover type. The approach entailed an automatic change detection process. The method uses a hypothesis test to determine where landcover change has occurred. Many methods of change analysis use a threshold method, though hardly any use a hypothesis-based test for such landcover change analysis, to the best of my knowledge. The developed method was validated based on data collected over a large area in Ontario's Clay Belt.

The study is also significant because NDVI and LST are used together to predict landcover changes. To my knowledge, such experiments are not widely seen in literature, though the features have been used individually for the purpose of predicting landcover change. The result from this study indicates that it is a valuable area to continue developing since its outcomes are similar to existing and established change detection methods.

This aspect of the study was published in Ituen, I & Hu, B. (2021) An Automatic and Operational Method for Land Cover Change Detection Using Spatiotemporal Analysis of MODIS Data: A Northern Ontario (Canada) Case Study," *ISPRS International Journal of Geo-Information*, vol. 10.

6.2.2 On Exploring Effective Methods for Land Use and Landcover Change Detection

From a literature survey conducted, it is rare to encounter studies which consider multi-sensor image classification for estimating soil carbon flux, which could in turn increase the accuracy of carbon estimates. My work adopted the fusion of multiple satellite datasets to generate landcover class knowledge. In exploring the object-based Random Forest classification, the Simple Non-Iterative Clustering (SNIC) algorithm was utilised. Spectral and textural features from the optical and SAR sensors were inputs to the classifier, and the classification accuracy was over 89%. However, the pixel-based classification yielded a higher overall accuracy than the object-based classification.

The multisource landcover classification technique endeavoured to use the best available data for the area's training and subsequent land classification. Furthermore, permanent water body maps were employed to avoid the effects of seasonality in taking training or validation samples from the water bodies. Also addressed was (i) a way to mitigate the problems of an area facing perpetually cloudiness (using the percentile composites); and (ii) performing annual change analysis without gathering new training samples when classified data from existing regional or national datasets are unavailable. When a comparison was made on the forest modeller for carbon and GHG emission estimates using nominal data as input vs. the more accurate input data for the region, the total ecosystem carbon was overestimated by 15.32% when generic data was used. Similarly, the generic data overestimated the SOC and the GHG emissions by 24.11% and 31.84% respectively. These results illustrate the need for employing carbon estimate methods such as the one proposed in this study – to generate more accurate results that will assist decisionmakers in their roles.

Regarding the change detection methods explored, I developed an effective disturbance-informed change detection methodology to address the data limitations in the study area. Typically, remote sensing studies only use remote sensing data for change assessment. However, that might be inadequate where continuous or regularly updated data is lacking, as sometimes occurs in the Clay Belt. Although many existing classified map products are used, my method does not rely solely on remote sensing data. Instead, I use the best available data input from all the products. If there is a conflict between classified maps, remote sensing data such as Lidar or other higher

resolution data could be used to determine the land cover type. The Progressive Adaptive Method I created also draws on nascent technology such as machine learning and Google Earth Engine's cloud computing to determine the trend of landcover and land use changes.

This study's contribution to scientific knowledge will aid in predicting the related GHG emissions over large areas when land usage is transformed. With these emission estimates, decisionmakers can choose the most sustainable routes in land transformation or development.

6.2.3 On Characterising the Impact Land Use Change has on GHG Emissions and Soil Carbon Stock

I integrated two carbon models to generate a valuable methodology for estimating SOC and GHG emissions in natural forests and converted land, e.g., cropland, shrub/pasture, land for livestock usage, etc. I also contributed to reducing uncertainty in the estimation of the emissions by increasing the accuracy of input data, particularly the input of landcover or land uses. The primary source of the landcover and land use data came from remote sensing data. To demonstrate the method's effectiveness, the estimated soil carbon amounts from four pasture sites and four farm sites were compared with field survey data. All the results were within standards of error, except the site where incomplete information regarding the livestock was received.

It should not be assumed that the conversion to cropland will necessarily produce more carbon and GHG than a livestock farm, but the crop rotation chosen for this study as well as their management have affected the outcome. In addition, periods of fallow were incorporated in the simulations. Not surprisingly, the emissions were reduced in the fallow years. Furthermore, there was less sequestration during the fallow years. If portions of the Clay Belt are converted to croplands, the farmers might choose a more diversified agri-food production. That, in turn, will affect the soil carbon and GHG emissions which arise. Further research will examine different land management scenarios following initial conversion from forested land, e.g., a reversion to forest cover if agricultural land is abandoned. Such findings are significant as they support a comprehensive understanding of the changes which could occur as Ontario's forests are converted into agricultural land over the coming years.

The publications generated from this aspect of my study were:

Ituen, I & Hu, B. (2024) “Assessing the Impact of Land Conversion on Carbon Stock & GHG Emissions,” *Land*

Ituen, I & Hu, B. (2021) “Estimating carbon and greenhouse gas emissions in remote regions of Canada,” *ISPRS Conference*.

Another contribution which I made to the body of knowledge was the creation of applications for users to explore the historical landcover changes in their region. The first app I created I dubbed The Landcover App. Users can choose any year from 2000 to 2023 to observe their region of interest. A user can see how a landcover class could shrink or increase over the different years, and perhaps which region has more landcover changes than others. The app also provides a link to my publication in the *IJGI* journal [70], which will give the reader a background on the MODIS IGBP landcover classes as well as change detection using remote sensing tools.

The second application is named the Forest Change App. My publication of change detection using statistical methods [70] drew on the forest change prediction from the Hansen et al. dataset [218] to validate my work. The Forest Change App permits analysis of remotely sensed images to characterise forest extent and change. The app graphically displays the predictions made for five of the forests in the Clay Belt – the Gordon Cosens forest, the Hearst forest, the Abitibi forest, the Romeo Malette forest, and the Night Hawk forest. The app is prepared with data from 2000 to 2020, and the user will select the year of interest to examine the forest data. The options the user can choose from are forest loss for the year, the forest gain, or the percent tree cover in that year. To aid in quantifying the data being visualised as maps, the user can select the region of interest and then view the data (e.g., of forest loss that year) in a graphical form. The graph will compare the forest loss for the year to the forest loss of previous years, to give context to the data. Furthermore, the Forest Change App graphs could buttress the changes a user observes of the landcover changes from the Landcover App. (The Forest Change App is based on a time series of Landsat images of 30m resolution.)

The creation of these user-friendly applications (which have remote sensing data at the backend) was done within the theme of this study – providing an interactive tool which is functional, easy to use and deploy, cost-effective, and able to be used in remote regions. After users

explore their area and surrounding regions for the landcover changes, their interest and awareness of the impacts of land use change on GHG emissions could increase. Furthermore, the apps could be used in discussions for predicting future emissions due to current or planned landcover changes. Thus, the target audience of these applications is not only government officers and policymakers, but also eco-enthusiasts and farmers interested in their region's land transformation.

Overall, my research and findings have contributed to the knowledge base on improving existing landcover change detection methods and developing methodologies for characterising the impacts those changes have on emissions and soil carbon stock. My research has demonstrated the usefulness of advanced remote sensing technology whereby multisource and multitemporal remote sensing data provides reliable input to reduce the uncertainty in carbon model estimates.

6.3 Future Research

It is desired that the current work will be a launching point for further study in the subjects raised herein. The following research directions are suggested for future research:

- i. A significant focus of this study was how to best integrate remote sensing data and existing products such as the geospatial data obtained from government databases to answer important questions under climate change. Although beyond the scope of this work, it would be worth examining further the complexity of the drivers of disturbances. Furthermore, the difference in soil carbon sequestered based on the severities of different disturbances could be explored as a future study.
- ii. Future studies could examine the extension of the proposed classification and change detection methods to other data sources from different satellites. As higher-resolution images become available and provide further historical context, data from satellites such as PlanetScope or Worldview can be explored for more accurate landcover change detection. Assessments can be made to determine if the current methodology would be applicable, or what amendments need to be made (e.g., integration of data on the GEE platform) to ensure high-accuracy inputs to carbon and GHG models.
- iii. An importance-informed landcover classification method can be investigated whereby the features of low importance (as ranked by the Random Forest classifier) are identified to be

excluded from further processing, whilst maintaining high overall accuracy of classification. That could simplify classification and change detection efforts, which is necessary for analysts with fewer resources. It would also be beneficial if there is limited access to high-fidelity data across the study region.

