MAPPING AND MONITORING HEALTH CONDITIONS OF WETLANDS USING REMOTE SENSING TECHNIQUES

A thesis submitted to the University of Dublin, Trinity College in candidature for the Degree of

Doctor of Philosophy in Civil, Structural and Environmental Engineering

by

Saheba Bhatnagar

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Under the supervision of Professor Bidisha Ghosh

and Professor Laurence Gill

October 2021

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signature:

Date: 20.10.2021

Acknowledgements

Great things never come from the comfort zone, and that is what these three years of PhD has mainly taught me – to take risks and to explore.

At almost the end of this 3-year long journey, I am thankful to a lot of people. First and foremost, I would like to thank my family - my papa (Mr Hari Bhatnagar) and momma (Mrs Priti Bhatnagar), my sister Sofia, Prabal jiju and my two cute nephews (Kartikeya and Agastya) for their love, support and encouragement throughout my life. You all are my backbone, and you are the reason I could take this bold step of doing a PhD.

I would sincerely like to thank my supervisors Dr Bidisha Ghosh and Prof Laurence Gill, for being wonderful mentors, for inspiring me to go an extra mile, for being helpful, kind and supportive. Without their professional help and creativity, this PhD was not possible. Under their guidance I successfully overcame many difficulties and learnt a lot. Their zeal for perfection, and conviction has inspired me to always aim high. I also would like to thank Dr Mahesh K Sha for introducing me to the new S-5p data. His technical help was very beneficial for understanding methane mechanism and exploring the new field. I want to express my gratitude towards the members of 'Ecometrics' project and Environmental Eng for imparting knowledge and helping me understand this new field of wetlands and ecology.

During my PhD, I had a chance to meet and work with some wonderful people at Trinity College Dublin and in Ireland. I would like to thank all of my PhD and postdoc friends and peers for all the fun during lunch/tea break sessions and Friday pints; thanks to my landlords for being so warm and welcoming. I would also like to thank all my friends from school, OIST and IIST for always calling, and checking on me, and basically keeping me sane. Talking to you guys, sharing funny memes has been such a stress-burster for me. I would specifically mention my friends Ayushi, Pulkit, Shipra, Ella and Niti for always being my constant support.

Last but definitely not least, I would like to thank Ankur Khurana, for all his love and care while I was working on the thesis. Thanks for being so understanding, for being such a great listener, my problem-solver and my best friend. I would not have finished this thesis without your unconditional support and technical help.

<div align="center">आप सभी को बहुत बहुत धन्यवाद |</div>

Abstract

Wetlands provide essential ecosystem services for the world, but in recent years, due to direct human activities as well as global warming and other climatic variations, there has been significant ongoing reductions in both wetlands' quantity and quality. Vegetation monitoring is key to assessing the overall health and dynamics of a wetland. Multiple vegetation surveys such as the national bog survey, national fen survey, etc. have been carried out in previous years to map and understand the extent of vegetation on these wetlands. These manual surveys, however, are often time-consuming and require a large number of people. Also, once surveyed, it is unlikely that such an effort will be repeated regularly to update the map. The remoteness and inaccessibility of many wetlands are also limiting factors. Hence, there is a growing recognition of RS techniques as a cost-effective and viable alternative to field-based ecosystem monitoring. This study aims to identify and monitor the environmental conditions of the wetlands using machine learning (ML) techniques.

The study begins with a pixel-based approach to map vegetation communities across raised bogs and fens using ensemble classifiers such as bagged tree (BT). Due to turf-cutting and other practices, the boundary of such wetlands is often ill-defined, and therefore the study initially develops a boundary delineation algorithm. Using edge detection techniques such as entropy filtering, canny edge detection, and Lazy snapping, the wetlands' boundaries were successfully delineated. The pixel-based approach applied initially only takes into account the spectral properties of the area. Therefore, the study was further extended to segment-based learning using graph cut maximum *a-posteriori* (MAP) segmentation. This takes into account the contextual information on top of the spectral information. This segmentation process acted as a post-classification smoothing for the wetland maps. The algorithm was tailored for land-cover, especially vegetation identification and hence, was termed Mapping Vegetation Communities (MVC) algorithm. A total of up to 18 classes were mapped, using multi-date satellite data, inside 13 wetlands using the MVC algorithm with an average accuracy of 84% for the years 2017 and 2018. The algorithm works very well for larger vegetation communities, but some small communities were not adequately identified due to the restriction of spatial resolution (10 m) of Sentinel-2 data. Therefore, drones or unmanned aerial vehicles (UAVs), otherwise known as drones, were employed to gain high spatial resolution.

Drones provide very high spatial resolution and flexibility in temporal resolution. In order to confirm the applicability of these methods for mapping vegetation inside wetlands, a comprehensive comparison was made between multiple DL and ML algorithms.The study reveals that DL provides higher accuracy compared to ML by approximately 2%, but also is time and cost-intensive. Hence, the choice of the algorithm should be application dependant.

The study then extends the ability of remote sensing-based monitoring of wetlands by

combining the high spatial resolution of drones with the global coverage of the satellite data to create seasonal maps of vegetation communities within the wetlands. This nested methodology incorporates geo-referenced land-cover maps, scaled at the drone resolution level and up-sampled to S2 imagery level through interpolation. A colour correction technique was introduced in the pipeline to improve consistency between drone image capture sessions. The proposed framework has been evaluated on various wetlands across Ireland, and results are presented herein for an ombrotrophic peatland, Clara Bog. Additionally, to automate the process, a majority voting technique was applied to seasonal images. The obtained monthly maps were united to produce a more precise annual vegetation map of the wetland for the year 2019. The application of this method thereby reduces the number of field surveys typically required to assess long-term ecological change on wetland habitats.

Wetlands are known to be the largest natural source of methane (CH_4), and the atmospheric lifetime of CH_4 is about $9 \pm$ two years which makes it a good target for climate change mitigation. In this study, the use of the methane data along with the retrieved surface albedo (SA) from the recently launched European Space Agency's Sentinel-5 Precursor (S-5p) mission has been explored over Canada. The sensitivity of the methane emissions from wetlands over Canada was seen in the total column methane measured by S-5p, which was then used for the land type classification. The data from 2018 and 2019 were used separately to create individual maps for the two years and compared to the reference ground truth. It was seen that spring, and autumn-time CH4 measurements are high, and the lowest values are during the summer. As the area covered by bog and fen are small and mixed with different land types (e.g., marsh, swamp, forest), the low variation in sensitivity of the CH4 data makes it difficult to determine these land types. However, six of the ten land types were identified with great confidence. Amongst them are two major wetland types in Canada (marsh and swamp), covering a significant area of the country. Also, identification of forest land type is significant in monitoring the area covered by forest and its change over time. This is an entirely new use of the S-5p CH_4 product, and the study showed the high potential of the data with applications in land type identification. The S-5p CH_4 data can be applied to multiple areas which are yet to be explored.

Finally, this study has also investigated a 28-year hydrological record and water levels on four turloughs in the west of Ireland (1989 to 2017) with respect to their ecohydrological metrics using statistical analysis. For each vegetation community, the metric was defined using the flood duration and flood depth, as well as global radiation and temperature as a proxy for the time of the year when the floodwaters first recede and the vegetation can emerge. These trends were compared with the latest RS map produced using the MVC algorithm, and it was seen that the key communities stayed intact despite some extreme flood events over the past 20 years. The metrics were further refined using hierarchical clustering for the range of parameters. Such an ecohydrological metric is beneficial for forming policies and defining pressures associated with drainage and other land use activities that may take place.

Contents

List of Figures

List of Tables

List of Tables

List of Abbreviations

This list describes several abbreviations that will be later used within the body of the document.

AdaBoost	Adaptive Boosting
AF	Active Flush
ALOS	Advanced Land Observing Satellite
ANN	Artificial Neural Network
AOF	Area of flight
ARB	Active Raised Bog
ARVI	Atmospherically corrected vegetation index
ATBD	Algorithm theoritical baseline document
BOA	Bottom of atmosphere
BoT	Boosted Tree
BT	Bagged Tree
C	Central
CA	Class accuracy
CAI	Cellulose absorption index
CART	Classification and Regression Trees
CF	Calcareous Fen
CH4	Methane
CHM	Canopy height model
CNN	Convolutional Neural Network
CVI	Chlorophyll Vegetation Index
CWCS	Canadian wetland classification system
CWI	Canada wetland inventory
DEM	Digital Elevation Model
DenseNet	Dense Network
DL	Deep Learning

DSM	Digital surface model
DT	Decision Trees
DTM	Digital terrain model
ENVISAT	Environmental Satellite
EPA	Environmental Protection Agency
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ETM	Enhanced Thematic Mapper
EU	Europe
EVI	Enhanced vegetation index
EW	Extra Wide
FLAASH	Fast line of sight atmospheric analysis of spectral hypercubes
FN	False Negatives
FOV	Field of view
FP	False Positives
GB	Gradient Boosting
Gcut	Graph cut
GEE	Google Earth Engine
GHGs	Green house gases
GIS	Geographical Information System
GLCM	Gray level co-occurrence matrix
GNDVI	Green normalised difference vegetation index
GPP	Gross Primary Productivity
GPS	Global Positioning System
GRD	Ground Range Detected
GWB	GroundWater Body
GWDTE	Ground water dependent terrestrial ecosystem
HH	Horizontal Horizontal Polarisation
HV	Horizontal Vertical Polarisation
IDB	Index database
IR	Infrared
IRWG	Infrared working group
ISODATA	Iterative self organising data analysis technique
IW	Interferometric Wide
JERS	Japanese Earth Remote Sensing Satellite

KGE	Kling Gupta efficiency
kNN	k-nearest neighbour
L1	Level-1
L2	Level-2
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
LUE	Light Use Efficiency
LULC	Land use land cover
LULUCF	Land use land use change and forestry
LWIR	Long wave infrared
M	Marginal
MA	Model accuracy
mAOD	mean above ordnance datum
MAP	Maximum a-posteriori
Matlab	MathWorks Laboratory
ML	Machine Learning
MNDWI	Modified normalised difference water index
MODIS	Moderate Resolution Imaging Spectro-radiometer
MRS	Multi-resolution segmentation
MSI	Multispectral Instrument
MSS	Multi-spectral scanner
MVC	Mapping Vegetation Communities
MWIR	Middle wave infrared
NASA	National Aeronautics and Space Admininstration
NDACC	Network for detection of atmospheric composition change
NDVI	Normalised difference vegetation index
NDWI	Normalised difference water index
NIR	Near Infrared
NSE	Nash Sutcliffe efficiency
NVC	National Vegetation Classification
OA	Overall accuracy
OBIA	Object based image analysis
OLI	Operational Land Imager
OS	Ordnance survey
PALSAR	Phase Array type L-band SAR

PBIA	Pixel based image analysis
PDF	Probability density function
PSRI	Plant senescence reflective index
Radar	Radio Detecting and Ranging
Radarsat	Radar satellite
RBF	Radial basis function
REP	Red Edge Position
ResNet	Residual Network
RF	Random Forest
RGB	Red, Green, Blue
RS	Remote Sensing
S-5p	Sentinel-5 Precursor
S1	Sentinel-1
S2	Sentinel-2
SA	Surface Albedo
SAR	Synthetic Aperture Radar
SAVI	Soil adjusted vegetation index
SC	Sub-central
SegNet	Segmentation Network
sfm	structure from motion
SKNN	Subspace KNN
SLAR	Side Looking Airborne Radar
SLC	Scan line corrector
SLC	Single look complex
SM	Sub-marginal
SMSC	Submarginal and subcentral
SNAP	Sentinel Application
SPOT	Systeme Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SVM	Support vector machine
SWIR	Short wave infrared
TA	Trasnfer accuracy
TCCON	Total carbon column observing network
TIR	Thermal Infrared
TIRS	Thermal Infrared Sensor

TM	Thematic Mapper
TN	True Negatives
TOPS	Terrain Observation with Progressive Scan
TP	True Positives
TROPOMI	Tropospheric monitoring instrument
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
UNet	U Network
USGS	United States Geological Survey
UV	Ultraviolet
UVIS	Ultraviolet and Visible
UVN	Ultraviolet, Visible and NIR
VGG	Visual Geometry Group
VH	Vertical Horizontal Polarisation
VI	Vegetation index
VNIR	Visible and NIR
VOG	Vogelmann Index
VRE	Vegetation Red Edge
VV	Vertical Vertical Polarisation
WAD2M	Wetland area dynamics for methane monitoring
WFD	Water Framework Directives
WI	Water Index
WV	WorldView
XCH4	Total column methane
XGB	eXtreme Gradience Boosting

List of Abbreviations

List of Abbreviations

Chapter 1

Introduction

"A question that sometimes drives me hazy: am I or are the others crazy?"

— Albert Einstein

Contents

1.1 Background

A wetland is an area covered or surrounded by water for the majority of a year. The wetlands are essential for various reasons, providing many ecosystem services (De Groot et al. 2012). They regulate the groundwater flow, which helps in flood management; they act as a buffer to reduce other disasters such as hurricanes, tsunami, as the coral reefs in wetlands reduce the speed of the wave; they provide a vital link between land and water and they help control soil erosion and purify the water. Wetlands can be artificial or natural and are home to a wide diversity of vegetation, aquatic and terrestrial species.

Apart from hosting vegetation and animal species, wetlands are also needed to grow staple products like rice. Despite these known advantages, wetlands are still facing threats and undergoing constant degradation. In order to reverse this direction of global wetland loss, the Ramsar Convention, held in Iran, was signed in 1971 (Xu et al. 2019). It is a global treaty to protect wetlands and also the first environmental convention in the world (Erwin 2009). Approximately 171 countries are a part of the Ramsar Convention covering ≈ 299 million ha area (Xu et al. 2019).The Ramsar convention (2011) provides a definition of wetlands as: "areas of marsh, fen, peatland or water, whether natural or artificial, permanent or temporary, with water that is static or flowing, fresh, brackish or salt, including riparian and coastal zones adjacent to the wetlands and islands or bodies of marine water deeper than six meters at low tide lying within the wetlands."

In Europe, wetlands cover approximately 7.3 million ha of the area, which is around 2% of the entire land area. Figure 1.1 shows the wetlands as a share of the total land in Europe.

Wetlands as a share of total land area in the EU Member States, 2015
(EU = 100)
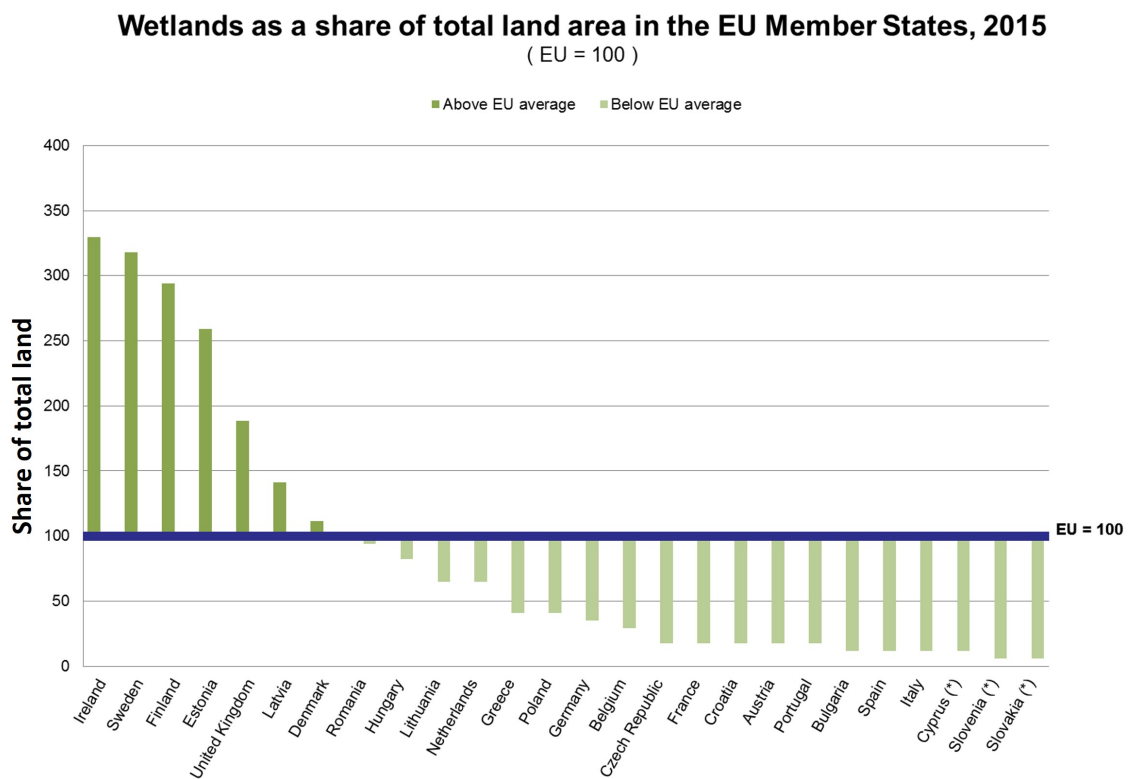
Figure 1.1: Wetlands as a share of total area, Europe, source *List of wetlands as a share of total land area* 2020

It can be seen that Ireland, Sweden, Finland, Estonia, and the United Kingdom contain the most significant contributions of wetlands (given their area) in Europe. In Ireland, wetlands cover a big area, but due to decades of turf cutting, drainage, fertilisation, etc., there has been a significant decrease in their spatial extent. For example, bogs were once spread across an area of 300,000 ha out of which only 18,000 ha remain today. These wetlands are a great source of peat which is harvested as fuel for electricity and heating purposes. Hence, there is a pressing need for restoration and monitoring of the wetlands.

1.2   Wetlands In Ireland

The scope of the thesis is on natural wetlands, in particular groundwater- dependent ecosystems (GWDTE) and priority habitats as listed in EU Habitats Directive, i.e., fens and turloughs as GEDTEs and precipitation-fed raised bogs. In 2004 under the EU Habitats Directive (92/43/EEC), in Ireland, 132 groundwater bodies were assessed for risk and marked as "priority" habitats. Figure 1.2 presents a schematic of the breakdown of the wetlands.
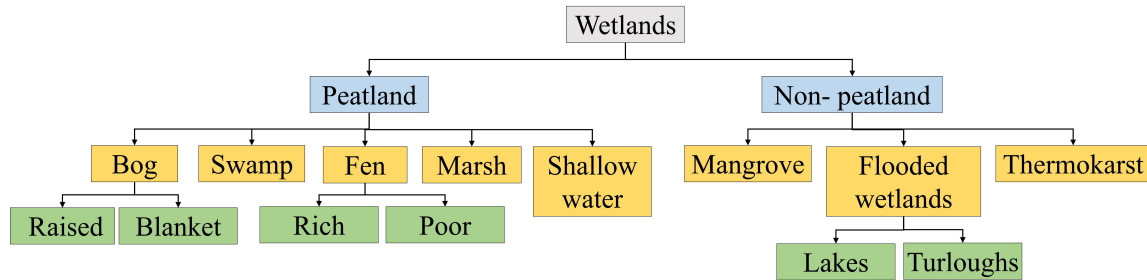
Figure 1.2: Types of wetlands in literature being reviewed

A peatland is an area of little or more vegetation and a naturally accumulated peat layer at the surface (Joosten & Clarke 2002). Peatlands include raised and blanket bogs, swamps, fens, marshes and shallow water bodies (see Figure 1.2). Peatlands contain unique flora containing mostly sphagnum species, which varies from location to location. Figure 1.3 represents some of the species found in peatlands. These species and more are found in every type of peatland.

Figure 1.3: Example of vegetation species found in peatland, source *Plant List* 2013

Peatlands are spread all across Ireland, covering $\approx 16.2\%$ of the whole country. They mainly consist of bogs and fens, as shown in Figure 1.4.
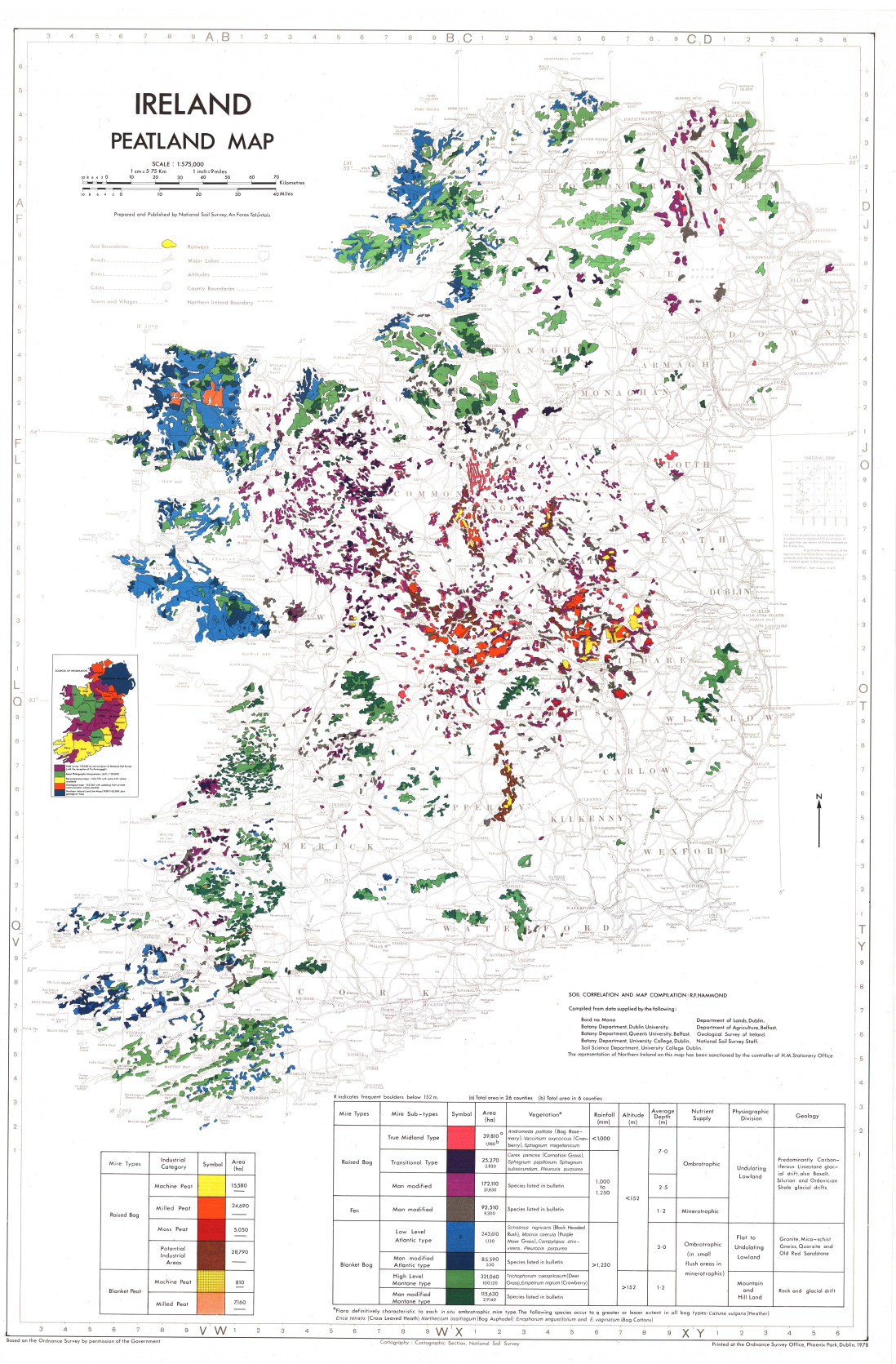
Figure 1.4: Peatlands in Ireland, source (O'Connell 2012)

Peatlands can be distinguished from other landscapes based on morphology and landscape position. Peatlands in the form of raised bogs have a raised-dome like structure and are acidic in nature. Another type of peatland called fens are situated in depressions and are usually alkaline with higher pH (Minasny et al. 2019). Peatlands can be used for multiple applications (Kimmel & Mander 2010) with some listed below.

- As a biomass producer for agricultural use, dairy, forestry, etc.

- As an energy source for extraction of fuel and growing horticultural media.

- As a carbon storage unit (peatlands are known carbon sequesters when in good condition and store up to a third of the world's total soil carbon (Maltby Barker, 2009)).

- As a water reservoir, providing some natural flood retention.

- As a climate regulator. As one of the most significant terrestrial C components, peat influences the direction and magnitude of carbon cycle-climate feedbacks (Minasny et al. 2019).

- For supporting biodiversity as it is a unique habitat for rare and endemic species.

A brief description of the wetlands under consideration in this research study, along with their distribution in Ireland, is given below.

- Raised Bogs
  Raised bog wetlands are discrete, dome-shaped masses of peat occupying former lakes or shallow depressions in the landscape. The active area of Sphagnum growth in a raised bog (i.e. ARB) should be accumulating peat. Hence, the ecological health of a bog can be assessed by its ability to form and accumulate peat. In healthy raised bogs, the surface layer, known as the acrotelm is made up by an intact continuous layer of Sphagnum moss. This layer has a large effect on the hydrology of the raised bogs, ensuring permanent saturation to within a few decimetres of the soil surface, which is necessary for the survival of the moss. Hence, the ecological health of the bog (i.e. whether it is peat-forming or in a degraded state) can be determined by the nature of its surface vegetation community composition (which can be divided into different characteristic "ecotopes") in relation to its water table and topography.
  The bogs occur throughout the Irish midlands and in some parts of County Galway, Kerry, Clare and Mayo. Figure 1.5 shows the extend and coverage of the raised bogs in Ireland (Mackin et al. 2017).

Figure 1.5: Bogs in Ireland, source *Wetland Survey Ireland* 2020

- Fens
  Fens are peaty habitats, often fed by precipitation, groundwater and surface water (Goodall & Gore 1983]). Unlike a bog, a fen is rich in minerals like calcium and iron (Beltman et al. 1996), and hence, vegetation communities in fens are clearly different from bogs. A fen could be rich or poor on the basis of its chemical composition: alkaline conditions means that a fen is considered rich, whereas acidic conditions mean a fen is considered to be poor. Alkaline fens and Calcareous fens (with *Carex davalliana*) often occur together and are considered as calcareous fens (Kimberley & Coxon 2013). Alkaline fens are dominated by sedges and rushes, whereas calcareous fens with *Carex davalliana* are wet, and more oligotrophic (Kimberley & Coxon 2013). Calcareous fens are among the most species-rich wetland (Van Diggelen et al. 1996). Some communities in the fen, such as meadows and wet grasslands, can appear after manipulation of the water table, mostly due to human interference, which can be used to define the drained area of the fen (Van Diggelen et al. 1996). Fens, in comparison to raised bogs, are more evenly spread across the country. They are mostly found in mineral-enriched places like the limestone regions of Ireland. The majority of fens are found in County Mayo followed by Counties Galway and Monaghan. Figure 1.6 depicts the spread of the fens across the counties in Ireland.

Figure 1.6: Fens in Ireland, source *Wetland Survey Ireland* 2020

- Turloughs
  Turloughs are almost unique to Ireland and can be defined as depressions in karst areas, seasonally flooded mostly by groundwater and supporting vegetation and soils characteristic of wetlands (Skeffington et al. 2006). The flooding period usually starts in October and lasts most of the winter and spring. The flooding duration is a function of the underground karst conduit network, the rainfall and the local basin topography and so is unique to every site (Waldren et al. 2015). With the advent of summer, the lake starts to dry, and vegetation communities start to emerge. The functioning of the turlough is mainly dependent on the characteristic periodic flooding, which plays a big role in the biological diversity of the turloughs (Naughton et al. 2012). A turlough when dry, can have a grassy appearance due to presence of sedges, with a gradation of communities down into the lower part of the turlough basin (where water is sustained the longest) to more wetland associated species such as small sedges, silverweed, and meadowsweet. Hence, a range of different vegetation communities can be associated with a topography which seems to be linked to the typical flood duration statistics across the year. Turloughs are mostly located in the west of Ireland with the majority occurring in County Galway. Figure 1.7 shows the extend and locations of turloughs in Ireland.

Figure 1.7: Turloughs in Ireland, source *Wetland Survey Ireland* 2020

1.3 Current practice in Wetland Monitoring

The Environmental Protection Agency (EPA) is responsible for monitoring the status of Groundwater Dependent Terrestrial Ecosystems (GWDTEs) under the Water Framework Directive (WFD). However, new and refined approaches, such as incorporating ecosystem services, are needed to assess the status of GWDTEs to reduce the pressures associated with the quality and quantity of groundwater from the supporting groundwater body (GWB). Understanding the hydrology of wetland systems is essential for successful management and the research conducted to date demonstrates that the ecology of groundwater-dependent wetlands is fundamentally reliant on the supporting hydrogeology.For example, the hydrogeological damage that has occurred at Clara Bog (County, Offaly), Pollardstown Fen (County Kildare) and Rahasane turlough (County Galway), is reflected in corresponding changes in ecology, demonstrating that understanding the eco-hydrogeological connectivity and environmental supporting conditions is vital for the conservation of such wetlands (Kilroy et al. 2008). EPA projects such as the 'Ecometrics' project that this study is part of, fills the linkage gap between eco-hydrology and environmental supporting conditions of the GWDTEs by introducing an appropriate metric by which changes can be monitored. This PhD thesis corresponds to the work package 'C' in the Ecometrics project.

A conceptual understanding of the eco-hydrogeology of GWDTEs and their associated groundwater systems is central to the implementation of the WFD and the Groundwater Directive (EC, 2009). A water supply mechanisms approach, based upon a series of concep-

tual hydrological models with an ecological overlay, has been used in the UK (Whiteman et al. 2009) and Ireland (Kimberley & Coxon 2013), to represent the dominant pathways of water movement from GWBs into GWDTEs and used to describe potential ecological responses to changes in groundwater quality and quantity. However, while it is understood that groundwater is vital in maintaining the environmental supporting conditions required to sustain wetland habitat and species, there is currently no measured (or easily obtained) metric available that can be used to assess the ongoing status of GWDTEs and their associated GWBs.

For wetland vegetation classification, there are strict phytosociological principles and rules that are to be followed. A study by White & Doyle 1982 gives detailed information about the wide variety of Irish vegetation communities, including a list of all the vegetation species recorded in Ireland. Another study by Fossitt 2000 presents a comprehensive hierarchical classification of Irish habitats in its entirety. Apart from exclusive Irish phytosociology, researches like The National Vegetation Classification (NVC) (Rodwell & nature conservation committee GB) conducted for vegetation in Great Britain has also been referred time to time (Waldren et al. 2015).

1.3.1 Manual mapping of wetlands

The most common way to assess a wetland is to carry out a field-based mapping exercise, which has occurred on wetlands all across Ireland over the years. For mapping vegetation, a quadrant of $1m^2$ is taken, and the number of leaf-types (fronds) present inside it are recorded (Figure 1.8). In accordance with the size of the vegetation community, the quadrat size is changeable. For a smaller communities estimates are carried out based on counts using $25cm^2$.

Figure 1.8: Wetland manual survey using $1m^2$ quadrat, source: Mackin et al. 2017

The number of quadrats used is also a function of the community size. All across the wetland, these quadrats are placed at various locations to ensure an accurate identification and recording of the community-density. Using "Domin scores", the plant species present

and their cover-abundance is recorded for each quadrat. The Domin scores are given in Table 1.1.

Domin score	Range of abundance
+	single individual (small)
1	1-2 individuals (larger than +)
2	<1%
3	1-4%
4	4-10%
5	11-25%
6	26-33%
7	34-50%
8	51-75%
9	76-90%
10	01-100%

Table 1.1: Domin scores and range of abundance, source Waldren et al. 2015

A survey led by Fernandez et al. 2014 named " Raised Bog Monitoring and Assessment " done in 2013 has mapped 32 raised bogs in Ireland, whilst another nation-wide survey "National Fen Survey" by Foss & Crushell 2008 carried out in 2008 has mapped more than 100 fens. Waldren et al. 2015 have mapped 22 turloughs across Ireland in 2007. These surveys took years to complete and were both resource and time-consuming. For monitoring the wetlands, regular surveys really need to be carried out, which is extremely challenging given the time and resource involvement. Due to this, remote sensing (RS), with its increasing use in broader hydrological applications, is becoming an increasingly investigated and accepted tool for the identification and classification of wetlands (Mahdavi et al. 2018).

1.3.2 Remote Sensing

As discussed previously, a crucial part of wetland conservation should be being able to monitor the ongoing health of the wetland. Remote sensing provides invaluable information to characterise and measure wetland states, condition, and functioning. Given the current need for up-to-date information, as well as the widespread coverage of wetland maps, satellite RS has been demonstrated to be the most efficient and cost-effective method for this purpose (Ozesmi & Bauer 2002). Remote sensing is a scientific tool to observe and study the Earth and its components (the land surface, the oceans, the atmosphere) and their dynamics from space.
Every satellite image comprises of the intensity value of an area captured at different wavelengths. These wavelengths are known as spectral bands. Therefore, a satellite image is an intensity matrix giving multi-dimensional spectral information. In order to analyse the satellite data, many image processing algorithms and machine learning (ML) techniques have been proven useful.
RS approaches offer significant advantages for wetland ecosystem management; they have the potential to describe wetland extent, structure and condition quickly, accurately and at a large spatial scale. The increased availability and quality of RS data in recent decades has made it a valuable resource for wetland management with enormous potential for habitat mapping, tracking restoration/degradation and monitoring of spatial and tempo-

ral changes. Therefore, a significant part of the thesis (and also the Ecometrics project) is to investigate the use of RS data for the identification and characterisation of crucial GWDTE types; turloughs, fens and also non-GWDTE - raised bogs. It should be noted that in order to have a common phrasing in the thesis, the term wetland has been used in place of GWDTE/non-GWDTE.

1.4   Research Objectives

The environmental conditions of wetland can be reflected to its ecology, which is given by the veg- etation composition present in that area. To date, in Ireland, the monitoring of the environmental conditions in most of the wetlands is done manually by field surveys. That can be very time consuming and is often disrupted. Hence, there is a need to automatise this monitoring process. This research aims to fulfil this gap by establishing a linkage between remotely sensed images and the ecology of the wetlands using machine learning (ML) and deep learning (DL) technqiues. The overall aim of this research is to come up with an approach that can help to understand the ecological conditions of the wetlands and to automatise the entire vegetation mapping process and reduce the amount of fieldwork being done. The individual objectives by which this aim will be realized are as follows:

1. Wetlands constitute a large area in the Irish landscape. The shape and size of these wetlands is a key indicator of their ecological condition. It is also an important characteristic when it comes to land management and urban planning. However, due to these wetlands' amount and sparse location, it is challenging to record their shape manually. Freely available satellite data along with machine learning algorithms can be used to keep track of the boundary of the wetlands. Hence, the first research objective is:
   To identify the boundary of a wetland in Ireland utilising open-source satellite data employing ML based algorithms.

2. After the delineation, to monitor the health of the wetland, identification of vege- tation inside the wetland is necessary. The type of vegetation in a wetland gives an idea of the nutritional quality of the area and can help in long term wetland maintenance. Therefore, the second objective is:
   To create a robust segmentation based technique for mapping vegetation communi- ties within the boundary delineated wetlands, using open-source satellite data, and cross-checking the accuracy against field-based surveys.

3. The freely available satellite data is widely available but lacks a high spatial res- olution, making detailed mapping (species level) of a wetland challenging. As an alternative, drone imagery can be used to obtain high-resolution images and identify vegetation species in an area. Hence, the third objective is:
   To analyse the applicability of high-resolution drone imagery for mapping vegetation communities within wetlands using both ML and DL techniques.

4. The satellite data provides global coverage and multispectral information, whereas the drone imagery provides high spatial resolution and flexibility to capture images. To benefit from both satellite and drone, it is essential to use them together. There- fore, the fourth objective is:
   To develop a combined ML based monitoring regime using satellite and drone im-

agery for automating the field-based surveying process. Moreover, analysing the minute changes in the vegetation communities across seasons using a specific change detection error metric.

5. Methane emissions from wetlands make up 20-30% of the global methane emissions. With the current technology and presence of satellite-based methane data, these emissions could be monitored. Hence, the fifth objective is:
   To analyse the effect of gaseous emissions (in particular methane) on detection of wetlands (and other land types) using open-source satellite data and ML algorithms.

6. Apart from ecology, the hydrology of the area also affects the health of the wetlands. The ground-water dependent wetlands, such as turloughs, are dependant on hydrology for their ecological behaviour. Therefore, the last objective of this study is:
   To analyse the hydrological trends in flooded wetlands such as turloughs using statistical data analysis, and to compare these trends with the latest satellite-based maps.

1.5 Literature review

This section gives a brief overview of the existing and past research done in the field of wetland remote sensing. The search engine used for gathering all the literature was "Engineering Village" by Elsevier. An initial search was done with the keywords "wetland remote sensing" for the time 2000-2020. This resulted in more than 2000 journal and conference papers, books, and user manuals. To constrict the search, keywords "peatland" and "satellite" were also added. The search was also made limited to published peer-reviewed papers, and duplicate articles using compendex and inspec (articles having the same abstract) were removed. This resulted in around 200 papers out of which 60 research papers were studied and analysed to form this literature review.
Other than the generalised literature review, there was a need to review additional literature specific to the objectives of this thesis. This additional literature is presented in the introduction of Chapters 2, 3, 4, 5, and 6.
This entire section has been divided into six subsections. Section 1.5.1 explains various types of wetlands where RS techniques have been applied. Section 1.5.2 gives a description of their locations. The RS data used for this purpose have been discussed in Section 1.5.3. Section 1.5.4 gives the critical objectives of the paper in terms of application, such as classification, flood detection, etc. Section 1.5.5 describes the features/bands used to conduct the studies. Furthermore, Section 1.5.6 describes the methods and algorithms that have been applied for wetland analysis using RS.

1.5.1 Types of wetlands

Peatlands

70% of the papers under review have concentrated on peatlands using RS techniques. Peatlands are a vast reservoir of carbon and mainly found in humid areas of the northern and southern hemisphere (Li, Xu & Zhao 2014, Reschke et al. 2012), for example, peatlands in Amazonian Peru (Draper et al. 2014). Hence, adding to the importance of mapping the wetlands. Peatlands can be further divided into categories such as bogs, fens, swamp, marsh, etc. (Mahdavi et al. 2018). Beamish 2014 explains there are various types of peat soils-mixtures such as wet and dry peatland soils spread across the raised and

blanket bogs, meres and fens, and afforested peat. These soils have different radiometric properties which are essential to identify and maintain for conservation. The main types of peatland are described in the following sections.

- Bogs
Bogs are ombrotrophic wetlands which are mainly covered by moss and sphagnum species. These wetlands are fed by precipitation and are generally soft and wet. Bogs can be differentiated into raised bogs and blanket bogs on the level of peat formation and mechanisms (Glatzel et al. 2020) (see Figure 1.9).

<div align="center">(a)   (b)</div>

Figure 1.9: Types of Bogs (a) Raised Bog (County, Offaly) (b) Blanket Bog (County Galway), photo by Saheba Bhatnagar

Raised bogs are mostly a mosaic of microforms (Lehmann et al. 2016) and usually have a dense layer of sphagnum moss, with the presence of ridges (up to 50 cm) and trees varying up to 3 m in height (Kohv et al. 2017). It is essential to maintain water table to a certain level in every bog if ARB conditions are to persist, and therefore, any drainage system from the bogs need to be appropriately reviewed and monitored, especially in the case of a restored bog (D'Acunha et al. 2018). Based on their ground condition, the bog complexes can be further broken down into virgin, drained, productive, old cutaway and recent cutaway (Knoth et al. 2013, McGovern et al. 2000. Studies like (McGovern et al. 2000, Knoth et al. 2013, Cole et al. 2014, Lehmann et al. 2016, D'Acunha et al. 2018 and Kohv et al. 2017) have concentrated solely on analysis of bog wetland using RS. Other details about the studies are discussed in the later sections.

- Fens

In contrast to bogs, fens are the mires of small sedge and brown moss communities which are recharged and drained by groundwater (Weiss & Crabtree 2011) as shown in figure 1.10. They generally appear on permanently waterlogged soils, with minimum fluctuations in the amount of peat and water level (Kopel et al. 2016). Fens can be further divided into rich (alkaline) and poor (acidic) depending on the nature of their recharge environment.

Figure 1.10: Scragh bog (alkaline fen), County Westmeath, photo by Saheba Bhatnagar

In the literature being reviewed, studies like (Weiss & Crabtree 2011, Kopel et al. 2016, Bourgeau-Chavez et al. 2017, and Koch et al. 2017) have mainly concentrated on alkaline or rich fens using RS techniques (see Section 1.5.3, 1.5.6). These are mainly located in areas overlying limestones. The vegetation in rich fens is mainly dominated by helophytic species, i.e. *Phragmites australis, Bolboschoenus maritimus, Schoenoplectus tabernaemontani* and *Carex acutiformis*, with grassland (Koch et al. 2017).

- Swamp and Marsh
  Another common category of peatlands is swamps and marshes (figure 1.11). Swamps and Marshes could be both mineral and organic (Mahdavi et al. 2018).

Figure 1.11: (a) Swamp (Maumee Lake Plain, Michigan, photo by Michael A. Kost) (b) Marsh (Ócsa, Hungary, photo by Gy Büttner)

These are wetlands covered with 75% or less water (Li & Chen 2005). A major difference between a marsh and a swamp is the herbaceous nature of marsh, whereas swamps

often have woody vegetation of up to 1m in height (Mahdavi et al. 2018). Vegetation communities like herbaceous, Carex, reeds, rushes are found in the marsh, and coniferous and deciduous trees, shrubs, moss are commonly found in swamps. Shallow water is one more category of peatlands encountered in the literature. Wetlands like marsh are also considered for analysis of ecological status (Díaz-Delgado et al. 2019), in conjunction with land cover mapping applications along with shallow water community (Grenier et al. 2008, Li & Chen 2005, Dingle Robertson et al. 2015, Touzi et al. 2015, Amani et al. 2017, Mahdianpari et al. 2017, Mahdianpari et al. 2018, and Amani et al. 2019). Shallow water consists of floating or submerged aquatic plants and reeds, with vegetation cover less than 25% (Grenier et al. 2008, Amani et al. 2017, Mahdianpari et al. 2017).

Non-peatlands

The main type of wetlands (non-peat based) found in literature being reviewed are mangrove wetlands, thermokarst wetlands and flooded wetlands such as turloughs. The turloughs fall under the scope of this thesis, and therefore are discussed in particular detail.

(a)   (b)
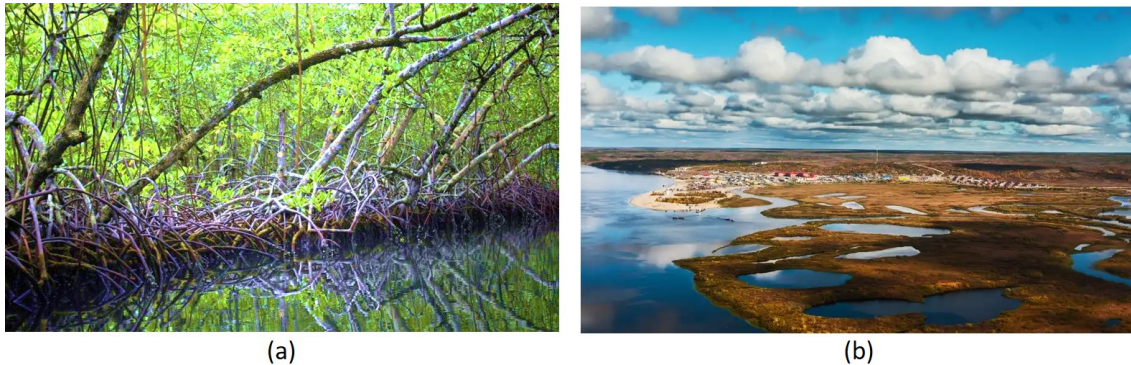
Figure 1.12: (a) Mangrove (India, photo by Ravi Sharma) (b) Thermokarst (Siberia, Russia, photo by Sharon Omondi)

- Mangrove
  Mangrove ecosystems (see Figure 1.12a) are the shrubs and forests that grow along coastal lines, river banks in subtropical and tropical climates all around the world (Wang et al. 2019b). Mangroves prefer a humid climate, freshwater inflow that brings in abundant nutrients and silt and grow luxuriantly in alluvial soils (Kathiresan & Bingham 2001). (Wang et al. 2019b) gives a detailed review exclusively on trends of RS in mangrove ecosystem monitoring. Other studies like (Longépé et al. 2011, Minasny et al. 2018) have also discussed mangroves and their importance in carbon management.

- Thermokarst
  Thermokarst lakes (see Figure 1.12b) and vegetated drained lake basins that occupy large areas of sedimentary permafrost lowlands (Jones et al. 2012, Regmi et al. 2012). They are known to be major global methane sources, and therefore, need to be monitored (Skeeter et al. 2020). The vegetation in Seward Peninsula is classified as Bering Tundra with *Ericaceous* shrubs, *Betula nana, Salix spp., Rubus chamaemorous, and Sphagnum fuscum, and Polytrichum strictum* common on the uplands, and *Calamagrostis canadensis, Salix planifolia, Salix palustre, Salix lanata, Carex aquatilis, Eriophorum angustifolium, E. scheuzeri, Sphagnum riparium, Hylacomnium splendens*, and *Tomenthypnum nitens*

common in drained lake basins (Jones et al. 2012).

- Turloughs
  Turloughs are intermittent lakes, located in karst limestone areas, occurring predominantly in Ireland. They are similar in flood dynamics to pools which are found more extensively in karst regions across the world. The flooding period usually starts in October and lasts most of the winter and spring. The flooding duration is a function of the underground karst conduit network, the rainfall and the local basin topography and so is unique to every site. With the advent of summer, the lake starts to dry, and vegetation communities start to emerge. The functioning of the turlough is mainly dependant on the characteristic periodic flooding, which plays a big role in the biological diversity of the turloughs (Naughton et al. 2012, McCormack et al. 2020). Figure 1.13 shows Blackrock turlough, County Galway, which is a classic example of flooding in turloughs. Apart from turloughs, flooded lakes, deep water also been covered in the literature (Amani et al. 2019, Berhane et al. 2018a, and Skeffington et al. 2006).

(a)	(b)

Figure 1.13: Blackrock Turlough (County, Galway) (a) Dry (b) Flooded, photo by Laurence Gill

A turlough when dry, can have a grassy appearance due to presence of sedges, with a gradation of communities down into the lower part of the turlough basin (where water is sustained the longest) to more wetland associated species such as small sedges, silverweed, and meadowsweet. Hence, a range of different vegetation communities can be associated with a topography which seems to be linked to the typical flood duration statistics across the year (Waldren et al. 2015).

Non-Wetlands

There are multiple studies that have used RS techniques to map other land types apart from wetlands. The main ones that are often located near to wetlands are discussed below. There have also been studies into various non- vegetation land types such as bare land or barren land, rocky areas, roads and urban settlements which are not discussed.

Figure 1.14: (a) Agricultural land in Egypt (source Maruyama et al. 2017) (b) Uplands in Kilimanjaro (source Sangeda et al. 2014)

- Agricultural lands
  Agricultural lands or croplands (for example, Figure 1.14a) are the areas cultivated or harvested for crop productions. They are often affected by grazing, fertilisers, and hence, the quality of soil is always subjective. Such lands are also analysed in conjunction with wetlands in most studies (White et al. 2017, Zhang et al. 2016, Minasny et al. 2018, Mahdianpari et al. 2018, and Amani et al. 2019). This is due to the location of croplands adjacent to peatlands, and hence, in a large-scale regional mapping study, all the classes are considered (White et al. 2017, Mahdianpari et al. 2018, and Amani et al. 2019). The locations, RS data and techniques used by these studies are discussed in detail in Sections 1.5.2-1.5.6.

- Uplands
  Apart from croplands, another region most commonly present adjacent to wetlands can be uplands. The uplands are usually the areas at higher altitudes than wetlands, consisting of denser vegetation. The type of vegetation, and their water holding capacity has an effect in terms of feeding the downstream wetlands. Many studies like (Touzi et al. 2009, Jones et al. 2012, White et al. 2017, and Montgomery et al. 2019) have analysed uplands as well in their study of wetlands for applications discussed in Section 1.5.4.

- Forests
  Forests are another land type frequently located near to wetlands. Forests can be of many types, such as boreal, temperate, tropical, etc. and cover a vast amount of area of the world. Forest can also be classified according to their age (Simkin et al. 2020) (see Figure 1.15). Forests can also divided as per the tree species using satellite data as described in (Breidenbach et al. 2020).

Figure 1.15: Forests in Finland (a) Pristine (b) Mature, source Simkin et al. 2020

There are instances where wetlands are located just on the edge of forests, making the differentiation between the two difficult using satellite imagery (Amani et al. 2019). Therefore, many studies have considered mapping forests at par with wetlands when mapping at a large scale (regional/national). Studies like (Kopel et al. 2016, Wang et al. 2019b, Mahdianpari et al. 2018, and Amani et al. 2019) mentions about forests.

1.5.2 Location of the Wetlands

Wetlands constitute 6% of the entire globe (Junk et al. 2013). The location of wetlands varies across the continents. Ramsar has designated special status to a few wetlands per country. According to Ramsar (2018) Africa has 39 sites, Asia has 32 sites, Europe has 110 sites, South America with 20 sites, and North America has 21 sites (Gardner & Finlayson 2018) under special status. Peatlands in specific have a tendency to grow in area with high rainfall, and low temperatures, Pittock et al. 2015 describes how the peatlands are distributed around the globe (figure 1.16).
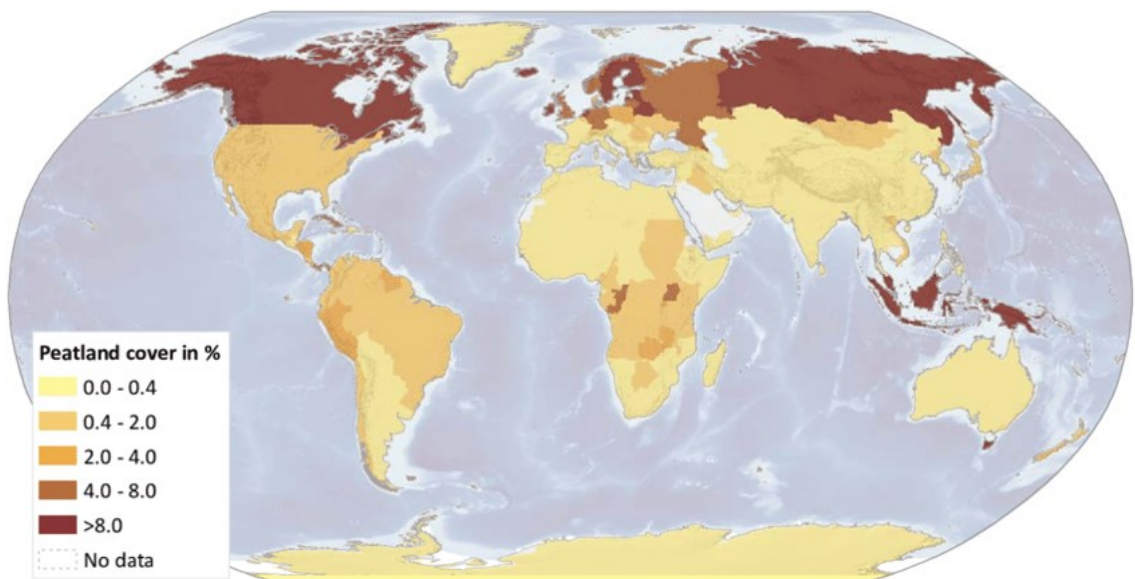
Figure 1.16: Global distribution of peatland, source Pittock et al. 2015

In the literature that was being reviewed (Appendix A), the studies on peatlands were mainly located in North America (42%). There were many studies on Canadian peatlands, for example, Ottawa (Touzi et al. 2009), Ontario (Dingle Robertson et al. 2015, White et al. 2017), Quebec (Dissanska et al. 2009), Newfoundland and Labrador (Amani et al. 2017). There are many studies which have analysed wetlands at a national scale in Canada (Merchant et al. 2016, Mahdianpari et al. 2017, Berhane et al. 2018b, Amani et al. 2019). In the United States of America, the wetland studies were from Alaska, like Yukon flats (Weiss & Crabtree 2011), Seward Peninsula (Regmi et al. 2012), Texas (Gou et al. 2015), and studies like (Bourgeau-Chavez et al. 2015, Rahman et al. 2017) concentrate on the analysis of wetlands at a national scale. The studied wetlands located in Europe continent were around 22% of those reviewed. Studies like (McGovern et al. 2000, Beamish 2014, McCormack et al. 2020) concentrated on wetlands in Ireland. Whereas, studies like (Beamish 2014, Crichton et al. 2015, Cole et al. 2014) investigated wetlands in the UK. In mainland Europe, studies are mainly from Germany (Knoth et al. 2013), Poland (Kopel et al. 2016, Szporak-Wasilewska et al. 2015), Romania (Díaz-Delgado et al. 2019). The whole of Asia covers around 20% of the wetland studies under review, with more than 10% carried out in China (Sun et al. 2019, Hu et al. 2020, Li et al. 2019, Wang et al. 2019a, Balogun et al. 2020). Studies like (Reschke et al. 2012, Berhane et al. 2018a, Berhane et al. 2018b) have focused on wetlands in Russia. Many studies like (Jaenicke et al. 2008, Minasny et al. 2018) have looked at Indonesian wetlands, and around 5% in South America like Amazon (Draper et al. 2014), and Chile (Cabezas et al. 2016).

1.5.3 Types of Data

There are a variety of data used in remote sensing. The data typically used for analysing wetlands remotely can be divided into two main categories, satellite and aerial (incl. capture using the unmanned aerial vehicle (UAV)), followed by several sub-categories, as shown in Figure 1.17. Ho 2009 describes that data can be also be divided with respect to sensors: Active and Passive. Active sensors emit the signal to the ground, and the signal reflected back is recorded as the data, for example, Radio detection and ranging (RADAR) and Light detection and ranging (LiDAR). In comparison the passive sensors does not have their own source of illumination and often rely on the sun to provide energy to record the data, for example, optical and infrared satellites, and optical UAV. Scope of this thesis is mainly focused on the usage of passive data, but first all types of remote sensing data are discussed in this review section.
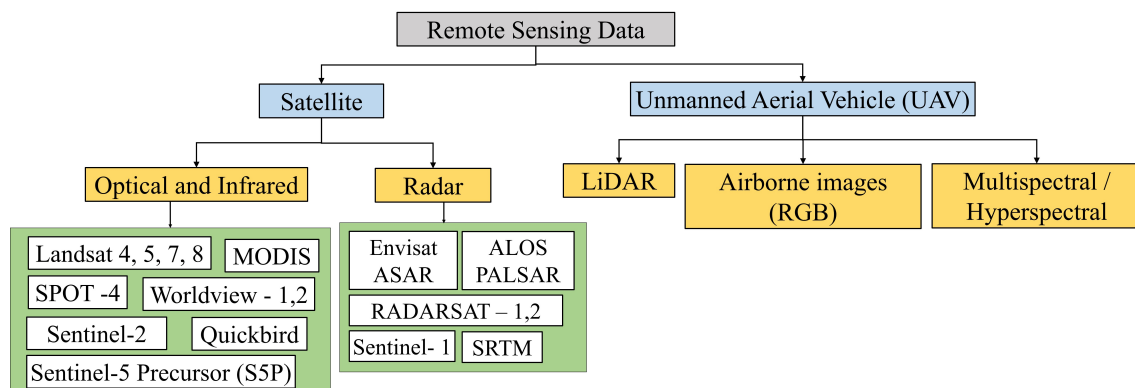
Figure 1.17: Types of RS data in literature being reviewed

Satellite Data

- Optical and Infrared Satellite
  Optical and infrared (IR) satellite data are most commonly used for landcover mapping. These satellites operate at the wavelengths between 400-15000nm. This range includes the visible (400–700 nm), the near-infrared (NIR, 700–1100 nm), the short infrared (SWIR, 1100–2500 nm), the mid-wave infrared (MWIR, 2500–7500 nm) and the long-wave infrared (LWIR, 7500–15000 nm) spectral regions.

  Optical and IR remote sensing involve the analysis of electromagnetic radiation captured after reflecting from the surface or ground. Different objects react differently to different wavelengths. Therefore, by studying the spectral characteristics of the objects, differentiation between them is possible. Vegetation and water bodies constitute of colour, chlorophyll, mesophyll, minerals, dissolved organic matter, etc., substances which all have an impact in the formation of the overall spectral signature. For chlorophyll detection, the vegetation absorbs red and blue wavelengths and reflects green (hence, its visual colour). Similarly, mesophyll is identified using the reflectance curve at NIR wavelengths. For understanding the water absorption characteristics in vegetation, SWIR is useful. There is a dip at 1400, 1900 and 2500 nm wavelengths depicting absorption. Which means that vegetation containing water, and water bodies such as wetlands can be identified and delineated using such RS data.

  There are around 50 optical and IR satellites orbiting around the earth, with multiple new missions already in the pipeline. Figure 1.18 gives a consolidated timeline of satellites from the very beginning (1970) to the near future, as described in Houborg et al. 2015. Some
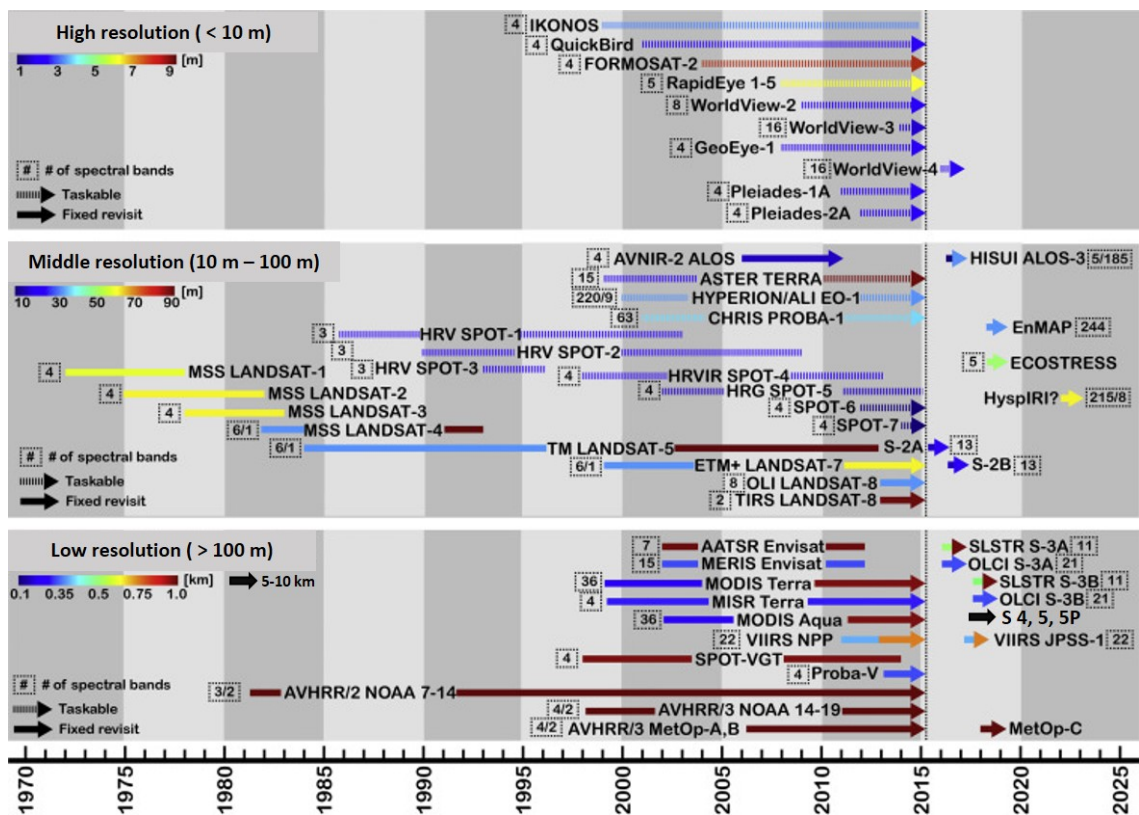
Figure 1.18: Timeline of existing and prospective optical and IR satellites from 1970 to 2025

other passive satellites like Sentinel-5 series (S5 and S5-Precursor (S-5p)) were recently launched (in 2018) by European Space Agency (ESA) for monitoring the greenhouse gases (GHGs) on a global scale. The main type of satellite data encountered in the literature being reviewed is discussed below.

– Moderate Resolution Imaging Spectroradiometer (MODIS)
The Moderate Resolution Imaging Spectroradiometer (MODIS) is an earth observation instrument launched aboard the Terra (1999) and Aqua (2002) satellites. It produces one of the most widely used open-source satellite data on a global scale (Friedl et al. 2010), as shown in Figure 1.19. MODIS (flying on-board with Terra and Aqua) has a swath of around 2330 km viewing width. It acquires data in 36 spectral bands ranging between wavelengths 415 nm to 14235 nm (Savtchenko et al. 2004). It has a spatial resolution varying from 1km to 250m depending on the wavelength and location with a temporal resolution of 1 day.
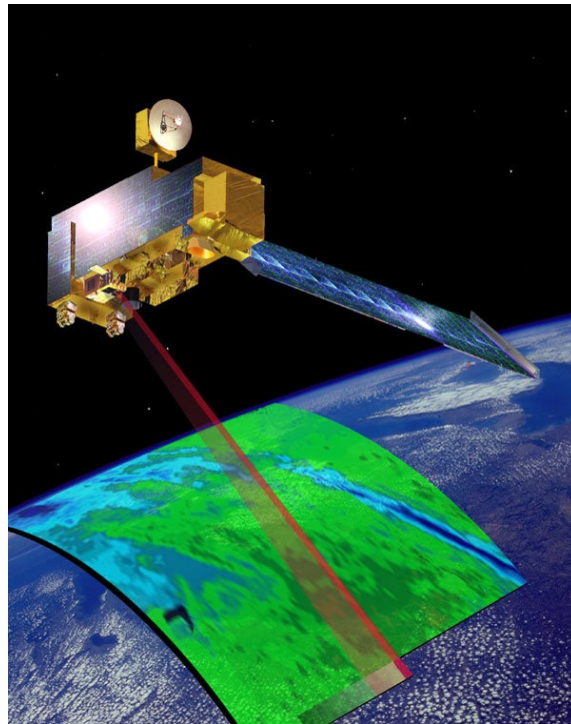
Figure 1.19: MODIS satellite covering images, source Cassardo et al. 2014

Studies like (Weiss & Crabtree 2011), (D'Acunha et al. 2018) have used MODIS data to analyse wetlands at a large country-wide scale due to the coarse resolution of the data. NASA provides more than 30 products from MODIS, including raw imagery, land products, ocean products, atmospheric products and cryosphere products (Savtchenko et al. 2004). The products are available for free and can be directly downloaded from (`https://modis.gsfc.nasa.gov/data/dataprod/`).

– Landsat series
The land-satellite, i.e., Landsat, was launched by NASA first in 1972, Landsat-1. It was followed up by Landsat-2 in 1975, and Landsat-3 in 1978. In 1982, Landsat-4 thematic mapper (TM) was launched. Before that, all the Landsat satellites were using a multi-spectral scanner (MSS). With the introduction of TM, new bands in the wavelength range

of SWIR was explored (Markham et al. 2004). Both Landsat-4 and 5 were very similar, with Landsat-5 being better in radiometric corrections (Markham et al. 2004). With further advancement in RS technologies, there was a considerable improvement in spatial resolution. An Enhanced Thematic Mapper (ETM+) was placed on-board with Landsat 7, launched in 1999. Due to a scan line corrector (SLC) failure in 2003, there have been gaps found in data collected by Landsat-7. The gaps in data range from none at the centre to 14 pixels at the extreme edges of the scan (Markham et al. 2004). Landsat-9 (to be launched in 2021) would replace Landsat-7. Lastly, Landsat-8, launched in 2013, carries Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) instruments. It has a resolution similar to Landsat-7 ETM+, with an overlapping swath of $185km \times 180km$ and extra spectral bands to for aerosol analysis, and cirrus cloud contamination detection. All Landsat satellites are sun-synchronous, placed at an altitude of 705km. Figure 1.20 depicts the Landsat satellites.

(a) (b) (c)
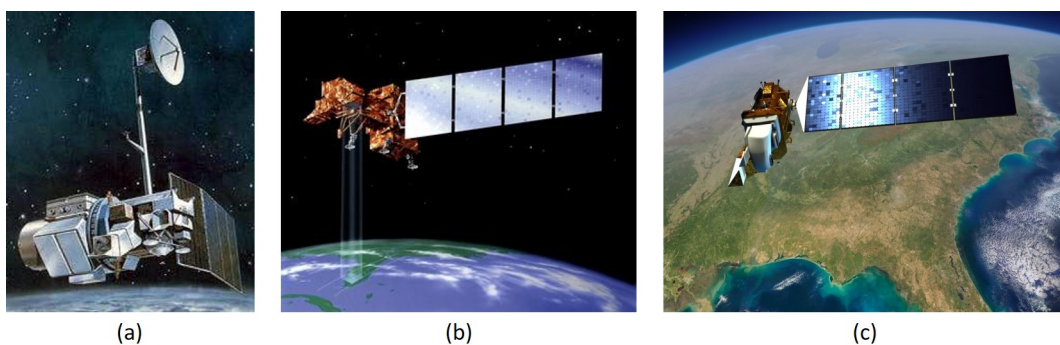
Figure 1.20: Landsat satellite (a) Landsat 4,5 (b) Landsat 7 (c) Landsat 8, source *United States Geological Survey* 2020

Landsat TM data was used in studies like (McGovern et al. 2000), (Regmi et al. 2012), and (Hu et al. 2020). Table 1.2 gives the details about resolutions and swath of the Landsat series. Landsat 4, 5 Multispectral Scanner (MSS) has a resolution of 60m, but Landsat Thematic Mapper (TM) has an enhanced spatial resolution of 30m visible+NIR (VNIR). The presence of ETM+ increased the spatial resolution of NIR wavelength to 15m and enhanced the spatial resolution of TIR wavelengths to 60m. Studies like Li & Chen 2005, Grenier et al. 2008 and Bourgeau-Chavez et al. 2015 have used Landsat 7 ETM+ data for characterising wetlands in Canada and USA. Other studies like Crichton et al. 2015, Draper et al. 2014, and have used Landsat 7 ETM+ for analysing wetlands in the UK and Amazonia respectively. Whereas more recent studies have applied Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) successfully for wetland analysis (White et al. 2017, Hu et al. 2020, Amani et al. 2019, Li et al. 2019, Wang et al. 2019a, and Balogun et al. 2020).

	Landsat 4	Landsat 5	Landsat 7	Landsat 8
Spectral bands	4 VNIR, 2SWIR, 1 TIR	4VNIR, 2 SWIR, 1 TIR	4VNIR, 2 SWIR, 1 TIR, 1 PAN	4VNIR, 2 SWIR, 2 TIR, 1 PAN, +2
Spatial Res (m)	30 (VNIR, SWIR), 120 (TIR)	30 (VNIR, SWIR), 120 (TIR)	30 (VNIR, SWIR) 60 (TIR), 15 (PAN)	30 (VNIR, SWIR) 100 (TIR), 15 (PAN)
Temporal Res (days)	16	16	16	16
Swath (km)	185	185	183	185 x 180

Table 1.2: Spectral, Spatial, Temporal resolution and swath of Landsat series

– Sentinel-2

In parallel to Landsat series, in 2015 the Sentinel series was launched by ESA. Sentinel-2 (S2) Multispectral Instrument (MSI), is the followup mission to the Système Pour l'Observation de la Terre (SPOT) and Landsat series, aimed at providing betterment in remote sensing products (Malenovskỳ et al. 2012). S2 offers satellite images with a resolution of 10 to 60 m (Drusch et al. 2012). In comparison to Landsat-8 OLI/TIRS, S2 has a better spatial resolution, better spectral resolution in the near-infrared region, three Vegetation Red Edge (VRE) bands with 20 m spatial resolution, but does not offer thermal data (Kaplan & Avdan 2017). Figure 1.21 describes the spectral range of S2 MSI.
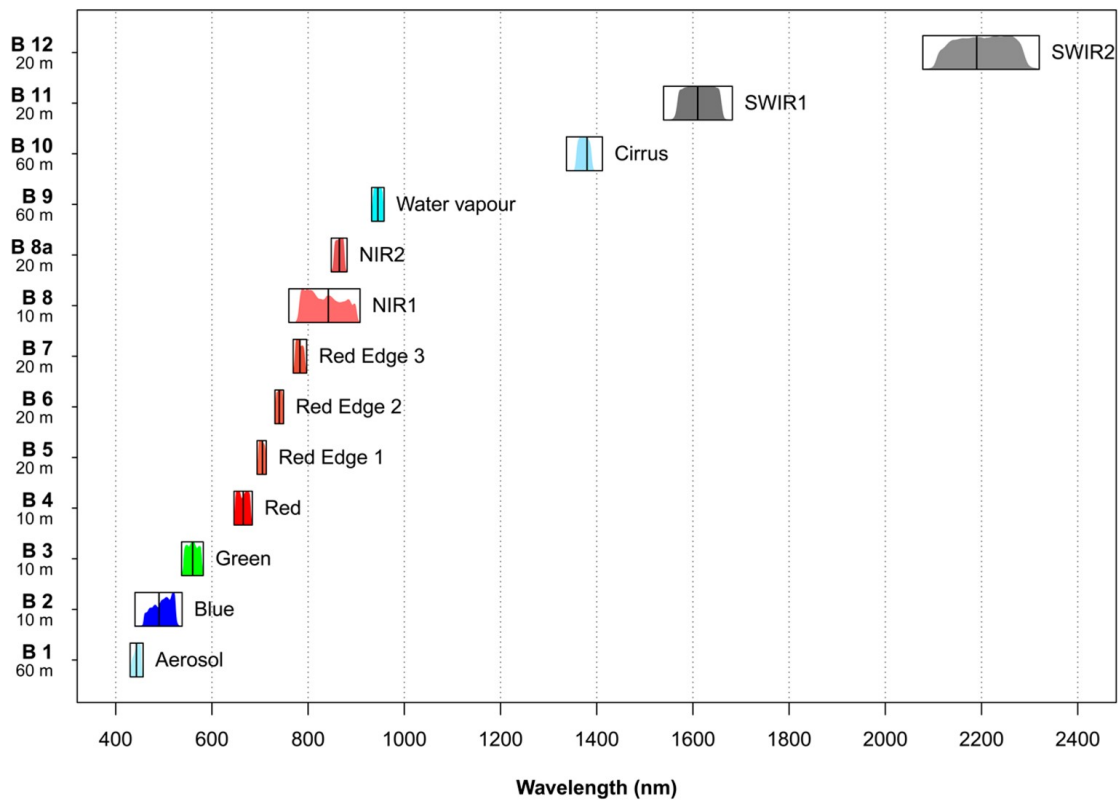
Figure 1.21: Sentinel-2 MSI spectral bands, source Immitzer et al. 2016

In 2017, Sentinel-2B was also launched. This has improved the overall temporal resolution of the satellite, 5 days at the equator and 2-3 days at mid-latitudes using the 2 satellites (Figure 1.22). The S2 orbits in a sun-synchronous orbit (i.e., fixed relative to the sun's position) at an altitude of 786 km, making its swath=290km which is bigger than Landsat-8. The S2 data can be downloaded, free of cost from (`https://scihub.copernicus.eu/`).
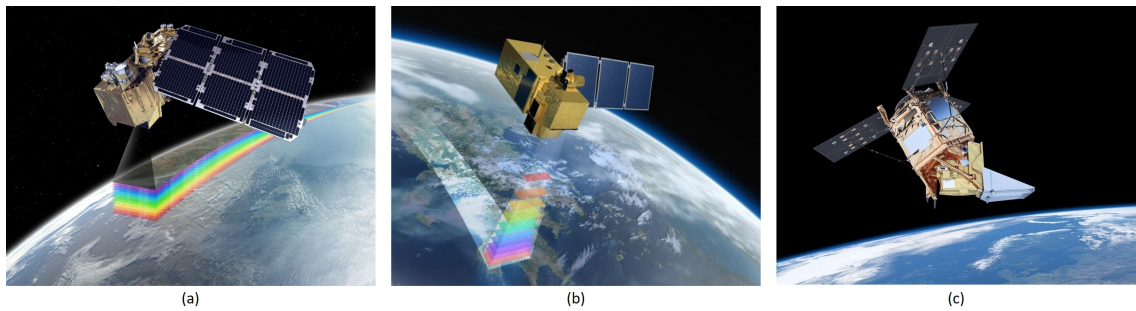
Figure 1.22: (a) Sentinel-2A (b) Sentinel-2B (c) Sentinel-5P, source *European Space Agency* 2020

Multiple studies like (Kaplan & Avdan 2017, Zhang et al. 2019, Díaz-Delgado et al. 2019) have successfully applied sentinel-2 for wetland-vegetation and water analysis. There have been multiple studies integrating data from both Landsat, and S2 for land cover mapping and analysis (Mandanici & Bitelli 2016, Skakun et al. 2017). Studies like Korhonen et al. 2017 states that the presence of 3 VREs in the S2 MSI data improved the estimation of biophysical parameters of vegetation. In the literature being reviewed, the main studies on wetland classification concentrating on S2, are from the past two years, and therefore, this data needs more exploration.

– Sentinel-5P

The Sentinel 5P launched by ESA consists of a satellite bus, a payload-carrying Tropospheric Monitoring Instrument (TROPOMI) and a ground station (Veefkind et al. 2012). The S-5p was launched in October 2017 and is operational for the timeframe 2017-2023.The S-5p covers the globe daily, with 2600 km swath width at a time, with a 1 day temporal resolution (Figure 1.22 c). The spatial resolution of S-5p is 5.5×7 km. The main aim of the S-5p is to monitor air quality and climate activities using products derived from wavelengths ranging from ultraviolet (UV) to short wave infrared (SWIR). Table 1.3 gives the spectral range, and level-1, 2 products of S-5p.

Wavelength (range nm)	Spectrometer	L1 Product	L2 Products
$270 - 300$	UV	Radiance Band 1	Vertical Column (HCHO,
300 - 320	UV	Radiance Band 2	SO2, NO2, O3), O3 Vertical
$320 - 405$	UVIS (UV, Visible)	Radiance Band 3	profile, Aerosol index, layer
$405 - 500$	UVIS	Radiance Band 4	height, Cloud parameters, L2
$675 - 725$	NIR	Radiance Band 5	O3 Tropospheric vertical
$725 - 775$	NIR	Radiance Band 6	column
$2305 - 2345$	SWIR	Radiance Band 7	Vertical column (CO, CH4)
$2345 - 2385$	SWIR	Radiance Band 8	Vertical column (CO, CH4)
270-775	UVN (UV,VIS,NIR)	Irradiance UVN	Vertical column (CO, CH4)
$2305 - 2385$	SWIR	Irradiance SWIR	Vertical column (CO, CH4)

Table 1.3: Spectral details about Level-1 and Level-2 products of Sentinel 5P

Studies like Lunt et al. 2019 have used S-5p data for estimating the methane trends in Africa. Out of all the available L2 products, this thesis concentrates on the vertical column CH4 retrieved using the SWIR wavelengths for land type identification, and methane trend analysis for the year 2018, 2019, see Chapter 6.

– Commercial satellites

Other than open-source, multiple commercial satellites SPOT-4, SPOT-5, Worldview 1, 2, Quickbird, GeoEye, IKONOS have also been launched. The data from these satellites are usually very high-resolution, and hence, also expensive. However, SPOT- 4 provides a spatial resolution compatible to Sentinel-2 at 10m (panchromatic) and 20m for (VNIR) (Grenier et al. 2008, Montgomery et al. 2019) but the current version SPOT-5 has a spatial resolution of $10 - 2.5$m (Niculescu et al. 2016). Studies such as Davranche et al. 2010 discuss the application of SPOT-5 successfully for wetland monitoring using classification trees.

The industry of commercial satellites has grown very rapidly (Dolgopolov et al. (2018)). DigitalGlobe is an American company, founded in 1992, which operates a range of high spatial resolution satellite sensors, such as IKONOS (1999), QuickBird (2001), WV1 (2007), GeoEye1 (2008), WV2 (2009), WV3 (2014), shown in figure 1.23. Apart from Digital-Globe, there are around 175+ PlanetScope and 13 SkySat in orbit. The sensors and satellites from Planet, including RapidEye, can capture images anywhere on Earth daily at 3m and 72cm resolution (Chen et al. 2019).

Figure 1.23: DigitaGlobe sensor satellites, source *DigitalGlobe* 2020

All these satellites are sun-synchronous. WV2 satellite provides a high temporal and spatial resolution of $1.7 - 3.7$ days and 0.5m, respectively. On the other hand, the satellite QuickBird has a temporal resolution of 1-3.5 days and spatial resolution varying from 0.65m to 2.9m. This discrepancy in resolutions is at nadir and off-nadir points. The details such as spectral bands, a nominal spatial resolution, orbit altitude, and swath about these satellite sensors are listed in table presented in Figure 1.24.

	Ikonos	Quickbird	GeoEye-1	Worldview-1	Worldview-2
Spectral Resolution (nm)	445–900 (P) 445–516 (B) 506–595 (G) 632–698 (R) 757–853 (N)	450–900 (P) 450–520 (B) 520–600 (G) 630–690 (R) 760–900 (N)	450–800 (P) 450–510 (B) 510–580 (G) 655–690 (R) 780–920 (N)	400–900 (P)	400–450 (C1) 585–625 (Y) 705–745 (RE) 860–1040 (N)
Spatial Resolution (m)	1.83 × 1.57	2.9 × 2.62	1.84 × 0.46	3.6 × 2.5	4.3 × 2.5
Panchromatic resolution (m)	0.82	0.65	0.46	0.50	0.46
Multispectral resolution (m)	4	2.44/1.63	1.64	N/A	1.85
Accuracy specification (m)	9/15	23	3	6.5	6.5
Temporal Resolution (revisit days)	3	1–3	2.1 (0.59 m GSD) 2.8 (0.50 m GSD) 8.3 (0.42 m GSD)	1.7 (1 m GSD) < 5.9 (0.55 m GSD)	1.1 (1 m GSD) < 3.7 (0.52 m GSD)
Swath width (km)	11.3	16.8/18	15.3	17.7	16.4
Spatial coverage (km)	50 × 112 (M) 11 × 120 (St)	16.8 × 16.8/11.2 × 11.2 (M) 16.8/11.2 × 360 (St)	50 × 300 (M) 28 × 224 (St)	60 × 110 (M) 30 × 110 (S)	96 × 110 (M) 48 × 110 (St)
Operational Altitude (km)	681	450/482	681	496	770
Data availability	Sep 1999–Jan 2015	Oct 2001–Jan 2015	Sep 2008–	Oct 2007–Jan 2015	Oct 2009–
Mission life (years)	16	14	7–15	> 11	7.25

Figure 1.24: Description of DigitalGlobe sensor satellites, source Chen et al. 2019

The high spatial resolution of these satellites is also due to their low attitude and smaller swath. The introduction of red-edge bands in WV 2, 3 have increased the applicability of this data for very detailed analysis. Studies like Dingle Robertson et al. 2015, Berhane et al. 2018b, and Wang et al. 2019a have applied WV1, 2 for wetland analysis in parts of Canada, Russia and China respectively. Whereas, other studies like Dissanska et al. 2009, Berhane et al. 2018a have used data from QuickBird satellite. IKONOS satellite, decommissioned in 2013, has also been used for wetland vegetation analysis (Crichton et al. 2015). A study by Pu & Landry 2012 suggests that due to the addition of four new bands in WV2 (yellow, red-edge, NIR2), it outperforms IKONOS in tree-species detection. Another study by McCarthy & Halls 2014 which compared WV, QuickBird and IKONOS states WV2 performs the best for mapping island habitats.

- RADAR Satellite
  RAdio Detection And Ranging (RADAR) remote sensing, also known as active microwave remote sensing is when the satellite emits microwave radiation which hits the ground and is reflected to the sensor. Rather than temperature (as in the case of IR satellites), the microwave RS depends on the structure of the ground and the amount of signal reflected from it. Therefore, it can penetrate the clouds, and clear images are obtained even on a cloudy day. Radar was initially developed in the 1950s, called SLAR (Side-Looking Airborne Radar) which was limited by the physical size of the antenna. As a solution to this limitation, SAR (Synthetic Aperture Radar) was developed. SAR is not dependent on the physical antenna size as it "synthesises" a very broad antenna by sending multiple signals consecutively. Figure 1.25 shows the timeline of the RADAR satellites from the early 2000s to the present.
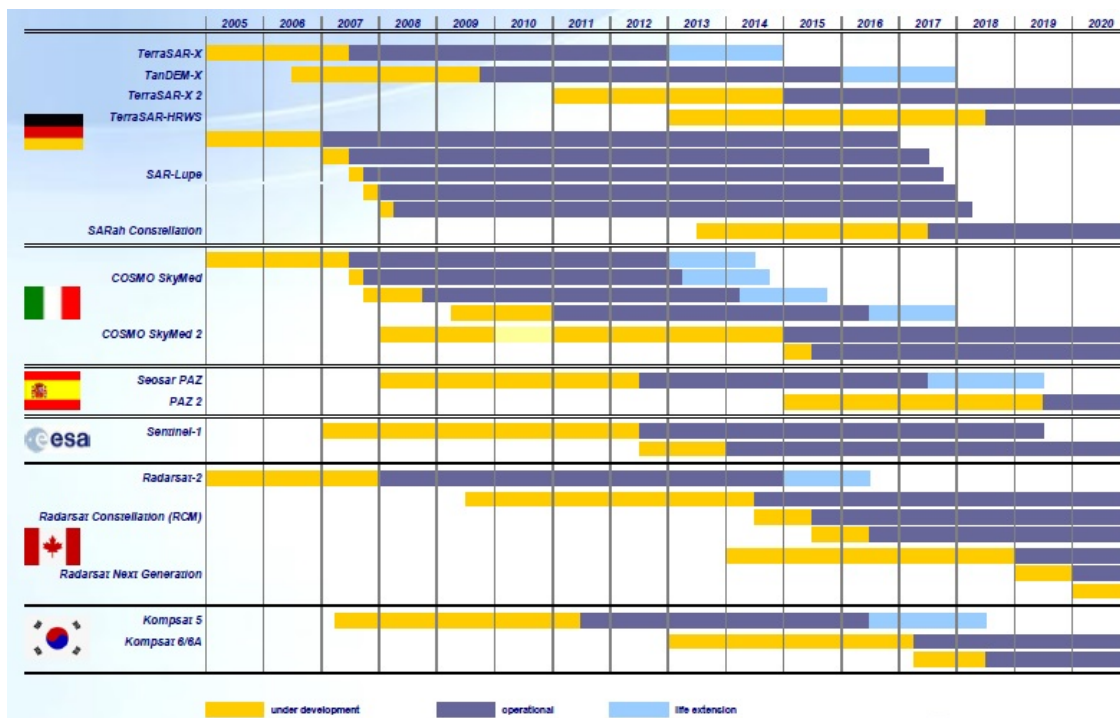
Figure 1.25: Timeline of RADAR satellites (2005 – 2020), source Döngi 2011

There are multiple operating modes of SAR (Moreira et al. 2013). Strip map mode is the mode where the antenna stays at a fixed position either 90° or slightly tilted at an angle. In ScanSAR mode the signal from the antenna sweeps periodically and therefore, covers a much larger area. In spotlight mode, similar to an actual spotlight, the area under consideration is illuminated for a longer period of time. Therefore, for this small area under consideration, much better resolution is achieved. Apart from these conventional modes, the ESA's Sentinel-1 also operates on other modes using Terrain Observation with Progressive Scan (TOPS) SAR (TOPSAR) techniques (De Zan & Guarnieri 2006). These modes are Interferometric Wide swath (IW) mode (which is a combination of scanSAR and TOPSAR), Extra Wide (EW) swath mode (similar to IW, but covers a swath of 400km), and wave mode (it is a type of single stripmap mode with alternating elevation beam at a fixed on/off duty cycle). There are multiple polarisation-modes of a SAR as well, which are discussed in Section 1.5.5.

– Open-source
Radar satellites like Environmental Satellite (ENVISAT) Advanced Synthetic Aperture Radar (ASAR) (decommissioned) was launched by ESA in 2002 into a sun-synchronous orbit at 800 km altitude. This satellite operated at C-band (wavelength $\approx 0.06m$) with an alternating spatial resolution of 30m (Reschke et al. 2012). ENVISAT has also been used for global monitoring. The study by Bartsch et al. 2009 utilised Global Monitoring (GM) mode in C-HH polarisation with 1 km resolution for monitoring the wetlands.
The ENVISAT was decommissioned in 2012. In its place, Sentinel-1 (S1) was launched in 2014. Sentinel-1 is an open-source c-band satellite with a temporal resolution of 12 days and an alternating spatial resolution of 20m. The swath covered by S1 depends on its mode of operation. In IW mode, the swath is 250km, in SM mode 80km, in wave mode 100km and 400km in EW mode. The three products of S1 are as follows.

28

1. Single Look Complex (SLC)
   The SLC is mainly used for producing DEM and topographic analysis. This product does not retain amplitude information.

2. Ground Range Detected (GRD)
   The GRD contains amplitude information, and is hence, used for vegetation mapping, soil moisture detection, land cover analysis.

