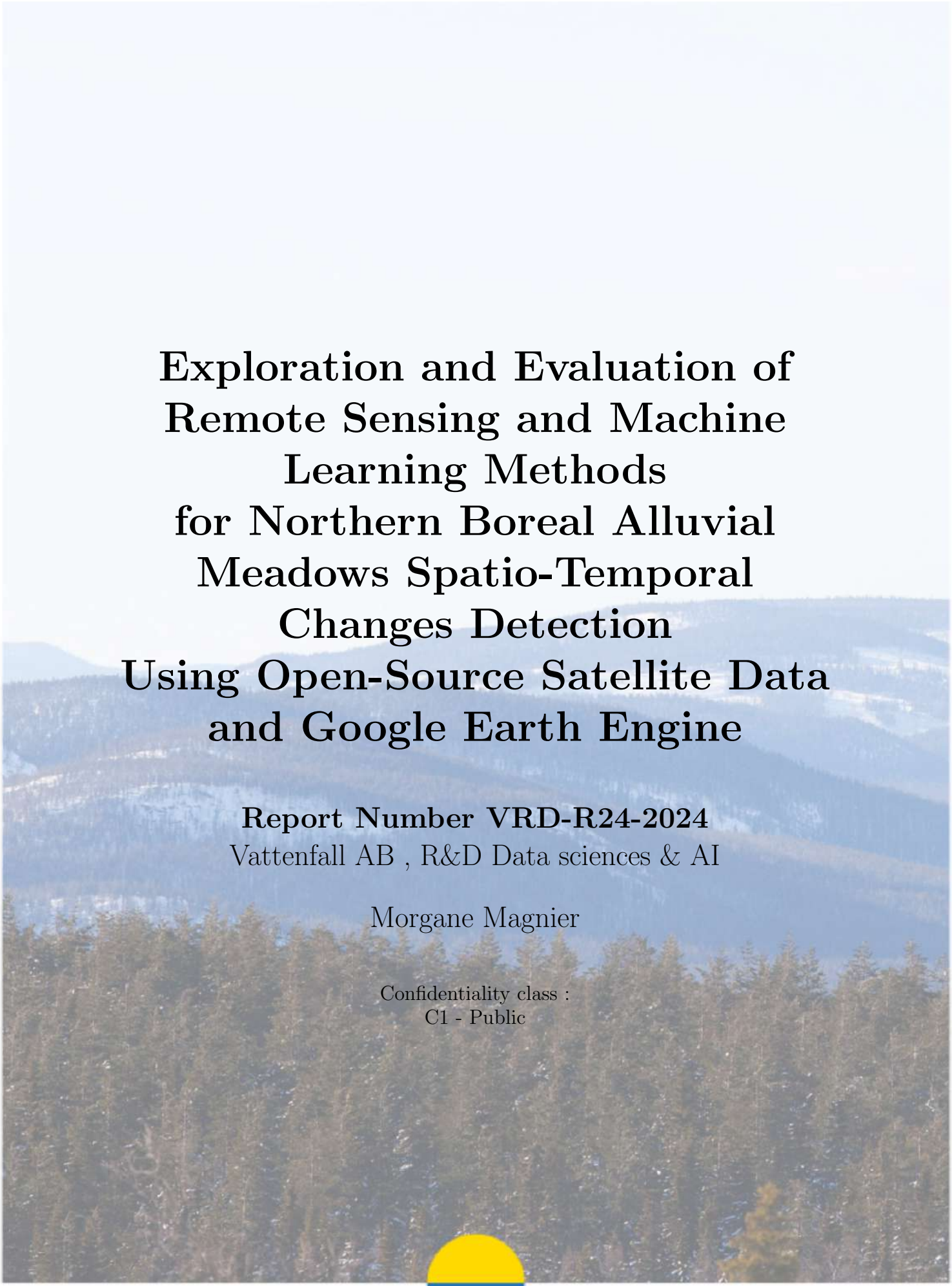

Exploration and Evaluation of Remote Sensing and Machine Learning Methods for Northern Boreal Alluvial Meadows Spatio-Temporal Changes Detection Using Open-Source Satellite Data and Google Earth Engine

Report Number VRD-R24-2024

Vattenfall AB , R&D Data sciences & AI

Morgane Magnier

Confidentiality class :
C1 - Public

VATTENFALL

Abstract

This master thesis investigates the potential of open-source satellite data for detecting habitat type changes in Northern Boreal Alluvial Meadows. The larger goal of the project is to explore the relationship between this habitat and the regulation of river water levels upstream of a hydraulic dam operated by Vattenfall within the Bredforsen Natura 2000 site.

The study employs various remote sensing and computer vision techniques, including spectral indexing, land cover classification using both unsupervised and supervised pixel-based algorithms, and statistical methods using the cloud based planetary processing open source platform, Google Earth Engine (GEE). These tools are evaluated for their effectiveness in detecting specific spatio-temporal changes and extracting long-term, exploitable time series data.

The analysis is conducted using open-source optical satellite data with limited spatial resolution (10, 25, and 30 meters), sourced from missions such as Landsat (5, 7, 8, and 9), and Sentinel-2. Additionally, satellite radar data from the Sentinel-1 SAR missions is utilized to a lesser extent. Supplementary datasets, including freely available land cover data (e.g., Dynamic World), digital elevation models, as well as climate data from European Centre for Medium-Range Weather Forecasts are used to support model training and analysis.

While no definitive conclusions could be drawn regarding the correlation between water level regulation and habitat changes, the study finds that the various tools complement each other in providing a comprehensive analysis of the relationship between water levels and vegetation changes.

Supervised classification emerges as the most effective method for examining the relationship between changes in Northern Boreal Alluvial Meadows, scrub and shrub encroachment, and water levels. However, its accuracy would benefit from additional field data, particularly regarding Northern Boreal Alluvial Meadows, croplands, shrubs and scrubs, and grasslands, to enhance classification results.

Spectral indexing proves valuable for preliminary studies and as input for classification models, as it helps rule out certain hypotheses related to vegetation changes, such as global shifts in vegetation or correlations with regional temperature or precipitation changes. However, it can lacks the precision needed in terms of spectral resolution, spatial resolution and ground based data to distinguish between different vegetation land cover types without additionnal data, especially when using Landsat sensors. NDVI density clustering is useful for detecting spatial changes but is limited by variability and the lack of statistical validation, making it less suitable for extracting long-term time series data.

The spatial resolution of open-source satellite data (10m and 30m) appears adequate for this application. However, the complexity and heterogeneity of Northern Boreal Alluvial Meadows, particularly the reflectance variability of pixels depending on plant species composition, remains unclear and warrants further exploration. While open-source optical imagery is crucial for long-term historical studies, it has limitations in quality and quantity, which may necessitate the fusion of multiple data sources to produce reliable time series, particularly in the context of fluctuating water levels. Radar data from Sentinel-1 proves valuable for monitoring water levels and has potential applications in ice monitoring.

Contents

Acronyms

DEM Digital Elevation Model. 7

DW Dynamic World. 5, 11

L5 Landsat-5. 4

L7 Landsat-7. 4

L8 Landsat-8. 4

L9 Landsat-9. 4

LULC Land Use Land Cover. 5, 28

MNDWI Modified Normalized Difference Water Index. 13

NDMI Normalized Difference Moisture Index. 13

NDVI Normalized Difference Vegetation Index. 13

NIR Near Infrared. 13

VI Vegetation Indices. 13

List of Figures

List of Tables

1 Acknowledgements

I would like to express my gratitude to Moshiour Rahaman, my company supervisor, and Mauro Dalla Mura for their invaluable assistance with my thesis. I am also deeply thankful to Dag Westland, my manager, for his support with all administrative matters and for being a great source of encouragement throughout this process.

I would like also to thank my friends from the SICOM class for their valuable advice on the technical and writing aspects.

I extend my heartfelt thanks to all my colleagues at Vattenfall, especially Anne, Ozan, Tara, Mia, and many others who made my time there both enjoyable and enriching. Their advice and support whenever needed were greatly appreciated. I am also thankful to the ski trip organizers, who provided me the opportunity from the very beginning to get to know my colleagues and explore Sweden.

A special thanks goes to Dag and Moshiour for selecting me for this project, as well as to my parents for their financial support during these six additional months. Their support allowed me to experience Sweden and Vattenfall—a journey that was both pleasant and fulfilling, and one I hope to continue in the future.

2 Introduction

2.1 Background

2.1.1 Hydroelectric Power and Environmental Impact

Hydroelectric power is widely regarded as a clean and renewable energy source. However, the construction and operation of dams can significantly disrupt natural water flows and ecosystems. The practice of hydropeaking, which involves rapid and frequent changes in water flow for optimal power production, exacerbates these disruptions by affecting both water flows and habitats. Previous research has highlighted various negative impacts, including habitat fragmentation, blocked migration paths for organisms, and habitat degradation [3].

To address these issues, the concept of environmental flows has been implemented globally. In Sweden, the EU Water Framework Directive [1], has led to significant changes in water management practices, including the revision of hydropower plant licenses to balance environmental and energy goals. Despite these measures, concerns about the ecological health of river systems persist.

2.1.2 Riparian Zones: Ecological Significance

Riparian zones, the areas along the banks of rivers and streams, are among the most biologically diverse ecosystems. These zones act as transition areas between aquatic and terrestrial environments, hosting species from both. The diversity of habitats in riparian zones supports high biodiversity. Riparian vegetation plays crucial roles in stabilizing riverbanks, controlling water temperature through shading, and reducing erosion [3].

Floodplains, an integral part of riparian zones, are characterized by their varied landscape and frequent flooding. These dynamics create a mosaic of habitats that support a wide range of species, making floodplains biodiversity hotspots recognized within the Natura 2000 network [3].

2.1.3 Challenges to Northern Alluvial Meadows

In particular, Northern alluvial meadows are specific river floodplains that are typically ice-covered in winter. These meadows are characterized by regular spring flooding and the impact of grazing and mowing, which decreased in the 20th century due to changes in agricultural practices such as the abandonment of hay production and the intensification of agriculture. Consequently, these areas became encroached by woody plants, reducing biodiversity [3].

Recognizing the ecological importance of alluvial meadows, they were included in the Natura 2000 network at the end of the 20th century. Restoration efforts, including mowing and grazing, have been implemented to prevent further decline. However, the long-term maintenance of these meadows also heavily depends on natural water flow conditions. In Sweden, where hydroelectric power contributes to 50 percent of the country's power generation, controlled water flow and damming significantly alter these conditions, impacting ecosystems that rely on specific water regimes [3].

2.1.4 Relevance to Vattenfall

Vattenfall AB, a major hydropower operator, is committed to environmental policies that include protecting nature and biodiversity. The company operates several dams and is subject to EU directives and hydropower plant licenses. Vattenfall AB seeks to minimize the environmental impacts of its operations while meeting energy production needs. Therefore, developing methods to maintain water flows that protect and restore Northern Alluvial Meadows is highly relevant to the company. The project is being conducted within Vattenfall's R&D division (located in Vattenfall main office, in Solna), specifically within the Asset and Analytics team, which is responsible for conducting projects related to data science in general, as well as numerous projects linked to biodiversity.

2.1.5 Study Area: Bredforsen Nature Reserve

Bredforsen, located in central Sweden within the counties of Gävle and Uppsala at the lower Dal River (Dalälven), serves as the study site for this research (See Figure 1. Scale of the satellite image won't be mentioned in the following section, as similar images will be used). This nature

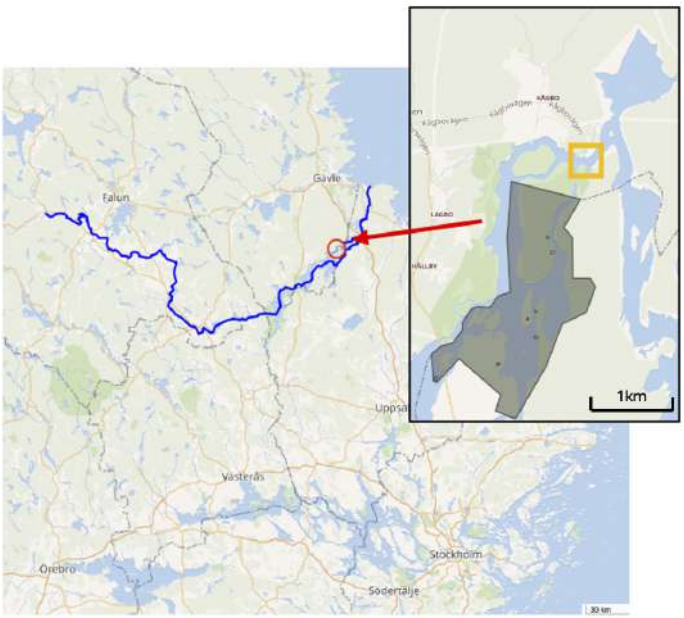

Figure 1: Location of Bredforsen area (in Red). In yellow, the localisation of Vattenfall hydropower plant. In blue, the Dal River.

reserve includes diverse habitats, with a significant presence of Northern Alluvial Meadows, which are vital for maintaining biodiversity. The reserve is part of the Natura 2000 network and is directly affected by the operations of the Söderfors power plant, owned by Vattenfall AB. The hydrological management practices in this area have critical implications for the ecological health of these meadows. Seasonal spring flooding and water flow regulation are key factors influencing the vitality of the riparian zones, which include extensive floodplain areas supporting grasses, herbs, and other vegetation.

Based on previous hydrological and biological studies concerning Northern Alluvial Meadows, several annual floodings have been implemented over the past years. It is now essential to understand their impact to develop a more precise management method regarding the regulation of water levels controlled by the Söderfors dam. The adjustments required by recent legislation on hydropower plant licenses aim to improve environmental outcomes in such sensitive areas.

2.1.6 Monitoring Northern Alluvial Boreal Meadows

According to European Commission Directive 92/43/EEC (2008), for Management of Natura 2000 habitats Northern Boreal alluvial meadows 6450 [2],

"The Swedish Environmental Protection Agency (Naturvårdsverket) has defined criteria for evaluating the conservation status of Northern Alluvial Boreal Meadows (Naturvårdsverket, 2005):

- The surface area (in hectares) that meets the definition of 'Nordic boreal alluvial meadows'.
- The percentage of well-managed meadow.
- Canopy coverage of trees and scrub (target is approximately 5 percent, with a range between zero and 30 percent for specific sites).
- Minimum area to be flooded annually (target is approximately 50 percent).
- Minimum percentage of monitoring study plots containing characteristic vascular plants of the habitat (examples include Bartsia alpina, Equisetum fluviatile, Parnassia palustris, Menyanthes trifoliata, Pedicularis palustris, and Pinguicula vulgaris).

The SEPA recommends a monitoring program that includes:

- Vegetation height, at 6-year intervals.
- Canopy cover, at 18-year intervals.
- Scrub encroachment, at 6-year intervals.

- Characteristic plant species on a minimum of 30 study plots along permanent transects, at 12-year intervals."

2.1.7 Advantages of Satellite Data in Environmental Monitoring

As highlighted in the previous section on monitoring Northern Alluvial Boreal Meadows, the Swedish Environmental Protection Agency (Naturvårdsverket) has established detailed criteria for evaluating the conservation status of these habitats, which require consistent and accurate data collection over time. Remote sensing via satellite imagery is an invaluable tool for this purpose, as it significantly reduces the need for extensive fieldwork, which is often costly and labor-intensive. Instead, satellite data provides continuous, high-frequency observations, allowing for efficient monitoring of key indicators like vegetation height, canopy cover, and flood extents.

The reflectance data captured by satellite sensors, including visible and near-infrared wavelengths, contains rich information about surface conditions. This enables precise discrimination and identification of changes in land cover, making it possible to monitor ecological shifts with a high degree of accuracy.

Open-source satellite data from platforms like Landsat, and Sentinel-1, which have been available for decades, are crucial for studying long-term biodiversity changes. These datasets provide a unique resource for tracking temporal trends, especially when historical field data are lacking. Moreover, their open-source nature ensures they are accessible at minimal cost, making them a cost-effective solution for large-scale environmental monitoring.

In addition, Landsat, Sentinel and MODIS satellite sensors have been successfully used in a range of applications, including deforestation monitoring, wetland change detection, water quality assessment, and flood dynamics analysis [6, 7, 9]. These applications further demonstrate the critical role of satellite data in supporting conservation and land management efforts.

2.2 Aim, Scope, and Limitations of the Thesis

The primary aim of this thesis is to explore and evaluate the use of various satellite data and AI tools for monitoring biodiversity, with a specific focus on Northern Alluvial Meadows. This study is part of a broader initiative to understand the relationship between water flows and the biodiversity of riparian zones. Additionally, the research investigates potential applications of these technologies in ice monitoring, flood detection, and dam management.

The scope of this research involves exploring and comparing different remote sensing and machine learning methods for detecting changes in Northern Boreal Alluvial Meadows within the Bredforsen nature reserve. The study emphasizes the use of open-source optical satellite data from Sentinel-2 (10m resolution) and Landsat-5 (L5),Landsat-7 (L7),Landsat-8 (L8),Landsat-9 (L9) (30m resolution), as well as radar data from Sentinel-1. The integration of these multiple satellite datasets aims to enhance the accuracy and reliability of change detection analyses, providing essential insights for conservation and management strategies.

However, the study acknowledges several limitations. The exclusive use of open-source satellite data with limited spatial resolution (10m for Sentinel-2 and 30m for Landsat 7 and 8) may impact the granularity and precision of the results. The research is confined to the Bredforsen nature reserve, limiting the generalizability of the findings to other regions. The evaluation of results and methods is conducted from a data science perspective, without rigorous input from biological experts, which may limit the ecological validity of the conclusions. Furthermore, as an initial exploration, the study lacks extensive field data and biodiversity knowledge, focusing more on the comparison and implementation of different methods rather than their improvement and statistical accuracy.

2.3 Data and platform used

2.3.1 Data used

The satellite data used in this thesis is summarized in Table 1.