- iv. This work also only considered the collection of training points from datasets of 30m spatial resolution or less. As a future assessment, classification efforts can be conducted by finding training data from higher-resolution datasets such as Dynamic World. This data product is a near-real-time land use/land cover classification dataset developed from the 10m Sentinel-2 imagery by leveraging deep learning.
- v. For a future study, carbon amounts in the forests will be calculated not only up to the point of harvest but also afterwards: calculations of carbon and the mitigation potential could be improved if analyses assume that carbon is emitted at harvest, or that the products created from the harvested wood are not carbon neutral. Some of the emissions in wood products being tracked in carbon models include carbon from harvesting trees or crops to their final decay, from thinning to logwood, pulp for paper, and slash (plants/logs left on site, thus entering the soil pool). A modified version of the CBM model (the model used in the current study) makes provision to estimate carbon in harvested wood products [210].
- vi. This study did not consider temporal differences, for instance, not considering if there had been two fire disturbances within a year. A disturbance that occurs frequently within a model time step should be simulated. In addition, I chose an aspatial model for this study. This model only simulates stand- and landscape-level forest dynamics. If the spatially explicit model was used, I could simulate the carbon and GHG emissions to recognise adjacencies and spatial relationships. This modelling was also based on plots, where each plot is assumed to be a homogenous area. While the aspatial CBM carbon model is set in units of hectares, the input remote sensing data for classification and disturbance is at a different scale, e.g. 30m spatial resolution. Classification and change detection were done at those scales to obtain the best available input data to the models – the high-resolution data for accurate landcover class distinction. Effort was taken to ensure input data accurately represented the region under study with the classification and disturbance data input to the model. Future work will seek to examine additional strategies to address the spatial and temporal mismatch and thereby reduce uncertainties.

- vii. Another future study direction would be to focus on specific landcover types in considering the effects of land use/landcover transformation on soil carbon and GHG emissions. For instance, peatlands are characterised as having large soil organic stock. This landcover type could be investigated. Interestingly, one-third of the world's peatlands are in Canada's terrestrial ecosystems [37]. Similarly, Wetland is a significant land cover type in Canada: it has been estimated that 13% of Canada's terrestrial area is covered by wetlands [187], and 25% of the global wetlands are in Canada [188]. (Another study's land classification estimated Canada's landcover to be 36% wetland [186]. However, wetlands are dynamic, even with the different seasons; hence, the discrepancy in classifying this cover type is unsurprising.) Honing in on landcover types such as wetland and peatland could yield new knowledge regarding the impact of changes to these cover types, related to carbon sequestration and atmospheric emissions.
- viii. It was acknowledged that artificial neural networks deliver very high accuracy in landcover/land use classification. However, it has also been observed that creating a training dataset for these networks is "cumbersome and delicate" [145], and several times the training areas are selected through photo-interpretation [145]. Since the goals of this study were to have an easily deployable method and provide ease of operation of a classification and change detection algorithm that yields highly accurate inputs for carbon modellers, a deep learning solution is currently not viable to meet those goals. Therefore, traditional machine learning was finally chosen as it is not dependent on a complex method nor massive amounts of training data, which eco-enthusiasts, farmers, etc. might not have resources for – as, apart from policymakers, that is whom this work is targeted at. Future investigations can be pursued to leverage artificial networks for carbon and GHG simulations.

Overall, I seek to explore further areas in which remote sensing data can improve the input to carbon models to obtain more accurate and accessible carbon estimates, in support of climate change mitigation and adaptation planning.

REFERENCES

- [1] Beef North, “Cow herd expansion in Northern Ontario.” [Online]. Available: <https://www.beefnorth.com/cochrane>.
- [2] Legislative Assembly of Ontario, “Report Commission of Enquiry Kapuskasing Colony – 1920.” [Online]. Available: <https://archive.org/details/kapuskascomm00onta/page/n7/mode/2up?view=theater>.
- [3] Ontario Grain Farmer, “Farming the Great Clay Belt,” *Ontario Grain Farmer*, 2017.
- [4] Government of Ontario, “Minister’s Order: Northern Livestock Pilot Program.” [Online]. Available: <https://www.ontario.ca/page/ministers-order-northern-livestock-pilot-program>.
- [5] J. Leal, “Province makes progress in Claybelt livestock expansion,” *Ontario Farmer*, Feb. 2018.
- [6] Ontario Ministry of Agriculture Food and Rural Affairs (OMAFRA), “Proposed Pilot Sites,” 2018. [Online]. Available: <http://www.omafra.gov.on.ca/english/policy/nlppilotsites.htm>.
- [7] Ontario Ministry of Agriculture Food and Rural Affairs (OMAFRA), “Northern Livestock Pilot Action Plan.” [Online]. Available: <http://www.omafra.gov.on.ca/english/policy/nlpap.htm#sustain>.
- [8] R. Jiang *et al.*, “Impacts of land use conversions on soil organic carbon in a warming-induced agricultural frontier in Northern Ontario, Canada under historical and future climate,” *J. Clean. Prod.*, vol. 404, p. 136902, Jun. 2023, doi: 10.1016/J.JCLEPRO.2023.136902.
- [9] X. Wei, M. Shao, W. Gale, and L. Li, “Global pattern of soil carbon losses due to the conversion of forests to agricultural land,” *Sci. Reports 2014 41*, vol. 4, no. 1, pp. 1–6, Feb. 2014, doi: 10.1038/srep04062.
- [10] IPCC, “AR4 Climate Change 2007: The Physical Science Basis,” 2007.
- [11] M. A. Wulder *et al.*, “The global Landsat archive: Status, consolidation, and direction,” *Remote Sens. Environ.*, vol. 185, pp. 271–283, Nov. 2016, doi: 10.1016/J.RSE.2015.11.032.
- [12] Z. Zhu and C. E. Woodcock, “Continuous change detection and classification of land cover using all available Landsat data,” *Remote Sens. Environ.*, 2014, doi: 10.1016/j.rse.2014.01.011.
- [13] S. R. S. Aldhshan and H. Z. M. Shafri, “Change detection on land use/land cover and land surface temperature using spatiotemporal data of Landsat: a case study of Gaza Strip,” *Arab. J. Geosci.*, vol. 12, no. 14, 2019, doi: 10.1007/s12517-019-4597-4.
- [14] R. Coffey, “The difference between ‘land use’ and ‘land cover.’” [Online]. Available: https://www.canr.msu.edu/news/the_difference_between_land_use_and_land_cover.