3. Ocean
   The ocean product is mainly used for ocean studies, and ocean wind field analysis.

S1 been actively applied for wetland analysis (Minasny et al. 2018), especially flood inundation studies (McCormack et al. 2020) and change detection (Whyte et al. 2018). There have been multiple studies combining S2, and S1 data (Bartsch et al. 2009, Slagter et al. 2020) as it enhances temporal and spectral resolutions.
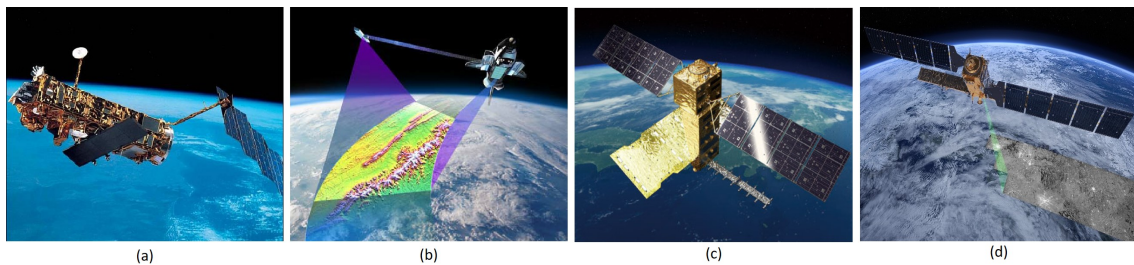
Figure 1.26: RADAR open-source satellites (a) ENVISAT (source *European Space Agency* 2020) (b) SRTM (source *United States Geological Survey* 2020) (c) ALOS PALSAR (source *Japan Aerospace Exploration Agency* 2020) (d) Sentinel-1 (source *European Space Agency* 2020)

Another satellite-like Space Shuttle Radar Topography Mission (SRTM) operates on both C-band and X-band SAR. It was launched in 2000 and is still operational. It covers a swath of 225km and provides digital elevation model (DEM) between 60° north and 56° south with a spatial resolution of 10m and temporal resolution of 11 days. DEMs provided by SRTM have been used in multiple studies like (Jaenicke et al. 2008, Draper et al. 2014, Bourgeau-Chavez et al. 2015, and Minasny et al. 2018).
Apart from the popular C-band, satellites like Advanced Land Observing Satellite (ALOS) - Phased Array type L-band Synthetic Aperture Radar (PALSAR) uses L-band SAR, which is capable of capturing images day and night. The DEM from ALOS, ALOS 2 are made free for public use by Japan Aerospace Exploration Agency (JAXA) and has been applied in various studies for wetland analysis (Longépé et al. 2011, Touzi et al. 2015, Bourgeau-Chavez et al. 2017, Mahdianpari et al. 2017, Minasny et al. 2018). This satellite provides a spatial resolution of 7-44m (high-resolution mode), and 100m (at scanSAR mode), the swath covered by it varies from 40-300km depending on the mode of operation. Figure 1.26 shows the above-mentioned open-source radar satellites.

– Commercial
   Other than open-source, there are multiple commercial radar satellites available (Figure 1.27). RADAR satellite (RADARSAT)-1, launched in 1995, was Canada's first commercial satellite. The satellite was C-band, operating at narrow and wide scanSAR mode, with resolution varying from 50-100m. This satellite was decommissioned in 2013. The RADARSAT-2 is a C-band satellite with a spatial resolution varying from 10 to 100m

depending on the mode of operation. Consecutively, the swath width also varies from 18-500km. It was launched in 2007 and has been used widely in wetland analysis (Dingle Robertson et al. 2015, Mahdianpari et al. 2017, Merchant et al. 2016, White et al. 2017, and Montgomery et al. 2019). TerraSAR-X is another radar satellite which captures images at X-band SAR with a spatial resolution of 1m (high-resolution spotlight mode) to 18.5m (scanSAR mode) (Regmi et al. 2012, Mahdianpari et al. 2017). Both C and X band
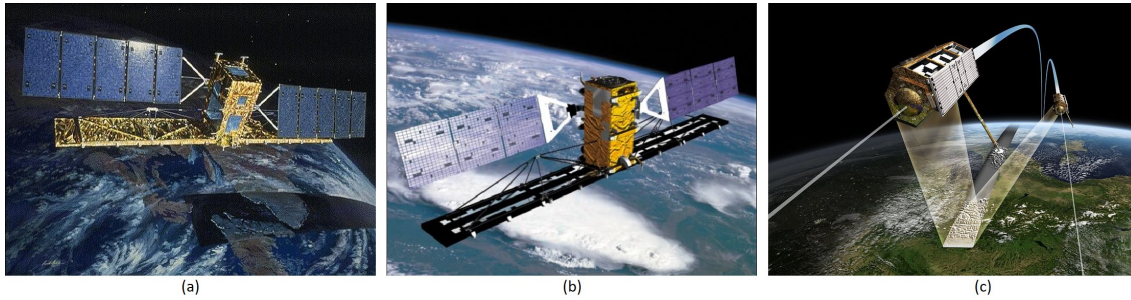
Figure 1.27: RADAR commercial satellites (a) RADARSAT-1 (b) RADARSAT-2 (c) TERRASAR-X, source *European Space Agency* 2020

radar are excellent and resistant to adverse weather conditions with X band having higher resolution due to higher frequency. L band, although it is at low frequency, is easier to design and its production is cost-effective when compared to C or X bands. Many studies have applied combinations of all these satellites for wetland analysis (Kim et al. 2013). Figure 1.28 shows an image captured using European Remote Sensing Satellite (ERS), Japanese Earth Remote Sensing Satellite (JERS) and Radar Satellite (RADARSAT) for the same area of Florida. There are multiple RADAR satellites present, the frequency band, and polarisation of some of them are presented in Figure 1.29.
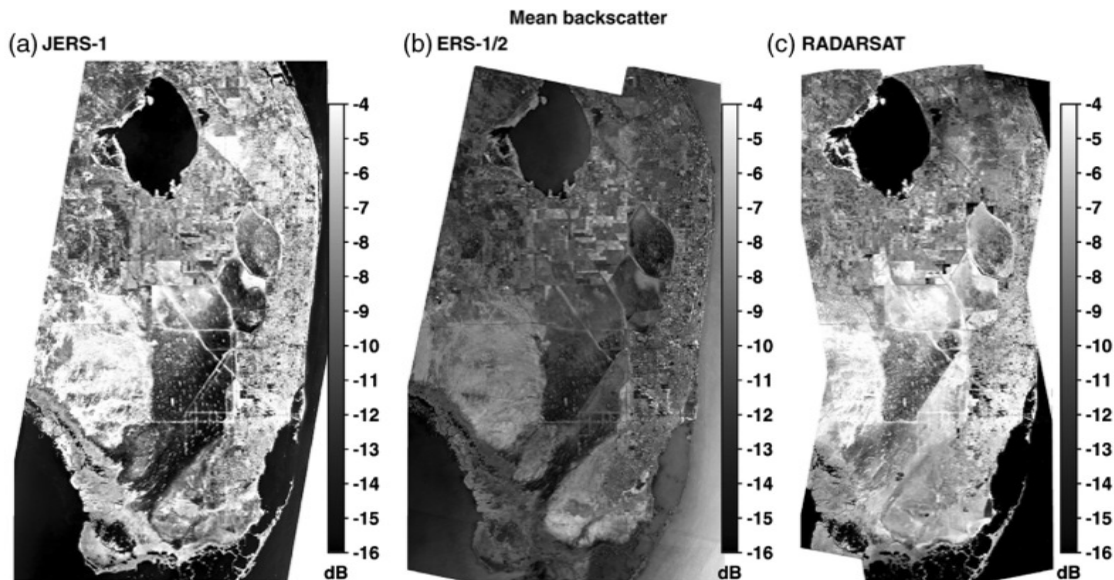
Figure 1.28: SAR mean backscatter images of South Florida (a) Mean SAR backscatter images of JERS-1, (b) ERS-1/2, and (c) RADARSAT, source Kim et al. 2013

SENSOR	FREQUENCY BAND (POLARIZATION)	COMMENTS	INSTITUTION, COUNTRY
Seasat	L (HH)	First civilian SAR satellite, operation for only ca. three months	NASA/JPL, USA
ERS-1/2	C (VV)	European Remote Sensing Satellites (first European SAR satellites)	ESA, Europe
J-ERS-1	L (HH)	Japanese Earth Resource Satellite (first Japanese SAR satellite)	JAXA, Japan
SIR-C/ X-SAR	L & C (quad) X (VV)	Shuttle imaging radar mission, first demonstration of spaceborne multi-frequency SAR	NASA/JPL, USA DLR, Germany ASI, Italy
Radarsat-1	C (HH)	First Canadian SAR satellite, swath width of up to 500 km with ScanSAR imaging mode	CSA, Canada
SRTM	C (HH+VV) and X (VV)	Shuttle Radar Topography Mission, first spaceborne interferometric SAR	NASA/JPL, USA DLR, Germany ASI, Italy
ENVISAT/ ASAR	C (dual)	First SAR satellite with Transmit/Receive module technology, swath width up to 400 km	ESA, Europe
ALOS/PalSAR	L (quad)	Advanced Land Observing Satellite (Daichi), swath width up to 360 km	JAXA, Japan
TerraSAR-X/ TanDEM-X	X (quad)	First bi-static radar in space, resolution up to 1 m, global topography available by end of 2014	DLR/Astrium, Germany
Radarsat-2	C (quad)	Resolution up to: 1 m × 3 m (azimuth × range), swath width up to 500 km	CSA, Canada
COSMO-SkyMed-1/4	X (dual)	Constellation of four satellites, up to 1 m resolution	ASI/MiD, Italy
RISAT-1	C (quad)	Follow-on satellite (RISAT-1a) to be launched in 2016, RISAT-3 (L-band) in development	ISRO, India
HJ-1C	S (VV)	Constellation of four satellites, first satellite launched in 2012	CRESDA/CAST/ NRSCC, China
Kompsat-5	X (dual)	Korea Multi-Purpose Satellite 5, resolution up to 1 m	KARI, Korea
PAZ	X (quad)	Constellation with TerraSAR-X and TanDEM-X planned	CDTI, Spain
ALOS-2	L (quad)	Resolution up to: 1 m × 3 m (azimuth × range), swath width up to 490 km	JAXA, Japan
Sentinel-1a/1b	C (dual)	Constellation of two satellites, swath width up to 400 km	ESA, Europe
Radarsat Constellation-1/2/3	C (quad)	Constellation of three satellites, swath width up to 500 km	CSA, Canada
SAOCOM-1/2	L (quad)	Constellation of two satellites, fully polarimetric	CONAE, Argentina

Figure 1.29: Summarising description about RADAR satellites, source Moreira et al. 2013

UAV Data

In this study, UAV data mainly refers to airborne images captured using UAVs, and pilot-aircraft-missions. These airborne images provide flexibility in spatial, spectral and temporal resolution, which often is not the case with satellites due to their fixed sensor and orbits (Bhatnagar et al. 2020b). Since the last decade, UAVs and drones have become very popular due to their compact sizes, and lower cost.

The two main types of drones are fixed-wing drones and multi-rotor drones which are available in similar price ranges (Floreano & Wood 2015). Depending on the manufacturer, both types of drones are capable of carrying a range of different sized sensors. Due to a simpler design and less complex mechanism, fixed-wing drones have longer flight time when compared to multi-rotor drones. However, fixed-wing rotor drones need to have a runway for launching, which is why multi-rotor drones are generally preferred when the application area is compact or unknown. Figure 1.30 depicts various kinds of UAVs or drones available in the market.

Figure 1.30: Multiple UAVs available in the market, source Shamshiri et al. 2018

In the literature being reviewed, the UAV data were differentiated on the sensor (or camera) used by the UAV. Studies like (Koch et al. 2017) have used a traditional RGB camera for analyses of wetland, whereas, for more refined analyses have been carried out using an infrared sensor (Knoth et al. 2013, Lehmann et al. 2016). The infrared sensor provides better information than traditional RGB, and is useful when analysing wetland vegetation. Another study by Díaz-Delgado et al. 2019 have used multispectral UAVs from detailed mapping of inundation levels, water turbidity, depth and aquatic plant cover. For species-level detailed analysis, studies like (Cole et al. 2014, Erudel et al. 2017) have applied hyperspectral sensor. The hyperspectral sensor provides very high spectral resolution in addition to the high spatial resolution from the UAV. More about UAV flight

analysis and integration of data is discussed in Chapter 4 and 5.

Finally, studies like (Szporak-Wasilewska et al. 2015, Langlois et al. 2017, Rahman et al. 2017, and Kohv et al. 2017) have used Light detection and ranging (LiDAR) for wetland analysis. The LiDAR data is collected in the form of point cloud, with multiple returns. Using these returns, many features, such as digital terrain model (DTM), digital surface model (DSM), digital elevation model (DEM), canopy height model (CHM), and 3D model can be calculated (Hamdan et al. 2015). Figure 1.31 shows the LiDAR point cloud and rasterised aerial view of park vegetation.
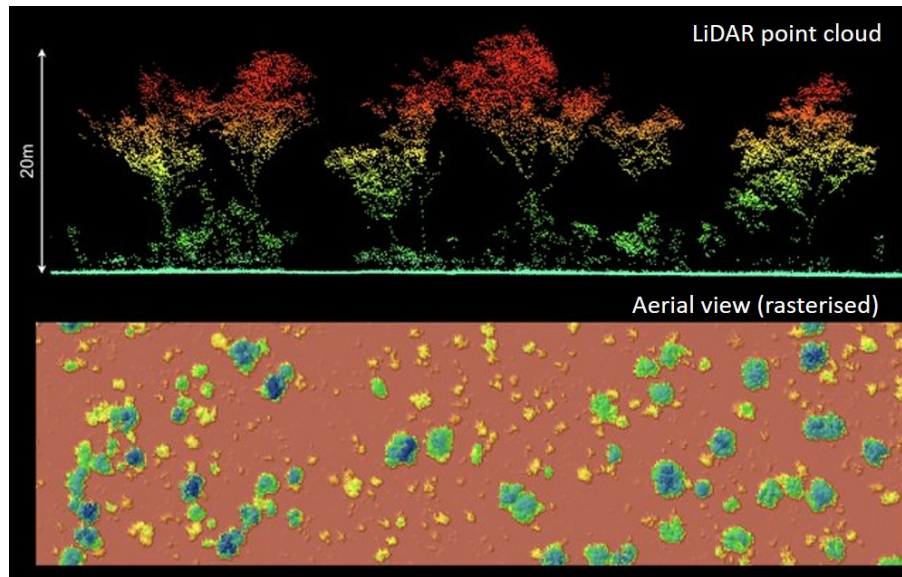
Figure 1.31: LiDAR point cloud (top) and aerial view of rasterised canopy height model interpolation, Territory Wildlife Park (Australia), source Levick et al. 2019

The 3D information provided by LiDAR is helpful in identifying many features which are spectrally similar and hence, are missed in the multispectral analysis. The acquisition of DEM used to be expensive for most environmental authorities and private users, but with LiDAR-derived DEM, the topography analysis has become more frequent and convenient (Crichton et al. 2015, Merchant et al. 2016, Koch et al. 2017, Montgomery et al. 2019). In the next subsection, a brief account of some types of applications applied in wetland RS is discussed.

1.5.4 Types of Applications

Wetland analysis can be divided into multiple objectives and applications. The main applications presented in the literature being reviewed are shown in Figure 1.32.
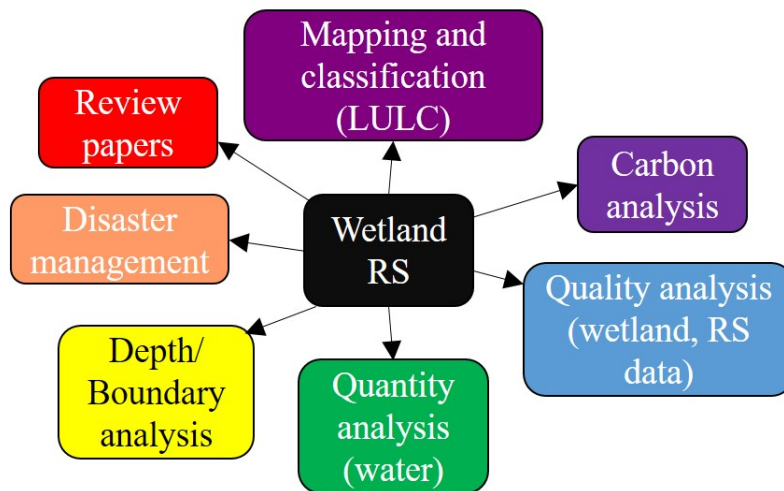
Figure 1.32: Types of applications in literature being reviewed

More than half of the literature focus on mapping various types of wetlands, wetland-vegetation/species, or land use land cover (LULC) analysis. This could be further divided into studies focusing on wetland mapping, which includes mapping vegetation and types of wetlands or peatlands (Dissanska et al. 2009, Beamish 2014, Reschke et al. 2012, Li, Xu & Zhao 2014, Bourgeau-Chavez et al. 2015). For example, (Merchant et al. 2016) have mapped flat bogs, channel fens, tree-dominated fens type peatlands. Similarly, there are studies which have mapped a bog complex into multiple classes like an uncut bog, acidic region, cutaway bog, virgin bog (McGovern et al. 2000). There are also studies which have concentrated on mapping a range of groundwater dependent ecosystems (GWDEs) (Gou et al. 2015). There are studies which have focussed more on specific vegetation communities of alkaline fen (Kopel et al. 2016), peatland vegetation (Erudel et al. 2017). Studies like (Montgomery et al. 2019, Berhane et al. 2018b) have presented wetland vegetation classification on the basis of aquatic and non-aquatic vegetation. Another study has identified tree-species (a total of 16 species) in a wetland (Sun et al. 2019). Apart from specific wetland mapping, many studies have concentrated on mapping broader land types, including forests (Amani et al. 2019), roads (Mahdianpari et al. 2018), uplands (White et al. 2017), croplands (Zhang et al. 2016, Mahdianpari et al. 2018, White et al. 2017), and barren land (Mahdianpari et al. 2018, Amani et al. 2019). Also, there are studies which have done a comparative analysis of the pixel-based and object-based classification techniques for wetland mapping (Hu et al. 2020). Furthermore, Wen & Hughes 2020 presents a comparison between various ensemble classification techniques such as Bagging, Boosting, and Stacking.

As stated earlier, wetlands are one of the significant sources of carbon emissions and so studies like Jaenicke et al. 2008 and Draper et al. 2014 have investigated the amount of carbon stored in the peatlands using RS techniques. Another study by Crichton et al. 2015, focused on understanding the carbon balance in the peatland ecosystem. Since, methane (CH4) is one of the major greenhouse gases released by wetlands, Lehmann et al. 2016 have investigated the change in methane fluxes over time in the peatlands.

Another application where wetland RS is used, is quality analysis. The data quality here means analyses of remote sensing data for understanding the healthy ground-condition of the wetland. A study by Regmi et al. 2012 talks about the after-effects of drainage

characterisation in the peatlands. Another study Díaz-Delgado et al. 2019 have diagnosed water turbidity, depth, aquatic plant cover to understand the ecological status of the wetland. Wetlands are often degraded due to natural or human-made events, and studies like Knoth et al. 2013 and Hu et al. 2020 talk about restoration and degradation status of the wetlands using remotely sensed data. Study like O'Connell et al. 2013 has also characterised the importance of radiometric corrections on Landsat data for the correct analysis of the wetland.

The amount of water present the wetlands, specifically peatlands, is critical and needs to be monitored. Constant flooding due to poor drainage can lead to loss of nutrients and change in the vegetation of the ecosystem. D'Acunha et al. 2018 has analysed the effects of rewetting a bog wetland. Other studies like (Kohv et al. 2017, Koch et al. 2017, Rahman et al. 2017) have used remote sensing data to monitor the water level in the peatland. Similarly, Weiss & Crabtree 2011 have used aerial images to derive the amount of surface water level present in the wetland. The depth of the peat fluctuates with the amount of water content. For example, sphagnum moss present in the raised bogs can soak water eight times as much as its weight which can lead to incremental changes in height and area of the community. This fluctuation is vital to analyse to understand the true extent of the active part of such a wetland. Langlois et al. 2017 has analysed bog boundaries in Canadian wetlands. Other studies Jones et al. 2012, Minasny et al. 2018 have examined the peat thickness using SAR data. Apart from this, other studies like Szporak-Wasilewska et al. 2015 have focused on generating a digital terrain model (DTM) of a river floodplain.

Studies like Li et al. 2019 and McCormack et al. 2020 have analysed the flooding levels in lakes and turloughs using optical and SAR data. Another study done by Touzi et al. 2015 has analysed the effect of fire on the peatland using SAR data. There have also been studies that have investigated tree species (Wang et al. 2019b, Chasmer et al. 2020), sensor type (Guo et al. 2017), and classification techniques (Mahdavi et al. 2018).

1.5.5 Types of features of sensors

The satellite sensors can capture information beyond just visible (RGB) wavelengths. The ability to capture information at infrared and microwave wavelengths have revolutionised remote sensing data analysis. Similarly, at UAV level, the data is no longer just limited to RGB images. Hyperspectral sensors, LiDAR point cloud have made much it possible to analyse a region with extreme detail. The features are either direct data from the satellite sensor or a derivative of it. Here, features represent both bands and indices used for wetland analysis. Figure 1.33 gives a breakdown of the features used in the literature being reviewed.
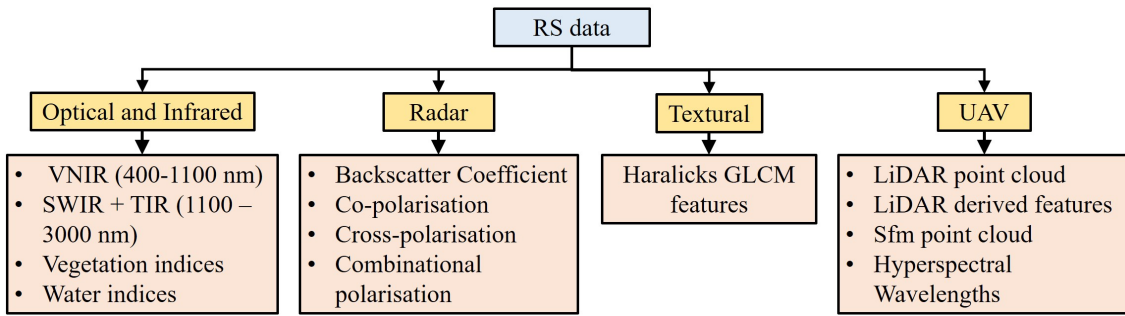
Figure 1.33: Types of features in literature being reviewed

1. Optical and Infrared features

The multispectral sensors acquire information in various spectral bands, including the visible and near-infrared, shortwave infrared, and thermal IR parts of the electromagnetic spectrum (Mahdavi et al. 2018). The wavelength region captured by optical and infrared sensors ranges from 400-3000nm, with $\approx 400-700nm$ in the visible range, $\approx 700-1100nm$ in NIR, $\approx 1100-1400nm$ lies in thermal IR (TIR), and $\approx 1400-3000nm$ in the SWIR. Each of these wavelength ranges is important in wetland analysis, as discussed in detail in this section. Figure 1.34 shows a classic spectral signature from wetland vegetation. Every
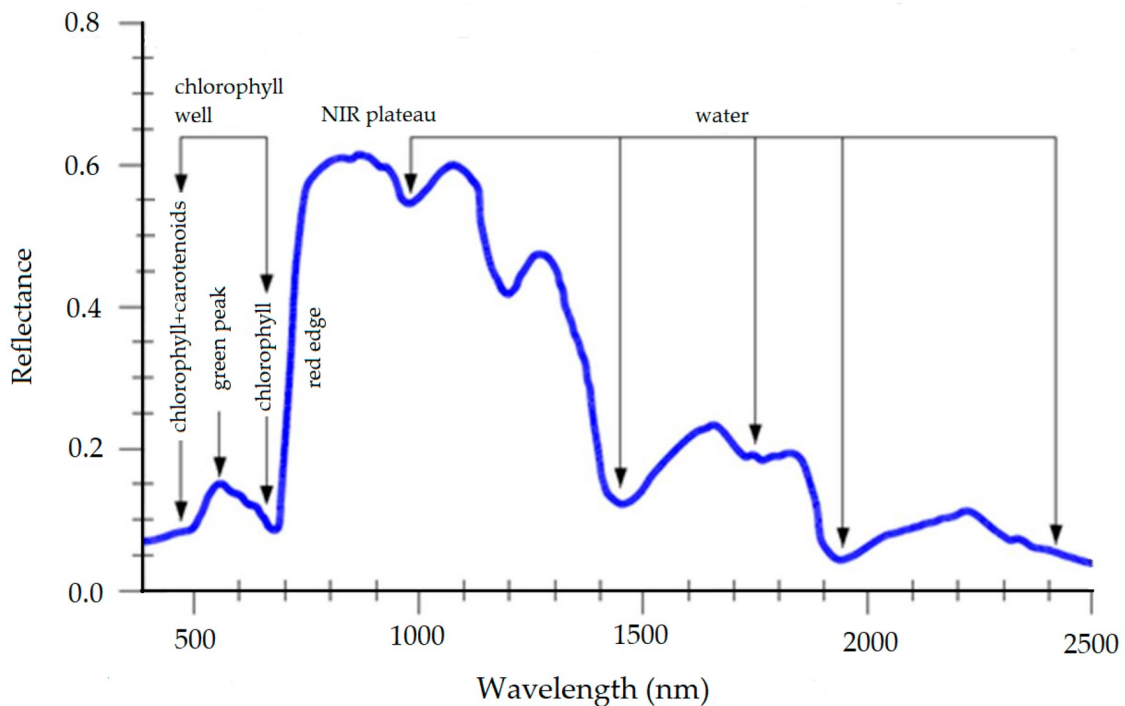
Figure 1.34: Reference spectral signature of vegetation, source (Moroni et al. 2019)

vegetation has a unique spectral signature, which follows the same pattern depending on its ground-condition as per the spectral signature shown in Figure 1.30. The absorption in the blue and red wavelengths is due to chlorophyll. Chlorophyll-a absorbs the light at violet-blue wavelengths, and chlorophyll b absorbs the light at blue-red wavelengths. Both of these reflects at the green wavelength. The rise at NIR wavelength depicts the health of the vegetation. The higher the difference between reflectance value at red and

NIR wavelengths, i.e., red-edge wavelengths, the better is the health of the vegetation (Hoffer & Johannsen 1969). The water absorption of the vegetation is depicted mainly in the SWIR region with dips at 1400nm, 1900nm, and 2400 nm. The water absorption region roughly starts after 1300nm (Hoffer & Johannsen 1969). The same vegetation but with different health and moisture conditions will have a different spectral signature. By the spectral signature curve, it is clear that there is a definite relationship between the wavelengths. To correctly identify the information out of these relationships, spectral indices were formed (Jackson 1983). The spectral index reveals many patterns of the surface which are hard to conceive using direct reflectance values. The spectral indices based on the region of spectra and application can be either focused on vegetation (VI) or water (WI). There are more than 500 indices present and recorded in the index database (IDB) (Henrich et al. 2009), and there is always a possibility of developing more indices, depending on the availability of data and application-type. Table 1.4 gives some examples of VIs and WIs present in the literature, with their equation. The BLUE, GREEN, RED, NIR and SWIR wavelengths are the same as mentioned in Figure 1.24.

Index name	Equation
⋆ Optical and Infrared	
Normalised difference vegetation index (NDVI)	$\frac{NIR-RED}{NIR+RED}$
Atmospherically corrected vegetation index (ARVI)	$\frac{NIR-RED-\gamma(RED-BLUE)}{NIR+RED-\gamma(RED-BLUE)}$
Chlorophyll vegetation index	$NIR(\frac{RED}{GREEN^2})$
Enhanced vegetation index	$2.5\frac{NIR-RED}{(NIR+6RED-7.5BLUE)+1}$
Soil adjusted vegetation index	$(\frac{NIR-RED}{NIR+RED+L})(1+L)$
Green NDVI	$\frac{NIR-GREEN}{NIR+GREEN}$
Normalised difference water index (NDWI)	$\frac{GREEN-NIR}{GREEN+NIR}$
NDWI 2	$\frac{NIR-SWIR}{NIR+SWIR}$
Modified NDWI	$\frac{GREEN-SWIR}{GREEN+SWIR}$
⋆ Hyperspectral	
Cellulose absorption index (CAI)	100(0.5(2030nm + 2210nm) - 2100nm)
Plant senescence reflective index (PSRI)	(678nm - 500nm) / 750nm
Vogelmann Index (VOG)	(734nm - 747nm) / (715nm + 726nm)
Red Edge Position (REP)	$700 + 40\frac{((670nm+780nm)/2)-750nm}{740nm-700nm}$

Table 1.4: Vegetation and water indices present in the literature being reviewed

In the literature, the Normalised difference vegetation index (NDVI) is extensively used vegetation index to analyse the health of the vegetation in wetlands (Li & Chen 2005, Jones et al. 2012). Other than NDVI, indices like enhanced vegetation index (EVI), tasselled cap vegetation, soil adjusted vegetation index (SAVI), evapotranspiration (ET) are also applied. O'Connell et al. 2013 and D'Acunha et al. 2018 have used a combination of NDVI,

SAVI, EVI, and ET along with NIR bands for bog wetland analysis. Due to the presence of water on the wetland surface, normalised difference water index (NDWI) has also been used by studies like Li et al. 2019. A combination of NDVI, NDWI, and EVI was used by (Gou et al. 2015) for mapping GDEs; the presence of three indices increases the ability of the algorithm to distinguish between visually similar land types. NDWI is also used for the identification of water and non-water areas by Weiss & Crabtree 2011. Another study by Balogun et al. 2020 has used a combination of 16 vegetation indices along with Landsat 8 bands for detection of oil spills in Johor, China Sea. The study states that chlorophyll vegetation index (CVI), modified difference water index (MDWI), NDVI, green chlorophyll index (GCI), green-NDVI (GNDVI), atmospheric resistance vegetation index (ARVI) and SAVI are appropriate for both vegetation and wetland impact assessment Balogun et al. 2020.

2. Radar features

As opposed to optical bands and indices, in radar RS, the features depend on the structure and radiation emitted from the surface. In this case, the polarisation of the SAR images plays a crucial role. In the active microwave (radar) RS, the sensor emits electromagnetic signals which then reaches the surface and is reflected. For example, in radar, the water surface acts as a reflector which scatters the signal away from the sensor. This makes the water area much darker than the non-water areas. Due to higher wavelength, and surface reflectance, the radar can also detect vegetation directly beneath the water to some extent (Martinis & Rieke 2015). The wavelengths and frequencies used in radar can be seen in Figure 1.35.
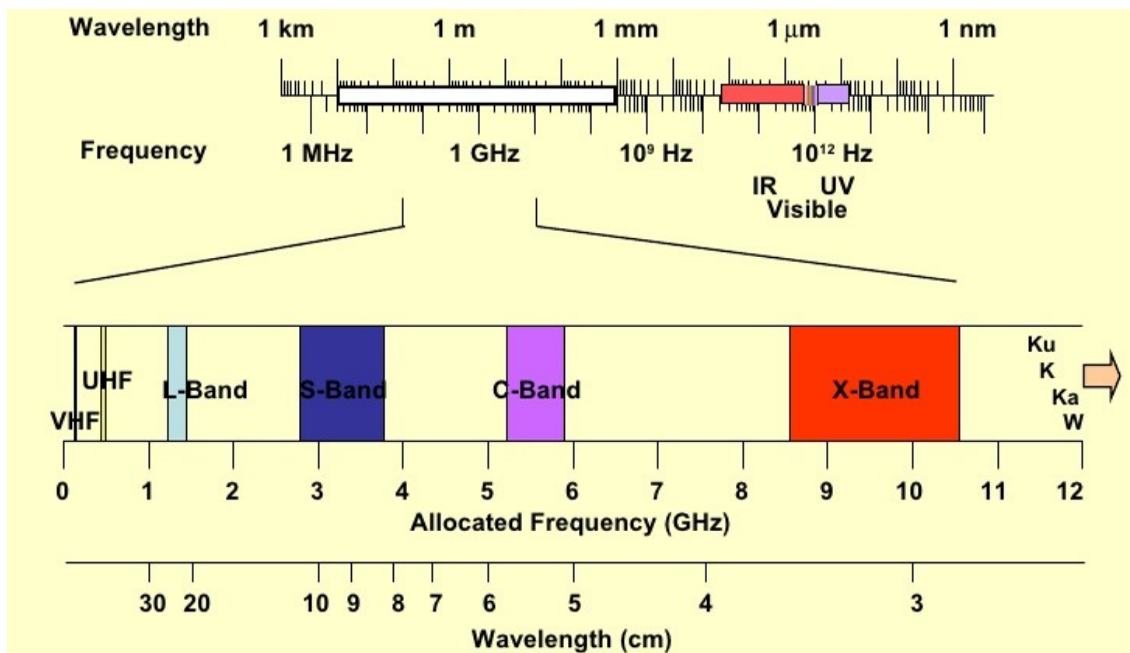
Figure 1.35: Wavelengths and frequencies in Radar

For vegetation and wetland analysis, C-band SAR (≈ 5.6 cm) is beneficial. Using C-band SAR, bushes, leaves and small branches can be identified (Hess & Melack 2003). Using X-band SAR ($\approx 3cm$), surface scatters from the top layer of the forest canopy is higher. Therefore, X band is useful for differentiating between flooded and non-flooded forests

(Martinis & Rieke 2015). Both X band and C band are also used to detect water under the marsh, i.e. the swamp type of vegetation (Martinez & Le Toan 2006). The L band (≈ 24.6), due to its longer wavelength and better penetration is used for detecting water under forest canopies (Martinez & Le Toan 2006).

The brightness of an object in the SAR image is related directly to its radar backscattering coefficient ($\sigma°$) (Hess & Melack 2003). The $\sigma°$ varies as a function of the surface roughness, the incidence angle, and the soil moisture (Hess & Melack 2003). The smooth surface reflects very little signal, and hence appear dark, whereas, urban areas, dense forests, etc. have higher reflectance, and they appear brighter in the radar image. The amount and type of signal transmitted and received by the sensor depend on its polarisation. Both horizontal and vertical polarisations are widely used in the literature for wetland analysis. They are HH (the signal transmitted and received horizontally); VV (the signal transmitted and received vertically); HV (the signal transmitted horizontally and received vertically); and VH (the signal transmitted vertically and received horizontally). A backscattering matrix is formed using these four polarisations (Mahdavi et al. 2018). Ouarzeddine et al. 2005 presents Cloude/pottier parameters like anisotropy, entropy, and incidence angle for SAR data analysis. Table 1.5 gives features that can be retrieved using the radar data along with the polarisations.

Feature	Description
Backscattering matrix	Gives basic decomposition of the scattered signal
Backscatter coefficient ($\sigma°$)	Gives information about the surface being imaged
	It has three parameters: polarisation, frequency and incidence angle
Polarisation	Gives information on various layers of the surface
Polarisation types:	
HH	details water bodies, used for inundation studies
HV or VH	details vegetation, fire, rough objects, etc.
VV	Similar to HH, more useful for detection of oil spills
$\frac{HH}{VV}$	Details the soil moisture information
Frequency	Penetration of the incident signal
	Details the relative roughness of the surface
Incidence Angle	Defines the swath of the sensor

Table 1.5: Radar features in the literature being reviewed

A study like (Touzi et al. 2009) demonstrates the use of C Band - HH polarisation with cloude-pottier parameters for wetland mapping in Canada. Another study (Bourgeau-Chavez et al. 2017) has used the L-Band SAR image with cloude-pottier parameters for analysing the effect of fire on wetlands. The study suggests that usage of L band reveals the healthy part of the bog covered by trees and bushes. (Bourgeau-Chavez et al. 2009) suggests that HH polarisation is useful for wetland delineation due to its better penetration through canopies (Mahdavi et al. 2018). Similarly, the co-polarisation (HH, VV) are useful in inundation studies (McCormack et al. 2020). The cross-polarisation (HV) has more sensitivity towards biomass (Bourgeau-Chavez et al. 2009) and therefore is used in (Draper et al. 2014) along with L band HH, and DEM for carbon measurements (either gas emissions or to quantify the biomass) from the wetlands in Amazon, Brazil. For land cover mapping, many studies have used a combination of both optical and radar data. Studies like (Bourgeau-Chavez et al. 2015) have used NDVI with HH and HV polarisation for peatland mapping, whereas, (Dingle Robertson et al. 2015) used all four polarisations, entropy, anisotropy, with NDVI and DEM for land cover mapping in Canada.

3. Textural features

When analysing various landforms, texture also plays an important role. Texture is an intrinsic property of virtually all surfaces, and it contains essential information about the structural arrangement of surfaces and their relationship to the neighbouring environment (Haralick et al. 1973). For example, an area with higher irregularities will have a coarser texture, whereas a smooth surface will be more homogeneous in texture. Haralick et al. 1973 states that each band either it is optical, infrared, or microwave can be treated as a gray-scale image and a gray level co-occurrence matrix (GLCM) can be derived for it. Figure 1.36 shows the texture-homogeneity in a delta located in southeastern Russia. It can be seen that homogeneity is higher for lake, streams, and textural information is higher in areas with vegetation.
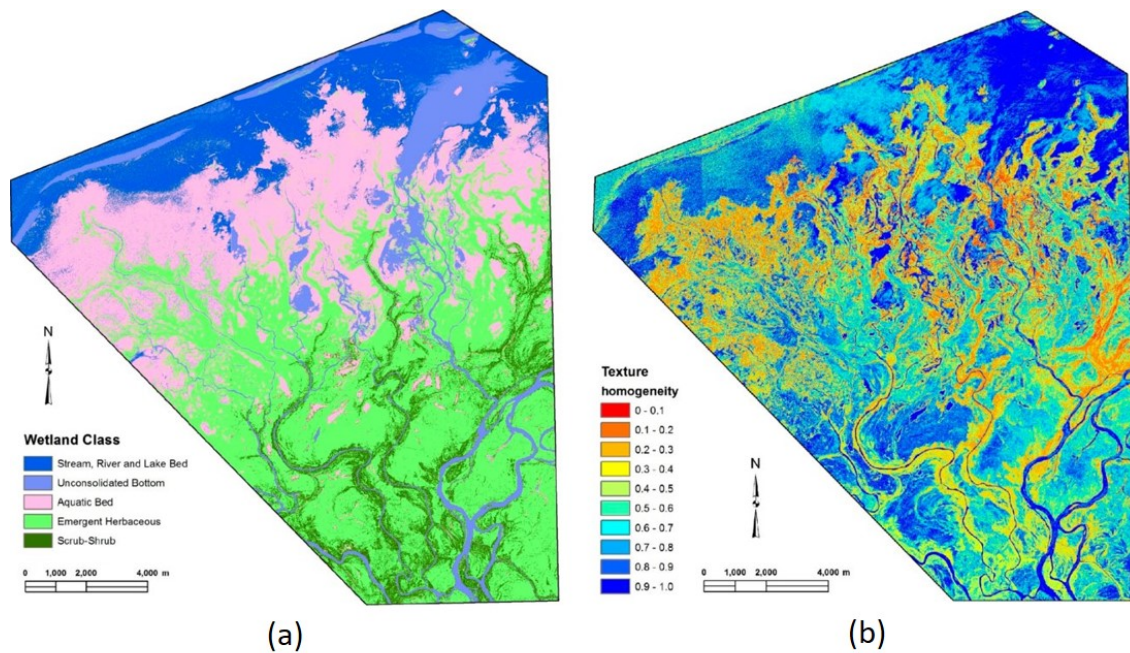
Figure 1.36: Selenga River Delta (a) Classified image (b) Homogeneity texture measure based on GLCM features, source Lane et al. 2014

In the literature, there are various studies which have taken advantage of the surface properties by using textural properties (Haralick et al. 1973). Grenier et al. 2008 proposed using a gray-level co-occurrence matrix (GLCM) as a method of quantifying the spatial relation of neighbouring pixels (Löfstedt et al. 2019). The most commonly used GLCM features in the literature are contrast, correlation, entropy, energy, homogeneity, and dissimilarity (Koch et al. 2017). Studies like Longépé et al. 2011 and Hu et al. 2020 have calculated textural features of Landsat images with NDVI for wetland mapping and change detection. Studies using high-resolution data have used textural features with NDVI, NDWI for mapping 22 wetland types (Berhane et al. 2018b). The main GLCM features are summarised in table 4.1 (Chapter 4).

4. UAV features

The UAV data is of a broad category; the multispectral camera has the same parameters as described in the optical and infrared section. Likewise, the textural features from the UAV are the same as described above. The main features specific to UAVs are:

- Light Detection And Ranging (LiDAR) point cloud
  The LiDAR data points are discrete by space and unorganised due to the nature of the geometry of the scanning device and the target interaction. Due to their non-grid representation and ill-defined boundaries, they are called a point cloud (Ramiya et al. 2017). The point cloud data from LiDAR is the representation of the surface with x (row), y (column) and z (return) parameters. This 3D information is useful when modelling the topography of a wetland (Jaenicke et al. 2008). The latest LiDAR sensors are able to acquire multiple return data at very high hit density. The multiple returns allow a detailed characterisation of the surface under consideration. The first return captures the topmost layer of the surface, mostly containing tree-canopies, urban settlements, etc. The later returns capture more detailed information such as tree-branches, stem, smaller vegetation under the top layer, etc. (Dalponte et al. 2009). The multiple returns can be utilised to form DSM. The last return is used to capture the DTM, that is the surface topography just about mean sea level (Hamdan et al. 2015). The DEM term is often used as an alternative to DSM and DTM, the return (second or third) can be decided to form LiDAR-derived DEM, which is widely used (Szporak-Wasilewska et al. 2015). Lastly, Canopy height model (CHM) is also a popular feature of LiDAR. CHM represents the height of the tree, i.e., the distance from the surface to the top layer of the tree. Therefore, it is not the elevation of the surface but the relative elevation of trees. Among others, studies like Zhu et al. 2019 have used it in inundation studies for mangrove wetlands. These various LiDAR-derived parameters are shown in Figure 1.37.
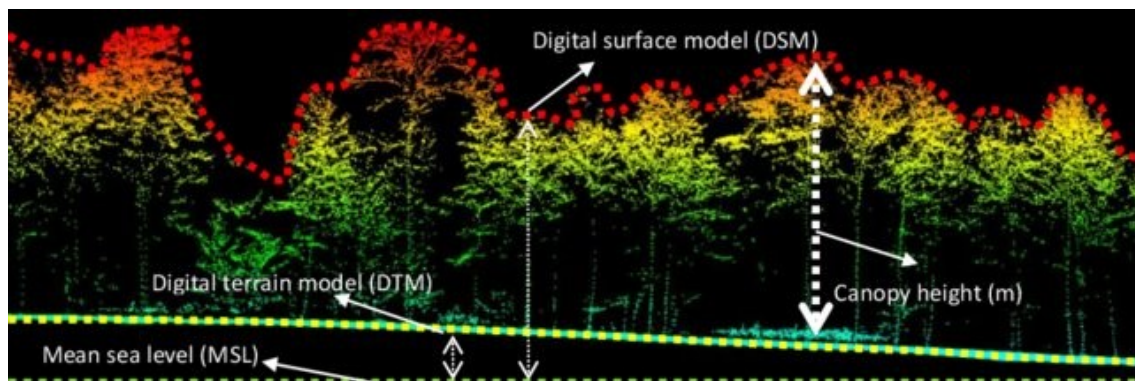
Figure 1.37: LiDAR point cloud distribution, Diptecorp forest, Malaysia Hamdan et al. 2015

- Structure from motion (sfm)
  Similar to LiDAR, 3D modelling can be done using a structure from motion (sfm) technique. The sfm uses stereo images obtained from photogrammetry and UAV. Figure 1.38 describes the basic principle behind the sfm process. The same point is captured multiple times, and when an image is captured from multiple angles, using feature mapping, the entire structure or scene can be recreated. It is particularly useful for 3D modelling in the absence of the expensive LiDAR sensor. Studies such as Kohv et al. 2017 have used SFM for water boundary mapping, for example.
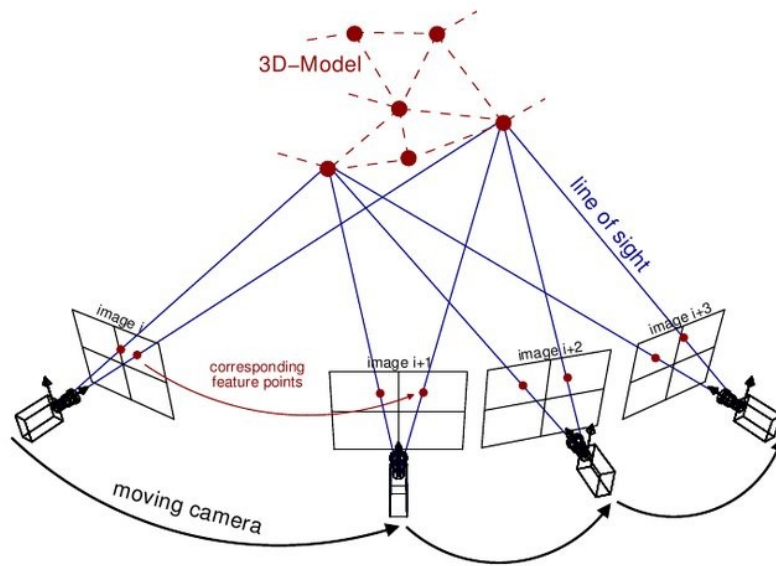
Figure 1.38: Structure from motion recreation of 3D model, source Van Riel 2016

- Hyperspectral sensor
  The hyperspectral sensor captures images not just at specific wavelengths but throughout the spectrum. Compared to a regular multi-spectral image, a hyperspectral image has a higher spectral resolution with a narrower bandwidth. This enables hyperspectral images to distinguish between spectrally close communities hence, improving the capability of multi-spectral image analysis. There are several hyperspectral satellites such as Hyperion, Prisma, etc. but in the literature being reviewed, the hyperspectral sensors are mainly used with UAVs, and hence, are discussed here. The hyperspectral images are captured as hyperspectral cubes containing continuous information of the surface (Figure 1.39). Due to this high spectral resolution, species-level mapping is also possible. Studies like Cole
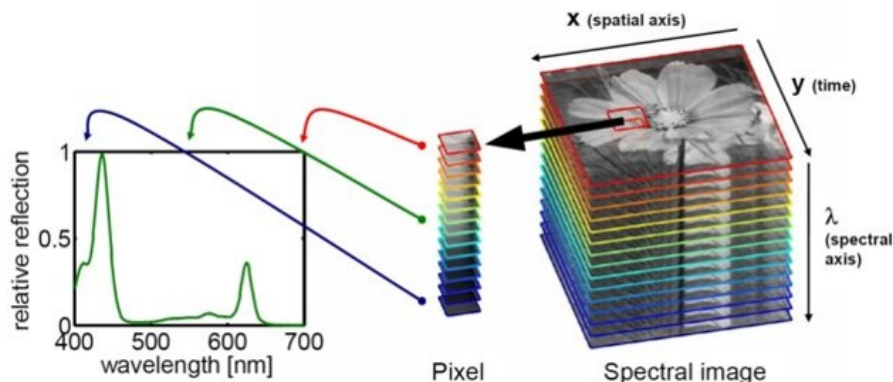
Figure 1.39: Hyperspectral image cube and wavelengths, source Polder et al. 2013

et al. 2014 have used Plant Senescing Reflection Index (PSRI), Photochemical Reflectance Index (PRI), Cellulose Absorption Index (CAI), Vogelmann Red Edge Index (VOG) and Red Edge Position (REP) for mapping vegetation cover in the wetland at cm scale.

1.5.6 Types of methods and algorithms

The methods applied for wetland analysis can be broken down into three main categories: regression analysis, classification analysis and statistical analysis. Out of the total of 60 studies reviewed, around 90% of them have used classification, followed by 10% regression and modelling. It should be noted that the studies under review mainly are from the last two decades, and with the advent of ML techniques – it has been highly applied. Figure 1.40 shows the categorisation of the methods being used in the literature.
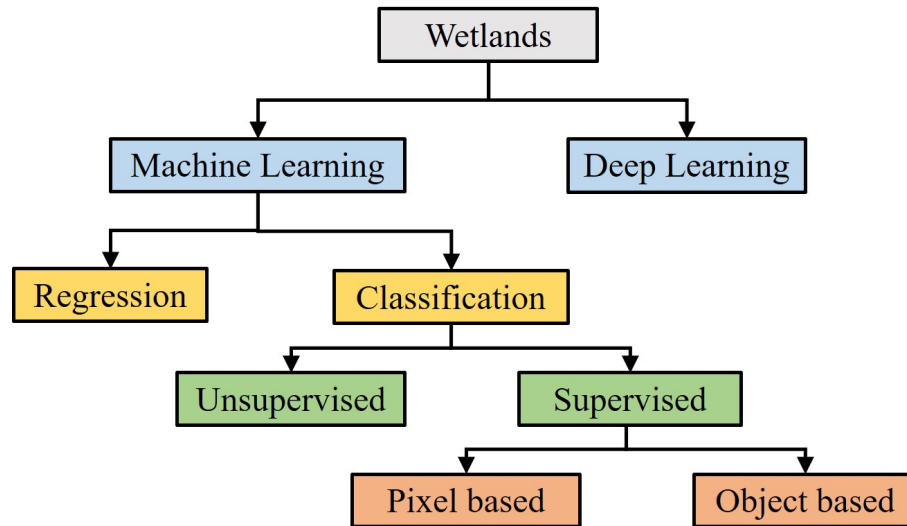
Figure 1.40: Algorithms and methods applied in literature being reviewed

Machine Learning

Machine learning can be defined as a set of algorithms, that can be applied on big data, to detect patterns or predict future trends, or to perform uncertain decision making (Murphy 2012). It can be divided into two major categories: Regression and Classification.

1. Regression Analysis
   The regression methods are used to set a relationship between two or more variables in the data. Usually, one of the variables is a dependent variable, and the other is the independent variable. Therefore, regression analysis is also applied when predicting trends in the data (Wu et al. 2006). There are multiple types of regression methods; some of them are discussed below (Figure 1.41).

   (a) Linear Regression (Montgomery et al. 2012)
   In linear regression, the relationship between x and y is linear, and it is assumed that y is continuous in nature. It can be described as

   $$y = \beta_1 x + \beta_2 \tag{1.5.1}$$

   where β is a constant. For multiple linear regression, there are multiple independent variables.

   (b) Logistic Regression (Hosmer Jr et al. 2013)
   In logistic regression, the probability of an event that is dependent on other variables is calculated. Therefore, the output always lies in between 0 and 1, and hence, logistic
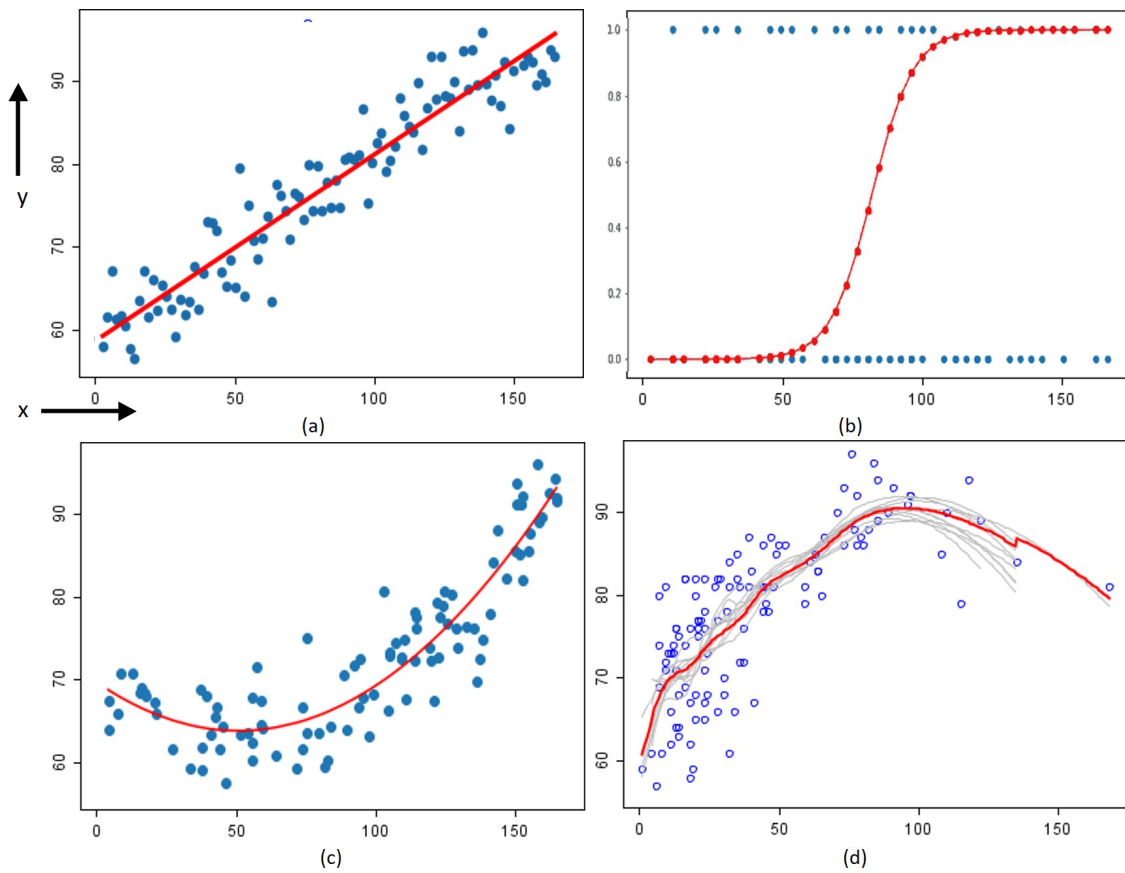
Figure 1.41: Examples of Regression on random dataset (a) Linear (b) Logistic (c) Polynomial (d) Bootstrapping

regression is also commonly used for classification. For n degree, it can be described as

$$y = \frac{1}{(1 + e^{-(\beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... + \beta_n x^n)})} \tag{1.5.2}$$

(c) Polynomial Regression (Breidt & Opsomer 2000)
   In polynomial regression, a non-linear relationship between x and y is formed using a polynomial combination of x and predicting y. For n degree, it can be described as

$$y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... + \beta_n x^n + k \tag{1.5.3}$$

(d) Bootstrap for Regression (Horowitz 1998)
   Bootstrapping is the process of dividing the data into multiple samples with repetition. For each of these sample spaces, a separate regression analysis is done, and the end result is given by combining (majority voting/averaging) the results of each sample spaces.

Studies like Regmi et al. 2012 have used linear regression to find a relationship between NDVI, backscatter (X band SAR), and drained thermokarst lake basins. It states that X-band SAR outperformed NDVI due to its ability to detect soil moisture. Linear modelling has also been used for classification (Díaz-Delgado et al. 2019). Another study like Zhang

et al. 2016 has used linear regression for hydrological response analysis between various wetland types. Regression has also been used for prediction of a plant's vascular health, Cabezas et al. 2016 have applied bootstrapping regression using random forest (RF) for analysis of the wetland plants.

2. Classification Analysis
For analysis of the health, and to monitor the wetland, classification techniques are also widely applied.

- Unsupervised Classification
In unsupervised methods, the user provides no labels, and the classification happens only based on the grouping of pixels having similar properties together, this is also known as clustering. There are two types of unsupervised algorithms used in the literature, Kmeans and Iterative Self-Organising Data Analysis Techniques (ISODATA).

(a) K-means (MacQueen 1967)
In K-means, the K stands for the number of clusters. Therefore, one of the key inputs is the number of clusters to be formed. This is the only information which is required, and the clustering process is entirely unsupervised. The data (pixels) are initially divided into random clusters. The distance (Euclidean/Mahanalobis/etc.) from every pixel to the centre of the cluster is calculated. The pixels are allocated to the cluster whose distance is minimum of all the cluster centres (figure 1.42). This is done iteratively until the desired number of clusters are formed. Studies like Gou et al. 2015 and Salinas et al. 2020 have used k-means for wetland classification.
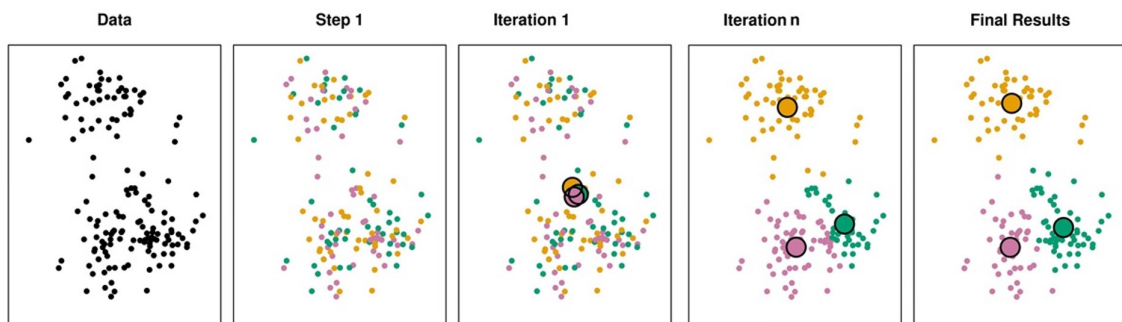
Figure 1.42: K-means clustering for 3 clusters, n iterations, source: *K-means clustering* 2018

(b) ISODATA (Ball & Hall 1965)
In ISODATA clustering, the initial division of data is done based on the distance between the pixels. Then the standard deviation of each cluster and the distance between each cluster centre is calculated. Based on the smaller distance, and lower deviation, the clusters are either merged or split. The merge/split criteria are based on the initial threshold given by the user as an input. For example, user's are asked to input a range of the number of clusters they expect, or range of exclusion distance, or closeness criteria. The process re-iterates until the threshold is reached, and the desired clusters are formed. Studies like Weiss & Crabtree 2011, Bourgeau-Chavez et al. 2001, Kitchingman 2017 have used ISODATA clustering for wetland patterns and communities analysis.

Figure 1.43 represents unsupervised classification done on a rice field in China. This shows

that there is not much difference in the two images, and for analysing the vegetation pattern, both work equally well (Zhong et al. 2011).

(a)   (b)   (c)
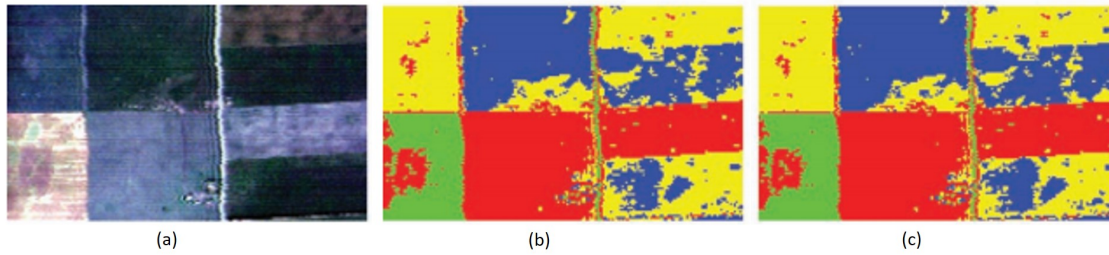
Figure 1.43: Unsupervised classification (4 clusters) of a rice field, China (a) Multispectral image (b) K-means (b) ISODATA; source (Zhong et al. 2011)

- Supervised Classification
  As opposed to unsupervised, in supervised classification, the user gives the initial input of training data and labels. This enables the algorithm to have a reference for the classification of the rest of the pixels. The supervised method is further divided into pixel-based (PBIA) and object-based image analysis (OBIA) or segmentation.

  – PBIA: Pixel-based image analysis methods analyse the value of each pixel in an image without considering the spatial or contextual information of the surrounding pixels (Mahdavi et al. 2018). Therefore, each pixel is treated as a separate entity, and the classification is based only on the features. These features are mentioned in section 1.6.5 Since no neighbourhood information is involved, the output is often pixelated and hard to interpret (figure 1.44).

  – OBIA: OBIA is the process of segmenting an image into spectrally and spatially homogeneous objects by incorporating the spectral, geometric, and other features of those objects for classification (Mahdavi et al. 2018, Benz et al. 2004, Salehi et al. 2012).

A detailed review presented by Mahdavi et al. 2018 states that there is no single classification algorithm that can be considered to be optimal for all applications and, thus, the desired algorithm should be selected based on the objectives and study area. Studies like Hu et al. 2020 and Dissanska et al. 2009 depict that OBIA is followed by classification. Therefore when it comes to classifiers, both PBIA and OBIA uses similar classifiers, with the main classifiers used in the literature discussed below.

(a) Maximum likelihood (Duda 2001)
Maximum Likelihood classification is based on Bayesian statistics. For its application, the data are assumed to have a normal distribution. The entire dataset (x) is divided into multiple numbers of classes. A probability distribution function (PDF) is defined for each class containing the data. The data (pixel/object) is then assigned the label with the highest value of the PDF (Duda 2001), as shown in Figure 1.45. The equation for the PDF is given as

$$PDF(x) = \frac{1}{\sqrt{(2\pi|\epsilon|)}} e^{\frac{-1(x-\mu)^T \epsilon^{-1}(x-\mu)}{2}} \tag{1.5.4}$$

where x is the divided data, μ is mean of x, and ϵ is the covariance matrix of x.
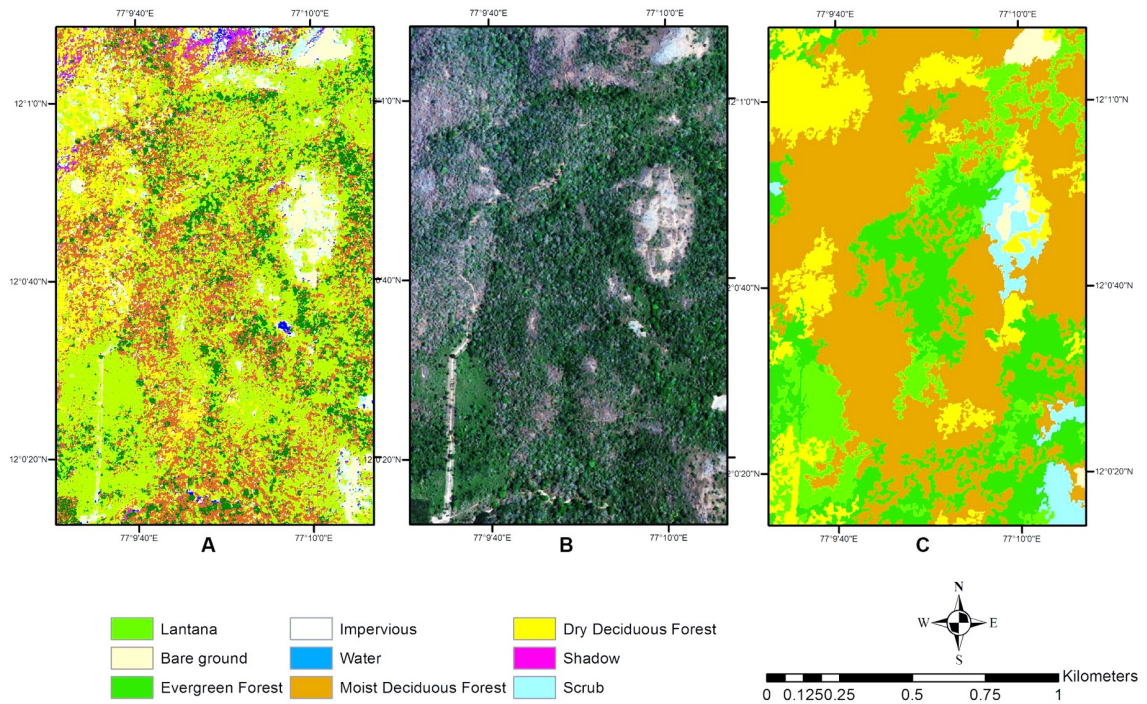
Figure 1.44: Classifications in Biligiri Rangaswamy Temple Tiger Reserve, India. (a) Pixel-based classified image (b) Geo-Eye multispectral image (c) Object-based classified image, source Niphadkar et al. 2017
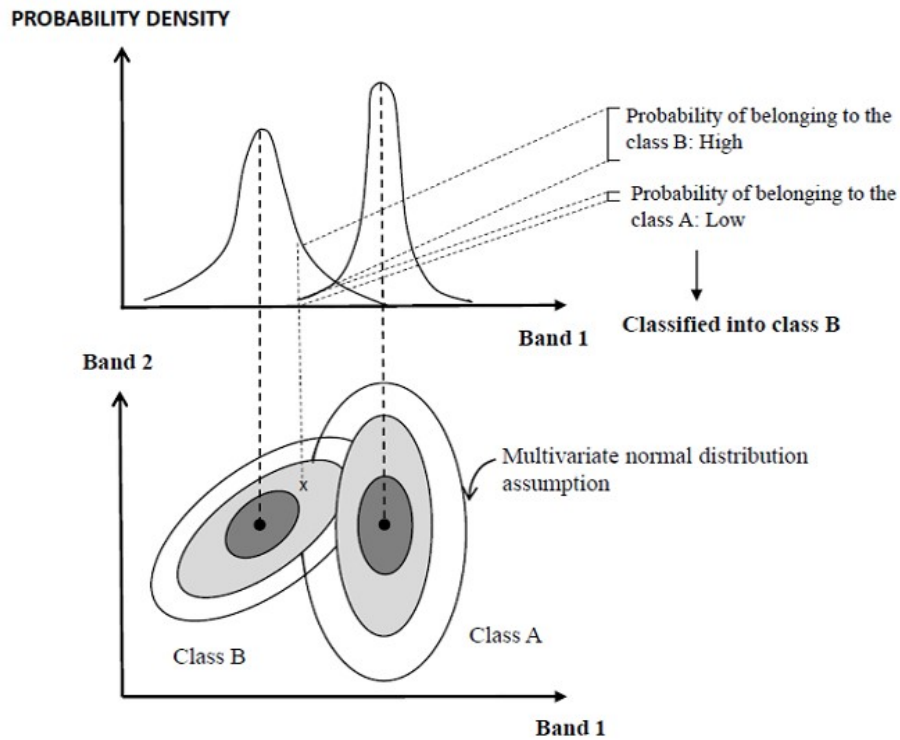
Figure 1.45: Maximum likelihood classification of data 'x' into classes (A, B), source(Valero Medina & Alzate Atehortúa 2019)

47

Many studies like McGovern et al. 2000 and Li, Xu & Zhao 2014 have used ML classifier for mapping peatlands. Equally, studies like Touzi et al. 2015 have used ML classifier to identify the damages in wetlands due to fire. Another study by Kohv et al. 2017 has used it for the monitoring water level in a bog with RGB and LiDAR data. The study states that for better identification of the water level, an additional NIR band is essential.

(b) K-Nearest Neighbour (KNN) (Laaksonen & Oja 1996)

The idea of kNN is simple; if the adjacent pixel/object belongs to a certain class, then the pixel (x) under consideration is most likely to belong to the same class as well. In the case of the classification, 'k' nearest neighbours are considered to decide the classification of x. It is based on the distance of unclassified pixels/objects from training data in a feature space. The most commonly used distance is Euclidean or Mahalanobis. Because the kNN algorithm decision is based on the surrounding adjacent pixels, it is also suitable for data with more intersection and overlapping. From Figure 1.46, shows that if the value of k is 3, then x belongs to class A, but if the value of k is 6, then x belongs to class B.
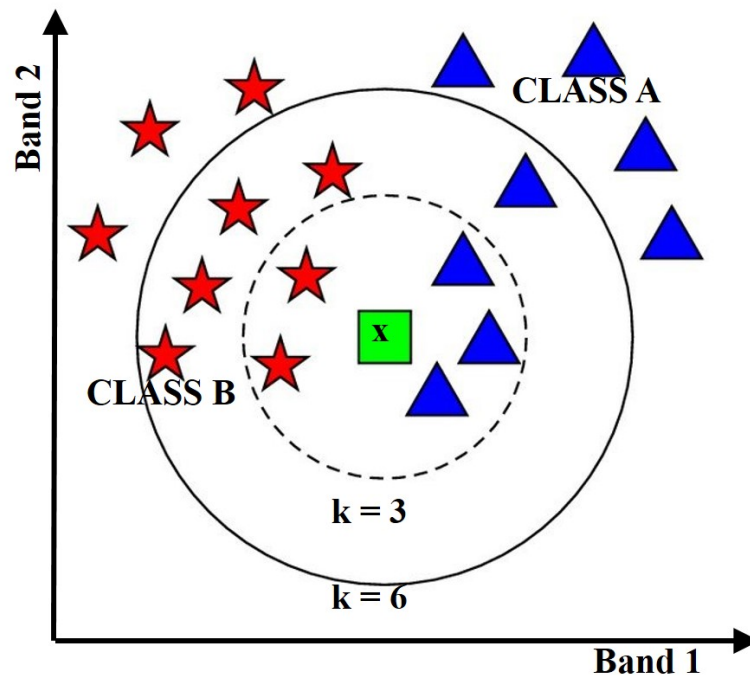
Figure 1.46: kNN classification of data 'x' into classes (A, B)

Hu et al. 2020 have applied KNN for wetland degradation analysis. Another study by Wang et al. 2019a has used KNN along with other classifiers for mapping coastal wetlands.

(c) Rule-based (Lawrence & Wright 2001)

Rule-based classification is based on predefined criteria set by the user. Every pixel/object is differentiated in various categories based on 'if-then' rules:

– if Band1(x) > 0.5 $\bigcup$ Band2(x) < 0.8, then x → Class A

– if Band1(x) < 0.5 $\bigcup$ Band2(x) > 0.8, then x → Class B

– if Band1(x) = 0.5 $\bigcup$ Band2(x) = 0.8, then x → unclassified

A study by Li & Chen 2005 used rule-based classification for mapping 5 types of wet-lands. Another study (Knoth et al. 2013) have used rule-based OBIA for determining the condition of the bog, and its restoration.

(d) Support Vector Machine (SVM) (Cortes & Vapnik 1995)
SVM is a non-parametric algorithm. SVM can perform a non-linear classification by implicitly mapping the data into high-dimensional feature spaces using kernels. The kernel in SVM is used to transform the data from input to the feature space. SVM defines a hyperplane to maximise the distance between the training samples of two classes, and then, classifies the other pixels/objects based on this hyperplane (Qian et al. 2015). The pixels/objects are mapped such that the data belonging to separate categories are divided by a proper gap that is as wide as possible, and this gap is called margin. Any new unclassified data is also mapped into the same space, and the class it belongs to is decided on the basis on the margin. The data points closer to the hyperplane are called Support vectors. Figure 1.47 describes the SVM hyperplanes and support vectors. SVM is applied both in regression and classification applications.
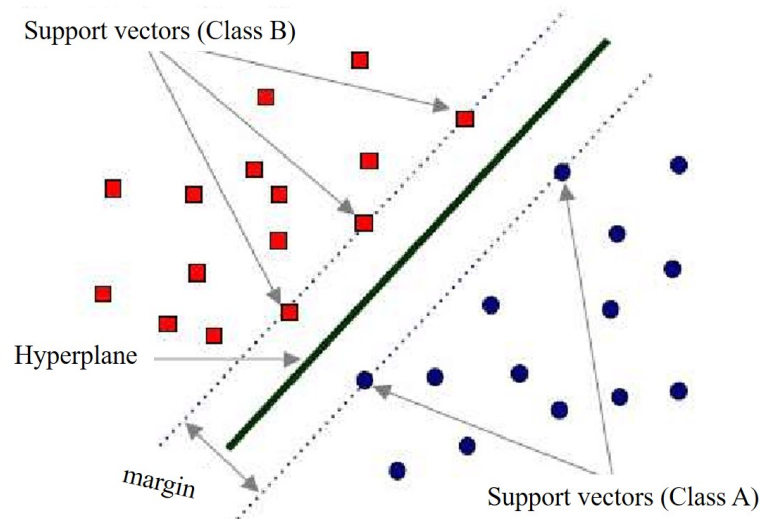
Figure 1.47: SVM classification of 'x' in classes (A, B)

Radial basis function (RBF) is one of the most commonly applied kernels in SVM. Apart from this, linear, non-linear, polynomial and sigmoid kernels can also be used depending on the data and nature of the application. One of the major advantages of SVM is that it is less sensitive to the amount of training data and can result in a higher classification accuracy given relatively small training data (Qian et al. 2015). Studies like Longépé et al. 2011 and Merchant et al. 2016 have used SVM for mapping peatlands. Other studies like Draper et al. 2014 and Koch et al. 2017 have applied it for carbon level emissions and water level estimations respectively.

(e) Decision Trees (DT) (Breiman et al. 1984)
Decision Trees belongs to the category of classification and regression trees (CART), i.e., DT can be used for both regression and classification. Every tree has nodes (branched) by which the input data (pixel/object) is divided into mutually exclusive groups based such that each group has the most homogeneous objects. The decision-making node is called the decision node, and the classified node is called the leaf node. The leaf node can

no further be divided, as it now belongs to a class. A DT may have multiple sub-trees depending on the type of data. There is multiple splitting criterion used in DT, such as Gini's criterion (discussed in Chapter 4). DT is fast, simple, and flexible; it is a valid classification technique for large and non-linear data (Figure 1.48).
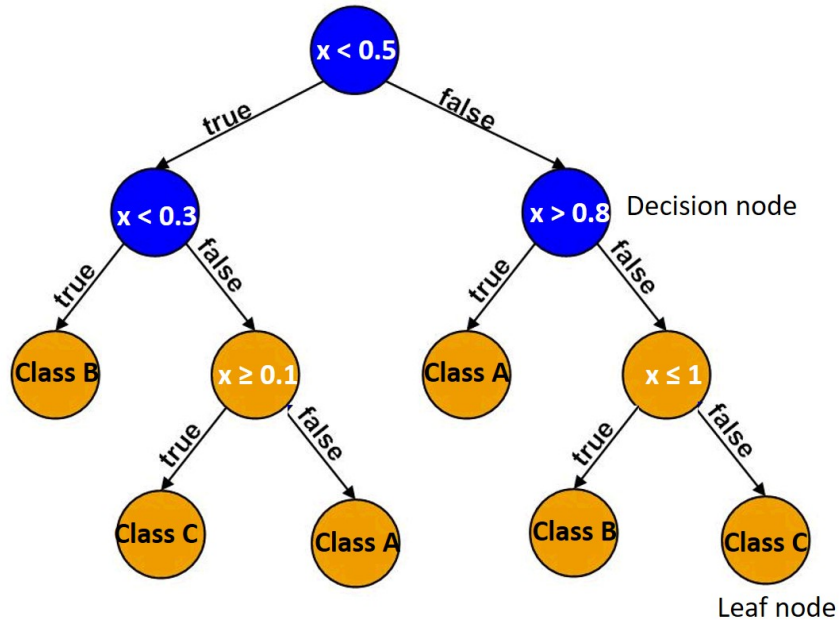
Figure 1.48: Decision tree classification of 'x' in classes (A, B, C)

A study by (Reschke et al. 2012) have used DT for assessing the capability of using C band SAR data for peatland mapping. Another study (Rahman et al. 2017) used DT for mapping groundwater table in peatlands. For mapping, a bigger wetland area, study like (Berhane et al. 2018b) has used SVM, DT and random forest (RF). The study suggests that the ensemble RF classifiers outperform SVM and DT for classification of wetlands.

(f) Ensemble Classifiers (Dietterich 2000)
Ensemble classifiers, as clear by the name, is a combination of multiple classifiers either applied in parallel or series. It is aimed at reducing variance. Ensemble classifiers can be divided based on their learning technique and combination, as shown in figure 1.49.
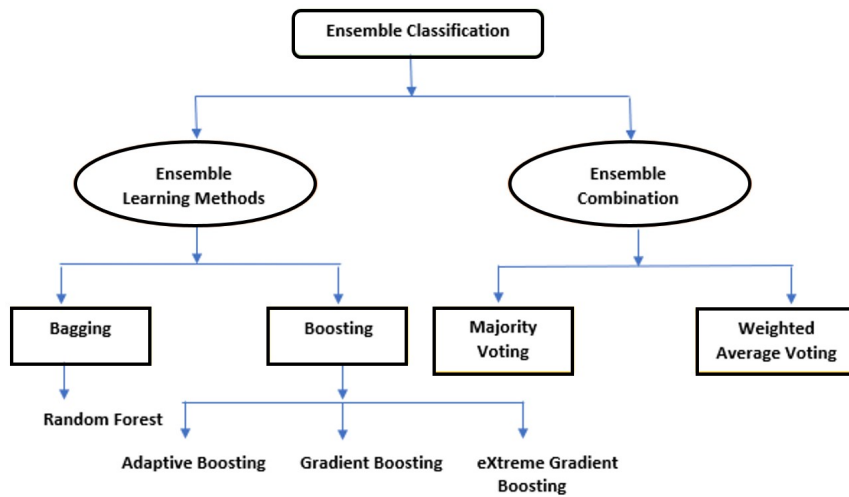
Figure 1.49: Types in Ensemble classification

– Bagging (Breiman 1996)
In Boostrap AGGregatING or Bagging the data is divided into subsamples (with or without replacement) using the technique 'bootstrapping'. Each of these subsamples is classified separately, and the final label is obtained using an ensemble combination method such as majority voting or averaging voting. Bagged Tree (BT) and Random Forest (RF) are examples of bagging. Both belong to the category of CART and are an extension of DT (Mahdavi et al. 2018). The only difference in them is that RF also randomly selects a subset of the classification variables while classification (Jozdani et al. 2019). A common structure of ensemble learning via bagging is shown in Figure 1.50.
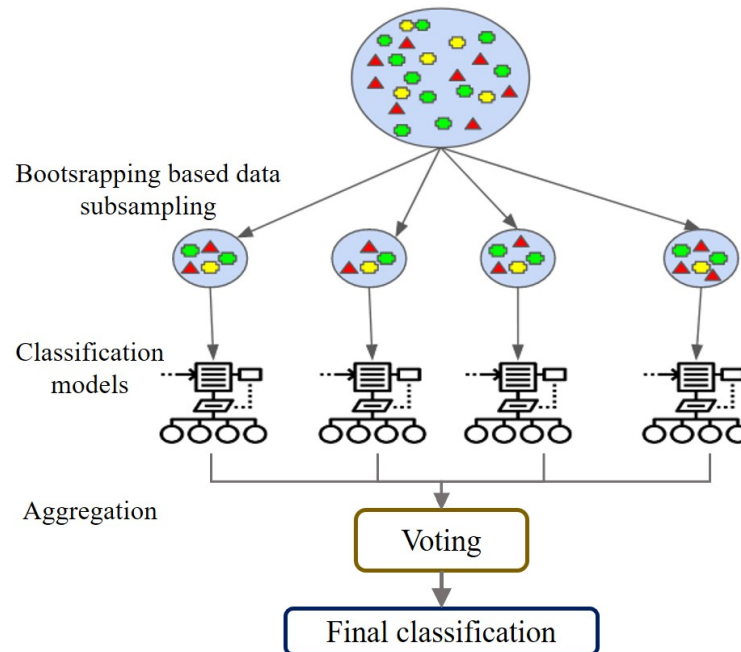
Figure 1.50: General workflow of classification using Bagging

– Boosting
In boosting, instead of growing in parallel, the ensemble learning is done sequentially. This means that each model is like an improved version of the previous model by adjusting parameters such as the weight of the classes in the data. It continues until the lowest error rate is achieved. Adaptive boosting (AdaBoost) is a generic type of boosting. AdaBoost was corrected on loss function; the model is known as Gradient Boosting (GB), which was further improved in computation by the recent eXtreme GB (XGB) (Jozdani et al. 2019). The general workflow of boosting is shown in Figure 1.51.
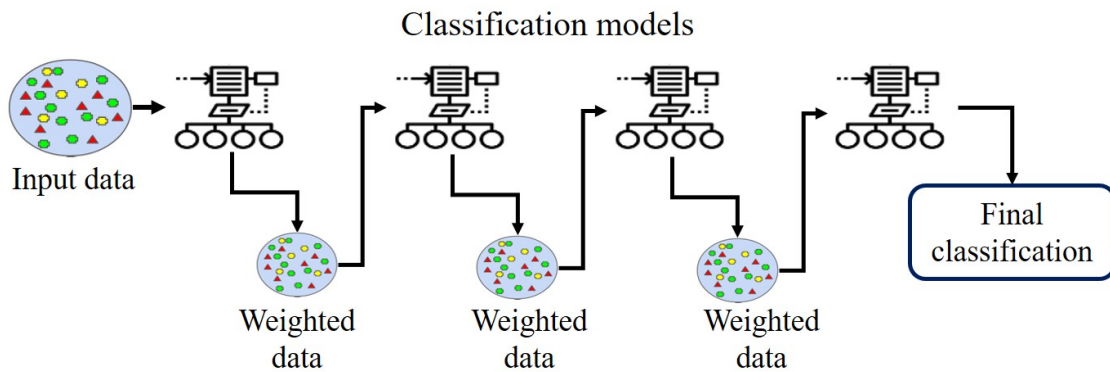
Figure 1.51: General workflow of classification using Boosting

– Majority Voting
In majority voting, the decision is made by taking the mode of all the decisions made by the classifiers.

– Weighted Average Voting
In average voting, the decision is made by taking the mean of all the decisions made by the classifiers. The decision is weighted proportionally to the performance of the model in the set of ensemble models.

The most popular ensemble classifier is Random forest (RF). It is also a non-parametric classifier, and therefore the distribution of the data makes no difference. It has been widely applied in mapping wetlands such as alkaline fen (Kopel et al. 2016), peatlands (Bourgeau-Chavez et al. 2015), lakes (Berhane et al. 2018a), aquatic wetland habitats (Berhane et al. 2018b), agricultural land and uplands (White et al. 2017), and multiple wetland types (Bourgeau-Chavez et al. 2017, Amani et al. 2017, Amani et al. 2019, Wang et al. 2019a). It is also applied for mapping thickness of the peat (Minasny et al. 2018), and polluted and non-polluted wetlands due to oil spillage (Balogun et al. 2020). Other than RF, (Wen & Hughes 2020) has used ensemble classifiers like the bagged tree (BT), boosted tree (BoT), stacked, etc.; a total of a combination of 17 ensemble classifiers. The study states classifiers with additional methods to increase ensemble diversity such as BT, RF show comparably high predictive power.

(g) Multi-resolution segmentation (MRS) (Comer & Delp 1995)
MRS is a segmentation method present mainly in Definiens Cognition Network Technology software 'eCognition'. It takes into consideration the parameters like image layer weight, scale, colour, shape, contrast and compactness for forming objects or segments from pixels.

It is an iterative algorithm, and segments are made until the threshold (provided by the user) is reached. Figure 1.52 shows transition of a satellite image into proper segments.

(a)   (b)   (c)
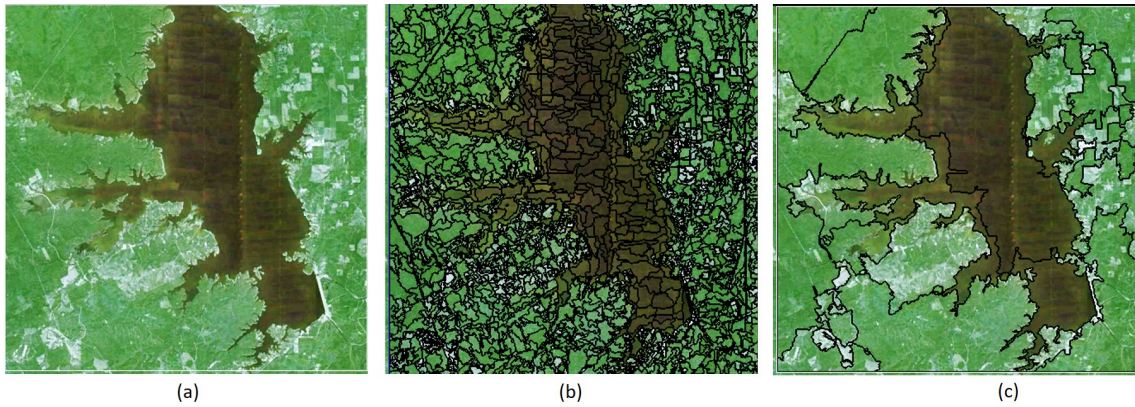
Figure 1.52: Multiresolution Segmentation using eCog (a) Satellite Imagery (b) Segments formed using watershed segmentation (c) Combined segments using multiresolution segmentation

Studies like (Grenier et al. 2008, Dissanska et al. 2009) have used it for mapping peatlands in Canada. It is also applied for the generation of DTM (Szporak-Wasilewska et al. 2015) and investigating gaseous fluxes (Lehmann et al. 2016).