Satellite data used in environmental monitoring are generally divided into two main types: optical and radar data. Optical sensors capture data in visible, near-infrared, and shortwave infrared bands, which are typically used for analyzing vegetation health, land cover classification, and water quality. These sensors generally offer a spatial resolution ranging from 10 to 60 meters and can have

Data Source	Data Type	Period	Interval	Spatial resolution	Spectral bands number	Spectral bands range
Sentinel-1 (S1)	Radar	Since 2014	6-12 days	10 m	1 band	5.405GHz (C-band)
Sentinel-2 (S2)	Optical	Since 2015	5 days (Two instruments)	10 m	12 bands	442.3nm (Blue-VIS) to 2202.4nm (SWIR))
Landsat 5	Optical	1984–2012	16 days	30 m	7 bands	450 nm (Blue) to 2350 nm (SWIR)
Landsat 7	Optical	Since 1999	16 days	30 m	7 bands	450 nm (Blue) to 2350 nm (SWIR)
Landsat 8	Optical	Since 2013	16 days	30 m	7 bands	435 nm (Blue) to 2350 nm (SWIR)
Dynamic World	Land Cover Classification	Since 2015	Daily	10 m		

Table 1: Summary of Data Sources Used

anywhere from 3 to 13 spectral bands, depending on the specific satellite (e.g., Landsat, Sentinel-2,). Radar sensors, on the other hand, such as those on Sentinel-1, use microwave frequencies to capture data regardless of weather conditions and lighting, making them ideal for monitoring floods, land subsidence, and ice movement. Radar data generally have a spatial resolution ranging from 5 to 20 meters [7],[6].

Dynamic World (DW), and ESA World Cover are both Land Use Land Cover (LULC) products [8], of 10m resolution, derived from Sentinel-2 data.

2.3.2 Platform Used

Monitoring changes in Northern Boreal Alluvial meadows requires extensive historical data over long periods. Cloud-based platforms like Google Earth Engine (GEE) are ideal for this task due to their ability to efficiently handle and process large volumes of data. GEE stores both data and processing tools on large server farms ("internet clouds") and provides access through APIs. The platform offers both JavaScript and Python APIs for accessing and analyzing vast remote sensing datasets. For this study, we utilized the Python API.

GEE is widely used for large-scale data processing due to its ability to access and quickly process extensive datasets without needing to download them locally. Its comprehensive data catalog and numerous available algorithms facilitate easy access to remote sensing tools, enabling the construction of long-term sequence data and trend analysis. Temporal aggregation methods in GEE reduce cloud interference, allowing the creation of long-time series of satellite image data. This makes GEE an invaluable tool for time series analysis and other related applications [42, 50].

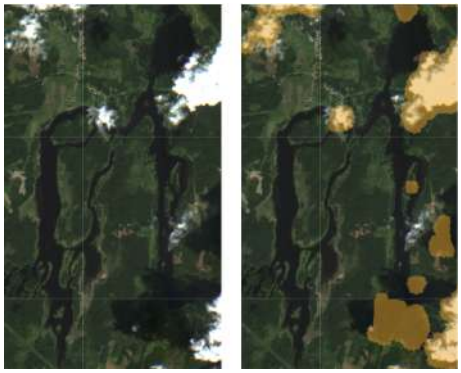

Figure 2: Clouds masking preprocessing Sentinel 2 data illustration in Bredforsen. On the left, RGB bands of S2 cloudy image. On the right, Clouds detected in orange by preprocessing method.

3 Preprocessing

3.1 Optical imagery preprocessing

Landsat missions and Sentinel-2 data provide both optical multi spectral data at two main levels of processing: Level 1 (L1) and Level 2 (L2). Other intermediate or lower levels exist, but won't be presented or used in this thesis.

Level 1 (L1) products from both Landsat and Sentinel-2 are raw satellite images corrected for geometric distortions, ensuring accurate location mapping. These L1 products, which include radiometric calibration and orthorectification using ground control points and digital elevation models (DEMs), still contain atmospheric effects. Level 2 (L2) products for both satellite systems involve additional processing to correct for atmospheric effects such as scattering and absorption, yielding more accurate surface reflectance values. Landsat L2 data are processed using algorithms like the Landsat Surface Reflectance Code (LaSRC), while Sentinel-2 L2 data are processed using the Sen2Cor processor. Usually, we have to be aware that both Sentinel-2 L2 and Landsat-8 L2 are not available for every assets limiting amount of data available or requiring more filtering in collection to avoid unpreprocessed data.

In particular, some more preprocessing steps are still necessary on L2 products, such as clouds - clouds shadows and snow (sometimes) masking.

Some functionalities of GEE platform can make a first filtering by proposing the clouds percentage of an asset in each product, but as we are studying a very local area representing a small part of assets, this method might in under optimal to keep as much data as possible.

3.1.1 Clouds Masking

On most of L2 products, added bands are available, informing directly on clouds / snows probability bands and mask. Those products are to be manipulated with precaution as some bands are empty for some products (available in only a time range, or geographic range etc), masking any clouds from a certain assets.

Bands used: `QA_PIXEL` for Landsat 5, 7, 8, and 9 [15], the Cloud Probability Image Collection for Sentinel-2, and the `SLC` band for cloud shadows [**?**].

For Landsat reflectance values to be similar between different sensors, coefficients from has also been added on ETM sensors (Landsat 4,5,7), tho their utility is a bit blurry depending on sources [11][36].

3.1.2 Clouds Filtering

After cloud masking, it is crucial to ensure that the remaining images contain enough valid data to produce reliable time series and compute indices such as surface reflectance or median values. In our case, because we are focusing on a very specific subset of satellite scene (Bredforsen area is measuring around 1,500 hectares, as a s2 images tile is 100x100km), relying on the global `CLOUDSPERCENTAGE` property proposed by GEE producst of the entire image is not suitable. This is because it can lead to discarding too many images that, despite high overall cloud coverage, still contain useful data within our area of interest.

To address this, we implemented a more localized filtering process. We calculate the number of valid pixels within our specific area of interest and discard images where the number of valid

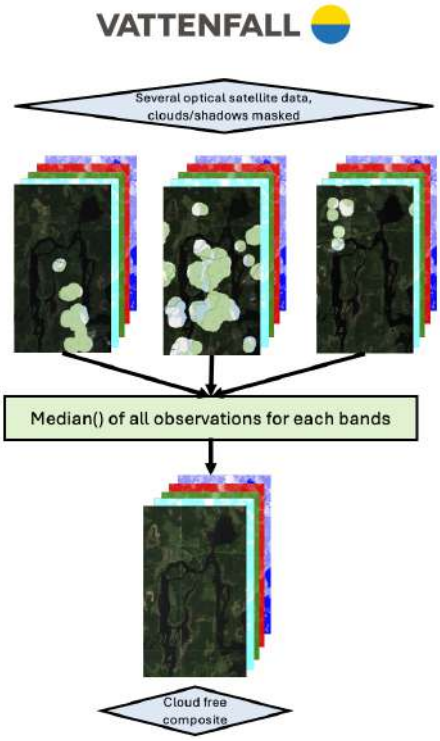

Figure 3: Composite creation from Sentinel 2 data illustration in Bredforsen

pixels falls below a certain threshold, which is determined based on the specific requirements of our study. This approach ensures that we retain as much useful data as possible without compromising the quality of the time series analysis.

3.1.3 Composite and aggregation methods

To avoid a lack of data, an additional common thing to do after cloud masking is creating composite from different assets, close in time acquisition; Different composition methods have various effects, especially on classification, and usually, temporal aggregation methods are better than selecting only cloudless images [22]. Composite creation is illustrated Figure 3. The choice of the number of observations and the temporal scale to optimally aggregate the data, especially when studying spatio-temporal changes, is a complex and extensive area of research. This topic will be briefly discussed and explored in the following sections.

Finally, images were also, in most case, normalized (Adjusting images to account for differences in acquisition times, angles, or sensor characteristics) and whenever needed, resampled (Aligning images to a common spatial resolution or projection).

3.2 SAR Preprocessing

Sentinel-1, collects C-band SAR imagery at various polarizations and resolutions. The data from Sentinel-1 is available in two main formats: Single Look Complex (SLC) and Ground Range Detected (GRD). The Level-1 GRD products aer the only currently supported by GEE platforms, they provide intensity ("detected") data, which is sampled in ground range [17].

The European Space Agency (ESA) preprocesses the Level-1 GRD data by applying initial corrections, including updating metadata with precise satellite positioning, applying radiometric calibration, and removing border noise to ensure data quality.

The data available in GEE platforms countains further preprocessing that is carried out using the Sentinel-1 Toolbox [18], [19]. This includes updating orbit metadata, removing thermal noise, applying additional radiometric calibration, and performing terrain correction to convert the data to terrain-corrected backscatter coefficients (σ), using Digital Elevation Model (DEM).

Backscatter coefficients are also accurately represented in decibels (dB), making it suitable for detailed analysis. In our study, we will utilize this preprocessed Sentinel-1 using Interferometric Wide (IW) swath mode data with VV (Single co-polarization, vertical transmit/vertical receive) and VH (Dual-band cross-polarization, vertical transmit/horizontal receive) polarization bands,

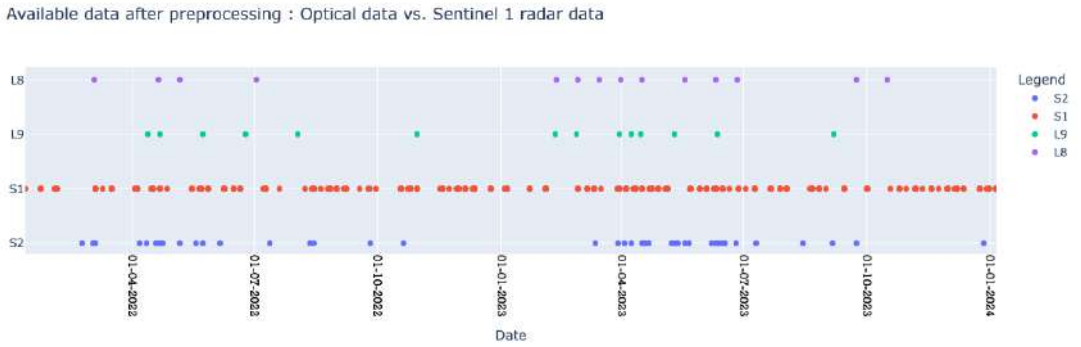

Figure 4: Available data after preprocessing in 2022-2024

which provide high-resolution SAR imagery.

As inspired by Gulacsi and Kovacs work [23], we also managed to add some other preprocessing steps. Wind filtering, using ERA5 dataset, the fifth-generation ECMWF atmospheric reanalysis [5], as well as performing incident angle normalization and applying a refined Lee filter for speckle noise filtering founded in Sentinel 1 toolbox [21].

4 Wetlands detection

Northern Boreal Alluvial Meadows are characterize by an annual spring flooding. Therefore, detection of lands that are flooded part of the year is a first important step in Northern Boreal Alluvial Meadows changes detection.

4.1 Material and Methods

Spring flooding, as well as artificial flooding from dam water flow management, occurs sporadically in Bredforsen over a period of 6 to 12 days, according to a manual study of images. Given this constraint, the choice of data for studying wetlands—areas that are flooded for part of the year—depends on the frequency of available observations.

As shown in Figure 4, Sentinel-1 indeed provides a higher observation frequency compared to optical data, such as that from Landsat missions and Sentinel-2, which are often hindered by significant cloud cover in the Bredforsen area. Sentinel-1 data are particularly effective for water detection due to their Synthetic Aperture Radar (SAR) technology, which allows for imaging regardless of cloud cover and lighting conditions. Additionally, according to András Gulácsi and Ferenc Kovács (2020), Sentinel-1 SAR imagery has demonstrated high accuracy in water cover detection in herbaceous wetlands (such as the study area) [23].

For this study, we use Sentinel 1 data acquired between October 2014 and July 2024 (No Sentinel data has been acquired before this date, we therefore uses all data available). The data collection has been filtered and preprocessed as described in 3.2.

A workflow of wetlands detection is presented Figure 5.

4.1.1 Water detection

To identify water surface on our S1 images, we applied a binary clustering using WEKA K-Means algorithm (Enhancement of the traditional K-Means by initializing cluster means from random sample) already implemented in GEE platform [38]. Using the default Euclidean distance function, we set the number of clusters to 2 and, following the work done by A. Gulácsi and F. Kovács (2020), we used both VH and VV polarization bands as inputs [23].

We trained the model with a substantial dataset of 10,000 random pixels. To identify clusters corresponding to water, we calculate for both clusters VV and VH polarization bands mean backscatter values. We identified the cluster of water with identifying the higher mean backscatter values of both bands.

The validation of the unsupervised clustering method is done by using Google Dynamic dataset, providing a very good F1 score for water body classification (0.93, calculated using Recall and Pre-

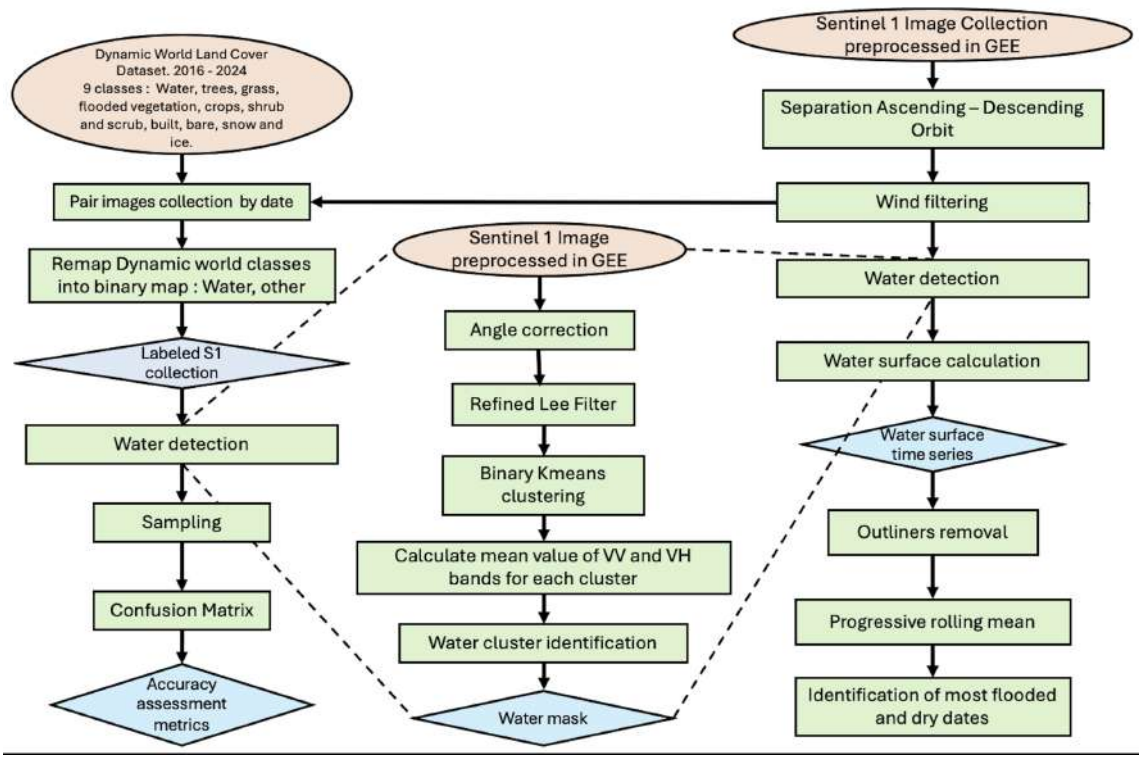

Figure 5: Wetlands detection workflow

cision provided by [12]). The Dynamic World maps were first reclassified into a binary classification (1 for water and 0 for all other classes) and then matched with S1 classified images based on acquisition dates.

4.1.2 Wetlands detection

After determining the water clusters, we calculated the surface of the area covered by water for each observation dates. In order to identify the area where spring/annual flooding is occuring, we identify, for each year, the water surface maximum and water surface minimum dates. To simplify the subsequent analysis and account for changes over time, we decided to extract a single mask, disregarding any variations in flooding that may occur between different years. To extract the final mask of wetlands area, we select the two S1 images from dates identified, and simply subtract the water cluster from the most flooded date from the water cluster of the most dry date.

4.2 Results and discussion

Figure. 6 illustrate the water detection. Some doubts were raised regarding the usefulness of the Refined Lee Filter, but no further studies were conducted, and we decided to keep the filter as is. We applied this detection on more than 5000 images with a pre-windy image removal and without. We preferably perform it on each observations and not creating composites, to have a dense time series of water surface, considering short regular flooding occurring. Still, this is to the price of less accurate water detection, with partly noise.

In total, 5,000 samples from 20 images were used to evaluate the confusion matrix. The analysis yielded an accuracy score of 0.91, which was considered precise enough to confidently use the time series for identifying images suitable for creating the wetlands mask.

Finally, removing wind filter and doing post processing on surface time series seemed to give better results.

Figure 7, present water body surface detected in Bredforsen. As we are using each observations, and didn't explored further in the preprocessing, some post processing of the time-series, such as removal of outliers, using classical `Pandas` The Python library, along with the application of a progressive rolling mean, has been utilized to effectively interpret the data. Various climate variables has been added, using ERA5 dataset [5], such as Daily Snow Cover, Daily Ice Depth,

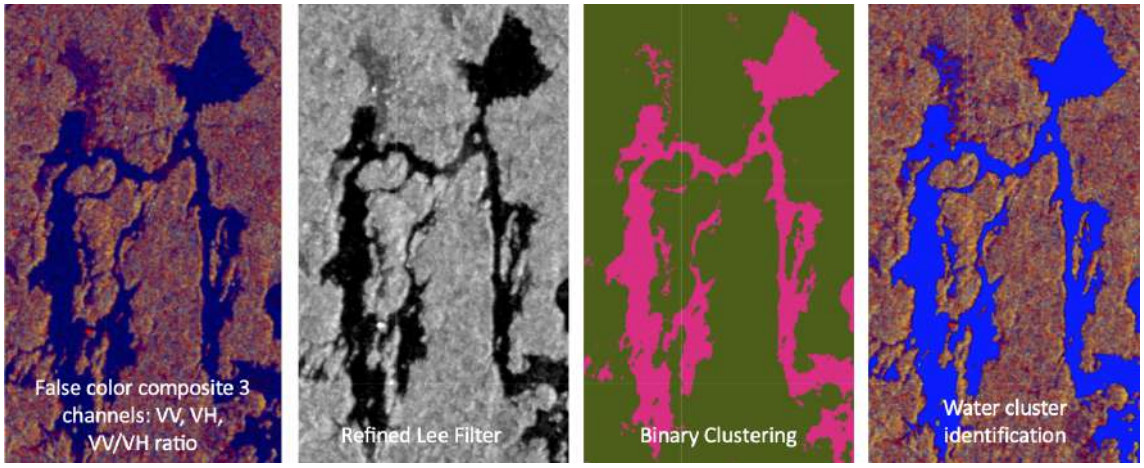

Figure 6: Water detection. In order, from left to right. A false color 3 channels composite using VV, VH and VV/VH band. SAR image after Refined Lee Filter. Result of binary kmean clustering. Water mask after identification, in blue.