- [15] Canada Centre for Remote Sensing, “Fundamentals of Remote Sensing,” 1982.
- [16] R. O. Duda, P. E. Hart, and D. G. Stork, “Pattern Classification and Scene Analysis 2nd ed. Part 1: Pattern Classification,” 1995.
- [17] Y. Sohn and N. S. Rebello, “Supervised and Unsupervised Spectral Angle Classifiers.”
- [18] M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, “A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information,” *Eur. J. Remote Sens.*, vol. 47, no. 1, pp. 389–411, 2014, doi: 10.5721/EuJRS20144723.
- [19] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, “A review of supervised object-based land-cover image classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, 2017, doi: 10.1016/j.isprsjprs.2017.06.001.
- [20] G. M. Foody, M. Pal, D. Rocchini, C. X. Garzon-Lopez, and L. Bastin, “The Sensitivity of Mapping Methods to Reference Data Quality: Training Supervised Image Classifications with Imperfect Reference Data,” *ISPRS Int. J. Geo-Information* 2016, Vol. 5, Page 199, vol. 5, no. 11, p. 199, Nov. 2016, doi: 10.3390/IJGI5110199.
- [21] D. Phiri and J. Morgenroth, “Developments in Landsat Land Cover Classification Methods: A Review,” *Remote Sens.* 2017, Vol. 9, Page 967, vol. 9, no. 9, p. 967, Sep. 2017, doi: 10.3390/RS9090967.
- [22] S. Talukdar *et al.*, “Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review,” *Remote Sens.* 2020, Vol. 12, Page 1135, vol. 12, no. 7, p. 1135, Apr. 2020, doi: 10.3390/RS12071135.
- [23] D. Lu and Q. Weng, “A survey of image classification methods and techniques for improving classification performance,” *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, 2007, doi: 10.1080/01431160600746456.
- [24] J. Zhang and G. M. Foody, “Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: Statistical and artificial neural network approaches,” *Int. J. Remote Sens.*, vol. 22, no. 4, pp. 615–628, 2001, doi: 10.1080/01431160050505883.
- [25] M. A. Shalan, M. K. Arora, S. K. Ghosh, M. A. Shalan, M. K. Arora, and S. K. Ghosh, “An evaluation of fuzzy classifications from IRS 1C LISS III imagery: A case study,” *Int. J. Remote Sens.*, vol. 24, no. 15, pp. 3179–3186, Aug. 2003, doi: 10.1080/0143116031000094791.
- [26] A. Tassi and M. Vizzari, “Object-Oriented LULC Classification in Google Earth Engine Combining SNIC, GLCM, and Machine Learning Algorithms,” 2020, doi: 10.3390/rs12223776.
- [27] L. Qu, Z. Chen, M. Li, J. Zhi, and H. Wang, “Accuracy Improvements to Pixel-Based and Object-Based LULC Classification with Auxiliary Datasets from Google Earth Engine,” *Remote Sens.* 2021, Vol. 13, Page 453, vol. 13, no. 3, p. 453, Jan. 2021, doi: 10.3390/RS13030453.
- [28] M. Vizzari, “PlanetScope, Sentinel-2, and Sentinel-1 Data Integration for Object-Based Land Cover Classification in Google Earth Engine,” *Remote Sens.*, vol. 14, no. 11, p.

2628, Jun. 2022, doi: 10.3390/RS14112628/S1.

- [29] P. Kempeneers, F. Sedano, P. Strobl, D. O. McInerney, and J. San-Miguel-Ayanz, “Increasing robustness of postclassification change detection using time series of land cover maps,” *IEEE Trans. Geosci. Remote Sens.*, 2012, doi: 10.1109/TGRS.2011.2181854.
- [30] S. P. Teng, Y. K. Chen, K. S. Cheng, and H. C. Lo, “Hypothesis-test-based landcover change detection using multi-temporal satellite images - A comparative study,” *Adv. Sp. Res.*, 2008, doi: 10.1016/j.asr.2007.06.064.
- [31] W. Kleynhans, J. C. Olivier, K. J. Wessels, B. P. Salmon, F. Van Den Bergh, and K. Steenkamp, “Detecting land cover change using an extended kalman filter onMODIS NDVI time-series data,” *IEEE Geosci. Remote Sens. Lett.*, 2011, doi: 10.1109/LGRS.2010.2089495.
- [32] A. Singh, “Review Article: Digital change detection techniques using remotely-sensed data,” *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989, doi: 10.1080/01431168908903939.
- [33] Y. Bazi, S. Member, L. Bruzzone, S. Member, and F. Melgani, “An Unsupervised Approach Based on the Generalized Gaussian Model to Automatic Change Detection in Multitemporal SAR Images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, 2005, doi: 10.1109/TGRS.2004.842441.
- [34] M. Al Mamun, M. Nazrul, I. Mondal, and B. Ahmed, “Change Detection-aided Single Linear Prediction of Multi-temporal Satellite Images.”
- [35] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, “PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data,” *Int. J. Remote Sens.*, 2008, doi: 10.1080/01431160801950162.
- [36] I. Ituen and B. Hu, “A multisource data approach for change and disturbance mapping of Ontario’s Clay Belt towards more accurate carbon and emissions estimation (Submitted),” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2024.
- [37] C. Sothe, A. Gonsamo, J. Arabian, W. A. Kurz, S. A. Finkelstein, and J. Snider, “Large Soil Carbon Storage in Terrestrial Ecosystems of Canada,” *Global Biogeochem. Cycles*, vol. 36, no. 2, p. e2021GB007213, Feb. 2022, doi: 10.1029/2021GB007213.
- [38] Province of New Brunswick, “Our Pathway Towards Decarbonization and Climate Resilience.”
- [39] United Nations Climate Change, “What is REDD+?” [Online]. Available: <https://unfccc.int/topics/land-use/workstreams/redd/what-is-redd>.
- [40] Government of Canada, “Net-zero emissions by 2050 - Canada.ca.” [Online]. Available: <https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/net-zero-emissions-2050.html>.
- [41] United Nations Climate Change, “What is the United Nations Framework Convention on Climate Change? | UNFCCC.” [Online]. Available: <https://unfccc.int/process-and->

- meetings/what-is-the-united-nations-framework-convention-on-climate-change.
- [42] CBC News, “Canada pulls out of Kyoto Protocol | CBC News,” 12-Dec-2011. [Online]. Available: <https://www.cbc.ca/news/politics/canada-pulls-out-of-kyoto-protocol-1.999072>.
 - [43] Government of Canada, “Regulatory Framework Air Emissions for Air Emissions,” 2007.
 - [44] Government of Canada, “Pan-Canadian Framework on Clean Growth and Climate Change - Canada.ca.” [Online]. Available: <https://www.canada.ca/en/services/environment/weather/climatechange/pan-canadian-framework.html>.
 - [45] Government of Canada, “2030 Emissions Reduction Plan: Clean Air, Strong Economy - Canada.ca,” 2023. [Online]. Available: <https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/climate-plan-overview/emissions-reduction-2030.html>.
 - [46] Environment and Climate Change Canada, “National inventory Report 1990 - 2020: Greenhouse Gas Sources and Sinks in Canada,” 2022.
 - [47] United Nations Climate Change, “Key aspects of the Paris Agreement.” [Online]. Available: <https://unfccc.int/most-requested/key-aspects-of-the-paris-agreement>.
 - [48] M. A. Wulder, J. C. White, R. A. Fournier, J. E. Luther, and S. Magnussen, “Spatially Explicit Large Area Biomass Estimation: Three Approaches Using Forest Inventory and Remotely Sensed Imagery in a GIS,” *Sensors*, vol. 8, pp. 529–560, 2008.
 - [49] C. Boisvenue and J. C. White, “Information needs of next-generation forest carbon models: Opportunities for remote sensing science,” *Remote Sens.*, vol. 11, no. 4, 2019, doi: 10.3390/rs11040463.
 - [50] G. P. Asner *et al.*, “High-fidelity national carbon mapping for resource management and REDD+,” *Carbon Balance Manag.*, vol. 8, no. 1, pp. 1–14, Jul. 2013, doi: 10.1186/1750-0680-8-7/FIGURES/6.
 - [51] M. Halme and E. Tomppo, “Improving the accuracy of multisource forest inventory estimates by reducing plot location error - A multicriteria approach,” *Remote Sens. Environ.*, vol. 78, no. 3, pp. 321–327, 2001, doi: 10.1016/S0034-4257(01)00227-9.
 - [52] H. Franco-Lopez, A. R. Ek, and M. E. Bauer, “Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method,” *Remote Sens. Environ.*, vol. 77, no. 3, pp. 251–274, 2001, doi: 10.1016/S0034-4257(01)00209-7.
 - [53] E. Tomppo, M. Nilsson, M. Rosengren, P. Aalto, and P. Kennedy, “Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass,” *Remote Sens. Environ.*, vol. 82, no. 1, pp. 156–171, Sep. 2002, doi: 10.1016/S0034-4257(02)00031-7.
 - [54] A. Baccini, M. A. Friedl, C. E. Woodcock, and R. Warbington, “Forest biomass estimation over regional scales using multisource data,” *Geophys. Res. Lett.*, vol. 31, no. 10, May 2004, doi: 10.1029/2004GL019782.