Deep Learning

In the last decade, there has been a surge in the usage of deep learning (DL) techniques or artificial neural networks (ANN) in RS (Hoeser & Kuenzer 2020). DL can be defined as a sequence of fully connected layers, as more layers are stacked, the model becomes deeper, and more complex features can be learned. DL can be further divided into labelling and semantic segmentation. The labelling gives one label for the entire object ignoring the background and nearby pixels (Figure 1.53).
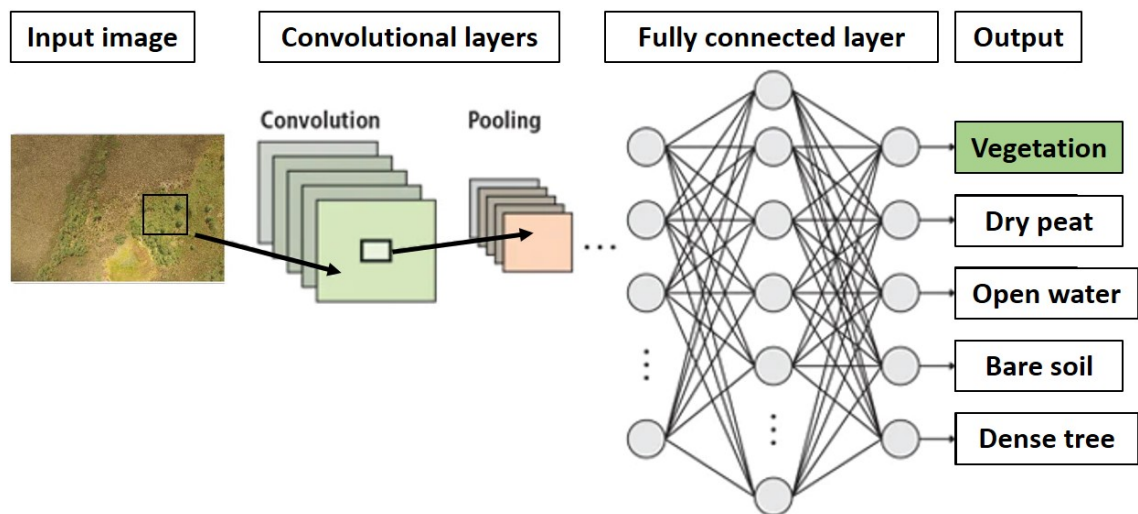
Figure 1.53: CNN based image labelling

Whereas, in semantic segmentation, every pixel has a label much like classification. The deep learning in literature being reviewed is applied using convolutional neural networks (CNNs). The various layers in CNNS such as convolution layer, pooling layer, etc. are discussed in detail in Chapter 4. CNNs have caused a leap in pattern recognition. Here each neuron is connected to only a local region of the input, making the network faster and less prone to overfitting for a large dataset. In addition, the same parameters are used in more than one place in CNN, making the model efficient statistically and computationally. In wetland mapping, studies like Mahdianpari et al. 2018 have compared multiple architectures like DenseNet121, InceptionV3, VGG16, VGG19, Xception, ResNet50, and InceptionResNetV2 with ML classifiers like SVM and RF. The study states that DL models outperform ML models. DeLancey et al. 2020 has compared various DL techniques for large wetland area mapping. The studies like Hoeser & Kuenzer 2020 give a full review of the DL architectures and their applications. Also, more details on DL can be found in Chapter 4 of this thesis. Hence, based on the literature reviewed, research gaps were identified and the scope of the thesis was framed, which is discussed in the next section.

1.6   Scope of the research

The novelty and scope of the research is as follows

- In the literature, it was seen that fewer studies have focused on the boundary delineation of wetlands. Given that peatlands typically have ill-defined boundaries, a boundary delineation algorithm was created using edge detection techniques.

- As can be seen from the literature, multiple open-source satellites like Landsat-8 and Sentinel-2 provides global data at a medium spatial resolution. In order to understand the ecological condition of the wetland, mapping of the vegetation communities was necessary. Therefore, a comparative analysis of optical data and ML classification algorithm was performed.

- The knowledge was extended from pixel-based to area-based using neighbourhood information and a graph cut maximum a-posteriori segmentation. It was applied along with the ensemble (bagged tree) pixel-based classifier. This combination of Bagged Tree + Graph cut was named as 'Mapping Vegetation Communities' (MVC) algorithm. It was used for temporal mapping of the vegetation in wetlands.

- Limited availability (due to clouds) and low-spatial resolution of satellite data is one of the major challenges for this type of research, especially in Ireland. Due to the limitation in resolution, small vegetation communities are often not mapped correctly. Therefore, the research was extended to evaluating the use of high-resolution UAV data. Due to the quantity of UAV data, it was possible to use DL techniques on it. Therefore, both ML and DL techniques were applied for segmentation of drone imagery.

- It is hard to achieve the spatial coverage provided by satellite imagery using a drone alone, mainly due to the limitation in battery life and other logistic issues. Therefore, a method to combine drone and satellite imagery was created, such that the best of both spatial and spectral imagery can be achieved. This method, named the 'nested drone-satellite approach', was applied on wetlands of Ireland but is a generalised procedure, applicable globally. It was followed by a change detection error-metric calculation, to detect changes in vegetation across seasons.

- From the literature, it was clear that the wetlands contribute to the emission of greenhouse gases (GHGs), and in particular methane (CH_4) which has a disproportional impact of global warming compared to carbon dioxide. However, no studies carried out so far appear to have explored the possibility of mapping wetlands using total column CH_4 data provided by the satellite. Therefore, to bridge this gap, mapping using S-5p data was performed.

- It is speculated that years of flooding and change in climatic patterns can lead to change in the distribution of wetland vegetation communities. To verify this, a statistical study was performed on hydrological data of ephemerally flooded turloughs located in the west of Ireland in order to evaluate the ecohydrological conditions that different vegetation communities require.

1.7 Study areas (wetlands)

As part of this study, a total of 16 wetlands (five bogs, three fens, eight turloughs) were analysed. This section provides the descriptions of the main wetlands used to create the algorithms (discussed in the following chapters).

- Clara bog
  Clara bog, County Offaly, is one of the largest of the remaining active raised bogs in Ireland. It spreads to over 250 ha in area and provides unique habitats for various vegetation species. Based on ecotopes, the bog is classified as Central, Subcentral, Submarginal, Marginal, and Active/Inactive Flush.
  The active accumulating peat conditions occur within the Central and Subcentral ecotopes, which are generally located at the centre of the bog, with central usually surrounded by the subcentral ecotope (Schaaf & Streefkerk 2002). Active flush and soaks, which are areas of focused surface water flow, typically have very wet conditions and are dominated by bog mosses of the genus Sphagnum (Mackin et al. 2017, Schaaf & Streefkerk 2002). Conversely, marginal and submarginal ecotopes, along with inactive flushes, are characterised by poor quality vegetation (Schaaf & Streefkerk 2002). However, the marginal bog vegetation, though not actively growing, is an integral part of the whole bog complex. The Clara bog was used mainly for developing the mapping algorithms for bogs. Other bogs used for testing the algorithms had similar properties as Clara bog, which are listed below.

  - Mongan bog, County Offaly

  - Monivea bog, County Galway

  - Killyconny bog, County Meath

  - Knockacollar bog, County Laois

- Scragh bog (Fen)
  In Scragh bog (fen), the predominant vegetation community is Rich Fen and Transition Mire. The Rich fen communities are *Carex approprinquata*, with a range of tall herbs such as *Cirsium palustre, C. heterophyllum, Trotlius europaeus*, and *Sanguisorba officinalis* (Adam et al. 1975). The peat substrate found here is generally alkaline. The transition mires are associated with open waters and quaking bogs. This community reflects the actual succession from fen to bog (Kimberley & Coxon

2013). This fen also consists of non-peat forming marsh communities which are generally quite nutrient-rich. The Scragh bog fen, located in County Westmeath, was used mainly for developing the mapping algorithms for fens. Other fens used for testing the algorithms had similar properties as Scragh fen, which are listed below.

- Toryhill fen, County Limerick

- Ballymore fen, County Westmeath

- Blackrock Turlough
Blackrock turlough located in County Galway consists of a variety of vegetation. There are mainly 12 vegetation communities which have been agglomerated into four broad communities of the vegetation species by Waldren et al. 2015 as follows. The first broad community is formed by grouping the *Poa annua – Plantago major* community and *Eleocharis acucularis* community. The second broad community consists of *Carex nigra – Ranunculus flammula* communities and *Agrostis stolonifera – Glyceria fluitans* communities. The third broad community consist of *Agrostis stolonifera – Ranunculus repens* communities, *Agrostis stolonifera – Potentilla anserine – Festuca rubra* community, *Potentilla anserine – Potentilla reptans* communities, and *Fillipendula ulmaria – Potentilla erecta – Viola sp.* community. *Lolium grassland* falls is classified as another community (broad community 4) in this study. Communities such as woodland, scrubs, etc. are not a part of the broad communities and are analysed separately. Out of the four broad communities, community 4, which contains dry grassland community is the driest group whereas, and community 2 is the wettest. Identification of species under these communities depicts the amount of wetness, fertility and stress tolerance of the turlough.The Blackrock turlough was used mainly for developing the mapping and hydrological monitoring algorithms for turloughs. Other turloughs used for testing the algorithms had similar properties as Blackrock turlough, which are listed below.

- Knockaunroe turlough, County Clare

- Turloughmore turlough, County Clare

- Roo West turlough, County Galway

- Lough Aleennaun, County Clare

- Coy turlough, County Galway

- Caherglassaun turlough, County Galway

- Garryland turlough, County Galway

1.8   Organisation of the thesis

The main work of the thesis is presented in six chapters following this chapter.

- Chapter 2: Monitoring environmental supporting conditions of a raised bog using remote sensing techniques.
In order to monitor the ecological condition of the raised bog, it was important to analyse the correct boundary of the bog and also, to map vegetation in it. In **Chapter 2**, a bog boundary delineation algorithm using edge-detection techniques, was introduced. This was followed by a comparative analysis of the open-source

satellite data from Landsat-8 and Sentinel-2 for classification of the raised bog. Due to better spatial resolution, and the presence of 3 new red-edge bands, Sentinel-2 was deemed to be better for mapping wetlands in Ireland. This was done using ML pixel-based algorithms, SVM, Subspace KNN, and Bagged Tree (BT). It was seen that for wetland mapping, the ensemble BT gave the best results.

- Chapter 3: Mapping Vegetation Communities inside Wetlands Using Sentinel-2 Imagery in Ireland.
  From Chapter 2, the choice of satellite data and the classification algorithm was clear. However, to map the vegetation communities in all the wetlands present across Ireland, spectral analysis on its own was not sufficient. To make the methodology more robust, a new mapping vegetation community (MVC) algorithm was tailored using the combination of existing BT classifier and graph cut (maximum-a-posteriori) algorithm. This segmentation algorithm uses the contextual information along with the spectral information of the pixels. Therefore, in **Chapter 3**, an extension of boundary delineation algorithm to other wetland types is presented. This was followed by a successful implementation of MVC algorithm for mapping vegetation across all three wetland types (bogs, fens and turloughs) over the time period of 2 years, 2017 and 2018.

- Chapter 4: Drone Image Segmentation Using Machine and Deep Learning for Mapping Raised Bog Vegetation Communities.
  The optical data provided from open-source satellite often is limited due to the presence of clouds. In a temperate climate like Ireland, this can pose a considerable challenge. Also, the spatial resolution of satellite often leads to pixel-mixing and eventual misclassification between communities. Therefore, to avoid this – high-resolution drone imagery was used (DJI Inspire 1). The drone data can be collected at the user's convenience and provides very detailed (spatial resolution = 1.7 cm) information. This data is applied for mapping vegetation communities in wetlands. **Chapter 4** presents a detailed comparison between ML and DL techniques using drone imagery of a raised bog. Both ML and DL techniques are also discussed in detail, and it was seen that DL outperforms ML methods in classification accuracy by 4% but is both time and resource-intensive. Keeping in mind the bigger picture, ML was deemed to be a better choice for mapping vegetation in unsurveyed wetlands.

- Chapter 5: A Nested Drone-Satellite Approach to Monitoring the Ecological Conditions of Wetlands.
  In Chapter 5, it was seen that the DJI drone has limitations linked to its battery life, and therefore only a limited area can be captured at a time. On the other hand, Sentinel-2 captures images globally, but in low-spatial resolution. Whereas, the drone has a very high spatial resolution. In order to get the best of areal coverage and spatial resolution, **Chapter 5** presents a detailed and cost-effective methodology to use drone and satellite together.
  Along with the combining methodology, a new error metric to analyse the minute change that occurs due in the ground in a time-series framework was formulated. The error metric includes information about similarity, area, orientation and extent of the vegetation communities. The study successfully implements the combination of drone and satellite imagery for producing seasonal and annual maps of the wetland.

- Chapter 6: Exploring the capability of Sentinel-5P data for land-type classification.

Analysing gaseous information was necessary for a complete study of wetlands using remote sensing techniques. For this, in **Chapter 6**, we have used the recently launched Sentinel-5 Precursor (Sentinel-5P) methane product. The S-5p methane product is highly sensitive to clouds, and therefore, the study of wetlands was performed for Canada instead of Ireland. Along with methane, S-5p's by-product surface albedo (SA) was also used for analysing the land type. The study successfully indicates that the total column methane and land-type are in a quasi-equilibrium state and, that the new S-5p data with SA can be beneficial in picking sensitivities in land types across the globe.

- Chapter 7: Ecohydrological metrics for vegetation communities in turloughs.
  Initially, in **Chapter 7** a data-analysis approach of analysing the effect of hydrological data on the wetland, in specific turlough has been made. The turloughs selected for this study have a similar water-quality index and a similar vegetation pattern. The turloughs were last surveyed in 2008, but the hydrological data on water level has been collected for the last 28 years. Due to the presence of open-source, mid-resolution Sentinel-2 data, it was possible to have a satellite survey of the turloughs. The trends in hydrology were compared statistically based on metric-parameters like flooding depth (in meters), flood amount (in months per year), global radiation and temperature of the turlough. Based on the metric, clustering of the years was done using 28 years data from 1989 – 2017. A similar analysis was done for the years 2008-2017 using the satellite surveyed map, and the trends were compared. It was seen that the vegetation communities remain intact, and the metric- parameters are not changed significantly in the last ten years. Such a metric indicates a clear relationship between the vegetation communities present across the four turloughs.
  Lastly, **Chapter 8** presents the conclusions of the thesis. A summary of the works presented along with the critical assessment of the work is presented. Some recommendations about future work using the methodologies developed in the thesis are suggested in this chapter.
  Apart from the main thesis, all the extra information is provided in the Appendix.

1.9 List of publications

1. Bhatnagar, S., Ghosh, B., Regan, S., Naughton, O., Johnston, P., Gill, L. (2018). Monitoring environmental supporting conditions of a raised bog using remote sensing techniques. Proceedings of the International Association of Hydrological Sciences, 380, 9-15. DOI: `https://doi.org/10.5194/piahs-380-9-2018`

2. Bhatnagar, S., Gill, L., Regan, S., Naughton, O., Johnston, P., Waldren, S., Ghosh, B. (2020). Mapping vegetation communities inside wetlands using Sentinel-2 imagery in Ireland. International Journal of Applied Earth Observation and Geoinformation, 88, 102083. DOI: `https://doi.org/10.1016/j.jag.2020.102083`

3. Bhatnagar, S., Gill, L., Ghosh, B. (2020). Drone Image Segmentation Using Machine and Deep Learning for Mapping Raised Bog Vegetation Communities. Remote Sensing, 12(16), 2602. DOI: `https://doi.org/10.3390/rs12162602`

4. Bhatnagar, S., Gill, L., Regan, S., Waldren, S., and Ghosh, B. (2020). A nested drone-satellite approach to monitoring the ecological conditions ofwetlands. ISPRS Journal of Photogrammetry and Remote Sensing. [UNDER REVIEW - REVI-

SIONS]

5. Bhatnagar, S., Sha, M.K., Gill, L., Langerock, B., and Ghosh, B.; Sensitivity of methane emission to variations in land types detected from space. Nature Geoscience. [SUBMITTED]

6. Bhatnagar S., Ghosh B., Waldren S., Coxon C., Johnston P., Sharkey N., Naughton O., Morrissey P., Gill L.; Ecohydrological metrics for vegetation communities in turloughs. Hydrological Sciences Journal. [SUBMITTED]

7. Bhatnagar, S., Ghosh, B., Regan, S., Naughton, O., Johnston, P., Gill, L. (2019, January). Remote Sensing Based Ecotope Mapping and Transfer of Knowledge in Raised Bogs. In Geophysical Research Abstracts (Vol. 21).

8. Bijkerk, E., Bhatnagar, S., Coxon, C., Johnston, P., Regan, S., Waldren, S., Gill, L. (2020, May). Investigating hydrological and biogeochemical controls on Irish alkaline fen habitats. In EGU General Assembly Conference Abstracts (p. 19765).

9. Bhatnagar, S. et al., Trinity College, University of Dublin, Ireland (Sep., 2018). Monitoring environmental supporting conditions of a raised bog using remote sensing techniques. European Space Agency Land Remote Sensing. ONLINE [POSTER]

10. Bhatnagar S., et al., Trinity College, University of Dublin, Ireland (Dec., 2018). Mapping a raised bog-wetland using OBIA. Irish Earth Observation Symposium.[POSTER]

Chapter 2

Mapping Vegetation Communities - Pixel based approach

"A person often meets his destiny on the road he took to avoid it. "

— Master Oogway, *Kung Fu Panda-1*

Contents

This chapter corresponds to the paper published on 18 December 2018, in "Proceedings of International Association of Hydrological Sciences" [1].

2.1   Introduction

A bog is a type of wetland which primarily depends on rainfall for water and nutrients. Bogs can be categorised as blanket bog and raised bog. Raised bogs are discrete, raised, dome-shaped masses of peat occupying former lakes or shallow depressions in the landscape (Fossitt 2000). They occur throughout the midlands of Ireland (Felicity Hayes-McCoy 2017), and in this study, we focus on one of the most extensive raised bogs in Ireland, Clara Bog, Co. Offaly. Monitoring wetland structure and function typically requires recurrent site visits, which can be prohibitively labour intensive, costly and time-consuming. Given the current need for up-to-date information, as well as the widespread

coverage of wetland maps, satellite RS has been demonstrated to be the most efficient and cost-effective method for this purpose.

The current state of art primarily focuses on different mapping types of wetlands (Mahdavi et al. 2018). The mapping is done using different wavelengths and spectral response of the objects. Vegetation has a higher response in near-infrared (NIR) region, and this study seeks to map the within-site distribution of key ecological communities present in the bog habitat. Satellite imagery-derived vegetation indices can be effectively used for assessing the vegetation status of an ecosystem. Vegetation communities present within an ecosystem are defined as ecotopes. The Normalised Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Atmospherically Resistant Vegetation Index (ARVI) are the most effective vegetation indices stated in the literature (L. Wiegand et al. 1991). Soil moisture is not directly derivable from the optical bands. The Normalized Difference Water Index (NDWI) using near-infrared (NIR) and short-wave infrared (SWIR) bands give an indication of wetness of the surface inferred as soil moisture. These indices can be used to provide a clearer picture of vegetation and water extent in an area.

Topography plays a vital role in analysing an ecosystem. It gives an accurate idea of elevation difference present between various plant communities. Light detection and ranging (LiDAR) provides with point cloud information which can be used to deduce the topography of an area. In this study, an airborne LiDAR-derived DEM is used for analysing the topographic extent of the bog.

For analysis of the RS data, various machine learning tools have proven to be useful (Lu & Weng 2007). There are many state-of-the-art segmentation and classification algorithms available. It is necessary to make full use of the advantages of different algorithms on the basis of multi-feature fusion, so as to achieve better segmentation effect (Yuheng & Hao 2017). Hence, in this study, a combination of segmentation algorithms is deployed. Classification accuracy is tested using various classifiers, namely, SVM, Bagged Tree and Subspace KNN. SVM can be tuned using the value of optimisation parameter, i.e., the kernel used and hence, overfitting can be avoided. Bagged Tree, Subspace KNN are ensemble classifiers. An ensemble learner is robust, and less resource is required for tuning the parameters. Here, a comparative study on the performance of the classifiers is carried out using freely available Landsat 8 OLI and Sentinel 2 MSI data for monitoring ecological condition and mapping ecotopes present inside the bog.

2.2 Materials and Methodology

2.2.1 Study Region and Datasets

The site selected for this study is one of the largest bogs in Ireland, Clara Bog (Figure 2.1), more details about the bog are presented in Section 1.7.

For the best description of the bogs following open-source data is used:

1. Sentinel-2 Multispectral Instrument Level 2A (S2-MSIL2A)

S2MSIL2A has bottom-of-atmosphere (BOA) reflectance in cartographic geometry. The granules also called tiles, are 100x100km^2 ortho-images in UTM/WGS84 projection. The L2A-BOA product is atmospherically corrected and ready to use (Gatti & Bertolini 2013) and is accessed from https://scihub.copernicus.eu/. The area at test lies under tile id - T29UNV. Sentinel-2 has a total of 12 bands out of which nine bands are used for analysis in this study (Band 2-8A, Band 11).

(h)

Clara bog, Co. Offaly

Figure 2.1: Sentinel-2 MSI spectral bands, source Immitzer et al. 2016

2. Landsat 8 Combined (LC08)

Landsat 8 carries two push-broom instruments: The Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) in UTM/WGS84 projection. Atmospherically corrected, ready to use data is accessed from http://earthexplorer.usgs.gov/. The area at test lies under path 207, row 23. In this study, 12 bands are used for analysis (Band 1-11 + Pixel quality assessment (QA)) (*Landsat-8 (L8) Data Users Handbook* 2016).

Images used in the study, are acquired by S2 and L8 on the same date (20th June 2017). Image from Sentinel-2 is resampled to 10 m (appropriate bands), and image from Landsat 8 is resampled to 30 m. The footprints for both the satellites can be seen in Figure 2.2.
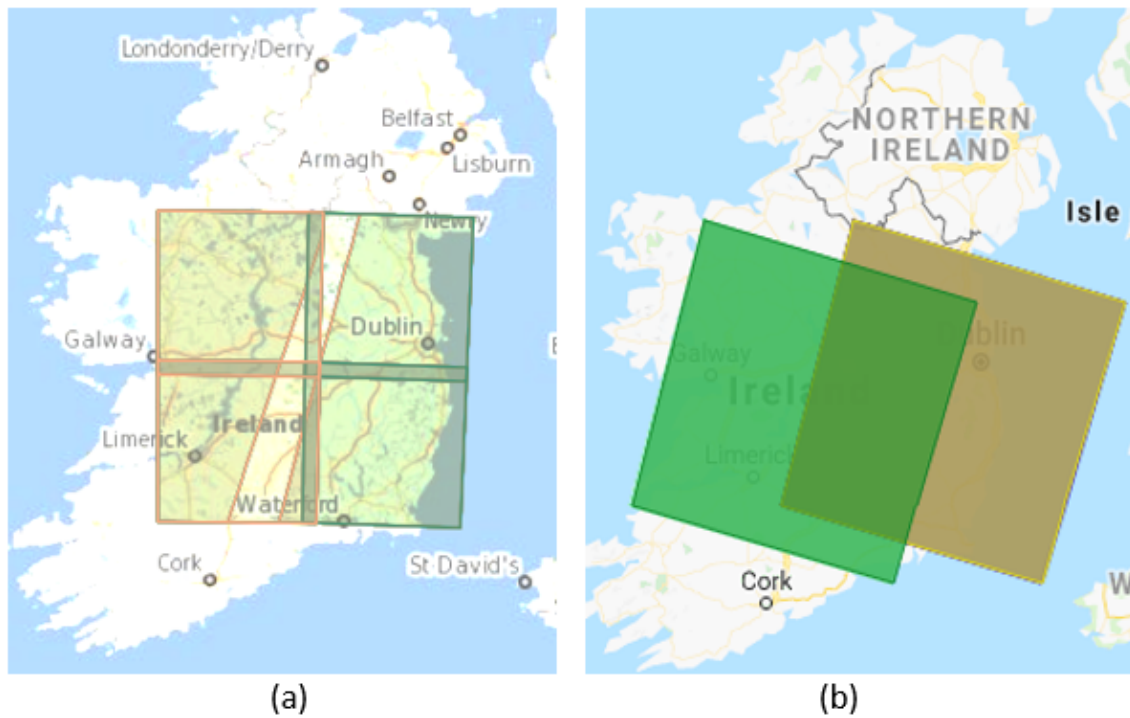
Figure 2.2: Footprints over Clara, Co. Offaly, Ireland (a) Sentinel-2, L2A, Tile_Id - 29UNV (https://scihub.copernicus.eu) (b) Landsat-8, OLI+TIRS, Path 207, Row 023 (https://earthexplorer.usgs.gov/)

2.2.2 Methodology

The methodology used is described in the following flowchart (Figure 2.3). First, it is necessary to delineate the ecosystem from the surrounding area in order to minimise the effect of outliers; this is achieved using segmentation algorithms. The delineated ecosystem is further divided into vegetation communities or ecotopes using a set of ensemble classifiers, namely Bagged Tree and Subspace KNN along with SVM.
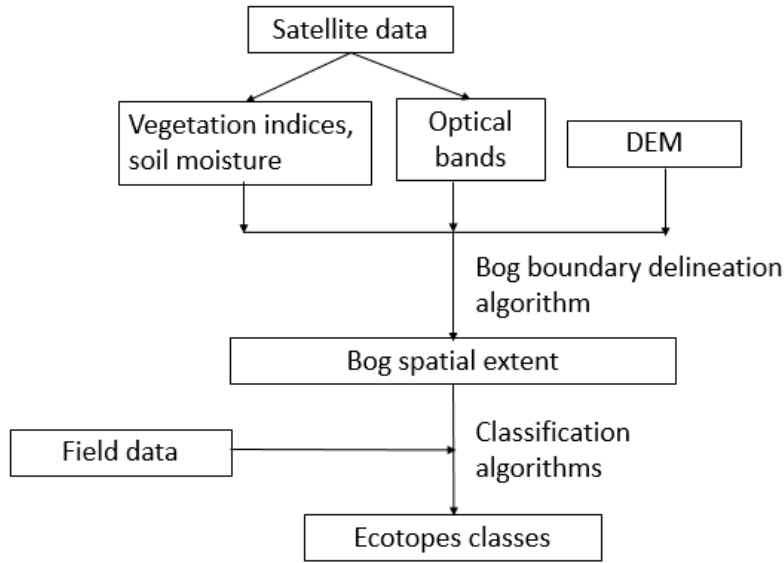
Figure 2.3: Methodology Flowchart

2.2.3 Vegetation Indices and Soil Moisture

The vegetation indices used in this study are:
1) Normalized Difference Vegetation Index

$$NDVI = (NIR - Red)/(NIR + Red) \tag{2.2.1}$$

NDVI indicates the amount of vegetation, distinguishes vegetation from the soil, minimizes topographic effects, etc.
2) Soil Adjusted Vegetation Index

$$SAVI = ((NIR - Red)/(NIR + Red + L)) * (1 + L) \tag{2.2.2}$$

where L is a soil correction factor.
3) Atmospherically Resisted Vegetation Index

$$ARVI = (NIR - RB)/(NIR + RB) \tag{2.2.3}$$

where RB is a combination of the reflectance in the Blue (B) and Red R channels RB = R-$\gamma(B - R)$ and γ depends on the aerosol type.
4) Soil Moisture: Normalized Difference Water Index

$$NDWI = (NIR - SWIR)/(NIR + SWIR) \tag{2.2.4}$$

Values range from -1, very low moisture level, to 1 very high moisture level.
Hence, there are a total of 5 extra layers i.e., NDVI, SAVI, ARVI, NDWI ,and DEM along with satellite bands which are fed into the algorithm as input characteristics. Therefore, for Sentinel-2 data set there is a total of 14 layers and for Landsat 8 there is a total of 17 layers.

2.2.4 Bog Boundary Delineation

Delineation of the wetland extent was carried out using three algorithms in conjunction:

1. Entropy Filtering: measures the relative change in entropy for the detection of edges. All the areas with potential objects are thus highlighted.

2. Canny Edge Detection: Initially, the intensity gradient is measured according to which background pixels are removed, i.e. only thin lines depicting edges remain. The algorithm uses two thresholds (upper and lower) to accept the pixels as the edge.

3. Graph Cut Segmentation: Divides every pixel into the foreground (source) and background (sink) based on probability and neighbourhood information. Two key steps then follow, first is to construct the graph, and the second is to produce min-cut (i.e. max flow).

Since the ecosystem and surrounding areas contain distinct vegetation, the NDVI image is used as the base image (Bi) for defining the bog boundary for Sentinel data, and band 8 (panchromatic) for Landsat 8 dataset.

Bog Boundary Delineation : *Algorithm*

$EBi \leftarrow$ Entropy_filter(Bi)
$CBi \leftarrow$ CannyEdge_Detect(Bi, EBi)
$GCi \leftarrow$ Graph_Cut(Bi, CBi)

Graph Cut image (GCi) is the delineated image. From GCi, spurious regions are removed on the basis of thresholding (Figure 2.4).
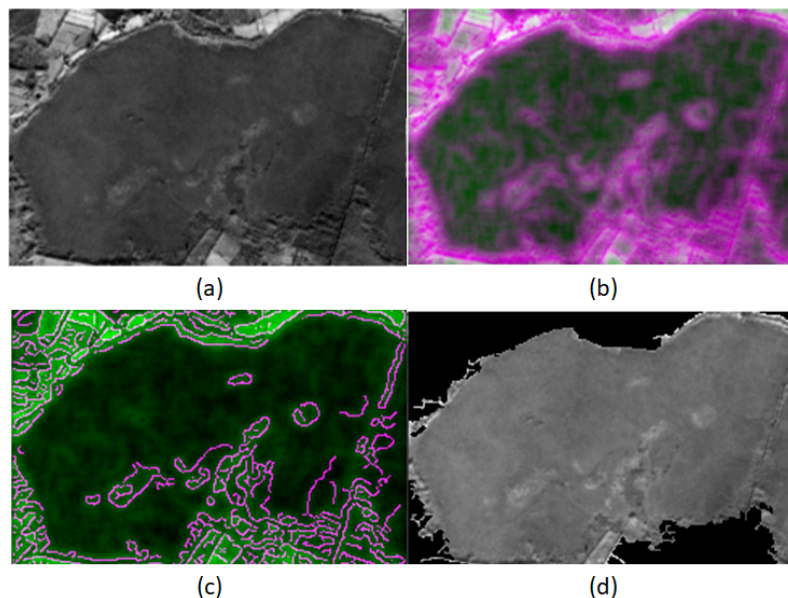
Figure 2.4: (a) Original Clara Image (b) Entropy Filter Image (c) Canny Edge Image (d) Boundary Delineated Image

2.2.5 Ecotope Identification and Classification

In this study, we have explored the applicability of pixel-based, supervised classification on a raised bog. The classifier is first trained on a subset of available data and then tested on a new location (test) to predict the classes (ecotopes) present. Here, we are doing a direct transfer of pixel-based knowledge from the training area to the testing area. The following classifiers are used for this purpose-

1. SVM - These are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training data, an SVM training algorithm builds a model that assigns new data to one of the two categories, making it a non-probabilistic binary linear classifier (Cortes & Vapnik 1995).

2. Bagged Tree – Ensemble, supervised classifier. It approaches to combine several machine learning techniques into one predictive model in order to decrease the variance; hence, tuning the prediction into an expected outcome (Thoma 2017).

3. Subspace KNN – Similar to Bagging, subspace KNN is an ensemble method to reduce the correlation between estimators (Ho 1998).

2.2.6 Validation

The result achieved by the classification algorithms was verified using field derived ecotope map (Figure 2.5). These maps were manually surveyed by NPWS. The location of the area being surveyed was recorded using a hand-held global positioning system (GPS), later trasferred into a geographical information system (GIS) software. The manual mapping techniques are discussed in detail in Section 1.3.1, the maps can be downloaded from https://www.npws.ie/maps-and-data.

Initially, a classification model is created using training data. For each of the classifier and data set, Model Accuracy (MA), Test Accuracy (TA) and Kappa Coefficient (kappa) are measured. The model accuracy (MA) is measured using 5-fold validation of training data. The test accuracy (TA) is the test accuracy when the model is applied to the testing data.

Ground truth is divided into:

Case 1: Training (50%) and Testing (50%)

Case 2: Training (70%) and Testing (30%)
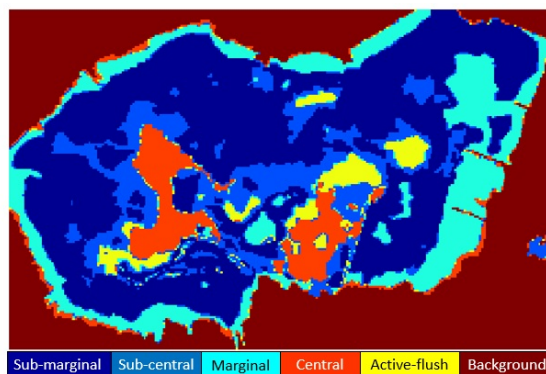
Figure 2.5: Ground truth with 5 ecotopes

2.3 Results and Discussion

2.3.1 Results

The results obtained using the aforementioned algorithms are validated using field de-
rived ecotopes. The spatial location of the ecotopes is not weather dependent, compared
to satellite imagery and corresponding vegetation indices which change with respect to
weather and other environmental conditions. This study primarily highlights the condi-
tion of the raised bog during the summer season. TA signifies the scope of transferring
the knowledge gained from the first half (Train) to identify ecotopes in the second half
(Test). Table 2.1 states the accuracy achieved using the aforementioned methodology:

	MA %	TA %	kappa
Sentinel-2 Case 1:			
SVM	73.96	32.40	0.1302
Bagged Tree	83.63	53.46	0.3424
Subspace KNN	81.65	52.95	0.3370
Sentinel-2 Case 2:			
SVM	71.71	36.36	0.2050
Bagged Tree	83.38	65.23	0.5013
Subspace KNN	62.02	36.50	0.1922
Landsat-8 Case 1:			
SVM	77.82	26.76	0.1305
Bagged Tree	87.27	53.87	0.3883
Subspace KNN	87.22	51.97	0.3635
Landsat-8 Case 2:			
SVM	72.92	29.06	0.1711
Bagged Tree	85.55	52.69	0.3539
Subspace KNN	85.00	52.71	0.3549

Table 2.1: Accuracies for all the cases
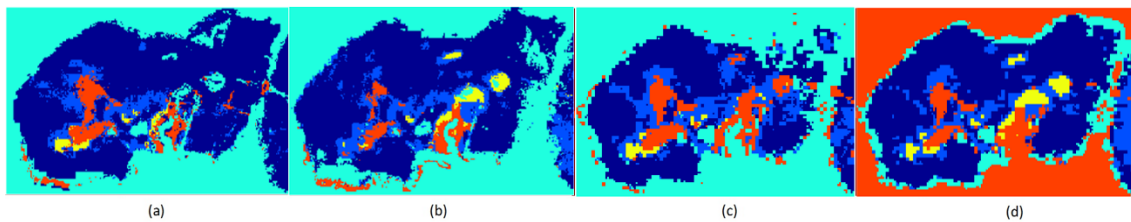
Figure 2.6: SVM Classified Image: (a)Sentinel-2 Case 1 (b)Sentinel-2 Case 2 (c)Landsat-8
Case 1 (d)Landsat-8 Case 2
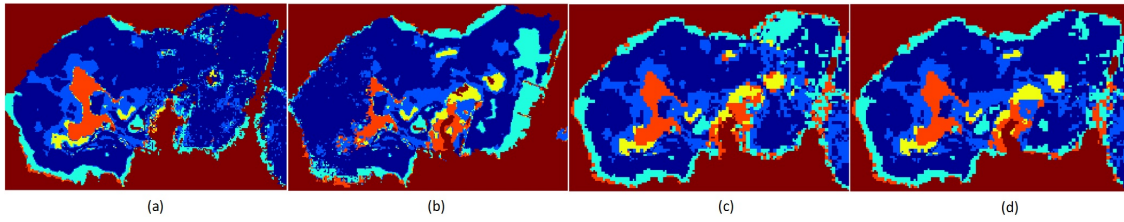
Figure 2.7: Bagged Tree Classified Image: (a)Sentinel-2 Case 1 (b)Sentinel-2 Case 2 (c)Landsat-8 Case 1 (d)Landsat-8 Case 2
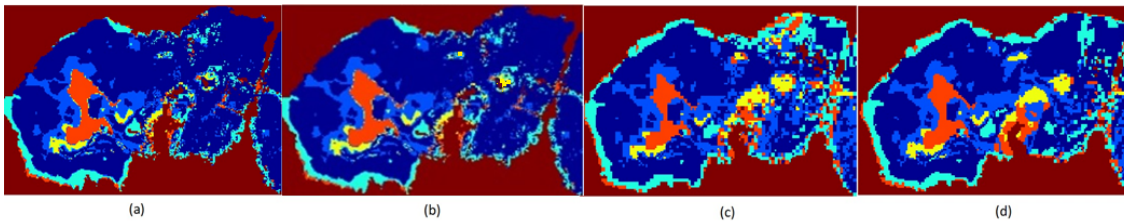
Figure 2.8: Subspace KNN Classified Image: (a)Sentinel-2 Case 1 (b)Sentinel-2 Case 2 (c)Landsat-8 Case 1 (d)Landsat-8 Case 2

2.3.2 Discussion

In this study, we have studied the ecological conditions of a raised bog using data from Sentinel-2, Landsat-8, their vegetation derivatives, and DEM. The key points in the result are discussed below –

- The use of a single algorithm for boundary delineation leads to the formation of smaller, non-connected objects, and hence, the bog is not delineated correctly. Using the entropy filter, canny edge detection, and Graph Cut in conjunction proved to be an effective way of delineating a complex structure from a middle-resolution image.

- SVM achieved the highest accuracy in the delineation of the submarginal ecotope but was not viable for marginal, active flush or background. The classifier has confused between marginal, active flush and background, giving 0% class accuracy in both datasets (Figure 2.6). SVM has a major drawback of tuning the parameters. In this study, the parameters were kept constant for both datasets, which increased the chances of overfitting; hence, the low accuracy.

- Ensemble classifiers (BT, SKNN) show similar results due to the fact that the ensemble methods are generally consistent (in terms of their effect on accuracy) (Figure 2.7, 2.8) (Maclin & Opitz 2011).

- The test (OA) accuracy is also highly dependent on the number of training pixels (Table 2.2).
  Total pixels in the image (Sentinel 2 MSI; 160*245) = 39200. The Submarginal ecotope is most correctly classified by all the classifiers followed by Marginal. Since other ecotopes

68

are present in much lower quantities (compared to submarginal and background), they are not identified correctly.

Name of Ecotope	Number of pixels
Submarginal	13771
Subcentral	4633
Marginal	4913
Central	1192
Active Flush	2443
Background	12248

Table 2.2: Number of pixels per class - S2

- Similarity between the signatures of the classes:

Ecotope Pair	JM distance
Submarginal and Subcentral	0.22
Submarginal and Marginal	0.85
Subcentral and Central	0.67
Central and Activeflush	1.48

Table 2.3: JM distance between ecotopes

Jeffries-Matusita (JM) distance is a widely used method for feature selection in multiclass problems (Davis et al. 1978). The values of JM distance (Table 2.3) between the ROI pixels depict low spectral separability between the ecotopes. These values were measured using 100 points from each ecotope-pair. A lower value means low separability and higher value shows higher spectral separability between ecotope pair. A higher value of JM distance is desirable for better identification of classes (Whelley et al. 2014).

2.4 Conclusions

In this study, we studied the application of mid-resolution satellite data for classification of a raised bog. The study initially describes a competent way of boundary delineation using a series of edge detection techniques. Vegetation indices along with soil moisture and DEM information are used as features to train the classification algorithms.Bagged tree (BT) classifier proved to be the best classifier for classification of the raised bog providing better accuracy than SVM or SKNN. Transfer of knowledge directly from train location to test location is not achieved effectively due to limitations in data-resolution and the amount of input training pixels. The study suggests that transfer of knowledge is effective between similar ecosystems when there is a distinct difference in the distribution of various ecotopes and pixels can be unmixed.

Chapter 3

Mapping Vegetation Communities - Segment based approach

"The limits of the possible can only be defined by going beyond them into the impossible."
— Arthur C. Clarke

Contents

This chapter corresponds to the paper published on 14 February 2020, in "International Journal of Applied Earth Observation and Geoinformation" [2].

3.1 Introduction

Wetlands, both natural and man-made, comprise approximately one-quarter of the total area of Ireland (*Ireland's Wetlands* 2000). They are known to provide a critical function concerning climate change, biodiversity, hydrology, and human health (Bureau 2001). Despite these known ecosystem services, the total wetland area in Ireland (as well as globally) has reduced significantly (more than 10%) over the past few decades due to human interference (Maltby & Acreman 2011). Hence, there is a pressing need for protecting such sites in Ireland.

In response to pressures related to land management, climate change and impacts from restoration activities, etc., ecological monitoring of the wetlands is necessary. However, ecological surveys are difficult to carry out due to the limitations in socio-economical resources, and the often secluded location of the habitats. Due to this, remote sensing (RS), with its increasing use in wider eco-hydrological applications, is becoming a more commonly utilised tool for the identification and classification of wetlands (Mahdavi et al. 2018).

Previous RS studies carried out on wetlands have mainly focused on mapping and distinguishing the different types of wetlands in the form of peatlands, marsh, swamp, bog, fen, etc. In Ireland, similar studies (Connolly 2019, Nitze et al. 2015) have mapped Irish wetlands using optical data. However, to date, few studies have carried out detailed community mapping inside the wetlands to determine their ecological conditions. The study by Lehmann et al. 2016 mapped sphagnum species in a peat bog along with dead vegetation and lichens using high-resolution images from an unmanned aerial system (UAS). Another study by Knoth et al. 2013 used visible and infrared images from two unmanned aerial vehicles (UAV) to map a bog complex. Apart from high-resolution images, free Landsat data has been used to accurately map 6 classes inside a raised bog (Crichton et al. 2015). Sentinel-2 (S2) time series data was used by Rapinel et al. 2019 to map 7 broad communities in a grassland. The authors have mapped 5 vegetation communities in a single raised bog in Ireland (Bhatnagar et al. 2018). The aforementioned studies and other available literature mainly focussed on communities identified within a single wetland type. In this study, we have developed a generalised methodology for detailed mapping of vegetation communities present within multiple types of wetlands. Instead of expensive but high-resolution UAV data, the study was implemented using time series of freely available Sentinel-2 data. The methodology can be applied to all types of wetlands and has been tested on wetlands present in Ireland.

For detailed species mapping, studies (Förster et al. 2017, Koch et al. 2017) have often used support vector machine (SVM) to classify time-series S2 data. Studies carried out by Amani et al. 2017, Millard & Richardson 2015, and Nitze et al. 2015 depict the advantage of using ensemble classifiers for wetland classification. A study done by Hird et al. 2017 for large area wetland mapping, applies boosted regression trees using GEE and multi-source satellite data. Other studies like Mahdianpari et al. 2019, Amani et al. 2019 illustrates the usage of ensemble classifiers via GEE for Canadian wetland mapping at a large scale. The study done on Clara bog by the authors (Bhatnagar et al. 2018) illustrated the use of BT applied on pixels for mapping vegetation communities within the wetlands. Consequently, a comparative analysis of the state-of-the-art classifiers was performed. The ensemble method called bootstrap aggregating, aka Bagging Trees (BT) proved to be the best technique. Although effective, this technique tends to produce some spurious pixels compromising the accuracy of the map. A possible solution to overcome these errors is to partition the image into homogenous closed groups or segments based on proximity (area). Hence, pixels are no more a single entity but part of an enclosed segment. This study extends the BT classification results to form a basis for image segmentation using contextual information. Maximum-a-posteriori estimation, commonly known as Graph Cut (GCut) segmentation, has been used for final vegetation mapping using image segmentation. The overall customised combination of the aforementioned algorithms used to precisely map the vegetation communities is termed as the Mapping Vegetation Communities (MVC) algorithm.

The study, therefore, presents a methodology and results of using the MVC algorithm on the satellite images to map vegetation communities in a bog, fen, and turlough wetlands

for multiple days between June 2017 and October 2018. The methodology presented can be followed for automatisation of the health monitoring of different wetlands over time which should be able to warn of progressive degradation or, alternatively progress in terms of restoration activities.

3.2 Wetland Description

Groundwater dependant terrestrial ecosystems (GWDTEs) include a wide range of wetlands, which are home to many endangered species (Kimberly and Coxon, 2013). This study concentrates mainly on two types of GWDTEs (see, Table 3.2.1) and raised bog wetlands. The later although not traditionally classified as GWDTEs require high water levels that may also be supported by the interaction between the water in the peat and the underlying regional groundwater (Regan et al. 2019). A total of 13 wetlands are mapped in this study consisting of 5 raised bogs, 5 turloughs, and 3 fens. These wetlands are located in the west and middle of Ireland, in over 6 different counties. The details of the wetlands are given in Table 3.1, with a further discussion on 3 specific case studies in Section 3.4.

	Wetland	Wetland description	Name	Location	County	Area(ha)
1	Raised bog	Discrete, dome-shaped masses of peat occupying shallow depressions.	Clara	53.327, -7.620	Offaly	250
		The ecological health of the bog can be determined by the nature	Mongan	53.320, -7.950	Offaly	205
		of its surface vegetation community composition (termed ecotopes)	Monivea	53.361, -8.680	Galway	160
		in relation to its water table and topography.	Killyconny	53.789, -6.969	Meath	191
			Knockacollar	53.969, -7.539	Laois	130
2	Calcareous fen	Peaty habitats, often fed by precipitation,	Scragh bog	53.579, -7.360	Westmeath	22.8
		groundwater and surface water (Goodall & Gore 1983).	Toryhill	52.538, -8.686	Limerick	0.95
		These wetlands are continually wet and rich in biodiversity	Ballymore	53.494, -7.637	Westmeath	11
3	Turloughs	Depressions in karst areas, seasonally flooded mostly by groundwater.	Blackrock	58.121, -8.741	Galway	143
		The ephemeral inundation plays a big role in the biological	Knockaunroe	52.988, -9.032	Clare	42.5
		diversity of the turloughs (Naughton et al. 2012).	Turloughmore	54.066, -8.692	Clare	21
			Roo West	53.967, -8.917	Galway	28
			Lough Aleennaun	53.007, -9.118	Clare	10.7

Table 3.1: Description of wetlands used in the study

3.3 Material and Method

Sentinel-2 (S2) data was used to map vegetation communities on the wetlands by performing supervised classifications using the BT and image segmentation using MAP. The S2-level 2A bottom of atmosphere (BOA) reflectance images were used for multiple dates over two years (2017, 2018) and the results were verified using field derived vegetation maps. An overview of the workflow is shown in Figure 3.1.
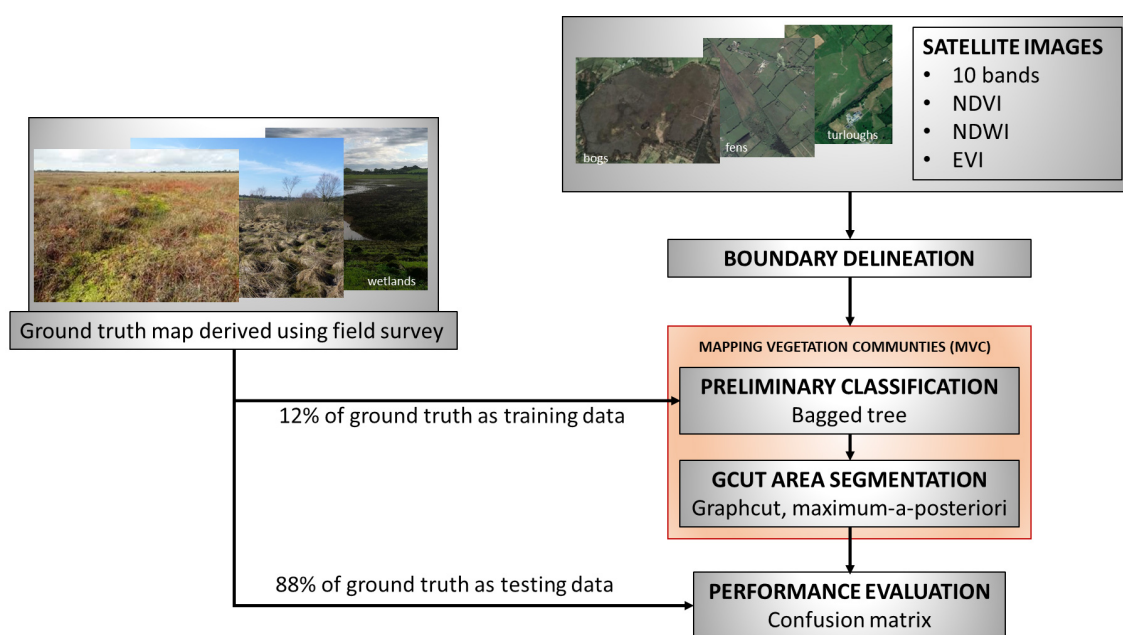
Figure 3.1: Overall methodology used to map vegetation communities in wetlands

3.3.1 Sentinel-2 Data

The Sentinel-2 Multispectral Instrument Level 2A (S2-MSIL2A) images used are bottom-of-atmosphere (BOA) reflectance in cartographic geometry. The L2A-BOA product is atmospherically corrected and ready to use (ESA Sentinel online, 2019) and is accessed from the Copernicus Open Access Hub (*Copernicus Open Access Hub* 2018). The areas that were used lie under tile IDs - T29UNV, T29UNU, T29UPV of Sentinel-2 footprints (Figure 3.2). Pre-processing of the images was carried out using software SNAP v.6.0 (*SNAP - ESA Sentinel Application Platform v6.0.5* 2018) which includes resampling, sub-setting, up-sampling of the bands acquired at 20 m to obtain a stack of spectral bands at 10 m resolution.
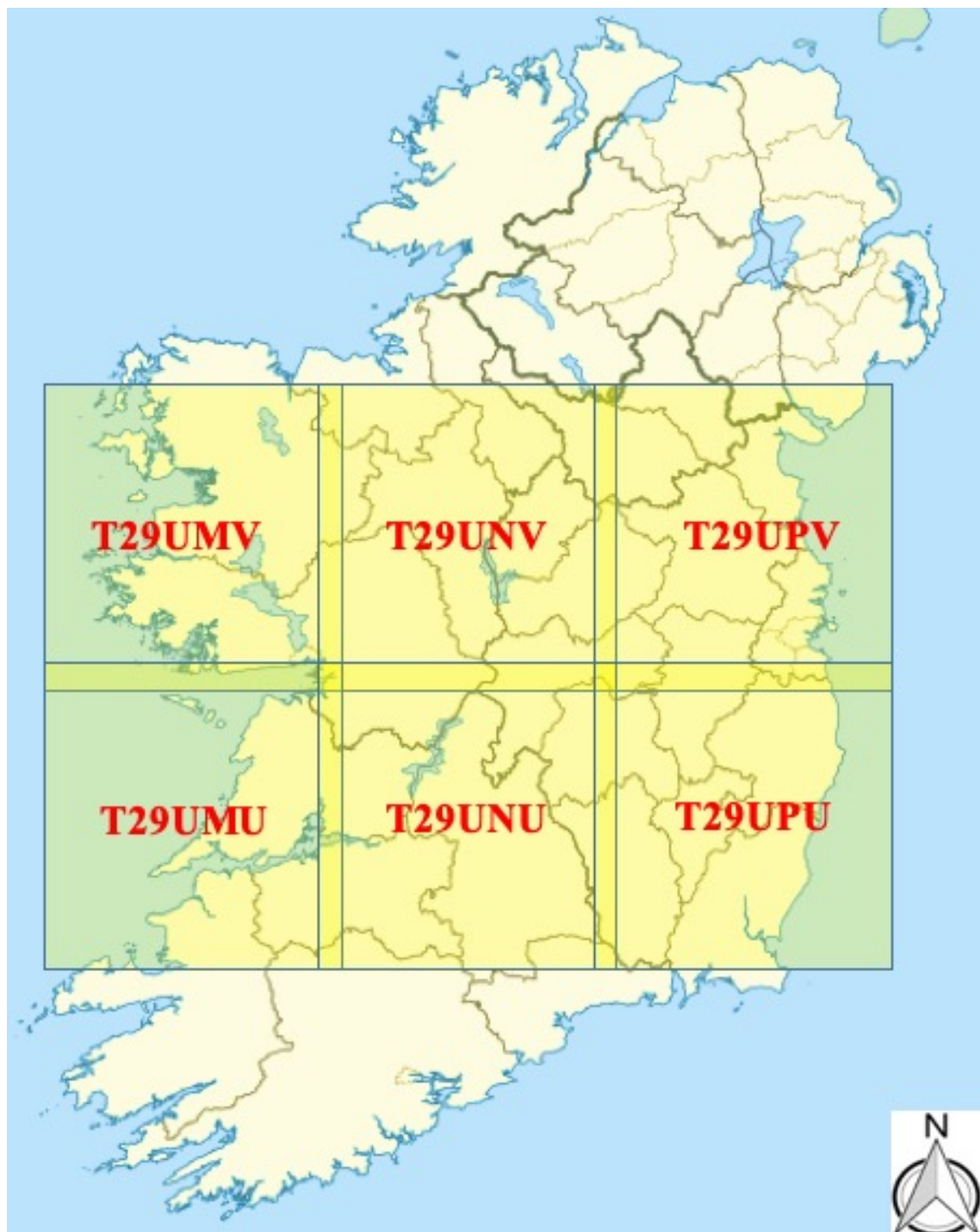
Figure 3.2: Sentinel 2 tiles over the study areas

S2 Level 2 has a total of 12 bands out of which 10 bands (Bands 2-8A and Bands 11-12) have been used for analysis in this study (Gatti & Bertolini 2013) as shown in Table 3.2. These bands were chosen due to their spectral significance and compatible spatial resolutions.

BANDS	Blue	Green	Red	VRE1	VRE2	VRE3	NIR	NarrowNIR	SWIR 1	SWIR 2
Wavelength μm	0.49	0.56	0.665	0.705	0.740	0.783	0.842	0.865	1.610	2.20
Spatial Resolution (m)	10	10	10	20	20	20	10	20	20	20

Table 3.2: Sentinel 2 bands description

The images used for all the wetlands analysed in this study were captured at multiple times throughout 2017 and 2018. For optimum analysis, only the cloud-free images were chosen, which restricted the sample size. A total of 7 dates were selected where both the bogs and fens being studied were out of the impact of clouds. Since turloughs are flooded during the winters, only summer images were chosen for its analysis. The main reason for starting the study from 2017 was the availability of S2 Level-2 products with Level-2A (bottom of atmosphere reflectance) images only available from March 2017. The months for which the imagery was captured for bogs and fens are June 2017, December 2017, February 2018, April 2018, June 2018, July 2018, and October 2018. For turloughs, the non-inundated months were chosen, which are June 2017, June 2018, and July 2018.
Along with the 10 spectral bands, 3 additional vegetation and water indices were used to enhance the classification process. The details about these indices are presented in Table 3.3.

Index	Formula	Key significance
NDVI (Liu & Huete 1995)	NDVI=(NIR-Red)/ (NIR+Red)	$1 \geq NDVI > 0.1$ represents the presence and status of vegetation.
EVI (Gao et al. 2000)	$EVI = \frac{2.5 \times (NIR - RED)}{((NIR + 6.0 \times RED - 7.5 \times BLUE) + 1.0)}$	$0.8 \geq EVI \geq 0.2$, improves NDVI on high leaf area index (LAI) or chlorophyll.
NDWI (Gao 1996)	NDWI = (NNIR – SWIR) / (NNIR + SWIR)	$NDWI \geq 0.5$ detects open water and ranges much less for vegetation. It highlights soil moisture and wet-vegetation communities.

Table 3.3: List of vegetation and water index used for mapping

3.3.2 Boundary Delineation

For mapping and monitoring the wetlands, a boundary delineation algorithm was applied, as detailed in Bhatnagar et al. 2018, using NDVI as the base image (B_i) to separate wetlands from non-wetland areas (see Figure 3.3), for more details on method see Section 2.2.4. It must be noted that this technique is useful for the delineation of bogs and fens; however, turloughs are much harder to distinguish. The turloughs' boundary lacks consistency based on the season, and therefore, the delineation was carried out using the ground truth shapefiles, which were field-surveyed manually by a team of ecologists at NPWS, Ireland (*Maps and Data : National Parks Wildlife Service* 2019). Further in this study, the 3-dimensional image of size $a \times b \times u$ was transformed into a 2-dimensional image using Eq. 1,

$$a \times b = n \qquad (3.3.1)$$

with the final delineated image represented as $\mathbf{I}^{n \times u}$, where u = 13 (number of bands) and n is the number of pixels ($a \times b$). The 10 spectral bands, NDVI, EVI, and NDWI, were used as features (layers) in this study and were defined in a feature space U = (n,u).
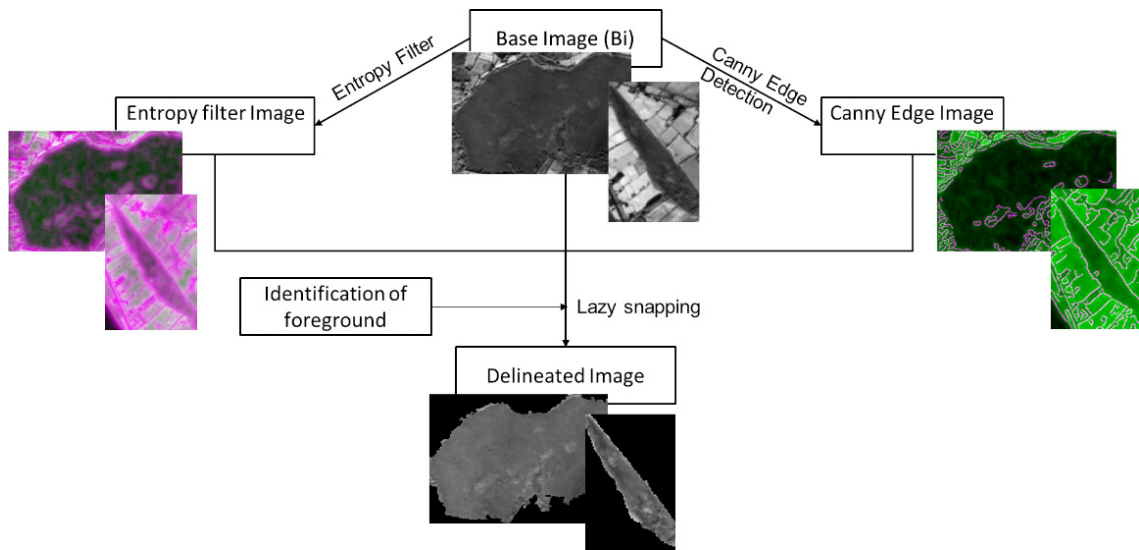
Figure 3.3: Boundary delineation, Clara bog

3.3.3 Mapping Vegetation Communities (MVC)

The MVC algorithm is based on the GCut and BT classification algorithms. The steps of the MVC algorithm are described below. The algorithm was written using MATLAB v.2018b (*MATLAB, The MathWorks* 2019).

1. Preliminary Classification

   For performing classification, multiple state of the art classifiers were tested for a given dataset. edIn this study, the amount of training data directly corresponds to the amount of field work and manual assessment done by ecologists. The resolution of imagery is 10m, therefore, x% of 100 pixels is equal to x ha of field work. To reduce the amount of field work, the amount of training data was to be decreased. After a heuristic search, the ground truth was sampled randomly into training [Z] (12%) and testing [$\hat{Z}$] (88%), the training data containing a mixture of uniformly distributed data from each class.

   The set of classifiers include two kernel-based classifiers: (i) an SVM classifier, used with radial basis function (RBF) kernel (Cortes & Vapnik 1995) and Naïve Bayes (Rish 2001) with Gaussian kernel and (ii) a Euclidean distance-based classifier: k-Nearest Neighbour (kNN) with k=2 (Guo et al. 2003). Two ensemble classifiers were also compared: (i) Bagged Tree (BT) (Bauer & Kohavi 1999), and (ii) Random Forest (RF) (Liaw & Wiener 2002) with 30-subspaces each. The classifier giving the best accuracy was chosen for further analysis. Table 3.4 presents a comparison of the classifiers (for the Clara bog dataset) for different criteria, namely model (algorithm) accuracy (5-fold validation accuracy), misclassification cost (based on 0 for correct and 1 for incorrect classification across the 8000 random pixels), prediction speed (number of observations predicted per second), and training time (the amount of time the classifier took for training in seconds).

	Model accuracy	Misclassification cost	Prediction speed	Training time
SVM	78%	1757	$\approx$4900 obs/sec	23.16 s
kNN	74.8%	2017	$\approx$13000 obs/sec	04.22 s
Naïve bayes	71.5%	2283	$\approx$880 obs/sec	43.96 s
BT	80.4%	1567	$\approx$21000 obs/sec	08.74 s
RF	78.3%	1976	$\approx$ 42000 obs/sec	10.26 s

Table 3.4: Comparison of state-of-the-art classification techniques for Clara bog dataset

2. Likelihood estimation

   In the next step, the likelihood related to every pixel is determined. For the image $\mathbf{I}^{n \times u}$ and feature space U, the posterior probability can be defined as $Pr(F|U, I)$ (JACKSON & RAGAN 1975). For the given posterior probability, the MAP estimate is given by (Veksler & Zabih 1999):

$$argmax_F(Pr(U, I|F)Pr(F)) \tag{3.3.2}$$

   where $Pr(U, I|F)$ is the likelihood function, and Pr(F) is the prior over the classes in F.

3. edCalculation of data and smoothness functions

   In this step the data and smoothness function for $U \epsilon I^{n \times u}$ is calculated. The data function $D^{(n \times k)}(p_k, U)$ measures the cost of assigning the classified label p_k to the pixel p for a given feature space U in the image $I^{n \times u}$. $D^{(n \times k)}(p_k, U)$ can be expressed following Boykov et al. 2001:

$$D^{(n \times k)}(p_k, U) = ||U(p_k - R(p)||^2 \tag{3.3.3}$$

   where $R(p)$ is the observed reflectance intensity vector of the pth pixel.

   In this study, neighbouring pixels were encouraged to have the same class using a smoothness regularisation function $V_{p,q}^{k \times k}(p_k, q_k)$. This signifies the cost of assigning the classified label p_k, q_k to adjacent pixels, p,q and was used to impose spatial smoothness. It is defined using the Potts model (Chen et al. 2007, Boykov et al. 2001) as follows:

$$V_{p,q}^{k \times k}(p_k, q_k) = c \times exp(- \triangle (p, q)/\sigma) \times T(p_k \neq q_k) \tag{3.3.4}$$

   where $\triangle(p, q) = |R(p) - R(q)|$ denoting the difference in the reflectance value vector of p and q. c is a smoothness factor ($c > 0$) which was determined by trial and error to be 0.75, a value that was well suited for all wetland types. σ is used to control the contribution of $\triangle(p, q)$ to the penalty ($\sigma > 0$). The value of σ depends on the standard deviation of the neighbouring pixels (p,q): the larger $\triangle$ between two neighbouring pixels, the greater the likelihood they have to be partitioned into two separate segments (Chen et al., 2007). T = 1 if $p_k \neq q_k$ and 0 otherwise. $V_{p,q}^{k \times k}(p_k, q_k)$ was defined with respect to the number of classes (k) so as to maximise the gradient between the pixels of differing classes.

   The $D^{(n \times k)}(p_k, U)$ used for this study was the confidence scores vector of the size $n \times k$. It is the posterior probability linked with each pixel for every class. The $V_{p,q}^{k \times k}(p_k, q_k)$ i.e., the smoothness function was defined with respect to the number of labels (k) present so as to maximise the gradient between the pixels of differing labels. It was a matrix of size $k \times k$.

4. Estimation of energy ($E(p_k, U)$)

   One of the main functions of machine learning is to encode dependencies between variables LeCun et al. 2006. Study by Boykov et al. 2001 explains energy (loss) minimization can be interpreted directly as posterior maximising. Using probability functions from previous

steps, we get the energy function as:

$$E(p_k, U) = \sum_{p\epsilon n} D^{n\times k}(p_k, U) + \sum_{p,q\epsilon n} V_{p,q}^{k\times k}(p_k, q_k) \qquad (3.3.5)$$

5. Energy minimisation using α-expansion
   The energy function was defined using Eqn 3.3.5. The α-expansion min cut based integer optimisation algorithm was used for minimising energy (Boykov et al. 2001, Li et al. 2011).This algorithm yields very good approximations to the segmentation and is quite efficient from a computational point of view, as described in Boykov et al. 2001.

6. Forming discrete segments
   The final step is the comparison of energies of each pixel calculated in step 3 with its surrounding pixels. The pixels are classified such that the energy is minimised i.e., the maximum posterior probability per class per pixel is maintained. The loop (steps 3, 4 and 5) continues until all the pixels are agreed upon and discrete segments are formed (using α-expansion). The final mapped image is defined by $[\hat{F}(k)]$.

MVC : *Algorithm*

---

for every pixel in the mapped image F having class p_k
$q_k \leftarrow argmin(E(p_k, U))$
If $E(q_k, U) < E(p_k, U)$:
$q_k = p_k$
Else
Repeat until satisfied: $q_k \leftarrow argmin(E(p_k, U))$

3.4 Performance Evaluation

Field surveys were commissioned by ecologists in the National Parks and Wildlife Service (NPWS), which were made available for analysis and mapping (NPWS, 2019). These field-derived maps were used as the ground truth (GT) in this study, see,Section 2.2.6. The details on manual mapping of the vegetation is discussed in Section 1.3.1. It must be noted that the MVC algorithm was carried out for all the dates using training data [Z] located at the same location. The final image $[\hat{F}(k)]$ was then validated against the testing data $[\hat{Z}]$. The confusion matrix was used as an accuracy measure with the rows representing the ground truth and the columns representing the predicted classes by the MVC algorithm. Precision and sensitivity along with the overall accuracy (OA %; Eqn. 3.4.1) and kappa (κ) was also calculated for every vegetation community present in all the wetlands (Labatut & Cherifi 2012).

$$OA = \frac{TP+TN}{TP + FP + FN + TN} \qquad (3.4.1)$$

where TP = true positive, FP = false positive, FN = false negative, TN = True negative

3.5 Case studies

The MVC algorithm is validated for a total of 13 wetlands over the period of June 2017-October 2018, from which this study features three case studies as illustrative examples (see locations in Figure 3.4).
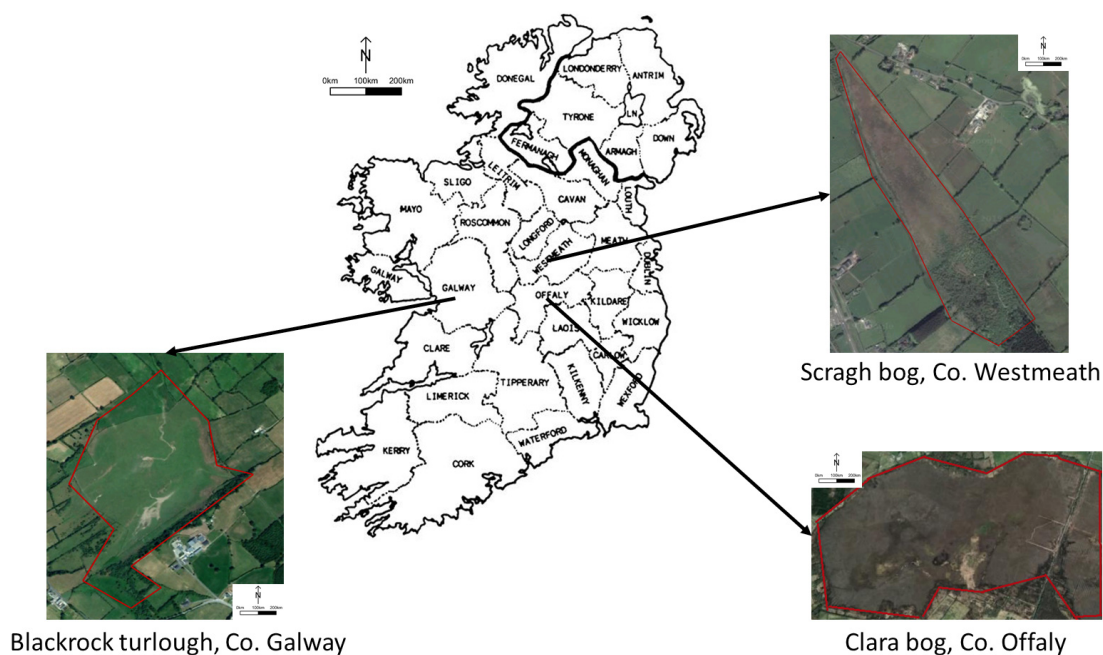
Figure 3.4: Location of wetlands used for the case-studies in Ireland

For mapping, the spectral signature of the vegetation communities present in the wetlands, the S2 image, and the ground truth maps was overlaid. Only pure pixels belonging to the vegetation communities were selected. The median reflectance value for every community was plotted against the wavelengths under consideration (Figure 3.5, 3.7, 3.9). In this study, the S2 image spectra were analysed at specific wavelengths set out in Table 3.2. The detailed discussion of spectral signatures at each wetland site along with the MVC algorithm results is discussed in the subsections below.

3.5.1   Case study 1 - Clara Bog

The key ecotopes in Clara bog are based on a vegetation classification system developed by the NPWS to characterise the different conditions of a bog from the ecologically pristine active raised bog (ARB) down to a degraded status, are the Central ecotope, Subcentral ecotope, Active Flushes and soaks, Submarginal ecotope, and Marginal ecotope. More details on Clara bog vegetation can be found in Section 1.7.

Spectral Profile

The spectral signature of Clara bog can be seen in Figure 5 where the dissimilarity between the Central and Marginal ecotope is notable. The wetness of Central, Subcentral and Active Flush ecotopes is evidenced by the low reflectance value in the SWIR wavelengths. From Figure 3.5, the Submarginal and Marginal ecotopes have similar reflectance values in NIR wavelengths, but they can be distinguished by the SWIR wavelengths. The higher reflectance value of the Marginal ecotope in SWIR wavelengths depicts its dry nature. In general, the reflectance for wavelengths greater than 1300 nm is inversely proportional to the amount of water content present in the plant (Ng et al. 2007). This is also dependent on the leaf thickness of the plant.
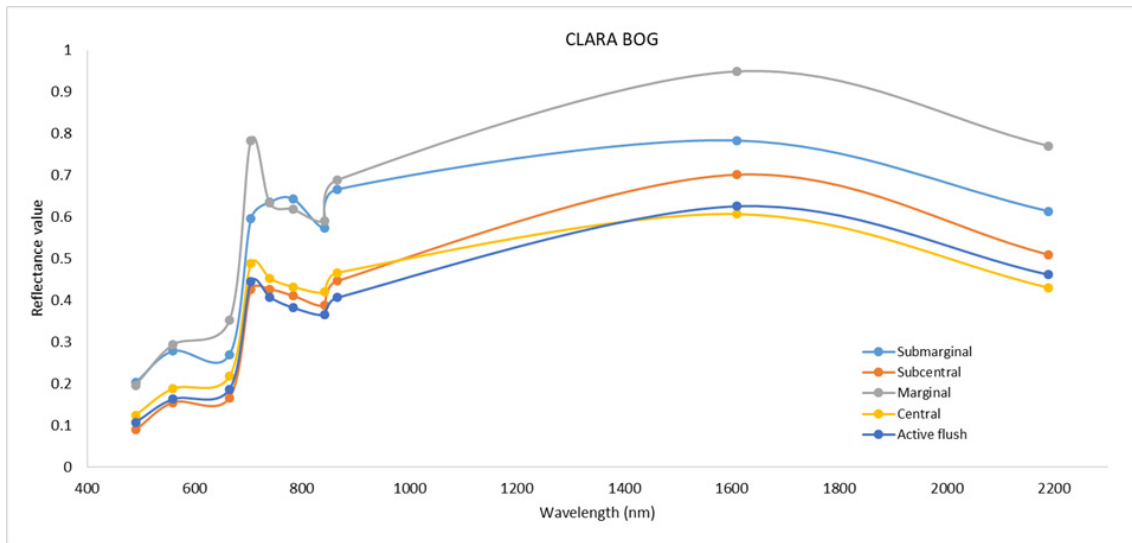
Figure 3.5: Spectral signature of different ecotopes in Clara bog, 30th June 2018

Vegetation Mapping in Clara Bog

The results obtained for Clara bog for the year 2017, 2018 using the MVC algorithm are presented in Figure 3.6 and Table 3.5. In Clara bog, the Central ecotope was very well-identified. This is because it is located in the wettest area of the raised bog, with indices such as NDWI giving a clear indication of its location. The Central ecotope also has a high precision value (see Table 3.5) which indicates that the misclassification of other communities as Central is low, justifying the above statement. The sensitivity value of the Submarginal is the highest out of all the ecotopes. However, during summer, the Subcentral ecotope starts to dry and integrate into Submarginal increasing false negatives (FN).
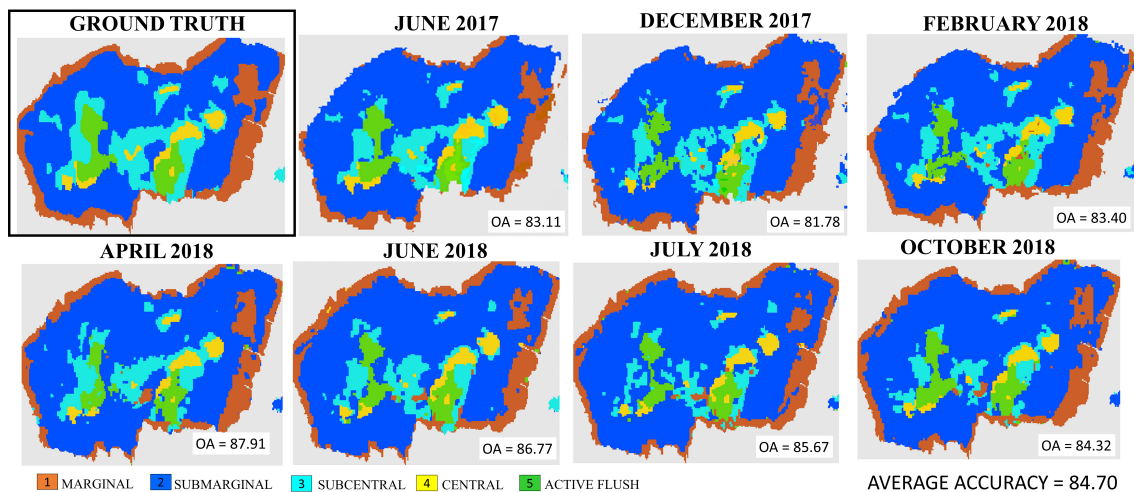
Figure 3.6: Vegetation Communities in Clara bog for the year 2017, 2018

	1	2	3	4	5
1	13553	79	107	38	38
2	1346	2530	45	142	55
3	883	14	2534	1	8
4	43	73	0	772	54
5	135	48	9	30	1275
Precision	84.9	92.2	94	78.5	89.1
Sensitivity	98.1	61.4	73.6	81.9	85.1
OA	86.77				
κ	0.79				

Table 3.5: Confusion matrix - Clara Bog, 30th June 2018 with 1. Marginal 2. Sub-marginal 3. Subcentral 4. Central 5. Active Flush

The Marginal ecotope is under segmented in images from 2017 and early 2018 but is identified well in summer 2018 (Figure 6), which was one of the driest summers in Ireland in over 30 years. It can be inferred that the degraded but vegetative part of the bog (Marginal ecotope) can be best identified in a dry environment using S2 images. The OA for Clara bog ranges from 81% to 87% for all the seasons, and for all seasons the crucial ecotopes such as Central, Subcentral were picked out accurately.

3.5.2 Case study 2 - Scragh Bog (fen)

Scragh bog located in midlands of Ireland is defined in detail in Section 1.7.

Spectral Profile

Figure 3.7 presents the spectral signature of Scragh bog (fen). It can be seen that the communities of wet willow-alder-ash woodland, marsh, pools, transition mires (quaking bog) and raised bog follow an overlapping pattern in the SWIR wavelengths depicting wetness in these communities. The Rich fen community, being the most nutrient-rich, has a high reflection at green and NIR wavelengths. Transition mires are generally associated with the wettest part of the fen (Foss & Crushell 2008), which is justified with its relatively low reflectance values at SWIR wavelengths. In contrast, Conifer woodland has a high value of reflectance in SWIR wavelengths suggesting that the community is dry. This community has high reflectance at green wavelengths (band 3) which is in agreement with the visual appearance of this community.
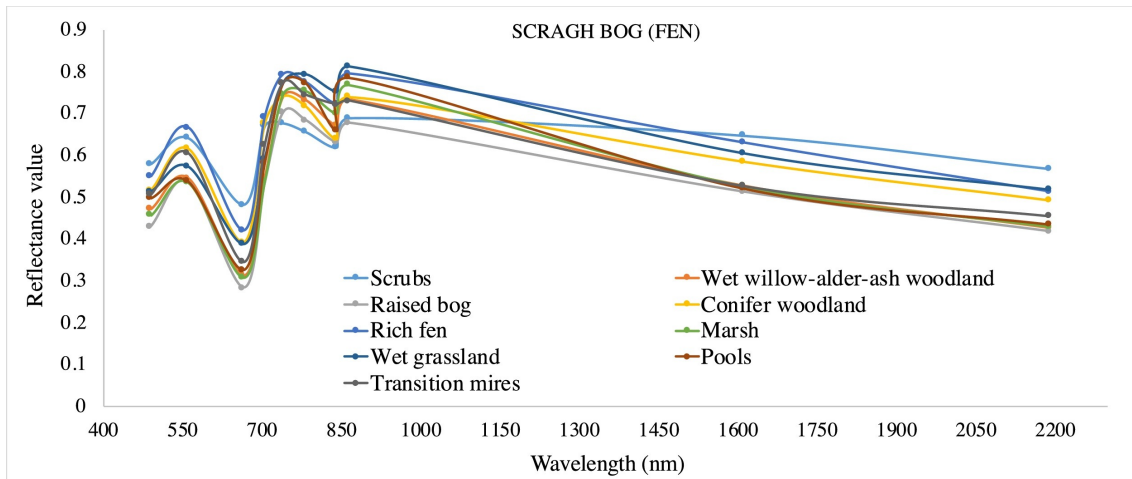
Figure 3.7: Spectral signature of different vegetation communities in Scragh bog, 30th June 2018

Vegetation Mapping in Scragh Bog

The results obtained for Scragh bog (fen) for the year 2017, 2018 using MVC algorithm are presented in Figure 3.8 and Table 3.6.
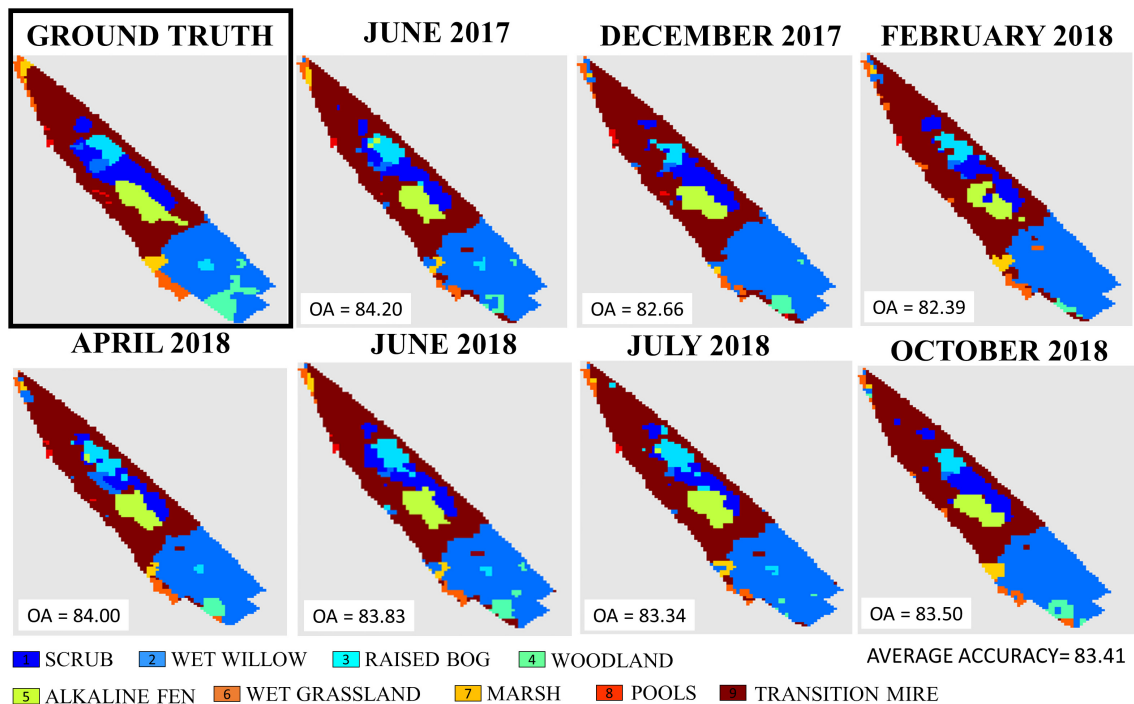
Figure 3.8: Vegetation Communities in Scragh bog for the year 2017, 2018

	1	2	3	4	5	6	7	8	9	10
1	165	5	1	3	10	5	0	1	1	12
2	28	106	1	0	24	3	0	1	4	3
3	24	5	25	0	24	2	0	1	1	2
4	4	10	0	67	54	3	0	1	0	5
5	2	10	0	4	687	10	2	2	1	1
6	0	3	0	0	78	100	0	5	1	9
7	0	0	0	0	17	3	33	4	0	9
8	0	0	0	0	35	1	0	48	1	9
9	0	2	0	0	8	0	5	0	51	50
10	3	0	0	0	7	1	0	0	0	754
Precision	73	75.17	92.59	90.54	72.77	78.12	82.5	76.19	85	88.29
Sensitivity	81.28	62.35	29.76	46.52	95.54	51.02	50	51.06	43.96	98.56
OA	79.62									
κ	0.68									

Table 3.6: Confusion matrix - Scragh Bog, 30th June 2018 for vegetation communities 1. Scrub 2. Wet willow 3. Raised bog 4. Woodland 5. Alkaline fen 6. Wet grassland 7. Marsh 8. Pools 9. Transition Mire

In Scragh bog (fen) (Figure 3.8), the transition mire, an important community, has a very low FN leading to high sensitivity values (Table 3.6). This means that the likelihood of the misclassification of transition mires as other classes/communities is very low. Another important community, Rich fen (alkaline fen) has been identified with high precision. The actual shape of the Rich fen has been best identified in images from June 2017, 2018. This is because of the difference in wetness between transition mire and Rich fen during summer, which is picked up using NDWI. Similarly, the raised bog is identified best in the summer images (June, July, etc.), compared to the wet grasslands which are identified best in winter images. This can be attributed to the fact that during the summertime, wetness decreases and communities start to merge with the willow-alder-ash woodland community (as can be seen in Figure 3.8). For the fen, the OA ranges from 82% to 84% for all the seasons, and for all seasons the crucial vegetation communities such as Rich fen community, transition mire were picked out accurately.

3.5.3 Case study 3 - Blackrock Turlough

Blackrock turlough, under dry conditions, consists of 12 vegetation communities, as defined in Section 1.7.