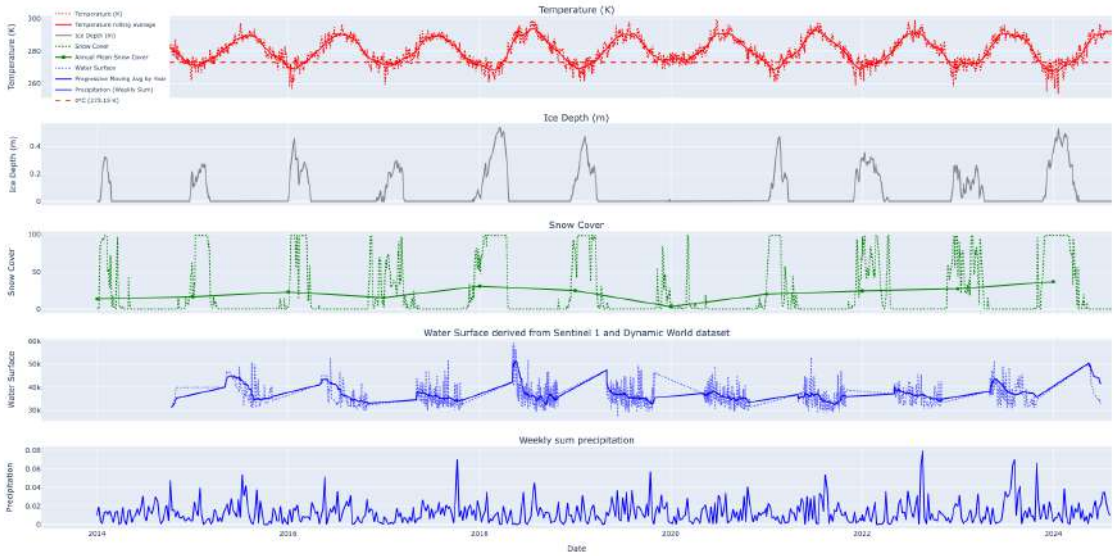

Figure 7: Water time series derived from S1 data, as well as other climate variables.

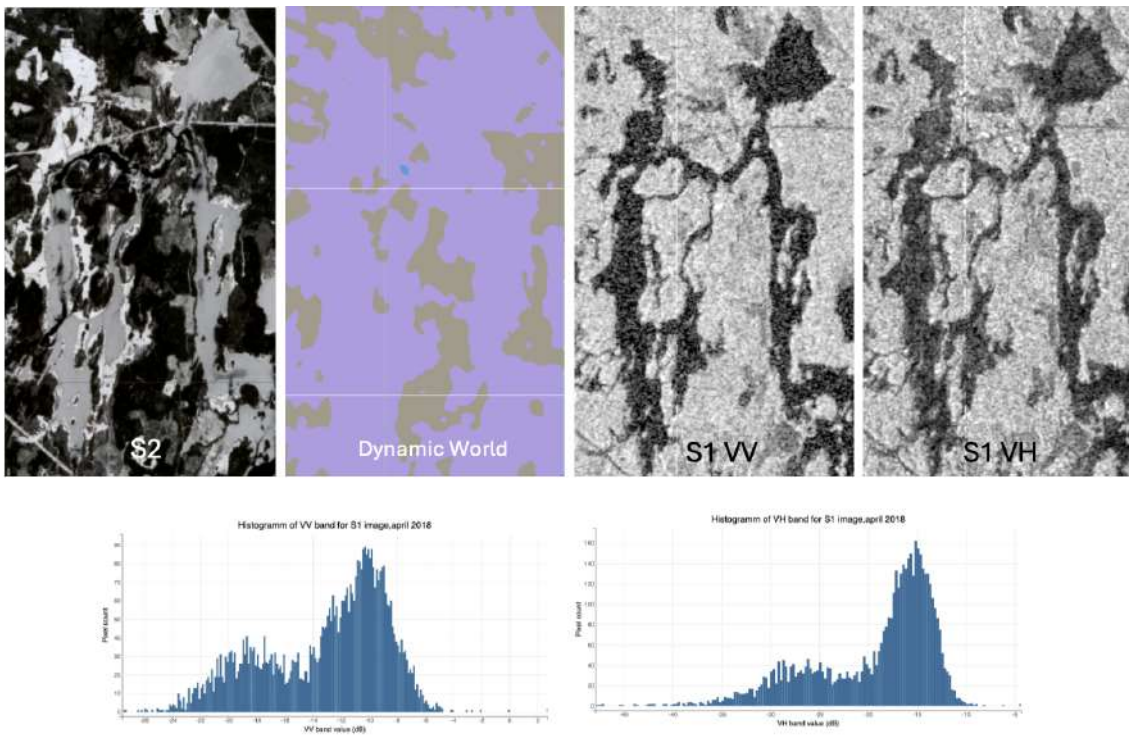

Figure 8: Different data types acquired on April 6th, 2018. From top to bottom and left to right: a S2 image, its corresponding Dynamic World classification map, the VV backscatter coefficient band from a S1 image, and the VH backscatter coefficient band from the same S1 image. Below these, histograms display the backscatter coefficient values for VV (left) and VH (right) from the S1 image.

Daily Temperature, and Daily precipitation aggregated in weekly sum.

A hypothesis can be made that the dynamics of water levels are correlated with ice depth and average snow cover, suggesting that significant ice and snow melt in spring leads to higher water levels in May. Precipitation data is used to verify that elevated water levels are not due to an unusually high precipitation period. Temperature data is employed to ensure consistency between snow cover, ice depth, and water levels, as data errors may occur. The artificial annual flooding has not been identified or distinguished from flooding caused by snow melt and ice melt in the images, requiring more detailed information from Vattenfall regarding its frequency, etc. A comparison of water levels between a river affected by a dam and one that is not affected is necessary, considering that the dam has likely been present for a long time, leading to similar dynamics. Alternatively, it is necessary to find a method to precisely study water levels prior to the implementation of artificial flooding.

In the future studies, we made the hypothesis that it is not necessary to have such a high frequency of observations to study the dynamics related to water levels and alluvial meadows. This frequency helped to clearly identify a study area considered favorable for the emergence of Northern Boreal alluvial meadows, with the hypothesis that this is the only area where the dam has an influence, which is a very limiting assumption.

We chose to detect water only during the extended summer season, specifically from May to October. Figure 8 presents various images acquired on the same day: an optical image from Sentinel-2 data (S2), the corresponding DW classification map, the 'VH' and 'VV' bands from Sentinel-1 (S1 VV and S1 VH), as well as histograms of the 'VV' and 'VH' backscatter values from the same image.

Although the work by Stonevicius et al. (2022) [24] demonstrated that using Sentinel-1 backscatter thresholds to detect ice in long sections of temperate rivers yielded reasonably good accuracy, especially considering the lower accuracy in areas where river flow and ice formation are influenced by large valley-dammed reservoirs like the Bredforsen rivers, we encountered difficulties in distinguishing snow, ice, and water. A image from S2, as well as its corresponding classification map from DW dataset, and a some S1 VV VH backscatter coefficient values acquired the same

date, are presented Figure 8. Tho the S2 image can confirme that the scenery in Bredforsen this day is a mix of snow, ice, and water/others land cover type, we found that most images from Bredforsen displayed a dual distribution of backscatter coefficients, making it impossible to discriminate between ice and water from a simple threshold. Additionally, the use of the RF model to classify water, snow, and ice using the Dynamic World dataset as reference data resulted in very poor accuracy, as demonstrated by the discrepancies between the S2 image and the DW classification map.

A more in-depth study, potentially including fieldwork, might help further investigate this issue and find a reliable method to discriminate between ice and water. Without removing the snow and ice season from the classification, the result was a mixed class of ice and water, leading to significantly overestimated water surface values in Bredforsen. Further studies are needed to determine whether using a single class for both water and ice could effectively lead to inaccuracies in the study, or if it can effectively be considered.

For the sake of simplicity, and also considering that the frequency of our dataset may lead to potentially missing significant high flooding events, we identify the most flooded and driest dates as those with the minimum and maximum water levels in the entire dataset, rather than defining wetlands for each year.

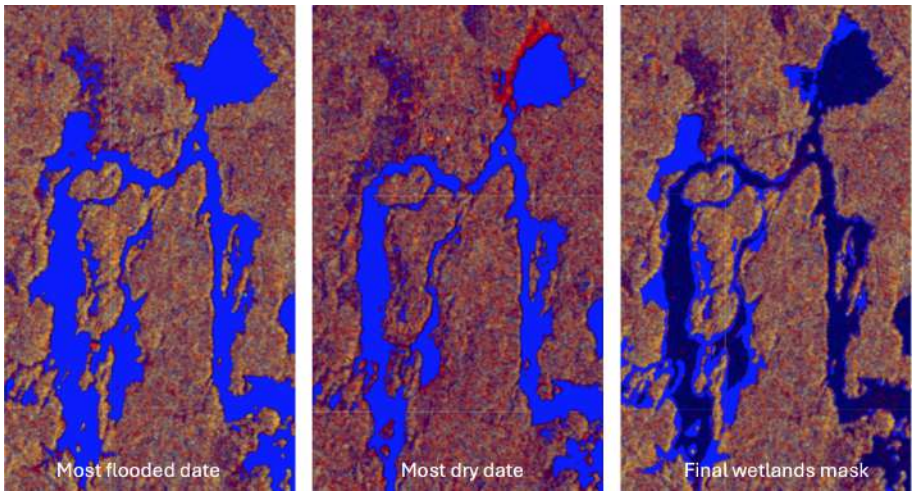

Figure 9: Wetlands mask extraction. To the left, water mask detected on the most flooded date. In center, water mask detected on the most dry date. To the right, wetlands mask obtained by substraction of the first two masks.

The final mask obtain is illustrated Figure 9. This mask would be used as the region of interest to study the changes.

This mask raised some questions, particularly regarding the limitations it imposes on the study. First, there are concerns about its accuracy, as the selection of dates is uncertain. The precision of the water detection is not perfect; the mask could miss undetected water in the driest image or might overestimate water in certain areas due to false positives. Additionally, there is uncertainty regarding the frequency of observations that detect water, which does not effectively account for the specific dates each year that correspond to flooded areas.

Alternative approaches, such as defining wetlands based on areas with a certain altitude level—that could easily become flooded part of the year and remain moist—could be considered. These areas could also allow for further delimitation of regions that might experience different frequencies of water presence and study the differences in species composition.

Although Northern Boreal Alluvial Meadows are typically defined as grasslands that are covered with ice in winter and flooded during part of the year, the spatial extent of the areas affected by changes in water cycles—and the presence of these meadows—might extend beyond the current mask.

Exploring the evolution of wetland areas over different years could be an avenue for future studies, particularly in examining the correlation between water dynamics and riparian zones.

5 Indexing

One of the key techniques in remote sensing is the use of spectral indices, which are mathematical combinations of pixel values from different spectral bands in satellite imagery. These indices enhance the detection and analysis of various surface features, such as vegetation, water bodies, and soil conditions, allowing researchers to derive meaningful insights from complex datasets.

A significant body of research has focused on developing vegetation indices (VIs), starting with the widely known Normalized Difference Vegetation Index (NDVI) [33]. Since its first implementation, the NDVI has become a cornerstone in vegetation monitoring due to its simplicity and effectiveness. NDVI is particularly valuable because it can indicate the presence or absence of vegetation and assess the health of vegetation by analyzing the Near Infrared (NIR) and Red bands, which characterize the spectral reflectance curve of vegetation.

Recent advancements in remote sensing have expanded the catalog of spectral indices available to researchers. Montero et al. (2023) [25] presented a standardized catalog of spectral indices aimed at enhancing remote sensing applications in Earth system research. This catalog, published in Scientific Data, serves as a valuable resource for researchers by detailing various indices and their applications. Additionally, Montero developed *eemont*, a Python package that extends GEE, facilitating more advanced remote sensing analyses [26].

As a first approach to detect spatio-temporal changes in Northern Boreal Alluvial Meadows, we utilize these spectral indices to detect and analyze vegetation changes over time in the Bredforsen area and its wetlands. We generated time series of index values and assessing spatial dynamics by comparing indices values between different images.

5.1 Materials and Methods

Remote sensing indices are crucial for monitoring vegetation and water dynamics. Historical studies, such as those by Lyon et al. (1998) [28], have demonstrated the effectiveness of the NDVI in detecting vegetation changes. However, more recent research, including the work of Xue and Su (2017) [29], highlights the importance of selecting specific Vegetation Indices (VI) based on the environmental context and research objectives. While NDVI is widely utilized, it is sensitive to canopy structure and photosynthesis, and it requires calibration to correct for soil and atmospheric effects.

To find the optimal spectral indices for our study area, we conducted a brief preliminary analysis to identify the most appropriate indices for monitoring vegetation and water dynamics in the Bredforsen area. This analysis, that won't be presented in this thesis, included studying indices visually, as well as doing a Component Principal Analysis of various indices times series derived from satellite images from wetlands scenery. If most vegetation indices informations could be contain in one, no other indices from literature, has been found particularly more relevant than the tradional NDVI. Finally spectral indices chosen for this study includes NDVI, which characterizes vegetation, Modified Normalized Difference Water Index (MNDWI), which characterizes water presence and humidity [30], and the Normalized Difference Moisture Index (NDMI), characterizing moisture in vegetation [34].

Thoses indices can be caculated using the following equations :

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}} \tag{1}$$

$$\text{MDNWI} = \frac{\rho_{\text{GREEN}} - \rho_{\text{SWIR}}}{\rho_{\text{GREEN}} + \rho_{\text{SWIR}}} \tag{2}$$

$$\text{NDMI} = \frac{\rho_{\text{NIR}} - \rho_{\text{SWIR}}}{\rho_{\text{NIR}} + \rho_{\text{SWIR}}} \tag{3}$$

Where :

- ρ_{NIR} is the reflectance in the Near Infrared band.
- ρ_{RED} is the reflectance in the Red band.
- ρ_{GREEN} is the reflectance in the Green band.
- ρ_{SWIR} is the reflectanc ein the Shortwave Infrared.

Temporal changes

To extract time series, we employed optical imagery from Sentinel-2, acquired between 2017 and July 2024, and Landsat missions, including Landsat 5, 7, 8, and 9, spanning from 1984 to the present. The multisource dataset was preprocessed according to the methodology detailed in Section 3.1. Specifically, monthly composites were created, and NDVI, MNDWI, and NDMI spectral indices were calculated on these composites.

To complement our analysis dataset, we incorporated monthly temperature and precipitation data for the Bredforsen area using ERA5 dataset, the fifth-generation ECMWF atmospheric reanalysis [5].

After preprocessing, composites from Landsat and Sentinel-2 were masked to extract the median index values exclusively within detected wetlands, as described in Section 4. The pixel values of the spectral indices (NDVI, MNDWI, and NDMI) were then aggregated into a single value per image, representing the median index values of each monthly composite within the wetlands.

Spatial changes

As spatial changes are made to studies related to the vegetation, only special changes regarding NDVI index and NDMI index were conducted. Summer composites were chosen for temporal aggregation. This choice was made because monthly or single observation NDVI values can vary between years due to fluctuations in the vegetation growing season, influenced by random climate factors or climate change dynamics. As we benefit from a big historical range of data, we choose to compute spatial changes between 1999 and 2023 using Landsat 7, and a comparison between Sentinel 2 and Landsat 8 change detection is made between years 2017 and 2023.

A most common way to study changes between two single band images is to subtract the two images, and apply a threshold, generally considered as the standard deviation of the difference value [52]. However, while this method is effective for detecting general changes, it doesn't seem to be adequate in our case. The standard deviation of the differences in spectral index bands may be very small, as we found that minimal changes have occurred in these indices in the wetlands of Bredforsen. A more suitable approach is to use the mean standard deviation of the two single-band images under consideration. Therefore, we consider that changes have occurred if they exceed the standard deviation of spectral index values, taking into account the specific sensor and type of temporal aggregation. This approach also ensures a more consistent threshold for detecting spatial changes, allowing for better comparison.

5.2 Results and Discussion

Temporal changes

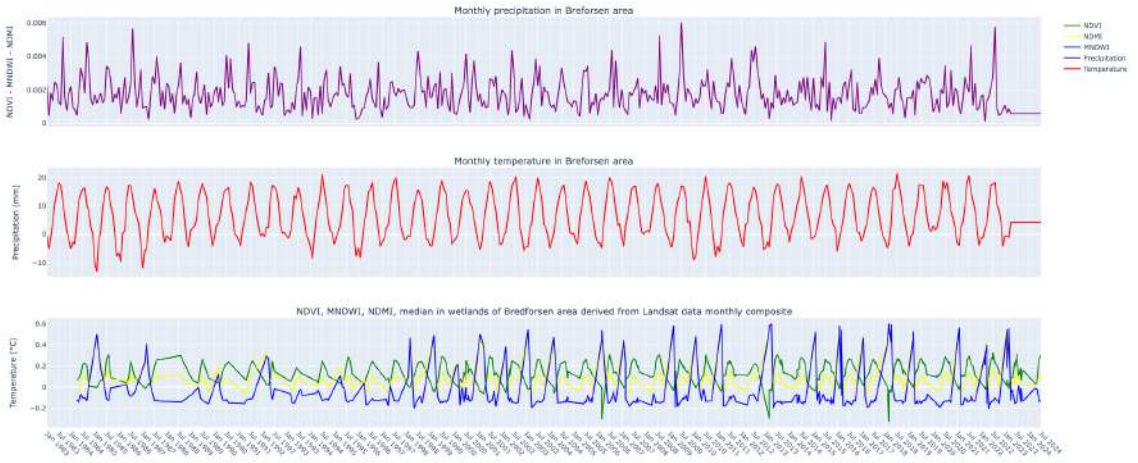

Figure 10: NDVI, MNDWI, NDMI time series in wetlands, derived from Landsat data monthly composite 1984-2024, as well as monthly temperature and precipitation time series derived from the ERA5 dataset (After April 2023: Precipitation and Temperature unavailable).