- [55] P. M. Barbosa, D. Stroppiana, J. M. Grégoire, and J. M. C. Pereira, “An assessment of vegetation fire in Africa (1981-1991): Burned areas, burned biomass, and atmospheric emissions,” *Global Biogeochem. Cycles*, vol. 13, no. 4, pp. 933–950, 1999, doi: 10.1029/1999GB900042.
- [56] R. H. Fraser and Z. Li, “Estimating fire-related parameters in boreal forest using SPOT VEGETATION,” *Remote Sens. Environ.*, vol. 82, no. 1, pp. 95–110, Sep. 2002, doi: 10.1016/S0034-4257(02)00027-5.
- [57] G. Popkin, “The hunt for the world’s missing carbon,” *Nat.* 2015 5237558, Jun. 2015.
- [58] D. Amarsaikhan and M. Saandar, “Fusion of multisource images for update of Urban GIS,” in *Image Fusion and Its Applications*, Y. Zheng, Ed. 2011.
- [59] D. Amarsaikhana, H. H. Blotevogel, J. L. van Genderenc, M. Ganzorig, R. Gantuya, and B. Nergui, “Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification,” *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 83–97, 2010, doi: 10.1080/19479830903562041.
- [60] Natural Resources Canada, “Canadian National Fire Database (CNFDB).” [Online]. Available: <https://cwfis.cfs.nrcan.gc.ca/ha/nfdb>.
- [61] G. Hulley, S. Veraverbeke, and S. Hook, “Thermal-based techniques for land cover change detection using a new dynamic MODIS multispectral emissivity product (MOD21),” *Remote Sens. Environ.*, vol. 140, pp. 755–765, Jan. 2014, doi: 10.1016/J.RSE.2013.10.014.
- [62] J. Batlle-Aguilar, A. Brovelli, A. Porporato, and D. A. Barry, “Modelling soil carbon and nitrogen cycles during land use change. A review,” *Agron. Sustain. Dev.* 2010 312, vol. 31, no. 2, pp. 251–274, Jun. 2010, doi: 10.1051/AGRO/2010007.
- [63] T. Huffman, J. Liu, M. McGovern, B. McConkey, and T. Martin, “Carbon stock and change from woody biomass on Canada’s cropland between 1990 and 2000,” *Agric. Ecosyst. Environ.*, vol. 205, pp. 102–111, Jul. 2015, doi: 10.1016/J.AGEE.2014.10.009.
- [64] S. Tremblay and R. Ouimet, “White Spruce Plantations on Abandoned Agricultural Land: Are They More Effective as C Sinks than Natural Succession?,” *For.* 2013, Vol. 4, Pages 1141-1157, vol. 4, no. 4, pp. 1141–1157, Dec. 2013, doi: 10.3390/F4041141.
- [65] W. A. Kurz *et al.*, “Carbon in Canada’s boreal forest-A synthesis,” *Environ. Rev.*, vol. 21, no. 4, pp. 260–292, 2013, doi: 10.1139/ER-2013-0041/ASSET/IMAGES/LARGE/ER-2013-0041F10.JPEG.
- [66] M. Thurner *et al.*, “Carbon stock and density of northern boreal and temperate forests,” *Glob. Ecol. Biogeogr.*, vol. 23, no. 3, pp. 297–310, Mar. 2014, doi: 10.1111/GEB.12125/SUPPINFO.
- [67] L. Moreau, E. Thiffault, D. Cyr, Y. Boulanger, and R. Beauregard, “How can the forest sector mitigate climate change in a changing climate? Case studies of boreal and northern temperate forests in eastern Canada,” *For. Ecosyst.*, vol. 9, p. 100026, Jan. 2022, doi: 10.1016/J.FECS.2022.100026.

- [68] H. Lee and J. (Eds. . Romero, “Climate Change 2023 Synthesis Report IPCC, 2023: Sections. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team.”
- [69] Natural Resources Canada, “Inventory and land-use change Forest inventory,” 2024. [Online]. Available: <https://natural-resources.canada.ca/climate-change/climate-change-impacts-forests/carbon-accounting/inventory-and-land-use-change/13111>.
- [70] I. Ituen and B. Hu, “An automatic and operational method for land cover change detection using spatiotemporal analysis of MODIS data: A Northern Ontario (Canada) case study,” *ISPRS Int. J. Geo-Information*, vol. 10, no. 5, 2021, doi: 10.3390/ijgi10050325.
- [71] I. Ituen, “Probabilistic risk and safety assessments in supercritical water reactors,” in *Canadian Nuclear Society - 31st Annual Conference of the Canadian Nuclear Society and 34th CNS/CNA Student Conference 2010*, 2010, vol. 1.
- [72] K. L. Webster, F. D. Beall, I. F. Creed, and D. P. Kreutzweiser, “Impacts and prognosis of natural resource development on water and wetlands in Canada’s boreal zone,” *Environ. Rev.*, vol. 23, no. 1, pp. 78–131, Jul. 2015, doi: 10.1139/ER-2014-0063/ASSET/IMAGES/LARGE/ER-2014-0063F11.JPEG.
- [73] U. Franko *et al.*, “CANDY Manual,” pp. 5–5, 2015.
- [74] Rothamsted Research, “Rothamsted Carbon Model (RothC).” [Online]. Available: <https://www.rothamsted.ac.uk/rothamsted-carbon-model-rothc>.
- [75] S. Nakamura *et al.*, “Validation of soil organic carbon dynamics model in the semi-arid tropics in Niger, West Africa,” *Nutr. Cycl. Agroecosystems*, vol. 89, no. 3, pp. 375–385, Apr. 2011, doi: 10.1007/S10705-010-9402-4/FIGURES/5.
- [76] C. Jie *et al.*, “Simulating the carbon storage of spruce forests based on the FORECAST model and remotely sensed data,” *Proc. - 2011 19th Int. Conf. Geoinformatics, Geoinformatics 2011*, 2011, doi: 10.1109/GEOINFORMATICS.2011.5981581.
- [77] W. A. Kurz *et al.*, “CBM-CFS3: A model of carbon-dynamics in forestry and land-use change implementing IPCC standards,” *Ecol. Modell.*, vol. 220, no. 4, pp. 480–504, Feb. 2009, doi: 10.1016/j.ecolmodel.2008.10.018.
- [78] F. Service, L. S. Heath, M. C. Nichols, J. E. Smith, and J. R. Mills, “FORCARB2: An updated version of the U.S. Forest Carbon Budget Model,” *Gen. Tech. Rep. NRS-67. Newt. Square, PA U.S. Dep. Agric. For. Serv. North. Res. Station. 52 p. [CD-ROM].*, vol. 67, 2010, doi: 10.2737/NRS-GTR-67.
- [79] O. R. Masera *et al.*, “Modeling carbon sequestration in afforestation, agroforestry and forest management projects: the CO2FIX V.2 approach,” *Ecol. Modell.*, vol. 164, no. 2–3, pp. 177–199, Jun. 2003, doi: 10.1016/S0304-3800(02)00419-2.
- [80] Natural Resource Ecology Laboratory, “The CENTURY Model.” [Online]. Available: <https://www2.cgd.ucar.edu/vemap/abstracts/CENTURY.html>.
- [81] N. H. U. Institute for the Study of Earth, Oceans and Space, “User’s Guide for the DNDC