Spectral profile

Figure 3.9 represents the spectral signature for Blackrock Turlough. As discussed in the previous section, the species in broad community 2 show a similar vegetative pattern in visible and NIR wavelengths and overlapping pattern in SWIR wavelengths. However, the *Carex nigra-Ranunculus flammula* community has higher reflectance value in NIR wavelengths, indicating a better condition of the vegetation. The vegetation species from broad community 1 differ in their spectra in vegetation red edge bands but exhibit similarity elsewhere. From the curve, the community *Filipendula ulmaria-Potentilla erecta-Viola sp. C* and *Potentilla anserina-Potentilla reptans* have lower reflectance in NIR wavelengths, indicating stress. The *Lolium grassland* has higher reflectance in the NIR wavelengths

depicting less stress in the vegetation and scrubs have a higher reflectance in SWIR wavelengths depicting lower water content in scrubs. This is in agreement with the field derived results stated in Waldren et al. 2015.
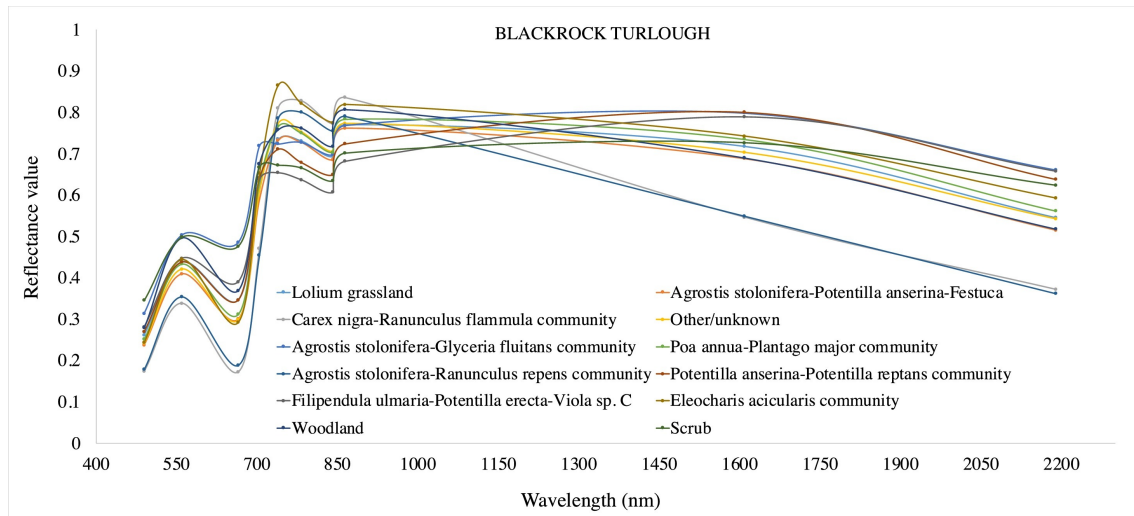
Figure 3.9: Spectral signature of different vegetation communities in Blackrock turlough, 30th June 2018

Vegetation mapping in Blackrock Turlough

The results obtained for Blackrock turlough for 2017 and 2018 using MVC algorithm are presented in Figure 3.10 and Table 3.7.
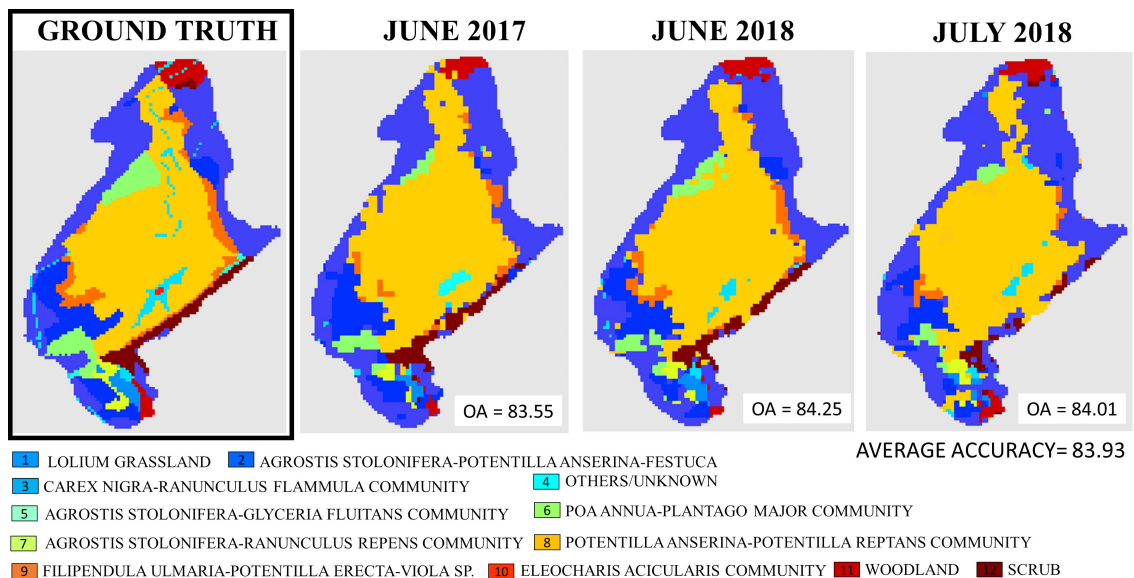
Figure 3.10: Vegetation Communities in Blackrock turlough for the year 2017, 2018

	1	2	3	4	5	6	7	8	9	10	11	12
1	1458	12	0	2	0	6	0	38	5	0	7	8
2	71	413	0	0	0	3	9	33	11	0	7	0
3	7	2	49	2	0	0	10	1	0	0	0	0
4	37	1	0	45	0	0	0	110	1	0	12	2
5	12	0	0	0	3	0	0	0	2	0	0	4
6	33	14	0	0	0	207	1	67	0	0	0	0
7	10	0	5	0	0	1	50	4	0	0	0	5
8	64	21	0	4	0	7	0	2268	10	0	2	4
9	53	7	0	0	0	0	0	69	174	0	0	13
10	0	0	0	4	0	0	0	3	0	0	0	0
11	50	0	0	0	0	0	0	1	0	0	115	0
12	41	0	0	1	0	0	0	18	2	0	0	188
Precision	79.4	87.8	90.7	77.5	100	92.4	71.4	86.8	85	0	80.4	83.9
Sensitivity	94.9	75.5	69.0	21.6	14.2	64.2	66.6	95.2	55	0	69.2	75.2
OA	84.2											
κ	0.81											

Table 3.7:   Confusion matrix – Blackrock Turlough, 30th June 2018 for vegetation communities 1. Lolium grassland 2. Agrostis stolonifera – Potentilla anserine – Festuca rubra 3. Carex nigra – Ranunculus flammula 4. Unknown community 5. Agrostis stolonifera – Glyceria fluitans 6. Poa annua – Plantago 7. Agrostis stolonifera – Ranunculus repens 8. Potentilla anserine – Potentilla reptans 9. Filipendula ulmaria – Potentilla erecta – Viola sp. 10. Eleocharis acicularis 11. Woodland 12. Scrub

The major vegetation species present in Blackrock belong to broad community 3, *Potentilla anserina – Potentilla reptans* which is surrounded mainly by *Lolium grassland* and woodlands. It can be seen that the shape of the *Potentilla reptans* community remains intact throughout the two years with high precision and sensitivity values associated. The *Lolium grassland* also has a very high sensitivity value with a high number of true positive (TP) values which are correctly identified. This community is typically dry and not often affected by floods. The species in broad community 3 also have a high precision value of around 85%. Broad community 1, which contains ruderal species, is partially identified. The species in broad community 2 are small in size but have high precision. Much of the *Agrostis stolonifera – Glyceria fluitans* community is misclassified as *Lolium grassland* hence, the low sensitivity. With small plant communities, the FN increases due to the coarse (10 m) resolution of the satellite image. Hence, it could be suggested that the small vegetation communities should be grouped and analysed as a part of a broad community using S2 images in the future. The OA for the Blackrock turlough is consistent at 84%, and for all months under consideration, the crucial communities such as *Lolium grassland, Potentilla reptans* were picked out accurately.

3.6   Discussion

The proposed methodology illustrates the use of open-source, mid-resolution S2 data for accurately mapping vegetation communities in the bogs, fens, and turloughs. Similar studies have mainly concentrated on a single wetland-type making the application limited (Knoth et al. 2013, Crichton et al. 2015, Lehmann et al. 2016). Using the proposed generalised methodology, up to 18 vegetation communities were mapped in various wetlands,

depicting its wide applicability. The mapping of the different vegetation communities can be used to infer the ecological health of the wetland, making habitat surveys simpler and more effective for ecologists. A total of 13 wetlands were mapped over time between 2017 and 2018 using the MVC algorithm. Instead of combining all the images (Rapinel et al. 2019), the study classifies each S2 image separately to monitor changes over time. The average accuracy of all the wetlands mapped is shown in Figure 3.11.

The wetlands mapped in this study vary significantly in shapes and sizes, making it challenging to identify small vegetation communities using pixels at 10 m resolution. The S2-L2A imagery comes with 3 specific vegetation red-edge bands, an extra narrow NIR band that facilitates the identification of vegetation communities. To create a generalised algorithm, 10 bands of S2 imagery were insufficient; hence, 3 extra indices were used. A combination of both NDVI and EVI, gave a clear distinction of the health of the vegetation communities present in the wetlands by measuring the chlorophyll amount and not just the greenness of the plant. To identify soil moisture content and the wetness of vegetation communities, NDWI was used, which led to better overall classification of the wetland.
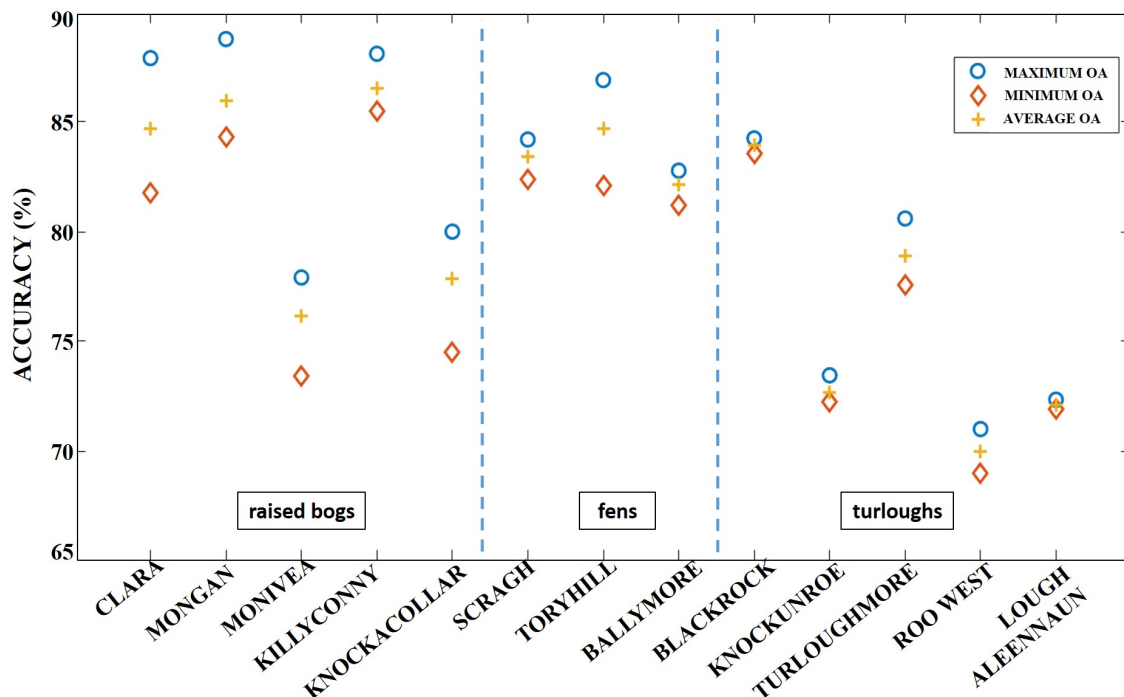
Figure 3.11: Maximum, minimum and average accuracy (%) for all 13 wetlands

For the raised bog, the study was successful in mapping the key ecotopes, Central, Active Flush, and Subcentral, which indicate the peat-forming areas within the bog along with the additional ecotopes. A total of up to 9 ecotopes were mapped using the proposed algorithm. The Central and Active Flush ecotope remains predominantly wet throughout the year. The usage of NDWI index makes these communities distinguishable from Subcentral, Submarginal and Marginal. The Marginal ecotope is the greener area making the NDVI index useful in its identification. The overall results prove that MVC algorithm presented in this study provides a promising solution for automatic field monitoring of these ecotopes.

For the limestone groundwater-fed wetlands, such as fens, the main challenge was to identify the transition zones such as the transition mire. The study was successful in identifying the transition areas as well as areas of raised bog within it. The use of SWIR wavelength and an additional NDWI index led to better identification of the wet communities. The transition mires (present in 2 out of 3 fen-sites) were mapped with high precision. High precision indicates low misclassification and the robustness of the MVC algorithm. A total of up to 9 communities were mapped in the 3 fen sites. Indices like NDVI and EVI led to better identification of the healthy (alkaline) area in the fen (Rich fen community) for all seasons.

Since turloughs are flooded throughout the winter period, the study was carried out using summer images of 2017 and 2018. A total of up to 18 vegetation communities were identified using the algorithm. The communities with high fertility, such as *Lolium grassland*, were present in all the turloughs. This community, although associated with dry conditions, has a higher level of nutrients and was identified with 72% average precision. EVI proved to be a better index for identifying grassland and scrubs due to its resilience towards chlorophyll saturation. Patches of water, pools, and flooded pavements were distinguished using NDWI (75% average precision). The study also suggests that the spatial extent of the different broad community groupings remain intact after seasonal flooding and can be identified using S2 imagery with high accuracy.

Figure 11 depicts that the wetlands with the bigger area and bigger sample pixel size have better OAs. Therefore, it can be suggested that the use of high-resolution images which will provide a higher sample size for all areas, even smaller communities may improve the algorithm performance. The reflectance value of the same vegetation communities differed in different wetlands. This was due to the change of S2-tile due to differing weather conditions, etc. However, the proposed MVC algorithm was robust and resilient to these errors.

Overall, the temporal study carried out on the wetlands using the satellite data over two years does not indicate any major ecological changes, thereby indicating its effectiveness. The highest accuracies of about 87% for the raised bog, 84% for the fen and 84% for the turlough were achieved. The size of vegetation communities played a role in the overall classification accuracy. If the size of the vegetation community was less than 30 pixels (0.3 ha), the precision of mapping decreased, leading to a decrease in OA (for example, for the turloughs of Roo West and Lough Aleennaun). A recent study byRapinel et al. 2019 has mapped more than 30 plant species with an average accuracy of 78% by merging the species in 7 classification communities. The present study achieved comparable OAs for mapping the small vegetation communities ($\geq$ 0.3 ha) without such limitations. The OA from each wetland showed good consistency when applied to images from different times and seasons within the 2 years. This consistency in the accuracy demonstrates the robustness of the MVC algorithm and further indicates that no significant impact due to human interference or other changes has taken place in these locations, which also agrees with the reality on the ground. However, it can be assumed that if the MVC algorithm is applied to satellite images over much longer time spans, then ecological changes in the wetlands with changes in shape and size of key vegetation communities could be picked up, thereby demonstrating the effectiveness of the proposed methodology as an efficient remote monitoring tool.

3.7 Conclusion

This study successfully provides a detailed mapping of communities within wetlands using mid-resolution, open-source S2 data which has previously been carried out predominantly using expensive hyperspectral data or high-resolution data. The results provide a good characterisation of the content (S2 imagery) in both the spectral and the spatial domain. The proposed MVC algorithm has high overall accuracies for all wetlands, as well as a high precision for key vegetation communities. The novelty of the proposed methodology is that it is applicable to vegetation communities with complex boundaries of varying sizes. As shown in the study, the methodology is robust and repeatable for all seasons, all kinds of wetlands with minimal supervision. The use of this desk-based study can reduce the need for extensive field-work to a great extent, enabling ecologists to monitor the bogs using freely available satellite data. Such data can also be used to assess the progress of any restoration schemes for re-establishing the conditions in a wetland or identify potential damage (due to unlicensed drainage practices for example) which can then be used as a method of early warning for the authorities.

Chapter 4

Drone Image Segmentation

"It is our choices, that show what we truly are, far more than our abilities."

— J.K. Rowling, Harry Potter and the Chamber of Secrets

Contents

This chapter corresponds to the paper published on 12 August 2020 in "Remote Sensing" journal [3].

4.1 Introduction

The use of drones for different types of vegetation classification has increased many folds over the last decade. This is due to the technological development of affordable and lightweight drones. With drones, a very high and flexible spatial resolution can be achieved, which is not possible with satellite imagery due to their fixed orbits. In many cases the area of wetlands can be relatively small, whereby satellite-based classification is not sensitive enough and can produce large errors. There are several ways to reduce the

error in satellite images, but most of them require extensive hyperspectral bands. However, another method to get detailed monitoring of small areas is to use unmanned aerial vehicles (UAVs), more commonly known as drones.

Most drones typically carry optical cameras (RGB) and occasionally can support a thermal sensor, but some drones can also support more expensive hyperspectral cameras. The presence of a thermal/hyperspectral sensor allows more details to be gathered and improves spectral resolution. However, the dilemma about spectral versus spatial remains unanswered. A study by Hirano et al. 2003 has used AVIRIS hyperspectral data with 224 bands and 20 m spatial resolution to detect invasive plant species (Colubrina asiatica (Brongniart, Adolphe Théodore) in Florida. Another study by Pengra et al. 2007 used Hyperion (30 m) hyperspectral data to detect Phragmites australis (Steudel, Ernst Gottlieb) in coastal wetlands and states that, due to low spatial resolution, the analysis was affected by pixel mixing. Therefore, apart from high spectral resolution, a proper spatial resolution is also required for monitoring vegetation communities closely. Drone images have much higher spatial resolution when compared to satellite images. Drones have been explicitly used for species detection (Alvarez-Taboada et al. 2017, Baena et al. 2017, Dvořák et al. 2015, Hill et al. 2017). A study by (Ruwaimana et al. 2018) states many advantages of using a drone over satellite imagery for identification of land cover communities such as water, land *Avicennia alba* (Blume, Carl (Karl) Ludwig), *Nypa fruticans* (Verh. Batav. Genootsch. Kunst.), *Rhizophora apiculata* (Blume, Carl (Karl) Ludwig), and *Casuarina equisetifolia* (Linnaeus, Carl). Drone imagery has also been applied for specific applications such as analysing vegetation under shallow water, tracking waterbirds, and their habitats (Chabot et al. 2018, Han et al. 2017). A study by Zheng et al. 2018 concluded that a thermal (infrared) sensor on its own performs comparable to an RGB sensor, but a multispectral sensor (with multiple spectral bands and indices) is required for the best analysis of nitrogen on rice fields. Multiple spectral sensors, however, although useful, are costly (Govender et al. 2007). Therefore, taking all the points into consideration, as an alternative to an expensive camera, an RGB camera was used in this study.

For the analysis of drone data, many techniques are available. The state-of-the-art techniques in drone image analysis consist of both machine learning (ML) and deep learning (DL). The study by Miyamoto et al. 2018 applies multiple ML algorithms including random forest (RF), SVM, and gradient boosting decision tree (GBDT) to classify trees, grasses, bare gravel/sand bed, and water surface. The study achieved a high accuracy of up to 98% using RF classifier on UAV images. ML algorithms have also been used for vegetation segmentation. Apart from ML, advanced deep learning techniques are also now being widely applied for drone image segmentation. A recent state-of-the-art review by Hoeser & Kuenzer 2020 shows the surge of applying DL in the field of RS. It also gives details about various convolutional neural network (CNN) models and suggests that $\approx 20\%$ of all studies since 2012, uses DL with UAV imagery. DL have been applied for various applications like urban land classification (Zhang et al. 2010, Montoya-Zegarra et al. 2015), forest cover classification (Dechesne et al. 2017), and wetland type classification (Cui et al. 2019, Jiang et al. 2019). A study by Kentsch et al. 2020 uses ResNet50 and UNet for classification of forest tree-species, and Nigam et al. 2018 has used transfer learning to get the best semantic segmentation of the aerial images AeroScapes dataset. Both suggest that the usage of transfer learning enhances the analysis. A study by Do et al. 2018 has utilised both ML (linear regression) and DL (neural network) for predicting water and chlorophyll content in citrus leaves. The study suggested that both ML and DL give comparable results for predictions using UAV images. From the literature, it is apparent that both ML and DL can be applied for drone image segmentation. However,

it is not clear which technique, the traditional state-of-the-art machine learning or the advanced deep learning is better for the identification of the communities. Therefore, in this study, we applied both ML and DL techniques for vegetation classification of different vegetation communities on a raised bog wetland. Our study also demonstrates the pros and cons of both methods. It also gives a clear insight into both the techniques and their applicability for future studies on vegetation identification.

4.2 Study Area and Materials

This study concentrates on a small part of the bog located in West Clara bog (as shown in Figure 4.1).
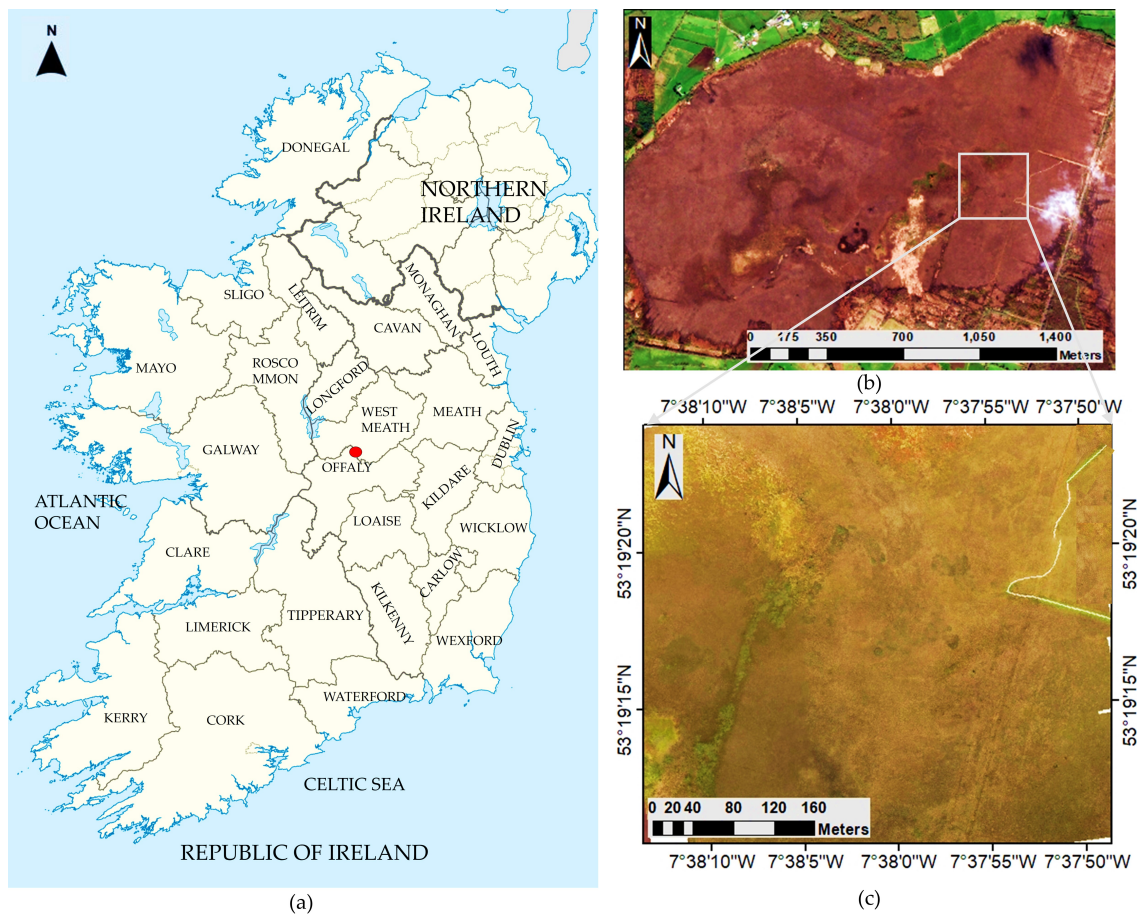
Figure 4.1: Study area: (a) map of Ireland (with the highlighted area: Clara Bog). (b) West of Clara Bog, County Offaly (with the highlighted area covered by drone). (c) Area covered by DJI Inspire 1^{TM} drone

The major ecotopes present in Clara bog are Central (C), Subcentral (SC), Submarginal (SM), Marginal (M), and Active flush (or flush) (AF). Out of all these ecotopes, the main focus is on the conservation of the active peat-forming areas (Bhatnagar et al. 2018, Bhatnagar et al. 2019, Bhatnagar et al. 2020a), which are considered to be C, SC, and AF ecotopes. More details on Clara bog vegetation can be found in Section 1.7.
For capturing high-resolution images, a DJI Inspire 1^{TM} drone was used. The camera used

with the drone was Zenmuse X3. It is an optical camera with 100–1600 ISO range (for photo) and 94° field of view (FOV). The lens is anti-distortion and autofocus (20mm of 35mm format equivalent). The aspect ratio, while clicking the images, was kept at 4:3. The images were captured on 21st April 2019 at around noon time. The highest temperature on the day was recorded at $19°C$. The height of the flight was $\approx 100m$, and the spatial resolution of the images captured was 1.8 cm. The drone mission was pre-loaded using Google maps in Pix4DCapture application to capture $\approx 8ha$ of the area using an iOS-12 device. The images were captured individually with 70% frontal and 80% sideways overlap at an average speed of 3 m/s. Figure 1c provides the drone imagery of the study area. For georeferencing, the drone imagery had geo-tags (lat-long locations) present in it. For better orientation, imagery was overlayed on high-resolution DigitalGlobe World Imagery (spatial resolution = 30cm) available as a base map in ArcMap v.10.6.1 (*ESRI, World 30cm Imagery* 2020). Using 'georeferencing' toolbox present in ArcMap, 3–4 ground control points (GCPs) were identified for every image, and projection was rectified to Geographic Coordinate System—World Geodetic System 84 (GCS WGS 84). In this study, C, SC, SM, M, and AF ecotopes were all captured using high-resolution drone imagery (Figure 4.2).
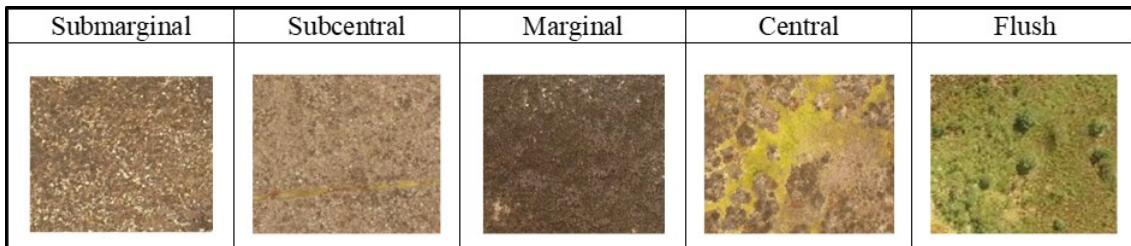
Figure 4.2: Ecotopes in Clara bog. Drone images, April 2019

The SM and SC ecotopes are highly homogenous and appear to be mixed throughout the bog (Bhatnagar et al. 2020a). These communities were therefore merged for the rest of the study. In total, around ≈ 75 images of dimension 3000×4000 were captured. Out of these images, 15 images were discarded due to differences in light intensity, motion blur, and camera tilt. The usable 60 images were divided into 70% training and 30% testing randomly, which is around 40 images for training and 20 images for testing. In order to have a correct idea of mapping accuracy, all the images were labelled for the four vegetation communities (M, SMSC, C, AF). For ML only a part of the labelled training data was required, whereas for DL fully labelled images were used. This is discussed further in Sections 3 and 4. For the creation of a training dataset, it is essential for all the images to have a similar intensity range. Depending on the lighting situation when the picture was taken, the colour properties may be changed, even though the textural properties remain unchanged. In a temperate climate like Ireland, this change in sunlight while capturing drone images is unavoidable. Therefore, going forward in future studies, the usage of colour correction techniques for drone images is recommended such that all the captured images can be used.

4.3   Segmentation Using Machine Learning

The segmentation of the images using machine learning techniques utilises combinations of intensity, colour, texture, and motion attributes to come up with hierarchical segments

(Shi & Malik 2000). The drone images used for this study have intensity and colour information. Although textural information is not present in the original image, textural features were subsequently calculated using the parameters mentioned in Table 1, (Feng et al. 2015). This was done by converting the RGB image into a grayscale image. The textural information presented in Table 4.1 was added as features along with the RGB layers. The entire computation of machine learning techniques and the steps described below was performed using MATLAB v.2019b using image processing toolbox (*MATLAB, The MathWorks* 2019).

Property	Description
Contrast	Intensity difference between pixels compared to its neighbour for the whole image (Tavares & Jorge 2015)
Correlation	Correlation of a pixel and its neighbour for the whole image (Schwenker et al. 2016)
Energy	Sum of squared elements in GLCM (Chai et al. 2011)
Homogeneity	Closeness of the distribution of pixels in the GLCM to its diagonal (Salem & Nasri 2009)
Mean	Mean of the area across the window
Variance	Variance of the area across the window
Entropy (e)	Statistical measure of randomness $e = -\sum(h \times \log_2(h))$ where h contains the normalised histogram counts
Range	Range of the area across the window (Wu et al. 2013)
Skewness (S)	Asymmetry of the data over the mean value (Mardia 1970) $S = E(p - \mu)^3/\sigma^3$, where μ is the mean of the pixel p, σ is the standard deviation of p, and E represents the expected value
Kurtosis (K)	Distribution to be prone to outliers (Mardia 1970); $K = E(p - \mu)^4/\sigma^4$

Table 4.1: Textural properties calculated using drone imagery

The segmentation technique used in this study, called graph cut, is based on max-flow min-cut (Stoer & Wagner 1997). This is done using posterior probabilities associated with every pixel for every class. In order to calculate the posterior probabilities, an initial classification of the drone images was carried out. Based on the texture and colour intensity, a total of 13 bands are used for further classification of the drone images. The type and choice of classifier used are discussed in the following subsection.

4.3.1 Choice of the ML Classifier

For efficient classification, the choice of the classifier is the most crucial decision that has to be made. Multiple studies have applied hyperspace based SVM (Ishida et al. 2018, Braun et al. 2010) for image classification. Other studies, like (Laliberte & Rango 2009), have used decision trees. Studies (Akar 2017, Meng et al. 2017) suggest that there is an advantage of using ensemble classifiers over other state-of-the-art classifiers. The most commonly used ensemble classifier consists of a tree model. The tree models are easy to understand and could be used for both classification and regression. There is no need for variable selection (since it is automatic) or variable transformation. They are robust to outliers and missing data, and particularly useful for large datasets.

In this study, in order to provide proper comparative analyses, the drone images captured

on 21st April 2019, were classified using multiple classifiers. The training dataset ($\approx 12k$ pixels from 40 images) was the input for all the classifiers. The classifiers were tested on model accuracy, misclassification cost (i.e., the total number of incorrectly identified pixels per 10,000 pixels), and training time (time taken by the classifier for training). The model accuracy for each ML model was calculated using 5-fold cross-validation for the entire 70% training dataset. This accuracy indicates the capability of the model to label the pixels correctly. The results (Table 4.2) describes all the classifiers and the corresponding accuracy metric. All the calculations were performed using MATLAB v.2019b (*MATLAB, The MathWorks* 2019).

The preliminary comparison was made using six classifiers, namely, decision trees (Friedl & Brodley 1997), naïve Bayes (Chen et al. 2007), discriminant analysis (Balakrishnama & Ganapathiraju 1998), SVM (Cortes & Vapnik 1995), k-nearest neighbour (KNN) (Laaksonen & Oja 1996), and random forest (RF) (Liaw & Wiener 2002). Based on the misclassification rate, model accuracy, and training time (see Table 4.2), RF was found to be best classifier.

Name	Parameter	Model Accuracy	Misclassification Cost	Training Time (s)
Decision trees	Max. no. of splits = 20; split criterion = Gini's diversity index	87.4	736	7.3
Discriminant analysis	Kernel = quadratic	89.4	618	8.6
Naïve Bayes	Kernel = Gaussian	78.3	1271	19.5
Support vector machine	Kernel = radial basis function (rbf) = 0.25	91.9	472	112.5
K nearest neighbour	No. of neighbours = 2; distance = Euclidean	91.0	528	378.8
Random forest	No. of trees (t) = 100 (1000 samples with repetition); total no. of splits = 5853	92.9	454	59.2

Table 4.2: Comparison of ML classification techniques

Random forest or bagging is a general-purpose procedure for reducing the variance of a predictive model. When applied on trees, the number of trees (t) is bootstrapped, each having a variance σ^2. In RF each tree can split on only a random subset of the samples (hence, the name). RF requires an attribute (sample) selection and a pruning method. Information gain ratio criterion (Ross 1993) and Gini Index (Breiman et al. 1984) are the most common attributes selection methodology. For this study, the Gini index criterion was used to decide the attributes. The Gini index (G) is given in Equation (1). Based on the value of G, the attribute was decided automatically.

$$G = \sum_{n} \sum_{i=1}^{N} (p_i \times (1 - p_i))_n \tag{4.3.1}$$

where p_i is the proportion of the pixel (i = 1 to N) belonging to a particular class n, i.e., it is the prior probability. A minimum of 10% of the entire ground truth image should be given as training and rest could be used for testing (Bhatnagar et al. 2020a). The samples were divided into 100 random subsets (with repetition), and for each tree, and the attributes (splitting criteria: which of the RGB bands) were decided using Equation (4.3.1). The final class selection for every pixel was made using majority voting. The workflow of the RF classifier is given in Figure 4.3.
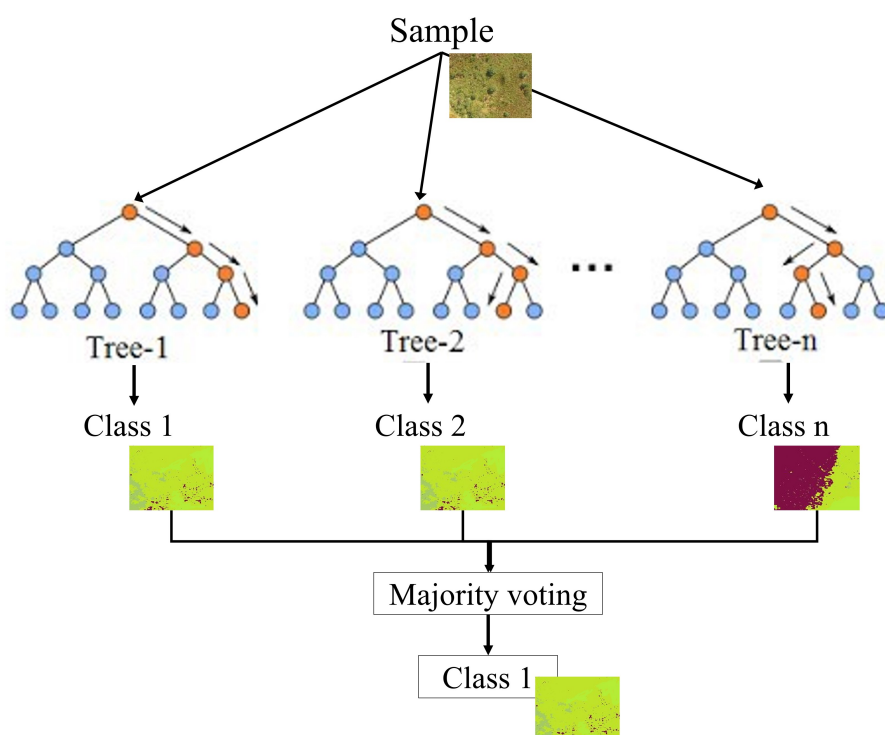
Figure 4.3: Workflow—random forest classifier

4.3.2 Segmentation

Once the drone images were classified, they were segmented using the maximum-a-posterior (energy minimisation) technique. The technique uses contextual (area) information to form proper segments from pixels. The formation of segments was done using a max-flow min-cut algorithm, commonly known as graph cut. This algorithm uses data cost and smoothness costs (Boykov & Kolmogorov 2004). The graph cut segmentation was performed in MATLAB v.2019b using MATLAB wrapper mex file function that enables the user to call C/C++ files (*Matlab Wrapper for Graph Cut* 2006).The method is described in detail in Section 3.3.1 The main difference for applying MVC algorithm here was the value of the smoothness factor (c). The $c > 5$ was chosen for the drone images. This can be compared to the optimum value of $c < 1$ when processing satellite images (Bhatnagar et al. 2020a). Therefore, it was seen that for a high resolution (1.8 cm), a higher value of c was required, whereas, when working with the 10 m spatial resolution from satellite images, a small value of c suffices. The results of the segmentation are further discussed in Section 4.5.1.

4.4 Segmentation Using Deep Learning

4.4.1 Parameters in Convolutional Neural Network

Convolution neural networks (CNNs or Covnets) have caused a step-change in pattern recognition progress. Here each neuron is connected to a local region of the input only, making the network faster and less prone to overfitting for a large dataset. Therefore, CNNs, when compared to traditional NNs, can have fewer parameters (Figure 4.4). More

details on the CNN architectures, and applications can be found in Pitié 2021.
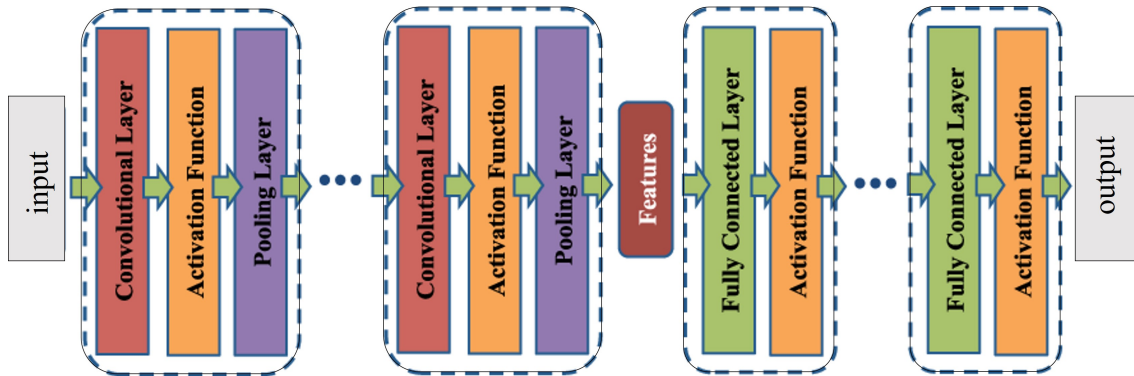
Figure 4.4: The general architecture of convolutional neural network (CNN)

Convolutional Layer

In CNN the convolutional layer is used instead of only fully connected layers. For visualising, convolution may look like a sliding window operation, but it is implemented as matrix multiplication.

Pooling Layer

The pooling layer downsamples the input by locally summarising the data in it. The two types of pooling are shown in Figure 4.5.
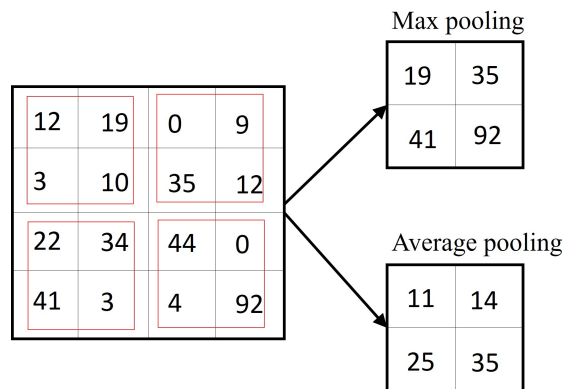
Figure 4.5: Types of pooling used in CNN

Of the two methods, max-pooling was used for this study, as it is a more efficient pooling technique (Masci et al. 2012).

Kernel Size

Kernels or the filters are used in order to down-sample the layers in CNN. It is preferable to use smaller kernels stacked on top of one another than using a large kernel (Li, Cai, Wang, Zhou, Feng & Chen 2014). Using smaller kernels decreases the number of parameters and also increases the nonlinearity. Therefore, in this study, the kernel of size 3×3 was used.

Stride

Stride defines by how much the kernel will move in the convolution layer. Here, stride =1 was used.

Padding

Padding is required to maintain the spatial resolution of the input image. The same padding was used in this study in order to maintain the same dimension between input and output.

Activation Function

The activation function (f(x)) defines the output for a given input. It also imparts non-linearity to the input.
Why do we need nonlinearity?
Combining linear functions yields a linear function; however, in order to compute more in-depth features, nonlinearity is required. With just linear functions, the model is no more expressive than a logistic regression model without any hidden layer. For this study, the ReLu activation function was used. Equation (4.4.1) describes the ReLu function.

$$f(x) = 0; x < 0; and f(x) = x; x \geq 0 \tag{4.4.1}$$

Softmax Classifier

Softmax Classifier is an activation function, typically used as the top layer (after a fully connected layer). It imparts probabilities of each input belonging to each output when there are more than two outputs. For the n number of classes, the Softmax activation (σ) can be defined by Equation (4.4.2).

$$\sigma(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{n} e^{x_k}}; j = 1, ..., n \tag{4.4.2}$$

Batch Normalisation

This layer normalises the hidden nodes before they are fed into an activation function.

Additional Parameters in CNN

An essential parameter in CNN is optimisation. Training a network can be considered to be an optimisation problem where the goal is to minimise the loss function. For faster and efficient processing, a subset of the data is taken one at a time, and therefore a stochastic gradient descent was used for optimisation in this study (Paine et al. 2013).
Another important parameter in CNN is regularisation. Regularisation can be done by adding a weight penalty term to the loss function (Equation (4.4.3)).

$$Loss = Loss + weight penalty(w) \tag{4.4.3}$$

L2 or ridge regularisation leads to the formation of small weights (Han et al. 2015). Additionally, L2 regularisation never causes a degradation in performance, even with the addition of kernels (Van Den Doel et al. 2013). Therefore, L2 regularisation was used in

CNNs for this study. For a given input x and its corresponding output $\hat{x}$ the regularisation function is given in Equation (4.4.4).

$$Loss = \sum_i (x_i - \hat{x}_i)^2 + \alpha \sum_i w_i^2 \qquad (4.4.4)$$

A third, important parameter for CNN architecture is the learning rate (LR). The LR is defined as the rate at which the weights are updated during the training of the network. An initial LR of 0.05 was used for this study.

Popular CNN Models

CNN models are formed using the combinations of parameters mentioned in the above sub-sections. The combinations of layers and the type of parameters used are often application-based and applied to solve a bigger problem. In this study, VGG16 (Simonyan & Zisserman 2014) and ResNet50 (He et al. 2016) were applied based on the work done by (Cheng et al. 2019, Qassim et al. 2018).

4.4.2 CNN for Semantic Segmentation

Semantic segmentation is a process of assigning a label to each pixel in an image such that pixels with the same label are connected via some visual or semantic property (Ghosh et al. 2019). In order to carry out semantic segmentation, the spatial information needs to be retained. Hence no fully connected layers are used, which is why they are called fully convolutional networks.

Moving from a fully connected to a fully convolution network

In the case of segmentation, an output layer is a group of 2D probability-maps of each pixel belonging to each class. These are known as score maps. The score maps are coarse as throughout the network; the information (image) has been down-sampled to absorb minute details. Therefore, to make the output compatible with the input in size, up-sampling is required.

Up-sampling can be done using either bilinear interpolation or cubic interpolation (or similar techniques). One way of up-sampling is via skip-connections or shortcut connections. Three key fully connected models, SegNet (Badrinarayanan et al. 2017), UNet (Ronneberger et al. 2015), and Pyramid Scene Parsing Network (PSPNet) (Zhao et al. 2017) are used in this study. A brief description of the models is given in the following subsections.

SegNet Model

SegNet works with encoder-decoder architecture, followed by a pixel-wise classification layer for multiple classes. Encoders extract the most relevant features from the given input. The decoder uses the information from encoder to up-sample the output (Figure 4.6).
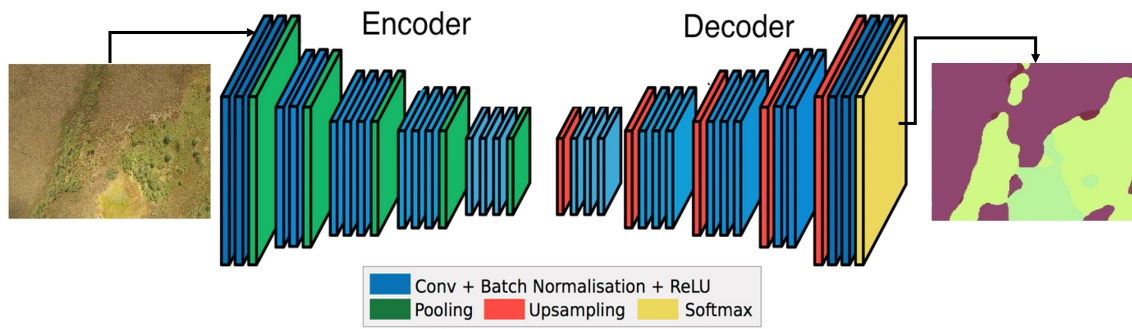
Figure 4.6: SegNet architecture for semantic segmentation for bog-ecotope semantic segmentation

The up-sampling technique used by the decoder is known as max-unpooling. Max-unpooling eliminates the need for learning to up-sample (as was required in skip-connections) as shown in Figure 4.7. Based on the location of the maximum value, the max-pooled values are placed. The remainder of the matrix is loaded with zeros. Convolution is done using any CNN models (as discussed in Section 4.4.1) using this layer.
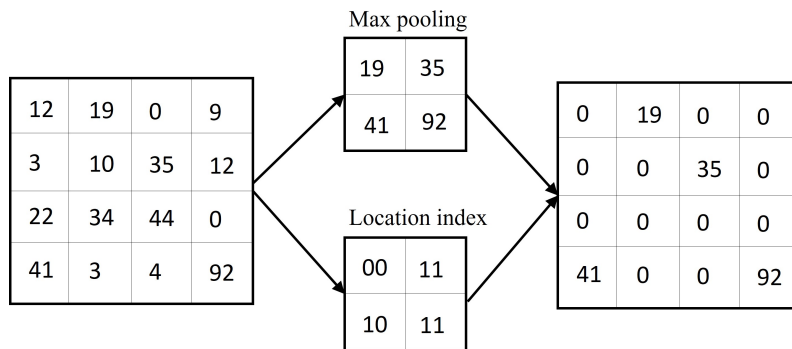
Figure 4.7: Pooling and unpooling for semantic segmentation

UNet Model

UNet network carries out the transpose convolution (encoder-decoder) and also uses skip connections (Figure 4.8).
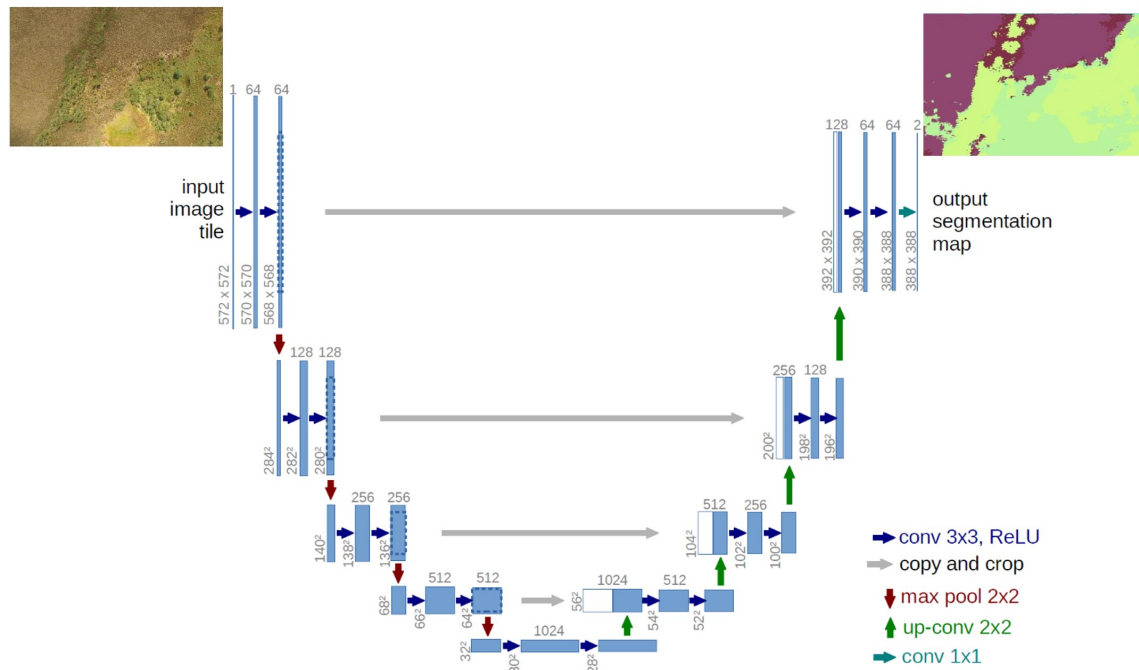
Figure 4.8: UNet architecture for semantic segmentation for bog-ecotope semantic segmentation

At every layer in the decoder side, the network finds a corresponding feature map (of the same size) from the encoder and adds (11 convolution) to the score map. This way, the size of the feature map is always in sync. Due to its architecture and depth, UNet is most widely used in biomedical image analysis.

PSPNet Model

PSPNet stands for Pyramid Scene Parsing Network. This network incorporates the scene and global features for scene parsing and semantic segmentation as shown in Figure 4.9.
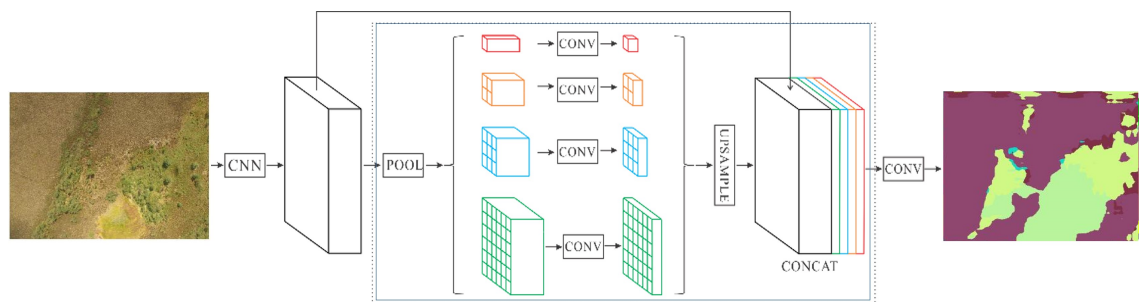
Figure 4.9: Pyramid Scene Parsing Network (PSPNet) architecture for semantic segmentation for bog-ecotope semantic segmentation

The pyramid pooling module in PSPNet fuses the features in four scales: coarse (1×1), 2×2, 3×3, and 6×6. The up-sampling done is a bilinear interpolation, and all the features are concatenated as the final pyramid pooling global feature (Zhao et al. 2017).

101

4.4.3 Methodology for the comparison between CNN models for the case study on raised bog drone images

Using the drone images captured on 21st April 2019, semantic segmentation using various CNN architectures was applied to identify and label the ecotopes present on Clara Bog. The entire computation was performed in python v3.7 (*Python Language Reference, version 3.7* 2020) using GPU (NVIDIA Tesla K40C 12GB CUDA), accessed remotely from trinity college high performing computer (TCHPC), and partly on google virtual machine (Tesla K40C 12GB). The study uses the repository in (Divamgupta 2019).

Training data preparation

In order to smoothly run the semantic segmentation, the preparation of training data was done as follows

1. Forty drone images were manually labelled using MATLAB Image Labeler app (*MATLAB, The MathWorks* 2019).

2. The labels (in .mat format) were converted into JPG.

3. The images and labels were resized in order to use the GPU memory efficiently and to speed up the process. For resizing, the images were shrunk in the order of 2^n such that the classes were clearly distinguishable. The resizing of the images was done using a bilinear interpolation technique.

4. The images were resized from 3000×4000 to $512 \times 1024(2^9 \times 2^{10})$ for further use. The size of the image is kept rectangular in order to maintain the aspect ratio of the original drone imagery. The ratio can be decided with respect to the application. For this study, to have a fair comparison between ML and DL methods, the size of the imagery was not reduced to smaller patches.
Alternatively, patches of the same size ($2^9 \times 2^{10}$) can be extracted with overlapping. For this study, the small patches did not cover all the ecotopes. In a single patch, at maximum, only two ecotope classes were covered. This is due to the large size of the raised bog in the application. Therefore, to incorporate the maximum number of ecotope classes in a single image and to avoid any information loss, resizing of the images was done (instead of extracting the patches).

5. After reshaping, the images were renamed such that the images and their corresponding labels can be identified.

Steps (2–5) were repeated for all 40 images having all four ecotope classes mentioned in step 1. The final training data consisted of 40 images (both RGB and labelled) of the size $2^9 \times 2^{10}$, which was fed to the CNN models described in the next subsection. The testing was carried out on the rest of the 20 images.

Models used for semantic segmentation

The models were created using a base network (tested on ImageNet) along with a segmentation architecture. The max number of *epochs* = 100, and images were shuffled at every epoch and a mini-batch size of 64. The loss between the labels given by the model and the actual (training) label at every epoch was calculated using cross-entropy loss described in

Equation (4.4.5).

$$crossentropy = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n}(\hat{x}_i log(x_{i,j})) + (1 - x_{i,j})log(1 - x_{i,j}) \tag{4.4.5}$$

where N is the total number of pixels, n is the total number of classes, x is the training label (input), and $\hat{x}$ is the output label as predicted by the models. Instead of training the network from scratch, one of the most common techniques is to use a pre-trained network. Pre-trained weights speed up the convergence process (to reach local minima, i.e., to overall minimise the loss). It is also considered better than random initialisation. For the four models listed below, 'ImageNet' dataset (Russakovsky et al. 2015) was used to initialise the weights. Other details are mentioned in detail in (Divamgupta 2019). The architecture for these models is shown in Figure 4.10.

- VGG16 base model with SegNet architecture.

- ResNet50 base model with SegNet.

- VGG16 with UNet.

- ResNet50 with UNet
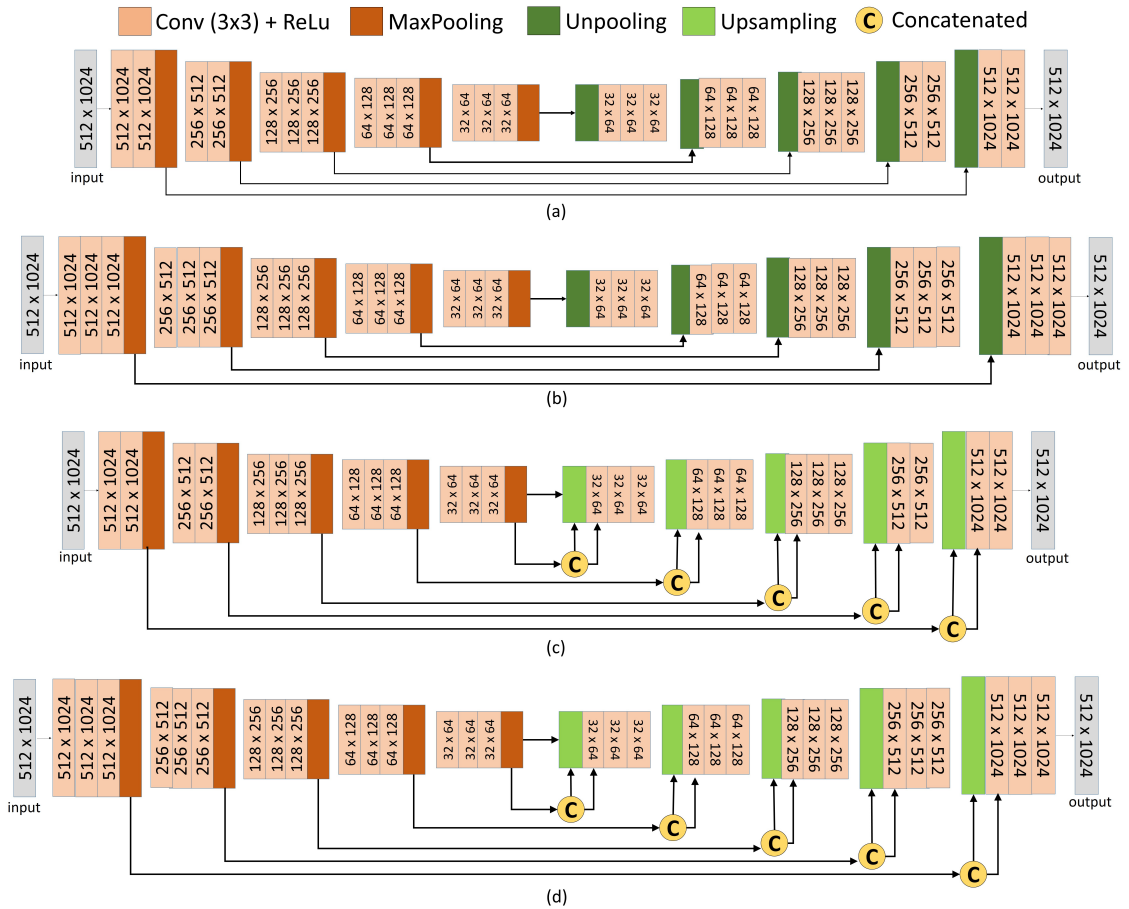
Figure 4.10: CNN Network architecture. (a) VGG16 + SegNet, (b) ResNet50 + SegNet, (c) VGG16 + UNet, (d) ResNet50 + UNet

Figure 4.10 a,b represents the SegNet architecture with VGG16 and ResNet50 as the base model, respectively. The left-hand side is the encoder, which has five blocks, and the layers are from the original VGG16 and ResNet models. Figure 10c,d represents the UNet architecture with VGG16 and ResNet50 models, respectively. The network uses the original layers from the VGG16, ResNet50, with the UNet architecture.

For a specific task of semantic segmentation, dedicated segmentation based dataset was also used for initialising the weights. For the PspNet, the pre-training was done using ADE20K data (Zhao et al. 2017), and Cityscapes dataset (Cordts et al. 2016). The ADE20K dataset has 21,200 images of various day to day scenes. The Cityscapes data contains images taken from video frames ($\approx 20,000$ coarse images) from 50 cities taken in spring, summer, and fall seasons. The models used are listed below, and the layers and architecture are described in Figure 4.9.

- PspNet trained on ADE 20K dataset.

- PspNet trained on Cityscapes dataset.

4.5   Results

Figure 4.11 depicts the segmentation results from both ML and DL techniques for a drone image (sized 512×1024) taken of Clara bog (the segmented images were combined to form an orthomosaic). The segmentation was carried out for four ecotope classes present in the drone image captured in the spring season. The accuracy and results are further discussed in this section.
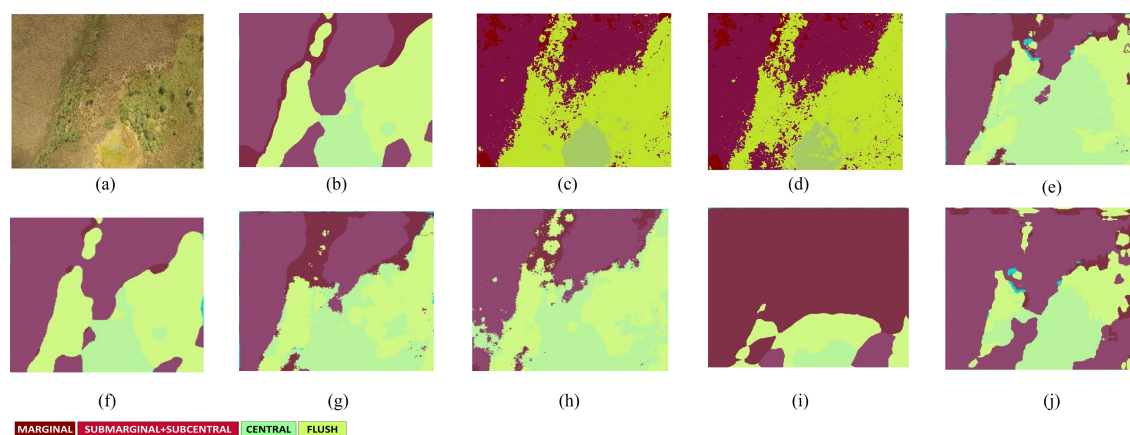
Figure 4.11: Segmentation results. (a) Drone image, (b) ground truth labelled image, (c) machine learning (ML) (random forest (RF) + Graph cut) segmentation using RGB features, (d) ML (RF + Graph cut) segmentation using RGB and textural features, (e) deep learning (DL) semantic segmentation using SegNet and VGG 16 model, (f) DL semantic segmentation using SegNet and ResNet50 model, (g) DL semantic segmentation using UNet with VGG16 model, (h) DL semantic segmentation using UNet with ResNet50 model, (i) DL semantic segmentation using PSPNet (Cityscapes), (j) DL semantic segmentation using PSPNet (ADE 20k)

4.5.1 Machine Learning

As discussed in Section 4.3, the ML classifiers were tested for model accuracy (5-fold validation), misclassification cost, and training time. Table 2 depicts the metric calculated over the entire 70% training data (as discussed in Section 4.3).

RF was chosen to be the best performing classifier, and further segmentation using Graph-cut algorithm was performed using the results from RF. The segmentation is a post-classification area based smoothing process. The final segmented image was checked against a fully manually labelled image to give overall accuracy (OA). The OA is the ratio of true positives (TP) with a total number of pixels (Equation (4.5.1)).

$$OA = \frac{TP{+}TN}{TP + FP + FN + TN} \tag{4.5.1}$$

where, TP = true positives, FP = false positives, FN = false negatives, and TN = true negatives. This was done for visible bands (RGB) and RGB + textural features. For proper comparison between ML and DL, the image was resampled from its original size (3000×4000) to a smaller scale (512×1024). The resampling of the image was done using bilinear interpolation (Andrews & Patterson 1976). Table 4.3 depicts the accuracies obtained by using a random forest classifier along with graph cut segmentation for both the sizes of the image. Since, the image used in segmentation using DL techniques is also resized, for an accurate comparison, the resized image (512×1024) was used in the further analysis.

	RGB Features	RGB + Textural Features
Original size (3000×4000)	83.3%	85.1%
Resized (512×1024)	82.9%	84.8%

Table 4.3: Segmentation accuracies using ML

As can be seen from Figure 4.11 b,c, there is not much difference in the segmentation using RGB with or without textural features. However, the textural features do add extra information and are known to be highly useful when there is a terrain variation in the scene. However, in this application, where the ecotopes under consideration are low-lying, homogenous communities, the addition of textural features did not improve accuracy very significantly—the OA only increasing by approximately 2%.

4.5.2 Deep Learning

The semantic segmentation using CNNs was performed for 100 epochs. The LR was decreased by a factor of 10 each time a model's accuracy was saturated. The overall accuracy performed on testing data (OA is also calculated for testing data) of all the models is shown in Figure 4.12.
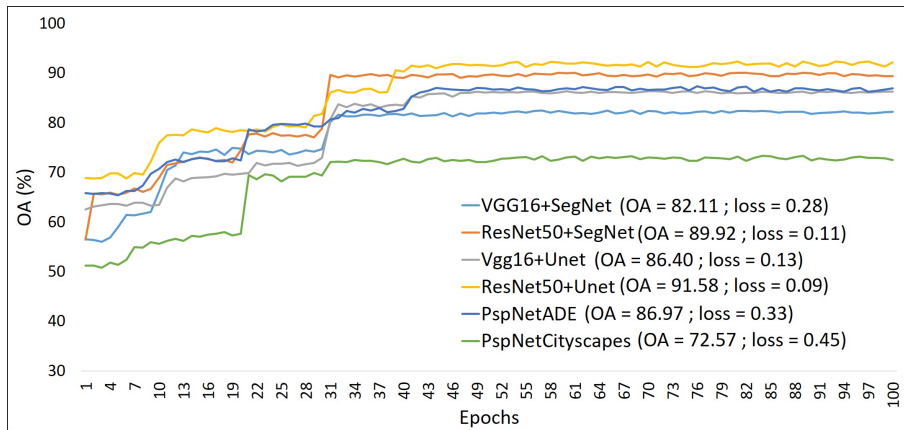
Figure 4.12: Overall accuracy over 100 epochs used for all CNN architectures for semantic segmentation of ecotopes in Clara Bog

There is a jump of an average $\approx 32\%$ in OA from the first to last epoch, with the PspNet model and ResNet50+SegNet showing the maximum increase in OA ($\approx 30\%, 25\%$ respectively). The cross-entropy loss decreased by an average of $\approx 28\%$ for the CNN models under consideration. This decrease happens by reducing the LR. Although accurate, a detailed analysis of per-class accuracy is required to make an informed decision about the best CNN architecture for the segmentation for this particular application in the identification of raised bog vegetation ecotopes. The per-class analysis is done to make sure there is no overfitting. As seen from Figure 4.11(i), a model can lead to overfitting, giving sufficient accuracy but incorrect classification.

Table 4.4 describes the confusion matrix for every community, and both ML and DL algorithms, which is discussed further in Section 4.6. Other accuracy checking parameters like Precision, Recall, and F1-score were also calculated for every community (ecotope) under consideration. Equations (4.5.2 - 4.5.4) give the formula to calculate the above-stated accuracy parameters.

$$Precision = \frac{TP}{TP + FP} \tag{4.5.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.5.3}$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.5.4}$$

	RF (RGB)				RF (RGB + TEXTURAL)				SEGNET + VGG16				SEGNET + RESNET50			
	M	SMSC	C	AF	M	SMSC	C	AF	M	SMSC	C	AF	M	SMSC	C	AF
M	58,405	1012	988	36,321	59,781	1598	1002	38,241	43,872	8854	2500	71,005	62,870	1631	835	16,360
SMSC	734	155,583	4979	2033	600	188,296	4608	587	7952	122,544	15,691	9514	3000	162,651	3005	2639
C	328	3862	77,939	44,321	256	4150	83,930	34,658	2831	18,529	77,369	73,108	967	4895	98,584	28,330
AF	38,010	3211	43,509	142,707	39,023	3079	32,584	112,589	65,896	21,251	64,211	99,833	18,470	14,110	10,358	148,383
Precision	0.59	0.94	0.61	0.62	0.59	0.96	0.68	0.66	0.35	0.79	0.45	0.40	0.77	0.95	0.74	0.78
Recall	0.60	0.95	0.61	0.63	0.60	0.95	0.69	0.66	0.36	0.72	0.48	0.39	0.74	0.89	0.87	0.76
F1 score	0.60	0.94	0.61	0.62	0.60	0.95	0.68	0.66	0.36	0.75	0.47	0.40	0.75	0.92	0.80	0.77

	UNET + VGG16				UNET + RESENET50				PSPNET ADE20K				PSPNET CITYSCAPES			
	M	SMSC	C	AF	M	SMSC	C	AF	M	SMSC	C	AF	M	SMSC	C	AF
M	82,589	9510	3258	36,951	73,897	1008	258	3371	36,351	9822	631	5311	128,890	78,353	36,118	63,001
SMSC	10,254	146,933	9800	19,759	4096	152,363	3690	15,892	15,200	210,052	6323	28,200	73,570	107,781	2988	4820
C	4523	12,967	96,582	35,489	982	5183	90,258	28,105	987	7921	96,587	3715	38,562	5815	32,510	12,377
AF	32,563	19,638	34,822	83,417	5101	14,852	22,110	155,446	3074	21,520	5300	127,296	58,360	7826	13,524	57,450
Precision	0.62	0.79	0.65	0.49	0.94	0.87	0.72	0.79	0.70	0.81	0.88	0.81	0.42	0.57	0.36	0.42
Recall	0.64	0.78	0.65	0.47	0.88	0.88	0.78	0.77	0.65	0.84	0.89	0.77	0.43	0.54	0.38	0.42
F1 score	0.63	0.78	0.66	0.48	0.91	0.87	0.75	0.78	0.67	0.83	0.89	0.79	0.43	0.55	0.37	0.42

Table 4.4: Segmentation accuracies using ML

107

4.6 Discussion

The study describes methods to map vegetation communities in a raised bog 'Clara Bog' located in Ireland using drone images from DJI Inspire 1^{TM} drone captured during the spring season. The size of the images were 3000×4000, and 40 images were used as training. Furthermore, both ML, DL algorithms were tested for the rest of the 20 images. The study shows that high-resolution (1.8cm) RGB images are adequate for mapping vegetation communities. However, a key challenge associated with RGB images is the change in intensity due to sunlight conditions, particularly in a temperate climate like Ireland, where sunlight levels are rarely constant for long. Therefore, in this study, all the images with significantly different light conditions were removed. The use of a colour correction technique could be a possible solution to this problem, which is a domain yet to be explored. Similarly, the addition of textural properties does create the challenge of increasing the computations (time and complexity). The segmentation is done using 13 features instead of three, thereby being more computationally expensive.

Initially, a comparative analysis of the state-of-the-art classifiers was performed (Table 4.2). It was seen that the RF ensemble classifier outperformed the other classifiers. The RF classifier uses bootstrapping for forming multiple trees leading to reduced possibilities of overfitting of the data. The SVM classifier with RBF kernel had similar accuracy and misclassification cost as RF, but with twice the training time. Hence, RF was deemed to be the best choice for drone image classification with model accuracy of 92%. As pixel-based segmentation often fails to take the contextual (area-based) information into account; therefore, to form segments based on area, graph cut segmentation was subsequently applied. Out of the 40 training images, only a part of labelled pixels ($\approx 12k$) was input to the ML model. The entire processing time of ML segmentation was ≈ 30 min.

This was done for both RGB and RGB + textural images. The images were resampled to $2^9 \times 2^{10}$ for proper comparison with deep learning algorithms (discussed later). It has to be noted that the aspect ratio of the imagery was maintained while resampling it. This was done mainly to keep the textural properties intact. The authors of (Tavares & Jorge 2015) explain that in order to capture textural properties, the size of the image/sliding window should be carefully chosen. Therefore, a decrease in the size of the image (or change in aspect ratio) can lead to a change in textural properties. Table 4.3 shows that the resampling using bilinear interpolation did not make a big difference in the OA. The resampled image with textural properties performs comparably to the original image. The OA with textural properties is also comparable to OA with just RGB for this application with a low-lying homogenous area of interest. Overall, the textural properties perform the best segmentation with both the original-sized image and the rescaled image.

From Figure 4.11 c,d, it can be seen that using textural properties, the ecotopes like SMSC and AF are differentiated better (see Table 4.4). Likewise, from Table 4, it can be seen TP for the C ecotope increases with the addition of textural properties but decreases for the AF ecotope. The decrease in misclassified pixels (FP, FN) between SMSC and AF has led to an increase in precision and recall for the SMSC ecotope. There is a definite increase in accuracy for the C and AF ecotopes by using textural features, whereas, the SMSC and M ecotopes are identified with similar precision, recall, and f1 score values for both the images.

The deep learning technique used for segmentation is semantic segmentation using CNN models. In this study, six different models were tested for the semantic segmentation to identify the different bog-ecotopes. The training data for the CNN models consisted of 40 images containing all the ecotopes in different orientations and lighting (brightness). The

size of the training data is a notable factor in this study, as for many applications, 100s of images are more usually required for training such CNN models. This study demonstrates the usage of minimum labelled training images for attaining the segmentation, given that 40 images seemed to be sufficient for this application as the weights were initialised using ImageNet dataset having 1000 different classes. This reduces the dependence on the extensive training dataset and also is faster (Liu et al. 2008). All these 40 images were resized to $2^9 \times 2^{10}$ for efficiently performing semantic segmentation. For an application involving a prominent area such as this, the classes are also sparsely located. Therefore, cropping or extracting patches from the images was leading to a reduction in classes (ecotopes) covered in an image. In order to make sure that the model identifies all the ecotopes, the images were resized. Nevertheless, for an application where the classes are located close enough (spatially), cropping/extracting patches can be a viable option.

The algorithms were run for 100 epochs, after which the accuracy was becoming saturated. The computation time was 700 min per model for 100 epochs. It was decided not to increase the number of epochs as it may lead to overfitting of the model (Grm et al. 2017). The LR was decreased with epochs when OA saturated. This decrease leads to faster convergence and an increase in accuracy. Without decreasing the LR, if the same LR is continued, one may still get high accuracy, but it would require a massive number of epochs; therefore, it is not recommended. There is an apparent increase in accuracy using DL methods when compared to ML methods. At the end of the epochs, it is clear that SegNet and UNet architecture with ResNet50 yield the best results for the semantic segmentation of bog-ecotopes. In comparison, the VGG16 base model has led to the over-classification of ecotopes such as M, AF. The VGG model has been shown to be effective when there is noise in the data but does not perform well when the brightness of the images changed (Yim et al. 2017). This explains the low accuracy of the model, as the images had different lighting due to variable weather conditions. Figure 4.11 (e-j) depicts the DL segmentation results. It can be seen that the segmentation using SegNet and UNet is similar for ecotopes like SMSC and C, but is different for AF and M ecotopes.

The study also demonstrates the use of transfer learning by using a segmentation specific pre-trained PspNet model. This model was pre-trained using ADE20K and cityscapes image set instead of widely applicable ImageNet. In our application, the usage of these segmentation datasets was not successful as the weights were calculated for a specific task of segmenting areas of traffic, cars, houses, pavements, etc. Additionally, due to the uniqueness of these communities, the weights transferred from the pre-trained models were not accurate. In order to use transfer learning, the model selected should be pre-trained using similar categories as the application.

For making the final decision of the best CNN architecture, the accuracy parameters for every ecotope were considered. Table 4.4 shows that the SMSC ecotope is identified quite well using all the CNN models, with the exception of the PspNet model pre-trained with cityscapes images. Using the base model ResNet50, the ecologically important, peat-forming communities (the SMSC, and C ecotopes) are better identified using SegNet than UNet. Using PspNet (ADE20K), the C ecotope was identified the best, although the OA of the model is low. Therefore, taking into consideration the OA, precision, recall, and f1-score of all the communities, SegNet architecture with the ResNet50 base model appears to be the best choice for drone image segmentation in relation to identifying raised bog vegetation types.

The best OA recorded from ML was 85%, and from DL was 91%. However, the most appropriate technique for this study was not decided just on the basis of OA. For applying the technique to new applications, other parameters cannot be ignored. For example, a

lot more training data was required for using DL as compared to ML. Similarly, time and hardware also play a significant role in deciding the best technique. Table 4.5 summarises the essential pros and cons of the two techniques.

PROS	CONS
Machine Learning	
1. Very fast training ($\approx$30 min per model)	1.Cannot be applied globally, i.e., does not work on augmented data.
2. Needs less training data ($\approx$12k labelled pixels in 40 images)	2. In order to re-use the same model,input has to be consistent with original images or to be modified.
3. No need of size alteration; accepts the input of any size	3. Can require parameterisation, manual handling.
4. No need for HPCs or GPUs; works well with CPU.	4. Textural and additional features have to be input separately; does not learn by itself.
5. Provides good accuracy (OA = 85%)	
Deep Learning	
1. Can be applied globally for multiple applications and works well with augmented data.	1. Slower training process ($\approx$700 min per model).
2. Works well with low resolution imagery.	2. Needs a considerable amount of training data or requires a pre-trained model ($\approx 48 \times 10^6$ labelled pixels in 40 images).
3. Learns textural features on its own, require no additional inputs.	3. Requires alteration of images with respect to size for faster computation and analysis.
4. Provides good results (OA = 91%)	4. It is recommended the use of GPUs with good RAM for running CNN efficiently.

Table 4.5: Pros and cons of ML vs. DL for mapping ecotopes, Clara bog

It is clear that there are many pros and cons of both techniques, as described in this study. The main idea behind using remote sensing techniques is to reduce the amount of manual fieldwork that is required for monitoring the wetlands. This includes minimising the training data given as input to the classifiers. Additionally, the idea is to automate the process in the simplest way possible, given that the availability of high performing computers or GPUs cannot always be guaranteed, in order to optimise the speed. Keeping in mind the above requisites, the ML technique is the clear choice for our application. Whilst DL techniques can be used once there is enough labelled data created from all the wetlands such that all the species are covered, in the case of a new wetland, which contains new species to be mapped, a clear indication of the species (with full coverage) is required. Therefore, DL is more advantageous to use for more global or applied applications, whereas for a more specific application such as this where not enough training data is available, ML can produce accuracies almost comparable to the DL.

Finally, the drone images are very high resolution and can be captured at any given time. However, although practical, drone image capturing does have certain limitations. The first limitation of using drones is the length of the battery life. For example, on average, the drone (such as DJI Inspire 1) the battery will only last approximately 15 min and so to cover a large area, many batteries are required. Therefore, in the future, in order to make the process cost-effective, drone images can be used in conjunction with the freely available satellite images. Satellite images give excellent coverage and proper temporal resolution meaning that the usage of drones and satellites together should be advantageous.

4.7 Conclusion

This study aimed at providing mapping of vegetation in wetlands using image segmentation. Using ML, a total of six different state-of-the-art pixel-based classifiers were compared, out of which, the best ML algorithm for the given dataset was shown to be the RF classifier (model accuracy = 92.9%). For ML image segmentation, RF classifier was

used with maximum a-posteriori graph cut segmentation. It was seen that accuracy is improved by $\approx 2\%$ after addition of textural features (OA = 85.1%) when compared to the original RGB image (83.3%), and ecologically important communities such as central ecotope were mapped better. Apart from ML, the study also describes image segmentation or 'semantic segmentation' using DL methods. A detailed account on the selection of variables for the DL segmentation was presented. The study was done on a combination of six architectures and base models. For the given dataset, the ResNet50 base model with both UNet (OA = 91.5%) and SegNet (OA = 89.9%) architecture performs very well. ResNet50+SegNet model was deemed best, as it was able to identify complex vegetation communities, such as SMSC ecotope better. It was seen that for mapping ecotopes in a raised bog, transfer of initial weights from wide-ranged ImageNet outperforms the segmentation-specific datasets like ADE20K or Cityscapes.

Overall, the accuracy of the DL was 4% higher than the ML methods. Additionally, the DL method does not require any colour correction or the addition of extra textural features. However, DL requires a large amount of initial labelled training data ($\approx 48 \times 10^6$ pixels). On the other hand, the ML algorithm requires much less training data ($\approx 12,000$ labelled pixels) and is much faster (≈ 30 times) when compared to CNNs. Therefore, in retrospect, for such a specific application as the wetland mapping application, it was considered that the ML approach is more suitable. This would be particularly useful for any un-surveyed wetland, where the minimum amount of information of the vegetation communities is required to produce accurate maps.

Chapter 5

A Nested Drone-Satellite Approach

"Sometimes the questions are complicated and the answers are simple."

— Carl Sagan

Contents

This chapter corresponds to the paper that has been submitted on 22 July 2020 in "ISPRS Journal of Photogrammetry and Remote Sensing" and currently is under review - ongoing corrections [4].

5.1   Introduction

The advent of Remote Sensing (RS) has made it possible to collect information and data over large areas at regular time intervals. Satellites such as the Landsat and Sentinel

series provide open-access, global land cover information. The Sentinel-2 (S2) satellite, introduced by European Space Agency (ESA) in 2015, provides data at 10 m, 20 m, and 60 m resolution at an interval of 5 days. This information can be used for a variety of purposes, including the monitoring of ecosystems (such as wetlands) and thus providing essential information to support wetland management for conservation or restoration.

The conservation of wetland ecosystems is critical both ecologically and economically, and remote sensing along with machine learning algorithms have been extensively applied for monitoring, mapping, and conservation purposes (Kampichler et al. 2010, Kuenzer et al. 2011, Murray et al. 2018). For mapping, such ecosystems, a study by Oostdijk et al. 2018, used high-resolution imagery. Apart from a single sensor, the fusion of satellite data from different sensors is also widely used to monitor the health of wetland ecosystems; for example, Montgomery et al. 2019, Mahdianpari et al. 2019 and Kaplan & Avdan 2018 have demonstrated the fusion of microwave and optical imagery with high accuracy. Open-source satellite imagery has been used to map a large area of Canadian wetlands (Amani et al. 2019, Mahdianpari et al. 2017); for example, Amani et al. 2019 illustrate mapping wetlands over a large area using 30,000 Landsat images.

In contrast, there are studies that have demonstrated the ability of remote sensing to distinguish between vegetation communities over much smaller areas, for example, the studies on small wetlands using multi-date S2 images in Ireland (Bhatnagar et al. 2018, Bhatnagar et al. 2020a). While S2 imagery can be beneficial at a global and regional scale, its 10 m resolution often exposes limitations when looking at spatially small wetland ecosystems with heterogeneous vegetation communities. The classification approach used in such cases is highly dependent on the availability and accuracy of ground truth data (Carlotto 2009), such as detailed ecological maps. Hence, for such ecosystems, the collection of ground truth information is also crucial.

Drones can capture images at a very high spatial resolution of a few cms, which can be used to identify small vegetation communities. The integration of the drone and satellite images for vegetation monitoring has recently been utilised for different applications. Studies by Murugan et al. 2017 and Agarwal et al. 2018 have used drone and satellite data for high precision agricultural applications for a detailed assessment of sugarcane production areas. Equally, Szantoi et al. 2017 have used a drone and Landsat images for identifying orangutan habitats such as reforested area, bare areas, forests, oil palm areas, etc. Other studies likeRäsänen & Virtanen 2019 have also deployed UAVs for the creation of ground truth maps, concluding that the UAV provides better data than satellite data for treeless vegetation mapping. For invasive woody species identification, Kattenborn et al. 2019 used drone imagery for preparing reference maps, using a combination of colour, texture, structure, and hyperspectral properties with microwave and optical data from Sentinel-1 and Sentinel-2, respectively. It suggests that the maximum mapping accuracy was achieved using a combination of hyperspectral, texture, and structure features. Gray et al. 2018 used spectral, textural, and LiDAR features to create coastal cover change maps. They confirmed the usage of drone imagery to be comparable with field-based manual mapping and hence should be explored. While capturing drone images, the weather conditions can significantly change the colour intensity or variations in an image. The difference in exposure and altitude of the sun leads to a change in the appearance of the communities captured by the RGB sensor. This variation has not been addressed in existing literature combining drone and satellite imagery, and consequently, it has been proven to be challenging to identify communities correctly (Åhlén 2005). Hence, in this study, a colour correction technique has been proposed as a pre-processing step to improve the consistency of the training data used for mapping the wetland. In this study, we pro-

pose a method to enrich training data used for satellite image classification by utilising drone images. Small areas of the wetland containing parts of all communities of interest were captured using the drone to provide high-resolution training data for comparative lower-resolution satellite images to improve the classification. This reduces the cost of field surveys and allows regular, automated, remote monitoring of the wetlands. The drone maps, provide the new ground truth for some parts of the wetlands for mapping the entire area using satellite imagery. The process can be repeated frequently due to low time and resource requirements leading to the development of seasonal maps. To create the annual map, an optimal multi-season labelling using a majority voting method was done for all the pixels. This automated procedure of using drone and satellite imagery together for producing temporal maps of the ecosystem is applicable for all wetland types and has been applied to a raised bog located in the midlands of Ireland.

5.2   Methodology

Ecological field surveys, carried out manually to define ground truth maps, inevitably contain some errors due to misclassifications and interpolations between quadrats. Such ground surveys take multiple days to cover a large land-area in comparison to drone surveys, which can be far more time and resource-efficient. Our proposed method utilises partial drone surveys of the region of interest to be used as training data to improve the classification of satellite image classification. This complex process has been described in the flow diagram of the proposed framework, presented in Figure 5.1. The details of the steps mentioned in the figure are discussed in detail in the following sub-sections.
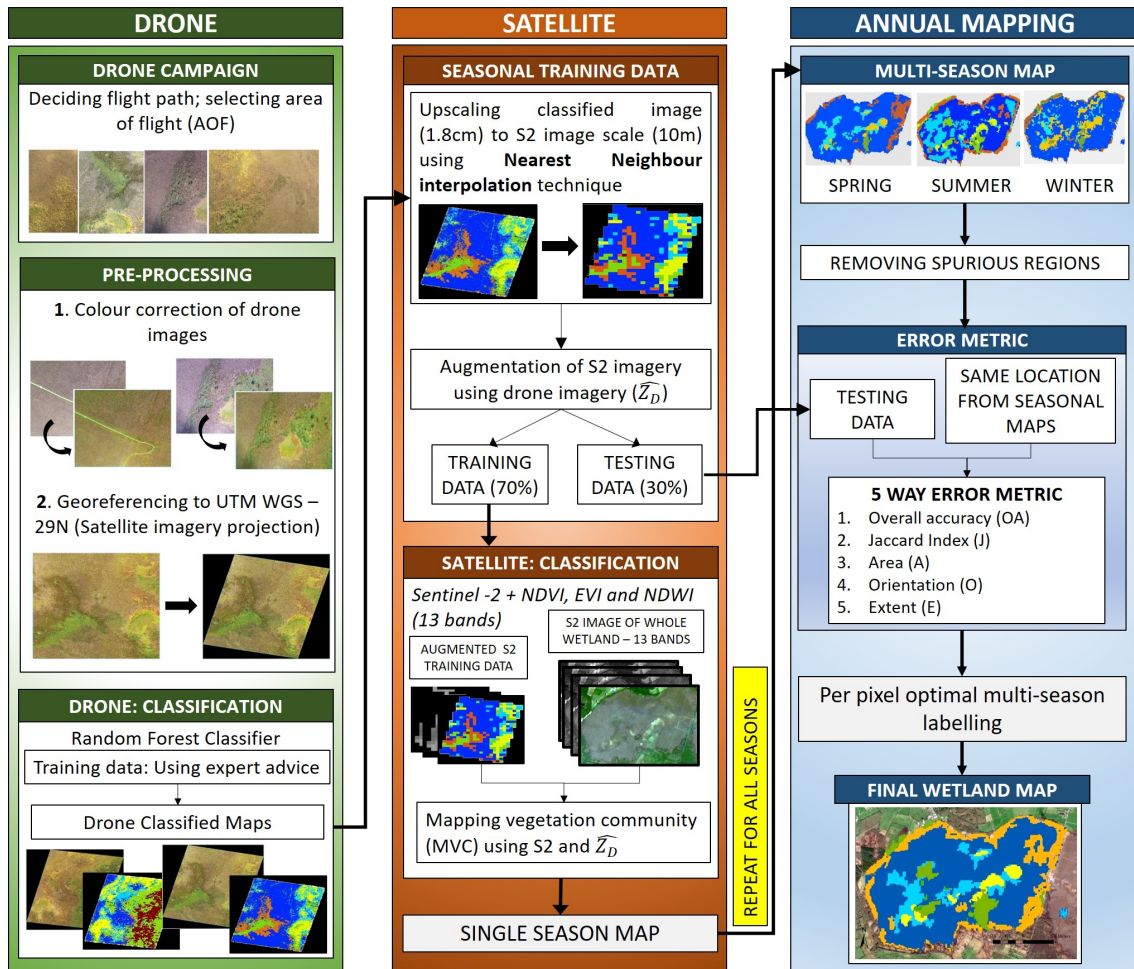
Figure 5.1: Methodology Flowchart

5.2.1 Drone Imagery

The extent of the individual wetlands, such as raised bogs, can vary from 10 to 1000ha or even more in some cases. To cover such large landmasses using UAV or drone, enough batteries, time, and favourable weather conditions are required. In the first step of the proposed framework, a plan for capturing the most informative image data, with space and time constraints, using drones is described considering technological, ecological, and geographical constraints.

5.2.2 Drone Technology

The two main types of drones are fixed-wing drones and multi-rotor drones, which are available in similar price ranges (Floreano & Wood 2015). Depending on the manufacturer, both types of drones are capable of carrying a range of different sized sensors. Due to a more straightforward design and less complicated mechanism, fixed-wing drones have longer flight times when compared to multi-rotor drones. However, fixed-wing rotor drones need to have a runway for launching, which is why multi-rotor drones are generally preferred when the application area is compact or unknown. For this study, a multi-rotor drone was used.

5.2.3 Drone Image Capture

The lifetime of the batteries that power the drones is the primary technological constraint. The choice of area of flight (AOF) needs to be defined such that all areas of interest (in our case representative areas of all vital vegetation communities present on a raised bog) are captured, optimising battery. The input of an expert in the field, such as an ecologist or consultation of field survey maps, will help in the choice of AOF.