Figure 10 present the time series of the three indices, such as NDVI, NDMI, MNDWI median from each month in the wetlands, as well as monthly precipitation and temperature from 1985 to today. A preliminary interpretation of these time series suggests that, despite some fluctuations,

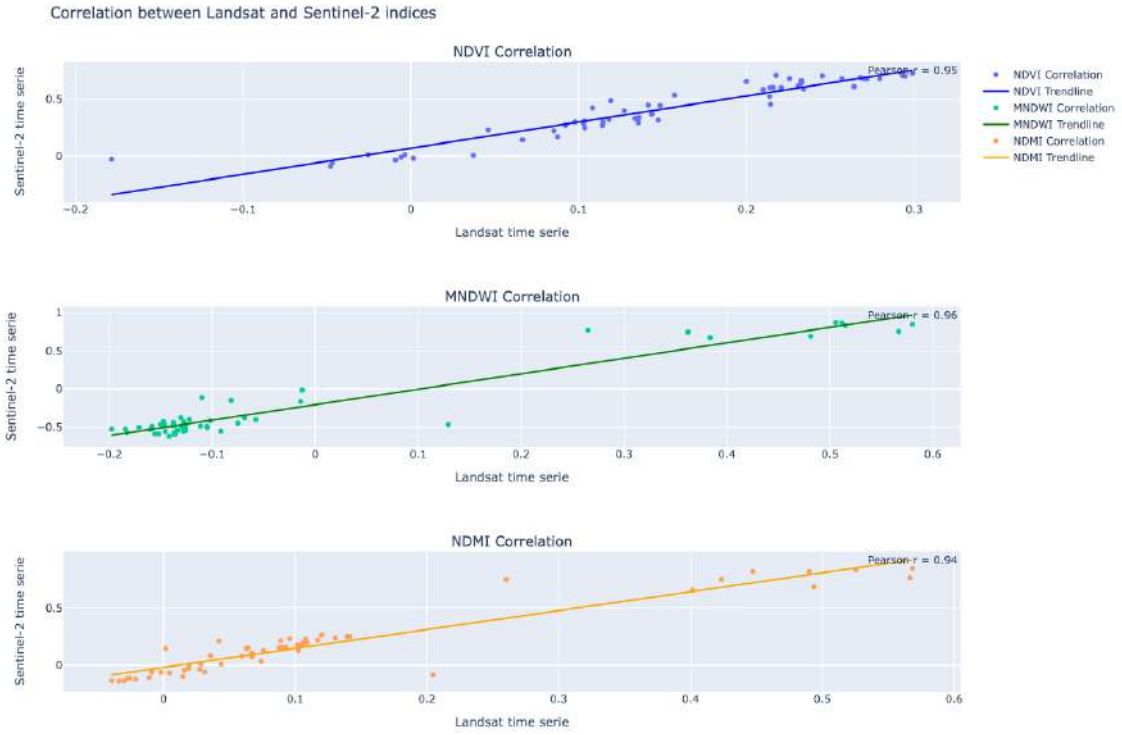

Figure 11: Correlation between NDVI (On top top), MNDWI (In center), and NDMI (On the bottom) time series from S2 and the ones from Landsat between 2017 and 2024.

	NDVI	**MNDWI**	**NDMI**	**Precipitation**	**Temperature**
NDVI	1.000000	-0.736620	-0.452250	0.364822	0.720530
MNDWI	-0.736620	1.000000	0.924362	-0.215553	-0.579363
NDMI	-0.452250	0.924362	1.000000	-0.084777	-0.371075
Precipitation	0.364822	-0.215553	-0.084777	1.000000	0.438366
Temperature	0.720530	-0.579363	-0.371075	0.438366	1.000000

Table 2: Correlation matrix (Pearson's coefficients) of spectral indices and climate variables for time series data from Figure 10.

there have been no significant long-term changes in the NDVI and NDMI values of the Bredforsen wetlands since 1985. The MNDWI index appears to have experienced fewer fluctuations between 1987 and 1997, which may warrant further investigation to explore a potential connection with the Bredforsen dam. Additionally, the precipitation data helps confirm that the observed increases in MNDWI are not merely due to higher precipitation events but indeed reflect changes in water levels within the wetlands.

Table 2 present the Pearson coefficient between the time series of the different variables. It shows a negative correlation (That can be also observed visually in Figure 10) between the MNDWI and NDVI indices. This negative correlation suggests that when water content in the wetlands increases, as indicated by higher MNDWI values, vegetation presence decreases, as shown by lower NDVI values. This relationship aligns with the expectation that increased water levels would submerge vegetation and reduce its surface visibility. As anticipated [37], the temperature data shows a positive correlation with NDVI, indicating that higher temperatures in the growing seasons are associated with increased vegetation health and growth. The simple Pearson coefficient might not be entirely appropriate if, for example, the correlation between two signals is offset in time. However, Figure 10 allows us to visually establish the essence of our conclusions.

Mandanici, E. and al. (2016), [27] showed that if in most cases, S2 and L8 sensors can be well combined, some issues can arise regarding near-infrared bands when S2 data are combined with both Landsat 8 and older Landsat images. As NDVI is using the NIR band, we didn't combine spectral indices values from Landsat and from S2. However, we used S2 for some validation of our conclusions;

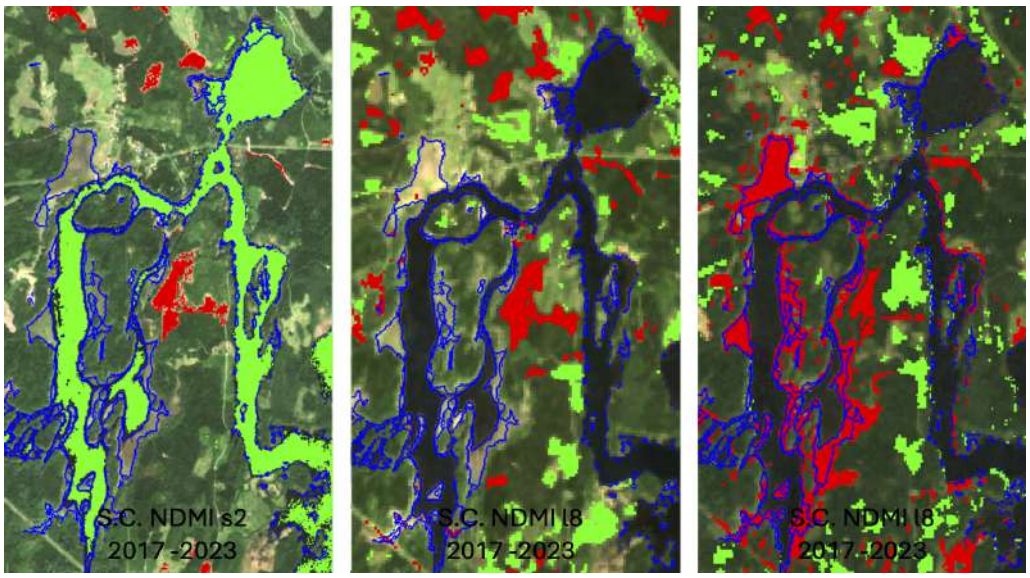

Figure 12: NDMI change detection. From left to right: change detection between 1) Sentinel-2 summer composites from 2017 and 2023, 2) Landsat 8 summer composites from 2017 and 2023, and 3) Landsat 7 summer composites from 1999 and 2023. Red indicates a significant negative change in NDMI values, while green indicates a significant positive change.Wetlands are delimited in blue.

Figure 11 presents the correlation between time series derived from two different sensors: Sentinel-2 (S2) and Landsat 7 & 8. The high Pearson coefficients for all three spectral indices (0.95, 0.96, and 0.94 for NDVI, MNDWI, and NDMI, respectively) indicate that the time series derived from Landsat do not exhibit biases related to: 1) the temporal aggregation differences between Landsat and S2, which could potentially lead to varying results, and 2) the impact of spatial resolution, which does not appear to affect the results at this stage of the data extraction.

Spatial changes

The standard deviation was calculated from 24 images captured by Landsat 7, 6 by Landsat 8, and 6 by Sentinel 2. Threshold detection for NDVI spectral indices was determined as follows: 0.11 for Landsat 7, 0.24 for Landsat 8, and 0.4 for Sentinel 2; however, significant variation in the standard deviation for Sentinel 2 was observed, leading to a potential reconsideration of the threshold selection, with the maximum thresholds for NDMI established as follows: 0.4 for Sentinel 2, 0.08 for Landsat 8, and 0.07 for Landsat 7.

A notice have to be made in the consideration of both indices for spatial changes, despite their high correlation in the time series shown in Table 2. This was decided following a brief visual analysis revealing that Landsat sensors were less sensitive to NDVI differences—such as those between wetlands and dense forests—compared to NDMI. However, the use of this index could be reconsidered in future studies.

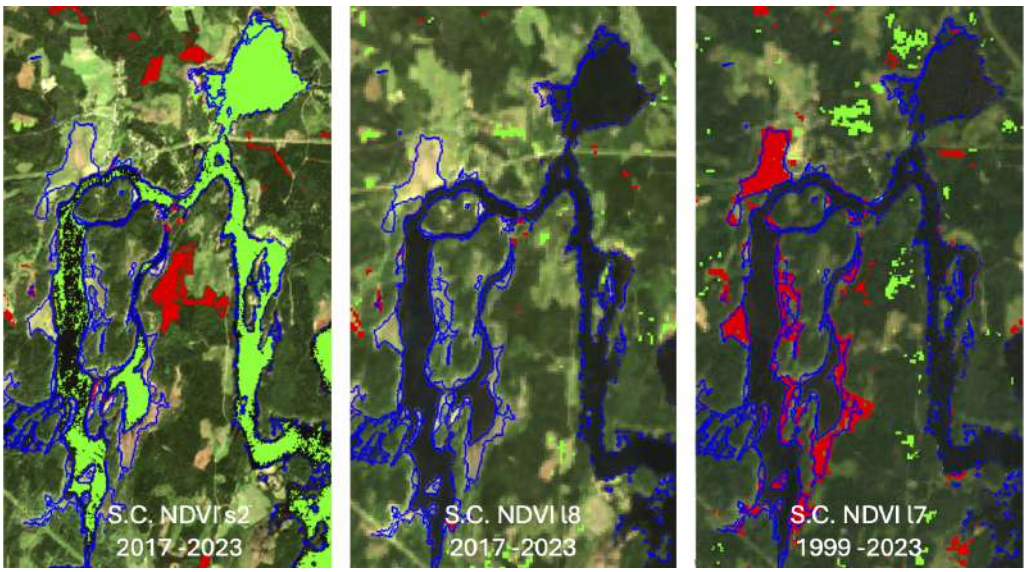

Figure 13: NDVI change detection. From left to right: change detection between 1) Sentinel-2 summer composites from 2017 and 2023, 2) Landsat 8 summer composites from 2017 and 2023, and 3) Landsat 7 summer composites from 1999 and 2023. Red indicates a significant negative change in NDVI values, while green indicates a significant positive change. Wetlands are delimited in blue.

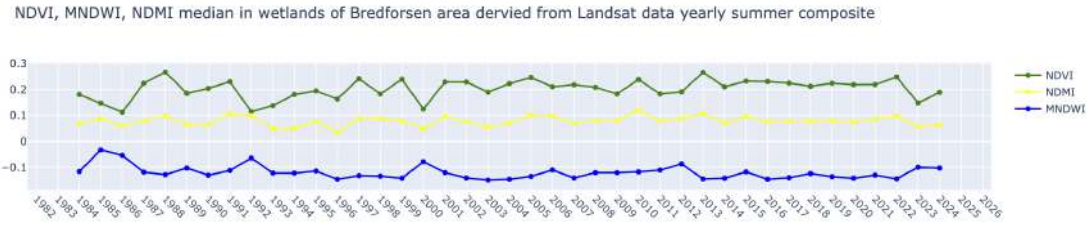

Figure 14: NDVI, MNDWI, and NDMI time series in wetlands derived from Landsat yearly summer composite 1984-2023.

Figure 13 and Figure 12 illustrate change detection between different dates; Sentinel-2 summer composites from 2017 and 2023,Landsat 8 summer composites from 2017 and 2023, and Landsat 7 summer composites from 1999 and 2023. Red indicates a significant negative change in NDVI values, while green indicates a significant positive change, according to the threshold fixed previously. Changes in NDVI values in water are not considered, as they can be influenced by various factors, relevant or not, that are beyond our expertise to interpret. We especially focus or interested on changes occured in the wetlands. If no big changes was detected between 2017 and 2023, with both NDVI and NDMI, A big dicrease of the NDVI and NDMI in the entire wetlands has been observed in the wetlands.

The conclusion regarding these changes remains limited. Indeed, if we consider Figure **??**, which presents the annual values of spectral indices (NDVI, MNDWI, and NDMI), we can observe that the median NDVI value for 2023 is particularly low compared to 1999 and does not reflect the general trend of NDVI, which has been relatively constant. This indicates that spatial changes between two dates cannot be used for accurate conclusions without prior knowledge of the broader-scale dynamics of the ecosystem.

Conclusions from both spatial and temporal changes should be considered preliminary with respect to the main subject of this thesis. While NDVI is a useful indicator of vegetation health, no studies have been conducted to understand the correlation between NDVI values and the Northern Boreal Alluvial Meadows ecosystem. In particular, Shrub encroachment, which is generally considered detrimental to the habitat type studied (Section 2.1.3), may not be adequately captured by NDVI median value. This raises the question of whether woody encroachment has a significant impact on NDVI values. If it does not, alternative techniques may be more appropriate for monitoring these changes.

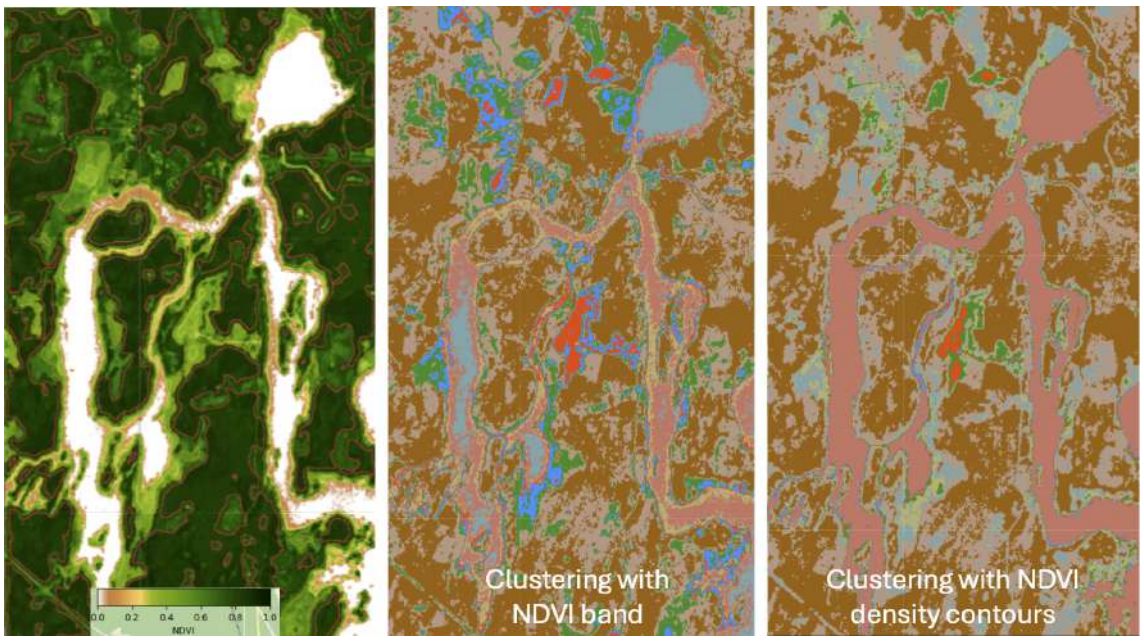

Figure 15: NDVI band from S2 image with it's density contours (Left), NDVI band clusters using only NDVI band as training (Center), and NDVI band clusters using NDVI density contours (Right).

A more comprehensive analysis and field investigation are needed to validate these findings and to further explore the complex dynamics at play in this ecosystem. The choice of indices was made based on a literature review and considering the limitations of spectral signature investigations in ecosystem studies, but it could also be subject to further development, as indexing remains the simplest and most accurate way to extract information from multi-spectral data.

6 NDVI density clustering

If no researchs has been found considering Landsat/Sentinel NDVI values to characterize precisely different type of vegetation, some studies has been found regarding NDVI value difference between different vegetation land cover type. For exemple, Seong and al. (2020) [32] showed that NDVI values of croplands and grasslands, and mixed forest had bigger annual variations than shrub, and evergreen forest, considering for exemple MODIS NDVI product. This study also showed that NDVI values from (AVHRR) instruments had great differences among land cover types, in particular, that forest has the highest values, between croplands and grass. Shrublands seems to have the biggest variance in NDVI values. Rama Ewing work (Master thesis, 2022) [35] proposed a method for Spatial changes detection in Alpine Meadows, using Unsupervised Clustering of NDVI Landsat bands, and NASA NDVI interpretation guide.

In order to precise to study made in Section 5, we propose a new approach to study the changes of different clusters of NDVI values, from which we make the hypothesis than certain NDVI values can be associated with a certain vegetation land cover type. This approach allows also a potential in understanding better the spatial dynamics in vegetation that occurs in Bredforsen.

6.1 Material and Methods

6.1.1 Material

For this method, we choose to use only yearly summer composite data, from Sentinel 2, Landsat 8, and Landsat 7 for simplification and to maintain consistency with considerations outlined in the previous section. While this approach reduces the granularity of information regarding vegetation dynamics, since different vegetation types might exhibit varying or more pronounced changes throughout the season, it [32], still captures the median NDVI values across these types.