- Model,” 2007.
- [82] European Forest Institute, “Modelling approach,” 2024. [Online]. Available: <https://efi.int/knowledge/models/efiscen/approach>.
 - [83] E. Milne *et al.*, “National and sub-national assessments of soil organic carbon stocks and changes: The GEFSOC modelling system,” *Agric. Ecosyst. Environ.*, vol. 122, no. 1, pp. 3–12, Sep. 2007, doi: 10.1016/J.AGEE.2007.01.002.
 - [84] G. P. Richards and D. M. W. Evans, “Development of a carbon accounting model (FullCAM Vers. 1.0) for the Australian continent,” *Aust. For.*, vol. 67, no. 4, pp. 277–283, 2004, doi: 10.1080/00049158.2004.10674947.
 - [85] A. F. Bouwman, T. Kram, and K. Klein Goldewijk, “Integrated modelling of global environmental change: an overview of Image 2.4,” 2006.
 - [86] A. Aaheim, R. K. Chaturvedi, and A. D. Sagadevan, “Integrated modelling approaches to analysis of climate change impacts on forests and forest management,” *Mitig. Adapt. Strateg. Glob. Chang.*, vol. 16, no. 2, pp. 247–266, Feb. 2011, doi: 10.1007/S11027-010-9254-X/TABLES/1.
 - [87] S. W. Running and E. R. Hunt, “Generalization of a Forest Ecosystem Process Model for Other Biomes, BIOME-BGC, and an Application for Global-Scale Models,” *Scaling Physiol. Process.*, pp. 141–158, Jan. 1993, doi: 10.1016/B978-0-12-233440-5.50014-2.
 - [88] J. A. Foley *et al.*, “An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics,” doi: 10.1029/96GB02692.
 - [89] I. Ituen and B. Hu, “Estimating carbon and greenhouse gas emissions in remote regions of Canada,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 43, no. B3-2021, pp. 575–580, 2021, doi: 10.5194/isprs-archives-XLIII-B3-2021-575-2021.
 - [90] Government of Ontario, “Provincial Satellite Derived Disturbance Mapping - Dataset - Ontario Data Catalogue.” [Online]. Available: <https://data.ontario.ca/en/dataset/provincial-satellite-derived-disturbance-mapping>.
 - [91] Natural Resources Information Portal, “Find a Plan.” [Online]. Available: https://nrip.mnr.gov.on.ca/s/fmp-online?language=en_US.
 - [92] Natural Resources Canada, “The science — and art — of forecasting wildland fires,” 2022. [Online]. Available: <https://natural-resources.canada.ca/simply-science/the-science-and-art-forecasting-wildland-fires/22766>.
 - [93] ESRI, “Ontario Wildfire Archive [Dataset].” [Online]. Available: <https://www.arcgis.com/home/item.html?id=6893d42d46c245598c0a4f100755a61c>.
 - [94] Government of Ontario, “Ontario GeoHub.” [Online]. Available: <https://geohub.lio.gov.on.ca/>.
 - [95] Natural Resources Canada, “Wildland fire and forest carbon: Understanding impacts of climate change.” [Online]. Available: <https://natural-resources.canada.ca/home/wildland-fire-and-forest-carbon-understanding-impacts-climate-change/24800>.

- [96] Government of Manitoba, “Province of Manitoba | agriculture - Why Do Farmers Burn?” [Online]. Available: <https://www.gov.mb.ca/agriculture/crops/crop-residue-burning-program/why-do-farmers-burn.html>.
- [97] British Columbia Public Service, “Agricultural burning - Province of British Columbia.” [Online]. Available: <https://www2.gov.bc.ca/gov/content/industry/agriculture-seafood/agricultural-land-and-environment/air-management/agricultural-burning>.
- [98] Ontario GeoHub, “Forest Disease Damage Event | Forest Disease Damage Event | Ontario GeoHub.” [Online]. Available: <https://geohub.lio.gov.on.ca/datasets/forest-disease-damage-event/explore?location=46.637948%2C-84.745000%2C4.00>.
- [99] Government of Canada, “Forest height in Canada 2011 - Open Government Portal.” [Online]. Available: <https://open.canada.ca/data/en/dataset/68b9f2b0-de5a-4571-a26d-473a15eedbf0>.
- [100] S. Dubayah, R., Hofton, M., Blair, J., Armston, J., Tang, H., Luthcke, “GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002 [Data set],” 2021. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LARSE_GEDI_GEDI02_A_002_MONTHLY#terms-of-use.
- [101] T. R. Loveland *et al.*, “Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data,” *Int. J. Remote Sens.*, vol. 21, no. 6–7, pp. 1303–1330, Jan. 2000, doi: 10.1080/014311600210191.
- [102] G. Grekousis, G. Mountrakis, and M. Kavouras, “An overview of 21 global and 43 regional land-cover mapping products,” *Int. J. Remote Sens.*, vol. 36, no. 21, pp. 5309–5335, 2015, doi: 10.1080/01431161.2015.1093195.
- [103] E. Bartholomé and A. S. Belward, “International Journal of Remote Sensing GLC2000: a new approach to global land cover mapping from Earth observation data GLC2000: a new approach to global land cover mapping from Earth observation data,” 2007, doi: 10.1080/01431160412331291297.
- [104] M. Friedl and D. Sulla-Menashe, “MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V061 [Data set],” 2022.
- [105] S. Fritz *et al.*, “Geo-Wiki: An online platform for improving global land cover,” *Environ. Model. Softw.*, vol. 31, pp. 110–123, May 2012, doi: 10.1016/J.ENVSOFT.2011.11.015.
- [106] S. Fritz *et al.*, “Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover,” *Remote Sens. 2009, Vol. 1, Pages 345-354*, vol. 1, no. 3, pp. 345–354, Aug. 2009, doi: 10.3390/RS1030345.
- [107] European Space Agency, “Land Cover CCI Product User Guide Version 2,” 2017.
- [108] European Space Agency, “OLCI Applications.” [Online]. Available: <https://sentiwiki.copernicus.eu/web/olci-applications>.
- [109] J. Wang, Y. Zhao, C. Li, L. Yu, D. Liu, and P. Gong, “Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250 m resolution,” *ISPRS J. Photogramm. Remote Sens.*, vol. 103, pp. 38–47, May 2015, doi: 10.1016/J.ISPRSJPRS.2014.03.007.

- [110] D. García-Álvarez *et al.*, “Global General Land Use Cover Datasets with a Time Series of Maps,” *L. Use Cover Datasets Valid. Tools*, pp. 287–311, 2022, doi: 10.1007/978-3-030-90998-7_15.
- [111] European Union’s Copernicus Land Monitoring Service, “CORINE Land Cover [Dataset].” [Online]. Available: <https://land.copernicus.eu/en/products/corine-land-cover/clc2018?tab=metadata>.
- [112] Eurostat, “LUCAS - Land use and land cover survey.” [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=LUCAS_-_Land_use_and_land_cover_survey.
- [113] Food and Agriculture Organization of the United Nations, “Africover - Land Cover Classification,” 1997.
- [114] U.S. Geological Survey, “National Land Cover Database.” [Online]. Available: <https://www.usgs.gov/centers/eros/science/national-land-cover-database>.
- [115] R. Latifovic and D. Pouliot, “Multitemporal land cover mapping for Canada: methodology and products,” *Can. J. Remote Sens.*, vol. 31, no. 5, pp. 347–363, 2005, doi: 10.5589/M05-019.
- [116] D. Pouliot, R. Latifovic, N. Zabcic, L. Guindon, and I. Olthof, “Development and assessment of a 250 m spatial resolution MODIS annual land cover time series (2000–2011) for the forest region of Canada derived from change-based updating,” *Remote Sens. Environ.*, vol. 140, pp. 731–743, Jan. 2014, doi: 10.1016/J.RSE.2013.10.004.
- [117] I. Olthof, R. Latifovic, and D. Pouliot, “Medium resolution land cover mapping of Canada from SPOT 4/5 data,” 2015.
- [118] C. F. Brown *et al.*, “Dynamic World, Near real-time global 10 m land use land cover mapping,” *Sci. Data* 2022 91, vol. 9, no. 1, pp. 1–17, Jun. 2022, doi: 10.1038/s41597-022-01307-4.
- [119] Spatial Thoughts LLP, “Enhancing Land Cover with Local Knowledge – Spatial Thoughts.” [Online]. Available: <https://spatialthoughts.com/2023/11/22/dw-mapathon-nairobi/#more-11245>.
- [120] (Eds.) Jeffrey A. Cardille, Nick Clinton, Morgan A. Crowley, David Saah, “Cloud-Based Remote Sensing with Google Earth Engine - Fundamentals and Applications.” [Online]. Available: https://docs.google.com/document/d/1UCB900oCdJERca-2WUeDlCu52MjPKJxETJ_jJcLM0bM/edit#heading=h.mm0ckpr9chy2.
- [121] J. M. Metsaranta, C. H. Shaw, W. A. Kurz, C. Boisvenue, and S. Morken, “Uncertainty of inventory-based estimates of the carbon dynamics of Canada’s managed forest (1990–2014),” *Can. J. For. Res.*, vol. 47, no. 8, pp. 1082–1094, 2017, doi: 10.1139/cjfr-2017-0088.
- [122] M. J. Bramley, S. Advisor, and N. Canada, “Canada’s Approach to Forest Carbon Quantification and Accounting: Key Concerns,” 2021.
- [123] G. K. Magnus *et al.*, “Growing our future: Assessing the outcome of afforestation