Designing the Flight Path

The battery life of the drone primarily restricts areal coverage. The drone flight time can be calculated as,

$$Flighttime = (capacity \times discharge)/Averagerange \tag{5.2.1}$$

where the capacity of the battery used by the drone is in milliamp-hour (mAh), discharge is the rate at which battery is discharged, (generally 80% for most of the drones), and the average range is the drone amp draw calculated in amperes (dependent on drone weight and battery voltage). An average multi-rotor drones such as DJI Inspire series, Spark, etc. have an average flight time of 15 minutes per battery.
The coverage is also dependent on the FOV of the drone sensor. Specifications such as focal length (f), FOV (θ), aspect ratio (r) are detailed by the manufacturer for each model of drone. The area covered by an image can be calculated as,

$$A = (2 \times h \times tan(\theta^\circ/2))/sqrt(1 + r^2) \tag{5.2.2}$$

$$B = (2 \times r \times h \times tan(\theta^\circ/2))/sqrt(1 + r^2) \tag{5.2.3}$$

where h (m) is the height at which the drone is flown, and A and B are the lengths of the sides of the image. The flight time and the corresponding coverage area can thus be calculated from Equations 5.2.1-5.2.3. For mapping purposes, it was beneficial to cover the most significant area with appropriate overlap adequate for mosaicking of the images. For effective mosaicking, a minimum of 80% frontal and 60% side overlap in the image was required (Nex & Remondino 2014). However, the amount of overlap should be adjusted according to the application. For forest or dense vegetation tracking, an overlap of 85% frontal and 75% sideways was recommended, whereas in flat or low-lying lands such as agriculture farms, wetlands, an overlap of 85% frontal, and 70% sideways was sufficient. Low lying snowy or sandy areas also need similar overlaps but higher contrast settings to capture more features.

Drone Image Pre-processing

Once the drone images were captured, the images were checked thoroughly to get rid of the blurred and unclear images. A library with only workable images was created.

- *COLOUR CORRECTION*
  While capturing the images from an RGB camera without any additional infrared/multispectral sensor, colour is critical for the identification of vegetation. If the intensity of the light changes significantly between images, the same community may have different colours in the different images depending on the camera exposure and amount of daylight at the time of capture. To avoid this discrepancy, a colour correction technique was applied to the images prior to analysis. Therefore, by using this correction, the lighting and weather

conditions for images that were captured days apart can be compensated for, assuming that the ground condition does not change significantly in the intervening time.

In this study, the colour correction technique based on an iterative linear transformation of probability density function (pdf) of the image and colour re-graining (Pitie et al. 2005). The method converted the RGB image in 1-dimensional space and then found the optimum pdf by projecting it iteratively into N-dimensional space. The colour correction algorithm is summarised below:

Let the original image be X, and the target image be represented by Y. The pdf of X and Y is represented by f(x) and g(y), respectively. C_x and C_y be cumulative pdfs of X,Y, where x,y are the components of RGB bands of the camera.

For N =1, the mapping function is given by Equation 5.2.4,

$$t(x) = C - 1(y(C_x(x))) \tag{5.2.4}$$

where, t is the one to one mapping function.

For $N \geq 2$

For k iterations

- Rotate the samples using a rotation matrix (R); $x_r \longleftarrow R_{x^k}$.

- Project the samples over axis i, giving the marginal f_i, g_i.

- For each axis i, find t_i to match f_i into g_i.

- Remapping the samples for all N dimensions.

- Rotate the samples to their original form $x^{(k+1)} \longleftarrow R^{-1} x_r$.

- k = k + 1; Repeat for until convergence.

For N dimensions, the final one to one mapping is given as Equation 5.2.5,

$$t(x) = x_j^\infty \tag{5.2.5}$$

where j is the jth pixel. The algorithm converges f,g when the operation was repeated for k = 20 rotations.

- *DRONE IMAGE MOSAICKING AND GEO-REFERENCING*
  Multiple drone images can be mosaicked together to create a single colour correct large image containing multiple communities in a single scene. This can be done based on location tags (GPS) or feature matching. Commercial software like Pix4DMapper provides precise mosaics and geo-referenced images based on locations as well. Feature matching includes the use of overlap information, translation, and rotations for effective mosaicking (Mallick 2002). Various image processing platforms (like MATLAB (MathWorks & Inc 2019), Adobe Photoshop (*Adobe Photoshop Lightroom Beta* 2019)) allow image mosaicking based on feature matching. For geo-referencing of the images, ground control points (GCP) are to be identified. The geo-referencing can be done using open-source QGIS (QGIS 2020) or ArcGIS (*ArcMap Desktop (Version 10.6.1)* 2019) platforms. For this study, the pre-processing steps, along with their results, are described further in section 5.4.

Drone Imagery Classification

Once the drone imagery was pre-processed, a semi-supervised classification of drone images was carried out. For identification of the vegetation communities, ecological experts were asked to select locations that provided good representations of each key community. In any single image, 3-5 training pixels (towards a centre of the extent of a class) were chosen for every class visible in the image. A set 16-neighbours, encasing pixels corresponding to every chosen training pixel, was used to create the training dataset. For a pixel p centred at (p_i, p_j) the 16 neighbourhood pixels lie within $(p_i\text{-}2, p_j\text{+}2), (p_i, p_j\text{+}2), (p_i\text{+}2, p_j\text{+}2), (p_i\text{+}2, p_j), (p_i\text{+}2, p_j\text{-}2), (p_i, p_j\text{-}2), (p_i\text{-}2, p_j\text{-}2), (p_i\text{-}2, p_j)$.

A Random Forest classification algorithm with 100 trees and three prediction variables (RGB bands) was then used to classify all the drone images. Random forest classification has been widely used for wetland and vegetation classification (Amani et al. 2017, Mahdianpari et al. 2017, Whitcomb et al. 2009, Zhang et al. 2019). It is a classification, and regression tree (CART) model, where the trees are split with respect to features (bands), and the final decision is made based on majority voting (Liaw & Wiener 2002). The 100 trees RF was performed on the samples formed using bootstrapping (with repetition). A random subset of one of the three predictor variables (out of R, G, B) was used for data partitioning at every node in the tree based on the predictor's ability to split the samples with the highest increase in uniformity. The classified drone images were denoted as Z_D, where D = number of drone images used. The algorithm was trained on 70% of the drone images, and the remaining 30% was used for testing the algorithm. The testing accuracy of the drone images defines the ability to use a drone to correctly identify the vegetation community at a given period /season.

Training Data for Satellite

To complement the satellite image, the Z_D was upsampled from the centimetre scale to 10 metres. The nearest neighbourhood interpolation technique was used for this purpose (Olivier & Hanqiang 2012). For a discrete image, the nearest neighbourhood interpolation works more effectively than bi-linear or cubic interpolation as the latter produces softer images. For every pixel p, an average of 5×5 neighbouring pixels was taken, and these 25 pixels were replaced by a mode-value of 25 pixels. There was no reduction in the area of the image; only the pixel-resolution of the image was changed. The upsampled images act as the new training data labels ($\hat{Z}_D$), which were used with the satellite imagery for overall classification. The effectiveness of the upsampling and corresponding loss of information was examined by calculating the ratios of the labels in the original and upsampled image (discussed in section 5.4.2). It has to be noted that in remote sensing terminology, going from high resolution to low resolution i.e., converting centimeters to meters, is referred to as upsampling. Downsampling is conversion of low resolution (meters) image to high resolution (centimeters).

5.2.4 Satellite Imagery

Using the labelled training data created by the drone imagery (as detailed in the previous sections), the satellite image was segmented. A cloud-free satellite image was selected which was captured at a time close to the date of drone survey. For a maritime temperate climate such as Ireland's, getting a cloud-free satellite image was a big challenge. Sentinel-1 (SAR) data can be used as an alternative, but it further reduces the spatial resolution to 20 m. This reduction in spatial resolution may lead to pixel mixing for small vegetation

communities leading to inaccurate mapping. Hence, Sentinel-2 was chosen to be the optimal satellite for further analysis. Details about S2 and its bands are given in Section 5.3.

Augmenting Satellite and Drone Imagery

The S2 imagery was augmented using $\hat{Z}_D$. As the resolution of ($\hat{Z}_D$) is 10 m, the entire S2 imagery was resampled to 10 m spatial resolution. This was carried out using a spatial analyst tool in ArcMap (*ArcMap Desktop (Version 10.6.1)* 2019). ($\hat{Z}_D$) was placed as a new layer in the S2 imagery and the locations of ($\hat{Z}_D$) were accurately matched. The augmentation makes sure that if the drone imagery was resampled to its original size, the corresponding S2 data could also be resampled to view detailed information.

Satellite Image Classification

For the classification of satellite imagery, a study done on Irish wetlands by Bhatnagar et al. 2020a has used the combination of ensemble Bagged Tree (BT) classifier, along with Graphcut segmentation. It was named mapping vegetation community (MVC) algorithm, and it was shown, for classification of wetlands, an accuracy of more than 80% was achieved. Therefore, this study has supported and used the MVC algorithm for the satellite image classification of the entire wetland. The augmented S2 imagery with ($\hat{Z}_D$) was used as training data with each 10m pixel labelled as described in the previous sections. Graphcut segmentation is a post-classification smoothing technique that used contextual information for the formation of clear segments. For more information on the method and its equation, see Bhatnagar et al. 2020a. This was important for a clear delineation of boundaries between the communities. The classified image of the wetland was denoted by Z_S, where S = season when the mapping was done.

5.2.5 Seasonal Mapping

To correctly monitor any natural changes in the wetland over time, such as vegetation growth over the seasons, it was essential to repeat the steps multiple times in a year. The steps mentioned in the section 5.2.1-5.2.4 were repeated for three seasons to obtain seasonal maps of the wetland (Z_S, s=1 (spring); 2 (summer) ; 3(winter)). The study like Van Deventer et al. 2019 has depicted the advantage of forming the final map using the combination of single seasonal maps. Therefore, in this study, the final map was created by performing a multi-seasonal majority voting for every pixel. To optimise the multi-seasonal labelling, it was necessary to consider seasonal change-detection. In retrospect, over an annual time frame, the changes in vegetation communities may not be significant, but at a seasonal scale, the changes are apparent. To detect the changes effectively, this study has introduced an error metric, as discussed below.

Removing Spurious Regions

Due to an increase in the detail of the training points ($\hat{Z}_D$), the classified map (Z_s) contained small regions, which made it difficult to track the exact boundary of the vegetation community. Hence, the sparse regions below 10 m^2 area were removed using an area-opening morphological operation (Vincent 1994), seasonal maps after this step were denoted as($\hat{Z}_S$).

Error Metric

The spatial changes in the vegetation communities across seasons include changes due to differential growth related to phenology (seasonal) and due to interference among species (facilitation, competition – operating over all time frames). The error metric, as detailed below, was calculated based on the segmented images ($\hat{Z}_s$) of the wetland for its minute analysis.

1. Overall Accuracy (OA)
   To validate test data against $\hat{Z}_s$, OA was calculated for every season, as shown in Equation 5.2.6.

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.2.6}$$

   where TP = True positive, FP = False positive, TN = True negative, FN =False negative.

2. Jaccard Similarity Index (J)
   It measures the similarity between the members of the two sets and reports the amount of similarity and distinction (Real & Vargas 1996), shown in Equation 5.2.7. Let the reference image = $\hat{Z}_D$; Classified image (for the similar region as reference image) $=\hat{Z}_s$, for s = 1,2,3 (spring, summer, winter)

$$J(\hat{Z}_D, \hat{Z}_s) = |\hat{Z}_D \bigcap \hat{Z}_s| / |\hat{Z}_D \bigcup \hat{Z}_s| \tag{5.2.7}$$

   where $\hat{Z}_D \bigcap \hat{Z}_s$ is the intersection of these two sets $\hat{Z}_D$ and $\hat{Z}_s$; $\hat{Z}_D \bigcup \hat{Z}_s$ is the union of these two sets $\hat{Z}_D$ and $\hat{Z}_s$, i.e., elements present in both ($\hat{Z}_D$ and $\hat{Z}_s$).

3. Area (A)
   Estimates the area of the vegetation community are selected. The area of every individual pixel is determined by looking at its 2×2 neighbourhood. Each pixel is part of four different 2×2 neighbourhoods, which indicates the change in overall growth/shrinkage of the community.

4. Orientation (O)
   The orientation gives the angle between the x-axis and the major axis of the ellipse. It can range from -90 to + 90 degrees, indicating the direction of the change of the community.

5. Extent (E)
   The extent indicates the total pixels present in the bounding box with respect to the total pixels present in the image.

The error metric allows seasonal discrepancies in the vegetation communities to be identified. Notably, along the boundary due to the spatial resolution of S2 imagery, a clear identification of the community can be tricky. During majority voting, some pixels do not have a definite label. For these pixels, the error-metric parameters gave an idea about the mutual change in the communities. The community with the least change in error-metric parameters (along the seasons) was selected as the dominant community. While majority voting, in case of an uncertain label, the pixel was labelled with the corresponding dominant community. This method allowed the proper selection of labels for optimal multi-season labelling and reduced the error due to pixel mixing between nearby communities, leading to a high level of accuracy.

5.3 Case Study

5.3.1 Study Area

Clara Bog, situated in County Offaly in the midlands of Ireland (see Figure 5.2), is one of the largest of the remaining active raised bogs in Ireland, for more details, see Section 1.7.
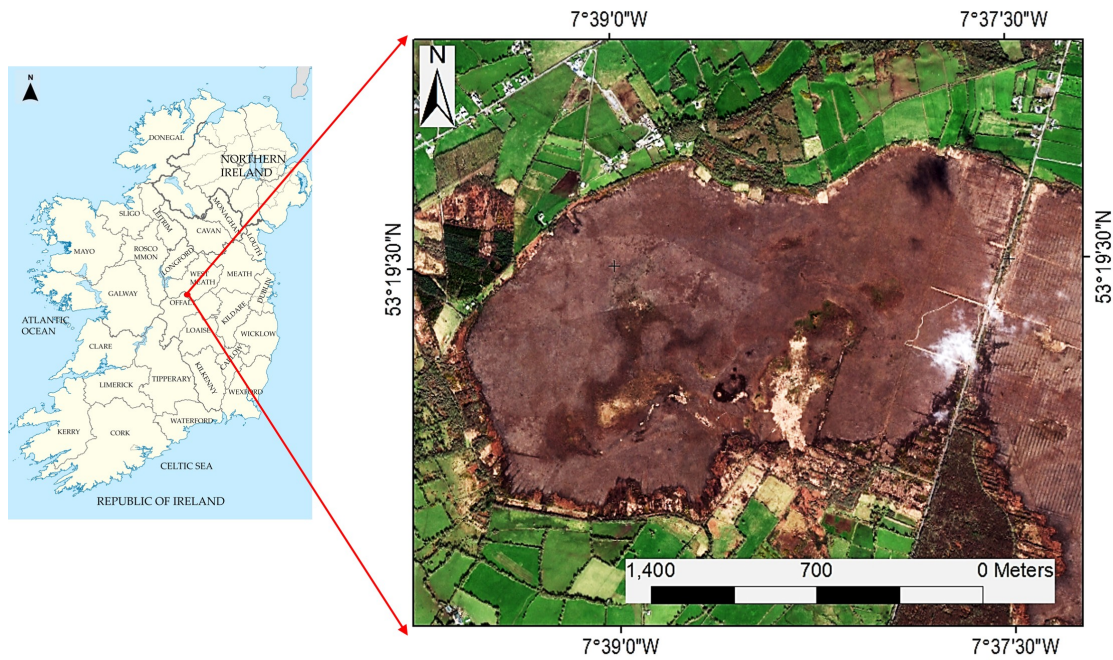
Figure 5.2: Location of Clara Bog, County Offaly

The ecotopes can change appearances by the season. Appearance is also dependant on the water content and capacity of the ecotope. The C and AF ecotopes are mostly wet areas characterised by "quaking" conditions underfoot, whereas SM and M ecotopes would be more firm. The difference in ecotope appearance with respect to seasons and their main properties and dominant species are shown in Figure 5.3 based on Fernandez et al. 2014. The drone images of the ecotopes depict a clear difference from spring to winter, thereby confirming the need to incorporate a temporal monitoring scheme for such regions.
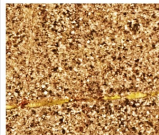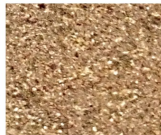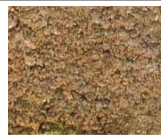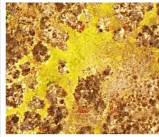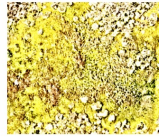
Ecotope | Description	Seasonal images		
	spring	summer	winter
SUBMARGINAL : Hard or soft surface, not quaking. Presence of algal mats, Narthecium hollows. Characteristics species: *Sphagnum papillosum, Sphagnum magellanicum, S. capillifolium, Trichophorum, Calluna vulgaris.*			
SUBCENTRAL : Soft surface, sometimes quaking. Only few hummocks, *Narthecium* hollows. Characteristics species: *Sphagnum magellanicum, Sphagnum papillosum,* occasional *Sphagnum cuspidatum.*			
MARGINAL : Consists both marginal and facebank (Kelly et al., 1995). Low water level, degraded micro-topography. Characteristics species: *Narthecium ossifragum, Sphagnum capillifolium, Trichophorum, Carex panacea*			
Central : Indication of high quality raised bog. Consists of pools, and high moss Characteristic species: *Sphagnum cuspidatum and S. magellanicum, Rhynchospora alba and Eriophorum angustifolium*).			
Active Flush : Wetter vegetation. Consists of pools, low hummocks and hollows. Characteristic species: *Aulacomnium palustre, Vaccinium oxycoccus, Empetrum nigrum and Andromeda polifolia*			

Figure 5.3: Description of species present in Ecotopes at Clara bog along with the seasonal drone images captured using DJI Inspire 1 with RGB Sensor

5.3.2   Material Used

Drone Data: DJI Inpsire 1^{TM}

The drone used in this study was a DJI Inspire 1^{TM} (DJI, Shenzhen, China) as detailed in Section 4.2. The Sentinel-2 Multispectral Instrument Level 2A (S2-MSIL2A) images used are bottom-of- atmosphere (BOA) reflectance in cartographic geometry.More details on S2 data are stated in Section 2.2.1. The wavelengths used in this study, along with its spatial resolution, are mentioned in Table 5.1.

BANDS	Blue	Green	Red	VRE1	VRE2	VRE3	NIR	NarrowNIR	SWIR 1	SWIR 2
Wavelength μm	0.49	0.56	0.665	0.705	0.740	0.783	0.842	0.865	1.610	2.20
Spatial Resolution (m)	10	10	10	20	20	20	10	20	20	20

Table 5.1: Sentinel 2 bands description

Along with the S2 bands, indices like normalised difference vegetation index (NDVI), enhanced vegetation index (EVI) have proven to be extremely useful for the detection of vegetation communities in various seasons (Bhatnagar et al. 2020a). For understanding the wetness of the ground, normalised difference water index (NDWI) was used.
For a clear understanding of the behaviour of the ecotopes across seasons, their spectral profiles were plotted (Figure 5.4). The median reflectance value for every community was plotted against the wavelengths under consideration mentioned in Table 5.1. The plots illustrate the similarity in the spectra between the ecotopes and explain the misclassifi-

cations between similar classes. Therefore, along with monitoring, there was a need to consider the spectral similarity of the ecotopes along the seasons. The wavelength 480 nm depicts the absorption of chlorophyll b, and 640nm depicts the absorption of chlorophyll a. For S2, this is depicted by blue and red bands. The water absorption happens in SWIR wavelength range at around 1400nm, 1900nm, and 2600nm. In the case of S2, the water absorption dip can be witnessed by using SWIR2 (2200 nm) wavelength.

The spectral signature of all the communities is highly overlapping, depicting an inbuilt similarity in the vegetation present inside the bog. The spectra of SM, SC ecotopes are overlapping for lower wavelengths (visible range), with slight differences along the seasons. The M ecotope is similar to SC, SM, for the spring season but is distinctive for summer and winter. It can be seen (through SWIR) that M is a dry ecotope for all the seasons, which is in agreement with the ground conditions. C, AF ecotopes are primarily wet throughout the year, as shown by their low value at SWIR2 in Figure 5.4(a, b, c).
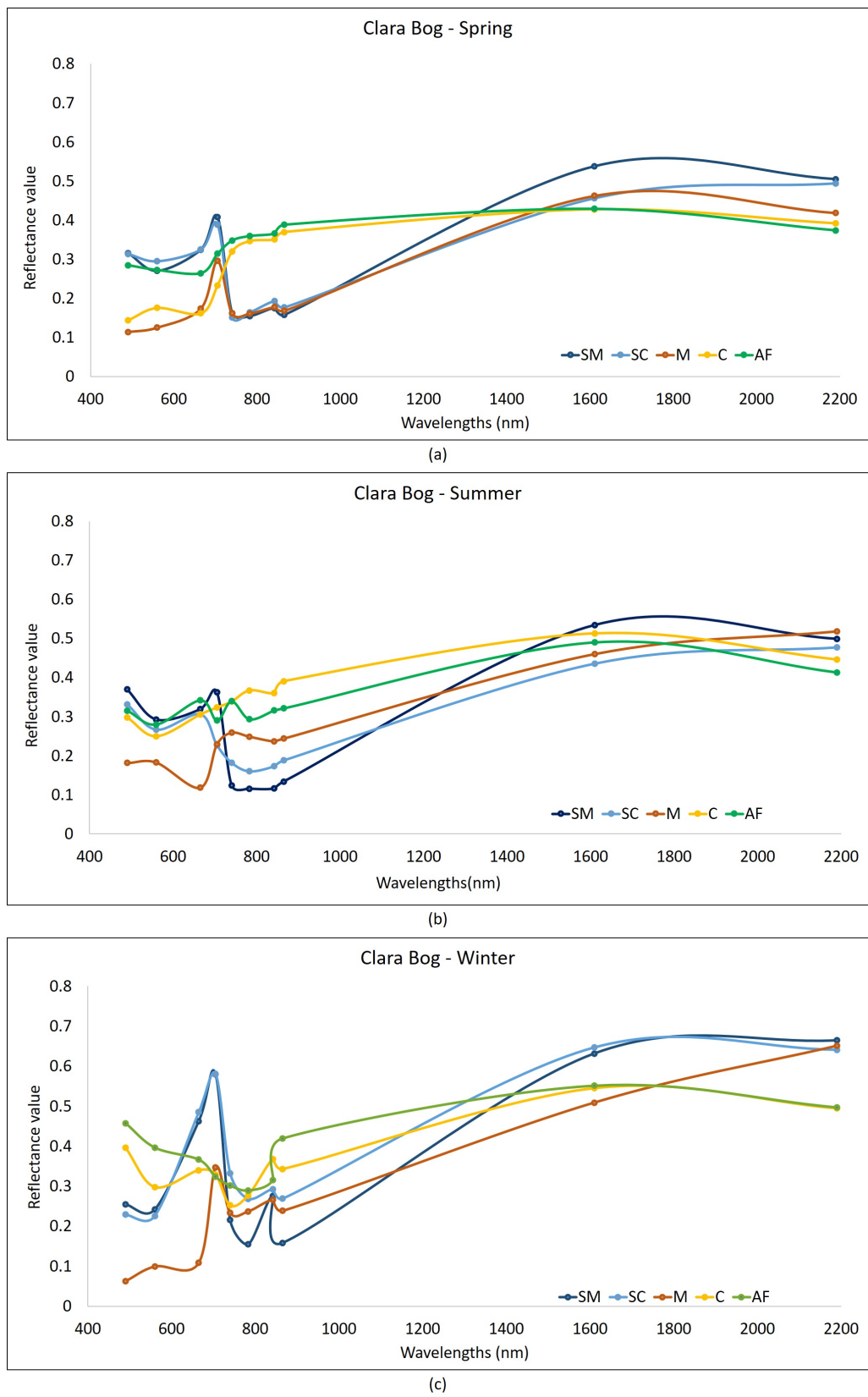
Figure 5.4: Temporal spectral signatures of ecotopes SM, SC, M, C, AF - (a) Spring (b) Summer (c) Winter

5.4 Results

In this section, the results of applying the proposed framework (Section 5.2) to the Clara Bog using the materials enlisted in Section 5.3.2 are described in detail.

5.4.1 Drone Image Capture

For this study, the total area captured was reduced to 8 ha due to high wind speed (which consumes more battery) at the location.
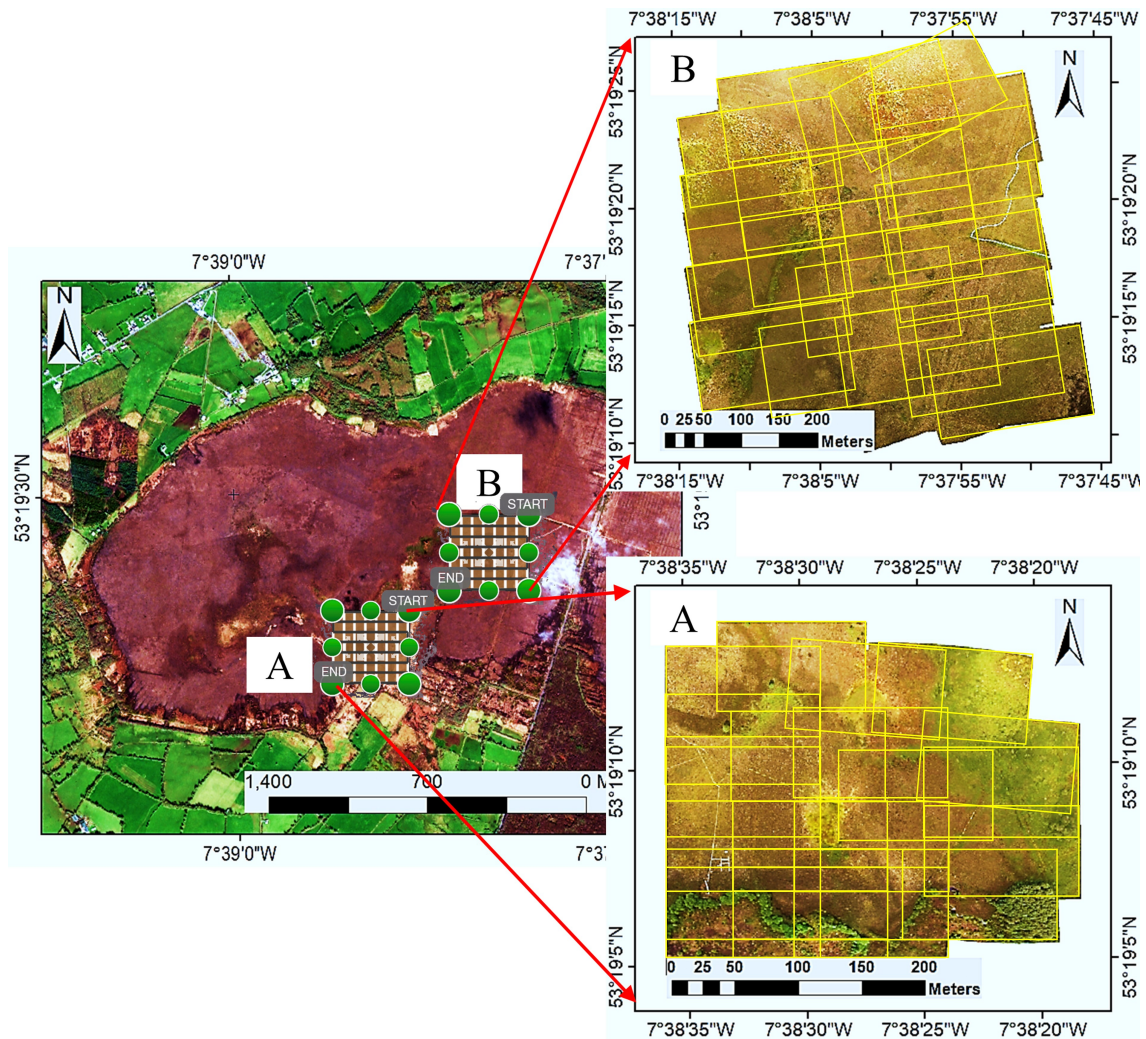
Figure 5.5: Area covered in the Pix4D Capture mission at Clara bog: A, B (mosaic of 20 drone images)

For acquiring the images, the open-source software application, "Pix4D Capture," was used. This application allows the AOF to be pre-planned, thereby providing an estimate of the flight time required. The experience on this raised bog suggested that each flight mission should be planned for a maximum of 8-12 minutes, as the wind drag, time taken for the drone for reaching to the mission's start point and landing require time at least 35% battery charge.

In Clara Bog, the AOF was decided with the aim of capturing the five main ecotopes present (Bhatnagar et al. 2018). With a frontal overlap of 70% and a sideways overlap of 80%, two different areas were covered using two batteries (Figure 5.5). The images were captured in overcast weather, or when the sun is at its maximum elevation, to minimise the effect of shadow. Out of the two regions captured, only area 'A' (Figure 5.5) was used for developing the training dataset. The images from area 'B' were used for testing purposes. In total, the mosaic is created using $\approx$ 20 images, but for further classification related analysis, only $\approx$ 5 images, containing all the ecotopes, per season were considered (using the pre-processing steps, as described below). The same areas were captured for three different dates, 18 May 2019, 4 August 2019, and 14 November 2019. The corresponding date for satellite imagery was 21 April 2019, 18 September 2019, and 12 November 2019. These dates depict the three seasons- spring, summer, and winter.

5.4.2 Drone Image Pre-processing

As described in section 5.2, the pre-processing consists of colour correction, drone image mosaicking, and geo-referencing. For colour correction, the target image was chosen to depict the best colour state, by manual inspection. This image consisted of all key ecotopes, such that a correct colour reference was created for all communities. All the other images were transformed into the colour template of the target image. In this study, N was taken as 20 for getting the best colour corrections. This method was very effective and yielded an appropriate colour correction. The effect of colour correction and subsequent mosaicking is evident in the study, and Figure 5.6 shows that after the application of colour correction, the images are reformed (to the target image colour palette). All the images had a similar colour range, and therefore, a single model was used for analysing them. The colour correction also helps in subsequent mosaicking by depicting clear similarity both in texture and intensity. Secondly, the panoramic mosaics of the images used in this study were created based on the feature-similarity and overlap in the images using Adobe Photoshop Lightroom version 7.3 software (*Adobe Photoshop Lightroom Beta* 2019). Lastly, the orientation of the mosaicked image was corrected and geo-referenced using the tools available in ArcMap v10.6.1 (*ESRI, World 30cm Imagery* 2020). For geo-referencing, 3-5 GCPs were identified. These points were matched with high-resolution 30 cm world-imagery as the base, and the projection of the images was rectified to UTM/WGS84.

5.4.3 Classification

Drone Imagery Classification

A total of 14 drone images ($\approx$ 5 images per season) were used to create maps of Clara bog. The RF classifier was used for classification with area 'A' as training and images from area 'B' as testing (Figure 5.5). The average model accuracy of mapping the drone images calculated using a 5-fold cross-validation method was 93.4%. The 5-fold validation technique involved dividing the entire training dataset into five subsets (with repetition) and checking accuracy for each subset (Fushiki 2011). The final accuracy is the average of all five. All the training images combined constitute around 8 ha in the area. Although the west of Clara bog is 250 ha, a previous study (Bhatnagar et al. 2020a) quantified the usage of a minimum amount of training data required to obtain a functional classification.
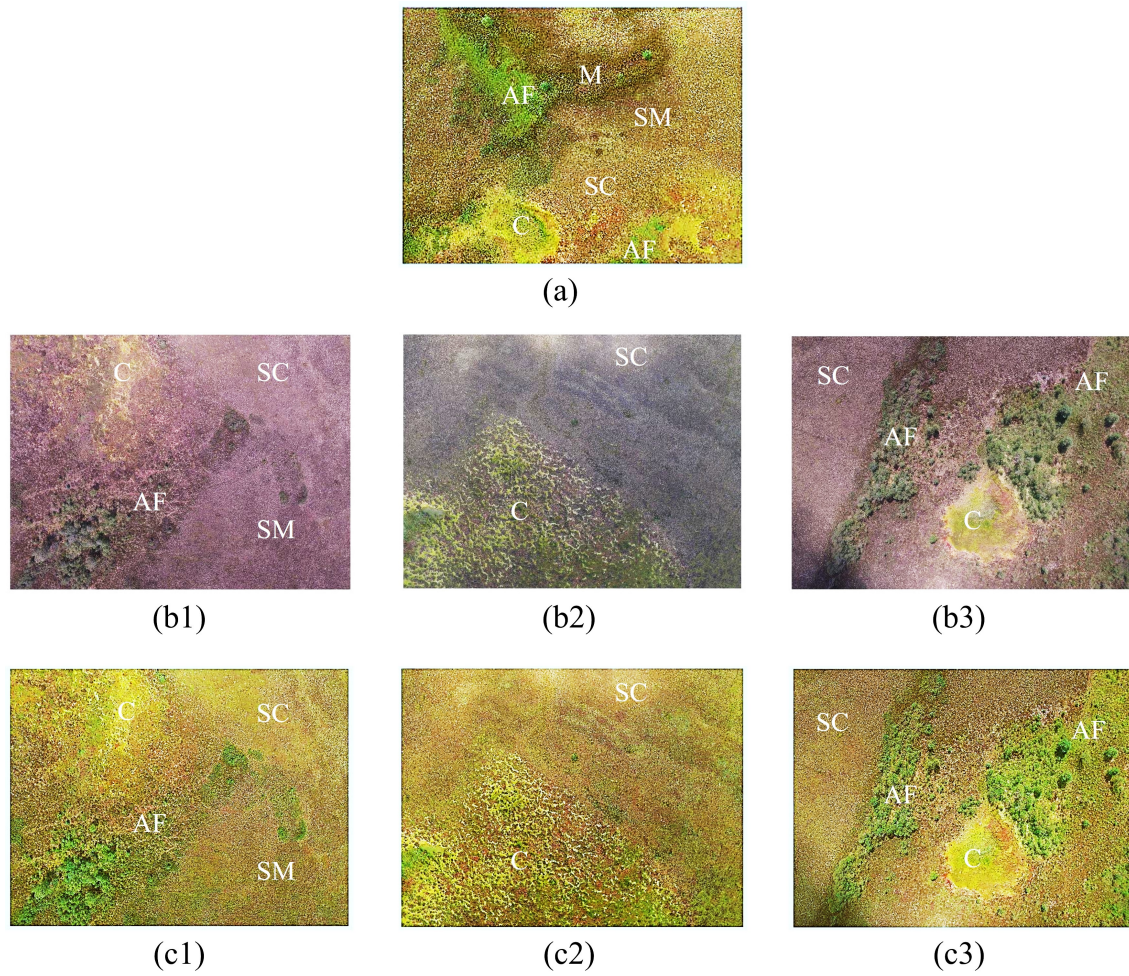
Figure 5.6: Colour correction using N dimensional pdf algorithm. (a) Target image (b) original drone images (c) colour corrected drone images. Ecotopes: AF = active flush; C = Central; M = Marginal; SC = Subcentral; SM = Submarginal

Training Data for Satellite Image Classification

The colour corrected and geo-referenced images were classified into five ecotope categories and upsampled to the S2 (10 m) spatial resolution, as shown in Figure 5.7. The "nearest neighbour interpolation" upsampling technique ensured that the areal composition of the vegetation communities stayed unaltered. The average areal composition of all the vegetation communities in the drone images and corresponding upsampled images is shown in Table 5.2, which shows that the upsampling process did not lead to any significant loss of information spatially. For better understanding the amount of error, the upsampled images were downsampled to their original scale (using the same – nearest neighbour interpolation technique) and were compared against the original drone images. There is an apparent error that was introduced when the original data was resampled. This error was controlled and reduced by choice of upsampling technique. Table 5.3 depicts the averaged confusion matrix and precision calculated for every community for the training images ('A').

(a)   (b)   (c)
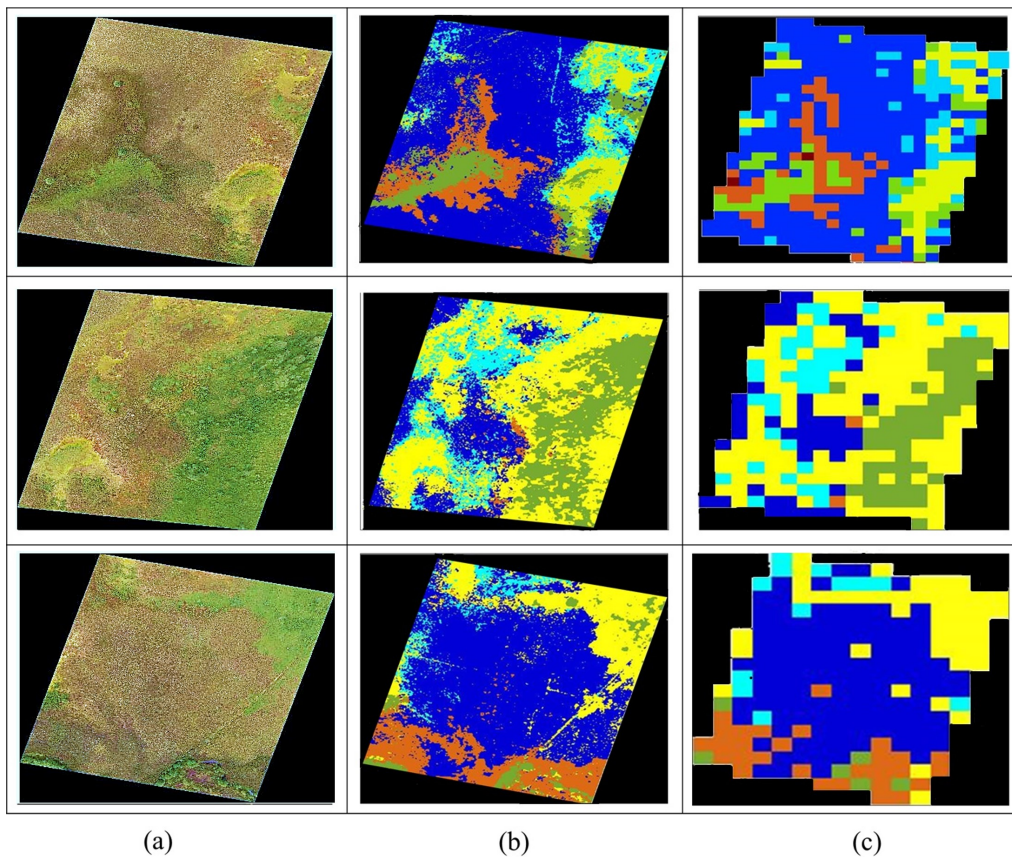
Figure 5.7: Upsampled images using the nearest neighbour upsampling technique (a) The original drone image (b) classified drone image (c) upsampled drone image

Areal(%)	SM	SC	M	C	AF
Original	54.96	25.67	6.68	7.54	5.15
Upsampled	58.45	25.78	5.56	7.88	2.33

Table 5.2: Areal coverage of ecotopes (%) in S2 and corresponding upsampled drone imagery

The segmented images were highly discrete, picking up the details at the cm scale. For this study, it can be seen that by using nearest neighbour interpolation, at an average (for every community), a precision of 89% was achieved (Table 5.3).

	SM	SC	M	C	AF
SM	1428852	16259	14412	491	6321
SC	15910	1023934	860	7320	136690
M	12664	632	95670	130	1068
C	488	7239	125	1274314	126678
AF	6407	147371	963 2	04941	2623402
Precision	0.97	0.87	0.86	0.90	0.87

Table 5.3: Confusion matrix of original drone classified and downsampled image (averaged for all images)

The precision was highest (97%) for SM which was the largest (spatially) found in Clara bog. Whereas M, with 86% precision, was spatially the smallest ecotope found in the area under consideration.

Satellite Image Classification: Seasonal Mapping

The overall accuracy (OA) and error metric was determined by comparing the upsampled classified drone imagery $\hat{Z}_D$ with classified S2 imagery for each season $\hat{\tilde{Z}}_s$. The OA was calculated using Equation 5.2.6.

Figure 5.8 depicts the average of the error metric parameters calculated for the test area. As an illustrative example, one test image (from area 'B') is also depicted in Figure 5.8. The test images were the RF-classified upsampled images, which were compared against the classified S2 imagery (of the same location) for every season.
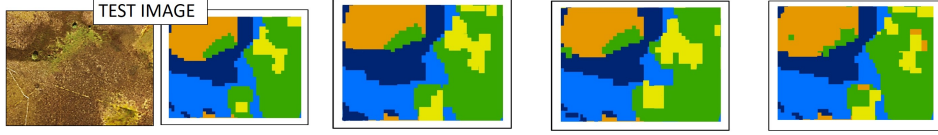
	SPRING	SUMMER	WINTER	
Date drone image acquired	18th May 2019	4th August 2019	14th November 2019	
Date satellite image	21st April 2019	18th Sept 2019	12th November 2019	
OA (%)	**86.6**	**87.2**	**81.2**	**Averaged Δ**
SM : Jaccard similarity	0.76	0.78	0.73	0.75
■ Δ (Area in m^2)	100	80	40	73.33
Δ (Orientation)	0.01	0.00	0.00	0.00
Δ (Extent)	0.04	0.01	0.00	0.01
SC : Jaccard similarity	0.81	0.78	0.85	0.81
■ Δ (Area in m^2)	50	40	20	36.66
Δ (Orientation)	0.02	0.01	0.02	0.16
Δ (Extent)	0.01	0.00	0.00	0.00
M : Jaccard similarity	0.71	0.88	0.66	0.75
■ Δ (Area in m^2)	70	90	110	80
Δ (Orientation)	0.02	0.01	0.33	0.12
Δ (Extent)	0.05	0.02	0.41	0.16
C : Jaccard similarity	0.82	0.78	0.71	0.77
□ Δ (Area in m^2)	90	100	120	103.33
Δ (Orientation)	0.00	0.20	0.05	0.08
Δ (Extent)	0.03	0.35	0.65	0.33
AF : Jaccard similarity	0.83	0.77	0.73	0.77
■ Δ (Area in m^2)	50	200	90	113.33
Δ (Orientation)	0.00	0.05	0.40	0.15
Δ (Extent)	0.20	0.57	0.71	0.49

Figure 5.8: Error Metric, OA of testing drone images against classified sentinel-2 imagery for five key ecotopes

The error metric presented in Figure 5.8 gives a definite idea about the seasonal mapping accuracy (OA) for mapping the ecotopes. A maximum of 87.2% OA was achieved for the summer imagery, which was quite close to the OA achieved in spring (86.6%). It also depicts the similarity index, change ($\triangle$) in the area, orientation, and extent for all seasons with respect to the $\hat{Z}_D$. All the four error parameters were given equal weightage, and an average of all was considered to make the final decision. A high value of the Jaccard similarity index was considered the best, and the $\triangle$ should be as low as possible. This was done for all the images, captured through all the seasons. As can be seen from Figure 5.8, and also based on ecological conditions on the bog, the ecotopes were given preference

as $SC > SM > C > AF > M$. This means, SC was given the highest weightage in majority voting, and M the least. The final result was the optimal-multi-season-labelled map, which consisted of the most certain label from all the seasons (Figure 5.9).
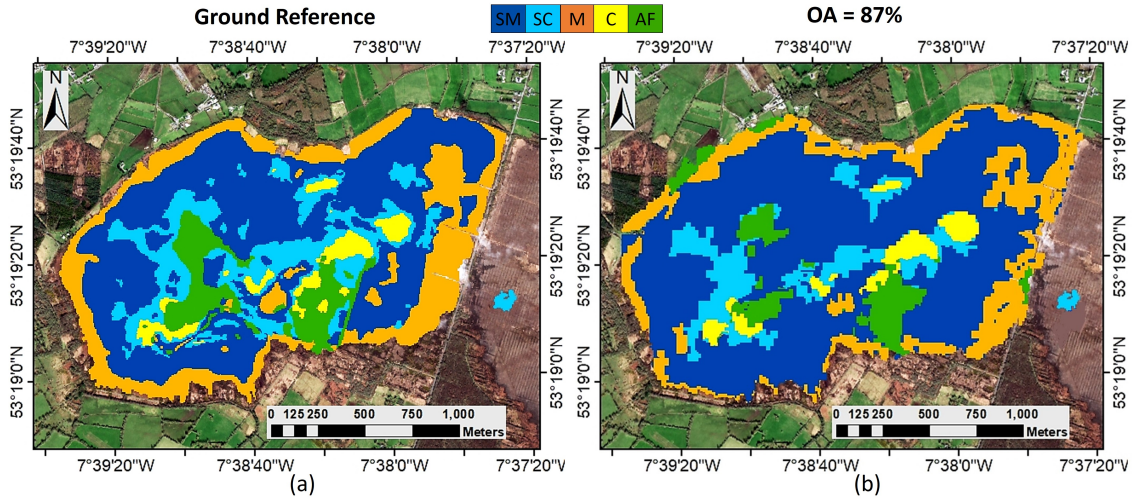
Figure 5.9: (a) Reference map (2018) Bhatnagar et al. 2020a) (b) Drone augmented S2 classified map (2019)

The entire wetland (Clara bog) was mapped using the augmented imagery via the MVC algorithm described in section 5.2. Figure 5.9(b) represents the final map created using the proposed method. When compared against Figure 5.9(a), which was a reference image taken from the study done by authors (Bhatnagar et al. 2020a), a mapping accuracy of 87% was achieved. This is a higher accuracy than using just Sentinel-2 imagery, as done by Bhatnagar et al. 2020a and without any field-survey information. This strengthens the case of the applicability and robustness of the methodology.

Apart from supervised, a firm argument can be the usage of unsupervised learning techniques altogether in place of the proposed technique. Therefore, an unsupervised K-means clustering analysis with K = 5 was performed on the S2 imagery of Clara bog across all the seasons. Based on the reference image (Figure 5.9(a)), the colour scheme and the clusters were assigned a label. For further confirmation, expert advice was also taken. Figure 5.10 depicts the K-means clustering with approximated labels.

The maps in Figure 5.10, when compared for accuracy against Figure 5.9(a), give an average accuracy of 62%. It can be clearly seen that ecotope C is identified very well; however, SM, SC ecotopes are highly misclassified. Therefore, it is clear that the use of drone imagery was essential for creating training data such that all the key ecotopes were identified and mapped accurately.
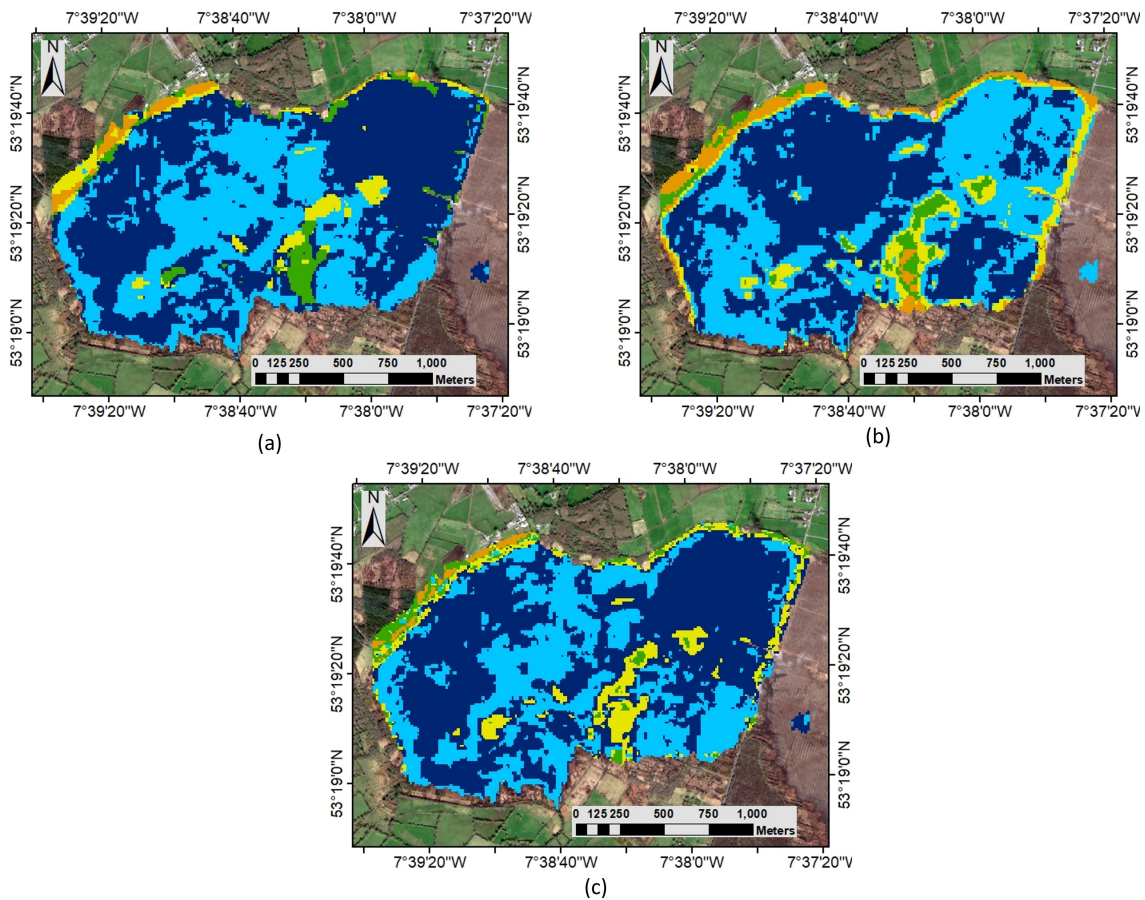
Figure 5.10: Clara bog unsupervised clustering (K = 5) of S2 imagery. (a) Spring (b) Summer (c) Winter

5.5   Discussion

The proposed methodology describes a robust and effective technique of using drone imagery as training data to augment satellite imagery for wetland classification. The main aim of the study was to reduce the dependency of wetland mapping on field surveys. The field surveys can be either done manually, or drones can be deployed for covering the entire area. For huge areas, both these methods can be both time and resource-intensive. Also, an inherent problem of the limited battery life of the drone still persists. Therefore, in this study, an alternative method where limited drone images were used with open-source satellite imagery for detailed mapping was presented. A minimum amount of drone images were captured ($\approx$ 5), with 70% frontal and 80% sideways overlap. In a case where the communities are scarce and located far apart, mosaicking will not be necessary and AOF, overlap should be decided such that they are correctly captured.

For a detailed analysis, a hyperspectral sensor can be used in place of the RGB sensor. These sensors capture images at multiple wavelengths and are specifically useful for mapping species. Although useful, these sensors are costly (pricing up to 50,000 Euros). Such an expensive sensor may not be affordable or available for many studies. Therefore, this study has provided steps to use a basic RGB camera for identifying intricate communities in a wetland. For compensating the presence of infrared bands (which are unaffected by

132

the lighting conditions), the study introduced the usage of colour correction techniques. This step improved the applicability of RGB images from different times and light conditions for consistent mapping. It ensured that all the images were used and tested against the same intensity palette even though the lighting conditions may vary amongst images. The images were then mosaicked and geo-referenced for further classification.

The initial drone image classification was done using RF (model accuracy 93.4%), which was upsampled using the nearest neighbour interpolation technique. The study shows that there was a limited loss of information during the upsampling (Table 5.2, 5.3). There are two unavoidable sources of error – model accuracy of drone classifier, and loss via upsampling. These errors were minimised by optimising the RF parameters while classification and optimising the size of the kernel for performing nearest neighbour interpolation. The upsampled imagery acted as new ground truth labels (training points) for the S2 data. The corresponding Sentinel-2 imagery was subset and augmented using the upsampled drone imagery. This step allows using open-source imagery (scaled 10m) with detailed labels. The classification of S2 was done using previously applied MVC algorithm (Bhatnagar et al. 2020a). The accuracy was checked for test images (area 'B'), and a maximum of 87.2% accuracy was achieved. This was repeated for three seasons for the year 2019; the OA of the final derived vegetation map compared to a field surveyed map was 87%.

The error metric was calculated by comparing previously unseen, upsampled drone imagery with MVC-classified S2 imagery (Figure 5.8). Such an involved metric gave a clear idea about minute changes that may occur seasonally but go unnoticed in a manual survey. This was because the characteristics of the ecotopes change depending on the season. Although the location of ecotopes remains the same, their appearance can be misleading and lead to potential misclassification. For example, the C ecotope was yellow during summer but was rather wet and brown in winter (Figure 5.3). The SM and SC ecotopes proved to be the most challenging to distinguish from each other as these communities are highly homogenous and often integrate. Therefore, a robust methodology using the differences between images taken over several different periods was necessary to characterise differences in the phenology of these ecotopes. This study successfully picks up the difference between the two ecotopes over three seasons, marking a pattern of seasonal vegetation growth and areal spread. For Clara bog, five key ecotopes were mapped. These ecotopes, in particular C, SC, and AF are usually wet and responsible for the formation and accumulation of peat. Due to a history of turf cutting, and drainage activities, these ecotopes are often damaged. It is of utmost importance to have precise monitoring of the ecotopes to understand and track any such activity.

A compounding problem with the field-surveyed vegetation maps was the subjective assignment of categories; the definition of Central, for example, in a raised bog, has always been subjective. There was no sharp boundary, and transitions to other zones were gradual. Hence, the precise mapping of vegetation manually was always going to remain somewhat instinctive. The vegetation changes temporally at different scales, annual directional shifts as communities expand or change or shift in area-occupied often due to interference (+ve and -ve) with other communities. Seasonal changes due to growth development and dieback, and even daily changes for some species where leaf angle varies both in relation to incident radiation and also by endogenous rhythms. The error metric, therefore, was a systematic change detection metric, providing a comparison of the minute spatial change between the images.

The error metric of the ecotopes depicts that across all seasons, the extent and orientation of the community stay intact. The SC ecotope was dominated by Sphagnum, which expands and contracts based on moisture conditions. This vegetation does not have a

hard boundary, and its interpretation can vary between maps generated manually. From Figure 5.9(b), it can be seen that SM, SC ecotopes appears to be blending. It was because the SM and SC ecotopes were also highly spatially integrated, leading to misclassification from the neighbouring wet communities because of the growth and death of characteristic species mentioned in Figure 5.3.

The ecotope M was mainly dry, and depending on the season, visibly could be green or brown. Sentinel-2's vegetation indices such as NDVI can give a clear delineation of this type of community. From Figure 5.4, it can be seen that the spectral signature was not consistent for smaller wavelengths in summer for the M ecotope. This justifies the change in the level of identification of this ecotope between the various seasons. The extent of C ecotope was the most interesting when evaluating the ecological health of such a wetland. It consists of moss species that hold water, though SC and AF also consist of water holding species (Figure 5.3). A water index, like NDWI, helps to identify and distinguish such wet communities. From Figure 5.4(c), it can be seen that the spectral signature of C has the highest dip in SWIR2 wavelength in winter depicting the water absorption of the community. The ecotope AF was identified correctly across all the seasons. This was due to a clear distinction using NDWI between the C and AF communities. A similarity of the spectral signature for SWIR wavelengths of AF with M in winter can also be seen (Figure 5.4(c)), this justifies the identification of the AF ecotope on the west of the bog (Figure 5.8).

The results presented so far clearly describe the necessity for drone and satellite image augmentation for a detailed temporal analysis. Using the proposed method, new patterns, and shifts in the community were seen, which were overlooked when using satellite imagery alone (Bhatnagar et al. 2020a). The study also compares the proposed methodology with unsupervised K-means clustering. Figure 5.10 shows that the C ecotope was very well-identified using unsupervised learning as well, but that the K means clustering has misclassified other ecotopes. Depending on the season, the wet communities such as C, and AF are also mixed and misclassified. The ecotopes SM, SC, are highly homogenous and spatially integrated vegetation types leading to high misclassification. This misclassification was solved using the error metric in the proposed methodology. The marginal ecotope was not identified using the unsupervised technique but was identified well using the proposed technique. Hence, this suggests, usage of the proposed methodology is beneficial in case of mapping an un-surveyed wetland and works better than mapping using just satellite data, therefore, improving the applicability of satellite imagery for wetland monitoring.

5.6   Conclusion

This study successfully demonstrates the suitability of using limited drone images for augmenting satellite imagery for improved overall classification of vegetation communities or ecotopes within the wetlands. The proposed methodology reduced the need to carry out a week-long field survey or complete drone survey lasting over multiple days to an hour-long drone survey of the chosen wetland.

A colour correction technique was introduced to create consistency between drone images captured under different light conditions and to maximise the use of captured drone images. The colour correction technique was crucial in utilising RGB images for creating ground truth or training data for vegetation communities which change with time-of-day, cloud cover, solar irradiance, and seasonal cycles. Also, this reduces the dependency on

battery life/flight time required for capturing images. This step was essential for a maritime temperate climate like Ireland.

The high-resolution drone images from the partial survey were classified using supervised RF classification with limited expert intervention and then upsampled to create labelled training dataset for satellite image (S2 imagery) classification, which were further classified using advanced semi-supervised MVC algorithm. This nested methodology improved increased classification accuracy (87%) of vegetation communities compared to the unsupervised method (62%), which is the only option in the absence of field data/ground truth. The proposed methodology proved to be a vital, resource-efficient tool to monitor the ecological health of wetlands with limited intervention, access, and prior knowledge.

Vegetation communities captured using RGB optical sensors using drone surveys provide season-specific information, and these colour and texture characteristics change every season. A five-factor error metric was identified to estimate the areal changes that occur in the communities within the year.

The proposed methodology is semi-automatic, with limited expert intervention. It is robust, and the general methodology can be applied to other types of wetlands, as has been seen in the previous paper by the authors. Overall, the study demonstrates the advantages of using inexpensive drone imagery combined with open-access satellite data to effectively classify wetland areas with high accuracy without the need of carrying out expensive field surveys or extended complete drone surveys.

Chapter 6

Methane Analysis of Canada using S-5P

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

— Marie Curie

Contents

This chapter corresponds to the paper that has been submitted on 15 October 2020 in "Nature Geoscience" and currently is under review [5].

6.1 Introduction

Methane (CH_4), after carbon dioxide (CO_2), is the second most important anthropogenic greenhouse gas contributing to climate change. Compared to CO2 it has a shorter atmospheric lifetime, of about 9 years (Prather et al. 2012), making it a favourable target for climate change mitigation. Atmospheric emissions and concentrations of CH_4 have increased continuously over the last decade (Saunois et al. 2020). Wetlands are known to be the largest natural source of CH_4 with an estimated average global emission, from "bottom-up" inventories/modelings approaches, of 149 Tg CH_4 yr^{-1} (range 102-182) during the past decade (2008-2017) (Saunois et al. 2020). This represents about 20%

of the total CH_4 emission sources estimated by such approaches. The wide variability in estimates is a result of the difficulty in defining CH_4 producing wetland areas and of parameterising terrestrial anaerobic conditions that drive CH_4 sources and conversely oxidative conditions leading to CH_4 sinks(Melton et al. 2013, Poulter et al. 2017, Wania et al. 2013). However, average CH_4 bottom-up wetlands emission estimates are lower than top-down emission estimates of 181 Tg CH_4 yr^{-1} (range 159-200) inferred by an ensemble of atmospheric inversions using an atmospheric constraint (Saunois et al. 2020). The difference between average wetland emissions from bottom-up and top-down estimates has increased from about 17 to 30 Tg CH_4 yr^{-1} in the recent global methane budget study (Saunois et al. 2020, Saunois et al. 2016), this difference for other natural emission sources (e.g. inland waters, geological, permafrost, vegetation, etc.) is 185 CH_4 yr^{-1} (Saunois et al. 2020). Reduction of the differences between the two estimate methods is of prime importance to constrain the global methane budget more accurately. About 5% of the atmospheric CH_4 uptake is by the methanotrophic bacteria present in unsaturated oxic soil, with the main sink being chemical reactions in the atmosphere (Saunois et al. 2020). The CH_4 emission contribution from land types is calculated as the product of emission flux density and the surface extent of CH_4 source/sink area (Bohn et al. 2015, Poulter et al. 2017). The seasonal and inter-annual variability of the areal extent of these land types is considered to be the main cause of uncertainty in calculating their absolute flux of CH_4 emissions, which is significant with respect to the global CH_4 budget (Poulter et al. 2017, Bohn et al. 2015, Desai et al. 2015). Here, we show for the first time that the sensitivity of CH_4 emission to variations in land types can be detected by a space instrument, i.e. TROPOspheric Monitoring Instrument (TROPOMI) which is onboard the European Space Agency's (ESA) Sentinel-5 Precursor (S-5p) satellite measuring daily global total column concentrations of atmospheric CH_4. We further show how this information can be used to identify the inherent land types responsible for such positive or negative emissions and finally help in the creation of a land type classification map for a region.

6.2 Satellite-based CH_4 total column data used in this study

The CH_4 total column measured by the satellite is a combination of CH_4 production, oxidation in the atmosphere (or soil uptake), and transport. S-5p measures CH_4 total column concentrations with a wide swath of 2600 km and a ground pixel of $7 \times 7km^2$ (upgraded to $5.5 \times 7km^2$ since August 2019) in exact nadir (see section 6.5.1) (Veefkind et al. 2012). The bias-corrected CH_4 total column data (Hasekamp et al. 2019) along with the retrieved surface albedo (SA) for the recommended quality filter were selected and spatially binned on a regular 0.05° grid (Section 6.5.1) for 2018 and 2019 in this study.

6.3 Selection of a favourable study region

Wetlands cover approximately 5.5% of the global land surface with an average areal extent of 8.0 to 8.4 million km^2 (Saunois et al. 2020). Apart from the ecological significance, wetlands store atmospheric carbon and act as a carbon sink. The CH_4 production in wetlands is influenced by the spatial and temporal extent of anoxia (water level in the soil), temperature, availability of substrate, and plant ecology (Valentine et al. 1994, Wania et al. 2010, Whalen 2005). Monitoring of these wetlands using remote sensing is a resource and time-efficient endeavour with significant ecological and environmental importance. A large section (25%) of world's remaining wetlands are located in Canada, covering

12.9% of Canada's terrestrial area (Warner & Rubec 1997, *Environment and Climate Change Canada. Canadian Environmental Sustainability Indicators: Extent of Canada's Wetlands.* 2016). Recently, Amani et al. 2019 created a Canada wetland inventory (CWI) using a composite of approximately 30,000 Landsat-8 surface reflectance images collected from 2016 to 2018. This method allows monitoring and mapping of wetlands every three years with 66% producer and 63% user accuracies. The CWI map included five wetland classes defined by the Canadian Wetland Classification System (CWCS) – bog, fen, marsh, swamp, and shallow-water – as well as other land types – forest, grassland, cropland, barren, deep-water, and snow (Figure 6.1a left inset). We have chosen Canada as our study region due to the presence of large wetland areas (and other land types), and the availability of CWI map for Canada for verifying our results.
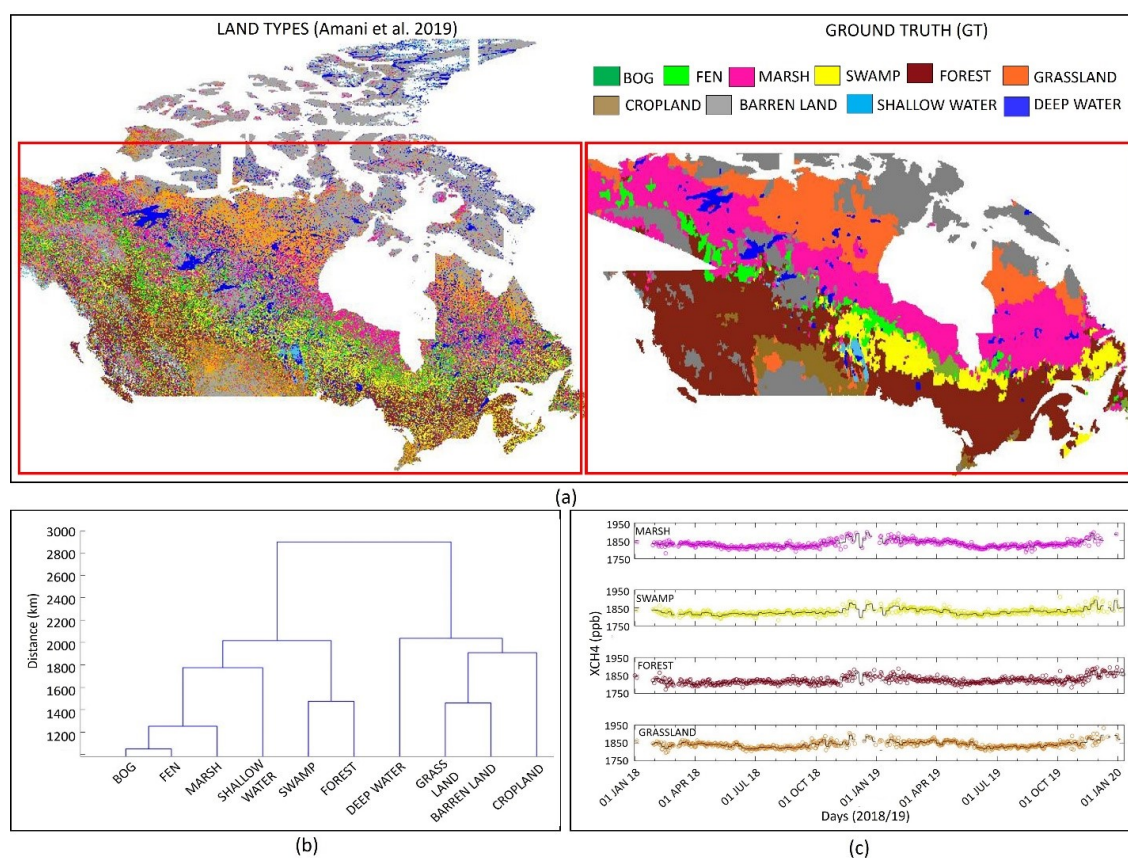
Figure 6.1: Land type classification map (ground truth) creation for Canada and its classification based on methane emission sensitivities. (a) (left) Land types in Canada as described in Amani et al. 2019 (10 classes) at 30 m spatial resolution and (right) ground truth (GT) created using the MODIS NDVI product and graph cut segmentation at 5.5km spatial resolution. (b) An unsupervised clustering of land types, Dendrogram depicting the inbuilt hierarchical relationship that exists in the data. (c) Time-series (Jan 2018 – Dec 2019) distribution of the CH4 (ppb) for marsh, swamp, forest, and grassland (dots), and their weekly average (solid line) – depicting the similarity and points of difference in the dataset.

The CWI map is available at a significantly higher spatial pixel resolution of 30 m compared to the binned S-5p resolution at 0.05° grid ($\approx 5.5km$). The CWI map was therefore

upscaled to a lower resolution map combining additional Moderate Resolution Imaging Spectroradiometer (MODIS) normalised difference vegetation index (NDVI) (Didan 2015) product utilising a classification-segmentation algorithm (Real & Vargas 1996) (Section 6.5.2). Some of the islands (far north) were not considered due to poor or none availability of S-5p data during long winter periods, in which the area was covered in snow/ice, and/or clouds limiting the satellite's view. The smoothed segmented map created for the selected area (Figure 6.1a right inset) was used as the ground truth (GT) in this paper (section 6.5.2).

6.4 Seasonal and spatial variations of S-5p CH_4 total column

Time-series of S-5p CH_4 total column concentrations over the four land types (marsh, swamp, forest, and grassland) are shown in Figure 6.1c, with gaps indicating missing data during November-January (winter months). In Figure 6.2, boxplots show the timeline of S-5p CH_4 total column for 2018/2019 (black/red) and the scatter over the individual land types (GT), and indicate little yearly difference between land types except for forest area with the lowest values.
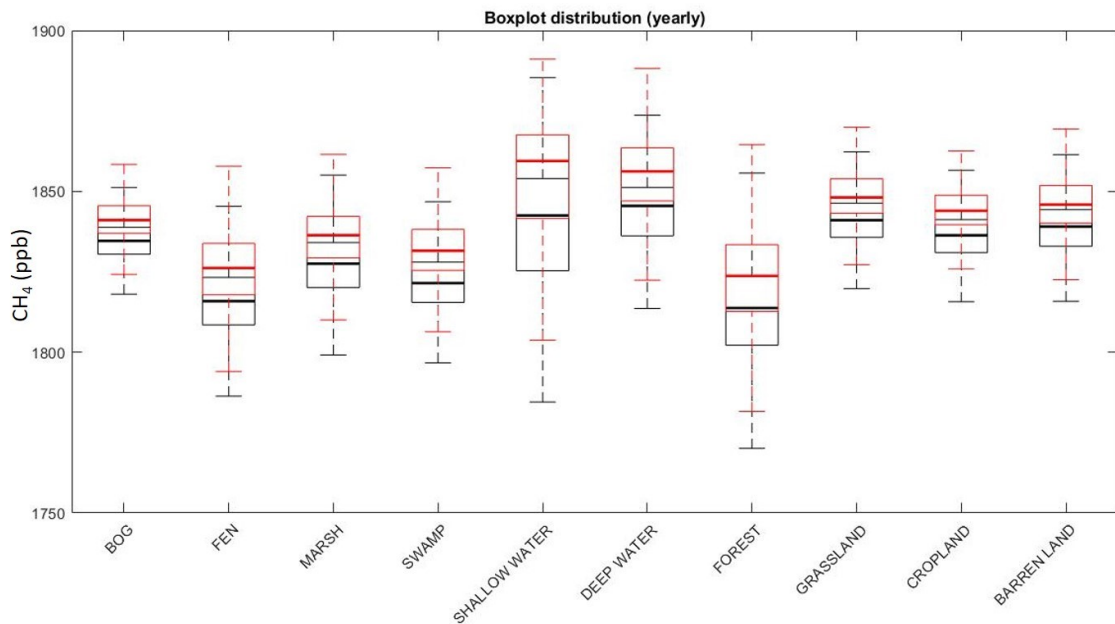
Figure 6.2: Boxplots made using the average of the yearly CH4 total column values over land types as captured by S-5p. It can be seen that all the communities have an average of 1825 ppb, and ranges from 1775 – 1900 ppb. Communities like bog, marsh, fen, swamp are almost identical (all being peatlands) and non-peatlands such as grassland, cropland, and barren-land also exhibit similar overall distribution. This also indicates that seasonality is important in proper identification.

Although the absolute total column values indicate lack of uniqueness, the entire dataset over two years reveal significant distinction as seen in the dendrogram plot of hierarchical clusters present within the dataset (Figure 6.1b). Within the two main clusters of the dendrogram, all wetlands and forest area are segregated from other land types. The uniqueness of S-5p CH_4 total column values from each land-type are signified by the

height of each dendrogram leg. Bog and fen appear indistinguishable and show features related closely to marsh. Swamp and forest also show inter-mixed features, most likely due to their spatial proximity. For similar reasons, grass and barren-land types show indistinguishable features indicating that CH_4 total column values from neighbouring land types may not be identifiable as distinct.

S-5p CH_4 total column and SA values were available for all key land types as classified in CWI map (Amani et al. 2019). Area covered by snow, as obtained using MODIS snow product, interfered with the capture and visibility of land types decreased in all cases with increased snow cover except for water surfaces, which were only visible when covered in snow (Figure 6.5).

6.5 Methods

The method for preparing Sentinel-5 Precursor (S-5p) methane (CH_4) total column and surface albedo (SA) data, along with the brief description of the machine learning (ML) algorithm, which was utilised to analyse the data is described here. The ML algorithm used was initially developed for identifying vegetation communities within wetlands using remote sensing (Didan 2015) and the steps for customising the algorithm for detecting the areal extent of land types analysing S-5p data is described here.

6.5.1 S-5p CH_4 total column and surface albedo data

S-5p with the TROPOspheric Monitoring Instrument (TROPOMI) onboard was launched on 13 October 2017 (Veefkind et al. 2012). It is orbiting the Earth in a near-polar sun-synchronous orbit at an altitude of 824 km with an ascending node equator crossing at 13:30 local time. TROPOMI is a nadir-viewing grating spectrometer measuring the solar radiation reflected by the Earth and its atmosphere in eight spectral bands from the ultraviolet (UV) to the short-wave infrared range (SWIR). S-5p has an orbit cycle of 16 days and covers the Earth with 14 orbits per day. The push-broom configuration with the imaging capabilities allows a wide swath of 2,600 km, which results in a daily global coverage. S-5p is one of European Space Agency's (ESA's) first atmospheric composition Sentinel missions relevant for air quality and climate monitoring. The primary atmospheric constituents measured are ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO), formaldehyde (CH_2O), methane (CH_4), sulphur dioxide (SO_2), aerosol, and clouds. The vertically integrated abundances of CH_4 are retrieved from the SWIR (2305-2385 nm) spectral channel3. The spatial resolution of the operational level 2 SWIR product was originally $7 \times 7 km^2$ in exact nadir and was increased to $5.5 \times 7 km^2$ on 6 August 2019. The operational processing to retrieve the column averaged dry air mixing ratio of CH_4 is performed by RemoTeC S5 algorithm (Hasekamp et al. 2019). The operational CH_4 total column product consists of a standard product and a bias-corrected product. The details of the bias correction is described in the Algorithm Theoretical Baseline Document (ATBD) (Hasekamp et al. 2019). The latest product version of the S-5p bias-corrected CH_4 total column data from Jan 2018 until Dec 2019 has been used in this study. This period also includes the data during the commissioning phase of the satellite (Jan – end April 2018). The quality of the data has been verified by the ESA mission performance centre (MPC) by performing validation against reference ground-based remote sensing networks of the Total Carbon Column Observing Network (TCCON) and the Infrared Working Group (IRWG) of the Network for the Detection of Atmospheric Composition Change (NDACC) (Sha & Langerock 2019, Lambert et al. 2020). As S-5p records solar

absorption measurements reflected by the Earth's surface and the atmosphere, measurements are not possible over larger parts of Canada during winter months (Nov-Jan).
The S-5p bias-corrected CH_4 total column values along with the retrieved surface albedo (SA) for qa value greater than 0.5 were selected and binned on a regular 0.05° grid to form the level 3 (L3) data. The harp component of the ESA atmospheric toolbox (`https://atmospherictoolbox.org`) was used to perform the latitude, longitude regridding where each S-5p pixel contributes to the regridded CH_4 value of the target grid cell if there was an overlap of the pixel and the grid cell. In case when multiple pixels are overlapping, a grid cell weighted average was taken using the overlap area as the weight.

6.5.2 Classification-segmentation machine learning algorithm

The level 3 regridded S-5p CH_4 total column and SA over Canada were analysed utilising a machine learning algorithm. The dendrogram created using the CWI describes the degree of dissimilarity between the clusters of land types. This dissimilarity was measured in the form of Euclidean distance between the centroids – depicting how close/far (in terms of CH_4 total column values) the land types exists. The workflow, including development of ground truth (GT) maps, creation of monthly yearly maps, and performance evaluation of the algorithm, is described in Figure 6.3.
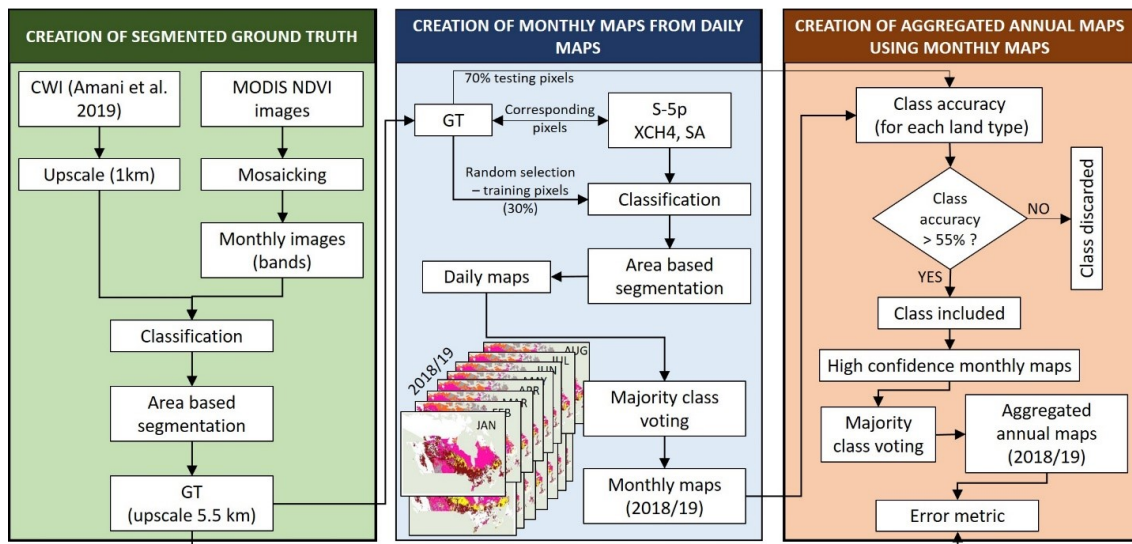
Figure 6.3: Flowchart showing the complete process from the creation of the ground truth (GT) to the creating of annual land type classification maps.

The first step of the analysis was the creation of a GT map for evaluating the sensitivity of S-5p CH_4 total column measurements to certain land types, especially wetlands. The basis of the GT map was the Canadian wetland inventory (CWI) created by Amani et al. 2019 and Moderate Resolution Imaging Spectroradiometer (MODIS) remote sensing data (Didan 2015). The spatial resolution of CWI, at 30 m per pixel resolution, is significantly higher resolution compared to the spatial resolution of the S-5p CH_4 data, which is on a 5.5 km grid. Therefore, the pixel-based (PB) CWI map was upscaled to lower resolution into a smoother segment based map combining relevant information from MODIS. The MODIS NDVI (Normalized Difference Vegetation Index) (MOD13A3) (Didan 2015) produces monthly NDVI maps at 1 km resolution with about 15 tiles covering the area

of interest (AOI) in Canada marked with a red rectangular box in Figure 6.1. All of the NDVI images were mosaicked using mean value (for the overlapping areas) for the 24 months of the study period. Therefore, using layer stacking a 3-dimensional image with 24 bands was created for the AOI. In order to create proper segments, the CWI map was also upscaled to 1 km resolution such that it is compatible with the MODIS NDVI product using nearest neighbour interpolation. The CWI map was segmented using features from NDVI maps utilising the proposed classification-segmentation algorithm (Bhatnagar et al. 2020a) described here.

Description of the algorithm

The decision trees such as classification and regression trees (CART) (Breiman et al. 1984) are simple classifiers, especially useful when dealing with a large amount of data. CART is often unstable to small deviation in the system. Therefore, to remove this instability and to make the trees robust and invariant, a method called "bootstrap aggregating" or "bagging" is used (Bauer & Kohavi 1999). Here, the samples are divided into T different subsamples (with repetition), and CART is applied separately for each subsample (Z). Therefore, for each pixel, the label is obtained using the plurality of the votes of T trees. This technique of application of CART using Bagging is known as Bagged Tree (BT) classification. For this study, T = 100 was applied to get an accurate result. Equation 6.5.1 gives the formula used for creating T trees using BT for each bootstrap sample Z^T, and n number of classes.

$$F(x) = \frac{1}{T} \sum_{b=1}^{T} z^b \qquad (6.5.1)$$

Where T=100 and F(x), $x \in n$ is the final classified map. This process was repeated daily for the years 2018, 2019. One of the trees is shown as an example in Figure 6.4. A PB classification only takes into account the reflectance value of a pixel and no contextual information. To detect the areal extent of certain land types, the study was extended to the traditional PB classification to segment-based (area-based) mapping using Graph cut min-cut (maximum a-posteriori) algorithm (Boykov et al. 2001). Maximising the a-posteriori is directly related to minimising the energy, i.e., reducing the data cost and the smoothness cost of the segments being formed. The data cost ($D(p, n_p)$) is related to assigning the best cost for a pixel p to have class n_p. The $D(p, n_p)$ used for this study was the confidence scores vector of the size $p \times n$. It is the posterior probability linked with each pixel for every class. Whereas, the smoothness cost ($Vp, q(n_p, n_q)$) is related to assigning the same labels to the neighbouring pixels (Boykov et al. 2001, Bhatnagar et al. 2020a) as per the number of classes (n) present so as to maximise the gradient between the pixels of differing classes. It was a matrix of size n×n. Maximising posteriori is directly related to minimising the energy (E) (Boykov et al. 2001), described in Equation 6.5.2, which was used for forming the best segments in this study.

$$E = \sum_{p} D(p, n_p) + \sum_{p,q} Vp, q(n_p, n_q) \qquad (6.5.2)$$

The smoothness parameter (part of the smoothness cost) plays a vital role in deciding the final segments and boundaries of the communities being formed. For a smaller region, such as a small wetland, smoothness parameter ranging from 0.5-1.75 is generally sufficient (Bhatnagar et al. 2020a). Whereas, for an area with more details (and gradient), a higher value of smoothness is required (Bhatnagar et al. 2020b). For this study, after a trial of
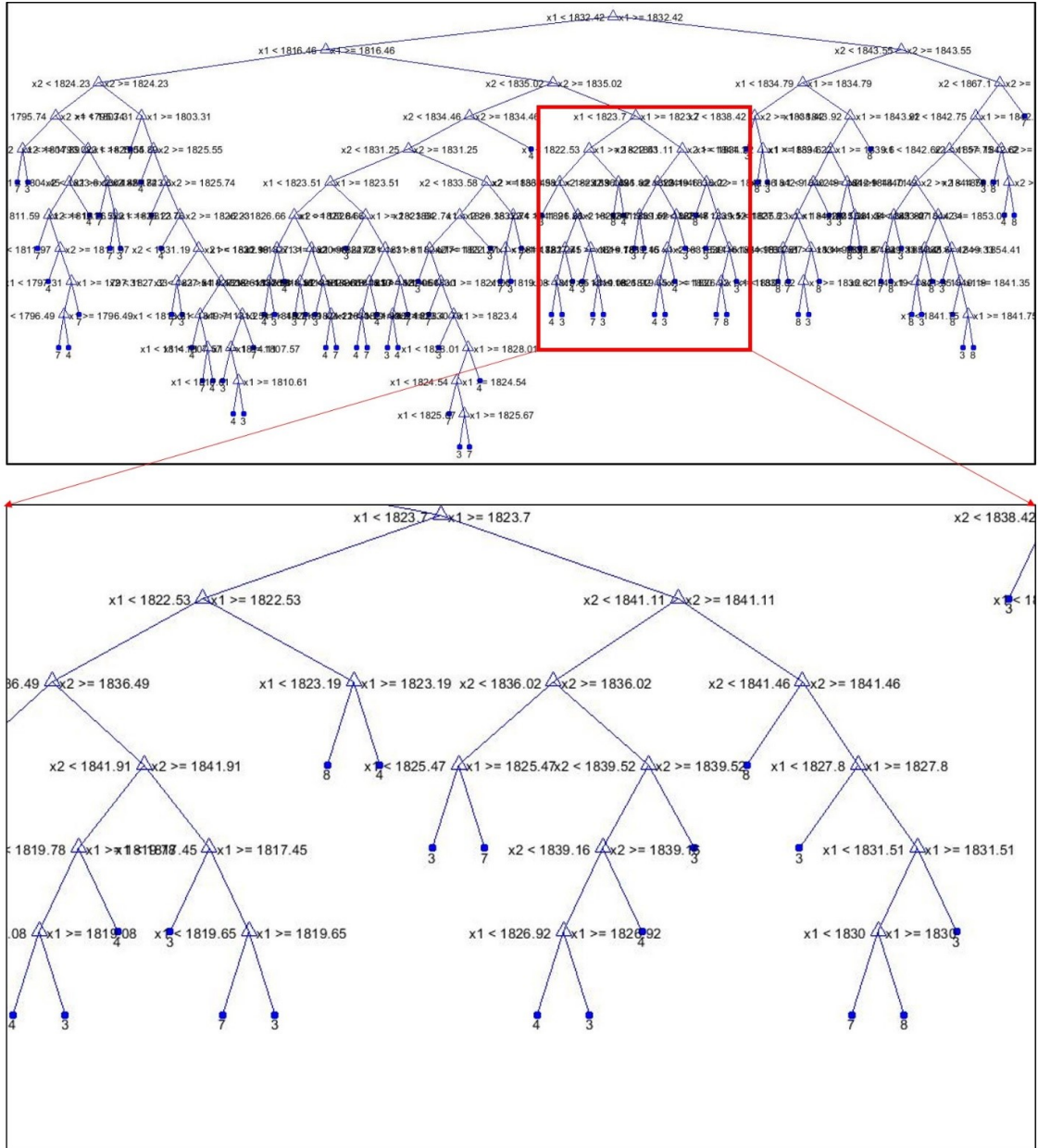
Figure 6.4: Bagged Tree Classifier (example of one tree (for 500 data points), and further leaf nodes zoomed). It can be seen that the data is very intermixed, and the division happens at a very minute level. This is challenging for the human eye to detect but the machine learning algorithm are able to pick these small details.

multiple smoothness parameter (c), c = 10 was finalised. A GT map was created using the BT classification, and subsequent Graph cut segmentation algorithm at 1 km pixel resolution. The map was upscaled to 5.5 km spatial resolution using nearest neighbour interpolation for upscaling it to the same resolution as the L3 S-5p data.

Using the segmentation model mentioned above, every pixel under AOI was mapped at least once every month. Therefore, a total of 679 daily maps were created for years 2018 and 2019. Using majority voting (Jimenez et al. 1999) for each pixel mapped in the daily maps - monthly maps were created, i.e., for every pixel $p \in N$ a class $x \in n$ would be assigned if,

$$\sum_{p=1}^{N} \widehat{F_p}(x) = \max_{x=1}^{n} \sum_{p=1}^{N} F_p(x) \tag{6.5.3}$$

where N are the total number of pixels, and $\widehat{F_p}(x)$ is the majority voted map at the end of each month for the years 2018, 2019. Pixels that were not mapped for any given day for that month were removed. Furthermore, only the covered/mapped area was used for further accuracy analysis for each land type (class). Class Accuracy (CA) is the ratio of the diagonal vector of the class under consideration with the total number of pixels belonging to the same class.

$$CA = \frac{TP}{TP + FN} \tag{6.5.4}$$

Eventually, a majority voting was applied on monthly maps (separately for two years) to obtain the final annual aggregated map. For every month, only the pixels with $CA \geq 55\%$ were selected to form a high-confidence map for that month and classes like bog, fen, deep water, shallow water were omitted in these maps due to low CA values. The monthly high-confidence maps were again combined (using majority voting) to form the final aggregated map for a year. The algorithm was trained using 30% randomly selected pixels from the GT map and the performance of the algorithms were tested on the remaining pixels (see Table 6.2 for confusion matrix). A kappa value of 0.69 was achieved, which are comparable to the kappa value (0.66) reported by Amani et al. 2019.