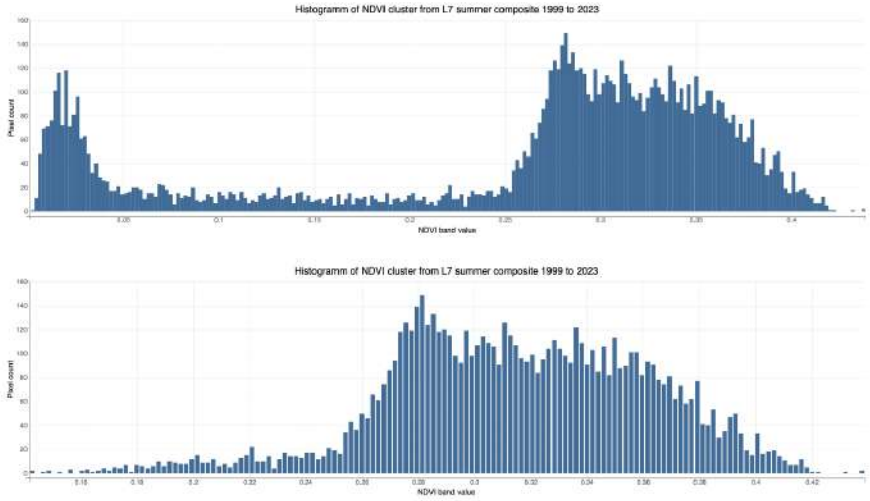

Figure 16: On top, Median histogram of NDVI derived from the yearly summer composites of L7 between 1999 and 2024. On bottom, the same histogram excluding pixels of the image corresponding to water bodies

6.1.2 NDVI density contours

We utilized a specific function from the `geemap` Python package to calculate NDVI contours, which characterize NDVI density across the landscape. These contours are illustrated Figure 15 (left image) presenting the NDVI band, and NDVI densty contours in red. We can identify that they are delineate different land cover types, such as dense forests and meadows. To quantify areas with similar NDVI values, we performed unsupervised clustering, using the WEKA Kmeans algorithm [38]. Using these NDVI contours as a training set. The resulting clusters, generated using only the NDVI band, showed improved delineation of land cover types compared to clustering without contour-based training, as illustrated in Figure 15, center image and right image. When the number of clusters is set equal to the number of density steps, the clusters align with the NDVI contours (Figure 15), as shown by the comparison between images on the left and the one on the right.

To ensure that the clusters were interpretable, we optimized the parameters for NDVI density steps, as well as the maximum and minimum NDVI values. Initially, we attempted to identify optimal parameters through visual interpretation, but observed significant fluctuations between different Landsat composites. Furthermore, histogram of NDVI values did not exhibit a Gaussian distribution. Figure 16 present de median histogram of Landsat yearly summer composite from years between 1999 and 2023, before removing water mask (On top) and after (On bottom). The prominent peak at low NDVI values corresponded to water bodies, while the rest of the values primarily represented vegetation, as the Bredforsen area contains little built-up land. We therefore choose to conduct a removal of water on our images before conducting the clustering, to focus especially on vegetation.

6.1.3 Water detection on optical images

To address this, we created a unique water mask for each dataset (Landsat 7, Landsat 8, and Sentinel-2) to exclude water bodies from our analysis The method is illustrated Figure 17. From image left to right; 1. Water spectral indices were calculated using `eemont` Python package, a total of 10 water spectral indices were calculated, to ensure an accurate classification and 2. A binary Kmeans clustering were process on those indices. Finally, 3. identification of the water cluster was made calculating the mean value of MNDWI band of each cluster.

6.1.4 NDVI clusters

We then established a fixed number of density steps by analyzing the median histogram of each dataset. A density step was defined as the standard deviation of the median histogram, with the minimum and maximum NDVI values set at three times the standard deviation. This method resulted in a total of seven clusters for each dataset.

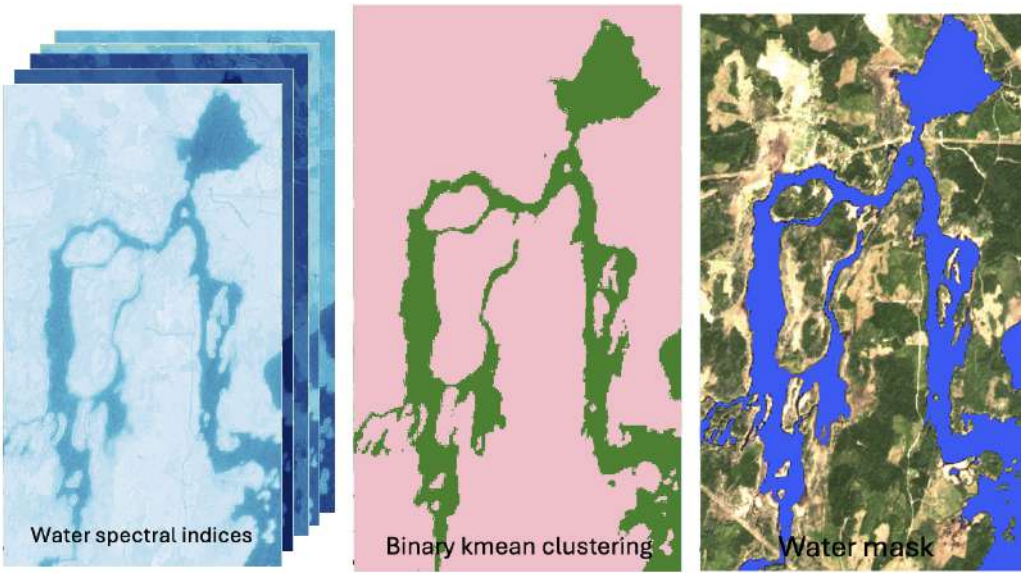

Figure 17: Illustration of water detection using optical satellite data. From image left to right; Water spectral indices, Result of binary Kmean clustering, Final mask after identification

After clustering, we needed to associate each cluster with a specific NDVI range. We calculated the mean NDVI for each cluster and ordered the clusters from 0 to 6, corresponding to increasing mean NDVI values. The range of NDVI values for these clusters, although subject to the uncertainty inherent in the K-means clustering method, tends to remain consistent across images from the same sensor. The method is equivalent to manually calculating the density and classifying according to different density levels.

6.2 Results and discussion

Nota bene. In this section, only spatial changes would be presented, as the algorithm used for cluster identification has encountered several issues related to client side operation when extended it to an `ImageCollection()`. This matter is mentioned as was experienced several times using the `.map()` function of GEE, and can complicate the development of certain functionalities.

Figure 18, 19, and 20, present NDVI density clustering of both composites chosen to compute spatial changes. Parameters for density contours has been found as followed; For S2, NDVI contours range from 0.4 and 0.8, with a 0.052 steps. For L8, NDVI contours range from 0.1 and 0.4, with a 0.041 steps. For L7, NDVI contours range from 0.15 and 0.4, with a 0.047 steps. The variety of colors in the change detection results relates to the intensity of these changes. For example, a pixel changing from the cluster with the lowest NDVI mean value to the cluster with the highest NDVI mean value can be seen as a positive high-intensity change (dark green). Conversely, a pixel changing from Cluster 2 to Cluster 1 indicates a negative small-intensity change (light red). This method provides a more accurate way to detect spatial changes compared to the results presented in the previous section. While the choice of density clusters might be revised later, taking into account the interpretation of the different ranges of NDVI values for each sensor, the detected changes are less dependent on the thresholds used. However, we could propose other decision criteria for the initial method using NDVI, such as the direct classification of different types of NDVI changes in the first section, which would, in fact, reduce the constraints associated with a single threshold and changes varying in intensity.

This method still appears to be more precise because it accounts for the dynamics of "objects" representing ecosystems with similar reflectance behaviors (in the red and NIR bands), thereby reducing the noise caused by single pixels exhibiting particular behaviors.

Figure 18 present similar results as Figure 13, With a negative change of NDVI value in the entire wetland area between 1999 and 2023. Differing from Figure 18 in the spatial changes between 2017 and 2023 using L8 or S2, Figure 19 and 20 present some small changes in the wetlands. If they are underestimated with Landsat sensors, and mostly positive changes (Figure 19), they are largely negative with S2 (Figure 19). Without actual quantification, these two figures offer a comparison between the precision of changes detected on S2 images, with a 10m resolution, and L8

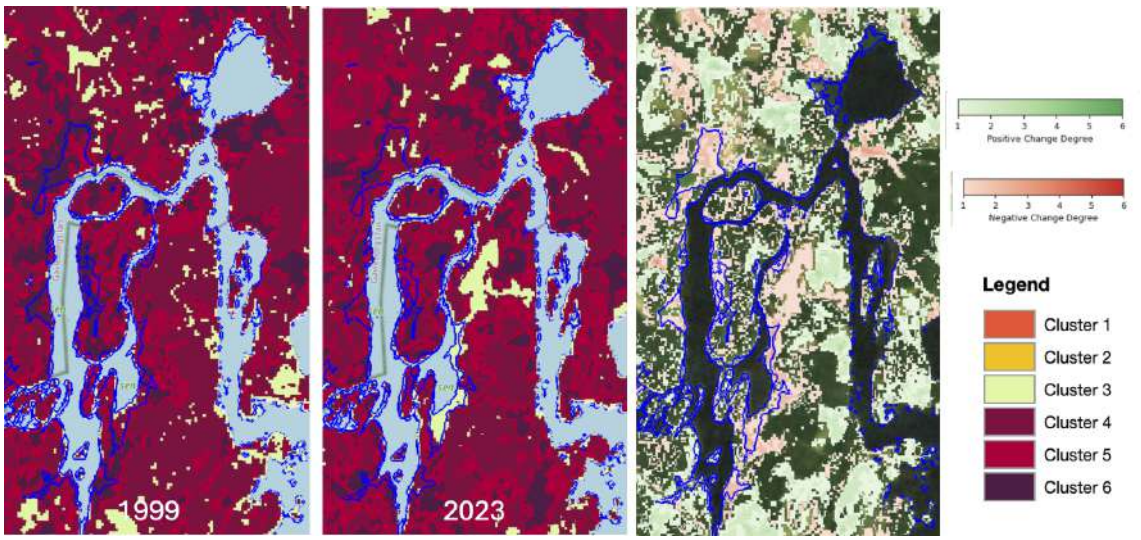

Figure 18: NDVI density clustering derived from Landsat 7 1999 summer composite (Left), 2023 summer composite (Center) and NDVI density spatial changes between the two. Legend of clusters are illustrated on the right, as well as level of changes.

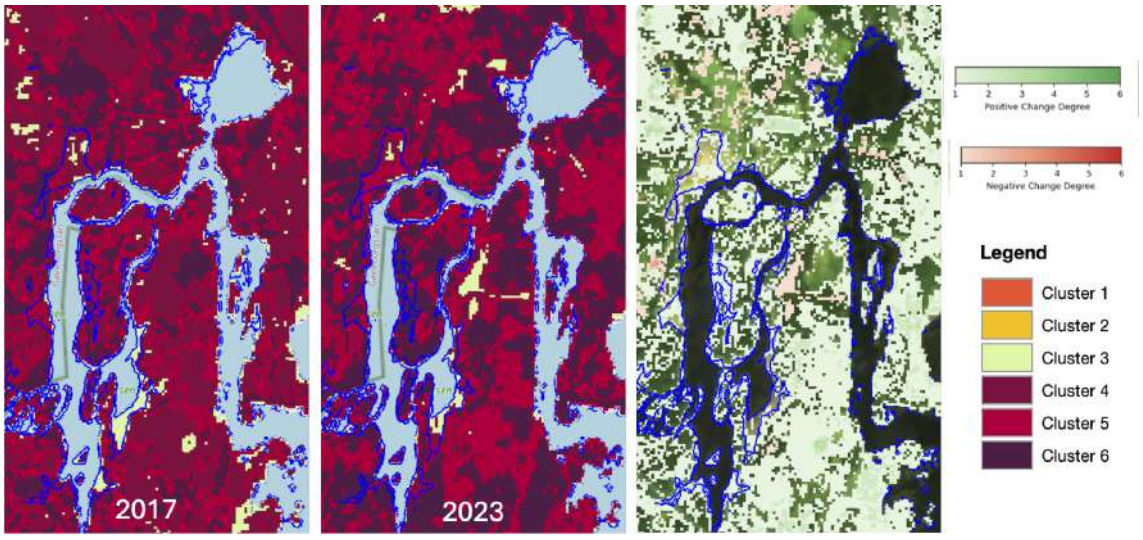

Figure 19: NDVI density clustering derived from Landsat 8 2017 summer composite (Left), 2023 summer composite (Center) and NDVI density spatial changes between the two. Legend of clusters are illustrated on the right, as well as level of changes.

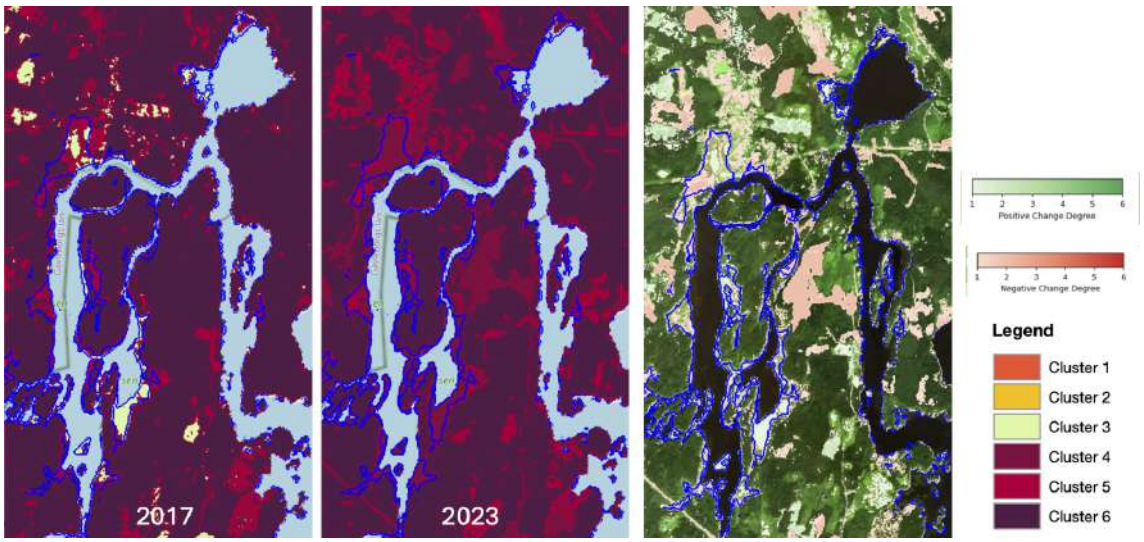

Figure 20: NDVI density clustering derived from Sentinel 2 2017 summer composite (Left), 2023 summer composite (Center) and NDVI density spatial changes between the two. Legend of clusters are illustrated on the right, as well as level of changes.

images, with a 30m resolution, which, moreover, have different radiometric behaviors. Although S2 appears to be more precise in detecting changes due to its higher spatial resolution, a more in-depth study, following the interpretation of these changes, could be conducted to determine whether the resolution of Landsat is sufficient for this task.

In the same way as the conclusion reached in the previous section, spatial changes should be interpreted considering the temporal context of the observed changes, to avoid limiting the analysis to specific cases, as we observed between the spatial changes from 1999 to 2023. Furthermore, the NDVI density method appears to be relevant for a precise analysis of vegetation dynamics changes, but the conclusions drawn can only be considered preliminary for this study without prior fieldwork analyzing the relationship between NDVI values and vegetation species.

7 Land cover classification

Another approach tried to detect spatio-temporal changes in Northern Boreal Alluvial Meadows is to detect the changes in the land cover type. As those meadows are expected to be encroached by woody plants (See section 2.1.3), we would like to understand the spatio-temporal changes in Bredforsen area such as proportion of trees, schrub and scrub, flooded meadows and what can appears as grass. For that, we explored several LULC classification models to ensure some time series of proportion of each land cover type, as well as spatial changes dynamics between to land cover images.

7.1 Material and Methods

In this study, we opted for a pixel-based classification approach for land use and land cover (LULC) mapping. This decision is motivated by several methodological and practical considerations. On one hand, pixel-based approaches, as demonstrated by Berhane et al. [39], offer simplicity and efficiency that are often sufficient for accurate classifications. Although object-based (OBIA) methods can provide advantages in certain contexts, they typically require more resources and complex decision-making, such as segmentation scale selection, which can be constraining in large-scale environments. Furthermore, some OBIA classification, using the well know Segment Anything Model [40], has been tried in Bredoforsen image and dismiss, as it requires important manpower to validate and interprete objects detected.

Random Forest (RF) is widely used for LULC classification due to its robustness, versatility, and effectiveness when dealing with complex and large datasets. RF is a non-parametric machine learning algorithm, meaning it does not make assumptions about the data distribution, making it particularly effective in handling non-linear data and variables with different characteristics

simultaneously [44, 45]. This distinguishes it from traditional algorithms like ISODATA and K-Means, which are sensitive to outliers and collinearity [39].

The advantages of RF include high classification accuracy, the ability to determine variable importance, and increased resistance to noise and outliers, thereby reducing the risk of overfitting [44]. RF works by combining the results of multiple decision trees (hence the name "random forest"), which helps to minimize errors and increase the reliability of the results [43]. Additionally, it offers great flexibility for various types of analyses, including regression, classification, and unsupervised learning [45]. In the context of Google Earth Engine (GEE) applications, RF is frequently used due to its fast execution and its ability to handle large datasets on a planetary scale, making this algorithm a popular and reliable choice for LULC classification [46].

Temporal aggregation, which involves combining satellite images taken at different times to create temporal composites, is crucial for LULC mapping, especially in regions where atmospheric conditions, such as cloud cover, can affect image quality [46]. By using techniques such as the median or mean of a set of images over a period of time, it is possible to reduce the impact of outliers and fill data gaps, leading to more reliable and complete results [43].

This approach is particularly advantageous for monitoring land cover changes over large areas because it captures specific seasonal patterns and better distinguishes between different LULC classes, even in complex environments [46]. Temporal aggregation also simplifies data processing by reducing the volume of data to be analyzed while maintaining high classification accuracy. For example, in studies where summer composites were used, the results showed higher accuracy compared to using single images, highlighting the effectiveness of this method in improving the quality of the LULC maps produced [39, 42].

However, it's essential to identify optimal combinations of satellite data and aggregation parameters to ensure that the accuracy of automated, temporally aggregated datasets matches that of traditional, manually selected image composites. For instance, combining optical and radar data can enhance classification accuracy.

For the supervised classification of Land Cover, we are exclusively utilizing data from the Landsat missions. If combining optical and radar data can enhance classification accuracy as demonstrated by Carrasco and al. study, [41], where multi-sensor datasets outperformed single-sensor datasets in land cover mapping, Landsat data is still preferred because it allows for the detection of changes over a time span that neither Sentinel-1 nor Sentinel-2 can adequately capture. Furthermore, combining dataset requires additional computational data sources, which was very limited in GEE. The exclusive use of SAR data, was also excluded as Sun and al. [43] suggest that Sentinel-1 does not significantly enhance classification accuracy compared to optical imagery from Sentinel-2. MODIS data was considered but ultimately excluded due to its lower spatial resolution, which does not provide the necessary precision for our analysis.