- programs in Ontario, Canada,” *For. Chron.*, vol. 97, no. 2, pp. 179–190, 2021, doi: 10.5558/tfc2021-019.
- [124] H. Temesgen, D. Affleck, K. Poudel, A. Gray, and J. Sessions, “A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models,” *Scand. J. For. Res.*, pp. 1–10, Feb. 2015, doi: 10.1080/02827581.2015.1012114.
 - [125] Environment and Climate Change Canada, “Canada’s Changing Climate Report,” 2019. .
 - [126] E. Hawkins and R. Sutton, “The Potential to Narrow Uncertainty in Regional Climate Predictions,” *Bull. Am. Meteorol. Soc.*, vol. 90, no. 8, pp. 1095–1108, Aug. 2009, doi: 10.1175/2009BAMS2607.1.
 - [127] Environment and Climate Change Canada, *National Inventory Report 1990 - 2021: Greenhouse Gas Sources and Sinks in Canada*. 2023.
 - [128] T. White, N. Luckai, G. R. Larocque, W. A. Kurz, and C. Smyth, “A practical approach for assessing the sensitivity of the Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3),” *Ecol. Modell.*, vol. 219, no. 3–4, pp. 373–382, 2008, doi: 10.1016/j.ecolmodel.2008.07.012.
 - [129] T. O’Hagan, “GEM Software,” 2017. [Online]. Available: <https://www.tonyohagan.co.uk/academic/GEM/index.html>.
 - [130] U. Franko, “Personal communication.” 2021.
 - [131] Google Earth Engine, “Synthetic Aperture Radar (SAR) Basics.” [Online]. Available: <https://developers.google.com/earth-engine/tutorials/community/sar-basics>.
 - [132] European Space Agency, “Overview of S1 Mission.” [Online]. Available: <https://sentiwiki.copernicus.eu/web/s1-mission>.
 - [133] European Space Agency, “Overview of S2 Applications.” [Online]. Available: <https://sentiwiki.copernicus.eu/web/s2-applications>.
 - [134] European Space Agency, “RapidEye Instruments.” [Online]. Available: <https://earth.esa.int/eogateway/missions/rapideye>.
 - [135] PlanetScope, “PlanetScope Instruments.” [Online]. Available: <https://earth.esa.int/eogateway/missions/planetscope#instruments-section>.
 - [136] Canadian Space Agency, “What is the RCM?” [Online]. Available: <https://www.asc-csa.gc.ca/eng/satellites/radarsat/what-is-rcm.asp>.
 - [137] Canadian Space Agency, “What is RADARSAT-2.” [Online]. Available: <https://www.asc-csa.gc.ca/eng/satellites/radarsat2/what-is-radarsat2.asp>.
 - [138] A. Ahrends *et al.*, “Current trends of rubber plantation expansion may threaten biodiversity and livelihoods,” *Glob. Environ. Chang.*, vol. 34, pp. 48–58, Sep. 2015, doi: 10.1016/J.GLOENVCHA.2015.06.002.
 - [139] S. M. Sterling, A. Ducharme, and J. Polcher, “The impact of global land-cover change on the terrestrial water cycle,” *Nat. Clim. Chang.* 2012 34, vol. 3, no. 4, pp. 385–390, Oct.

- 2012, doi: 10.1038/nclimate1690.
- [140] S. Niquisse, P. Cabral, Â. Rodrigues, and G. Augusto, “Ecosystem services and biodiversity trends in Mozambique as a consequence of land cover change,” *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.*, vol. 13, no. 1, pp. 297–311, Jan. 2017, doi: 10.1080/21513732.2017.1349836.
 - [141] A. Karpatne, Z. Jiang, R. R. Vatsavai, S. Shekhar, and V. Kumar, “Monitoring land-cover changes: A machine-learning perspective,” *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 8–21, Jun. 2016, doi: 10.1109/MGRS.2016.2528038.
 - [142] Y. L. Kong, Q. Huang, C. Wang, J. Chen, J. Chen, and D. He, “Long Short-Term Memory Neural Networks for Online Disturbance Detection in Satellite Image Time Series,” *Remote Sens. 2018, Vol. 10, Page 452*, vol. 10, no. 3, p. 452, Mar. 2018, doi: 10.3390/RS10030452.
 - [143] S. Srivastava, J. E. Vargas-Muñoz, and D. Tuia, “Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution,” *Remote Sens. Environ.*, vol. 228, pp. 129–143, Jul. 2019, doi: 10.1016/J.RSE.2019.04.014.
 - [144] C. Zhang and X. Li, “Land Use and Land Cover Mapping in the Era of Big Data,” *Land*, vol. 11, no. 10, p. 1692, Oct. 2022, doi: 10.3390/LAND11101692/S1.
 - [145] G. Cecili, P. De Fioravante, P. Dichicco, L. Congedo, M. Marchetti, and M. Munafò, “Land Cover Mapping with Convolutional Neural Networks Using Sentinel-2 Images: Case Study of Rome,” *L. 2023, Vol. 12, Page 879*, vol. 12, no. 4, p. 879, Apr. 2023, doi: 10.3390/LAND12040879.
 - [146] R. Naushad, T. Kaur, and E. Ghaderpour, “Deep Transfer Learning for Land Use and Land Cover Classification: A Comparative Study,” *Sensors 2021, Vol. 21, Page 8083*, vol. 21, no. 23, p. 8083, Dec. 2021, doi: 10.3390/S21238083.
 - [147] D. Zhang, Z. Liu, and X. Shi, “Transfer learning on EfficientNet for remote sensing image classification,” *Proc. - 2020 5th Int. Conf. Mech. Control Comput. Eng. ICMCCE 2020*, pp. 2255–2258, Dec. 2020, doi: 10.1109/ICMCCE51767.2020.00489.
 - [148] S. C. Hung, H. C. Wu, and M. H. Tseng, “Remote Sensing Scene Classification and Explanation Using RSSCNet and LIME,” *Appl. Sci. 2020, Vol. 10, Page 6151*, vol. 10, no. 18, p. 6151, Sep. 2020, doi: 10.3390/APP10186151.
 - [149] A. Alem and S. Kumar, “Transfer Learning Models for Land Cover and Land Use Classification in Remote Sensing Image,” *Appl. Artif. Intell.*, vol. 36, no. 1, Dec. 2022, doi: 10.1080/08839514.2021.2014192.
 - [150] P. Helber, B. Bischke, A. Dengel, and D. Borth, “Introducing EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification,” *Int. Geosci. Remote Sens. Symp.*, vol. 2018-July, pp. 204–207, Oct. 2018, doi: 10.1109/IGARSS.2018.8519248.
 - [151] A. M. Abdi, “Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data,” *GIScience Remote Sens.*, vol. 57, no. 1, pp. 1–20, Jan. 2020, doi: 10.1080/15481603.2019.1650447.