The accuracy of areal detection should not only be checked by typical PB evaluation metrics such as producer and user accuracy (Stoer & Wagner 1997). A set of error metrics linked with location and extent of land type detection was calculated for the annual maps for both years while comparing it with the GT of the captured region.

Error Metric

- Jaccard Similarity Index (J) measures the similarity between the members of the two sets and reports the amount of similarity and distinction.

- Area (A) estimates the total area of the selected land type. The area of every individual pixel is determined by looking at its 2×2 neighbourhood. Each pixel is part of four different 2×2 neighbourhoods, which indicates the change in overall growth/shrinkage of the community.

- Orientation (O) gives the angle between the x-axis and the major axis of the ellipse (covering the entire land type). It can range from -90 to + 90 degrees, indicating the direction of the change of the land type.

- Extent (E) indicates the ratio of total pixels present in the bounding box to the total pixels present in the image. The bounding box represents a box (rectangle/square) covering the major cluster of pixels present for a land type in an image.

The values of these metrics for both years were presented in Methods Table 1. The performance of the algorithm for detecting each land type showed the sensitivity of S-5p CH_4 total column and SA to the land type in 2018 and 2019 as presented in Table 6.1 , Table 6.2 and in Table 6.3.

6.6 Results and discussion

6.6.1 Detection of sensitivity of CH_4 emissions to land types

The S-5p CH_4 total column along with SA from gridded pixels were analysed together and separately using a classification-segmentation algorithm for each available day. Algorithm steps, training data (30%) and all details are described in section. The analysis generated daily maps from the testing data (70%) showing the extent of 10 different land types over 365 days in 2019 and 314 days in 2018 (51 missing days mostly during the S-5p commissioning phase). The daily maps during a calendar month were combined to create a monthly map where each pixel was identified as the land type using majority voting (Section 6.5.2). The time-series of the overall accuracy (OA) values for each land types calculated in comparison with the GT maps (Methods section provides details of algorithm performance evaluation) are shown in Figure 6.5 along with the SA for the respective land types. The OA improved with the inclusion of SA data for the majority of land types, especially for swamp and cropland when compared to the analysis considering only S-5p CH_4 data (Figure 6.5 all sub-plots). Land types with large areal extent, such as marsh, forest, grassland, swamp, and cropland showed high detectability (OA > 60%), while wetland types such as bog and fen showed low OAs due to low areal extent and proximity to other dominant land types. Bog, fen, and swamp were often misclassified due to their intermixed land distribution. In winter months, OA decreased due to the lack of S-5p data and were omitted from the plots. Marsh was detected with the highest OA, with variations in accuracy linked with lack of available pixels. Similarly, for grassland and barren-land, OAs were linked with the area covered by S-5p with grassland showing better detectability. The other key wetland-type, swamp, showed better detectability in spring and autumn compared to summer when it was misclassified as forest during the growing season. Forest was detected with reasonable OA which slightly reduced with melting of snow cover. The inclusion of SA improved the detectability of cropland significantly. Official S-5p CH_4 total column data is currently unavailable over water, and so, shallow and deep-water regions could only be detected when covered by snow/ice. The detectability of most dominant land types utilising CH_4 data thus indicates a significant difference and sensitivities of CH_4 emissions between land types.

The seasonal variations of the CH_4 emission sensitivities along with SA from land types are illustrated by the delineation of different land types during different seasons of 2018 (Fig. 6.6a) and 2019 (Fig. 6.6c). Land types detected with all OA values were plotted in monthly maps, with any missing or non-detectable pixels shown as white in Fig. 6.6. The sensitivities of the method for distinguishing betwee marsh, swamp, grassland, and barren-land were strongest during March-May while sensitivities to cropland and neighbouring land types were strongest during May-June. The yearly maps for 2018, 2019 were created using majority voting of the monthly maps during the calendar year, only including land types with $OA > 55\%$ (Fig. 6.6 b, d). The areal extent of marsh, swamp, forest, and grassland were identified with high confidence (Fig. 6.5), and in all cases with better detectability than traditional methods used in remote sensing using satellite imagery validated using field information (*Environment and Climate Change Canada. Canadian*
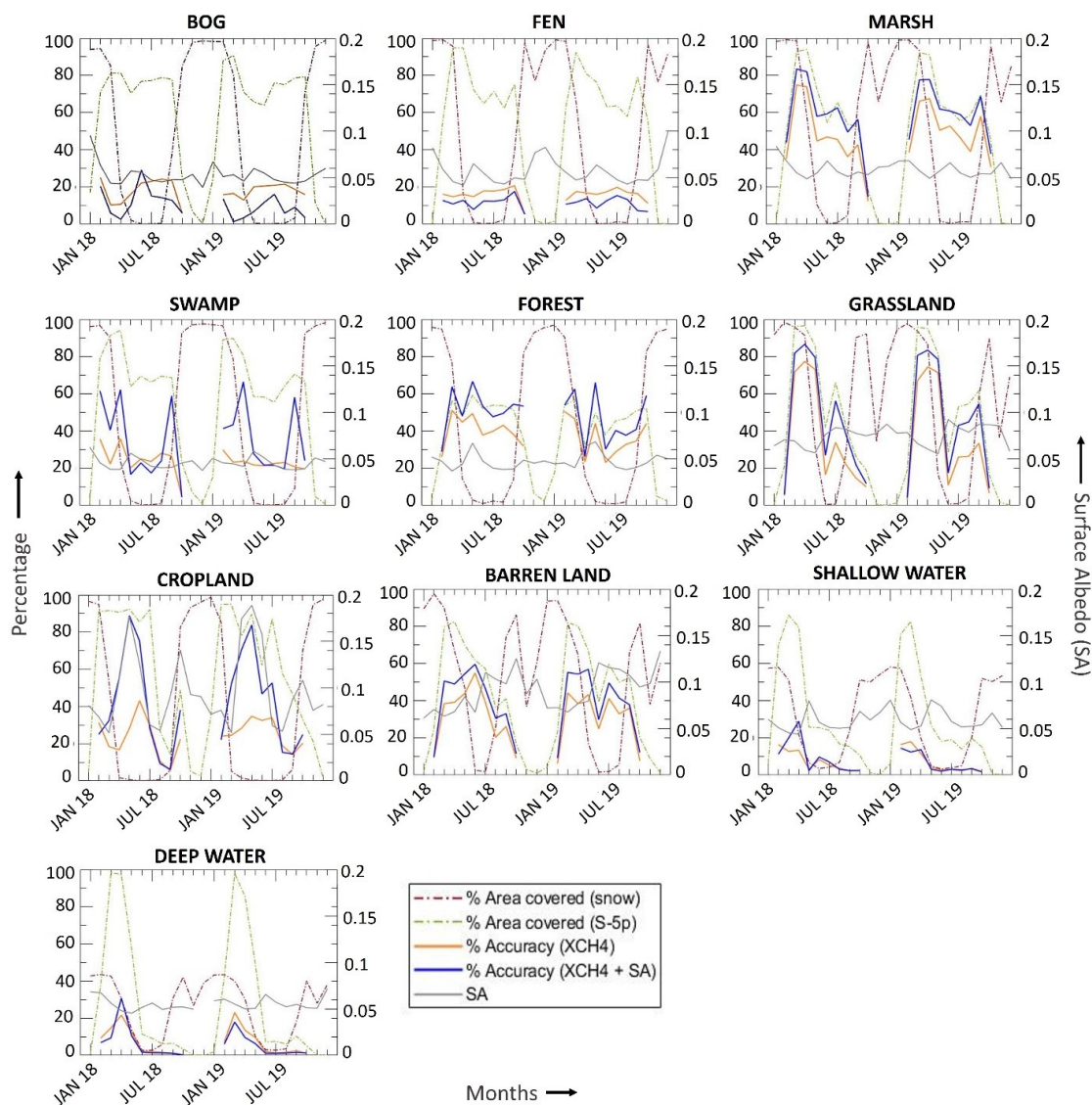
Figure 6.5: Time series plots of overall accuracies for different land types. Representation of % area covered by snow (dashed red), S-5p (dashed green), overall accuracy (in %, orange) achieved for land types classification using proposed methodology using S-5p CH4 (orange), and S-5p CH4 + SA products (blue) features (left y-axis) and surface albedo (right y-axis) value for each land type over the time period Jan 2018 – Dec 2019 (x-axis). The % area covered by snow is obtained using MODIS daily snow cover product (MOD10A1), the area covered by S-5p is the monthly average of the area captured by S-5p.
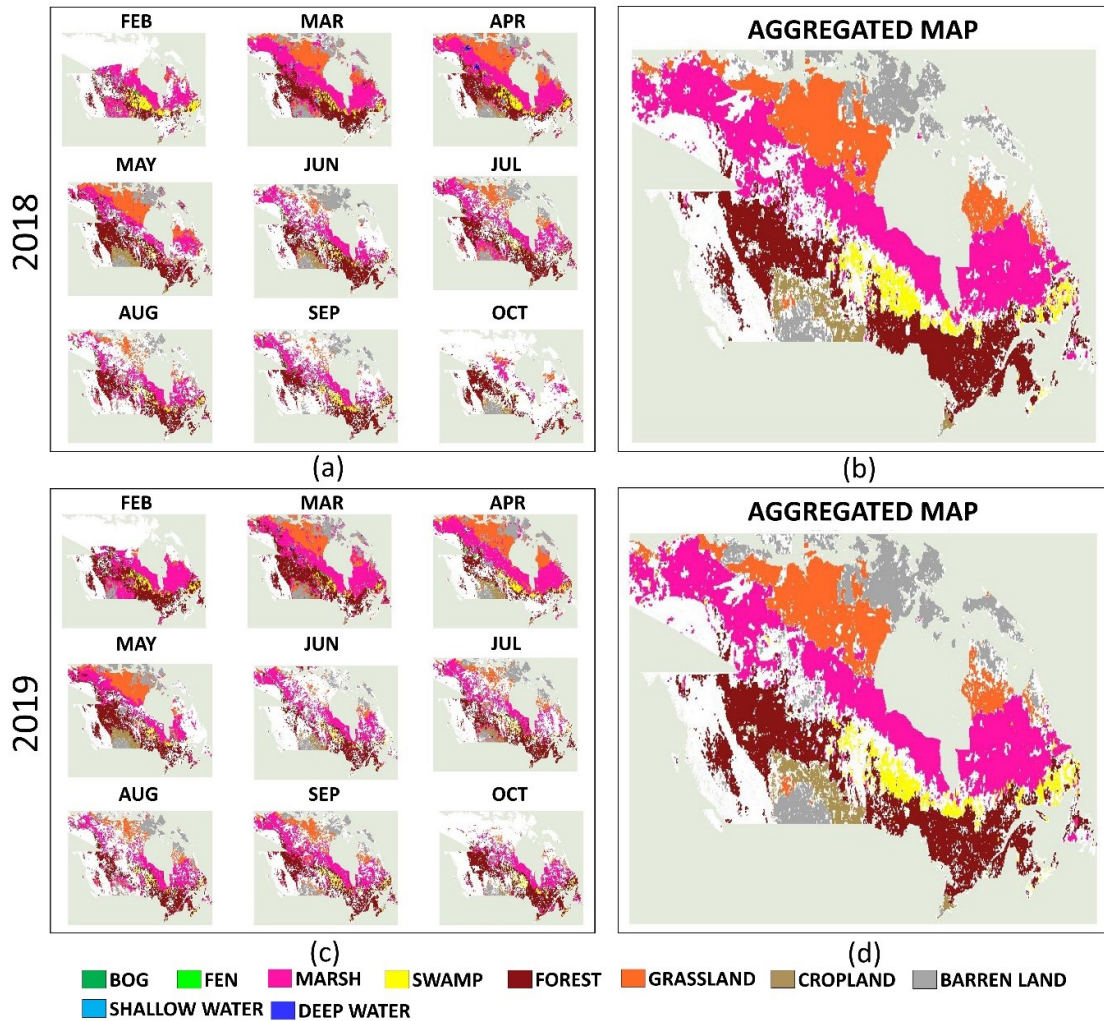
Figure 6.6: Land type maps created using CH4 and SA data as input. a/c, 2018/2019 – seasonal classified maps created using combining the daily images obtained from S-5p, the missing area (white/blank) was not covered by S-5p for that month. b/d, 2018/2019 – Aggregated maps created using the pixels with accuracy $\geq 55\%$ over the months (majority voted) for each year separately. These represent the area with high confidence for both 2018, 2019.

Environmental Sustainability Indicators: Extent of Canada's Wetlands. 2016, Amani et al. 2019) (Table 6.1, producer and user accuracies).

Aggregated Maps (2018/19)	Present Study		Amani et al. (2019)		Present Study	
	Identified region	Yr. 2018/19	Yr. 2017/18/19		Yr. 2018/19	
Land types	Area (1,000 km2)	% △ Area	Producer Accuracy (%)	User Accuracy (%)	Producer Accuracy (%)	User Accuracy (%)
MARSH	602.22	13.90/0.89	72.4	64.2	94.01/96.07	81.56/97.36
SWAMP	131.63	28.67/30.23	61.6	59.8	65.52/56.41	96.63/85.31
FOREST	544.98	9.02/12.23	70.6	75.9	87.02/86.37	96.16/100
GRASSLAND	354.01	2.99/9.67	62.2	85.1	90.23/89.17	87.44/100
CROPLAND	75.81	20.10/20.21	72.6	62.3	70.85/78.12	91.63/100
BARREN LAND	334.86	39.09/16.54	90	84.6	47.47/81.47	100/100

Table 6.1: Geometrical error metric and accuracies for the union of the area covered in aggregated maps of 2018, 2019 (confident land types). The % △ signifies the absolute change in the parameter value being identified in comparison to the original value of the parameter in the GT. △A signifies the change in the area. Producer and User accuracy (2018/19) is directly compared against accuracies (combined 2016-2018) presented in Amani et al. 2019.

	BOG	FEN	MARSH	SWAMP	SHALLOW WATER	DEEP WATER	FOREST	GRASSLAND	CROPLAND	BARREN LAND
					2018					
BOG	625	4	1656	158	1	1	928	22	9	115
FEN	5	1907	4833	458	1	2	4203	59	7	98
MARSH	18	64	63601	409	5	26	3700	2903	44	759
SWAMP	11	31	3707	9426	5	3	7915	102	28	164
SHALLOW WATER	0	1	204	114	222	3	184	23	39	77
DEEP WATER	1	2	1552	59	8	683	271	759	19	346
FOREST	15	45	4705	446	10	8	62585	314	369	556
GRASSLAND	2	4	3840	61	4	14	454	30272	510	2573
CROPLAND	3	3	3129	64	3	2	780	378	6263	806
BARREN LAND	5	11	6876	165	7	13	1575	3987	433	33134
					2019					
BOG	403	3	2282	92	0	1	917	81	30	21
FEN	2	1820	5347	176	0	3	4364	136	38	48
MARSH	10	59	69014	208	2	19	4266	2695	99	714
SWAMP	10	33	5033	8568	4	3	7853	82	70	70
SHALLOW WATER	0	1	261	10	143	2	149	123	47	25
DEEP WATER	0	4	1638	29	3	499	304	753	53	235
FOREST	9	38	5651	267	6	9	59373	437	353	341
GRASSLAND	1	3	4266	11	1	6	403	31057	455	2329
CROPLAND	1	1	2479	14	3	2	999	497	6628	986
BARREN LAND	2	10	6158	47	2	8	1737	4010	436	34145

Table 6.2: Confusion Matrix for 2018 and 2019. The x-axis describes the Predicted Class, and the y-axis describes the True Class.

The land types with large areal extent were generally detected well with high accuracies except in case of barren-land, which showed low producer accuracy indicating its low sensitivity to CH_4 emissions. The bog and fen wetlands with lower areal extent, were misclassified as marsh (Table 6.2). Similarly, some pixels in the swamp were misclassified as forest during summer periods (June-August). Land types other than wetland (cropland and barren-land) were best identified in summer with good boundary delineation and grassland, although adjacent to marsh, was well distinguishable throughout the year indicating the sensitivity to the difference CH_4 for these land types.

Aggregated Maps (2018/19)	2018/19	Identified region	2018/19	Identified region	2018/19
Land types	Jaccard	Orientation (°)	% △Orientation	Extent	% △Extent
MARSH	0.79/0.94	-11.61	1.89/0.04	0.104	2.61/1.20
SWAMP	0.68/0.54	-4.33	9.80/15.40	0.059	25.59/37.79
FOREST	0.85/0.88	-9.9	5.91/0.045	0.119	10.92/11.76
GRASSLAND	0.82/0.90	-9.86	3.74/0.82	0.092	1.62/10.27
CROPLAND	0.70/0.80	-15.5	4.87/8.68	0.027	10.29/22.86
BARREN LAND	0.51/0.83	-7.36	34.43/9.67	0.071	38.16/16.01

Table 6.3: Geometrical error metric and accuracies for the union of the area covered in aggregated maps of 2018, 2019 (confident land types). Jaccard represents the 2D similarity when maps are overlapped directly. $\triangle E$ change in extent, and $\triangle O$ change in orientation.

Detection of the areal extent of land types based on S-5p CH_4 total column combined with SA was done utilising the proposed machine learning (ML) algorithm where the efficiency of detection was investigated using a set of areal error metrics. Jaccard similarity index (J) (Real & Vargas 1996), area (A), orientation (O) and extent (E) (Table 6.1, 6.3). Good detection was seen for all six key land types while the variability of these metrics was mostly attributed to lack of availability of S-5p pixels, which were often due to inimical meteorological conditions.

6.7 Conclusions

This work demonstrates that the S-5p CH_4 total column data with a machine learning algorithm can reveal unique sensitivity to certain land types, especially marsh, forest and grassland. Analysing such CH_4 data along with derived surface albedo, the areal extents of six land types (following CWI), including two major wetland types (marsh and swamp) covering $\approx 60\%$ of the total wetland area of Canada, were identified for two consecutive years 2018, 2019. As the vegetation appearance of land types, especially wetlands, can vary seasonally, mapping it solely using aerial photography or satellite imagery may lead to errors because of a lack of consistent vegetation pattern(*Environment and Climate Change Canada. Canadian Environmental Sustainability Indicators: Extent of Canada's Wetlands.* 2016). The CWI generated using S-5p data in this study, is complementary to the traditional methods of land type identification showing daily, monthly, seasonal, and yearly changes (*Environment and Climate Change Canada. Canadian Environmental Sustainability Indicators: Extent of Canada's Wetlands.* 2016, Amani et al. 2019). These maps can be used by the WAD2M (Wetland Area Dynamics for Methane Modeling) to either verify or complement their data where measurements from other sources are not available. The study presents an entirely new application of satellite-based CH_4 product illustrating its potential for land type identification of large areas, monitoring, and studying the dynamic change over time as well as helping to constrain global methane emission models.

Chapter 7

Hydrological Modelling of Turloughs

"The individual is ephemeral, races and nations come and pass away, but man remains."

— Nikola Tesla

Contents

This chapter corresponds to the paper is submitted to IAHS Hydrological Sciences Journal [6].

7.1   Introduction

Wetlands can be recognised as transitional (both in space and time) ecosystems, or ecotones, between terrestrial and aquatic ecosystems (Mitsch & Gosselink 2015). The ecosystem services normally associated with wetlands are water supply and purification, flood

and erosion control, carbon storage and sequestration (and links to climate), and habitat preservation (Assessment 2005, Barbier 2011). However, about half of global wetland areas have been lost, and much of the remaining wetlands are degraded (Zedler & Kercher 2005). It is therefore crucial that accurate methods of assessing the ecological health of such wetlands are developed in order to feed into appropriate management decisions for their preservation and protection with respect to past, current and future anthropogenic impacts, including the potential impacts of a changing global climate. The presence or absence of a plant species can be used as a bioindicator and can provide information on the environmental conditions in the habitat in which it is found. One commonly used method in Europe of quantifying the relationship between indicator plant species and various environmental variables, for example, is the Ellenberg index (Ellenberg 1988; Hill et al. 1999). The type of vegetation/ecology in wetlands and its location is intimately entwined with the hydrological conditions, and hence, these ecohydrological relationships need to be properly understood and evaluated, from which healthy envelopes/metrics can then be defined for the different key wetland habitats. Such metrics can be used in order to assess the ecohydrological status of wetlands and monitor and manage their current and future existence. Regan et al. 2019, for example, have recently evaluated the ecohydrological envelope of different ecotopes on raised bogs in Ireland, using water table duration curves to show that active areas of sphagnum growth occur where water tables are within 0.1 m of the ground surface for approximately 90% of a given year.

This research focuses on ephemeral lakes that form in shallow depressions in karst areas, mainly in the west of Ireland, known as turloughs. Turloughs generally flood (and drain) from their lowest topographic point, often via estavelles linked to the main karst conduit networks (Naughton et al. 2012). This intermittent flooding of the basin produces a distinct hydrological gradient which produces a linked vegetation gradient. Such wetlands are designated priority habitats under the EU Habitat's Directive and are considered as groundwater-dependent terrestrial ecosystems (GWDTEs) under the EU Water Framework Directive (CEC 2000). The continually changing environment of turloughs means that they are more accurately considered as ecotones rather than ecosystems; that is, transitional zones between aquatic and terrestrial systems.

There have been many previous studies on turlough geomorphology, hydrology, ecology, and conservation importance (e.g. Coxon 1986; Skeffington et al. 2006) with summaries of turlough wetland plant and freshwater habitats and communities described in Goodwillie & Reynolds 2003. These were followed by a more integrated project (Waldren et al. 2015) studying the conservation status (vegetation habitats, invertebrates, water quality, hydrology etc.) of 22 turloughs. These were selected to provide a representative range across the different types of turlough found in Ireland primarily from the perspective of hydrogeological variation (conduit driven or shallow epikarst), depth and duration of flooding etc., but were further characterised later according to water quality, soil type and land-use.

Seasonally or intermittently flooded wetlands present a unique environment for plants that are able to live under such conditions, being affected by hydrology, disturbance, spatial heterogeneity and productivity (Pollock et al. 1998). From a hydrological perspective, duration (hydroperiod), depth and frequency of flooding have been shown to have the greatest effect on the ecology of wetlands (De Becker et al. 1999; Casanova & Brock 2000; Thompson & Finlayson 2001). More specifically to turloughs, the hydroperiod (and other linked metrics such as depth of flooding and the timing of the flood recessions) has major effects on both terrestrial and aquatic communities, strongly influencing the distribution of vascular plant species and the zonation of vegetation communities within turlough (Fin-

layson et al. 2018). Hence, this research has evaluated the spatial distribution of different vegetation communities on four turloughs within the same karst network in south Galway, Ireland in relation to their different hydrological conditions experienced over a 28-year period. These analyses have then been used to define broad ecohydrological metrics for the different vegetation community habitats.

7.2 Materials and methods

7.2.1 Study sites (geology, hydrology, water quality, land-use)

The four turloughs selected for study (Blackrock, Coy, Coole, Garryland and Caherglassaun) are located in an extensive conduit karst network catchment in south Galway in the west of Ireland (see Figure 7.1). This karst has formed in carbonate rock from the Carboniferous, which has experienced many episodes of karstification, most significantly during the most recent Cenozoic Era (Drew 2018), and is situated less than 100 m above current sea level.

The total catchment size is $\approx 500 km^2$ of which about one-third of the catchment area is formed by the Old Red Sandstone Slieve Aughty mountains (covered in peat and forestry) which is drained by three main rivers. These rivers feed rapid allogenic runoff down into the lowland karst network. The rest of the catchment receives autogenic rainfall recharge onto the karst which is covered by shallow limestone till. Several studies since have shown the interconnected nature of this south Galway karst conduit network which discharges out into the Atlantic Ocean at an intertidal spring at Kinvara (Naughton et al. 2018).

All four turloughs had been shown to have similar residence times (termed aggregation periods) during the initial 2007 / 2008 period of monitoring study of 38 to 67 days, which are much shorter than were calculated for many of the other turloughs monitored across the west of Ireland during that study (Waldren et al. 2015).

Blackrock, Coy, Caherglassaun and Garryland turloughs in this interconnected lowland karst network contain waters that are a mixture of soft water from rivers draining the Slieve Aughty mountains and hard water from the lowland calcareous parts of their catchments, yielding waters of relatively low alkalinity and also relatively high colour (due to the presence of humic and fulvic materials in drainage from peats). All four turloughs were deemed to be eutrophic-mesotrophic from total phosphorus measurements but mesotrophic- oligotrophic from Chl A measurements taken during the Waldren et al. 2015 study (Cunha Pereira et al. 2010). The development of algal biomass in these turloughs appears not to have been limited by phosphorus in these turloughs (compared to others studied), which was attributed to the more highly coloured waters of these turloughs causing more predominant light-limitations through the winter instead (Havens & Nurnberg 2004). More frequent sampling of these turloughs across the 2011 / 2012 hydrological year showed fairly consistent concentrations of nutrients both in time and between the four turloughs: total phosphorous ranging from 20 to 50 μgg/l, total dissolved phosphorous ranging from 5 to 30 μg/P, total nitrogen 0.25 to 1.2 mg/l. McCormack et al. 2016, although the nutrient concentrations were found to reduce over the flooded period with such nutrient loss processes thought to be occurring within the turlough systems themselves.

Finally, the soil types are very similar between the four turloughs comprised of mineral soils associated with till subsoils. Equally, the land-use is very similar between the four turloughs, with grazing mainly by cattle (with some sheep and horses) at a relatively low intensity during the summer.

7.2.2 Vegetation surveys

Field surveys

Vegetation field surveys were conducted over three field seasons, 2006, 2007 and 2008 across 22 turloughs (Waldren et al. 2015) using 1×1 m quadrats - the majority of vegetation being grassland or short herbaceous vegetation. A minimum of 5 relevés was recorded in each vegetation type. Within each relevé, the vascular plant species present and their cover-abundance were recorded using the Domin score. 28 vegetation communities were then described from multivariate analyses of the relevés taken across the 22 turloughs. These turlough plant community species and community identification keys were used in the field in 2008 for identifying and mapping the vegetation types in the four turloughs focussed on in this study (Blackrock, Coy, Caherglassaun and Garryland), which found 14 vegetation communities in at least one of these four turloughs. Trimble handheld GPS devices (Nomad and GeoExplorer models) were used for field recording and loaded with georeferenced images of all available aerial photos and Ordnance Survey (OS) maps. Boundaries between vegetation types were recorded along the estimated centre of the observed zone of transition between two types of community. They were recorded along the putative boundary at intervals of 5 m, 10 m, 20 m or 30 m, depending on particular local topography and spatial configuration of vegetation. Digital photographs were also taken at ground level in various locations to record the general topography, vegetation, water levels and various other features of the turlough as surveyed on the day; these were subsequently used to help improve the confidence of digital spatial representations of vegetation.

Vegetation community maps

ArcGIS software (*ArcMap Desktop (Version 10.6.1)* 2019) was used to generate digital vegetation maps using GPS data recorded in the field. Differentially-corrected data files were exported as ESRI shapefiles using GPS Pathfinder Office software. Point shapefiles were generated during export and loaded into ArcMap (*ArcMap Desktop (Version 10.6.1)* 2019) and a map file for the turlough was saved. Polyline shapefiles were created and added to the map to represent vegetation boundaries. OS vector maps were loaded to the map and the OS vector map lines within the turlough boundary which were needed to create vegetation polygons were copied to the previously created land parcel polyline dataset. Vegetation boundary polylines were then drawn to link all boundary points. Separate polygons were then each attributed to a vegetation type, using the information recorded in vegetation identification points or via a deductive process using field data and information from aerial and ground-level photographs.

Remote sensing

A more recent assessment of the vegetation distribution across the turloughs was made using remote sensing methods from Sentinel-2 (S2) imagery captured on 30th June 2018. All the S2 bands were resampled to 10 m, and additional normalised difference vegetation index (NDVI), enhanced vegetation index (EVI), and normalised difference water index (NDWI) were calculated. The initial pixel-based classification was done using an ensemble Bagged Tree classifier. The pixel-based classification only takes into account the spectral information and not the spatial relation that exists between the vegetation communities. Therefore, using the mapping vegetation community (MVC) algorithm described in

Bhatnagar et al. 2020a, the classification was extended to area-based segmentation. Final classification accuracy of approximately 85% was achieved (Bhatnagar et al. 2020a).

7.2.3 Hydrology

Field data (2007 to 2019)

A series of linked research projects have been carried out on the lowland karst network, to investigate the ecohydrology of the wetlands (Naughton et al. 2012; Gill et al. 2018), flood alleviation (Naughton et al. 2017; Morrissey et al. 2020), nutrient and other chemical fluxes through the turloughs (McCormack et al. 2016; Gill et al. 2018) and as well as freshwater discharges into Kinvara bay (McCormack et al. 2014). For this, the following hydrometeorological data have been collected with more detail given in the above references.

- Meteorology
  Rainfall data were collected from two tipping bucket rain gauges positioned at 70 mAOD and 150 m AOD in the catchment to assess the spatial distribution of rainfall. These data were then related to the Gort Derrybrien gauge operated by Met Eireann, which provided a longer dataset, in order to fill missing gaps.

- Water level
  Continuous water level data were collected using pressure transducers with in-built dataloggers at the base of four turloughs (in the lowland karst catchment between 2007 to 2018). These were recovered in the dry period each summer and downloaded and the data compensated against the variations in atmospheric pressure over the flooded period.

- Topography (Lidar etc.)
  Topographical data were derived from LiDAR mapping data for the catchment with a grid spacing of 2 m and a vertical accuracy of ± 0.15 m. Where such data were not available, further topographical survey data was obtained from manual surveys carried out using a Trimble 4700 GPS system with a minimum accuracy of 0.01 m (horizontal and vertical direction). These topographical data were combined with the available LiDAR data in ArcGIS, and a new integrated DEM was constructed using the Kriging method with a 2 m grid spacing. Depth-area-volume relationships for the turloughs and linked floodplains were derived from these DEMs, as required for the hydraulic model, (see Section 7.3.2).

Hydraulic model

A detailed semi-distributed hydraulic model of the karst network has been developed over many years for various different applications (ecohydrology, flood alleviation etc.) and was used to fill in any gaps in the turlough water level data but also to extend the turlough water time series back to 1989 when local rain monitoring started in the region. The model is built using the Infoworks CS drainage software due to its ability to model the hydraulics of the karst conduit network in both open channel and pressurised pipe flow. The model is described in detail in Gill, Naughton & Johnston 2013, Gill, Naughton, Johnston, Basu & Ghosh 2013 with the most recent update in Morrissey et al. 2020. The groundwater-surface water turlough dynamics were modelled as storage ponds in the software which were configured with the same depth-volume characteristics as the surface

topography, as derived from the detailed DEM, thereby giving an accurate profile of the exact depth across each turlough at any point in time on an hourly basis. Examples of the correlation of the model for the four turloughs against the water level field data for the four turloughs are shown in Supplemental Information, Figure 9.47 (as well as in the above references), all showing excellent model performance with Nash Sutcliffe and Kling Gupta efficiencies (NSE and KGEs) of > 0.9. This model has been used to extend the time series of the hydrological monitoring back to 1988 to yield 28 years of continuous water fluctuation data in the four turloughs.

7.2.4 Ecohydrology metrics

The key hydrological variables evaluated are depth, duration, frequency and timing of flooding. The model was used to define the hydrological conditions at all times across a 28-year period on the turloughs. Whilst the model produced data at an hourly time step, this was changed to an average daily time step (due to the slow movement of the water levels) in order to make the data processing quicker. From these water levels time series, the following metrics were derived for any point across each turlough over a hydrological year (which is taken from the beginning of October to the end of September in Ireland).

- Flood duration - the number of days in a hydrological year that any pixel area was covered in water, averaged out spatially across each community vegetation type.

- Flood depth - the average depth of water per day that any pixel area was covered in water. This was then averaged out spatially across the same vegetation type and expressed as an average value over a hydrological year.

- Flood frequency - the number of flood events experienced per vegetation community per hydrological year. A flood event was considered for each pixel if the area went from being dry to flooded and then remained flooded for at least 20 days. All the values for the pixels belonging to the same community were then averaged to form the flood frequency metric.

- Flood timing (in relation to the start of the growing season) – instead of using a date, or Julian day of the year, proxy variables of global radiation and air temperature at the time when the flooding stops and the vegetation were revealed for the first time in the hydrological year, were determined as more causative key vegetation variables. These were derived as the average values over a 30-day period (10 days before and 20 days after the pixels were first revealed).

The vegetation communities from the turlough maps (Section 7.2.2) were defined at a spatial resolution depending on the turlough size: Caherglassaun being the biggest was surveyed at 4×4 m, Blackrock at 3.5×3.5 m, Coy and Garryland being smaller at 2.5×2.5 m. For each of the defined communities, a binary image (2-D) was extracted from the vegetation survey map such that the spatial relationship between the points belonging to the same community is not lost. This spatial relationship is vital to understand the amount and duration of flooding that occurs in the region. For example, a community like Lolium grassland can stretch across wide areas of the turlough, with some part of the community in flood, and some part which is seldom flooded. The averaged value for the whole community was then measured using 'mean2' function, i.e., 2-D mean in MATLAB v.2019b (MathWorks & Inc 2019). This was repeated for all the communities across all the turloughs. Then, the metrics were combined and compared for common communities. The field vegetation surveyed map created in 2008 is used as the reference throughout the

study. The satellite-based study carried out subsequently by Bhatnagar et al. 2020a in 2018 was then used to assess whether there has been any significant change or shift in the turlough communities over the past ten years (2008 to 2018).

7.3 Results

7.3.1 Vegetation distribution

The vegetation maps derived in 2008 for the four turloughs are shown in Figure 7.1, showing the spatial distribution of the 15 different vegetation communities. The vegetation maps determined using remote sensing S2 imagery from June 2018 are shown in Figure 7.2, revealing a close match to the field survey maps.
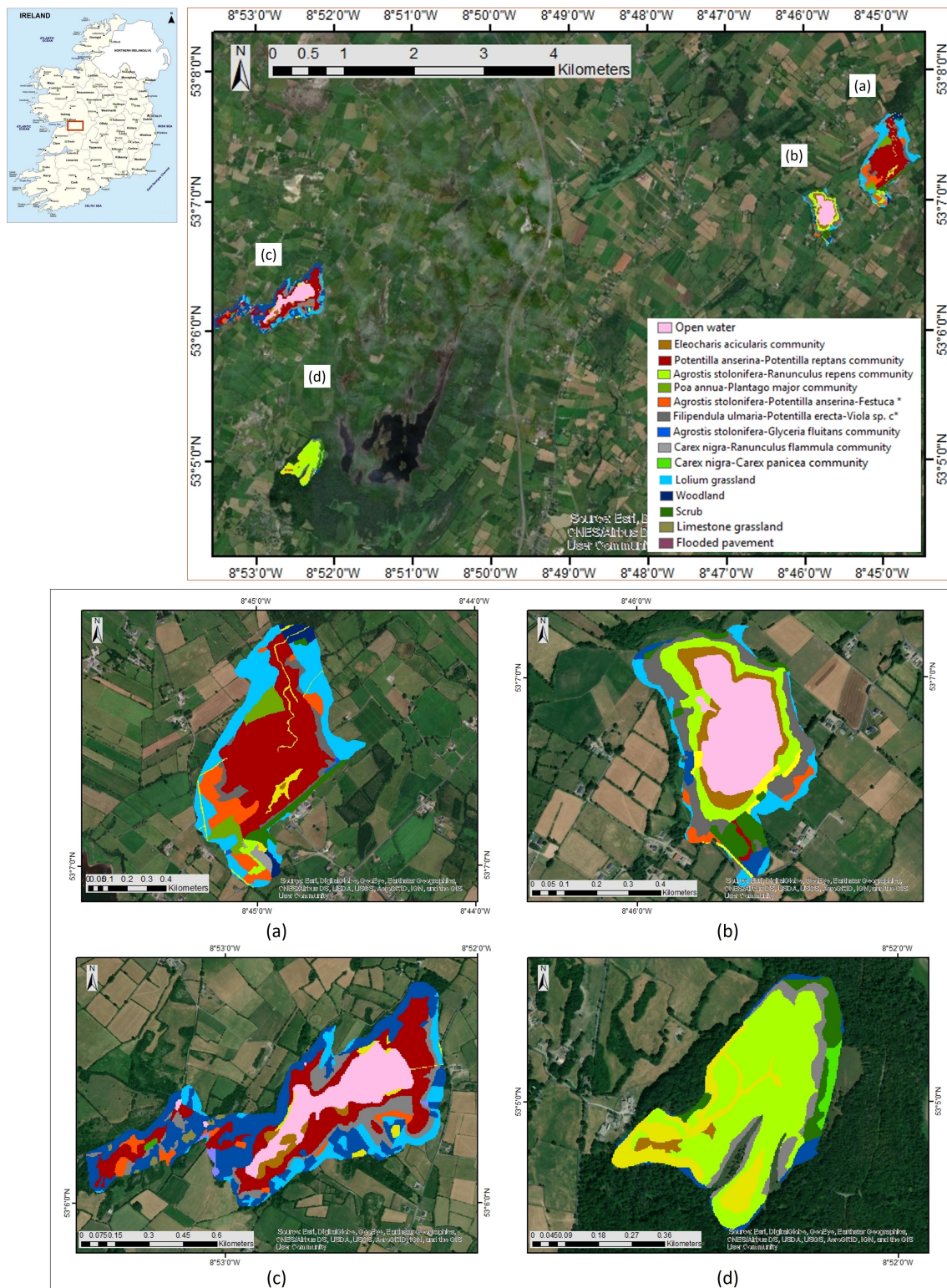
Figure 7.1: Vegetation communities for the four turloughs mapped in 2007 to 2009 (a) Blackrock, (b) Coy, (c) Caherglassaun and (d) Garryland.
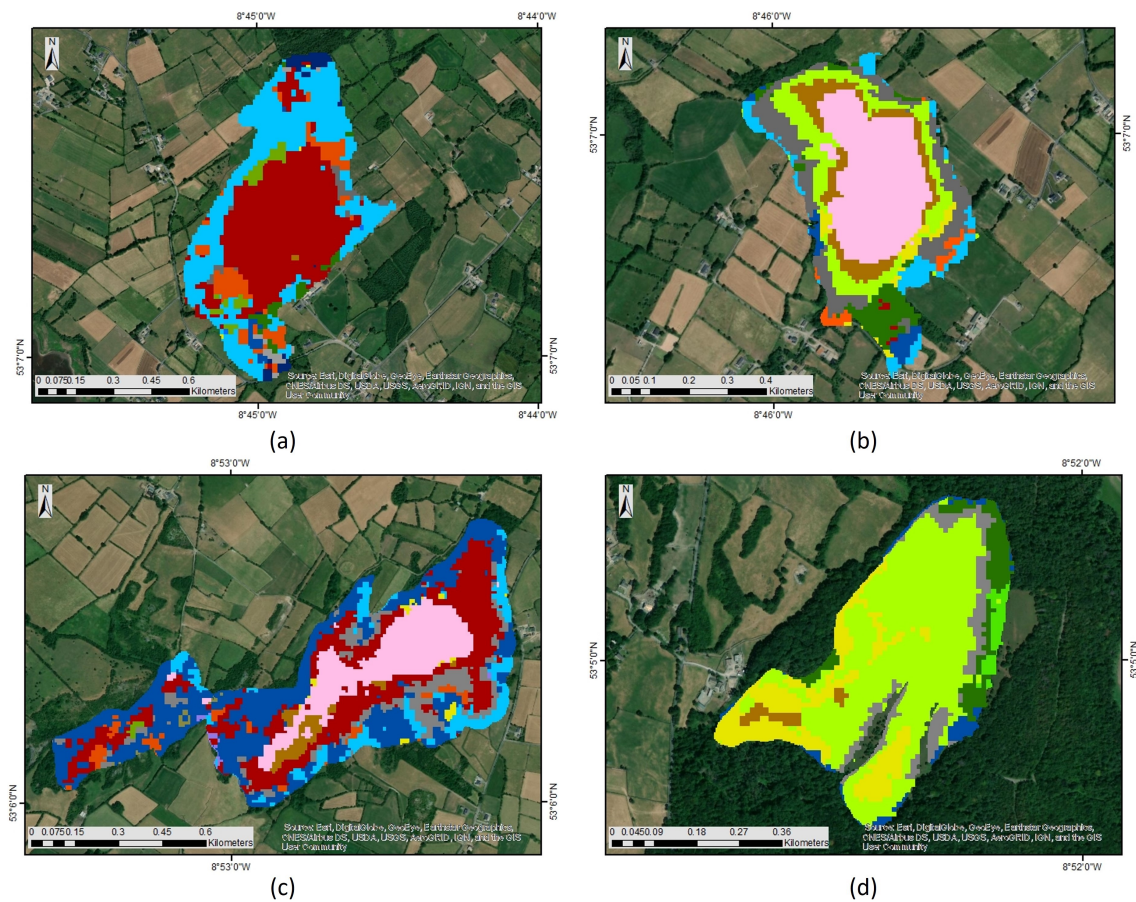
Figure 7.2: Vegetation communities for the four turloughs mapped using remote sensing from Sentinel-2 in 2018: (a) Blackrock, (b) Coy, (c) Caherglassaun and (d) Garryland.

The topography of four turloughs is shown in Figure 7.3, showing the steeper sides to Blackrock turlough, compared to the much flatter topography of Garryland turlough. Visually some correlation can be seen between the topography and vegetation communities and Figure 7.1, indicating the influence of flooding on vegetation location in these basins.
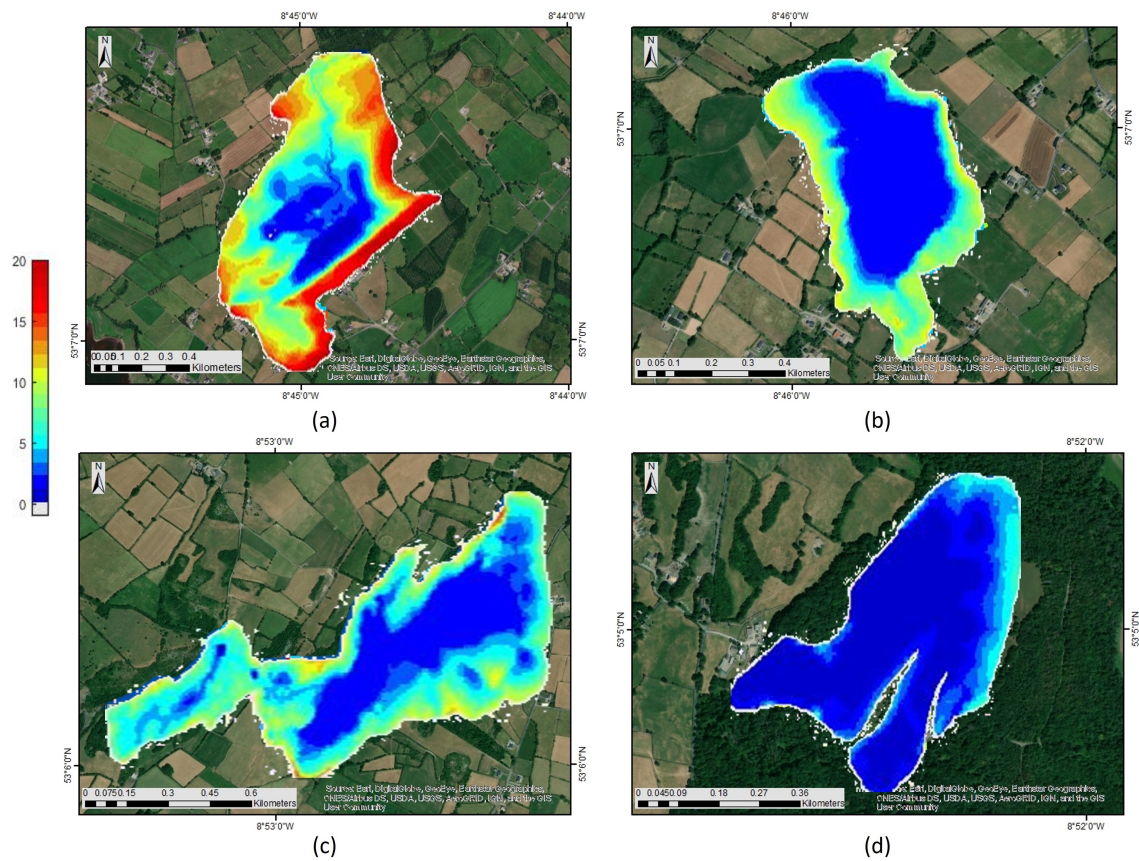
Figure 7.3: Turlough topographies in relation to their lowest point (a) Blackrock (b) Coy, (c) Caherglassaun and (d) Garryland (see Supplemental Information Figure 9.48 for actual topographies referenced to m AOD).

7.3.2 Hydrological characteristics

Figure 7.4 shows a sample 3-year period to show the nature of flood dynamics of the four turloughs over time. There was a particularly high level of flooding in November 2009, which caused much disruption in the area, with many properties, roads, railway routes inundated as well as extensive areas of farmland for several months. In general, the turloughs have one main flood once across the winter with small fluctuations. However, the 2008/09 hydrological year showed three moderate inundations that winter as well as some minor flood events in the summer 2009 period.
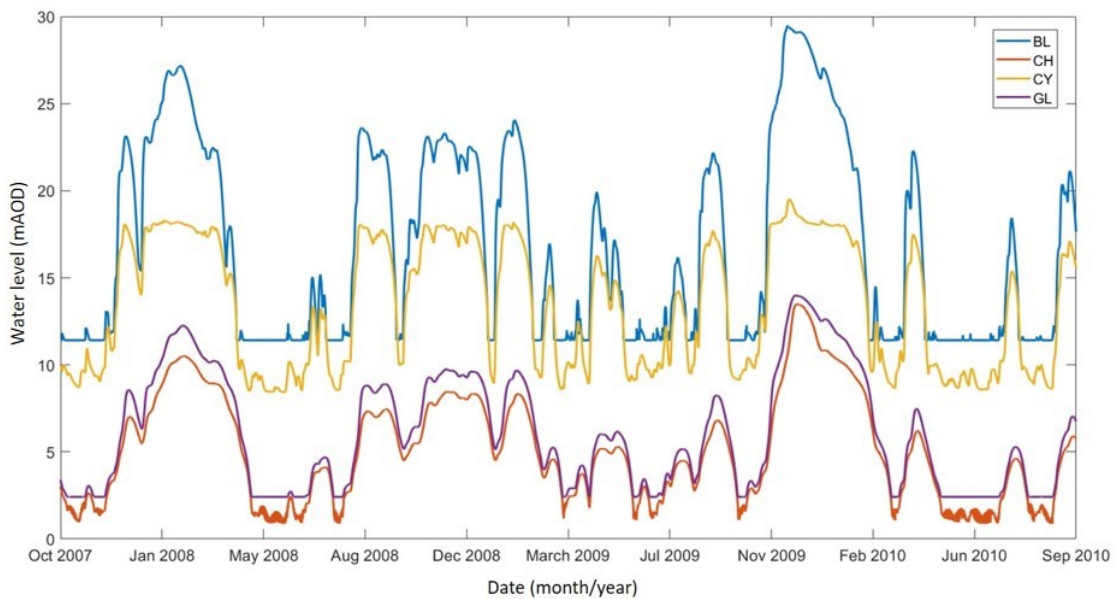
Figure 7.4: Time series of simulated water levels 2007 to 2010 for four turloughs

The average depth-duration plots for the four turloughs over the last 28 years is shown in Figure 7.5. This shows the much more flashy flooding dynamics of Blackrock turlough, which is the steepest in topography as well as being located at the start of the hydraulic network and therefore receiving a less damped hydraulic signal from the allogenic river inputs compared to the turloughs lower down; in comparison Garryland and Caherglassaun at the lower end of the system which show very similar curves. A comparison between the depth-duration profiles for 1997 to 2007 versus 2008 to 2018 is also shown on Figure 7.5(b) which is discussed later in Section 7.3.7 to assess whether any changes in vegetation over the period between the field survey and satellite survey can be linked to changes in hydrological regime between those two periods. The annual flood duration and flood depth spatial profile across the 28 years is shown for Blackrock turlough in Figures 7.6(a) and 7.6(b) (and in the Supplemental Info Figs 9.49 to 9.54 for the other three turloughs) revealing the difference between hydrological years. In particular years 1989/1990, 1991, 1994, 1995, 2009 and 2015/2016 can be seen to have much longer flood durations than many of the other years (and these were years when there was widely reported significant flood disruption in the winters in this area) compared to much drier years of 1997, 2006 etc.

7.3.3   Ecohydrology - flood duration

The flood duration statistics across the 28 years averaged for each of the key communities which were present in all or at least 3 out of 4 turloughs are shown in Figure 7.7. This shows wide fluctuation between different years but does equally show differences between communities. The variations in durations amongst and between the communities are shown more clearly in the boxplots in Figure 7.10(a) also.

160
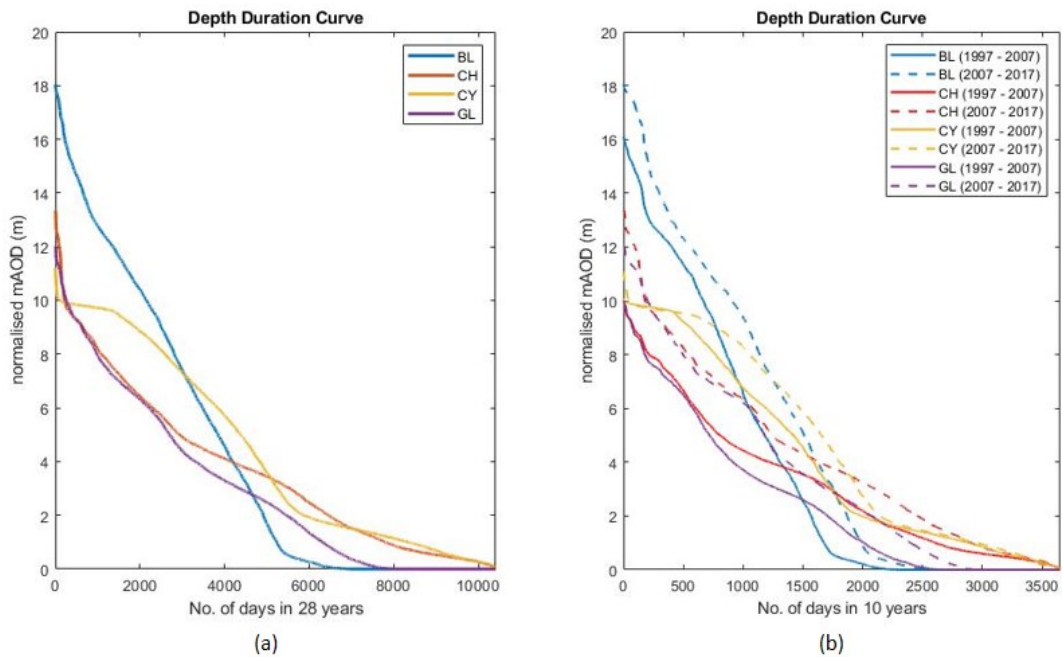
Figure 7.5: Flood depth-duration plots for four turloughs (a) across full 28-year dataset (b) comparing 10-year datasets (1997-2007 and 2008-2017).

7.3.4 Ecohydrology – flood depth

The flood depth statistics across the 28 years averaged for each of the key communities which were present in all or at least 3 out of 4 turloughs are shown in Figure 7.8. This shows wide fluctuation between different years but does equally show clear differences between communities. The variations in flood depths amongst and between the communities are shown in the boxplots in Figure 7.10(b) also.

7.3.5 Ecohydrology – flood timing

As described in Section 7.2.4, the timing of when the flood receded (and potential start of the growing season for each vegetation community) was determined by looking at the mean global radiation and mean temperature taking an average of the 30 days (10 days before it was first revealed and 20 days after). The global radiation statistics across the 28 years averaged for each of the key communities when coming out of flood, generally in the springtime, are shown in Figure 7.9 (with Figure 9.59 showing the equivalent air temperature relationships). Figures 9.55 to 9.58 in the Supplemental Information showing the air temperature and global radiation spatial plots for all four turloughs. This shows the higher elevation communities are exposed earlier in the year when the mean solar radiation (and average air temperature) is lower, as expected. The variations in average global radiation when first exposed from the flood waters amongst and between the communities are shown in the boxplots in Figure 7.10(c) also. The global radiation and the temperature have a direct relationship showing a correlation of 99% from regression analysis (see Section 7.3.6) as expected. The relationship between mean air temperature, global radiation, flooding depth, and duration is shown in Table 7.3.
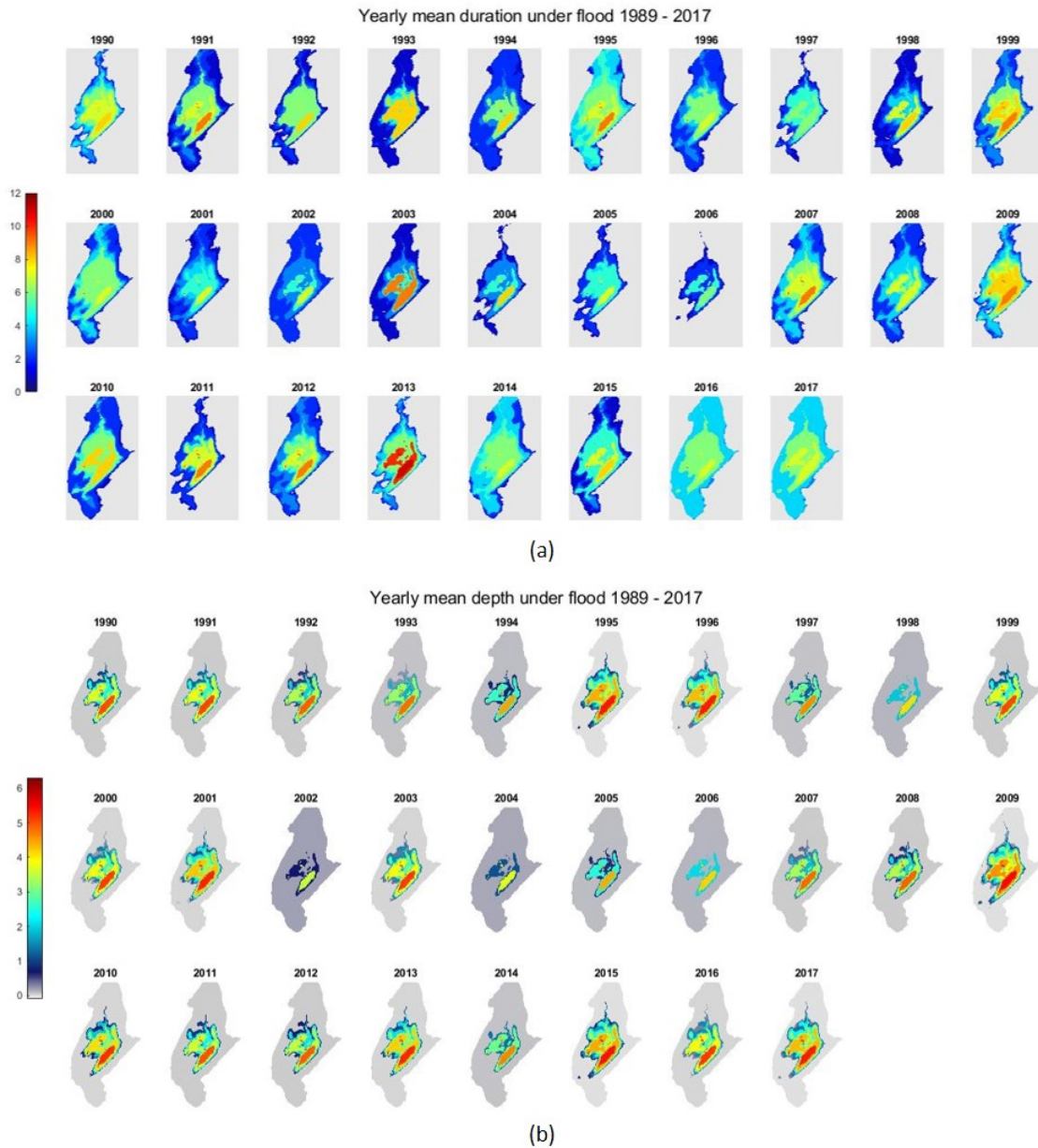
Figure 7.6: Annual (a) flood duration and (b) average flood depth spatial profiles for Blackrock turlough (see Figs. 9.49 to 9.51 for Coy, Garryland and Caherglassaun). Note, each annual figure represents a hydrological year from Oct to Sept with year written above referring to the end 9 months of each respective year.
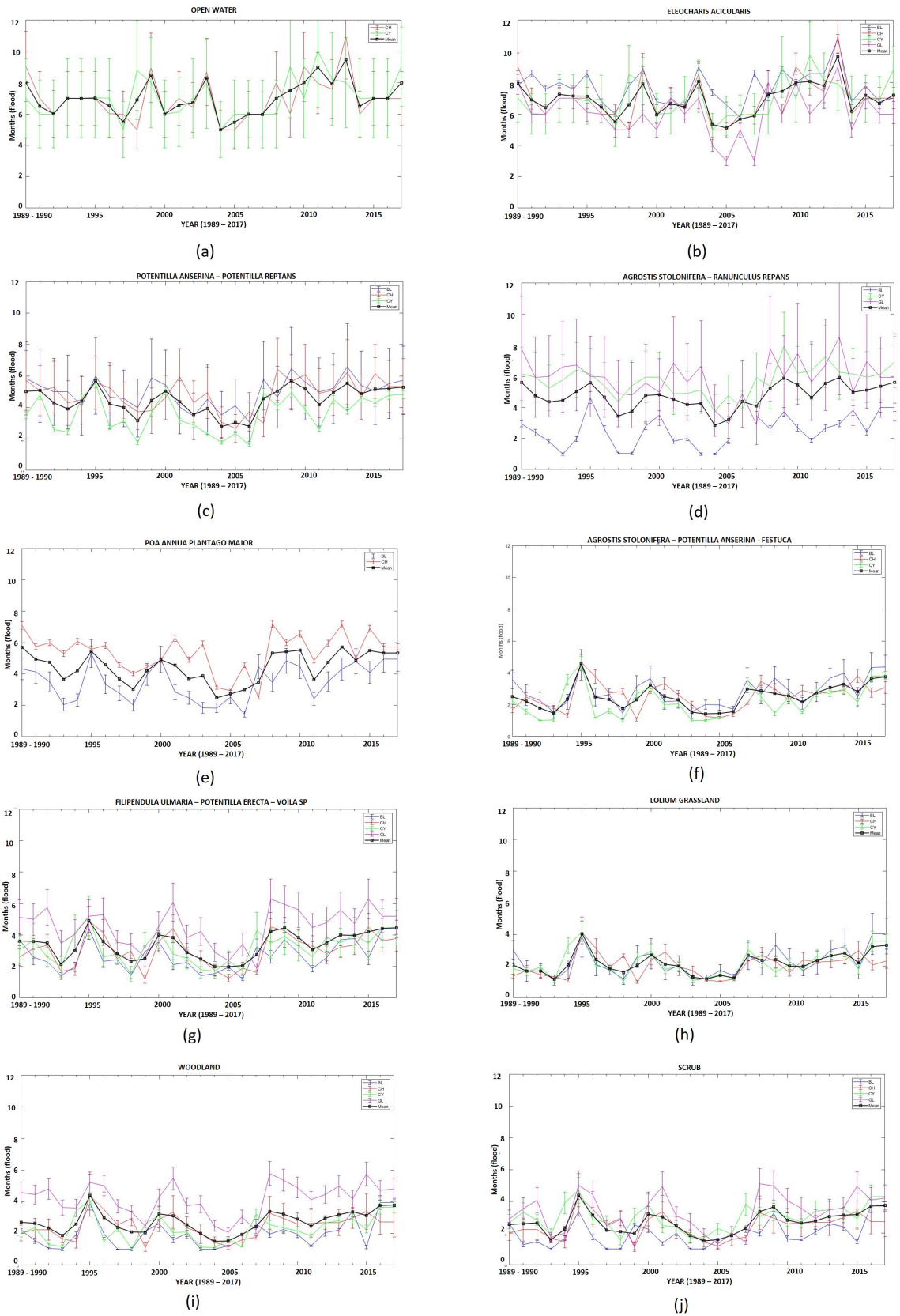
Figure 7.7: Mean flood duration across 28 years for key vegetation communities.
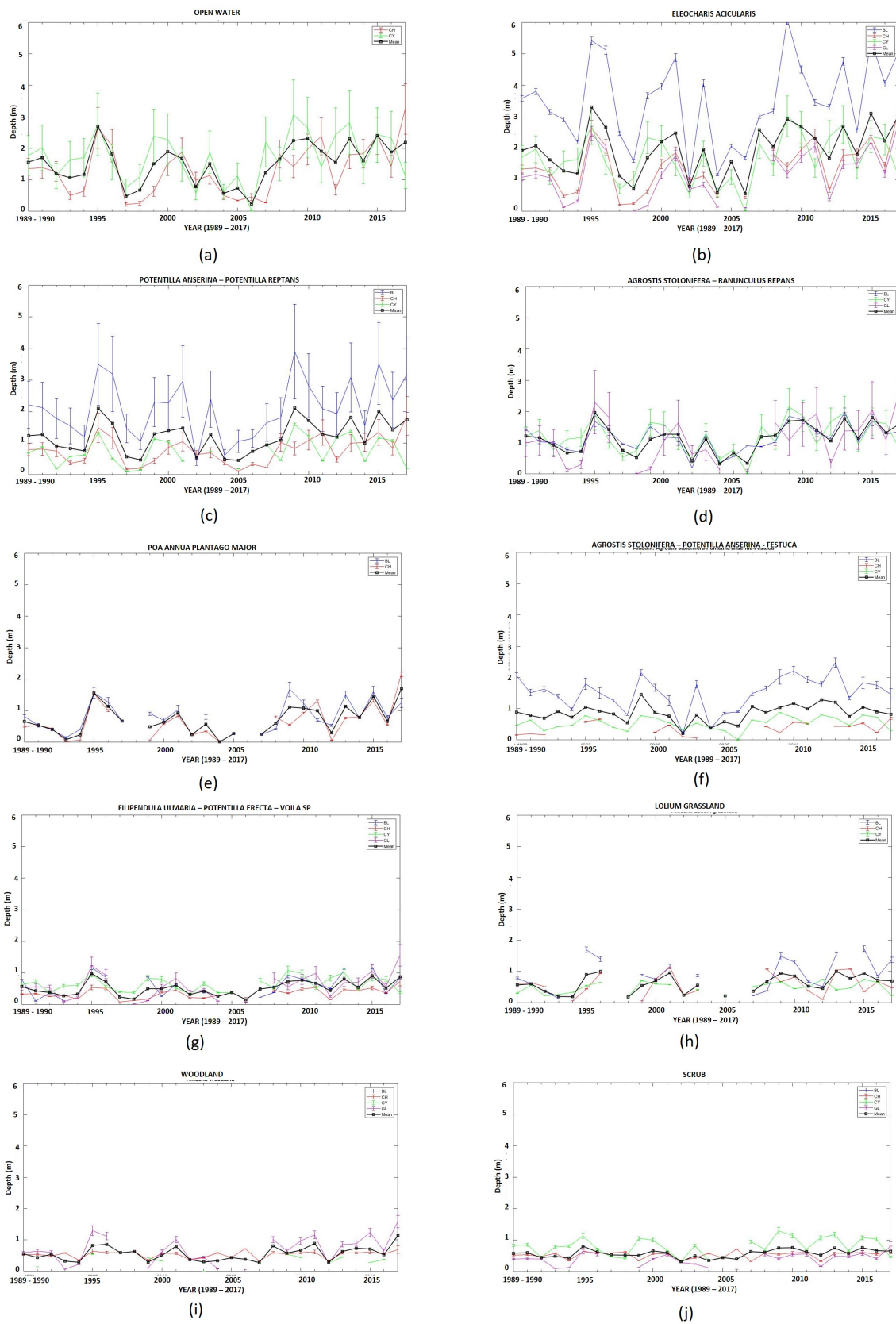
Figure 7.8: Mean flood depth across 28 years for key vegetation communities.
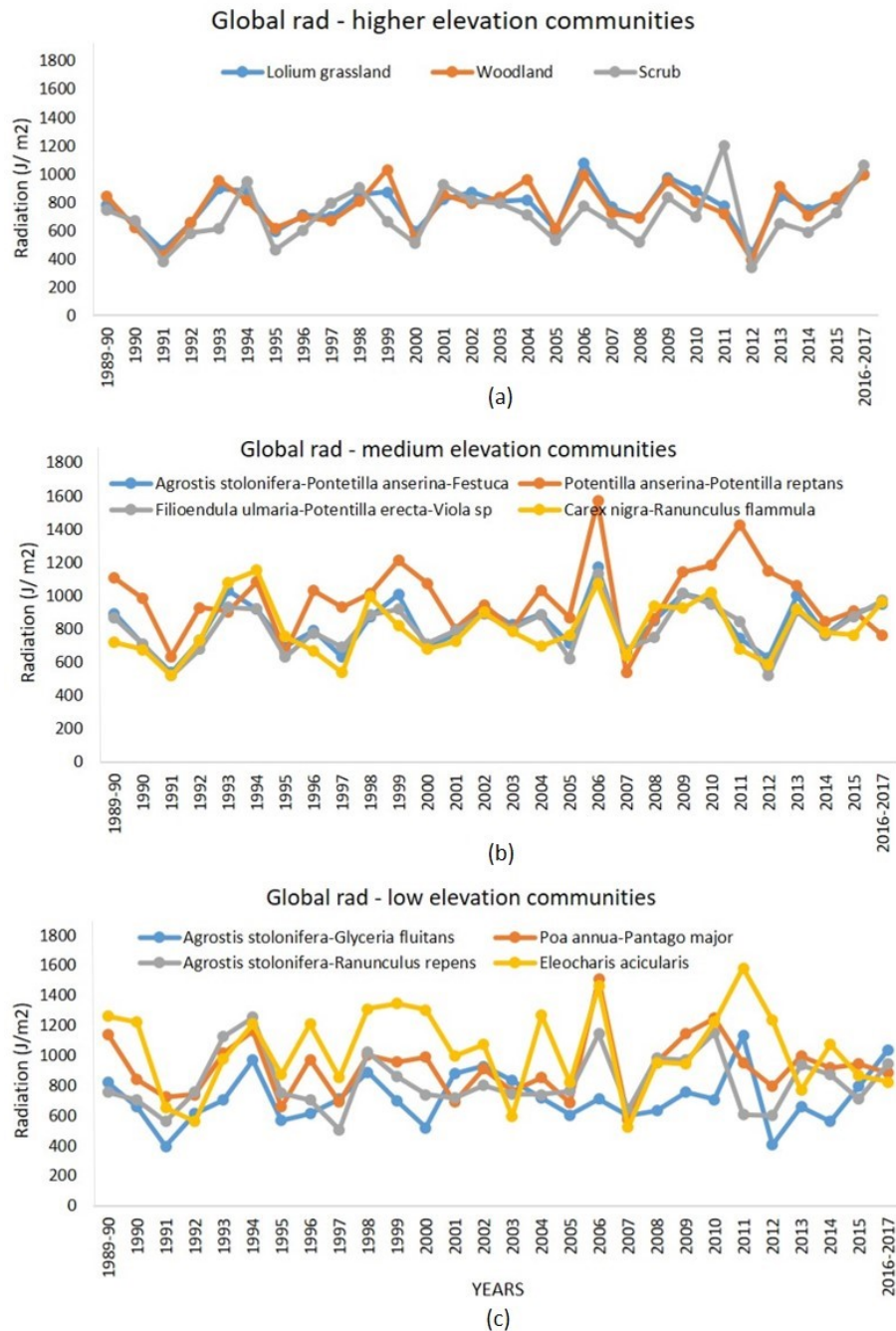
Figure 7.9: Mean global radiation for the week/month after which the different communities came out of flood for the year in Blackrock turlough.

7.3.6   Flood Frequency

Table 7.1 shows the average flood frequency experienced between the different vegetation communities across all turloughs. i.e. how many times a vegetation community went from being dry to flooded and stayed flooded for at least 20 days. The variations in flood frequency per year amongst and between the communities are shown in the boxplots in Figure 7.10(d) also. As can been seen most communities only experience a maximum of two flood inundations per year.

7.3.7   Ecohydrological metrics

A summary of the flood duration, depth, global radiation when coming out of flood and flood frequency metrics across all four turloughs for the different vegetation communities is presented in Table 7.1 (see Table 9.1 for a full breakdown of these statistics per turlough), and in Figure 7.10. The communities have been ordered according to the flood duration on the graphs and tables ranging from *Eleocharis acicularis* (experiencing the most flooded conditions in a year) to the flooded pavement community (experiencing the least amount of flooding).

	DEPTH (meters)				DURATION (months)				FREQUENCY (per year)				Global Radiation (J/m2)			
	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE
Open water	0.23	1.31	2.41	[0.65 − 1.24]	5	6.62	7.99	[6.20 - 7.91]	1.00	1.55	3.5	[1.00 - 1.95]	398	792	1519	[629 - 906]
Eleocharis acicularis	0.56	1.7	3.1	[0.49 − 1.26]	5.33	6.63	7.25	[6.26 - 7.38]	1.00	1.39	2.32	[1.25 - 1.70]	398	848	1578	[790 - 885]
Agrostis stolonifera-Ranunculus repens	0.06	0.95	1.86	[0.40 − 0.95]	2.84	4.58	5.61	[4.36 - 5.31]	1.00	1.4	2.81	[1.10 - 1.60]	391	795	1399	[700 - 890]
Poa annua-Pantago major	0.01	0.65	1.7	[0.26 − 0.80]	2.48	4.25	5.48	[3.81 - 5.27]	1.00	1.18	1.93	[1.00 - 1.65]	424	795	1337	[720 - 900]
Potentilla anserina-Potentilla reptans	0.5	1.07	2.01	[0.78 − 1.70]	2.79	4.14	5.29	[3.90 - 5.00]	1.04	1.4	3.19	[1.18 - 1.60]	383	775	1277	[716 - 874]
Filipendula ulmaria-Potentilla erecta-Viola sp	0.16	0.48	0.91	[0.42 - 0.71]	1.98	3.2	4.45	[2.80 - 4.05]	1.00	1.21	2.66	[1.15 - 1.50]	376	647	1148	[603 - 703]
Carex nigra-Carex panicea	0.04	0.2	0.42	[0.04 - 0.42]	1.13	2.85	4.44	[1.76 - 3.85]	1.00	1.14	1.97	[1.05 - 1.45]	403	539	654	[559 - 602]
Woodland	0.28	0.57	1.14	[0.47 - 0.77]	1.5	2.61	3.79	[2.30 - 3.20]	1.00	1.2	2.16	[1.09 - 1.45]	350	607	1110	[586 - 622]
Scrub	0.32	0.51	0.76	[0.39 - 0.62]	1.51	2.57	3.74	[2.25 - 3.18]	1.00	1.14	1.9	[1.09 - 1.42]	374	618	1071	[592 - 635]
Agrostis stolonifera-Pontetilla anserina-Festuca	0.21	0.71	1.24	[0.65 - 1.24]	1.42	2.27	3.74	[1.83 - 3.00]	1.00	1.29	2.05	[1.10 - 1.41]	385	681	1162	[610 - 700]
Limestone Grassland	0.34	0.52	0.73	[0.40 - 0.73]	1.59	2.28	3.05	[1.75 - 2.91]	1.00	1.06	1.6	[1.00 - 1.25]	403	539	654	[556 - 650]
Carex nigra-Ranunculus flammula	0.16	0.78	1.6	[0.26 - 1.40]	1	2.07	3.99	[1.70 - 3.25]	1.00	1.23	2.15	[1.00 - 1.20]	398	657	876	[600 - 766]
Lolium grassland	0.2	0.54	1	[0.31 - 0.59]	1.18	1.97	3.31	[1.95 - 2.59]	1.00	1.2	1.98	[1.02 - 1.24]	375	649	1122	[608 - 732]
Agrostis stolonifera-Glyceria fluitans	0.13	0.76	1.46	[0.55 - 1.20]	1	1.91	3.78	[1.44 - 2.80]	1.00	1.22	2.89	[1.00 - 1.35]	286	590	857	[560 - 754]
Flooded Pavement	0.43	0.43	0.43	[0.43 - 0.43]	1	1.78	3.14	[1.31 - 2.30]	1.00	1.78	3.14	[1.00 - 1.01]	403	538	634	[512 - 604]

Table 7.1: Metric describing averaged duration and depth for last 10 years using S2 maps for all communities over 4 turloughs.

Since the range of the communities stays more or less consistent throughout the years (see Section 7.3.7), regression analysis has been used to determine the relationship between the four key variables depth, duration, temperature and global radiation with respect to the vegetation type (see Figure 9.60). A variable regression technique was used, such that the best fit between the variables can be found. From Table 9.2 it can be seen that most variables are related linearly. However, variables such as duration-global radiation, and duration-temperature exhibit exponential relationships. The accuracy of the regression analysis is given by the R^2 value. The analysis shows that all four parameters are highly correlated, with the highest correlation between depth-duration, and global radiation–temperature. Therefore, all four parameters can be used to form metrics to frame the required ecohydrological condition of the turloughs.

A hierarchical spatial clustering analysis was then used on the 28-year dataset on the four turloughs on the key ecohydrological variables discussed heretofore (i.e. flooding depth, duration, temperature and global radiation) in order to identify clusters of similar hydrological years. The aim was to refine the ecohydrological metrics to what the different vegetation communities experienced in what might be considered to be more "normal" years. The hierarchical clustering enables the clusters to be formed based on relative distance. For this study, the hierarchical clustering was done using the formation of 'Dendrogram' using Matlab v.2019a (MathWorks & Inc 2019). The y-axis of the dendrogram represents the distance (in this case, Euclidean distance) or dissimilarity between the variables and the x-axis displays all the years. The variables here are the depth, duration, global radiation and frequency for all 28 years. The clusters are formed using agglomerative hierarchical cluster tree (Day & Edelsbrunner 1984). This was carried out using centroid linkage in Matlab v.2019a (MathWorks & Inc 2019), which is the distance (Euclidean) between the clusters (two or more). It is represented by Equation 7.3.1.

$$Dissimilarity = ||\overline{cluster_1} - \overline{cluster_2}||_2 \qquad (7.3.1)$$

where, Dissimilarity is the Euclidean distance between the clusters, and cluster represents the mean of the cluster.

In order to extract proper clusters from the linkage tree diagram, a threshold-cut-off line was drawn equal to the median of the dissimilarity (see red line on Figure 7.11). A total of 8 clusters were formed consisting of all 28 years: cluster 1(4,13, 15,5,7, 9, 17, 3, 19, 26, 28) ; cluster 2 (1, 25, 6, 14, 10, 21, 20); cluster 3 (11, 12, 27); cluster 4 (2, 16, 23); cluster 5 (24); cluster 6 (6); cluster 7 (18); cluster 8 (22). Cluster 1, with the highest number of years associated, appeared to pick up the years with least amount of extreme fluctuations and so was used to refine the ecohydrological metrics (flood depth, duration etc.) for all the vegetation communities across all four turloughs again (Table 7.2). In general, the refined statistics reveal slightly tighter range envelopes (for flood duration, temperature and global radiation) as might be expected), although the flood depth ranges for the different vegetation have extended somewhat. possibly indicating a more unimodal consistent flood peaks in the winter for these hydrological years.
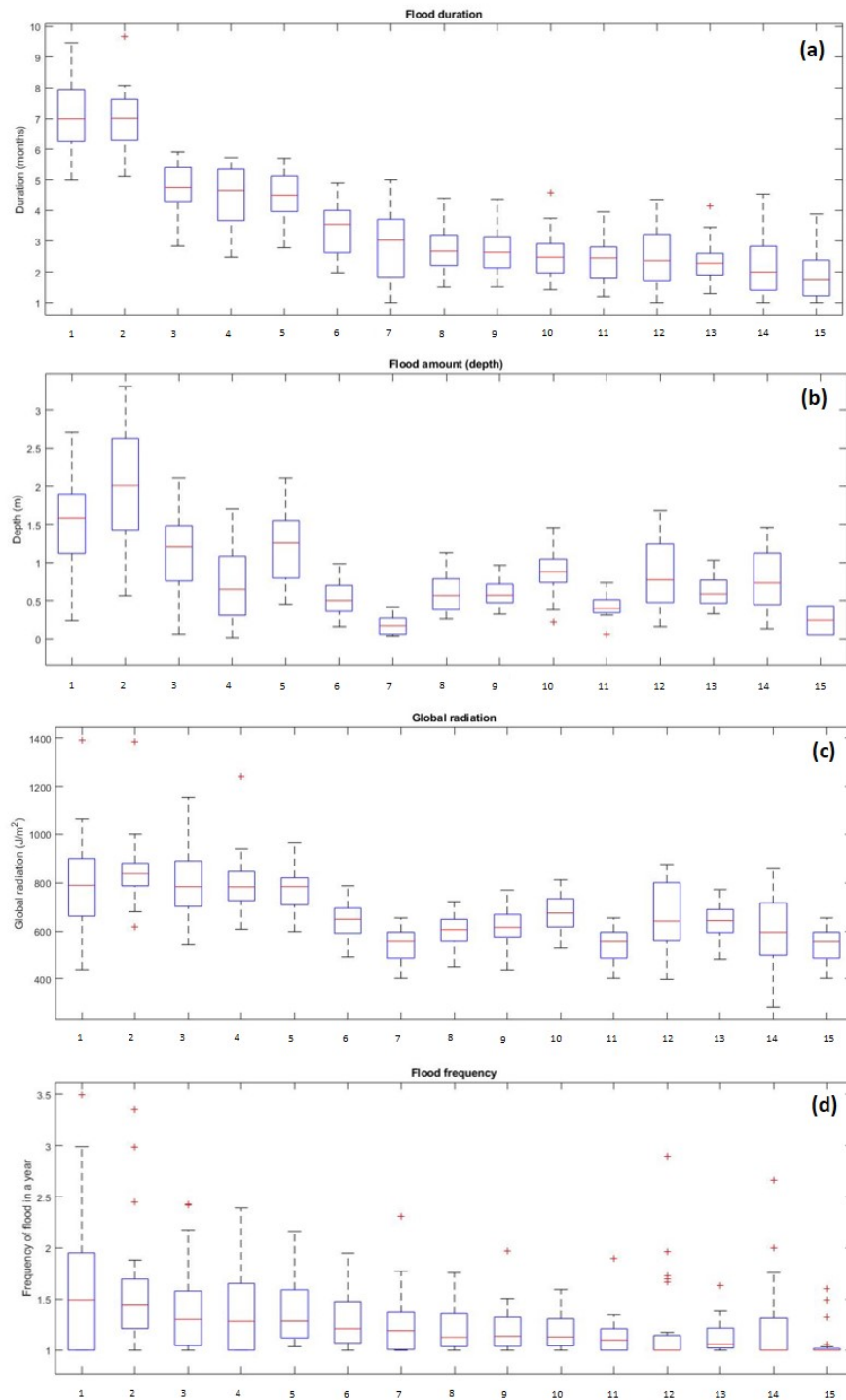
Figure 7.10: The statistics of four different ecohydrological metrics for the range of turlough vegetation communities averaged over the four turloughs (a) flood duration (b) flood depth (c) global radiation when the flood waters first recede and (d) flood frequency per year. x axis = vegetation communities: 1 = open water, 2 = *Eleocharis Acicularis*, 3 = *Agrostis stolonifera - Ranunculus repens*, 4 = *Poa annua - Plantago major*, 5 = *Potentilla anserina - Potentilla reptans*, 6 = *Filipendula ulmaria - Potentilla erecta - Viola sp.*, 7 = *Carex nigra - Carex panicea*, 8 = Woodland, 9 = Scrub, 10 = *Agrostis stolonifera - Potentilla anserina - Festuca*, 11 = Limestone Grassland, 12 = *Carex nigra - Ranunculus flammula*, 13 = Lolium Grassland, 14 = *Agrostis stolonifera - Glyceria fluitans*, 15 = Flooded Pavement

169

	DEPTH (meters)				DURATION (months)				FREQUENCY (per year)				Global Radiation (J/m2)			
	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE
Open water	0.23	1.97	3.31	[0.77 − 2.50]	5.11	6.92	9.67	[6.40 - 7.19]	1.00	1.57	3.49	[1.00 - 1.87]	440	792	1391	[600 - 811]
Eleocharis acicularis	0.56	1.52	2.70	[0.75 − 1.85]	5.00	7.01	9.46	[6.12 − 6.99]	1.00	1.54	3.35	[1.02 - 1.70]	618	848	1385	[723 - 807]
Agrostis stolonifera-Ranunculus repens	0.06	1.24	2.19	[0.55 − 1.62]	2.84	4.57	5.83	[4.10 − 4.87]	1.00	1.40	2.42	[1.00 - 1.58]	542	795	1152	[661 - 785]
Poa annua-Pantago major	0.01	0.71	1.70	[0.30 − 1.22]	2.48	4.44	5.72	[3.18 - 5.18]	1.00	1.40	2.39	[1.08 - 1.33]	608	795	1241	[680 - 732]
Potentilla anserina-Potentilla reptans	0.45	1.21	2.11	[0.71 − 1.50]	2.79	4.48	5.71	[3.25 − 4.88]	1.04	1.39	2.16	[1.02 - 1.52]	598	775	966	[669 - 739]
Filipendula ulmaria-Potentilla erecta-Viola sp	0.16	0.52	0.98	[0.44 - 0.80]	1.98	3.36	4.90	[2.18 - 4.04]	1.00	1.29	1.95	[1.00 - 1.21]	492	647	787	[570 - 630]
Carex nigra-Carex panicea	0.04	0.18	0.42	[0.22 - 0.47]	1.00	2.82	5.00	[1.55 − 4.03]	1.00	1.27	2.31	[1.00 - 1.21]	403	539	654	[490 - 556]
Woodland	0.26	0.59	1.13	[0.47 - 0.81]	1.50	2.74	4.40	[1.97 - 3.10]	1.00	1.20	1.76	[1.00 - 1.10]	452	606	722	[550 - 601]
Scrub	0.32	0.60	0.96	[0.49 - 0.65]	1.51	2.68	4.37	[1.91 - 3.15]	1.00	1.20	1.97	[1.01 - 1.12]	439	618	770	[556 - 619]
Agrostis stolonifera-Pontetilla anserina-Festuca	0.21	0.86	1.46	[0.54 − 0.85]	1.42	2.50	4.58	[1.60 − 2.72]	1.00	1.18	1.59	[1.04 - 1.10]	529	681	813	[609 - 662]
Limestone Grassland	0.06	0.42	0.73	[0.50 - 0.63]	1.20	2.37	3.95	[1.76 - 2.80]	1.00	1.14	1.90	[1.00 - 1.14]	403	539	654	[490 - 556]
Carex nigra-Ranunculus flammula	0.16	0.84	1.68	[0.48 - 1.36]	1.00	2.40	4.36	[1.03 − 2.36]	1.00	1.22	2.89	[1.00 - 1.40]	398	657	876	[539 - 658]
Lolium grassland	0.18	0.61	1.00	[0.20 - 0.75]	1.18	2.17	4.05	[1.34 - 2.33]	1.00	1.14	1.63	[1.03 - 1.10]	482	649	772	[583 - 637]
Agrostis stolonifera-Glyceria fluitans	0.13	0.78	1.46	[0.36 - 1.13]	1.00	2.20	4.54	[1.00 - 2.00]	1.00	1.21	2.66	[1.00 - 1.50]	286	591	858	[465 - 694]
Flooded Pavement	0.05	0.24	0.43	[0.43 - 0.43]	1.00	1.88	3.88	[1.15 - 2.46]	1.00	1.06	1.60	[1.00 - 1.20]	403	538	654	[490 - 556]

Table 7.2: Metrics describing the refined ecohydrological metrics for more 'normal' hydrological years – averaged over each community across 4 turloughs.
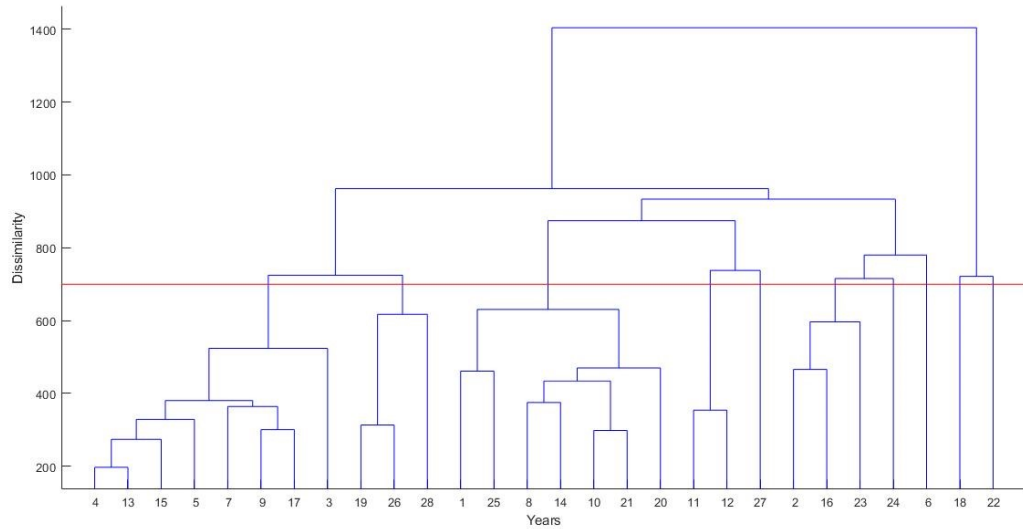
Figure 7.11: Dendrogram using four parameters (depth, duration, global radiation and frequency) to form clusters of 28 years.