As for sampling methods, three are typically employed: on-site investigation, visual interpretation, and referencing existing LULC products. [47].

Two sampling methods has been tried : One using visual interpretation and annotation of data helped with high resolution airplane data, as well as the use of LULC product such as the Dynamic World Land Cover.

ESA World Cover, another LULC product available in GEE, was considered but ultimately excluded. Work by Venter et al. [48] highlights that, firstly, ESA World Cover had lower global accuracy than other datasets (Dynamic World, ESRI). Secondly, ESA World Cover tended to overestimate grasslands compared to other datasets, which could potentially lead to misinterpretations in temporal classification maps. Additionally, ESA World Cover only provides maps on an annual basis, which limits the possibilities for more frequent temporal analyses. Although ESA World Cover's advantage lies in its ability to "characterize precise features," this was not relevant in our case, as we are primarily using Landsat 8 data. The level of detail provided by ESA World Cover is not particularly useful for detecting the features in the Bredfrosen landscape, based on our understanding of these ecosystems.

7.1.1 Approach 1: Using Hand annotated data.

Sentinel 2 and Landsat imagery and their 30m and 10m spatial resolution doesn't allow to easily identify visually land cover types, especially concerning our land cover surfaces studied, such as Northern Boreal Alluvial Meadows, and shrubs and bushes species, as well as flooded vegetation.

Airplane high resolution imagery from a summer 2023 flight has been provided by Vattenfall's biodiversity team, allowing some hand annotating data. This dataset allows a better hand made definition of land cover classes, that has been been defined as ; Trees, Northern Boreal Alluvial Meadows, Water, built, and Bare.

Training data has then been annotated using a summer composite from Landsat 8 data, and classes polygons (See Figure 22, Image from the left). A training samples dataset of 80% - 20% picked randomly has been made, and RF model has been trained one time, using optimal parameters proposed by Sun J. and al. (2023) [43], such as 800 trees, and 0.9 Bag Fraction.

7.1.2 Approach 2: Utilizing the Google Dynamic Dataset as reference data

A workflow diagram for Model 2 is presented Figure 21.

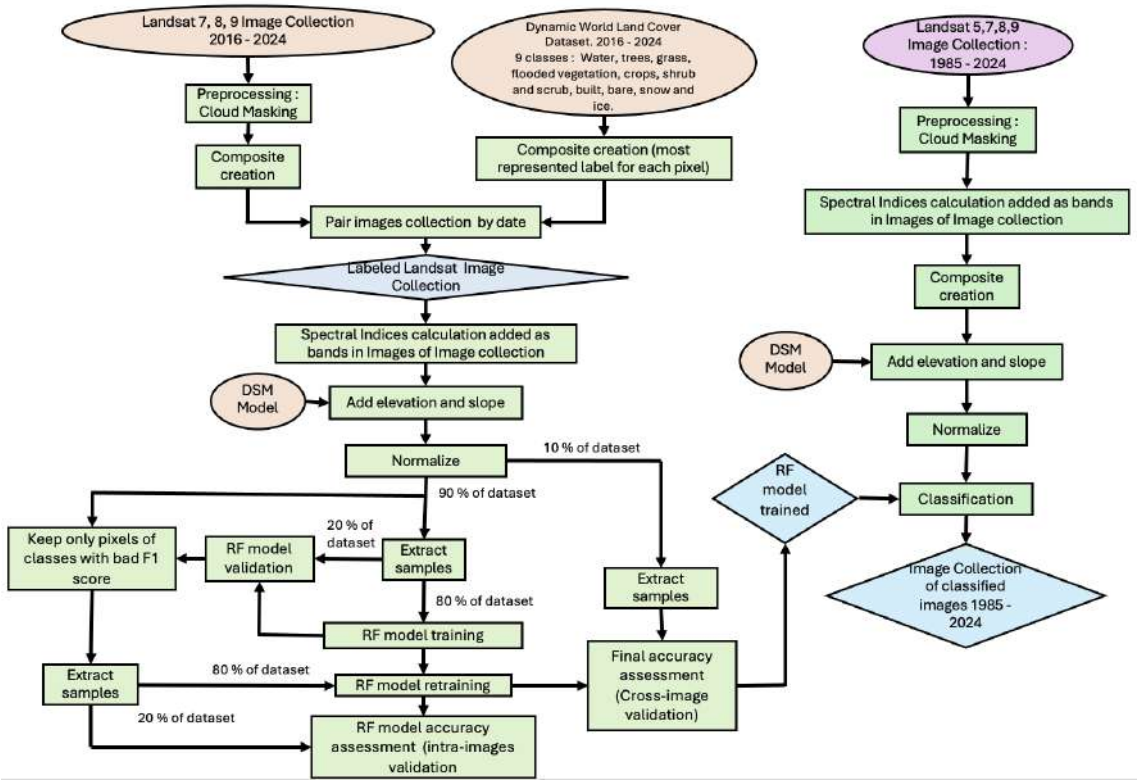

Figure 21: Workflow diagram of Model 2 for LC classification

Dynamic world land cover maps are available from 2015-06-27 to present, with a 2-5 days revisit time, as the dataset is based on predictions of Sentinel-2 L1C collection. This frequency is not considering cloud masking, that is done before prediction, and can conduct to a lot of data missing in certain regions. Dynamic world countains 9 classes, such as water, trees, grass, flooded vegetation, bare, schrub and scrub, built, and snow and ice.

Dynamic world description for classes of interest, is made as followed [12] ;

- **Grass**: Open areas covered in homogenous grasses with little to no taller vegetation. Other homogenous areas of grass-like vegetation (blade-type leaves) that appear different from trees and shrubland.
- **Flooded Vegetation**: Areas of any type of vegetation with obvious intermixing of water. Seasonally flooded areas that are a mix of grass/shrub/trees/bare ground.
- **Shrub & Scrub**: Mix of small clusters of plants or individual plants dispersed on a landscape that shows exposed soil and rock. Scrub-filled clearings within dense forests that are clearly not taller than trees. Appear grayer/browner due to less dense leaf cover.

Particular attention will be paid to changes in these land cover types, as the "Grass" and "Flooded Vegetation" classes could correspond to the Northern Boreal Alluvial Meadows in the Bredforsen area, while "Shrub and Scrub" may be considered as the problematic land cover type in this context 2.1.3.

We separated our approach in three models, each using different temporal aggregated data, in order to compare their performance.

Model 2.a We utilized a total of 23 annotated images from the Landsat series: 12 from Landsat 8, 5 from Landsat 7, and 6 from Landsat 9 (20 for training, 3 for validation). Each image contained 21,459 pixels at a 30-meter resolution within the region of interest. These images were annotated using the Dynamic Land Cover dataset as a reference, with labels corresponding to the acquisition dates of the Landsat images. The use of Sentinel-2 data for this model is not relevant, as the Dynamic World Land Cover dataset is already available for each Sentinel-2 observation.

Model 2.b Y. Piao and colleagues [47] proposed an approach for generating training samples for LULC supervised classification by defining criteria such as "complete consistency" and "temporal stability." This method reduces the number of pixels selected in the sampling process, mitigating computational limits within GEE, without increasing the risk of overfitting, particularly in terms of weather conditions.

Given the high variability of our classes and the lack of other Land Cover datasets that matched Google Dynamic World classes easily, we chose to base our criteria exclusively on temporal stability. Specifically, we retained pixels that exhibited the same class 80% of the time in our reference data. Since snow and ice classes were obviously not invariant, we chose to exclude the winter season from our training and validation datasets. The model was then trained as in Model 2.a, with an increased number of samples.

Model 2.c Using single image observations for Land Cover classification can lead to higher classification errors due to increased variance and noise in the data, which is sensitive to atmospheric and other conditions [46]. To mitigate this issue, we compared the performance of Model 2.a with a model trained on monthly composites. We utilized a total of 41 monthly composites from Landsat 7, 8 and 9 images from years 2022 and 2023 for training, and 15 from year 2021 for validation. Composite were created separated for each sensors. These images were annotated with monthly composite of Dynamic Land Cover dataset, created by using the `Mode()` aggregation method, selecting the most prevalent class for each pixel.

Model 2.d We also compared of Model 1.a with a model trained on yearly summer composites. We utilized a total of 4 summer composites from Landsat 7, 8 and 9 images from years 2020 to 2023, and 2 from year 2019 for validation. Composite were created separated for each sensors. These images were annotated with yearly summer composite of Dynamic Land Cover dataset, created by using the `Mode()` aggregation method.

For each models, we trained the Random Forest (RF) model using a fixed number of samples for each image. These samples were then concatenated and split into an 80% training and 20% validation dataset. Various hyperparameters, including the number of trees, the number of variables per split, the bagFraction, as well as the number of samples and input bands, were explored to find an optimal combination.

A commonly used sampling method in Google Earth Engine (GEE) is the `StratifiedSamples()` function, which helps prevent unbalanced training-validation datasets. However, due to an issue that remains unresolved at the end of this thesis, this method significantly limited the number of samples allowed when sampling across an entire collection, not just a single image. After investigating suitable hyperparameters to circumvent this limitation, we ultimately opted to use the `sample()` method. This approach enabled us to train more samples and trees. To address the problem of unbalanced classes, given the disproportionate representation of different surface classes in the Bredforsen area, we trained the model twice: first, including all classes; and second, excluding classes that were well classified—namely, trees and water—since they are the most represented in the Bredforsen area.

Accuracy assessement metrics

As we had a consequent number of reference data at disposition when using Dynamic World dataset, we choose to do the validation with two approachs; 1) Validation of pixel-classification that haven't been seen in the training, but from the same image as the pixels used for the training that we will call "*intra-image validation approach*". 2) Validation of pixel classification coming from images that any pixels haven't been seen in the training, that we will call "*cross-image validation approach*". This double validation ensure that the model can be well generalized to the whole Landsat dataset, without overfitting of atmospheric and weather conditions present in the training.

The accuracy of any land use and land cover (LULC) classifier depends heavily on factors such as the amount of input data, the characteristics of the study area, and the specific satellite sensor

used. Several studies have shown that different classifiers yield varying levels of accuracy across different regions and data sources [45].

The most commonly used indicators for accuracy assessment are the overall accuracy (OA), the kappa coefficient, and the F1 score.

Overall Accuracy (OA) is the ratio of correctly classified pixels to the total number of pixels classified. It provides a general measure of classification performance:

$$\text{Overall Accuracy} = \frac{C_p}{T_p} \times 100$$

where:

- C_p is the number of correctly classified pixels,
- T_p is the total number of pixels.

Kappa Coefficient is another widely used metric that assesses the agreement between the predicted classifications and the true classifications, adjusting for the agreement that might occur by chance. It ranges from 0 (no agreement) to 1 (perfect agreement):

$$\kappa = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \times x_{+i})}$$

where:

- N is the total number of observations,
- r is the number of rows in the confusion matrix,
- x_{ii} is the sum of diagonal elements (correct classifications),
- x_{i+} is the sum of row i,
- x_{+i} is the sum of column i.

User's Accuracy indicates the probability that a pixel classified into a given category actually represents that category on the ground:

$$\text{User's Accuracy} = \frac{\text{properly classified pixels of class}}{\text{total classified pixels of class}} \times 100$$

where:

- properly classified pixels of class is the number of pixels correctly classified for a given class,
- total classified pixels of class is the total number of pixels classified for that class.

Producer's Accuracy measures the probability that a reference pixel was correctly classified:

$$\text{Producer's Accuracy} = \frac{\text{properly classified pixels of class}}{\text{total reference pixels of class}} \times 100$$

where:

- properly classified pixels of class is the number of pixels correctly classified for a given class,
- total reference pixels of class is the total number of reference pixels for that class.

F1 Score is the harmonic mean of precision and recall, and it is particularly useful in cases where the dataset has imbalanced classes or where some classes are more significant than others, which is the case in Bredforsen's scenery.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

- Precision (or user's accuracy) is calculated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- Recall (or producer's accuracy) is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

In this study, these metrics were calculated using the confusion matrix embedded in Google Earth Engine (GEE), which provides a reliable and automated means of validating and evaluating classification accuracy [45]. F1 scores allow class-specific performance assessment. A particular attention has been made towards F1 scores of classes of interest, such as classes characterizing Northern Boreal Alluvial Meadows, and woody encroachment (See. **??**).

7.2 Results and discussion

7.2.1 Approach 1: Using Hand annotated data.

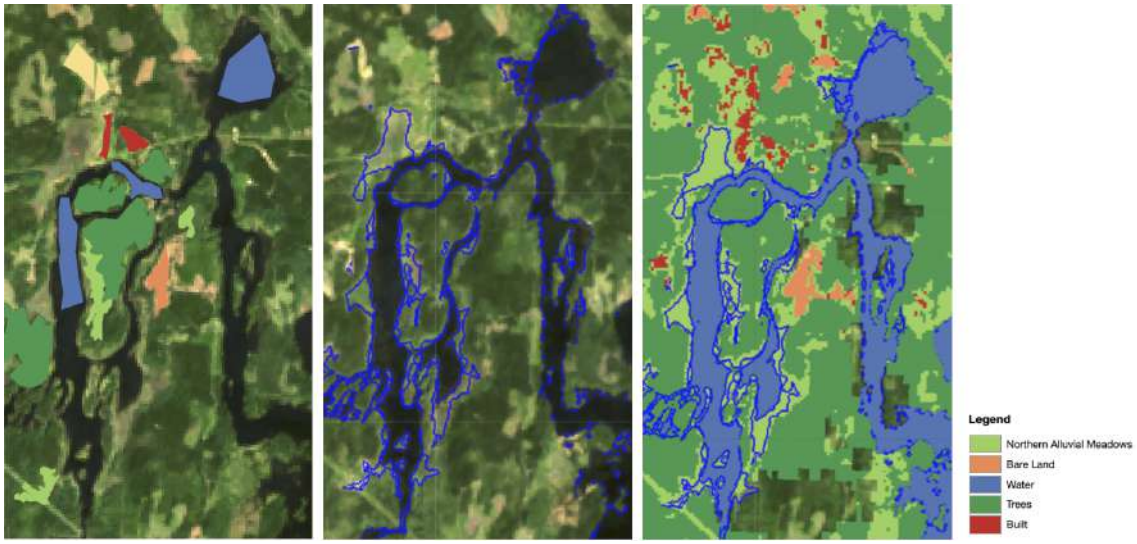

Figure 22: From left to right. Annotated polygons on L8 composite from summer 2023. L8 summer composite and wetlands (in blue). Classification of this composite using model 2.

Figure 22 illustrates the results from the classification of a single Landsat 8 image composite, using Model 2. Since the polygons were annotated for summer 2023, no cross-image validation could be performed. The intra-image validation yielded an overall accuracy of 0.95 and a Kappa score of 0.93.

However, several issues arise with this method. Firstly, our previous studies, which involved manually inspecting several yearly summer composites and their classification maps using Model 2, have shown that this approach can lead to overfitting to the atmospheric conditions present in the 2023 composite used for training. This hypothesis is further supported by the poor accuracy and Kappa score (0.76 and 0.6, respectively) found in cross-image validation for Model 1.d, which was trained under the same conditions—using a single yearly composite for training and applying it to yearly summer composites for other years. Secondly, the annotated data is limited both in quantity and in the range of classes defined, particularly in distinguishing between Northern Boreal Alluvial Meadows and shrub & scrub, as well as between trees and shrub & scrub.

Although the use of this model with the annotated data was found to be imprecise and irrelevant for studying changes, the results highlight potential challenges related to future field work. Firstly, relying on single-date field data could lead to over-fitting of the Model to specific atmospheric and weather conditions. Secondly, the ability of satellite data to discriminate, in this specific case, between Grasslands and Flooded Vegetation (or Northern Boreal Alluvial Meadows) and Shrub & Scrub, may be compromised. This matter is discussed in the next section.

7.2.2 Approach 2: Using Google Dynamic Dataset as reference data

Accuracy assessment

Kappa scores and F1 scores were calculated at each training step (1 and 2) as well as during the final validation.

The testing of parameter combinations was somewhat limited due to: 1) the computational constraints of Google Earth Engine (GEE), which restricted our ability to automate the testing of all hyperparameters combinations, and 2) the large number of parameters involved and the time required for testing.

For simplicity, key model parameters—such as the number of trees, bag fraction, and variables per split—were primarily tested using Model 2.a, trained on Landsat 8 annotated data, and were more or less fixed thereafter.

Several conclusions can be drawn regarding the hyperparameters of the model, particularly for Model 2.a trained exclusively with Landsat 8 data. The combination of 100 trees, 4 variables per split, a 0.9 bag fraction, and the use of 10 input bands (reflectance bands from Landsat 8, as well as NDVI, NDMI, MNDWI, elevation, and slope) with a sample size of 2,000 (80%-20% split) yielded the best results, with an overall accuracy of 0.92 and a Kappa score of 0.86. The limit of 100 samples was determined by GEE's computational constraints, given the 10 input variables and 2000 samples per image.

Considering a smaller number of input variables, increasing the number of trees from 100 to 200 or 300 did not significantly improve the results. Further increasing the number of trees from 300 to 500 resulted in a slight decrease in overall accuracy and Kappa score (from 0.877 to 0.86 after the first training). Increasing the number of samples from 2000 to 5000 improved the Kappa score from 0.91 to 0.92. However, no definitive conclusions were drawn due to the computational limitations of GEE. For example, increasing the number of trees from 500 to 600, while keeping the same number of samples and input variables, led to memory issues. Therefore, these results should be considered preliminary and may require further refinement if a method to overcome GEE's computational limits is found. We also tested different `tileScale`, in sampling method parameters, but this did not resolve the issue. Generally, increasing the number of input variables (bands) gave better classification results.

Model 2.a, trained exclusively with Landsat 8 data, produced significantly poorer results when compared to training with data from Landsat 8, Landsat 7, and Landsat 9. Specifically, when using the cross-image validation approach (using images from both three sensors), the model yielded a Kappa score of 0.63 and an overall accuracy of 0.77, compared to a Kappa score of 0.73 and an overall accuracy of 0.88 when trained with data from all three sensors. This result was predictable, but still necessary to mentioned, as it could have been, for computational manner, interesting to train the model only on Landsat 8, if the results weren't significantly different.