- [152] Z. Zhu *et al.*, “Continuous monitoring of land disturbance based on Landsat time series,” *Remote Sens. Environ.*, vol. 238, p. 111116, Mar. 2020, doi: 10.1016/J.RSE.2019.03.009.
- [153] H. Ghassemian, “A review of remote sensing image fusion methods,” *Inf. Fusion*, vol. 32, pp. 75–89, Nov. 2016, doi: 10.1016/J.INFFUS.2016.03.003.
- [154] U. Gessner *et al.*, “Multi-sensor mapping of West African land cover using MODIS, ASAR and TanDEM-X/TerraSAR-X data,” *Remote Sens. Environ.*, vol. 164, pp. 282–297, Jul. 2015, doi: 10.1016/J.RSE.2015.03.029.
- [155] M. Lange, B. Dechant, C. Rebmann, M. Vohland, M. Cuntz, and D. Doktor, “Validating MODIS and Sentinel-2 NDVI Products at a Temperate Deciduous Forest Site Using Two Independent Ground-Based Sensors,” *Sensors 2017, Vol. 17, Page 1855*, vol. 17, no. 8, p. 1855, Aug. 2017, doi: 10.3390/S17081855.
- [156] Han-qiu Xu and Tie-jun Zhang, “Cross comparison of ASTER and Landsat ETM+ multispectral measurements for NDVI and SAVI vegetation indices,” *Spectrosc. Spectr. Anal.*, vol. 31, no. 7, 2011.
- [157] K. Didan, “MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set],” 2015.
- [158] Z. Wan, S. Hook, and G. Hulley, “MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set],” 2015.
- [159] W. Huete, Alfredo R., Justice, C. van Leeuwen, “MODIS VEGETATION INDEX ALGORITHM THEORETICAL BASIS v3,” 1999.
- [160] S. Odunuga and G. Badru, “Landcover Change, Land Surface Temperature, Surface Albedo and Topography in the Plateau Region of North-Central Nigeria,” *Land*, 2015, doi: 10.3390/land4020300.
- [161] S. Pal and S. Ziaul, “Detection of land use and land cover change and land surface temperature in English Bazar urban centre,” *Egypt. J. Remote Sens. Sp. Sci.*, 2017, doi: 10.1016/j.ejrs.2016.11.003.
- [162] J. Sentian and S. S. K. Kong, “The effect of land cover changes on surface temperature and precipitation in the Southeast Asia region,” *Adv. Sci. Lett.*, vol. 21, no. 2, pp. 181–184, 2015, doi: 10.1166/asl.2015.5849.
- [163] Y. Deng *et al.*, “Relationship among land surface temperature and LUCC, NDVI in typical karst area,” *Sci. Rep.*, 2018, doi: 10.1038/s41598-017-19088-x.
- [164] A. Majumder, P. K. Kingra, R. Setia, S. P. Singh, and B. Pateriya, “Influence of land use/land cover changes on surface temperature and its effect on crop yield in different agro-climatic regions of Indian Punjab,” *Geocarto Int.*, vol. 0, no. 0, pp. 1–24, 2019, doi: 10.1080/10106049.2018.1520927.
- [165] Y. Zhang and S. Liang, “Impacts of land cover transitions on surface temperature in China based on satellite observations,” *Environ. Res. Lett.*, vol. 13, no. 2, 2018, doi: 10.1088/1748-9326/aa9e93.

- [166] U. Demšar, P. Harris, C. Brunsdon, A. S. Fotheringham, and S. McLoone, “Principal Component Analysis on Spatial Data: An Overview,” *Ann. Assoc. Am. Geogr.*, vol. 103, no. 1, pp. 106–128, Jan. 2013, doi: 10.1080/00045608.2012.689236.
- [167] A. P. King and R. . Eckersley, Eds., *Statistics for Biomedical Engineers and Scientists*. Cambridge, MA: Academic Press, 2019.
- [168] B. Hu, G. Brown, C. Stirling, and J. Wang, “Land Cover Classification of Subarctic Wetlands Using Multisource Remotely Sensed Data,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 5073–5092, 2024, doi: 10.1109/JSTARS.2024.3361279.
- [169] J. R. G. . Hansen, M. C.; Potapov, P. V.; Moore, R.; Hancher, M.;Turubanova, S. A.; Tyukavina, A.; Thau, D.; Stehman, S. V.; Goetz, S. J.; Loveland, T. R.; Kommareddy, A.; Egorov, A.; Chini, L.; Justice, C. O.; Townshend, “High-Resolution Global Maps of,” vol. 850, no. November, pp. 850–854, 2013, doi: 10.1126/science.1244693.
- [170] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, “PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data,” *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008, doi: 10.1080/01431160801950162.
- [171] S. Seyum, G. Taddese, and T. Mebrate, “Land use land cover changes on soil carbon stock in the Weshem Watershed, Ethiopia,” *For. Res. Eng. Int. J.*, vol. 3, no. 1, 2019, doi: 10.15406/freij.2019.03.00074.
- [172] G. Sharma, L. K. Sharma, and K. C. Sharma, “Assessment of land use change and its effect on soil carbon stock using multitemporal satellite data in semiarid region of Rajasthan, India,” *Ecol. Process.*, vol. 8, no. 1, pp. 1–17, Dec. 2019, doi: 10.1186/S13717-019-0193-5/FIGURES/8.
- [173] Y. Zhou, A. E. Hartemink, Z. Shi, Z. Liang, and Y. Lu, “Land use and climate change effects on soil organic carbon in North and Northeast China,” *Sci. Total Environ.*, vol. 647, pp. 1230–1238, Jan. 2019, doi: 10.1016/J.SCITOTENV.2018.08.016.
- [174] L. Lai *et al.*, “Carbon emissions from land-use change and management in China between 1990 and 2010,” *Sci. Adv.*, vol. 2, no. 11, Nov. 2016, doi: 10.1126/SCIADV.1601063/SUPPL_FILE/1601063_SM.PDF.
- [175] P. Illiger *et al.*, “Estimation of regional soil organic carbon stocks merging classified land-use information with detailed soil data,” *Sci. Total Environ.*, vol. 695, Dec. 2019, doi: 10.1016/J.SCITOTENV.2019.133755.
- [176] S. Tadese, T. Soromessa, A. B. Aneseye, G. Gebeyehu, T. Noszczyk, and M. Kindu, “The impact of land cover change on the carbon stock of moist afro-montane forests in the Majang Forest Biosphere Reserve,” *Carbon Balance Manag.*, vol. 18, no. 1, pp. 1–20, Dec. 2023, doi: 10.1186/S13021-023-00243-Z/TABLES/12.
- [177] Google Earth Engine, “Sentinel-1 Algorithms.” [Online]. Available: <https://developers.google.com/earth-engine/guides/sentinel1>.
- [178] Ontario GeoHub, “Forest Resources Inventory.” [Online]. Available: <https://geohub.lio.gov.on.ca/pages/forest-resources-inventory>.