7.3.8 Change in communities over 10-year period

As detailed in Section 7.2.4, the field surveys were carried out in 2008 and then 10 years later another map of the turloughs was produced using the Sentinel-2 satellite approach (Figure 7.2) using mapping vegetation communities (MVC) algorithm presented in Bhatnagar et al. 2020a. A comparison between these two maps (and the spatial coverage of the different vegetation therein) has been made to assess if there have been any shifts in vegetation community spatial distribution which could then be linked to possible changes in hydrological regimes over that 10-year period. The depth-duration plots comparing the 10 year period up to the 2008 vegetation survey (1998 to 2008) and from then up to the satellite survey in 2018 (2008 to 2018) are shown on Figure 7.5(b) which shows that there does appear to have been more severe / longer lasting flooding over the past 10 years on all turloughs except Coy which is a turlough that is known to have a higher level overflow in the kart system which thereby maintains fairly uniform peak flood depths, which are unresponsive to additional rainfall. The 2008 to 2018 period has been marked by more extreme rainfall and weather extremities, particularly 2009 and 2015/16 with two exceptional periods of flooding.

A visual comparison of Figures 7.1 and 7.2, set out in Table 7.4, shows that majority of the communities appear to stay intact (for example Lolium grassland, *Potentilla anserina – Potentilla reptans*, etc.); even the smaller spatial extent communities like *Eleocharis acicularis* have been identified well in all four turloughs. It has previously been assumed in Section 7.3.6 for example, in terms of deriving the ecohydrological metrics that the vegetation communities essentially stay intact, with the 2008 field derived vegetation map used as the reference throughout.

Table 9.2 shows all the different hydrological metric data plotted for the two different time periods whilst Table 7.3 shows the Jaccard similarity and percentage change in the area of all the communities across turloughs between the vegetation survey (2008) and Sentinel-2 survey (2018). It has to be noted that the S2 bands were resampled to 10 m;

	BL		CH		CY		GL	
	Jaccard	△ Area	Jaccard	△ Area	Jaccard	△ Area	Jaccard	△ Area
Open water	NaN	NaN	0.86	173	0.94	134	NaN	NaN
Eleocharis acicularis	0.04	-14	0.64	24	0.7	8	0.52	-21
Potentilla anserina-Potentilla reptans	0.74	-219	0.69	331	0.25	130	NaN	NaN
Agrostis stolonifera-Ranunculus repens	0.12	-121	NaN	NaN	0.74	-32	0.8	-250
Poa annua-Pantago major	0.23	-1440	0.55	7	NaN	NaN	NaN	NaN
Agrostis stolonifera-Pontetilla anserina-Festuca	0.32	-975	0.49	-223	0.51	-155	NaN	NaN
Filipendula ulmaria-Potentilla erecta-Viola sp	0.18	-1775	0.51	-473	0.7	-234	0.4	-21
Agrostis stolonifera-Glyceria fluitans	0.014	-100	NaN	NaN	NaN	NaN	NaN	NaN
Carex nigra-Ranunculus flammula	0.41	-66	NaN	NaN	NaN	NaN	NaN	NaN
Carex nigra-Carex panicea	NaN	NaN	NaN	NaN	NaN	NaN	0.49	92
Lolium grassland	0.39	+3730	0.58	-432	0.55	-572	NaN	NaN
Woodland	0.44	-94	0.44	1011	0.56	-341	0.28	-704
Scrub	0.25	-827			0.66	-145	0.54	-129
Limestone Grassland	NaN	NaN	0.37	-133	NaN	NaN	NaN	NaN
Flooded Pavement	NaN	NaN	0.28	-184	NaN	NaN	NaN	NaN

Table 7.3: Jaccard similarity and change in the area (10 sq. m.) of all the communities across turloughs from vegetation survey (2008) to Sentinel-2 survey (2017).

therefore, the spatial resolution of the satellite-derived maps is 10 m instead of 1 m (as of the vegetation survey). Given the challenge with more coarse spatial resolution for the satellite images, 15+ communities have been identified. Classification accuracies of 85% for Blackrock, 89.5% for Caherglassaun, 90.2% for Coy, and 91.8% for Garryland were achieved, making an average accuracy for classification of turloughs to be 89%.

As can be seen from Table 7.3, the main trend seems to be reduction in areas of the communities located higher up the turlough slopes with some increases in the wetter communities in Caherglassaun, Coy and to a lesser extent Garryland. This perhaps corroborates the known increase in more flooded conditions over the past 10 years. However, communities like *Agrostis stolonifera-Glyceria fluitans* were not well identified using S2 imagery due to their small size, which is thought to be the reason for such low similarity and high percentage areal change. Other communities like limestone grassland and flooded pavement were also not identified well. This mainly depends on the condition of the community at the time when the satellite images were captured. Other than that, most of the key vegetation communities show $\approx 20\%$ of the change in the area. The main question is whether this is an actual change in the spatial coverage of the communities or whether it is more a function of the satellite resolution and pixel mixing.

7.4 Discussion

Four turloughs, in a linked conduit-dominated karst system, with fluctuating water levels throughout the year have been evaluated in terms of the spatial distribution of different vegetation communities and their respective hydrological conditions experienced in such fluctuating wetland systems. The four ephemeral wetlands were chosen as they had similar water quality, soils and land use, and so the changes in hydrology parameters should provide the key variables differentiating between the different communities.

It is known that some species can tolerate a range of soil moisture/flooding, and are usually found almost throughout such turlough basins, for example *Agrostis stolonifera, Potentilla anserina*, and *Ranunculus repens*. whilst others have a more restricted range due to stricter habitat requirements, such as aquatic species which occur only in permanent water bodies.

Located in the lower parts of the turloughs is the Eleocharis acicularis community. This is

not a common community across Ireland, with restricted distribution in a limited number of turloughs making it of very high conservation value. It forms on relatively small patches on drying mud near water, usually at the very base of the turlough. The results here show that it typically experiences (and therefore requires) 6 to 7 months of inundation per year at depths of 0.75 to 1.85 m.

The next set of communities all experiencing similar average flood durations between 3 to 5 months per year on average are the *Agrostis stolonifera-Ranunculus repens, Poa annua-Plantago major* and *Potentilla anserina-Potentilla reptans* communities. The *Agrostis stolonifera-Ranunculus repens* community is found widespread across different turloughs. The community is relatively short (c. 25 cm) forb-dominated sward. This community was found in the upper to middle zones of the turlough basins in the NPWS project (Waldren et al. 2015), with a mean Ellenberg Wetness value 6.7, indicative of damp but not wet soils. The *Potentilla anserina-Potentilla reptans* community is a herb-dominated community, with a mean sward height of c.10 cm and is usually located in the middle to the bottom of the flooding gradient with mean Ellenberg Wetness value of 6.1, indicative of a damp site. The metrics in these four turloughs show that both communities appear to exist in locations with very similar ranges of flood duration and depth to the *Agrostis stolonifera-Ranunculus repens* community. The *Poa annua-Plantago major* community was found in areas where the integrity of the soil had been damaged through poaching, allowing the large proportion of ruderal species found in this type to colonise. The species list consists of perennials that can rapidly colonise from the surrounding grassland. More widely across turloughs this community was found on trampled ground in the upper reaches of the turlough basins; as suggested by the mean Ellenberg Wetness value of 5.9. Whilst it was found in areas with similar flood durations In these four turloughs to the other two communities, the range of flood depths it typically experiences seems to be a lot lower.

Next in order of flood duration is the *Filipendula ulmaria-Potentilla erecta-Viola* community which is a herb-rich community occurring in the middle of the flooding gradient. This was found to have a mean Ellenberg Wetness score of 6.2 across all turloughs, indicating that it occurs in damp sites. It is an important community is as it contains Viola persicaria and hybrids, a Red Data Book species.

There are then several different communities found in the middle to upper zones of the turloughs which are hard to separate using the ecohydrological variables: these include the *Carex nigra-Carex panicea, Carex nigra-Ranunculus flammula*, Woodland, Scrub, *Agrostis stolonifera-Potentilla anserina-Festuca rubra*, Limestone grassland, Lolium grassland and *Agrostis stolonifera–Glyceria fluitans* communities. Across a wider set of turlough these were all found to exist in areas with lower Eilenberg Wetness indices from 5 to 6 indicating damp but not constantly wet, substrate, with the exception of the *Agrostis stolonifera – Glyceria fluitans* and *Carex nigra-Ranunculus flammula* communities which more broadly seem to be located at the base or near the bottom of the turloughs, in areas that are likely to retain some standing water throughout the season, with mean Ellenberg value for Wetness is $\approx 8+$ (i.e. on wet sites). This did not seem to be the case in this selection of turloughs, although the communities were only found in very localised areas on the upper slopes of turloughs which perhaps could be near to springs, creating locally wet conditions in the soils. Most of these communities are typically found in grazed areas.

Finally, there is the flooded pavement community that occurs on exposed limestone pavement at the upper fringes of turloughs where open limestone pavement abuts the flood zone. The mean Ellenberg Wetness value of 5.7 is indicative of slightly damp soils. This community is of high conservation value, especially as habitat for *Potentilla fruticosa*, a species which is rare throughout the British Isles and largely restricted to the fringes of

some turloughs in Ireland. The results from these four turlough suggest that they survive in areas that are only flooded from 1 to 2 months per year.

The NPWS funded project on a wider range of turloughs (Waldren et al. 2015) concluded that duration of flooding and total phosphorus in the flood water of turloughs were the environmental variables most closely associated with the distribution of vegetation communities and vascular plant species and are therefore likely to be the most important ecological drivers of turlough vegetation. Given that these four linked turloughs analysed in this paper all had similar levels of phosphorus (soil, type, land use grazing etc.) then this analysis does indeed show that flood duration appears to be the key differentiating hydrological variable. There appears to be little to separate the communities with respect to flood frequency – all usually experiencing a regime of one to two floods per year. Equally the global radiation (and air temperature) of the time of year when the floods start to abate does not revel such large differences between the communities on average, indicating that perhaps this is not such a key parameter to the vegetation within the constraints of the hydrological regime in these four turloughs.

Finally, this research has attempted to demonstrate how this overall methodology could be augmented with the use of remote sensing data in order to provide a more regular way of surveying the vegetation communities, thereby allowing any changes to be picked up and attributed to changes in hydrological regime (whether anthropogenic or not). This could be used as an effective way of monitoring such groundwater dependent terrestrial ecosystems (GWDTEs). Whilst the satellite images did seem to have picked up some changes in the spatial coverage of the communities which could possibly be linked to slighter wetter conditions over the past 10 years, it is difficult to tell whether this is more a artefact of the accuracy of the relatively low resolution of the Sentinel-2 satellite for this type of image analysis, or whether it is actually linked to real changes in spatial distribution on the ground. However, in the future ever more high-resolution satellite data and/or with the use of drone/ satellite combination to improve accuracies (Bhatnagar et al. 2020a, Bhatnagar et al. 2020b), such an approach should yield further insights into the hydrological metrics affecting different vegetation communities.

7.5   Conclusion

This paper has evaluated the ecohydrology of intermittent wetlands, using the turloughs found in karst areas (and mainly found in Ireland) being extreme exemplars of such ephemeral flooding environments. A methodology of how to derive different ecohydrological variables associated with the spatial distribution of different vegetation communities has been outlined. Such metrics can then be evaluated alongside a wider mix of variables such as water quality (particularly nutrients), soil type, land-use, etc.in order to understand the habitat requirements for such plant communities and their associated ecological systems.

The analysis on these four turloughs on the same karst network has revealed distinct differences between vegetation communities, from *Eleocharis acicularis* found at the base of the turlough typically experiencing 6 to 7 months of inundation per year compared to the limestone pavement community at the top fringes of the turloughs only flooded from 1 to 2 months per year.

Finally, an approach that uses remotely sensed data to provide an assessment of whether there have been changes in the spatial distribution of the communities has been presented. A change in the spatial coverage of the communities with a slight decrease in "drier" com-

munities and increase in "wetter" communities seems to correlate with statistically wetter conditions in the turloughs measured over the past 10 years, compared to 10 years prior to that. However, more research is needed in this field to assess whether such changes are skewed as a result of the relatively low resolution of the Sentinel-2 satellite compared to the field survey used for this comparison over a 10-year period.

Chapter 8

Conclusions

"The opposite of a correct statement is a false statement but the opposite of a profound truth may well be another profound truth."

— Niels Bohr

Contents

8.1 Critical Reflection and Inferences

8.1.1 Summary of the thesis

Wetlands are important to study and monitor as they have a direct impact on the global carbon cycle and related climate change challenges, as well as providing many other ecosystem services. Up to now in Ireland, the mapping and monitoring of wetlands have been carried out manually by field visits to collect data. This process is both time and resource-intensive. Motivated by this, the study has formed methodologies to automatise the wetland mapping procedure by employing RS data along with ML and DL techniques.

The initial idea was to come up with a way to utilise the freely available satellite data for mapping the wetlands. Sentinel-2 with dedicated spectral bands for vegetation identification and 10 m spectral resolution was selected as the best choice of satellite data. Multiple spectral indices were calculated using the spectral range of S2. One of the first tasks was to detect the exact boundaries of the wetlands and therefore, a boundary delineation algorithm was created. This was followed by performing pixel-based classification on all the wetland types under consideration (raised bogs, fens and turloughs) to map various vegetation communities. It was shown that pixel-based methods perform well, but given the Sentinel-2 spatial resolution, spectral information on its own was not sufficient.

176

Therefore, the study then explored the contextual information present in wetlands in the form of neighbourhood pixels segmentation. These segments were applied on top of the spectral classification, which led to the removal of spurious, uncertain pixels, and more optimal mapping of wetland vegetation communities was achieved.

The study was extended to using drone imagery for a more detailed and flexible time-series analysis. Due to the presence of a high enough number of drone images, both ML and DL techniques could be applied for vegetation mapping. Considering the time, cost, and resources, ML techniques proved to be better for mapping unsurveyed Irish wetlands. The drone imagery has high spatial resolution but lacks spectral resolution in comparison to the satellite images. Also, the limited battery life of the drone made it challenging to capture and survey big wetlands. Therefore, keeping in mind the global coverage and spectra of the S2, a technique to augment S2 using drone imagery was formulated. Since the drone had only an RGB sensor, a colour correction technique was applied such that all the images captured from the drone can be utilised for further analysis. This nested approach allows the usage of drone imagery as the new ground truth for satellite images. It was seen that by using this methodology, subtle seasonal changes in the vegetation communities were getting picked up. A direct pixel-based confusion matrix was insufficient to quantify these changes. Therefore, an error metric was tailored using geometrical properties to detect these seasonal changes. This is a generalised methodology which can be applied globally for any application, not just wetlands in Ireland.

It is known that wetlands are one of the largest contributors of methane in the environment. Methane data from the recently launched Sentinel-5 Precursor was used to analyse the effect of methane emissions from wetland and use it for the identification of land types. However, since the S-5p methane product is available for cloud-free scenes, and has a relatively low spatial resolution of 5.5 km, not enough data is available for Ireland to carry out any meaningful analysis. The methane analysis of wetlands was, therefore, trialled out for Canada, given its much bigger spatial expanse and range of different temperate wetlands. This analyses using S-5p showed that different land-types could be detected across the whole of Canada. It was seen that the S-5p total column methane measurements are able to differentiate the positive and negative emission sensitivities from different land types.

Finally, the vegetation maps which were mapped using S2 data can also act as a new GT for verifying existing trends in hydrology. As an extension of previous studies done on turloughs, which mainly focussed on real-time hydrological data, this study has explored the possibility of using RS maps with time-series of hydrological data to form ecohydrological metrics for different vegetation communities in a turlough. It was seen that the RS maps could help in the identification of vegetation communities and the corresponding flooding depth, duration, and other parameters which indicate of the start of the growing season (temperature and global radiation), which can be updated on a regular basis. This is particularly useful for the formation of policies with regards to potential future impacts on such wetlands from anthropogenic disturbances (e.g. water abstractions, flood alleviation schemes etc.) and/or changes in climate which may affect the flooding patterns and dynamics in such ephemeral wetlands.

8.1.2 Key conclusions and inferences

- The boundary delineation algorithm was created using a combination of 3 edge detection techniques, namely entropy filtering, canny edge detection, and lazy snapping. It was seen that the ensemble of these methods along with the NDVI as the base image, gave very good delineation. For a large amount of wetlands, this is an automated way to

differentiate between 'wetland' and 'non-wetland' area. It was tested successfully on bogs and fens in order to define each wetland's extent such that only the correct area inside the wetland is analysed further. It should be noted that the algorithm did not work well for turloughs, as expected, as they do not have a steady boundary, and unlike bogs/fens merge into grasslands when not-flooded.

- In the pixel-based (PB) approach, it was seen that giving large training data leads to overfitting in case of PBIA (Chapter 2). Given that the aim of the study was to reduce the amount of fieldwork, and the training data is directly related to the amount of field area that has to be surveyed, a minimum (yet sufficient) amount of training data (percentage) was quantified. For the wetlands in Ireland, it appears that 8-12% of training data is enough to get good pixel-based accuracy.

- For the S2 data, all the key pixel-based classifiers like SVM, Naïve Bayes, KNN, BT, and RF were applied and examined for model accuracy, misclassification cost, prediction speed and training time. BT gave the best model accuracy with the least misclassification cost and training time. Therefore, for S2-PBIA, BT was chosen as the appropriate classifier.

- One of the contributions of the study is the introduction of contextual information along with spectral information. The segment-based study was done using a BT classifier (pixel-based) with Graph cut segmentation (maximum *a-posteriori*) min-cut alpha expansion algorithm. This combination, termed an MVC algorithm, was used to map 13 wetlands and up to 18 vegetation communities over time between 2017 to 2018. Given the spatial resolution of the data and size of the communities, an averaged OA of 85% was achieved.

- It was also seen that the addition of extra spectral bands such as SAVI, ARVI, NDVI, EVI, NDWI adds much more information to the original satellite image. The NDVI and EVI give an account of vegetation health, whereas, NDWI gives the idea about the soil moisture of the wetland. The vegetation communities often have the same appearance, and it is important to utilise their full spectra to get clear identification. These additional spectral indices help to differentiate between spectrally similar communities.

- It was also seen that the accuracy of MVC algorithm decreases slightly for smaller wetlands. This is because of the inevitable pixel-mixing that happens at a 10 m spatial resolution. Often the vegetation communities were small or sparsely distributed. Therefore in a 10 m patch, multiple communities can exist. The reflectance value captured by the satellite is the average value taken from the area. This does not give a clear difference between the sparse/small communities. Therefore, for precise mapping of smaller wetland communities, higher resolution data was required.

- In order to obtain higher-resolution data, a DJI Inspire-1 drone with Zenmuse X3 RGB camera was used for species-level mapping. The aspect ratio of the images was 4 : 3, which makes the size of the image 3000×4000 (area = 0.2 ha). The maximum flying time (per battery) was 15 minutes. However, due to wind-drag, battery consumption in take-off and landing, the maximum flight time was reduced to 12 minutes. In 12 minutes, an approximate area of 8-10 ha was captured. This was the main challenge while surveying large wetlands, as only a part of the wetland was captured every time.

- Previous studies have applied both ML and DL on drone images, and it was unclear which method would work well for Irish wetlands. Therefore, a detailed comparison of the two methods was made. It was seen that for ML, Random forest (RF) is the best classifier for drone imagery. The RF classification was done using textural information (Table 4.1),

with a total of 13 bands (3 RGB and 10 GLCM feature bands). It was seen that for Clara bog, the communities are low-lying and do not have a big difference in texture. Nevertheless, there was an increase of 2% by addition of the textural information on the overall classification. It was also seen that reducing the size of the image (from 3000×4000 to $2^9 \times 2^{10}$) did not make a big difference in terms of overall mapping accuracy (since the aspect ratio was maintained).

- The DL method outperformed ML by approx 4% OA with SegNet+ResNet50 architecture, giving the best results. For training DL models, all the images were resized (as mentioned above) and manually labelled. This requires a clear understanding of the area, and so for a new area, it might not be practical. For the computation of DL methods, a graphics processing unit (GPU) was required, housed in the Trinity College High Performing Computer Centre (TCHPC). GPU is expensive, and not commonly available for everyone. Therefore, for an unsurveyed wetland, the ML technique with around 85% accuracy but the requirement for much fewer training points was deemed to be the more appropriate choice.

- To take advantage of both satellite and drone data, a nested drone-satellite approach was created. The idea was to use a minimum number of drone images (with an affordable RGB camera/sensor) and temporally map a significantly bigger area. A detailed methodology of capturing drone images, georeferencing and mosaicking was given. An additional crucial step was developed where colour correction was applied. This step, often ignored in the literature, is necessary for a temperate climate like Ireland. Multiple colour correction techniques, including histogram correction, intensity grading, etc. were tested and the method suggested by Pitie et al. 2005 gave the best results. The drone imagery was then classified using the RF classifier, and the classified image was upsampled to 10 m using nearest neighbour interpolation. For discrete data, the nearest neighbourhood interpolation was shown to be the best upsampling technique. The upsampled image acts as a new GT for the satellite image. Using this new GT, the whole wetland can then be classified. This was done temporally across three seasons in 2019 for a raised bog. A final 2019 map was then obtained using majority voting of the results from the seasonal maps. While majority voting, in order to pick correct labels, the monthly maps were checked for additional geometrical features mentioned in the error metric.

- It was seen that the appearance and the behaviour of vegetation communities change from season to season. In Clara bog, such vegetation growth and die-back were identified using the error metric. This error metric also suggests that pixel-based comparison of classified images can be misleading and other criteria such as orientation, area, and the extent of the community are equally essential to confirm the change detection.

- Field studies have shown that there is a difference in methane emission rates from bogs, fens, marsh, swamp, etc. Therefore, all these wetlands, along with other land types, can also be differentiated and identified based on their methane emissions. For this, a new study based using ML based classification were tested with S-5p CH4 along with the retrieved surface albedo data over Canada and were able to identify six of the ten land types in the CWI. Amongst them are two major wetland types in Canada (marsh and swamp) covering significant area of the country. The other land types identified are forest, cropland, grassland and barren land. Monitoring the area covered and its change over time is crucial for land use, land use change and forestry (LULUCF). The land surface maps from S-5p data can be generated based on the CH4 positive or negative emission sensitivities detected from space on a daily, monthly, or yearly basis.

- As the vegetation content of the land types and in specific wetlands vary seasonally, mapping it using aerial photography or satellite imagery can lead to errors because of a lack of consistent vegetation pattern. The classification map generated using the proposed algorithm is complementary to the traditional methods of land type identification using aerial photography or satellite imagery. These maps can be used by the WAD2M (Wetland Area Dynamics for Methane Modeling) to either verify or complement where measurements from other sources are not available. This is a complete new use of satellite based CH4 product and the study showed the high potential of the data for land type identification of large areas, its monitoring and studying the dynamic change over time.

- The CH4 is the column-averaged dry-air mole fractions of methane and not just surface emissions. The total column methane is a combination of methane emitted locally and methane brought into the column due to wind-transportation from other places. Since the land types are getting identified with high accuracy using this satellite data, this, therefore, demonstrates that the CH4 emissions from the wetland types do have some significant effect on the total column. This is particularly useful for sensitivity analysis of land types. To our knowledge, this is the first every study using S-5p CH4 and SA product for quantifying local emissions and identification of land types. The results suggest that the data has a vast potential to be applied globally for analysing the effects of land types on methane emissions.

- Finally, a data analysis approach was applied for identifying the flooding trends in the ephemeral turlough wetlands using both field-surveyed map (2008) and a latest RS based map (2017). It was seen that for key communities, the spatial extent of communities remained mostly constant in multiple turloughs. The study has also provided clear eco-hydrological metrics defining the range of duration, depth, and other parameters which indicate of the start of the growing season (temperature and global radiation), experienced by each vegetation community (averaged over multiple turloughs).

8.1.3   Limitations of the research

1. Influence and importance of the amount of training data
   The amount of training data plays an essential role in the overall identification of vegetation communities. A completely unsupervised study can show the patterns present on the ground, but it is never certain. Therefore, some amount of supervision needs to be provided for mapping the wetlands.

   How much training data is enough?

   For making the model robust, and easily applicable, it should require the least supervision or parameterisation. For S2, each pixel is $10m^2$ area; therefore, to get 1000 pixels as training, an area of $10,000m^2$ would need to be mapped manually. As can be seen from Chapter 2, giving 50%-70% of the data for training leads to overfitting. Therefore, from Chapter 3 onwards a maximum of 12% training data was used to map Irish wetlands – which fetched good results. However, it should be noted that this amount is still a lot for a wetland with a big area, and especially for unsurveyed wetlands.
   Also, the training data, if not correctly distributed, can bring an uncanny bias in the model. For example, if the area of the communities varies significantly, the training data has to be divided such that any class is not favoured. In this study, all the training data was randomly initialised taking the same % from each class, hence, making it uniform. It

was seen that when just 8% of training data was appropriately distributed, the accuracy is high (Chapter 5).

2. Accuracy dependence on the ground truth data
   For cross-checking the satellite and drone maps, ground truth data collected in the field was used. The ground truth surveys were done at a particular time of the year and did not depict the seasonality in the vegetation. Therefore, when the seasonal satellite or drone maps are checked against the ground truth, there can be some discrepancy. This does not necessarily mean that the classified map is wrong; it only suggests that the classified map is different from the ground truth map. Therefore, it can be suggested that accuracy can be only as good as the ground truth. This also explains the need for regular updating of the ground reference maps, if possible.

3. Lack of data due to cloud coverage
   In Ireland, one of the major challenges was the unavailability of the satellite data due to high cloud coverage (especially in the west of Ireland). This is particularly problematic when looking for seasonal changes. An alternative is to use SAR (S1) data instead, but it further reduces the spatial resolution. For vegetation analysis, S2 works very well, but due to the lack of cloud-free images, the study was not able to do a complete monthly mapping of vegetation every year.
   Similar to S2, S-5p XCH4 product is derived using IR band, and hence, is sensitive to clouds. Even though S-5p has a daily temporal resolution, there were only a few months for which the data was available for the whole of Ireland. Due to insufficiency of data, the modelling of S-5p (XCH4) was not feasible to be carried out for Ireland.

4. Limitations of training data lead to a problem in applying DL.
   Whereas for ML, the aim was to reduce the input training data, for the DL methods, it is never a feasible option. In order to get a robust model, a large quantity of training data (images) was required. A time series of satellite imagery, as can be seen in Chapter 3, contains a maximum of 9 images. Therefore, DL could not be applied to satellite imagery. With the use of the drone, a large number of images were captured, which was a sufficient amount of training data for DL, along with transfer learning. For this, all the images were fully labelled manually. The labelling process is time and resource consuming. Moreover, it requires expert supervision to label all the classes. For an unsurveyed wetland, creating such fully labelled training data would be very challenging. Therefore, ML techniques were overall deemed more appropriate (Chapter 4).

5. Non-transferability of the model
   One of the fundamental limitations, which comes out from the study, is that every wetland has a different functioning in terms of water level, nutrient level, etc. It is rare that two wetlands located close to one another also have exact same phenology. As can be seen from the spectral signature plots (Chapter 5), the same communities also change significantly over various seasons. All this suggests that a new (fresh) training of the model is required for correct identification of the land type. A single model trained for one wetland, cannot be applied directly to another wetland. This is partly also because of the change in satellite imagery. The spectra being captured by the satellite depends on the sun-zenith angle, temperature, and other atmospheric parameters. These parameters change significantly from East to West coast. Atmospheric correction models such as sen2cor (for S2 data), Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH), and others are available. For the data being applied in the study (S2), sen2cor correction was already applied.

Moreover, even then a difference in the spectral signature of the same community located in multiple locations was seen. In order to curb this problem, a binary classification for identification of the ecologically significant 'central' ecotope on raised bogs was taken (Appendix B). This technique was also limited to the bogs captured in the same tile and was not robust.

A similar trend was noticed using S-5p data, where the classification model that was trained on the east of Canada and was then applied to the West of Canada. The model fails, as the emissions from the west are different from the east. Therefore, it suggests that the model needs training data from both west and east, or it cannot be successfully applied.

8.2   Recommendations for future research

8.2.1   Guidelines for GWDTE health monitoring using RS in Ireland

Updating an already surveyed wetland map

- For updating an area's map, the old map can be used as it is with recent S2 imagery.

- The imagery has to be pre-processed, as described in Chapter 3 (re-sampling, subsetting).

- Along with the default spectral bands, additional vegetation indices (bands) such as NDVI, EVI, SAVI, ARVI; and water indices such NDWI, NDWI2 are to be calculated. The additional bands are to be stacked along with the default bands of S2 for further usage.

- 8-12% of training data can be given as an input to the training model (MVC algorithm), which is sufficient for producing correct maps.

- The classified map can be cross-checked with the original GT map, along with error-metric calculation (as described in Chapter 5).

- The areas with the severe changes, where a community is not identified at all or the $\triangle A$, $\triangle O$, $\triangle E$ is large, should be further surveyed to confirm.

- For this surveying, a drone can be flown only over the uncertain area, and the experts can review the collected imagery for confirming the status of the vegetation community.

- In case the drone imagery does not give sufficient information or idea about the community, a manual field-survey by ecologists can be performed in that area.

- It has to be noted that the field survey required is to be done only on the uncertain area – therefore, it still saves a lot of time and resources as compared to surveying the whole wetland.

Mapping an unsurveyed wetland

- For an unsurveyed wetland, using the satellite imagery, an unsupervised (k-means) clustering can be performed. This will give an idea about the potential classes which are present in the ground.

- The value of 'k' in k-means can vary between 5-10 (hit and trial) depending on the size of the wetland.

- For linking the vegetation to the clusters, old maps of similar wetlands can be utilised. Also, many wetlands, such as raised bogs, have an existing trend. For example, in the raised bog, any cluster formed in the middle is most likely to be central/subcentral. The boundary would usually consist of marginal/facebank. Therefore, a preliminary map can be prepared this way.

- For a much better understanding, the nested drone-satellite approach mentioned in Chapter 5 can be applied.

- Here using just 3-4 drone images, the whole of wetland can be mapped. Usually, these points are hard to pick up by using satellite data alone.

- Only a few points are required to be identified as a community (by the expert), and the rest of the training data is automatically created using neighbourhood information.

- This way an unsurveyed wetland can be mapped just in a day, saving a lot of time and resource.

Lastly, the satellite-based MVC algorithm developed in this study could be applied by National Parks and Wildlife Services, Ireland for mapping wetlands and other habitats across Ireland. All the codes and data has already been shared with NPWS to start using the methodology. As the methodology uses open-source data and can be applied free of cost, that can help NPWS to update the maps of wetlands which are inaccessible due to land access / logistical reasons. This will also allow them to have temporal monitoring, instead of surveying in 5 years due to time and cost involved.

8.2.2 More recommendations

1. Gross Primary Productivity (GPP) is one of the critical measures of carbon balance which represents the total carbon uptake through photosynthesis per unit of area. GPP determines the amount of atmospheric CO2 sequestered into biomass and expedites the microbial decomposition as well as the production of methane and dissolved organic carbon. Thus, detailed knowledge of GPP dynamics in space and time and key drivers affecting it is imperative for improving our predictions of peatland ecology, biogeochemistry and carbon balance in response to global change. This is currently being done for raised bogs in Ireland (as part of the EPA funded SmartBog project). The summary of the paper (which is currently under review) is as follows:
Carbon flux models based on remotely sensed data have been significantly utilised to estimate Gross Primary Productivity (GPP) for agriculture, forestry and grassland sites. The spectral signature of sphagnum mosses is characteristically different as compared to the vascular vegetation. The water content, vegetation communities and the reflectance of the bogs are different from these other ecosystems. Therefore, there is a need to develop a vegetation index, especially for bogs. Hence, it is crucial to evaluate the relationship between measured and modelled GPP for a raised bog ecosystem using Light Use Efficiency model and S2 satellite imagery. It is also essential to develop a vegetation index for raised bog ecosystem based on a combination of existing vegetation indices and field measurements. The performance of the developed index should be validated under typical

(normal) and exceptional weather scenarios.

2. The DL has excellent potential for species-level identification using both satellite and drone data. Therefore, in future, there is a need to simplify the application of DL by introducing a training library specific to vegetation communities. The major challenge with the limited satellite and drone images is the formation of the training data. Therefore, such an open-repository where participants can globally upload images of multiple vegetation communities would be very useful for training the CNN model.

3. The Sentinel-5P data provides extensive information about GHGs at a very high (1 day) temporal resolution. Such information is highly beneficial for disaster management, where quick response and analysis is required, for example, in events of a forest fire. This data can be applied to identify reductions in wetland areas, forest areas and/or agricultural areas globally.

4. The hydrological trend analysis should be extended to turloughs with different water quality indices. It is known that water quality, as well as the hydrological dynamics studied in this project, have a direct impact on the type of vegetation communities present in an area. Therefore, a new hydrological metric taking into consideration poor and nutrient-rich turloughs need to be created.

Bibliography

Adam, P., Birks, H., Huntley, B. & Prentice, I. (1975), 'Phytosociological studies at malham tarn moss and fen', *Yorkshire, England. Plant Ecology* **30**(2), 117–132.

Adobe Photoshop Lightroom Beta (2019). last accessed on 15 November 2019.
   URL: *http://labs.adobe.com/technologies/lightroom/*

Agarwal, A., Singh, A., Kumar, S. & Singh, D. (2018), Critical analysis of classification techniques for precision agriculture monitoring using satellite and drone, *in* '2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS', IEEE, p. 83–88.

Akar, (2017), 'Mapping land use with using rotation forest algorithm from uav images', *European Journal of Remote Sensing* **50**(1), 269–279,.

Alvarez-Taboada, F., Paredes, C. & Julián-Pelaz, J. (2017), 'Mapping of the invasive species hakea sericea using unmanned aerial vehicle (uav) and worldview-2 imagery and an object-oriented approach', *Remote Sensing* **9**(9), 913.

Amani, M., Mahdavi, S., Afshar, M., Brisco, B., Huang, W., Mohammad Javad Mirzadeh, S., White, L., Banks, S., Montgomery, J. & Hopkinson, C. (2019), 'Canadian wetland inventory using google earth engine: the first map and preliminary results', *Remote Sensing* **11**(7), 842.

Amani, M., Salehi, B., Mahdavi, S. & Granger, J. (2017), 'Spectral analysis of wetlands in newfoundland using sentinel 2a and landsat 8 imagery', *Proceedings of the IGTF* .

Andrews, H. & Patterson, C. (1976), 'Digital interpolation of discrete images', *IEEE Transactions on Computers* **100**(2), 196–202.

ArcMap Desktop (Version 10.6.1) (2019).

Assessment, M. E. (2005), *Ecosystems and Human Well-being: Synthesis*, Island Press, Washington, DC.

Badrinarayanan, V., Kendall, A. & Cipolla, R. (2017), 'Segnet: A deep convolutional encoder-decoder architecture for image segmentation', *IEEE transactions on pattern analysis and machine intelligence* **39**(12), 2481–2495.

Baena, S., Moat, J., Whaley, O. & Boyd, D. (2017), 'Identifying species from the air: Uavs and the very high resolution challenge for plant conservation', *PloS one* **12**(11).

Balakrishnama, S. & Ganapathiraju, A. (1998), 'Linear discriminant analysis-a brief tutorial', *Institute for Signal and information Processing* **18**, 1–8.

Ball, G. H. & Hall, D. J. (1965), Isodata, a novel method of data analysis and pattern classification, Technical report, Stanford research inst Menlo Park CA.

Balogun, A.-L., Yekeen, S. T., Pradhan, B. & Althuwaynee, O. F. (2020), 'Spatio-temporal analysis of oil spill impact and recovery pattern of coastal vegetation and wetland using multispectral satellite landsat 8-oli imagery and machine learning models', *Remote Sensing* **12**(7), 1225.

Barbier, E. B. (2011), 'Wetlands as natural assets', *Hydrological Sciences Journal* **56**(8), 1360–1373.

Bartsch, A., Wagner, W., Scipal, K., Pathe, C., Sabel, D. & Wolski, P. (2009), 'Global monitoring of wetlands–the value of envisat asar global mode', *Journal of Environmental Management* **90**(7), 2226–2233.

Bauer, E. & Kohavi, R. (1999), 'An empirical comparison of voting classification algorithms: Bagging, boosting, and variants', *Machine learning* **36**(1-2), 105–139.

Beamish, D. (2014), 'Peat mapping associations of airborne radiometric survey data', *Remote Sensing* **6**(1), 521–539.

Beltman, B., Van den Broek, T., Bloemen, S. & Witsel, C. (1996), 'Effects of restoration measures on nutrient availability in a formerly nutrient-poor floating fen after acidification and eutrophication', *Biological Conservation* **78**(3), 271–277.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. & Heynen, M. (2004), 'Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information', *ISPRS Journal of photogrammetry and remote sensing* **58**(3-4), 239–258.

Berhane, T. M., Lane, C. R., Wu, Q., Anenkhonov, O. A., Chepinoga, V. V., Autrey, B. C. & Liu, H. (2018a), 'Comparing pixel-and object-based approaches in effectively classifying wetland-dominated landscapes', *Remote sensing* **10**(1), 46.

Berhane, T. M., Lane, C. R., Wu, Q., Autrey, B. C., Anenkhonov, O. A., Chepinoga, V. V. & Liu, H. (2018b), 'Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory', *Remote sensing* **10**(4), 580.

Bhatnagar, S., Ghosh, B., Regan, S., Naughton, O., Johnston, P. & Gill, L. (2018), 'Monitoring environmental supporting conditions of a raised bog using remote sensing techniques', *Proceedings of the International Association of Hydrological Sciences* **380**, 9–15.

Bhatnagar, S., Ghosh, B., Regan, S., Naughton, O., Johnston, P. & Gill, L. (2019), Remote sensing based ecotope mapping and transfer of knowledge in raised bogs, *in* 'Geophysical Research Abstracts', Vol. 21.

Bhatnagar, S., Gill, L. & Ghosh, B. (2020b), 'Drone image segmentation using machine and deep learning for mapping raised bog vegetation communities', *Remote Sensing* **12**(16), 2602.

Bhatnagar, S., Gill, L., Regan, S., Naughton, O., Johnston, P., Waldren, S. & Ghosh, B. (2020a), 'Mapping vegetation communities inside wetlands using sentinel-2 imagery in ireland', *International Journal of Applied Earth Observation and Geoinformation* **88**, 102083.

Bohn, T. J., Melton, J. R., Ito, A., Kleinen, T., Spahni, R., Stocker, B., Zhang, B., Zhu, X., Schroeder, R., Glagolev, M. V. et al. (2015), 'Wetchimp-wsl: intercomparison of wetland methane emissions models over west siberia', *Biogeosciences (Online)* **12**(11).

Bourgeau-Chavez, L., Endres, S., Battaglia, M., Miller, M. E., Banda, E., Laubach, Z., Higman, P., Chow-Fraser, P. & Marcaccio, J. (2015), 'Development of a bi-national great lakes coastal wetland and land use map using three-season palsar and landsat imagery', *Remote Sensing* **7**(7), 8655–8682.

Bourgeau-Chavez, L., Kasischke, E., Brunzell, S., Mudd, J., Smith, K. & Frick, A. (2001), 'Analysis of space-borne sar data for wetland mapping in virginia riparian ecosystems', *International Journal of Remote Sensing* **22**(18), 3665–3687.

Bourgeau-Chavez, L. L., Endres, S., Powell, R., Battaglia, M. J., Benscoter, B., Turetsky, M., Kasischke, E. S. & Banda, E. (2017), 'Mapping boreal peatland ecosystem types from multitemporal radar and optical satellite imagery', *Canadian Journal of Forest Research* **47**(4), 545–559.

Boykov, Y. & Kolmogorov, V. (2004), 'An experimental comparison of min-cut/max-flow algorithms for energy minimisation in vision', *IEEE transactions on pattern analysis and machine intelligence* **26**(9), 1124–1137.

Boykov, Y., Veksler, O. & Zabih, R. (2001), 'Fast approximate energy minimization via graph cuts', *IEEE Transactions on pattern analysis and machine intelligence* **23**(11), 1222–1239.

Braun, A., Weidner, U. & Hinz, S. (2010), 'Support vector machines for vegetation classification–a revision', *Photogrammetrie-Fernerkundung-Geoinformation* **2010**(4), 273–281.

Breidenbach, J., Waser, L. T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., Puliti, S. & Astrup, R. (2020), 'National mapping and estimation of forest area by dominant tree species using sentinel-2 data', *arXiv preprint arXiv:2004.07503* .

Breidt, F. J. & Opsomer, J. D. (2000), 'Local polynomial regression estimators in survey sampling', *Annals of statistics* pp. 1026–1053.

Breiman, L. (1996), 'Bagging predictors', *Machine learning* **24**(2), 123–140.

Breiman, L., Friedman, J., Stone, C. & Olshen, R. (1984), *Classification and regression trees*, CRC press.

Bureau, R. C. (2001), 'Wetlands values and functions. ramsar convention bureau'.

Cabezas, J., Galleguillos, M. & Perez-Quezada, J. F. (2016), 'Predicting vascular plant richness in a heterogeneous wetland using spectral and textural features and a random forest algorithm', *IEEE Geoscience and Remote Sensing Letters* **13**(5), 646–650.

Carlotto, M. (2009), 'Effect of errors in ground truth on classification accuracy', *International Journal of Remote Sensing* **30**(18), 4831–4849.

Casanova, M. & Brock, M. (2000), 'How do depth, duration and frequency of flooding influence the establishment of wetland plant communities?', *Plant Ecology* **147**, 237–250.

Cassardo, C., Cremonini, R., Ferrarese, S., Bechini, D. R. & Tabone, I. (2014), 'Comparison of meteorological observations and wrf model retrieved cloud top height for the jem-euso experiment'.

Chabot, D., Dillon, C., Shemrock, A., Weissflog, N. & Sager, E. (2018), 'An object-based image analysis workflow for monitoring shallow-water aquatic vegetation in multispectral drone imagery', *ISPRS International Journal of Geo-Information* **7**(8), 294.

Chai, H., Wee, L., Swee, T. & Hussain, S. (2011), 'Gray-level co-occurrence matrix bone fracture detection', *WSEAS Transactions on Systems* **10**(1), 7–16.

Chasmer, L., Mahoney, C., Millard, K., Nelson, K., Peters, D., Merchant, M., Hopkinson, C., Brisco, B., Niemann, O., Montgomery, J. et al. (2020), 'Remote sensing of boreal wetlands 2: Methods for evaluating boreal wetland ecosystem state and drivers of change', *Remote Sensing* **12**(8), 1321.

Chen, S., Cao, L., Liu, J. & Tang, X. (2007), Iterative map and ml estimations for image segmentation, *in* 'Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on', IEEE, p. 1–6.

Chen, Y., Guerschman, J. P., Cheng, Z. & Guo, L. (2019), 'Remote sensing for vegetation monitoring in carbon capture storage regions: A review', *Applied energy* **240**, 312–326.

Cheng, K., Cheng, X., Wang, Y., Bi, H. & Benfield, M. (2019), 'Enhanced convolutional neural network for plankton identification and enumeration', *PloS one* **14**(7).

Cole, B., McMorrow, J. & Evans, M. (2014), 'Empirical modelling of vegetation abundance from airborne hyperspectral data for upland peatland restoration monitoring', *Remote Sensing* **6**(1), 716–739.

Comer, M. L. & Delp, E. J. (1995), Multiresolution image segmentation, *in* '1995 International Conference on Acoustics, Speech, and Signal Processing', Vol. 4, IEEE, pp. 2415–2418.

Connolly, J. (2019), 'Mapping land use on irish peatlands using medium resolution satellite imagery', *Irish Geography* **51**(2), 187–204.

Copernicus Open Access Hub (2018).
URL: *https://scihub.copernicus.eu.*

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R. & Schiele, B. (2016), The cityscapes dataset for semantic urban scene understanding, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', p. 3213–3223.

Cortes, C. & Vapnik, V. (1995), 'Support-vector networks', *Machine Learning* **20**(3), 273–297.
URL: *https://doi.org/10.1007/BF00994018*

Coxon, C. E. (1986), *A study of the hydrology and geomorphology of turloughs*, Trinity College.

Crichton, K., Anderson, K., Bennie, J. & Milton, E. (2015), 'Characterizing peatland carbon balance estimates using freely available landsat etm+ data', *Ecohydrology* **8**(3), 493–503.

Cui, B., Zhang, Y., Li, X., Wu, J. & Lu, Y. (2019), Wetlandnet: Semantic segmentation for remote sensing images of coastal wetlands via improved unet with deconvolution, *in* 'International Conference on Genetic and Evolutionary Computing', Springer, Singapore, p. 281–292.

Cunha Pereira, H., Allott, N. & Coxon, C. (2010), 'Are seasonal lakes as productive as permanent lakes? a case study from ireland', *Canadian Journal of Fisheries Aquatic Sciences* **67**(8), 1291–1302.

D'Acunha, B., Lee, S.-C. & Johnson, M. S. (2018), 'Ecohydrological responses to rewetting of a highly impacted raised bog ecosystem', *Ecohydrology* **11**(1), e1922.

Dalponte, M., Coops, N. C., Bruzzone, L. & Gianelle, D. (2009), 'Analysis on the use of multiple returns lidar data for the estimation of tree stems volume', *IEEE journal of selected topics in applied earth observations and remote sensing* **2**(4), 310–318.

Davis, S. M., Landgrebe, D. A., Phillips, T. L., Swain, P. H., Hoffer, R. M., Lindenlaub, J. C. & Silva, L. F. (1978), *Remote sensing: The quantitative approach.*

Davranche, A., Lefebvre, G. & Poulin, B. (2010), 'Wetland monitoring using classification trees and spot-5 seasonal time series', *Remote sensing of environment* **114**(3), 552–562.

Day, W. & Edelsbrunner, H. (1984), 'Efficient algorithms for agglomerative hierarchical clustering methods', *Journal of classification* **1**(1), 7–24.

De Becker, P., Hermy, M. & Butaye, J. (1999), 'Ecohydrological characterisation of a groundwater-fed alluvial floodplain mire', *Applied Vegetation Science* **2**, 215–228.

De Groot, R., Brander, L., Van Der Ploeg, S., Costanza, R., Bernard, F., Braat, L., Christie, M., Crossman, N., Ghermandi, A., Hein, L. & Hussain, S. (2012), 'Global estimates of the value of ecosystems and their services in monetary units', *Ecosystem services* **1**(1).

De Zan, F. & Guarnieri, A. M. (2006), 'Topsar: Terrain observation by progressive scans', *IEEE Transactions on Geoscience and Remote Sensing* **44**(9), 2352–2360.

Dechesne, C., Mallet, C., Le Bris, A. & Gouet-Brunet, V. (2017), 'Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery', *ISPRS Journal of Photogrammetry and Remote Sensing* **126**, 129–145.

DeLancey, E. R., Simms, J. F., Mahdianpari, M., Brisco, B., Mahoney, C. & Kariyeva, J. (2020), 'Comparing deep learning and shallow learning for large-scale wetland classification in alberta, canada', *Remote Sensing* **12**(1), 2.

Desai, A. R., Xu, K., Tian, H., Weishampel, P., Thom, J., Baumann, D., Andrews, A. E., Cook, B. D., King, J. Y. & Kolka, R. (2015), 'Landscape-level terrestrial methane flux observed from a very tall tower', *Agricultural and Forest Meteorology* **201**, 61–75.

Díaz-Delgado, R., Cazacu, C. & Adamescu, M. (2019), 'Rapid assessment of ecological integrity for lter wetland sites by using uav multispectral mapping', *Drones* **3**(1), 3.

Didan, K. (2015), 'Mod13a3 modis/terra vegetation indices monthly l3 global 1km sin grid v006', *NASA EOSDIS Land Processes DAAC* .

Dietterich, T. G. (2000), Ensemble methods in machine learning, *in* 'International workshop on multiple classifier systems', Springer, pp. 1–15.

DigitalGlobe (2020), `https://www.digitalglobe.com/`. Accessed: 20 August 2020.

Dingle Robertson, L., King, D. J. & Davies, C. (2015), 'Object-based image analysis of optical and radar variables for wetland evaluation', *International Journal of Remote Sensing* **36**(23), 5811–5841.

Dissanska, M., Bernier, M. & Payette, S. (2009), 'Object-based classification of very high resolution panchromatic images for evaluating recent change in the structure of patterned peatlands', *Canadian Journal of Remote Sensing* **35**(2), 189–215.

Divamgupta (2019).
URL: *https://github.com/divamgupta/image-segmentation-keras.git*

Do, D., Pham, F., Raheja, A. & Bhandari, S. (2018), Machine learning techniques for the assessment of citrus plant health using uav-based digital images, *in* 'Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping III', Vol. 10664, International Society for Optics and Photonics, p. 106640.

Dolgopolov, A. V., Smith, P. M., Christensen, C. B., Stroup, T. & Jones, T. (2018), Analysis of the commercial satellite industry, *in* '2018 AIAA SPACE and Astronautics Forum and Exposition', p. 5296.

Draper, F. C., Roucoux, K. H., Lawson, I. T., Mitchard, E. T., Coronado, E. N. H., Lähteenoja, O., Montenegro, L. T., Sandoval, E. V., Zaráte, R. & Baker, T. R. (2014), 'The distribution and amount of carbon in the largest peatland complex in amazonia', *Environmental Research Letters* **9**(12), 124017.

Drew, D. (2018), 'Karst of ireland: Lanscape, hydrogeology, methods', *Geological Survey* .

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. et al. (2012), 'Sentinel-2: Esa's optical high-resolution mission for gmes operational services', *Remote sensing of Environment* **120**, 25–36.

Duda, R. (2001), 'Hart. p., e., and stork, dg," pattern classification"', *Nueva York: John Wiley and Sons* .

Dvořák, P., Müllerová, J., Bartaloš, T. & Brna, J. (2015), 'Unmanned aerial vehicles for alien plant species detection and monitoring', *International Archives of the Photogrammetry, Remote Sensing Spatial Information Sciences* **40**.

Döngi, F. (2011), Earth observation from space - the european landscape in the second decade, *in* 'Proceedings of SPIE, Remote Sensing 2011', Vol. 8181, Prague, Czech Republic. URL:.
URL: *http://spie.org/Documents/AboutSPIE/PDF/ERS11-plenary-Doengi.pdf*

Ellenberg, H. (1988), *Vegetation Ecology of Central Europe*, Cambridge University Press, UK.

Environment and Climate Change Canada. Canadian Environmental Sustainability Indicators: Extent of Canada's Wetlands. (2016), `www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=69E2D25B-1`. Consulted on 04 October 2020.

Erudel, T., Fabre, S., Houet, T., Mazier, F. & Briottet, X. (2017), 'Criteria comparison for classifying peatland vegetation types using in situ hyperspectral measurements', *Remote Sensing* **9**(7), 748.

Erwin, K. L. (2009), 'Wetlands and global climate change: the role of wetland restoration in a changing world', *Wetlands Ecology and management* **17**(1), 71.

ESRI, World 30cm Imagery (2020).
URL: *http://www.arcgis.com/home/*

European Space Agency (2020), `https://www.esa.int/`. Accessed: 20 August 2020.

Felicity Hayes-McCoy, W. (2017), *Dingle and its Hinterland: People, Places and Heritage*, the Collins Press.

Feng, Q., Liu, J. & Gong, J. (2015), 'Uav remote sensing for urban vegetation mapping using random forest and texture analysis', *Remote sensing* **7**(1), 1074–1094.

Fernandez, F., Connolly, K., Crowley, W., Denyer, J., Duff, K. & Smith, G. (2014), 'Raised bog monitoring and assessment survey 2013', *Irish Wildlife Manuals* (81).

Finlayson, C. M., Everard, M., Irvine, K., McInnes, R. J., Middleton, B., Van Dam, A. A. & Davidson, N. C. (2018), The wetland book: Structure and function, management and methods. i, Springer.

Floreano, D. & Wood, R. J. (2015), 'Science, technology and the future of small autonomous drones', *Nature* **521**(7553), 460–466.

Foss, P. & Crushell, P. (2008), 'Guidelines for a national fen survey of ireland survey manual. report to the national parks and wildlife service'.

Fossitt, J. A. (2000), *A guide to habitats in Ireland*, Heritage.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. & Huang, X. (2010), 'Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets', *Remote sensing of Environment* **114**(1), 168–182.

Friedl, M. & Brodley, C. (1997), 'Decision tree classification of land cover from remotely sensed data', *Remote sensing of environment* **61**(3), 399–409.

Fushiki, T. (2011), 'Estimation of prediction error by using k-fold cross-validation', *Statistics and Computing* **21**(2), 137–146.

Förster, M., Schmidt, T., Wolf, R., Kleinschmit, B., Fassnacht, F., Cabezas, J. & Kattenborn, T. (2017), Detecting the spread of invasive species in central chile with a sentinel-2 time-series, *in* '2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp', IEEE, p. 1–4.

Bibliography

Gao, B. (1996), 'Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space', *Remote sensing of environment* **58**(3), 257–266.
URL: *https://doi.org/10.1016/S0034-4257(96)00067-3*

Gao, X., Huete, A., Ni, W. & Miura, T. (2000), 'Optical–biophysical relationships of vegetation spectra without background contamination', *Remote Sensing of Environment* **74**(3), 609–620.
URL: *https://doi.org/10.1016/S0034-4257(00)00150-4*

Gardner, R. C. & Finlayson, C. (2018), Global wetland outlook: state of the world's wetlands and their services to people, *in* 'Ramsar Convention Secretariat'.

Gatti, A. & Bertolini, A. (2013), 'Sentinel-2 products specification document'.

Ghosh, S., Das, N., Das, I. & Maulik, U. (2019), 'Understanding deep learning techniques for image segmentation', *ACM Computing Surveys (CSUR* **52**(4), 73.

Gill, L., Naughton, O. & Johnston, P. (2013), 'Modeling a network of turloughs in lowland karst', *Water Resources Research* **49**(6), 3487–3503.

Gill, L., Naughton, O., Johnston, P., Basu, B. & Ghosh, B. (2013), 'Characterisation of hydrogeological connections in a lowland karst network using time series analysis of water levels in ephemeral groundwater-fed lakes (turloughs)', *Journal of hydrology* **499**, 289–302.

Gill, L. W., Babechuk, M. G., Kamber, B. S., McCormack, T. & Murphy, C. (2018), 'Use of trace and rare earth elements to quantify autogenic and allogenic inputs within a lowland karst network', *Applied Geochemistry* **90**, 101–114.

Glatzel, S., Worrall, F., Boothroyd, I., Drollinger, S., Moody, C. & Clay, G. (2020), Differences in peat formation between an atlantic blanket bog and a subcontinental raised bog, *in* 'EGU General Assembly Conference Abstracts', p. 7402.

Goodall, D. & Gore, A. (1983), Ecosystems of the world: mires-swamp, bog, fen and moor-regional studies, *in* 'Ecosystems of the world', Vol. 4, Elservier.

Goodwillie, R. & Reynolds, J. (2003), *Turloughs*, University College Dublin Press, Dublin.

Gou, S., Gonzales, S. & Miller, G. R. (2015), 'Mapping potential groundwater-dependent ecosystems for sustainable management', *Groundwater* **53**(1), 99–110.

Govender, M., Chetty, K. & Bulcock, H. (2007), 'A review of hyperspectral remote sensing and its application in vegetation and water resource studies', *Water Sa* **33**(2).

Gray, P., Ridge, J., Poulin, S., Seymour, A., Schwantes, A., Swenson, J. & Johnston, D. (2018), 'Integrating drone imagery into high resolution satellite remote sensing assessments of estuarine environments', *Remote Sensing* **10**(8), 1257.

Grenier, M., Labrecque, S., Garneau, M. & Tremblay, A. (2008), 'Object-based classification of a spot-4 image for mapping wetlands in the context of greenhouse gases emissions: the case of the eastmain region, québec, canada', *Canadian Journal of Remote Sensing* **34**(sup2), S398–S413.

Grm, K., Štruc, V., Artiges, A., Caron, M. & Ekenel, H. (2017), 'Strengths and weaknesses of deep learning models for face recognition against image degradations', *Iet Biometrics* **7**(1), 81–89.

Guo, G., Wang, H., Bell, D., Bi, Y. & Greer, K. (2003), Knn model-based approach in classification, *in* 'OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"', Springer, Berlin, Heidelberg, p. 986–996.

Guo, M., Li, J., Sheng, C., Xu, J. & Wu, L. (2017), 'A review of wetland remote sensing', *Sensors* **17**(4), 777.

Hamdan, O., Nor Azura, O., Nurul Dasani, A., Mohd Azahari, F. & Abd Rahman, K. (2015), 'Airborne lidar for estimating aboveground biomass in dipterocarp forests of malaysia'.

Han, S., Pool, J., Tran, J. & Dally, W. (2015), Learning both weights and connections for efficient neural network, *in* 'Advances in neural information processing systems', p. 1135–1143.

Han, Y., Yoo, S. & Kwon, O. (2017), 'Possibility of applying unmanned aerial vehicle (uav) and mapping software for the monitoring of waterbirds and their habitats', *Journal of Ecology and Environment* **41**(1), 21.

Haralick, R. M., Shanmugam, K. et al. (1973), 'Textural features for image classification', *IEEE Transactions on systems, man, and cybernetics* (6), 610–621.

Hasekamp, O., Lorente, A., Hu, H., Butz, A., de Brugh, J. & Landgraf, J. (2019), 'Algorithm theoretical baseline document for sentinel-5 precursor methane retrieval', *Netherlands Institute for Space Research* p. 67.

Havens, K. & Nurnberg, G. (2004), 'The phosphorus-chlorophyll relationship in lakes: Potential influences of color and mixing regime', *Lake and Reservoir Management* **20**, 188–196.

He, K., Zhang, X., Ren, S. & Sun, J. (2016), Deep residual learning for image recognition, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', p. 770–778.

Henrich, V., Götze, E., Jung, A., Sandow, C., Thürkow, D. & Gläßer, C. (2009), Development of an online indices database: Motivation, concept and implementation, *in* 'Proceedings of the 6th EARSeL Imaging Spectroscopy SIG Workshop Innovative Tool for Scientific and Commercial Environment Applications, Tel Aviv, Israel', pp. 16–18.

Hess, L. & Melack, J. (2003), Remote sensing of vegetation and flooding on magela creek floodplain (northern territory, australia) with the sir-c synthetic aperture radar, *in* 'Aquatic Biodiversity', Springer, pp. 65–82.

Hill, D., Tarasoff, C., Whitworth, G., Baron, J., Bradshaw, J. & Church, J. (2017), 'Utility of unmanned aerial vehicles for mapping invasive plant species: a case study on yellow flag iris (iris pseudacorus l', *International Journal of Remote Sensing* **38**(8-10), 2083–2105.

Hill, M., J.O., M., Roy, Bunce, D. & R.G.H. (1999), 'Technical annex -ellenberg's indicator values for british plants', *Centre for Ecology and Hydrology, Natural Environment Research Council* .

Hirano, A., Madden, M. & Welch, R. (2003), 'Hyperspectral image data for mapping wetland vegetation', *Wetlands* **23**(2), 436–448.

Hird, J. N., DeLancey, E. R., McDermid, G. J. & Kariyeva, J. (2017), 'Google earth engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping', *Remote sensing* **9**(12), 1315.

Ho, P.-G. (2009), *Geoscience and remote sensing*, BoD–Books on Demand.

Ho, T. K. (1998), Nearest neighbors in random subspaces, *in* A. Amin, D. Dori, P. Pudil & H. Freeman, eds, 'Advances in Pattern Recognition', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 640–648.

Hoeser, T. & Kuenzer, C. (2020), 'Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends', *Remote Sensing* **12**(10), 1667.

Hoffer, R. M. & Johannsen, C. J. (1969), 'Ecological potentials in spectral signature analysis', *Remote sensing in ecology* pp. 1–16.

Horowitz, J. L. (1998), 'Bootstrap methods for median regression models', *Econometrica* pp. 1327–1351.

Hosmer Jr, D. W., Lemeshow, S. & Sturdivant, R. X. (2013), *Applied logistic regression*, Vol. 398, John Wiley & Sons.

Houborg, R., Fisher, J. B. & Skidmore, A. K. (2015), 'Advances in remote sensing of vegetation function and traits'.

Hu, T., Liu, J., Zheng, G., Zhang, D. & Huang, K. (2020), 'Evaluation of historical and future wetland degradation using remote sensing imagery and land use modeling', *Land Degradation & Development* **31**(1), 65–80.

Immitzer, M., Vuolo, F. & Atzberger, C. (2016), 'First experience with sentinel-2 data for crop and tree species classifications in central europe', *Remote Sensing* **8**(3), 166.

Ireland's Wetlands (2000). Accessed 5 May 2018.

Ishida, T., Kurihara, J., Viray, F. A., Namuco, S. B., Paringit, E. C., Perez, G. J., Takahashi, Y. & Marciano Jr, J. J. (2018), 'A novel approach for vegetation classification using uav-based hyperspectral imaging', *Computers and electronics in agriculture* **144**, 80–85.

Jackson, R. D. (1983), 'Spectral indices in n-space', *Remote sensing of environment* **13**(5), 409–421.

JACKSON, T. J. & RAGAN, R. M. (1975), 'Bayesian decision theory and remote sensing', *Photogrammetric Engineering and Remote Sensing* **41**(9).

Jaenicke, J., Rieley, J., Mott, C., Kimman, P. & Siegert, F. (2008), 'Determination of the amount of carbon stored in indonesian peatlands', *Geoderma* **147**(3-4), 151–158.

Japan Aerospace Exploration Agency (2020), `https://global.jaxa.jp/`. Accessed: 20 August 2020.

Jiang, J., Liu, F., Xu, Y. & Huang, H. (2019), 'Multi-spectral rgb-nir image classification using double-channel cnn', *IEEE Access* **7**, 20607–20613.

Jimenez, L. O., Morales-Morell, A. & Creus, A. (1999), 'Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks', *IEEE Transactions on Geoscience and Remote Sensing* **37**(3), 1360–1366.

Jones, M. C., Grosse, G., Jones, B. M. & Walter Anthony, K. (2012), 'Peat accumulation in drained thermokarst lake basins in continuous, ice-rich permafrost, northern seward peninsula, alaska', *Journal of Geophysical Research: Biogeosciences* **117**(G2).

Joosten, H. & Clarke, D. (2002), 'Wise use of mires and peatlands', *International Mire Conservation Group and International Peat Society* **304**.

Jozdani, S. E., Johnson, B. A. & Chen, D. (2019), 'Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification', *Remote Sensing* **11**(14), 1713.

Junk, W. J., An, S., Finlayson, C., Gopal, B., Květ, J., Mitchell, S. A., Mitsch, W. J. & Robarts, R. D. (2013), 'Current state of knowledge regarding the world's wetlands and their future under global climate change: a synthesis', *Aquatic sciences* **75**(1), 151–167.

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H. & Arriaga-Weiss, S. (2010), 'Classification in conservation biology: a comparison of five machine-learning methods', *Ecological Informatics* **5**(6), 441–450.

Kaplan, G. & Avdan, U. (2017), 'Mapping and monitoring wetlands using sentinel-2 satellite imagery.', *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* **4**.

Kaplan, G. & Avdan, U. (2018), *Sentinel-1 and Sentinel-2 data fusion for wetlands mapping: Balikdami, Turkey*, International Society for Photogrammetry and Remote Sensing.

Kathiresan, K. & Bingham, B. L. (2001), 'Biology of mangroves and mangrove ecosystems', *Advances in marine biology* **40**, 84–254.

Kattenborn, T., Lopatin, J., Förster, M., Braun, A. & Fassnacht, F. (2019), 'Uav data as alternative to field sampling to map woody invasive species based on combined sentinel-1 and sentinel-2 data', *Remote sensing of environment* **227**, 61–73.

Kentsch, S., Lopez Caceres, M., Serrano, D., Roure, F. & Diez, Y. (2020), 'Computer vision and deep learning techniques for the analysis of drone-acquired forest images, a transfer learning study', *Remote Sensing* **12**(8), 1287.

Kilroy, G., Dunne, F., Ryan, J., O'Connor, A., Daly, D., Craig, M., Coxon, C., Johnston, P. & Moe, H. (2008), 'A framework for the assessment of groundwater-dependent terrestrial ecosystems under the water framework directive', *Environmental Research Centre Report Series* (12).

Kim, S.-W., Wdowinski, S., Amelung, F., Dixon, T. H. & Won, J.-S. (2013), 'Interferometric coherence analysis of the everglades wetlands, south florida', *IEEE transactions on geoscience and remote sensing* **51**(12), 5210–5224.

Kimberley, S. & Coxon, C. (2013), Evaluating the influence of groundwater pressures on groundwater-dependent wetlands, Strive report. [online] environmental protection agency. Available at:.
URL: *https://www.epa.ie/pubs/reports/research/water/STRIVE$_1$00$_w$eb.pdf*

Kimmel, K. & Mander, Ü. (2010), 'Ecosystem services of peatlands: Implications for restoration', *Progress in Physical Geography* **34**(4), 491–514.

Kitchingman, A. (2017), 'Mapping pseudoraphis spinescens using remotely sensed imagery', *Unpublished Client Report for Goulburn Broken Catchment Management Authority. Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, Heidelberg, Victoria* .

K-means clustering (2018). Accessed: 20 August 2020.
URL: *https://stackoverflow.com/q/51263331*

Knoth, C., Klein, B., Prinz, T. & Kleinebecker, T. (2013), 'Unmanned aerial vehicles as innovative remote sensing platforms for high-resolution infrared imagery to support restoration monitoring in cut-over bogs', *Applied vegetation science* **16**(3), 509–517.

Koch, M., Koebsch, F., Hahn, J. & Jurasinski, G. (2017), 'From meadow to shallow lake: Monitoring secondary succession in a coastal fen after rewetting by flooding based on aerial imagery and plot data.', *Mires & Peat* **19**.

Kohv, M., Sepp, E. & Vammus, L. (2017), 'Assessing multitemporal water-level changes with uav-based photogrammetry', *The Photogrammetric Record* **32**(160), 424–442.

Kopel, D., Michalska-Hejduk, D., Berezowski, T., Borowski, M., Rosadzisski, S. & Chormasski, J. (2016), 'Application of multisensoral remote sensing data in the mapping of alkaline fens natura 2000 habitat', *Ecological Indicators* **70**, 196–208.

Korhonen, L., Packalen, P., Rautiainen, M. et al. (2017), 'Comparison of sentinel-2 and landsat 8 in the estimation of boreal forest canopy cover and leaf area index', *Remote sensing of environment* **195**, 259–274.

Kuenzer, C., Bluemel, A., Gebhardt, S., Quoc, T. & Dech, S. (2011), 'Remote sensing of mangrove ecosystems: A review', *Remote Sensing* **3**(5), 878–928.

L. Wiegand, C., J. Richardson, A., E. Escobar, D. & H. Gerbermann, A. (1991), 'Vegetation indices in crop assessments', **35**.

Laaksonen, J. & Oja, E. (1996), Classification with learning k-nearest neighbors, *in* 'Proceedings of International Conference on Neural Networks (ICNN'96', Vol. 3, IEEE, p. 1480–1483.

Labatut, V. & Cherifi, H. (2012), 'Accuracy measures for the comparison of classifiers'. arXiv preprint arXiv:1207.3790. arXiv:1207.3790v1.

Laliberte, A. & Rango, A. (2009), 'Texture and scale in object-based analysis of sub-decimeter resolution unmanned aerial vehicle (uav) imagery', *IEEE Transactions on Geoscience and Remote Sensing* **47**(3), 761–770.

Lambert, J., Compernolle, S., Eichmann, K., de Graaf, M., Hubert, D., Keppens, A., Kleipool, Q., Langerock, B., Sha, M., Verhoelst, T. et al. (2020), 'Quarterly validation report of the copernicus sentinel-5 precursor operational data products# 06: April 2018–february 2020., s5p mpc routine operations consolidated validation report series, issue# 06, version 06.0. 1, 154 pp'.

Landsat-8 (L8) Data Users Handbook (2016).

Lane, C. R., Liu, H., Autrey, B. C., Anenkhonov, O. A., Chepinoga, V. V. & Wu, Q. (2014), 'Improved wetland classification using eight-band high resolution satellite imagery and a hybrid approach', *Remote sensing* **6**(12), 12187–12216.

Langlois, M. N., Richardson, M. C. & Price, J. S. (2017), 'Delineation of peatland lagg boundaries from airborne lidar', *Journal of Geophysical Research: Biogeosciences* **122**(9), 2191–2205.

Lawrence, R. L. & Wright, A. (2001), 'Rule-based classification systems using classification and regression tree (cart) analysis', *Photogrammetric engineering and remote sensing* **67**(10), 1137–1142.

LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M. & Huang, F. (2006), 'A tutorial on energy-based learning', *Predicting structured data* **1**(0).

Lehmann, J. R., Münchberger, W., Knoth, C., Blodau, C., Nieberding, F., Prinz, T., Pancotto, V. A. & Kleinebecker, T. (2016), 'High-resolution classification of south patagonian peat bog microforms reveals potential gaps in up-scaled ch4 fluxes by use of unmanned aerial system (uas) and cir imagery', *Remote Sensing* **8**(3), 173.

Levick, S. R., Richards, A. E., Cook, G. D., Schatz, J., Guderle, M., Williams, R. J., Subedi, P., Trumbore, S. E. & Andersen, A. N. (2019), 'Rapid response of habitat structure and above-ground carbon storage to altered fire regimes in tropical savanna', *Biogeosciences* **16**(7), 1493–1503.

Li, H., Xu, D. & Zhao, Y. (2014), 'Peatland area change in the southern altay mountains over the last twenty years based on gis and rs analysis', *Frontiers of earth science* **8**(4), 558–563.

Li, J., Bioucas-Dias, J. M. & Plaza, A. (2011), 'Hyperspectral image segmentation using a new bayesian approach with active learning', *IEEE Transactions on Geoscience and Remote Sensing* **49**(10), 3947–3960.

Li, J. & Chen, W. (2005), 'A rule-based method for mapping canada's wetlands using optical, radar and dem data', *International Journal of Remote Sensing* **26**(22), 5051–5069.

Li, L., Chen, Y., Xu, T., Shi, K., Liu, R., Huang, C., Lu, B. & Meng, L. (2019), 'Remote sensing of wetland flooding at a sub-pixel scale based on random forests and spatial attraction models', *Remote Sensing* **11**(10), 1231.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. & Chen, M. (2014), Medical image classification with convolutional neural network, *in* '2014 13th international conference on control automation robotics vision (ICARCV', IEEE, p. 844–848.

Bibliography

Liaw, A. & Wiener, M. (2002), 'Classification and regression by randomforest', *R news* **2**(3), 18–22.

List of wetlands as a share of total land area (2020). Accessed: 20 August 2020.
   URL: *https://ec.europa.eu/eurostat*

Liu, H. & Huete, A. (1995), 'A feedback based modification of the ndvi to minimize canopy background and atmospheric noise', *IEEE transactions on geoscience and remote sensing* **33**(2), 457–465.

Liu, Y., Starzyk, J. & Zhu, Z. (2008), 'Optimised approximation algorithm in neural networks without overfitting', *IEEE transactions on neural networks* **19**(6), 983–995.

Löfstedt, T., Brynolfsson, P., Asklund, T., Nyholm, T. & Garpebring, A. (2019), 'Gray-level invariant haralick texture features', *PloS one* **14**(2), e0212110.

Longépé, N., Rakwatin, P., Isoguchi, O., Shimada, M., Uryu, Y. & Yulianto, K. (2011), 'Assessment of alos palsar 50 m orthorectified fbd data for regional land cover classification by support vector machines', *IEEE Transactions on Geoscience and Remote Sensing* **49**(6), 2135–2150.

Lu, D. & Weng, Q. (2007), 'A survey of image classification methods and techniques for improving classification performance', *International Journal of Remote Sensing* **28**(5), 823–870.

Lunt, M., Palmer, P. I., Feng, L., Landgraf, J. & Veefkind, P. (2019), 'Towards improved tropical african methane emission estimates using tropomi xch 4 data', *AGUFM* **2019**, B23D–02.

Mackin, F., Barr, A., Rath, P., Eakin, M., Ryan, J., Jeffrey, R. & Fernandez-Valverde, F. (2017), 'Best practice in raised bog restoration in ireland. irish wildlife manuals, no. 99', *Ireland: Wildlife Service, Department of Culture* **54**.

Maclin, R. & Opitz, D. W. (2011), 'Popular ensemble methods: An empirical study', *CoRR* **abs/1106.0257**.
   URL: *http://arxiv.org/abs/1106.0257*

MacQueen, J. (1967), 'Some methods for classification and analysis of multivariate observations', *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob* **1**, 281–297.

Mahdavi, S., Salehi, B., Granger, J., Amani, M., Brisco, B. & Huang, W. (2018), 'Remote sensing for wetland classification: A comprehensive review', *GIScience & Remote Sensing* **55**(5), 623–658.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S. & Gill, E. (2019), 'The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform', *Remote Sensing* **11**(1), 43.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F. & Motagh, M. (2017), 'Random forest wetland classification using alos-2 l-band, radarsat-2 c-band, and terrasar-x imagery', *ISPRS Journal of Photogrammetry and Remote Sensing* **130**, 13–31.

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F. & Zhang, Y. (2018), 'Very deep convolutional neural networks for complex land cover mapping using multi-spectral remote sensing imagery', *Remote Sensing* **10**(7), 1119.

Malenovskỳ, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R. & Berger, M. (2012), 'Sentinels for science: Potential of sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land', *Remote Sensing of environment* **120**, 91–101.

Mallick, S. (2002), *Feature based image mosaicing*, Department of Electrical and Computer Engineering, University of California, San Diego.

Maltby, E. & Acreman, M. (2011), 'Ecosystem services of wetlands: pathfinder for a new paradigm', *Hydrological Sciences Journal* **56**(8), 1341–1359.

Mandanici, E. & Bitelli, G. (2016), 'Preliminary comparison of sentinel-2 and landsat 8 imagery for a combined use', *Remote Sensing* **8**(12), 1014.

Maps and Data : National Parks Wildlife Service (2019).
    URL: *https://www.npws.ie/maps-and-data.*

Mardia, K. (1970), 'Measures of multivariate skewness and kurtosis with applications', *Biometrika* **57**(3), 519–530.

Markham, B. L., Storey, J. C., Williams, D. L. & Irons, J. R. (2004), 'Landsat sensor performance: history and current status', *IEEE Transactions on Geoscience and Remote Sensing* **42**(12), 2691–2694.

Martinez, J. & Le Toan, T. (2006), 'Mapping of flood dynamics and vegetation spatial distribution in the amazon floodplain using mutitemporal sar data', *Remote Sensing of Environment* .

Martinis, S. & Rieke, C. (2015), 'Backscatter analysis using multi-temporal and multi-frequency sar data in the context of flood mapping at river saale, germany', *Remote Sensing* **7**(6), 7732–7752.

Maruyama, S., Shebl, S., Abdel-Gawad, K., Kubota, A., Shimizu, K., Ishikawa, N. & Mohamed, A. (2017), Agricultural production—cultivation techniques and farming, *in* 'Irrigated Agriculture in Egypt', Springer, pp. 225–254.

Masci, J., Meier, U., Ciresan, D., Schmidhuber, J. & Fricout, G. (2012), Steel defect classification with max-pooling convolutional neural networks, *in* 'The 2012 International Joint Conference on Neural Networks (IJCNN', IEEE, p. 1–6.

MathWorks, T. & Inc (2019), 'lazysnapping'. Available at:.
    URL: *https://uk.mathworks.com/help/images/ref/lazysnapping.html.*

MATLAB, The MathWorks (2019). Natick, Massachusetts, United States.

Matlab Wrapper for Graph Cut (2006).
    URL: *Shai Bagon. https://github.com/shaibagon/GCMex*

McCarthy, M. J. & Halls, J. N. (2014), 'Habitat mapping and change assessment of coastal environments: an examination of worldview-2, quickbird, and ikonos satellite imagery and airborne lidar for mapping barrier island habitats', *ISPRS International Journal of Geo-Information* **3**(1), 297–325.

McCormack, T., Gill, L., Naughton, O. & Johnston, P. (2014), 'Quantification of submarine/intertidal groundwater discharge and nutrient loading from a lowland karst catchment', *Journal of Hydrology* **519, Part B**, 2318–2330.

McCormack, T., Naughton, O., Johnston, P. & Gill, L. (2016), 'Quantifying the influence of surface water–groundwater interaction on nutrient flux in a lowland karst catchment', *Hydrology and Earth System Sciences* **20**(5), 2119.

McCormack, T., Naughton, O. et al. (2020), Remote sensing for monitoring and mapping karst groundwater flooding in the republic of ireland, *in* 'EGU General Assembly Conference Abstracts', p. 18921.

McGovern, E. A., Holden, N. M., Ward, S. M. & Collins, J. F. (2000), 'Remotely sensed satellite imagery as an information source for industrial peatlands management', *Resources, conservation and recycling* **28**(1-2), 67–83.

Melton, J., Wania, R., Hodson, E. l., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C., Beerling, D., Chen, G. et al. (2013), 'Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (wetchimp)'.

Meng, X., Shang, N., Zhang, X., Li, C., Zhao, K., Qiu, X. & Weeks, E. (2017), 'Photogrammetric uav mapping of terrain under dense coastal vegetation: An object-oriented classification ensemble algorithm for classification and terrain correction', *Remote Sensing* **9**, 1187.

Merchant, M. A., Adams, J. R., Berg, A. A., Baltzer, J. L., Quinton, W. L. & Chasmer, L. E. (2016), 'Contributions of c-band sar data and polarimetric decompositions to subarctic boreal peatland mapping', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10**(4), 1467–1482.

Millard, K. & Richardson, M. (2015), 'On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping', *Remote sensing* **7**(7), 8489–8515.
URL: *https://doi.org/10.3390/rs70708489*

Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., de Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H., Malone, B. et al. (2019), 'Digital mapping of peatlands–a critical review', *Earth-Science Reviews* **196**, 102870.

Minasny, B., Setiawan, B. I., Saptomo, S. K., McBratney, A. B. et al. (2018), 'Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands', *Geoderma* **313**, 25–40.

Mitsch, W. & Gosselink, J. (2015), 'Wetlands. 5th'.

Miyamoto, H., Momose, A. & Iwami, S. (2018), Uav image classification of a riverine landscape by using machine learning techniques, *in* 'EGU General Assembly Conference Abstracts', Vol. 20, p. 5919.

Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012), *Introduction to linear regression analysis*, Vol. 821, John Wiley & Sons.

Montgomery, J., Brisco, B., Chasmer, L., Devito, K., Cobbaert, D. & Hopkinson, C. (2019), 'Sar and lidar temporal data fusion approaches to boreal wetland ecosystem monitoring', *Remote Sensing* **11**(2), 161.

Montoya-Zegarra, J., Wegner, J. & Schindler, K. (2015), 'Semantic segmentation of aerial images in urban areas with class-specific higher-order cliques', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2**, 127–133.

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I. & Papathanassiou, K. P. (2013), 'A tutorial on synthetic aperture radar', *IEEE Geoscience and remote sensing magazine* **1**(1), 6–43.

Moroni, M., Porti, M. & Piro, P. (2019), 'Design of a remote-controlled platform for green roof plants monitoring via hyperspectral sensors', *Water* **11**(7), 1368.

Morrissey, P. J., McCormack, T., Naughton, O., Johnston, P. M. & Gill, L. W. (2020), 'Modelling groundwater flooding in a lowland karst catchment', *Journal of Hydrology* **580**, 124361.

Murphy, K. P. (2012), *Machine learning: a probabilistic perspective*, MIT press.

Murray, N., Keith, D., Bland, L., Ferrari, R., Lyons, M., Lucas, R. & Nicholson, E. (2018), 'The role of satellite remote sensing in structured ecosystem risk assessments', *Science of the Total Environment* **619**, 249–257.

Murugan, D., Garg, A. & Singh, D. (2017), 'Development of an adaptive approach for precision agriculture monitoring with drone and satellite data', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10**(12), 5322–5328.

Naughton, O., Johnston, P. & Gill, L. (2012), 'Groundwater flooding in irish karst: the hydrological characterisation of ephemeral lakes (turloughs', *Journal of Hydrology* **470**, 82–97.

Naughton, O., Johnston, P., McCormack, T. & Gill, L. (2017), 'Groundwater flood risk mapping and management: examples from a lowland karst catchment in ireland', *Journal of Flood Risk Management* **10**(1), 53–64.

Naughton, O., McCormack, T., Drew, D., Gill, L., Johnston, P., Morrissey, P. & Regan, S. (2018), 'The hydrogeology of the gort lowlands', *Irish Journal of Earth Sciences* **36**, 25–44.

Nex, F. & Remondino, F. (2014), 'Uav for 3d mapping applications: a review', *Applied geomatics* **6**(1), 1–15.

Ng, C., Toh, Y., Lam, C., Chang, C. & Liew, S. (2007), Effects of leaf water content on reflectance, *in* '28th Asian Conference on Remote Sensing', p. 631–636.

Niculescu, S., Lardeux, C., Grigoras, I., Hanganu, J. & David, L. (2016), 'Synergy between lidar, radarsat-2, and spot-5 images for the detection and mapping of wetland vegetation in the danube delta', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9**(8), 3651–3666.

Nigam, I., Huang, C. & Ramanan, D. (2018), Ensemble knowledge transfer for semantic segmentation, *in* '2018 IEEE Winter Conference on Applications of Computer Vision (WACV', IEEE, p. 1499–1508.

Niphadkar, M., Nagendra, H., Tarantino, C., Adamo, M. & Blonda, P. (2017), 'Comparing pixel and object-based approaches to map an understorey invasive shrub in tropical mixed forests', *Frontiers in plant science* **8**, 892.

Nitze, I., Barrett, B. & Cawkwell, F. (2015), 'Temporal optimisation of image acquisition for land cover classification with random forest and modis time-series', *International Journal of Applied Earth Observation and Geoinformation* **34**, 136–146.

O'Connell, J. (2012), Monitoring Changes to Irish Peatlands Using Satellite Remote Sensing, PhD thesis, University College Dublin.

O'Connell, J., Connolly, J., Vermote, E. F. & Holden, N. M. (2013), 'Radiometric normalization for change detection in peatlands: a modified temporal invariant cluster approach', *International Journal of Remote Sensing* **34**(8), 2905–2924.

Olivier, R. & Hanqiang, C. (2012), 'Nearest neighbor value interpolation', *Int. J. Adv. Comput. Sci. Appl* **3**(4), 25–30.

Oostdijk, M., Santos, M., Whigham, D., Verhoeven, J. & Silvestri, S. (2018), 'Assessing rehabilitation of managed mangrove ecosystems using high resolution remote sensing', *Estuarine, Coastal and Shelf Science* **211**, 238–247.

Ouarzeddine, M., Belhdj Aissa, A., Souissi, B., Belkhider, M. & Boulahbal, S. (2005), Polarimetric classification using the cloude/pottier decomposition, *in* 'ESA Special Publication', Vol. 586, p. 16.

Ozesmi, S. L. & Bauer, M. E. (2002), 'Satellite remote sensing of wetlands', *Wetlands ecology and management* **10**(5), 381–402.

Paine, T., Jin, H., Yang, J., Lin, Z. & Huang, T. (2013), 'Gpu asynchronous stochastic gradient descent to speed up neural network training'. arXiv preprint arXiv:1312.6186.

Pengra, B., Johnston, C. & Loveland, T. (2007), 'Mapping an invasive plant, phragmites australis, in coastal wetlands using the eo-1 hyperion hyperspectral sensor', *Remote Sensing of Environment* **108**(1), 74–81.

Pitie, F., Kokaram, A. & Dahyot, R. (2005), N-dimensional probability density function transfer and its application to color transfer, *in* 'Tenth IEEE International Conference on Computer Vision (ICCV'05', Vol. 1 (Vol. 2, IEEE, p. 1434–1439.

Pitié, F. (2021), 'Lecture notes on deep learning and its applications'.
URL: *https://frcs.github.io/4C16-LectureNotes/index.html*

Pittock, J., Finlayson, M., Arthington, A., Roux, D., Matthews, J. H., Biggs, H., Harrison, I., Blom, E., Flitcroft, R., Froend, R. et al. (2015), 'Managing freshwater, river, wetland and estuarine protected areas', *Protected area governance and management* pp. 569–608.

Plant List (2013), Version 1.1. Published on the internet. Accessed: 20 August 2020.

Polder, G., Pekkeriet, E. & Snikkers, M. (2013), 'A spectral imaging system for detection of botrytis in greenhouses'.

Pollock, M., Naiman, R. & Hanley, T. (1998), 'Plant species richness in riparian wetlands - a test of biodiversity theory', *Ecology* **79**, 94–105.

Poulter, B., Bousquet, P., Canadell, J. G., Ciais, P., Peregon, A., Saunois, M., Arora, V. K., Beerling, D. J., Brovkin, V., Jones, C. D. et al. (2017), 'Global wetland contribution to 2000–2012 atmospheric methane growth rate dynamics', *Environmental Research Letters* **12**(9), 094013.