Additionally, using monthly composites for classification resulted in better outcomes than classifying each single cloud-free observations with Model 2.a. This change improved the Kappa score of the final Model 2.a. hyperparameters and inputs combination from 0.73 to 0.76 in cross-image validation. However, it is important to consider that in the first case, the cross-image validation was performed on three images (with 2000 samples), whereas in the second case, 14 images were used for cross-image validation. This difference in the number of images could introduce variability in the scores, potentially affecting the results depending on the images selected.

Model that led to the better results is Model 2.d using yearly composite trained with 2020, 2021, 2022, and 2023 yearly summer composite composite, and validated on 2019 yearly summer composites (from L7 and L9) with a 0.87 kappa score for intra-image validation approach and 0.79 for cross-image validation approach. However, this result might again be considered in a sense that Model 2.c, cross-image validation were performed on 15 monthly composites (from L7 and L8), as Model 2.d cross-image validation were performed on only 2 yearly summer composite (from L7 and L8). A remark has to be made that those models hasn't been cross-image validated with L9 data. Model 2.b was found to be irrelevant, as training and validation on only temporally stable pixels led to an almost complete disappearance of the grass and shrub & scrub classes, which are crucial for the study. A more effective approach might involve selecting pixels using other LC datasets, similar to the methodology presented by Y. Piao and colleagues [47]. However, this method requires finding multiple LULC datasets with similar, yet relevant classes for the study, as well as sufficient temporal and spatial resolution.

Additionally, some training was conducted using only wetlands areas for the training samples, which yielded poor results, with Kappa scores below 0.6 in both cross-image validation and intra-image validation within wetlands, and an overall accuracy of less than 0.71 for both Model 2.c and 2.d. Several factors can be explained by; 1) The reduction in the study area, combined with the constraint of increasing the number of observations—limited by the frequency and temporal scale of DW—leads to a decrease in the number of training pixels, and 2) The inherent variability of wetland environments, which exhibit particularly fluctuating surface conditions.

As seen with Model 2.b, no pixels from classes present in wetlands demonstrated temporal stability. This underscores the challenges of accurately classifying land cover in such complex

landscapes.

Finally, it is important to consider that the accuracy results are also constrained by the accuracy of the reference data itself. Dynamic World, while valuable, shows varying accuracy across different land cover classes. For example, in the context of wetlands, Dynamic World achieved a user's accuracy of 34.9% for grasslands, 74.1% for flooded vegetation, and 55.6% for shrub & scrub [12]. The global agreement between Dynamic World classifications and expert consensus is 73.8%, but this figure can drop significantly for overlapping classes in complex regions such as wetlands.

Temporal changes

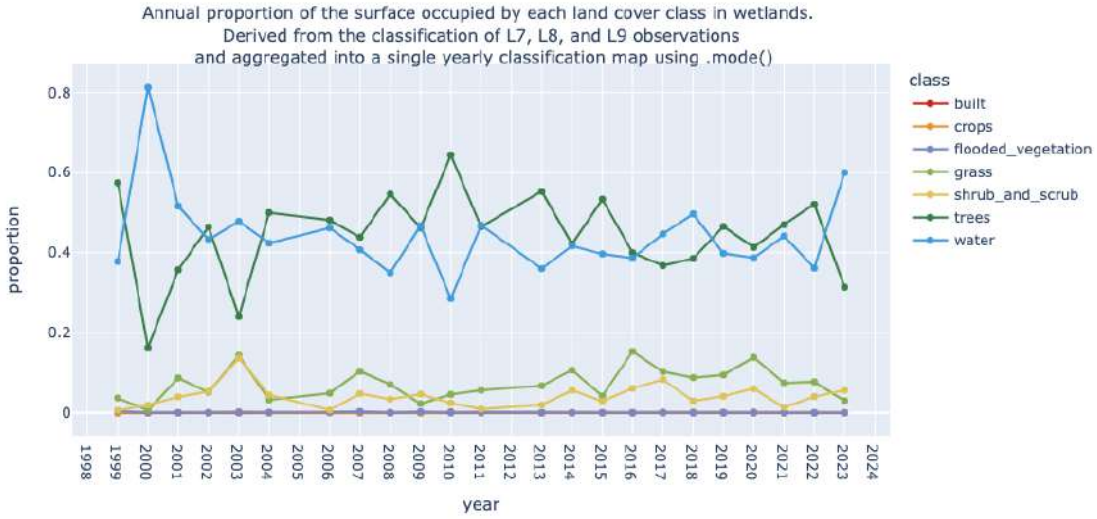

Figure 23: Time series of Land Cover classes from Model 1.a derived from classification of Landsat yearly summer composite.

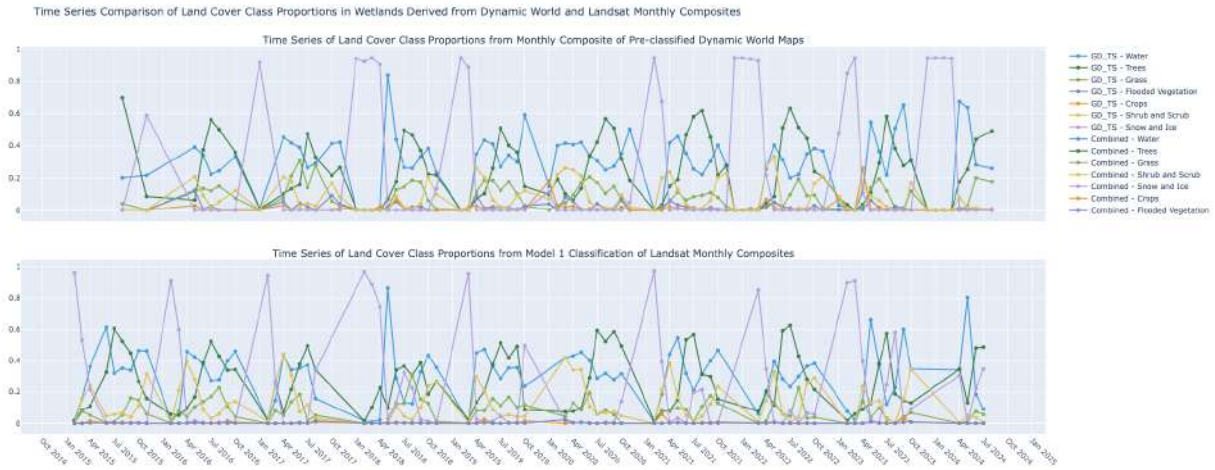

Figure 24: Time series of Land Cover Class Proportions in wetlands from 2015 to 2024. On top, derived from pre-classified Dynamic World map (GD-TS curves). At the bottom, (Combined Landsat curves) derived from Model 1 classification of Landsat Monthly Composite.

Figure 23 present time series of the proportions of each class defined by DW dataset in wetlands, derived from the classification of 24 yearly summer composites from Landsat sensors (7, 8, and 9) using Model 1.d. Figure 23 present time series of the proportions of each class defined by DW dataset in wetlands, derived on top; From monthly composite of preclassified DW classification maps. On bottom; From the classification of monthly composites from Landsat sensors (7, 8, and 9) using Model 1.c. Since we did not create composites from different sensors, an average surface area for each class was calculated for years where two or three composites from these sensors were available.

No significant trends in data quantity or variability between different years can be clearly identified. However, there is a noticeable trend of tree cover decreasing since 2010, following a previous period of growth.

There is ambiguity when interpreting the results using Model 2, particularly in determining whether woody or shrub and scrub encroachment should be attributed to an increase in the "trees" class, the "shrub and scrub" class, or both, potentially at the expense of the "grass" class. Furthermore, as defined in 2.1.6, the canopy coverage of trees and shrub should be less than 30%, ideally around 5%. Pixels with a 30m resolution that have been classified as trees might actually contain a mix of trees and grass with a canopy coverage lower than 30%, yet still be classified as trees.

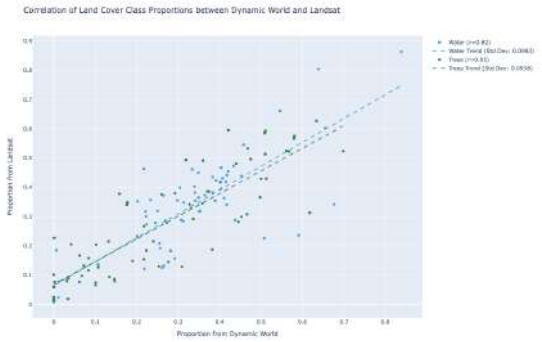

Figure 25: Correlation of grass and trees proportions time series in wetlands from Landsat and Dynamic World.

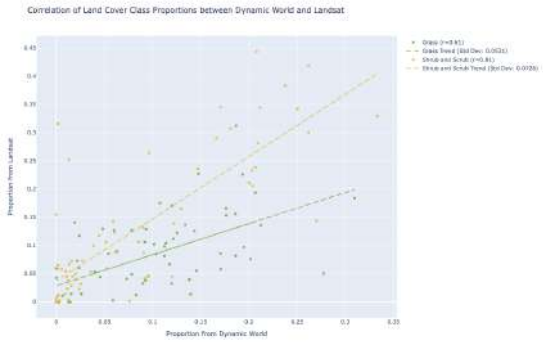

Figure 26: Correlation of grass and schrub and scrub proportions time series in wetlands from Landsat and Dynamic World.

The Dynamic World paper review [12] identified that 4,641 pixels classified as grass were actually trees, while 118,152 pixels of grass were misclassified as trees. Similarly, 478,891 pixels classified as trees were actually shrub and scrub, and 38,111 pixels of grass were misclassified as shrub and scrub. Considering the Landsat 30m resolution, these numbers could be even higher. This mis-classification highlights two key points: 1) Interpretation of changes may be overly pessimistic regarding woody encroachment, potentially overlooking some grassland areas in the study of changes; 2) Interpretation of changes may be overly optimistic regarding shrub and scrub encroachment.

Figures 26 and 25 present the correlation between the 2015-2023 time series of different classes proportions obtains with Landsat monthly composite classified with Model 2, and the one derived directly from the pre-classified Dynamic World dataset. These figures highlight the uncertainty of the time series obtained for the classes of grasslands and shrublands, compared to those obtained for trees and water. We identify three hypotheses for these uncertainties. The first two hypotheses have been demonstrated previously: (1) The lack of accuracy in Model 2 concerning the classification of these specific classes, and (2) The accuracy of DW classification concerning these specific classes, leading to less accurate classification in Model 2 when added. The last hypothesis could be considered or ruled out by more in-depth studies: the inability of Landsat sensors, in terms of resolution and spectral response, to effectively capture and classify these specific ecosystems.

Spatial changes

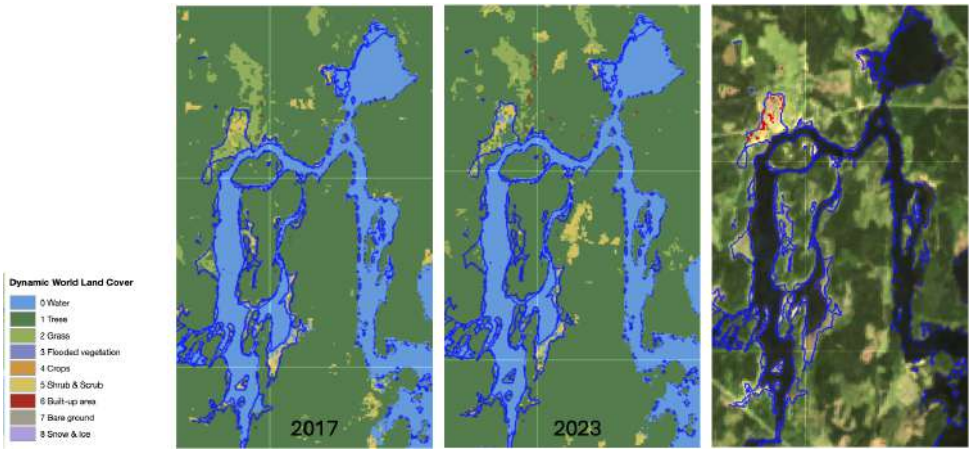

Figure 27: Classification map of 2017 (Left) and 2023 (Right), derived from Landsat 8 data classification. On the right, Landsat 2017 composite with spatial changes detection between 2017 and 2023 of shrub and scrub encroachment (in red) and grass apparition (in green).

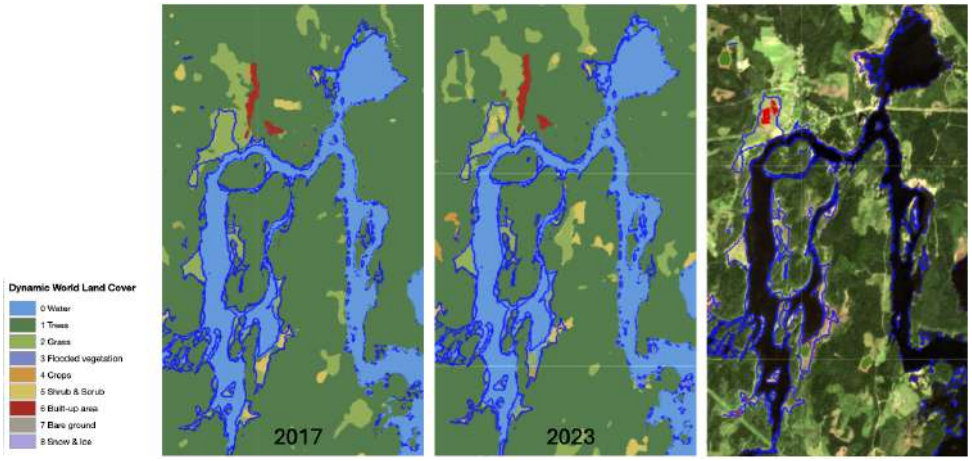

Figure 28: Classification map of 2017 (Left) and 2023 (Right), derived from Google Dynamic dataset. On the right, spatial changes detection of shrub and scrub encroachment (in red) and grass apparition (in green).

Figure 27 presents the spatial changes detected between the classification maps of 2017 and 2023, derived from Landsat 8 yearly summer composites classified using Model 1.d. Additionally, Figure 28 shows the spatial changes detection between the Dynamic World summer composite classification maps of 2017 and 2023. In both figures, the wetland boundaries, as detected in Section 4, are highlighted in blue. Spatial changes have been computed based on two criteria: 1) Grass class replaced by Shrub and Scrub classes between date 1 and date 2 is identified as negative change and displayed in red. 2) Shrub and Scrub classes replaced by Grass class between date 1 and date 2 is identified as positive change and displayed in green. Other changes are not considered in these maps. The background consists of the RGB colors from the 2017 yearly summer composite for both figures 28 and 27.

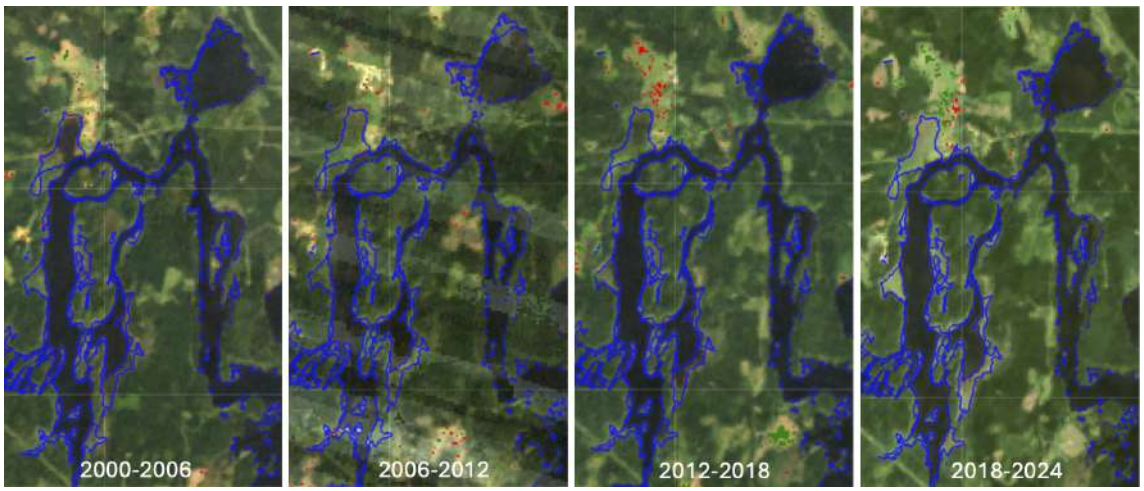

Figure 29: Spatial change detection of the class 'Grass' replaced by 'Shrub and Scrub.' From left to right: 2000 to 2006, 2006 to 2012, 2012 to 2018, and 2018 to 2024. Areas in red indicate the appearance of shrubs and scrub where grass was initially present. Areas in green indicate the appearance of grass where shrubs and scrub were initially present.

Finally, Figure 29 present the spatial change detection between classification maps every six years. Changes are presented as in Figures 28 and 27. The intervals has been chosen as the Swedish Environmental Protection Agency has advised, monitoring of scrub encroachment should be conducted at six-year intervals. If these Figures illustrate the potential of the methods presented in the context of monitoring Northern Boreal Alluvial Meadows, the changes detected in these images, which are almost nonexistent, cannot be considered accurate findings due to the considerations mentioned in the previous paragraph.

8 Discussion

The results obtained by the different methods showed significant contrasts. Although the effectiveness and relevance of these methods are evident for monitoring spatial changes in the Bredforsen landscape, the lack of data or field knowledge limits the ability to draw definitive conclusions about Spatio-temporal changes in Northern Boreal Alluvial Meadows.

All the methods used require further studies, especially for the specific case of Alluvial Meadows. For example, the NDVI clustering method requires field studies to understand the relationship between NDVI value clusters and the species present in these clusters. This would help validate or refute the relevance of this method for detecting changes in Northern Alluvial Meadows. However, the method could be used in a general study of vegetation health and dynamic in Bredforsen. Spatial changes as detected and presented in Figures 20,19, and 20, with dates and temporal aggregation adapted, could be a useful tool. Similarly, supervised classification accuracy suffer from lack of reference data. The exclusive use of a single LULC product has proven insufficient for complex and variable ecosystems such as wetlands. Other LULC products, like the LUCAS land cover dataset [14], or the combination of multiple LULC products, could be explored. From the work of [48], these alternatives could offer an interesting solution to field data. However, classification in the specific region of Bredforsen, which presents a relatively unique cold climate at the global ecosystem level, remains a challenge. Northern Boreal Alluvial Meadows, although resembling other more common habitats, remain a very localized habitat. Therefore, classification results may require field data to better understand the characteristics and surface types present in the studied landscape. In particular, the resolutions provided by satellite images do not always allow for visual assessment of these characteristics to validate the obtained classifications. However, the method present very good accuracy for water and trees class using Landsat sensors for classification, and is a great tool to study deforestation or water dynamics with a very good historical range (1985 - present).