- [179] Open Government Portal, “2020 Land Cover of Canada.” [Online]. Available: <https://open.canada.ca/data/en/dataset/ee1580ab-a23d-4f86-a09b-79763677eb47>.
- [180] Open Government Portal, “Annual Crop Inventory 2020.” [Online]. Available: <https://open.canada.ca/data/en/dataset/32546f7b-55c2-481e-b300-83fc16054b95>.
- [181] Agriculture and Agri-Food Canada, “AAFC Land Use Time Series 2000, 2005, 2010, 2015, 2020,” 2021. [Online]. Available: <https://open.canada.ca/data/en/dataset/fa84a70f-03ad-4946-b0f8-a3b481dd5248>.
- [182] Agriculture and Agri-Food Canada, “ISO 19131 AAFC Land Use – Data Product Specifications.” [Online]. Available: https://agriculture.canada.ca/atlas/data_donnees/landuse/supportdocument_documentdesupport/en/ISO_19131_AAFC_Land_Use_Data_Product_Specification.pdf.
- [183] Open Government Portal, “Dynamic Surface Water Maps of Canada from 1984-2023 Landsat Satellite Imagery.” [Online]. Available: <https://open.canada.ca/data/en/dataset/62de5952-a5eb-4859-b086-22a8ba8024b8>.
- [184] I. Olthof and T. Rainville, “Dynamic surface water maps of Canada from 1984 to 2019 Landsat satellite imagery,” *Remote Sens. Environ.*, vol. 279, p. 113121, Sep. 2022, doi: 10.1016/J.RSE.2022.113121.
- [185] T. Hermosilla, M. A. Wulder, J. C. White, and N. C. Coops, “Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes,” *Remote Sens. Environ.*, vol. 268, p. 112780, Jan. 2022, doi: 10.1016/J.RSE.2021.112780.
- [186] M. Amani *et al.*, “Canadian Wetland Inventory using Google Earth Engine: The First Map and Preliminary Results,” *Remote Sens. 2019, Vol. 11, Page 842*, vol. 11, no. 7, p. 842, Apr. 2019, doi: 10.3390/RS11070842.
- [187] Environment and Climate Change Canada, “Extent of Canada’s wetlands.” [Online]. Available: <https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/extent-wetlands.html>.
- [188] National Wetlands Working Group, “The Canadian Wetland Classification System,” 1997.
- [189] Open Government Portal, “Canadian Digital Elevation Model, 1945-2011.” [Online]. Available: <https://open.canada.ca/data/en/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333>.
- [190] D. Lu *et al.*, “Change detection techniques,” *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Jun. 2004, doi: 10.1080/0143116031000139863.
- [191] K. A. Adepoju and S. A. Adelabu, “Improving accuracy of Landsat-8 OLI classification using image composite and multisource data with Google Earth Engine,” *Remote Sens. Lett.*, vol. 11, no. 2, pp. 107–116, Feb. 2020, doi: 10.1080/2150704X.2019.1690792.
- [192] A. Tassi, D. Gigante, G. Modica, L. Di Martino, and M. Vizzari, “Pixel- vs. Object-Based Landsat 8 Data Classification in Google Earth Engine Using Random Forest: The Case

- Study of Maiella National Park,” *Remote Sens.* 2021, Vol. 13, Page 2299, vol. 13, no. 12, p. 2299, Jun. 2021, doi: 10.3390/RS13122299.
- [193] B. Lake, “Understanding soil pH.” [Online]. Available: https://www.dpi.nsw.gov.au/__data/assets/pdf_file/0003/167187/soil-ph.pdf.
- [194] Agriculture and Agri-Food Canada, “Soils of Canada.” [Online]. Available: https://agriculture.canada.ca/atlas/apps/aef/main/index_en.html?AGRIAPP=3&APPID=e87af05bd35848598994b13f45a24a25&WEBMAP-EN=c225cc78d5b142d58eacefae91cc535b&WEBMAP-FR=ad0b6822a33e411683f99979a1167efa&mapdescription=true&print=true&breadcrumb=b1,b2,b10,b3&a.
- [195] Ontario Ministry of Agriculture Food and Rural Affairs (OMAFRA), “Starting a Farm in Ontario,” 2018.
- [196] Ontario Ministry of Agriculture Food and Rural Affairs (OMAFRA), “Beef Cattle Farming Training Program for Indigenous people,” 2018. [Online]. Available: <http://www.omafra.gov.on.ca/english/policy/nlpresources.htm>.
- [197] J. Kent, “Agriculture in the Clay Belt of Northern Ontario,” *Can. Geogr. / Géographies Can.*, vol. 10, no. 2, pp. 117–126, Jun. 1966, doi: 10.1111/j.1541-0064.1966.tb00530.x.
- [198] OMAFRA, “Pasture Production, Publication 19,” 2015.
- [199] N. M. Mahowald *et al.*, “Interactions between land use change and carbon cycle feedbacks,” *Global Biogeochem. Cycles*, vol. 31, no. 1, pp. 96–113, Jan. 2017, doi: 10.1002/2016GB005374.
- [200] R. A. Houghton, “Interactions between land-use change and climate-carbon cycle feedbacks,” *Curr. Clim. Chang. Reports*, vol. 4, pp. 115–127, 2018, doi: 10.1007/s40641-018-0099-9.
- [201] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2021 – The Physical Science Basis*. Cambridge University Press, 2023.
- [202] Ontario Ministry of Natural Resources, “Managed Forests and Climate Change,” *Carbon*, 2022. [Online]. Available: <https://www.ontario.ca/page/managed-forests-and-climate-change>.
- [203] R. Ness, “Time to stop building infrastructure for yesterday’s climate Climate change is already having impacts that will get worse in the future Similar, but different — climate change will impact every region in its own way,” 2024. [Online]. Available: https://climateinstitute.ca/impacts-building-infrastructure-yesterday-s-climate/?utm_source=newsletter&utm_medium=email&utm_campaign=napr24i.
- [204] C. J. Bradshaw and I. G. Warkentin, “Global estimates of boreal forest carbon stocks and flux,” 2015, doi: 10.1016/j.gloplacha.2015.02.004.
- [205] D. F. Sandoval, J. J. Junca Paredes, K. J. Enciso Valencia, M. F. Díaz Baca, A. M. Bravo Parra, and S. Burkart, “Long-term relationships of beef and dairy cattle and greenhouse gas emissions: Application of co-integrated panel models for Latin America,” *Heliyon*,

- vol. 10, no. 1. 2024, doi: 10.1016/j.heliyon.2023.e23364.
- [206] H. Ritchie, “The carbon footprint of foods: are differences explained by the impacts of methane?,” 2020. [Online]. Available: <https://ourworldindata.org/carbon-footprint-food-methane#article-citation>.
 - [207] L. Zifei, “How Much Gas Do Beef or Dairy Cattle Produce,” *Kansas State Research and Extension*, pp. 1–4, 2015.
 - [208] P. Faverdin, H. Guyomard, L. Puillet, and A. Forslund, “Animal board invited review: Specialising and intensifying cattle production for better efficiency and less global warming: contrasting results for milk and meat co-production at different scales,” *Animal*, vol. 16, no. 1, p. 100431, 2022, doi: 10.1016/j.animal.2021.100431.
 - [209] V. MacPhail and A. Kyle, Jack: OMAFRA, “Rotational Grazing in Extensive Pastures,” 2012.
 - [210] I. Ménard, E. Thiffault, Y. Boulanger, and J. F. Boucher, “Multi-model approach to integrate climate change impact on carbon sequestration potential of afforestation scenarios in Quebec, Canada,” *Ecol. Modell.*, vol. 473, p. 110144, Nov. 2022, doi: 10.1016/J.ECOLMODEL.2022.110144.
 - [211] Copernicus Climate Change Service, “Copernicus: 2023 is the hottest year on record, with global temperatures close to the 1.5°C limit,” *Copernicus*, 2024. [Online]. Available: <https://climate.copernicus.eu/copernicus-2023-hottest-year-record#:~:text=2023 had a global average,highest annual value in 2016>.
 - [212] NASA, “Summer 2023 Was the Hottest Month on Record,” *NASA Earth Observatory*, no. August, 2023.
 - [213] P. L. PBC, “2023 In Images,” 2023. [Online]. Available: https://www.planet.com/pulse/2023-in-images/?mkt_tok=OTk3LUNISC0yNjUAAAGQJi1I-1BLF8O2GkXPwLesEYuef-Tg5a3ey2WRllcc20wvIExjWJwACgmsNaBMw05hVV2AaVLWqVomWakrwfKhZUIz607ZWl6aLy_th3vKJs.
 - [214] B. B. Feleke and N. Princewill, “More than 300 killed as heavy rains wreak havoc across East Africa,” 04-Dec-2023.
 - [215] B. N. Salahieh, J. Sutton, and L. Mascarenhas, “More than a third of the US population , from the Midwest to the East Coast , under air quality alerts from Canadian wildfire smoke,” 28-Jun-2023.
 - [216] I. Livingston, “Canada ’ s astonishing and record fire season finally slows down,” p. 2023, 18-Oct-2023.
 - [217] N. Opinion, S. Culture, L. Cecco, and W. Kurz, “Wildfires turn Canada’s vast forests from carbon sink into super-emitter,” *Toronto*, pp. 1–11, 22-Sep-2023.
 - [218] M. C. Hansen *et al.*, “High-resolution global maps of 21st-century forest cover change,” *Science (80-.)*, 2013, doi: 10.1126/science.1244693.