The results from the three studies highlighted that field data would also be essential to validate whether the spectral resolutions of the satellite sensors used can effectively capture the spectral characteristics of vegetation types like Northern Boreal Alluvial Meadows and distinguish them from other vegetation types. If this is not the case, other data sources should be considered for this application. In cases where field data do not allow for a multi-temporal study, and if the study associated with the use of satellite data does not require going back beyond the start of sensors like Sentinel-1 and Sentinel-2, some studies have highlighted the effectiveness of using multi-source data (optical and SAR) for classification [43]. This approach could therefore be further explored.

However, increasing the number of input variables and data is significantly limited by the resources available on the Google Earth Engine (GEE) platform. While GEE greatly simplifies the management of satellite data, its use may need to be reconsidered if a large amount of data is required to train the model for sufficient accuracy. This hypothesis may be made in a lack of experience toward and may not be the case regarding other works like Sun, J. and al. (2023) [43]. Other cloud platforms, like Planetary Computer, are emerging and could offer interesting alternatives.

Biases in the values (indices and spectral bands used for classification) related to the temporal aggregation of data represent an important statistical dimension to study. Although this issue has not been at the forefront of the current study, the temporal heterogeneity of available observations, combined with the low frequency of Landsat data, could generate uncertainties in the values extracted from these temporal data. This variability has been quantified, for example, in the work of Carrasco, and al. (2019) [41]. Figure 29 also illustrates the difference; The background images used for spatial changes detection are all 4 yearly summer composites from Landsat 7. The composite has been created using all observations with less than 30% clouds in Bredforsen restricted region, from may to september. In the first image, (2000-2006), we can identify some water in the wetlands, in an area very dry in 2018 for example. Image 2 (2006-2012) present some artefacts.

Moreover, relying on a single observation or a single month for the composite is not desirable in the study of inter-annual variations, as vegetation growth varies depending on climatic conditions, especially in regions where snow melt can be delayed or occur earlier in the season.

A solution to detect spatial changes while limiting biases related to noise in images is proposed through the MAD transformation. The MAD transformation was attempted during this master's

thesis but is not presented in this report, as all the work done has been entirely collected in a GEE tutorial on the subject [52]. The method appears particularly relevant for very precise change detection. In the context of this study, it seemed appropriate, as the interpretation of changes in wetlands—variable and complex ecosystems—appears particularly challenging.

Although the effectiveness of the proposed methods for Northern Boreal Alluvial Meadows remains to be demonstrated, they seem to offer comprehensive tools for extracting relevant data for analyzing spatio-temporal changes. Spatial changes detection, as in Figures 19, 20, 13,12, and 27, identify clearly deforestation for exemple. Other useful applications for Vattenfall, related to biodiversity, such as deforestation or hydrological monitoring of rivers, could then also be considered using the same methods. This would possibly require changing the data used according to the necessary spatial resolution, provided that the temporal constraints required for the analysis are studied beforehand. In particular, the three methods seem complementary. The indices time series, associated with climate data, can help provide a better understanding of the context of climatic dynamics and the different elements constituting the landscape (water, vegetation, construction) before conducting a more detailed analysis of changes related to a specific type of surface. NDVI clustering can complement the classification, especially in cases where some pixels may be misclassified.

In the event that field studies reveal an inability of satellite images with 30 or 10 m resolution to capture changes in Northern Boreal Alluvial Meadows, the work of [13] has proven the effectiveness of using UAVs (drones) for studying spatio-temporal changes in alpine meadows. This approach could therefore be adapted to Northern Boreal Alluvial Meadows. However, it should be noted that this solution would be significantly more expensive.

9 Conclusion

The study explored various remote sensing and machine learning tools to detect changes in Northern Boreal Alluvial Meadows. No correlation was found between water level dynamics and changes in these meadows. Moreover, no significant changes could be identify, likely due to the lack of field data and expertise in analysis.

In particular, the study concluded that;

1. The temporal and spatial resolution of satellite data (Landsat 7, 8, 9, Sentinel-2, Sentinel-1) is adequate for studying spatio-temporal changes in the Bredforsen region. 2. The utility of NDVI clustering has been founded complicated in time-series extraction. The use of the NDVI density for spatial changes has not been proven, tho it seems more interpretable for further studies than NDVI spatial change detection. More in depth studies and field work are needed, to determine link with NDVI values and vegetation species. 3. Temporal spectral indices and LC classification are complementary. They allow for precise interpretation of potential changes. a. Spectral indices are effective in ruling out changes related to global vegetation or climate changes. b. The precision of land cover (LC) classification is crucial for detecting spatio-temporal changes in Northern Boreal Alluvial Meadows. LC classification in a multi-temporal study must account for atmospheric variations between images, as they result in different spectral characteristics for the same surface type. This remains true even when considering temporally aggregated data and products that have already been atmospherically corrected. c. The choice of LULC products requires careful consideration. For changes specific to Northern Boreal Alluvial Meadows, the Dynamic World and ESA World Cover products lack the accuracy in the relevant classes for single-use but may be useful in combination with other data sources. d. Manually annotated data need to be reviewed and refined. It is necessary to add multi-temporal data to validate the classification. The spectral responses of species of interest (species present in Northern Boreal Alluvial Meadows and invasive species) captured by satellite sensors must be studied to validate their ability to differentiate between species. 5. The multi-temporal study of spectral indices and supervised classification as proposed in this study seem adequate for analyzing spatio-temporal changes in the Bredforsen region, provided the study is not limited specifically to Northern Boreal Alluvial Meadows. 6. GEE is effective for deploying large-scale methods with flexibility and speed. However, it limits control over debugging and resource management, which can complicate in-depth studies on machine learning algorithm parameters. 7. The study also focused on the precise detection of water in the region. Previous work had already proven the effectiveness of the methods and data used. This study confirmed their utility in the Bredforsen region, as well as their performance with a maximum observation frequency without temporal aggregation. The temporal frequency and spatial resolution of the data used (S1) seem adequate. For more in-depth studies of hydrological dynamics, a preliminary evaluation of the necessary observation frequency is still suggested.

References

[1] European Commission, *Water Framework Directive*, 2024. Available at: https://environment.ec.europa.eu/topics/water/water-framework-directive_en.

[2] European Commission, Directorate-General for Environment, Eriksson, M., *Management of Natura 2000 habitats Northern Boreal alluvial meadows 6450 – Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora*, European Commission, 2008. Available at: https://data.europa.eu/doi/10.2779/69388.

[3] Matouš Jimel, *Hydrological Control of Plant Species Composition and Distribution in Dal River Alluvial Meadows*, Master thesis in Biology, 30 hp, Master's Program in Ecology, 120 hp, Umeå University, Department of Ecology and Environmental Science, Spring 2023. Available at: https://www.umu.se/en/emg.

[4] NOAA, "Climate Forecast System Version 2 (CFSv2) 6-Hourly Products," Google Earth Engine, [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/NOAA_CFSV2_FOR6H. [Accessed: July 23, 2024].

[5] Google Earth Engine. ECMWF ERA5 Monthly Dataset. Available online: https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_MONTHLY#description (accessed on August 7, 2024).

[6] U.S. Geological Survey. Landsat Missions. https://www.usgs.gov/landsat-missions. Accessed on August 10, 2024.

[7] European Space Agency. Sentinel Missions. https://sentinels.copernicus.eu/web/sentinel/home. Accessed on August 10, 2024.

[8] Google Earth Engine. Dynamic World. https://earthengine.google.com/datasets/catalog/GOOGLE_DYNAMICWORLD_V1. Accessed on August 10, 2024.

[9] NASA Earth Observatory. MODIS. https://modis.gsfc.nasa.gov/. Accessed on August 10, 2024.

[10] European Space Agency. ESA WorldCover. https://worldcover2020.esa.int/. Accessed on August 10, 2024.

[11] Google Earth Engine. Is cross-sensor Landsat surface reflectance harmonization needed? **2024**. https://developers.google.com/earth-engine/faq#is_cross-sensor_landsat_surface_reflectance_harmonization_needed

[12] C. F. Brown, S. P. Brumby, B. Guzder-Williams, and A. M. Tait et al., "Dynamic World, Near real-time global 10 m land use land cover mapping," *Scientific Data*, vol. 9, no. 1, p. 251, Jun. 2022, doi: 10.1038/s41597-022-01307-4. [Online]. Available: https://doi.org/10.1038/s41597-022-01307-4. License: CC BY 4.0.

[13] Dawen Qian, Qian Li, Bo Fan, Xiaowei Guo, Yangong Du, and Guangmin Cao. Landscape pattern changes across alpine shrub meadows gradient in warm-season pastures on the Qinghai-Tibet Plateau. *Ecological Complexity*, 51:100979, 2022. DOI: 10.1016/j.ecocom.2022.100979.

[14] European Commission, Joint Research Centre (JRC). LUCAS - Land Use and Coverage Area Frame Survey. Available online: https://esdac.jrc.ec.europa.eu/projects/lucas (accessed on August 26, 2024).

[15] U.S. Geological Survey, "CFMask Algorithm," *Landsat Missions*, https://www.usgs.gov/landsat-missions/cfmask-algorithm. Accessed: Aug. 13, 2024.

[16] Google Earth Engine, *COPERNICUS_S2_CLOUD_PROBABILITY Dataset*. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY. Accessed: Aug. 13, 2024.

[17] Google Earth Engine, "SAR Basics," [Online]. Available: https://developers.google.com/earth-engine/tutorials/community/sar-basics. [Accessed: Jul. 23, 2024].

[18] Google Earth Engine, "Sentinel-1 Guide," [Online]. Available: https://developers.google.com/earth-engine/guides/sentinel1. [Accessed: Jul. 23, 2024].

[19] Sentinel-1 Toolbox, "GitHub Repository," [Online]. Available: https://github.com/senbox-org/s1tbx. [Accessed: July 23, 2024].

[20] A. S. Yommy, R. Liu, and A. S. Wu, "SAR Image Despeckling Using Refined Lee Filter," in *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2015, vol. 2, pp. 260-265, doi: 10.1109/IHMSC.2015.236.

[21] ESA, "Refined Lee Speckle Filter," GitHub Repository, [Online]. Available: https://github.com/senbox-org/s1tbx/blob/master/s1tbx-op-sar-processing/src/main/java/org/esa/s1tbx/sar/gpf/filtering/SpeckleFilters/RefinedLee.java. [Accessed: July 23, 2024].

[22] S. Talukdar, P. Singha, S. Mahato, Shahfahad, S. Pal, Y.-A. Liou, and A. Rahman, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sensing*, vol. 12, no. 7, p. 1135, 2020. doi: https://doi.org/10.3390/rs12071135.

[23] A. Gulácsi and F. Kovács, "Sentinel-1-Imagery-Based High-Resolution Water Cover Detection on Wetlands, Aided by Google Earth Engine," *Remote Sensing*, vol. 12, no. 10, p. 1614, 2020.

[24] Edvinas Stonevicius, Giedrius Uselis, and Dalia Grendaite. *Ice Detection with Sentinel-1 SAR Backscatter Threshold in Long Sections of Temperate Climate Rivers*. Remote Sensing, 14(7):1627, 2022. MDPI. https://doi.org/10.3390/rs14071627.

[25] D. Montero, C. Aybar, M. D. Mahecha, F. Martinuzzi, M. Söchting, and S. Wieneke, "A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research," *Scientific Data*, vol. 10, no. 1, p. 197, 2023.

[26] D. Montero, "eemont: A Python package that extends Google Earth Engine," *The Journal of Open Source Software*, vol. 6, no. 62, 2021. https://doi.org/10.21105/joss.03168

[27] E. Mandanici and G. Bitelli, "Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use," *Remote Sensing*, vol. 8, no. 12, p. 1014, 2016. Available: https://doi.org/10.3390/rs8121014.

[28] Lyon, J.G.; Yuan, D.; Lunetta, R.S.; Elvidge, C.D. A Change Detection Experiment Using Vegetation Indices. *Photogrammetric Engineering Remote Sensing* **1998**, 64, 143-150.

[29] Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *College of Mechanical and Electronic Engineering, Northwest AF University* **2017**. Available online: https://www.researchgate.net/publication/317146924_Significant_Remote_Sensing_Vegetation_Indices_A_Review_of_Developments_and_Applications (accessed on Aug 07 2024).

[30] Xu, H. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *International Journal of Remote Sensing* **2006**, 27, 3025-3033.

[31] Shou-Ye Xue, Hai-Yan Xu, Cui-Cui Mu, Tong-Hua Wu, Wang-Ping Li, Wen-Xin Zhang, Irina Streletskaya, Valery Grebenets, Sergey Sokratov, Alexander Kizyakov, and Xiao-Dong Wu. Changes in Different Land Cover Areas and NDVI Values in Northern Latitudes from 1982 to 2015. *Advances in Climate Change Research*, 12(4):456–465, August 2021. https://doi.org/10.1016/j.accre.2021.04.003.

[32] Noh-Hun Seong, Daeseong Jung, Jinsoo Kim, and Kyung-Soo Han. Evaluation of NDVI Estimation Considering Atmospheric and BRDF Correction through Himawari-8/AHI. *Asia-Pacific Journal of Atmospheric Sciences*, 2020. https://doi.org/10.1007/s13143-019-00167-0. Accessed August 12, 2024.

[33] Tucker, C.J. (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, 8, 127-150. http://dx.doi.org/10.1016/0034-4257(79)90013-0.

[34] Taloor, A.K., Manhas, D.S., & Kothyari, G.C. (2020). Retrieval of land surface temperature, normalized difference moisture index, normalized difference water index of the Ravi basin using Landsat data. *Advances in Climate Change Research*, 11(4), 100051. https://doi.org/10.1016/j.acags.2020.100051.

[35] Rama Ewing, *Spatiotemporal Change Detection of the Alpine Meadows at Holcomb Valley, San Bernardino Mountain National Forest, Using GIS and Remote Sensing Techniques*, Master's thesis, California State University, San Bernardino, December 2022.

[36] Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S. S., Egorov, A. V. Characterization of Landsat-7 to Landsat-8 Reflective Wavelength and Normalized Difference Vegetation Index Continuity. *GSCE Faculty Publications* **2016**, 34. `http://openprairie.sdstate.edu/gsce_pubs/34`

[37] Zhifang Pei, Shibo Fang, Wunian Yang, Lei Wang, Mingyan Wu, Qifei Zhang, Wei Han, and Dao Nguyen Khoi, "The Relationship between NDVI and Climate Factors at Different Monthly Time Scales: A Case Study of Grasslands in Inner Mongolia, China (1982–2015)," *Sustainability*, vol. 11, no. 24, article 7243, 2019. doi: 10.3390/su11247243.

[38] Google Earth Engine, "Earth Engine API: ee.Clusterer.wekaKMeans," 2023. [Online]. Available: `https://developers.google.com/earth-engine/apidocs/ee-clusterer-wekakmeans`. [Accessed: Aug. 22, 2024].

[39] Tedros M. Berhane, Charles R. Lane, Qiusheng Wu, Oleg A. Anenkhonov, Victor V. Chepinoga, Bradley C. Autrey, and Hongxing Liu, "Comparing Pixel- and Object-Based Approaches in Effectively Classifying Wetland-Dominated Landscapes," *Remote Sensing*, vol. 10, no. 1, article 46, 2018. doi: 10.3390/rs10010046.

[40] Meta AI Research, "Segment Anything Model (SAM)," 2023. [Online]. Available: `https://segment-anything.com`. [Accessed: Aug. 20, 2024].

[41] Luis Carrasco, Aneurin W. O'Neil, R. Daniel Morton, and Clare S. Rowland, "Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine," *Remote Sensing*, vol. 11, no. 3, article 288, 2019. doi: 10.3390/rs11030288.

[42] Salvatore Praticò, Francesco Solano, Salvatore Di Fazio, and Giuseppe Modica, "Machine Learning Classification of Mediterranean Forest Habitats in Google Earth Engine Based on Seasonal Sentinel-2 Time-Series and Input Image Composition Optimisation," *Remote Sensing*, vol. 13, no. 4, article 586, 2021. doi: 10.3390/rs13040586.

[43] Sun, J., Ongsomwang, S. (2023). Optimal parameters of random forest for land cover classification with suitable data type and dataset on Google Earth Engine. *Journal of Remote Sensing and GIS*, 12(1), 24-36.

[44] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93-104, 2012. [Online]. Available: `https://doi.org/10.1016/j.isprsjprs.2011.11.002`. [Accessed: August 4, 2024].

[45] Zafar, Z., et al. (2023). Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 85, 101361.

[46] Phan, T. N., Kuch, V., Lehnert, L. W. (2020). Land cover classification using Google Earth Engine and random forest classifier—The role of image composition. *Remote Sensing*, 12(15), 2411.

[47] Y. Piao, S. Jeong, S. Park, and D. Lee, "Analysis of Land Use and Land Cover Change Using Time-Series Data and Random Forest in North Korea," *Remote Sensing*, vol. 13, no. 17, p. 3501, Sep. 2021. [Online]. Available: `https://doi.org/10.3390/rs13173501`. [Accessed: August 4, 2024].

[48] Z. S. Venter, D. N. Barton, T. Chakraborty, T. Simensen, and G. Singh, "Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover," *Remote Sens.*, vol. 14, no. 16, p. 4101, Aug. 2022. doi: 10.3390/rs14164101.

[49] Masroor, M.; Avtar, R.; Sajjad, H.; Choudhari, P.; Kulimushi, L.C.; Khedher, K.M.; Komolafe, A.A.; Yunus, A.P.; Sahu, N. Assessing the Influence of Land Use/Land Cover Alteration on Climate Variability: An Analysis in the Aurangabad District of Maharashtra State, India. *Sustainability* **2022**, *14*, 642. https://doi.org/10.3390/su14020642

[50] Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

[51] Hu, Y.; Hu, Y. Land cover changes and their driving mechanisms in Central Asia from 2001 to 2017 supported by google earth engine. *Remote Sens.* **2019**, 11, 554. https://doi.org/10.3390/rs11050554

[52] Google Earth Engine, 'Introduction to IMAD: Image Differencing and Multivariate Alteration Detection - Part 1' [Online]. Available: https://developers.google.com/earth-engine/tutorials/community/imad-tutorial-pt1. [Accessed: August 21, 2024].