Prather, M. J., Holmes, C. D. & Hsu, J. (2012), 'Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry', *Geophysical Research Letters* **39**(9).

Pu, R. & Landry, S. (2012), 'A comparative analysis of high spatial resolution ikonos and worldview-2 imagery for mapping urban tree species', *Remote Sensing of Environment* **124**, 516–533.

Python Language Reference, version 3.7 (2020).
URL: *http://www.python.org*

Qassim, H., Verma, A. & Feinzimer, D. (2018), Compressed residual-vgg16 cnn model for big data places image recognition, *in* '2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC', IEEE, p. 169–175.

QGIS (2020), 'Qgis geographic information system'.
URL: *http://qgis.org*

Qian, Y., Zhou, W., Yan, J., Li, W. & Han, L. (2015), 'Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery', *Remote Sensing* **7**(1), 153–168.

Rahman, M. M., McDermid, G. J., Strack, M. & Lovitt, J. (2017), 'A new method to map groundwater table in peatlands using unmanned aerial vehicles', *Remote Sensing* **9**(10), 1057.

Ramiya, A. M., Nidamanuri, R. R. & Krishnan, R. (2017), 'Segmentation based building detection approach from lidar point cloud', *The Egyptian Journal of Remote Sensing and Space Science* **20**(1), 71–77.

Rapinel, S., Fabre, E., Dufour, S., Arvor, D., Mony, C. & Hubert-Moy, L. (2019), 'Mapping potential, existing and efficient wetlands using free remote sensing data', *Journal of environmental management* **247**, 829–839.

Real, R. & Vargas, J. (1996), 'The probabilistic basis of jaccard's index of similarity', *Systematic biology* **45**(3), 380–385.

Regan, S., Flynn, R., Gill, L., Naughton, O. & Johnston, P. (2019), 'Impacts of groundwater drainage on peatland subsidence and its ecological implications on an atlantic raised bog', *Water Resources Research* .

Regmi, P., Grosse, G., Jones, M. C., Jones, B. M. & Anthony, K. W. (2012), 'Characterizing post-drainage succession in thermokarst lake basins on the seward peninsula, alaska with terrasar-x backscatter and landsat-based ndvi data', *Remote Sensing* **4**(12), 3741–3765.

Reschke, J., Bartsch, A., Schlaffer, S. & Schepaschenko, D. (2012), 'Capability of c-band sar for operational wetland monitoring at high latitudes', *Remote Sensing* **4**(10), 2923–2943.

Rish, I. (2001), 'An empirical study of the naive bayes classifier', *IJCAI 2001 workshop on empirical methods in artificial intelligence* **3**(22), 41–46.

Rodwell, J. S. & nature conservation committee (GB), J. (2006), *National vegetation classification: users' handbook*, Joint nature conservation committee Peterborough.

Ronneberger, O., Fischer, P. & Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, *in* 'International Conference on Medical image computing and computer-assisted intervention', Springer, Cham, p. 234–241.

Ross, Q. (1993), 'C4. 5: programs for machine learning'.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. & Berg, A. (2015), 'Imagenet large scale visual recognition challenge', *International journal of computer vision* **115**(3), 211–252.

Ruwaimana, M., Satyanarayana, B., Otero, V. & Muslim, A. (2018), 'The advantages of using drones over space-borne imagery in the mapping of mangrove forests', *PloS one* **13**(7).

Räsänen, A. & Virtanen, T. (2019), 'Data and resolution requirements in mapping vegetation in spatially heterogeneous landscapes', *Remote Sensing of Environment* **230**, 111207.

Salehi, B., Zhang, Y., Zhong, M. & Dey, V. (2012), 'Object-based classification of urban areas using vhr imagery and height points ancillary data', *Remote Sensing* **4**(8), 2256–2276.

Salem, Y. & Nasri, S. (2009), Texture classification of woven fabric based on a glcm method and using multiclass support vector machine, *in* '2009 6th International Multi-Conference on Systems, Signals and Devices', IEEE, p. 1–8.

Salinas, J., Eggerth, M., Miller, M., Meza, R., Chacaltana, J., Acuña, J. & Barroso, G. (2020), Wetland mapping with multitemporal sentinel radar remote sensing in the southeast region of brazil, *in* '2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)', IEEE, pp. 669–674.

Sangeda, A. Z., Kahimba, F. C., Kashaga, R. A., Semu, E., Mahonge, C. P., Mkanda, F. X. et al. (2014), 'Testing of decision making tools for village land use planning and natural resources management in kilimanjaro region', *Open Journal of Soil Science* **4**(13), 446.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S. et al. (2016), 'The global methane budget 2000–2012', *Earth System Science Data* **8**(2), 697–751.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K. et al. (2020), 'The global methane budget 2000–2017', *Earth System Science Data* **12**(3), 1561–1623.

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J. & Nickless, D. (2004), 'Terra and aqua modis products available from nasa ges daac', *Advances in Space Research* **34**(4), 710–714.

Schaaf, S. & Streefkerk, J. (2002), Relationships between biotic and abiotic conditions, *in* 'Conservation and restoration of raised bogs', Dept. of the Env. and Local Government; Staatsbosbeheer, p. 186–209.

Schwenker, F., Abbas, H. & El Gayar, N. (2016), trentin, *in* E., ed., 'Artificial Neural Networks in Pattern Recognition: 7th IAPR TC3 Workshop, ANNPR 2016', Vol. Proceedings (Vol. 9896, Springer, Ulm, Germany.

Sha, M. & Langerock, B. (2019), 'S5p mpc vdaf and tccon4s5p ch4 validation results, presentation at the sentinel-5p third products release workshop'.

Shamshiri, R. R., Hameed, I. A., Balasundram, S. K., Ahmad, D., Weltzien, C. & Yamin, M. (2018), 'Fundamental research on unmanned aerial vehicles to support precision agriculture in oil palm plantations', *Agricultural Robots-Fundamentals and Application* .

Shi, J. & Malik, J. (2000), 'Normalised cuts and image segmentation', *IEEE Transactions on pattern analysis and machine intelligence* **22**(8), 888–905.

Simkin, J., Ojala, A. & Tyrväinen, L. (2020), 'Restorative effects of mature and young commercial forests, pristine old-growth forest and urban recreation forest-a field experiment', *Urban Forestry & Urban Greening* **48**, 126567.

Simonyan, K. & Zisserman, A. (2014), 'Very deep convolutional networks for large-scale image recognition'. arXiv preprint arXiv:1409.1556.

Skakun, S., Vermote, E., Roger, J.-C. & Franch, B. (2017), 'Combined use of landsat-8 and sentinel-2a images for winter crop mapping and winter wheat yield assessment at regional scale', *AIMS geosciences* **3**(2), 163.

Skeeter, J., Christen, A., Laforce, A.-A., Humphreys, E. & Henry, G. (2020), 'Vegetation influence and environmental controls on greenhouse gas fluxes from a drained thermokarst lake in the western canadian arctic', *Biogeosciences Discussions* pp. 1–25.

Skeffington, M. S., Moran, J., Connor, Á. O., Regan, E., Coxon, C., Scott, N. & Gormally, M. (2006), 'Turloughs–ireland's unique wetland habitat', *Biological Conservation* **133**(3), 265–290.

Slagter, B., Tsendbazar, N.-E., Vollrath, A. & Reiche, J. (2020), 'Mapping wetland characteristics using temporally dense sentinel-1 and sentinel-2 data: A case study in the st. lucia wetlands, south africa', *International Journal of Applied Earth Observation and Geoinformation* **86**, 102009.

SNAP - ESA Sentinel Application Platform v6.0.5 (2018).
URL: *http://step.esa.int*

Stoer, M. & Wagner, F. (1997), 'A simple min-cut algorithm', *Journal of the ACM (JACM)* **44**(4), 585–591.

Sun, Y., Huang, J., Ao, Z., Lao, D. & Xin, Q. (2019), 'Deep learning approaches for the mapping of tree species diversity in a tropical wetland using airborne lidar and high-spatial-resolution remote sensing images', *Forests* **10**(11), 1047.

Szantoi, Z., Smith, S., Strona, G., Koh, L. & Wich, S. (2017), 'Mapping orangutan habitat and agricultural areas using landsat oli imagery augmented with unmanned aircraft system aerial photography', *International journal of remote sensing* **38**(8-10), 2231–2245.

Szporak-Wasilewska, S., Mirosław-Świątek, D., Grygoruk, M., Michałowski, R. & Kardel, I. (2015), Processing of airborne laser scanning data to generate accurate dtm for floodplain wetland, *in* 'Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII', Vol. 9637, International Society for Optics and Photonics, p. 963720.

Tavares, J. & Jorge, R. (2015), Computational vision and medical image processing v, *in* 'Proceedings of the 5th Eccomas Thematic Conference on Computational Vision and Medical Image Processing (VipIMAGE 2015', CRC Press, Tenerife, Spain.

Thoma, M. (2017), 'Ensembles'.
URL: *https://martin-thoma.com/ensembles/*

Thompson, J. & Finlayson, C. (2001), Freshwater wetlands, *in* A. Warren & J. French, eds, 'Habitat Conservation: Managing the Physical Environment', John Wiley and Sons Ltd, Chichester.

Touzi, R., Deschamps, A. & Rother, G. (2009), 'Phase of target scattering for wetland characterization using polarimetric c-band sar', *IEEE Transactions on Geoscience and Remote Sensing* **47**(9), 3241–3261.

Touzi, R., Omari, K. & Sleep, B. (2015), Assessment of fire peatland damages using polarimetric l-band alos, *in* 'Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International', IEEE, pp. 1195–1198.

United States Geological Survey (2020), `https://www.usgs.gov/`. Accessed: 20 August 2020.

Valentine, D. W., Holland, E. A. & Schimel, D. S. (1994), 'Ecosystem and physiological controls over methane production in northern wetlands', *Journal of Geophysical Research: Atmospheres* **99**(D1), 1563–1571.

Valero Medina, J. A. & Alzate Atehortúa, B. E. (2019), 'Comparison of maximum likelihood, support vector machines, and random forest techniques in satellite images classification', *Tecnura* **23**(59), 3–10.

Van Den Doel, K., Ascher, U. & Haber, E. (2013), 'The lost honour of l2-based regularisation', *Radon Series in Computational and Applied Math* .

Van Deventer, H., Cho, M. & Mutanga, O. (2019), 'Multi-season rapideye imagery improves the classification of wetland and dryland communities in a subtropical coastal region', *ISPRS Journal of Photogrammetry and Remote Sensing* **157**, 171–187.

Van Diggelen, R., Molenaar, W. & Kooijman, A. (1996), 'Vegetation succession in a floating mire in relation to management and hydrology', *Journal of Vegetation Science* **7**(6), 809–820.

Van Riel, S. (2016), 'Exploring the use of 3d gis as an analytical tool in archaeological excavation practice'.

Veefkind, J., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H., De Haan, J., Kleipool, Q. et al. (2012), 'Tropomi on the esa sentinel-5 precursor: A gmes mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications', *Remote Sensing of Environment* **120**, 70–83.

Veksler, O. & Zabih, R. (1999), 'Efficient graph-based energy minimization methods in computer vision'.

Vincent, L. (1994), Morphological area openings and closings for grey-scale images, *in* 'Shape in Picture', Springer, Berlin, Heidelberg, p. 197–208.

Waldren, S., Gill, L., Allott, N., C. C., Irvine, K. & Johnston, P. (2015), Turlough hydrology, ecology and conservation, Unpublished report,, National Parks Wildlife Services. Department of Arts, Heritage and the Gaeltacht, Dublin, Ireland.

Wang, L., Jia, M., Yin, D. & Tian, J. (2019b), 'A review of remote sensing for mangrove forests: 1956–2018', *Remote Sensing of Environment* **231**, 111223.

Wang, X., Gao, X., Zhang, Y., Fei, X., Chen, Z., Wang, J., Zhang, Y., Lu, X. & Zhao, H. (2019a), 'Land-cover classification of coastal wetlands using the rf algorithm for worldview-2 and landsat 8 images', *Remote Sensing* **11**(16), 1927.

Wania, R., Melton, J., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C., Chen, G., Eliseev, A. V. et al. (2013), 'Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (wetchimp)', *Geoscientific Model Development* **6**(3), 617–641.

Wania, R., Ross, I. & Prentice, I. (2010), 'Implementation and evaluation of a new methane model within a dynamic global vegetation model: Lpj-whyme v1. 3.1', *Geoscientific Model Development* **3**(2), 565.

Warner, B. & Rubec, C. (1997), 'The canadian wetland classification system', *Wetlands Research Centre, University of Waterloo, Waterloo, Ontario* .

Weiss, D. J. & Crabtree, R. L. (2011), 'Percent surface water estimation from modis brdf 16-day image composites', *Remote sensing of environment* **115**(8), 2035–2046.

Wen, L. & Hughes, M. (2020), 'Coastal wetland mapping using ensemble learning algorithms: A comparative study of bagging, boosting and stacking techniques', *Remote Sensing* **12**(10), 1683.

Wetland Survey Ireland (2020). Accessed: 20 August 2020.
URL: *http://www.wetlandsurveysireland.com/*

Whalen, S. (2005), 'Biogeochemistry of methane exchange between natural wetlands and the atmosphere', *Environmental Engineering Science* **22**(1), 73–94.

Whelley, P. L., Glaze, L. S., Calder, E. S. & Harding, D. J. (2014), 'Lidar-derived surface roughness texture mapping: Application to mount st. helens pumice plain deposit analysis', *IEEE Transactions on Geoscience and Remote Sensing* **52**(1), 426–438.

Whitcomb, J., Moghaddam, M., McDonald, K., Kellndorfer, J. & Podest, E. (2009), 'Mapping vegetated wetlands of alaska using l-band radar satellite imagery', *Canadian Journal of Remote Sensing* **35**(1), 54–72.

White, J. & Doyle, G. (1982), 'The vegetation of ireland-a catalog raisonne', *Journal of Life Sciences Royal Dublin Society* **3**(2), 289–368.

White, L., Millard, K., Banks, S., Richardson, M., Pasher, J. & Duffe, J. (2017), 'Moving to the radarsat constellation mission: Comparing synthesized compact polarimetry and dual polarimetry data with fully polarimetric radarsat-2 data for image classification of peatlands', *Remote Sensing* **9**(6), 573.

Whiteman, M., Wheeler, B., Shaw, S., Lewis, T., Grout, M. & Tanner, K. (2009), 'Use of wetmecs typology to aid understanding of groundwater-dependent terrestrial ecosystems in england and wales', *Groundwater Quality Assessment and Monitoring. John Wiley & Sons* pp. 259–272.

Whyte, A., Ferentinos, K. P. & Petropoulos, G. P. (2018), 'A new synergistic approach for monitoring wetlands using sentinels-1 and 2 data with object-based machine learning algorithms', *Environmental Modelling & Software* **104**, 40–54.

Wu, Q., Li, H.-q., Wang, R.-s., Paulussen, J., He, Y., Wang, M., Wang, B.-h. & Wang, Z. (2006), 'Monitoring and predicting land use change in beijing using remote sensing and gis', *Landscape and urban planning* **78**(4), 322–333.

Wu, Y., Zhou, Y., Saveriades, G., Agaian, S., Noonan, J. & Natarajan, P. (2013), 'Local shannon entropy measure with statistical tests for image randomness', *Information Sciences* **222**, 323–342.

Xu, T., Weng, B., Yan, D., Wang, K., Li, X., Bi, W., Li, M., Cheng, X. & Liu, Y. (2019), 'Wetlands of international importance: Status, threats, and future protection', *International Journal of Environmental Research and Public Health* **16**(10), 1818.

Yim, J., Joo, D., Bae, J. & Kim, J. (2017), A gift from knowledge distillation: Fast optimisation, network minimisation and transfer learning, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', p. 4133–4141.

Yuheng, S. & Hao, Y. (2017), 'Image segmentation algorithms overview', *CoRR* **abs/1707.02051**.
URL: *http://arxiv.org/abs/1707.02051*

Zedler, J. B. & Kercher, S. (2005), 'Wetland resources: status, trends, ecosystem services, and restorability', *Annu. Rev. Environ. Resour.* **30**, 39–74.

Zhang, A., Sun, G., Ma, P., Jia, X., Ren, J., Huang, H. & Zhang, X. (2019), 'Coastal wetland mapping with sentinel-2 msi imagery based on gravitational optimized multilayer perceptron and morphological attribute profiles', *Remote Sensing* **11**(8), 952.

Zhang, C., Wang, L. & Yang, R. (2010), Semantic segmentation of urban scenes using dense depth maps, *in* 'European Conference on Computer Vision', Springer, Berlin, Heidelberg, p. 708–721.

Zhang, W., Yi, Y., Song, K., Kimball, J. S. & Lu, Q. (2016), 'Hydrological response of alpine wetlands to climate warming in the eastern tibetan plateau', *Remote Sensing* **8**(4), 336.

Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. (2017), Pyramid scene parsing network, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', p. 2881–2890.

Zheng, H., Cheng, T., Li, D., Zhou, X., Yao, X., Tian, Y. & Zhu, Y. (2018), 'Evaluation of rgb, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice', *Remote Sensing* **10**(6), 824.

Zhong, Y., Zhang, L. & Gong, W. (2011), 'Unsupervised remote sensing image classification using an artificial immune network', *International journal of remote sensing* **32**(19), 5461–5483.

Zhu, X., Hou, Y., Weng, Q. & Chen, L. (2019), 'Integrating uav optical imagery and lidar data for assessing the spatial relationship between mangrove and inundation across a subtropical estuarine wetland', *ISPRS Journal of Photogrammetry and Remote Sensing* **149**, 146–156.

Åhlén, J. (2005), 'Colour correction of underwater images using spectral data'.

Bibliography

Chapter 9

Appendix

All the extra results which could not be included in the papers (and main body of the thesis) are given in this section.

Part I

Appendix A

9.1 Literature Review

The literature review (table) can be found using this url:
https://drive.google.com/file/d/1oCSOhSn8cBtotqs-hQQgKPl2mFuFBB31/view?usp=sharing
Please email at *sbhatnag@tcd.ie*, in case the link does not get open. Thank you.

Figure 9.1: List of reviewed literature with the key points of the studies

Part II

Appendix B

9.2 Satellite Image Segmentation: Wetlands

9.2.1 Boundary Delineation

The use of a single algorithm for boundary delineation leads to the formation of smaller, non-connected objects, and hence, the bog could not be appropriately delineated. Using the entropy filter, canny edge detection and lazy snapping in conjunction proved to be an effective way of delineating a complex structure from a middle-resolution image. The base image used has to be picked out in a way that maximum variance can be achieved between bog and non-bog areas. Other bands such as NIR, vegetation-red edge can also be used to get a similar delineation. Clara, Monivea and, Mongan bogs were delineated better due to their bigger sizes. The percentage overlap results indicate that for the smaller bogs, the boundary delineation was not very accurate due to the coarse resolution of the image. Figure 7.1 shows the wetlands, along with their delineated boundaries.

Bog sites	(a) Clara	(b) Mongan	(c) Monivea	(d) Knockacollar	(e) Killyconny
% overlap	77	89	80	72	74
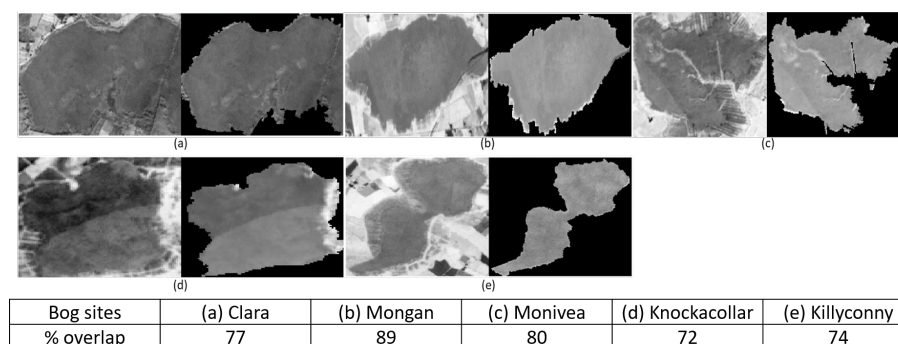
Figure 9.2: (1) Base image with (2) delineated image for (a) Clara bog (b) Mongan bog (c) Monivea bog (d) Knockacoller bog (e) Killyconny bog

9.2.2 Histogram Analysis

Histograms of the normalised intensity value of each image against the relative probability are mapped across the area of the bogs. The intensity value is normalised using the min-max normalisation technique. Figure 7.2 depicts the histogram of major ecotypes, averaged for all the ten bands given in Table 1, present in the 5 studied bogs. The ease by which the different ecotypes can be identified of ecotypes is dependent on their distribution (histograms). For normally distributed data, the classification is easier, as was found, for example, in Monivea bog (Figure 7.2(c)). Data in Monivea bog was closest to the normal distribution but was skewed: the central ecotype was positive off-centre exhibited a small positive skewness (0.3273) in comparison to marginal ecotype (-0.2472). In comparison, for Killyconny bog (figure 7.2(e)) the marginal (0.5275), sub-marginal (0.1934) ecotypes had a positive skew showing the difference in the distribution of the same ecotype in different bogs. Therefore, the portability of the algorithm from one bog to another was not feasible. A bi-modal histogram usually indicates a strong presence of two different groups, which is useful for binary-classifications. A histogram showing more than two peaks are difficult to detect using gaussian kernels. Multimodal histograms can result from the presence of mixed pixels towing to the coarse resolution of RS data (Figure 7.2 (a), (b), (d), and (e)). For example, the sub-central and sub-marginal ecotypes are spatially interconnected, which leads to an increment in mixed pixels and hence, their

histograms are similar and exhibit multiple modes. Outliers also lead to high kurtosis as can be seen from Clara (highest for marginal ecotype = 29.41), Mongan bog (highest for sub-central ecotype = 42.76) (Figure 7.2 (a3, a4, b3)). The peaks and troughs vary for every ecotype in every bog showing no pattern or similarity. This kind of distribution of the data is challenging for classification of the ecotypes present in the bog-wetlands.
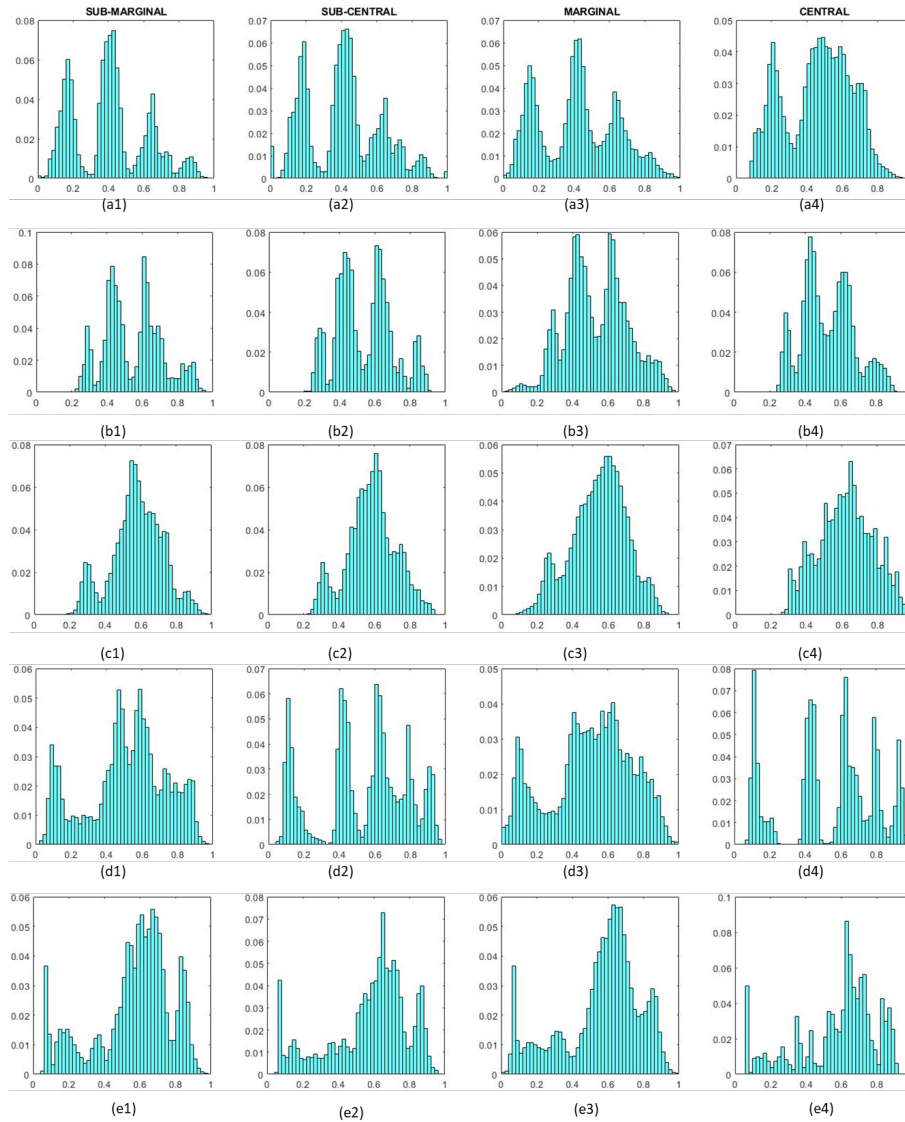
Figure 9.3: Histogram of Ecotopes - Raised Bogs

9.2.3 Spectral Knowledge Transfer

Due to the inaccessibility of many wetlands, there is a growing recognition that remote sensing techniques can be a viable and cost-effective alternative to field-based ecosystem monitoring. Wetlands encompass a diverse array of habitats, for example, fens, bogs, marshes, and swamps. In this study, we have concentrated on natural raised bogs, found in the Irish midlands. The objective of the study was to investigate the use of multispectral satellite imagery for delineating the extent of raised bogs and then monitoring their ecological composition in order to help with Ireland's obligations under the EU Habitats

Directive. This was carried out using open-source Sentinel-2 data. An initial study to delineate the boundary of the bogs using the combination of edge detection techniques was performed. Once the bog boundary was defined, the spectra from the delineated area were studied. Various vegetation indices along with soil moisture information and DEM were used as features to train the classification algorithm. An ensemble classifier Bagged Tree (BT) as a supervised pixel-based learner has been used for mapping the ecotopes. In order to acknowledge the spatial distribution of the ecotopes, we partitioned the bog into ecotope-objects via graph cut segmentation also known as MAP (maximum a posteriori) estimation. The results have been verified for 5 different bog-sites using field-derived ecotope maps which has shown that the addition of spatial knowledge enhances the overall accuracy. The study further extends to investigate the scope whether it is possible to transfer the knowledge contained in the classification algorithm from one bog-wetland to other wetlands. If this is possible it would mean that other bogs could be automatically and remotely mapped. Various bog-indices highlighting special areas in the bogs under consideration were created. For bogs located in close proximity, the knowledge transfer is carried out successfully for the active regions (i.e. the major peat forming areas) using a binary-linear discriminant. The study limits to closely located bogs due to unavoidable changes in environmental conditions which in turn changes the intensity values of ecotopes in far-away locations. Such a methodology that is just being developed may be able to significantly reduce the amount of field work required by ecologists on each wetland using freely available satellite data.

9.2.4 Spectral knowledge transfer - EGU presentation

CONCLUSIONS

SUPERVISED:

- Using 3 different edge detection techniques leads to better boundary delineation of complex areas.
- Using Bagged Tree classifier and a minimum amount of training data, we were able to map ecotopes in all 5 raised bogs.
- Integration of spatial information along with spectral is beneficial, object based (graphcut – MAP) classification increased overall accuracy to up to 95%.
- The algorithm developed is universal and all kinds of wetlands (bog, fen, turlough) can be mapped through it, given some supervision (ground truth required).

UNSUPERVISED:

- Binary classification is very effective and gives a correct estimation of ecotope location.
- Direct transfer of knowledge between bogs was possible using newly created *bog indices*.
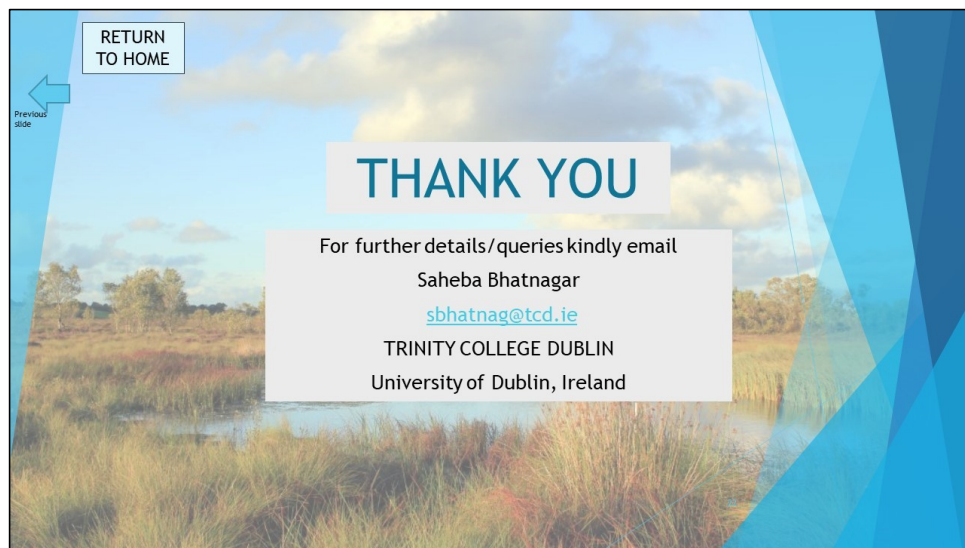
9.2.5 Comparison between spectral indices on Raised Bogs - winter and summer

This weather metric shows the value of correlation coefficient between spectral indices recorded using S2 data in February and in June for raised bogs. The correlation is very high, with some indices showing negative correlation (Figure 7.2.1). This is important, as these indices could be used to differentiate the seasonal data for the raised bogs.

	Correlation b/w JUNE and FEB 2018
BLUE (490nm)	0.8367
GREEN (560nm)	0.9098
Red (665nm)	0.9033
Veg Red Edge (705nm)	0.9661
Veg Red Edge (740nm)	0.9515
Veg Red Edge (783nm)	0.9468
NIR (842 nm)	0.9455
Narrow NIR (865 nm)	0.9549
SWIR-1 (1610 nm)	0.9736
SWIR-2 (2190 nm)	0.9664
NDVI	-0.4223
PVI	0.9414
SAVI	0.9326
tan2 (R, NIR)	0.9192
sin(G) * cos(NIR)	0.9131
(B4 − B6) /(B5 + B6)	0.9868
NDWI	0.6118
NDWI2	0.9062
tan(NDWI)	-0.7471
Tasselled Cap Vegetation	-0.6625
Overall	0.6999

Figure 9.4: Weather Metric

9.2.6 Graphic User Interface

A beta version of the boundary delineation and classifcation algorithm was made such that the ecologists at NPWS and EPA can use it as an interactive application. Please click on this url to see the video of the GUI `https://drive.google.com/file/d/0B9z0GU4XkCs_Z2duQlBYbnBqcmVZMDBDS1FCVHFMU2hKbmFR/view?usp=sharing`.

Additionally all the codes and GUI will be uploaded on github once the project is officially over.

9.2.7 MVC Algorithm on Wetlands

Classification maps, and confusion matrix of extra wetlands which were mapped using MVC algorithm (averaged for 2 years). The location and other information about these wetlands is given in Table 3.1.

1. Ballymore Fen

GROUND TRUTH   **MAP SEGMENTED IMAGE**

- SCHOENUS – CAREX FEN AND MENYANTHES POOL   2 CAREX – MENYANTHES TRANSITION MIRE
- 3 JUNCUS SUBNODULOSUS FEN   4 CALLUNA – ERIOPHORGUM BOG   5 CAREX- MEYANTHES TRANSITION MIRE AND FLIPENDULA - HOLCUS
- 6 SCRUB AND DRY GRASSLAND   7 FILIPENDULA- HOLCUS   8 ULEX SCRUB AND MOLINIA CUTAWAY

Averaged map of vegetation communities in Ballymore
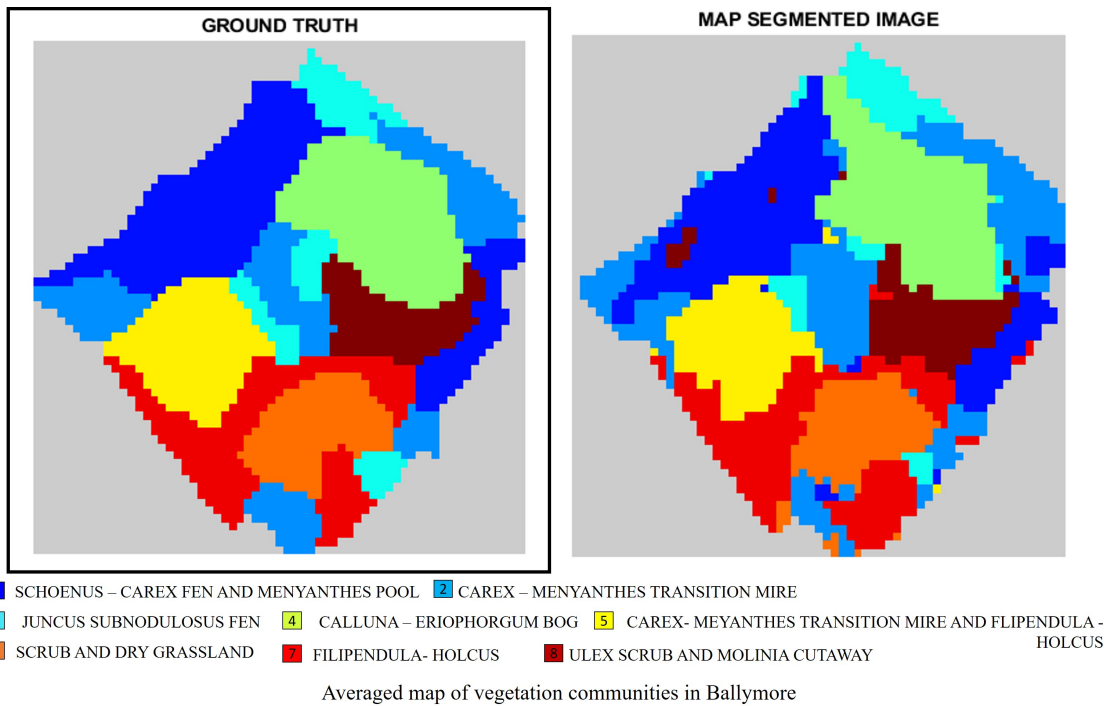
Figure 9.5: Ballymore Fen (a) GT (b) Classified Map

Averaged confusion matrix of Ballymore with
1. Schoenus-Carex fen and Menyanthes pool 2. Carex-Menyanthes transition mire 3. Juncus subnodulosus fen 4. Calluna-Eriophorum bog 5. Carex-Menyanthes transition mire and Filipendula-Holcus 6. Scrub and Dry grassland 7. Filipendula-Holcus 8. Ulex scrub and Molinia cutaway

	1	2	3	4	5	6	7	8
1	390	45	6	31	0	0	2	16
2	37	283	6	2	4	6	19	1
3	3	46	104	11	3	0	9	0
4	4	10	8	308	0	0	0	13
5	5	8	1	0	204	0	20	0
6	1	8	0	0	0	165	15	0
7	3	9	0	0	2	13	247	8
8	7	10	2	1	0	0	4	136
Precision	86.66	67.54	81.88	87.25	95.77	89.67	78.16	78.16
Sensitivity	79.59	79.05	59.09	89.79	85.71	87.3	87.58	85
OA	82.15							
Kappa	0.8							

Figure 9.6: Ballymore Fen: Confusion Matrix

2. Tory Hill Fen

GROUND TRUTH   MAP SEGMENTED IMAGE

1	FLIPENDULA FEN AND JUNCUS	2	FLIPENDULA TALL HERB FEN
3	FLIPENDULA AND MOLINIA FEN GRASSLAND	4	PHALARIS GRASSLAND	5	CLADIUM
6	SALIX WOODLAND	7	MOLINIA SCHOENUS FEN	8	IMPROVED GRASSLAND

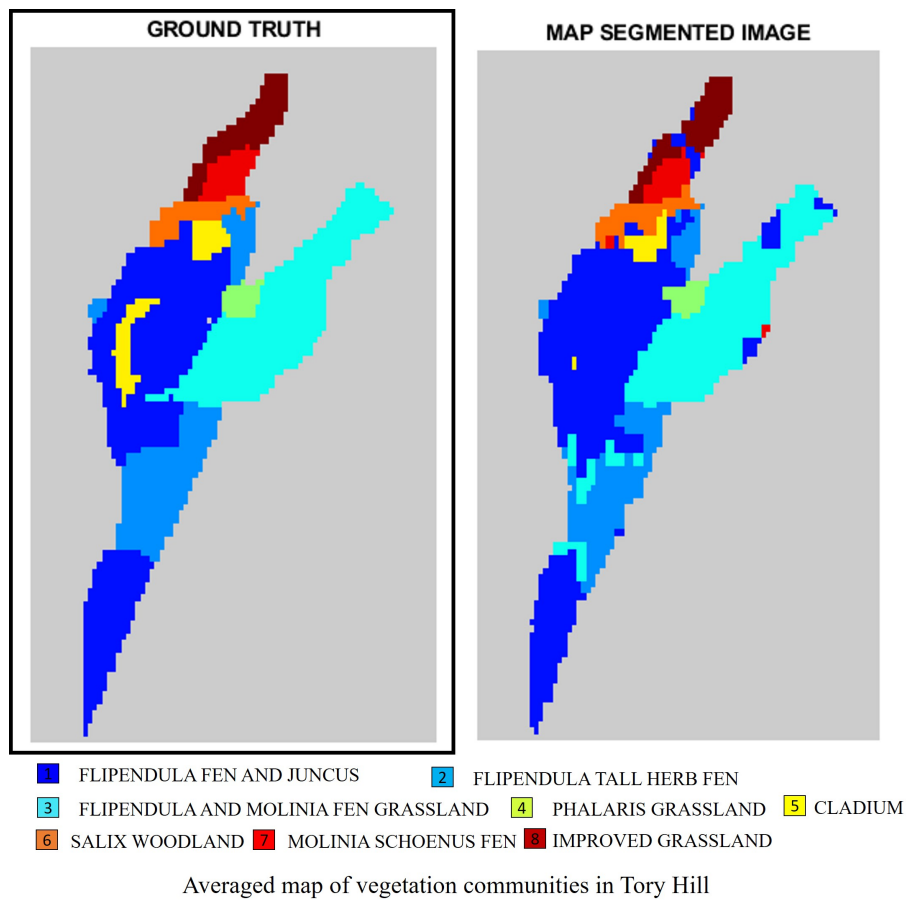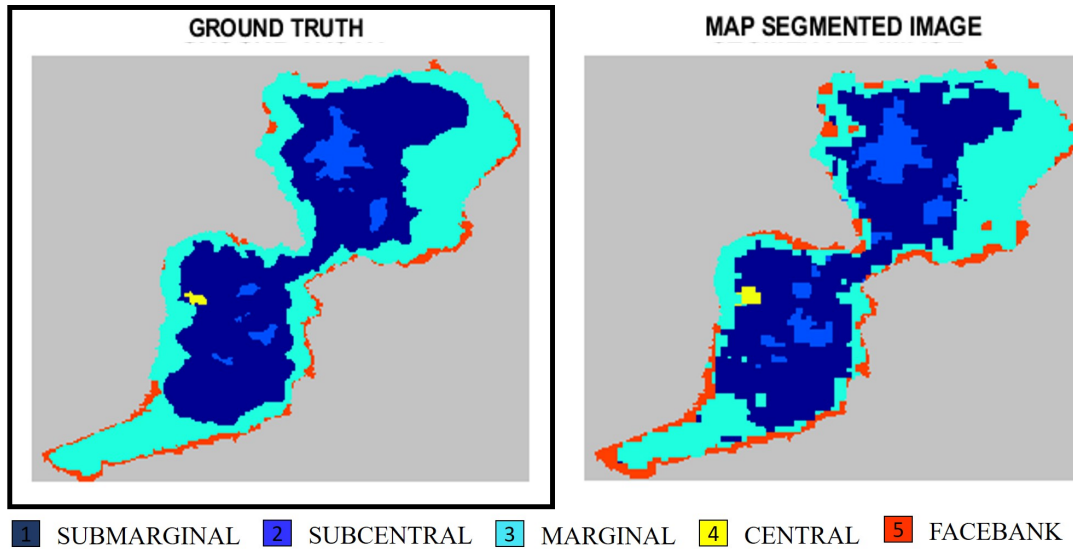Averaged map of vegetation communities in Tory Hill

Figure 9.7: Tory Hill Fen (a) GT (b) Classified Map

Averaged confusion matrix of Tory Hill with
1. Flipendula fen and Juncus 2. Flipendula tall herb fen 3. Flipendula and Molinia fen grassland 4. Phalaris grassland 5. Cladium 6. Salix woodland 7. Molinia – Schoenus fen 8. Improved grassland

	1	2	3	4	5	6	7	8
1	711	23	24	3	8	1	1	0
2	48	246	38	0	0	0	0	0
3	44	2	484	0	0	0	3	0
4	2	0	4	33	0	0	0	0
5	60	2	0	0	26	7	0	0
6	5	0	0	0	3	62	3	0
7	9	0	0	0	0	6	49	2
8	10	0	0	0	0	0	3	111
Precision	79.97	90.1	88	91.66	70.27	81.57	83.05	98.23
Sensitivity	92.22	74.09	90.8	84.61	27.36	84.93	74.24	89.51
OA	84.7							
Kappa	0.83							

Figure 9.8: Tory Hill Fen: Confusion Matrix

3. Killyconny Bog

237

GROUND TRUTH   **MAP SEGMENTED IMAGE**

1 SUBMARGINAL   2 SUBCENTRAL   3 MARGINAL   4 CENTRAL   5 FACEBANK

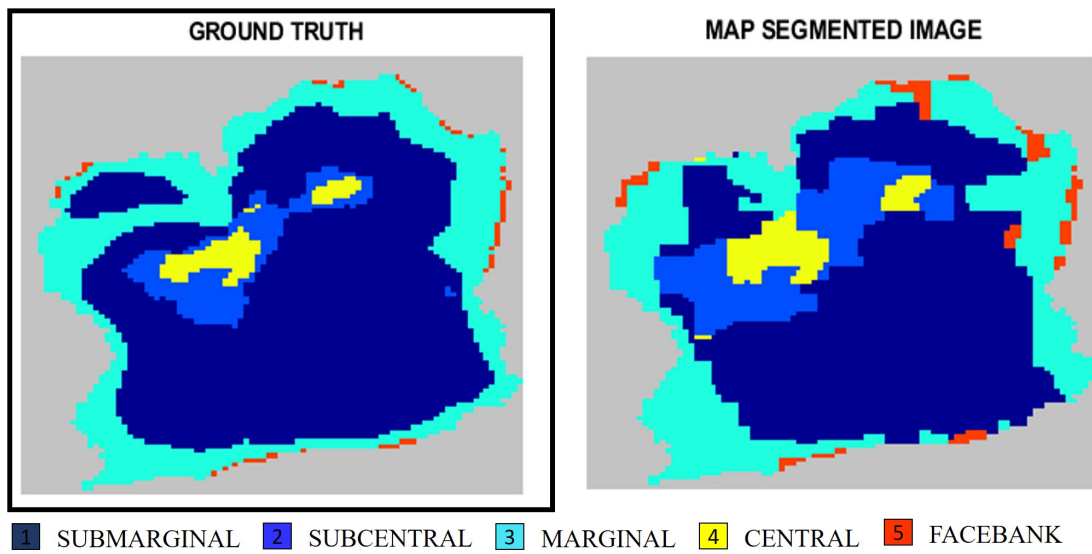Averaged map of vegetation communities in Killyconny

Figure 9.9: Killyconny Bog (a) GT (b) Classified Map

Averaged confusion matrix of Killyconny with ecotypes
1. Submarginal 2. Subcentral 3. Marginal 4. Central 5. Faceback

	1	2	3	4	5
1	10426	30	454	5	0
2	493	491	4	0	0
3	1385	3	7745	0	73
4	30	0	8	17	0
5	41	0	476	0	638
Precision	84.2	93.7	89.1	77.2	89.7
Sensitivity	95.5	49.6	84.1	30.9	55.2
OA	86.54				
Kappa	0.82				

Figure 9.10: Killyconny Bog: Confusion Matrix

4. Knockacollar Bog

Averaged map of vegetation communities in Knockacollar

Figure 9.11: Knockacollar Bog (a) GT (b) Classified Map

Averaged confusion matrix of Knockacollar with ecotypes
1. Submarginal 2. Subcentral 3. Marginal 4. Central 5. Facebank

	1	2	3	4	5
1	2589	379	389	34	8
2	24	286	0	40	0
3	253	24	1179	10	93
4	0	6	0	149	0
5	0	0	7	0	55
Precision	90.33	41.1	74.2	63.9	35.2
Sensitivity	76.1	81.7	75.6	96.1	88.7
OA	77.84				
Kappa	0.75				

Figure 9.12: Knockacollar Bog: Confusion Matrix

5. Mongan Bog

GROUND TRUTH MAP SEGMENTED IMAGE

1 SUBMARGINAL   2 SUBCENTRAL   3 MARGINAL   4 CENTRAL   5 FACEBANK

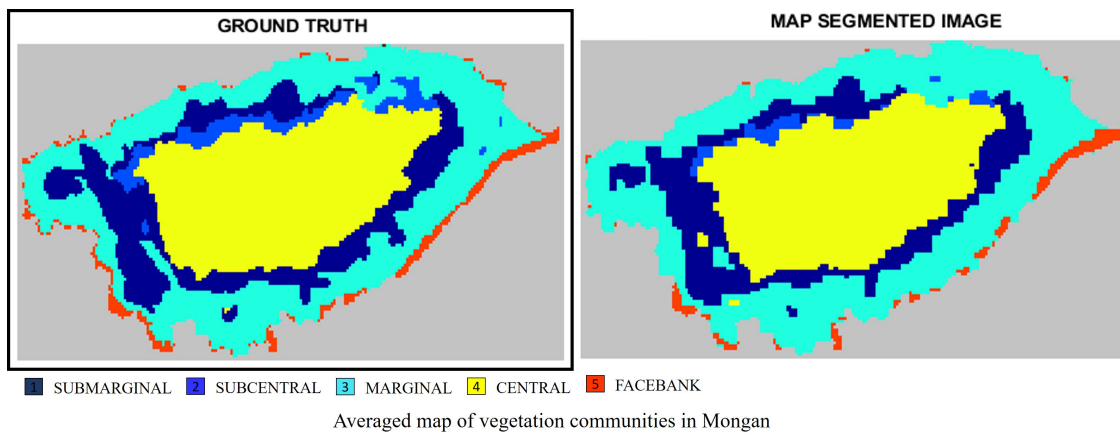Averaged map of vegetation communities in Mongan

Figure 9.13: Mongan Bog (a) GT (b) Classified Map

Averaged confusion matrix of Mongan with ecotypes
1. Submarginal 2. Subcentral 3. Marginal 4. Central 5. Facebank

	1	2	3	4	5
1	1859	10	536	155	0
2	147	160	127	149	0
3	257	0	4378	29	63
4	71	26	8	4307	0
5	0	0	217	0	291
Precision	79.6	81.6	83.1	92.8	82.2
Sensitivity	72.6	27.4	92.6	97.6	57.2
OA	85.96				
Kappa	0.83				

Figure 9.14: Mongan Bog: Confusion Matrix

6. Knockaunroe Turlough

1 UNKNOWN	2 LOLIUM GRASSLAND	3 SEDGE FEN	5 PERSICARIA AMPHIBIA	5 REEDBED	6 ELEOCHARIS PALUSTRIS – RANUNCULUS FLAMMULA

1 UNKNOWN 2 LOLIUM GRASSLAND 3 SEDGE FEN 5 PERSICARIA AMPHIBIA 5 REEDBED 6 ELEOCHARIS PALUSTRIS – RANUNCULUS FLAMMULA
7 OPEN WATER 8 LIMESTONE GRASSLAND 9 SCHOENUS NIGRICANS FEN 10 FLOODED PAVEMENT 11 CAREX NIGRA – CAREX PANICEA
12 FILIPENDULA ULMARIA – POTENTILLA ERECTA VIOLA SP. 13 AGROSTIS STOLONIFERA – POTENTILLA ANSERINA - FESTUCA
14 AGROSTIS STOLONIFERA – GLYCERIA FLUITANS 15 PEATY POND 16 SCRUB 17 WOODLAND 18 POTENTILLA FRUTICOSA

Averaged map of vegetation communities in Knockaunroe

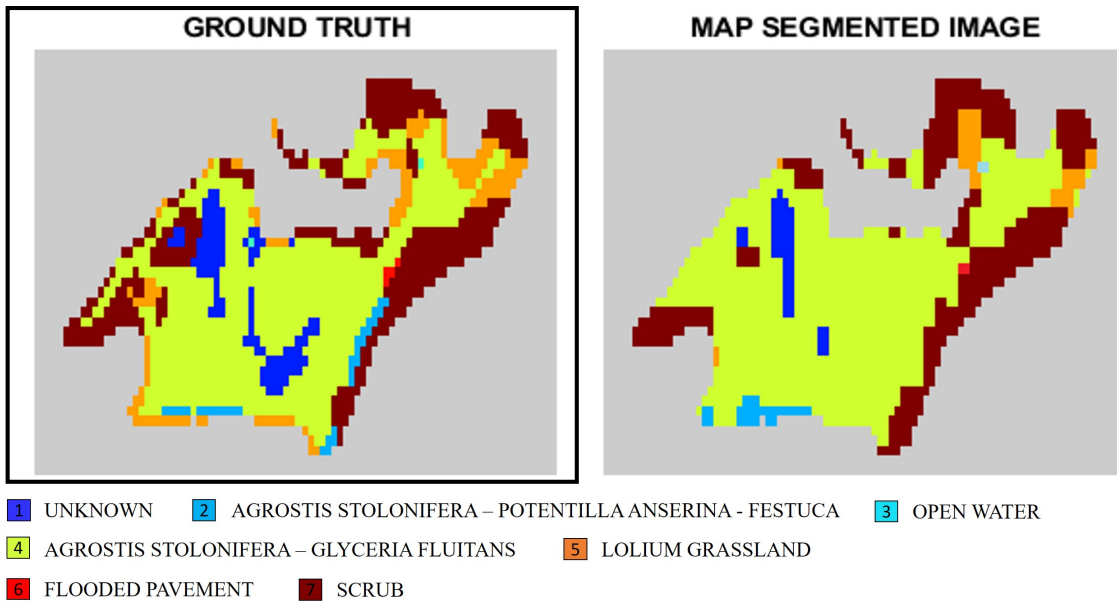Figure 9.15: Knockaunroe Turlough (a) GT (b) Classified Map

Averaged confusion matrix of Knockaunroe with
1. Unknown 2. Lolium grassland 3. Sedge fen 4. Persicaria amphibia 5. Reedbed 6. Eleocharis palustris – Ranunculus flammula 7. Open water 8. Limestone grassland 9. Schoenus nigricans fen 10. Flooded pavement 11. Carex nigra – Carex panicea 12. Filipendula ulmaria – Potentilla erecta – Viola sp. 13. Agrostis stolonifera – Potentilla anserine – Festuca 14. Agrosts stolonifera – Glyceria fluitans 15. Peaty pond 16. Scrub 17. Woodland 18. Potentilla fruticosa

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1049	187	32	0	68	324	24	0	0	34	2	0	8	0	0	6	0	0
2	100	3781	120	0	17	94	0	0	5	3	49	2	2	3	2	96	6	0
3	31	327	2115	10	129	669	0	2	1	315	48	5	1	0	6	67	0	7
4	0	47	5	104	0	21	0	0	0	10	9	4	0	0	0	9	0	0
5	37	39	52	0	1163	215	0	0	0	5	0	0	0	0	14	0	0	0
6	96	169	455	5	173	4889	5	0	8	507	8	0	0	0	0	29	0	0
7	6	0	1	0	0	36	80	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	25	0	129	0	297	0	0	0	0	7	0	0	4
9	0	10	7	0	7	38	0	0	28	0	1	0	0	0	0	9	0	0
10	1	35	155	0	4	540	0	25	0	4381	7	4	0	0	0	31	13	47
11	8	137	34	9	30	37	0	0	14	23	344	0	0	0	0	43	0	0
12	0	7	0	0	0	4	0	0	0	0	0	38	0	0	0	0	0	0
13	3	1	0	0	0	0	0	0	0	0	0	0	63	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0
15	0	5	6	0	5	0	0	0	0	0	0	0	0	0	50	0	0	0
16	5	172	90	15	8	142	0	0	14	59	11	2	0	5	0	785	0	0
17	3	19	6	0	0	21	0	0	0	34	0	0	0	0	0	0	112	1
18	0	0	10	0	0	15	0	3	2	211	0	0	0	0	0	3	2	147
Precision	78.34	76.6	68.49	72.72	72.5	69.15	73.39	81.13	40	74.51	71.81	69.09	85.13	74.19	69.44	72.35	84.21	71.35
Sensitivity	60.49	88.34	57.72	49.76	76.26	77.06	65.04	27.92	28	83.55	50.66	77.55	94.02	100	75.75	60.01	57.14	37.59
OA	72.67																	
Kappa	0.68																	

Figure 9.16: Knockaunroe Turlough: Confusion Matrix

7. Lough Aleenaun Turlough

GROUND TRUTH MAP SEGMENTED IMAGE

1 UNKNOWN 2 AGROSTIS STOLONIFERA – POTENTILLA ANSERINA - FESTUCA 3 OPEN WATER

4 AGROSTIS STOLONIFERA – GLYCERIA FLUITANS 5 LOLIUM GRASSLAND

6 FLOODED PAVEMENT 7 SCRUB

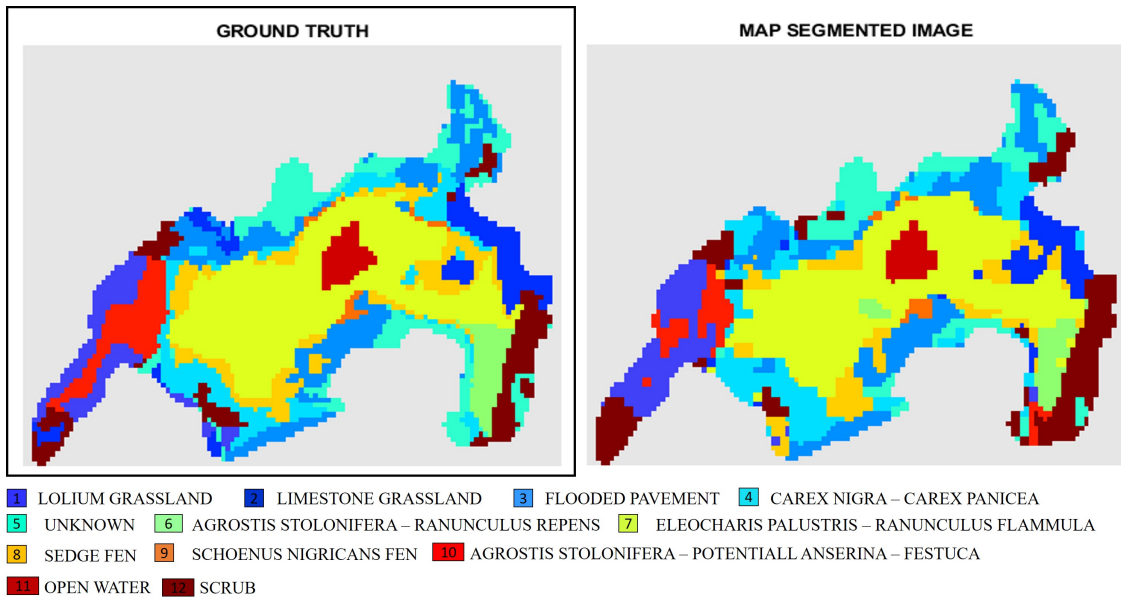Averaged map of vegetation communities in Lough Aleenaun

Figure 9.17: Lough Aleenaun Turlough (a) GT (b) Classified Map

Averaged confusion matrix of Lough Aleenaun with
1. Unknown 2. Agrostis stolonifera-Potentilla anserina-Festuca 3. Open water 4.
Agrostis stolonifera-Glyceria fluitans 5. Lolium grassland 6. Flooded pavement 7. Scrub

	1	2	3	4	5	6	7
1	40	0	0	60	0	0	0
2	0	12	0	6	0	0	15
3	0	0	1	1	0	0	0
4	5	5	0	668	6	0	58
5	0	9	0	68	30	0	34
6	0	0	0	2	0	1	1
7	1	0	0	101	16	0	262
Precision	86.95	46.15	100	73.56	57.92	100	70.81
Sensitivity	40	36.36	50	90.02	21.27	33.34	68.94
OA	72.12						
kappa	0.67						

Figure 9.18: Lough Aleenaun Turlough: Confusion Matrix

8. Roo West Turlough

242

1 LOLIUM GRASSLAND	2 LIMESTONE GRASSLAND	3 FLOODED PAVEMENT 4 CAREX NIGRA – CAREX PANICEA
5 UNKNOWN 6 AGROSTIS STOLONIFERA – RANUNCULUS REPENS		7 ELEOCHARIS PALUSTRIS – RANUNCULUS FLAMMULA
8 SEDGE FEN 9 SCHOENUS NIGRICANS FEN 10 AGROSTIS STOLONIFERA – POTENTIALL ANSERINA – FESTUCA		
11 OPEN WATER 12 SCRUB		

Averaged map of vegetation communities in Roo West

Figure 9.19: Roo West Turlough (a) GT (b) Classified Map

Averaged confusion matrix of Roo West with
1. Lolium grassland 2. Limestone grassland 3. Flooded pavement 4. Carex nigra – Carex panacea 5. Uknown 6. Agrostis stolonifera – Ranunculus repens 7. Eleocharis palustris – Ranunculus flammula 8. Sedge fen 9. Schoenus nigricans fen 10. Agrostis stolonifera – Potentilla anserine – Festuca 11. Open water 12. Scrub

	1	2	3	4	5	6	7	8	9	10	11	12
1	197	2	2	18	0	0	1	7	0	13	0	20
2	4	145	20	41	0	0	10	10	0	0	0	51
3	0	1	347	55	89	0	8	10	10	0	0	12
4	0	18	67	213	5	0	16	28	0	12	0	14
5	4	9	38	49	274	2	3	15	0	6	0	90
6	3	0	0	0	0	94	14	3	0	7	0	17
7	0	0	5	18	0	29	1068	45	3	0	10	0
8	0	24	54	59	0	0	62	180	0	0	0	2
9	0	0	4	0	0	0	13	0	19	0	0	0
10	87	2	0	4	1	0	0	4	0	106	0	15
11	0	0	0	0	0	0	14	0	0	0	85	0
12	0	1	0	32	10	1	1	7	0	1	0	289
Precision	66.77	71.78	64.61	43.55	72.29	74.6	88.26	58.25	59.37	73.1	89.47	56.67
Sensitivity	75.76	51.6	65.22	57.1	55.91	68.11	60	47.24	52.77	48.4	85.85	84.5
OA	70											
kappa	0.58											

Figure 9.20: Roo West Turlough: Confusion Matrix

9. Turloughmore Turlough

243
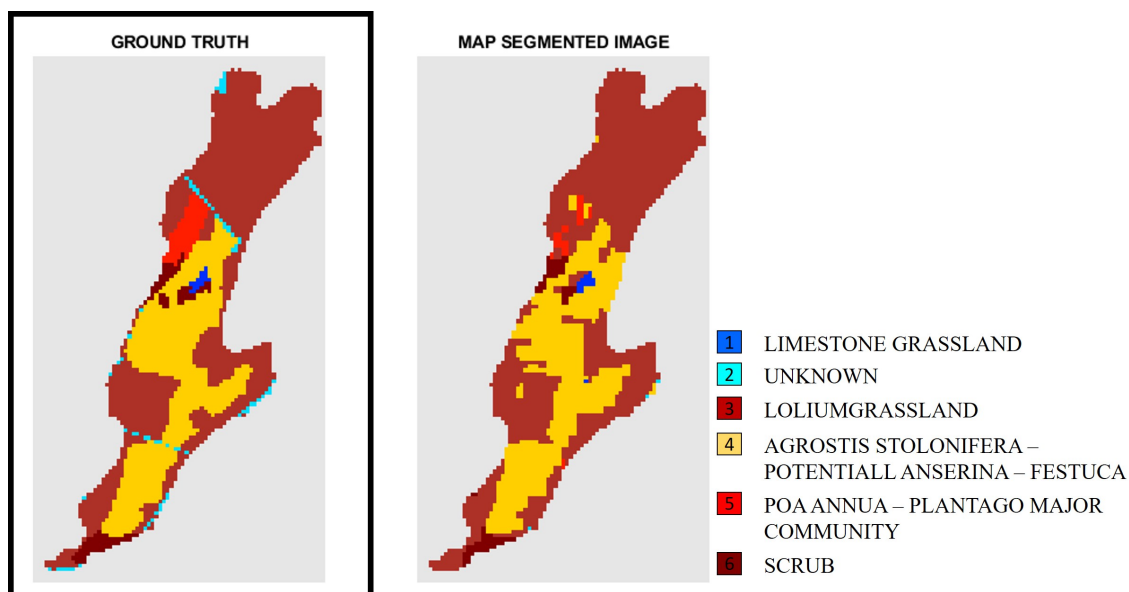
Averaged map of vegetation communities in Turloughmore

Figure 9.21: Turloughmore Turlough (a) GT (b) Classified Map

Averaged confusion matrix of Turloughmore with
1. Limestone grassland 2. Unknown community 3. Lolium grassland 4. Agrostis
stolonifera – Potentilla anserine – Festuca 5. Poa annua – Plantago 6. Scrub

	1	2	3	4	5	6
1	13	0	2	1	0	1
2	1	18	47	8	0	5
3	19	6	1658	181	23	24
4	21	2	211	780	4	18
5	1	0	40	15	46	8
6	2	9	19	23	1	75
Precision	22.8	51.42	83.86	77.38	62.16	57.25
Sensitivity	76.47	22.78	86.76	75.28	41.81	58.13
OA	78.91					
kappa	0.7					

Figure 9.22: Turloughmore Turlough : Confusion Matrix

Part III

Appendix C

9.3 Drone Imagery

9.3.1 Introduction to drone imagery

Before the drones came into the picture, the study was solely concentrating on satellite data (S2), with a spatial resolution of 10m. DJI Inspire-1, when flown at the height of 100m, had a spatial resolution of 1.8cm. The video (`https://drive.google.com/file/d/1lQgBH3f7ndXoCnvDqQo1yUzNgVRolMxF/view?usp=sharing`) gives a visualisation of the depth and detail of going from 10m to 1.8cm. It was made for Clara bog in June 2019.

Some other images, and movies while the drone surveys can be found here (`https://drive.google.com/file/d/1CGh2k5M1iMIOUAqFvkOVX4zD_-BpWS_w/view?usp=sharing`).

The resolution of the images being captured was also cross-checked manually by calculation. Aa drone image of Trinity College's rugby field was taken, which was later compared with a standard rugby field. This way, using the height of flight - the spatial resolution was measured.

(a) Image of Rugby field (Trinity College Dublin) taken by DJI Inspire 1

(b) Image of a standard rugby field for reference of area
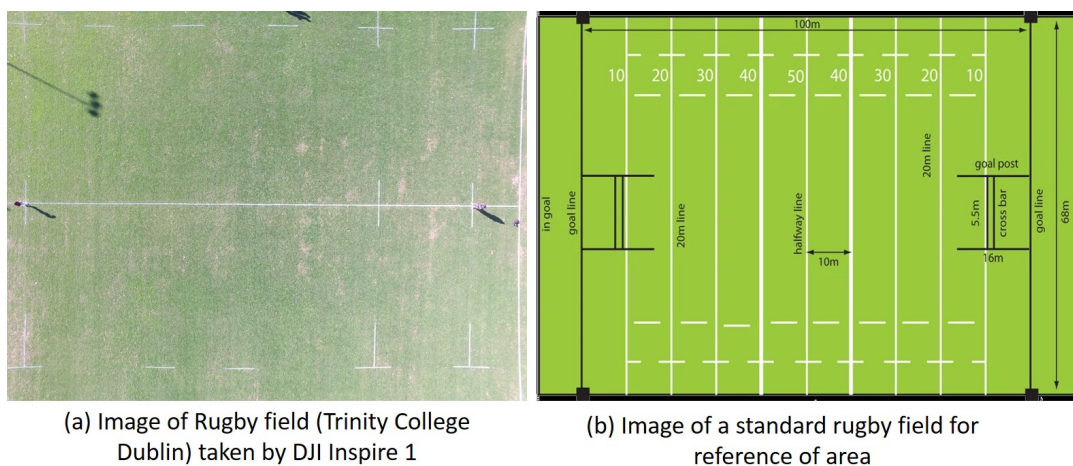
At a flying height of 28.5 m
Size of the image (pixel-wise)= 4000 x 3000
Size of the rugby field in the image (approx.) = 60 x 40 sq. m
➔ 1 pixel = 0.5 cm ➔For a height of 1m, resolution = 0.0175 cm
➔ For a height of Nm, resolution = 0.0175 x N cm == **0.018 x N cm**

Figure 9.23: Manually calculating spatial resolution of DJI Inspire 1

9.3.2 Drone image classification

- Scragh Bog
  Drone image segmentation of Scragh bog using the ML algorithm (as discussed in Chapter 4) is shown below.
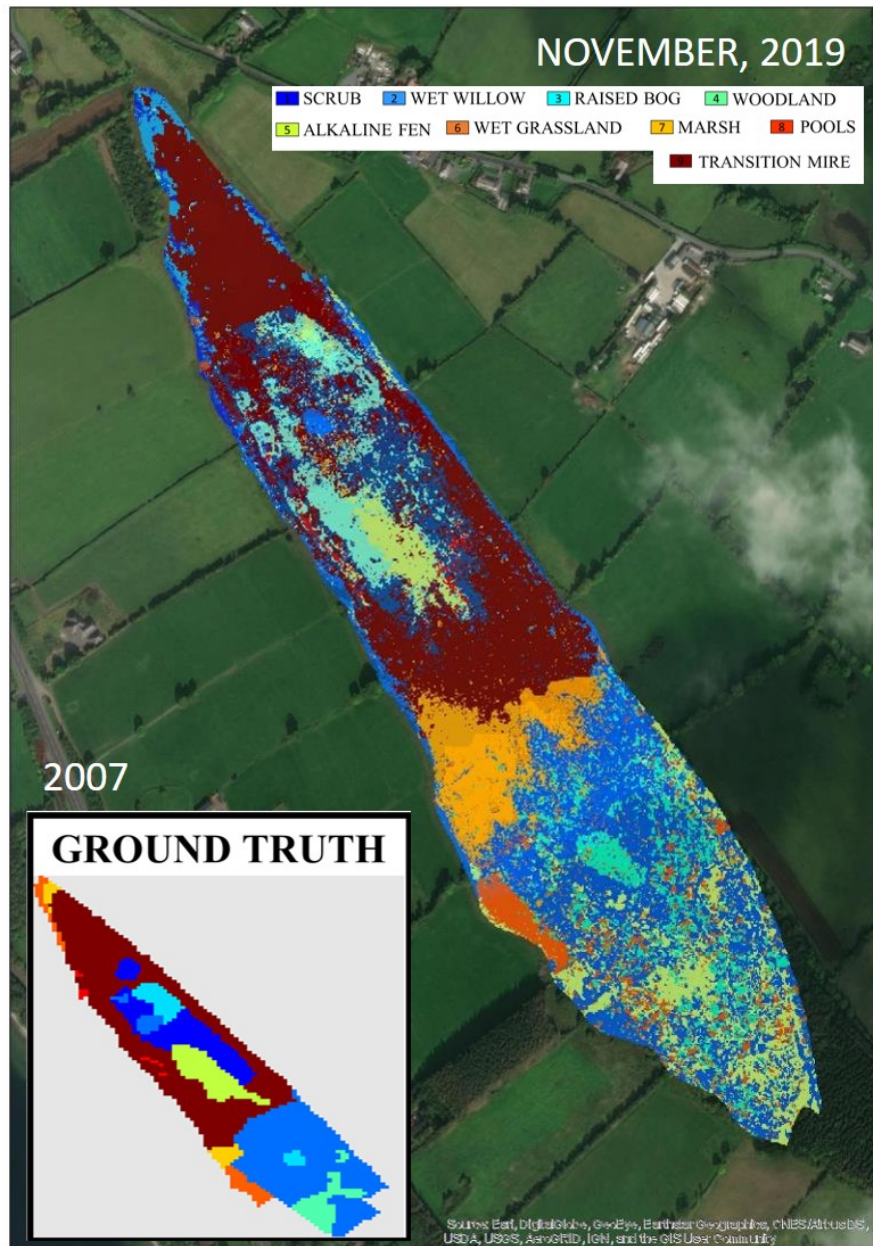
Figure 9.24: Scragh Bog- Drone vegetation survey

It can be seen that due to high level of detail, vegetation communities are sparsely identified. The segmentation can be made better using drone and satellite together.

- Ballymore Fen (drone + satellite)
  A similar trend, like Scragh bog, was seen in Ballymore fen. The Ballymore fen has a mosaic of various vegetation communities. If the GT is not updated for such amount of detail, the drone segmented image comes out very sparse, and hard to interpret. Therefore, for Ballymore, S2 image along with the drone image was used. This was done using the methodology explained in Chapter 5 for Clara bog.

249
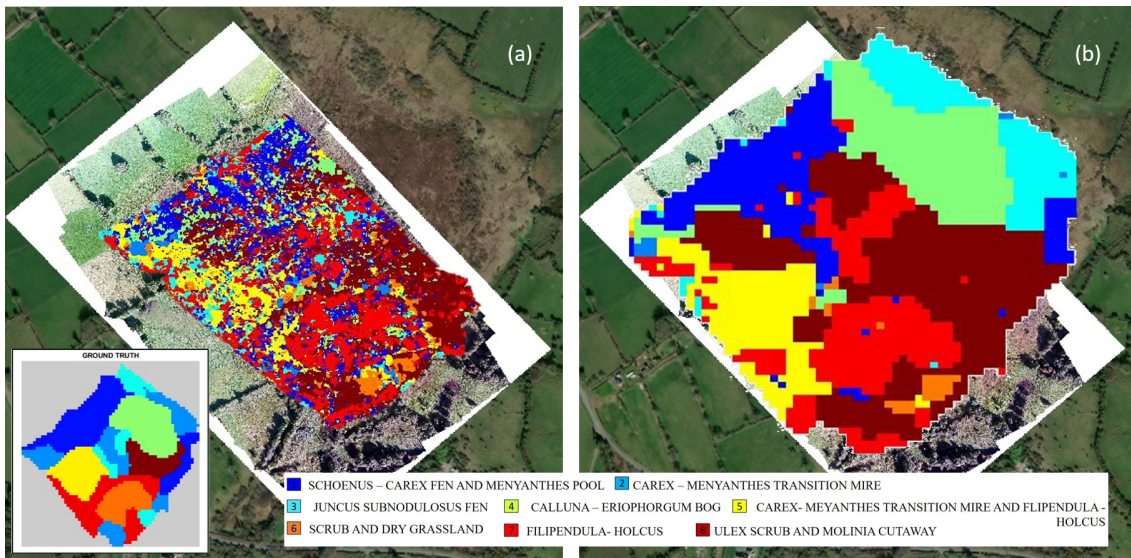
Figure 9.25: Ballymore Fen- (a)Drone (b)nested drone satellite vegetation mapping

It can be seen that, usage of S2 data along with drone imagery gives better clusters, and the map is better understandable.

9.3.3 Drone image classification (labelling) using CNN

Using the drone images captured on 16th August 2019, an image labelling CNN architecture was explicitly created to identify and label the ecotopes present on Clara Bog. The same architecture was also tested on Mongan bog. The results indicate that the architecture is very effective and can be applied globally to wetlands for species-level vegetation identification.

9.3.4 Creating data for training

The size of the images captured by the drone is 3000 x 4000 pixels, containing multiple ecotopes. The steps taken for creating the training dataset are as follows:

1. A smaller subset of the image containing only a single class was initially cropped out. The size of the cropped images varied from a few 100s to 1000s of pixels (depending on the area that has been selected).

Figure 9.26: Selecting smaller area per ecotope

2. The images were further divided into 128 x 128 x 3 such that a bigger set of images are created from each subset.
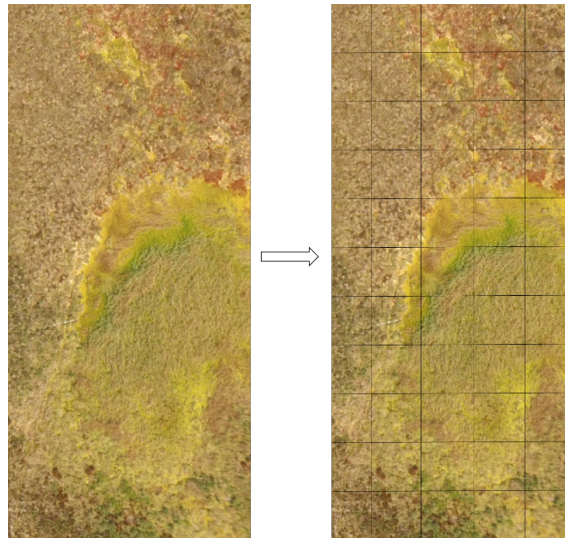
Figure 9.27: Dividing the whole area into smaller patches

3. Each of the small images is stored in a folder titled as the ecotope.

4. The process was repeated for 40 images, and a total of 1500 images per ecotope (class) was created.

The size of the image (128 x 128 x 3) was decided after inspection, as the images smaller than this did not have enough gradient - hence, this was defined as the minimum size required for further labelling. 75% of images from each of the ecotope were used for training, and the rest 25% was kept for testing the model.

9.3.5 CNN Architecture

The following simple architecture was created using 12 layers for labelling the images into 4 major categories (central, marginal, flush, and submarginal(SM) + subcentral (SC) =

(SMSC)). The SM and SC category was merged as these two ecotopes are highly homogeneous, and it was very difficult to distinguish between the two categories manually.

The network architecture is as follows, The optimisation algorithm used was SGDM with initial LR = 0.01, L2 regularisation. The max number of epochs = 4, and images were shuffled at every epoch and a mini batch size of 64.

1. Image input layer
2. Convolution 2D layer (kernel size = 3, No. of kernels stacked = 8, Padding = same)
3. Batch Normalisation
4. Activation Func (ReLu)
5. Max Pooling (stride = 2)
6. Convolution 2D layer (kernel size = 1, No. of kernels stacked = 16, Padding = same)
7. Batch Normalisation
8. Activation Func (ReLu)
9. Max Pooling (stride = 2)
10. Convolution 2D layer (kernel size = 3, No. of kernels stacked = 31, Padding = same)
11. Batch Normalisation
12. Activation Func (ReLu)
13. Fully connected layer (for 4 output classes)
14. Softmax classifier
15. Classification output layer

9.3.6 Results

The validation accuracy of 91.58% was achieved.The OA (test) = 88.5% .
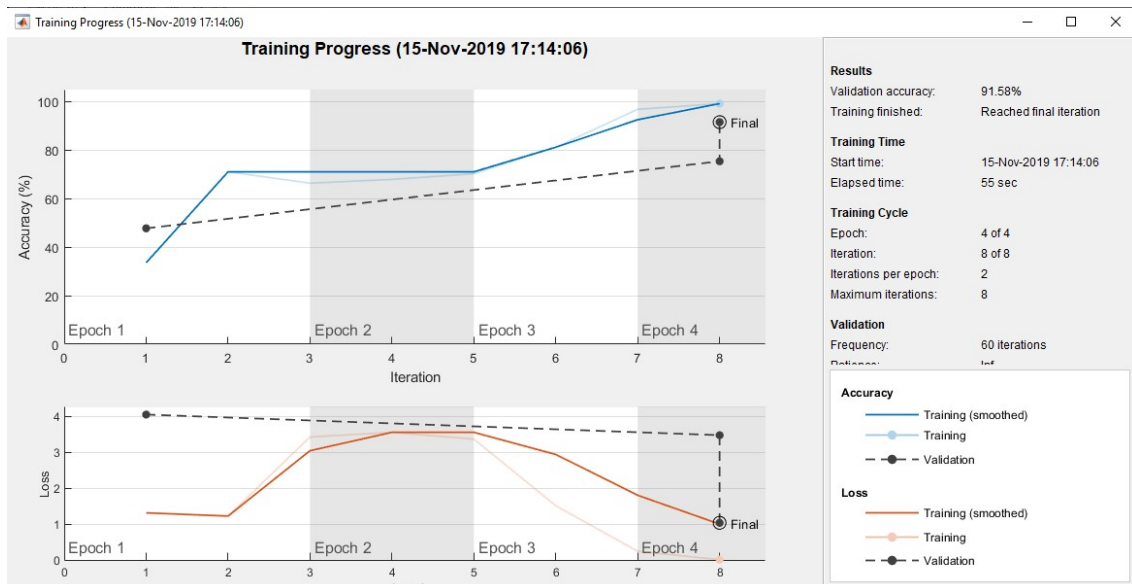
Figure 9.28: Training Progress Report, Matlab v 2019
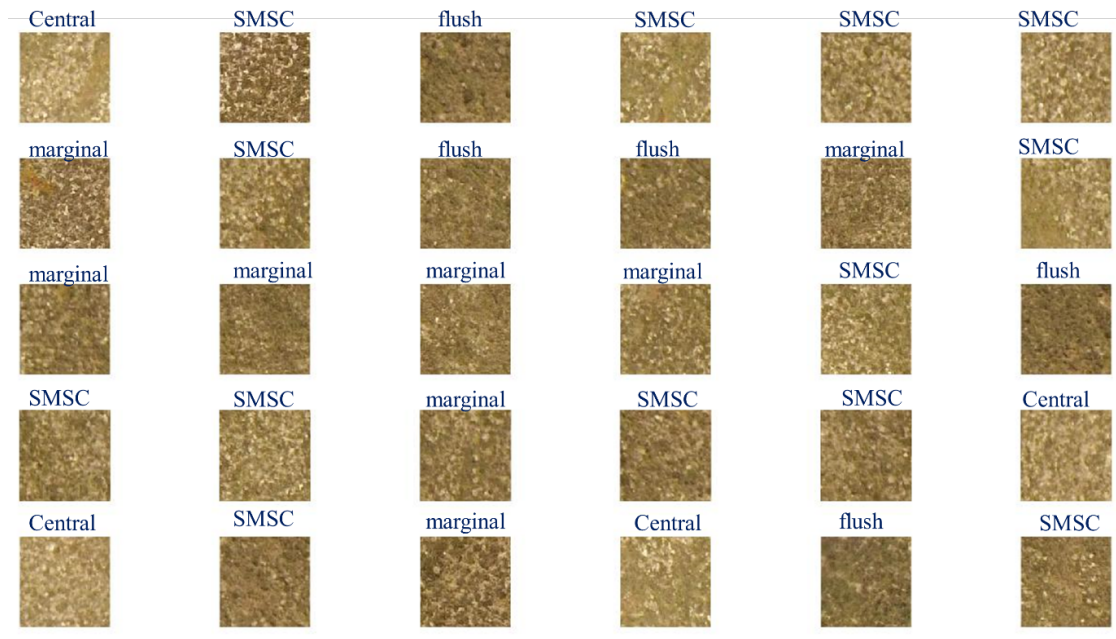
Figure 9.29: Labelled Patches

The central ecotope is mapped with the best accuracy, followed by flush ecotope. Marginal and flush ecotopes exist with overlapping boundaries, but there is a clear distinction between the two ecotopes. For SMSC, there has been some wrong identification, but the majority is identified correctly. The accuracy can be increased by increasing the training data and augmenting it.

The main advantage of using small images along with a small and straightforward architecture is the amount of training time. The whole process was done in less than a minute with high accuracy. This proves that the model is efficient and robust for the image labelling task, although the efficiency of the model is mainly dependent on the training data and an increase in training data shows an exponential increase in accuracy.

CNN can now be used for image classification, bounding box object detection, key-point prediction, etc. For a lot of global processes such as autonomous driving, image search engines, augmented reality etc. a complete understanding of the whole scene in an image is required. Therefore, something more than just labelling was required. Hence, in the main study, semantic segmentation was done.

9.3.7   3D modelling using drone imagery

Using the structure from motion (sfm) technique, there were some wetlands for which 3D models were reconstructed. The sfm technique essentially uses the stereo images to create an estimated 3D model of the area. Since the stereo images that are taken has a time gap in between them, this model may not represent exact information.

For this study, we took stereo images of Clara Bog, and Scragh Bog (Alkaline fen part). These images are presented in the figures below.

- Clara Bog
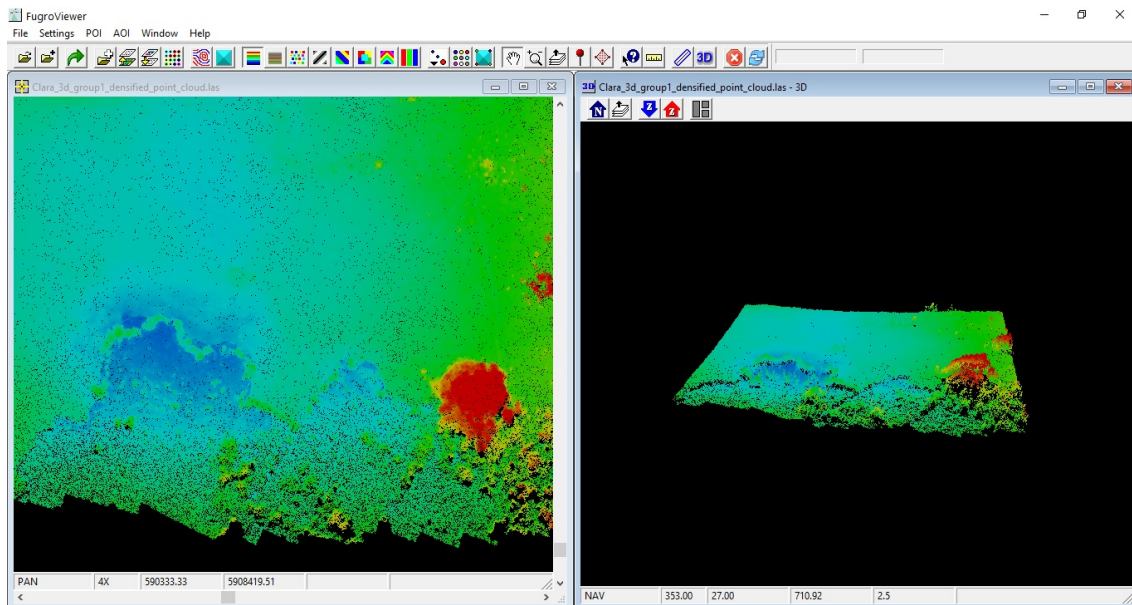
Figure 9.30: Clara Bog (west) - Facebank, Marginal and Submarginal ecotopes as point cloud

- Scragh Bog
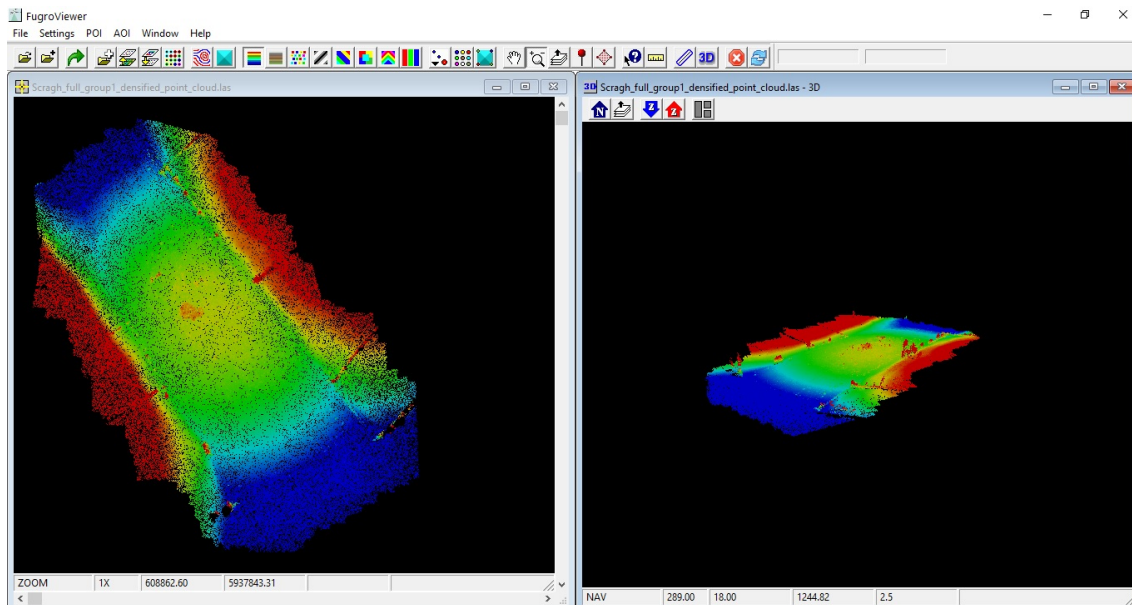
Figure 9.31: Scragh Bog (full area) as point cloud

The Alakline fen part of the Scragh bog is swelled (as it has absorbed water) and the minute vegetation on it can be clearly seen.
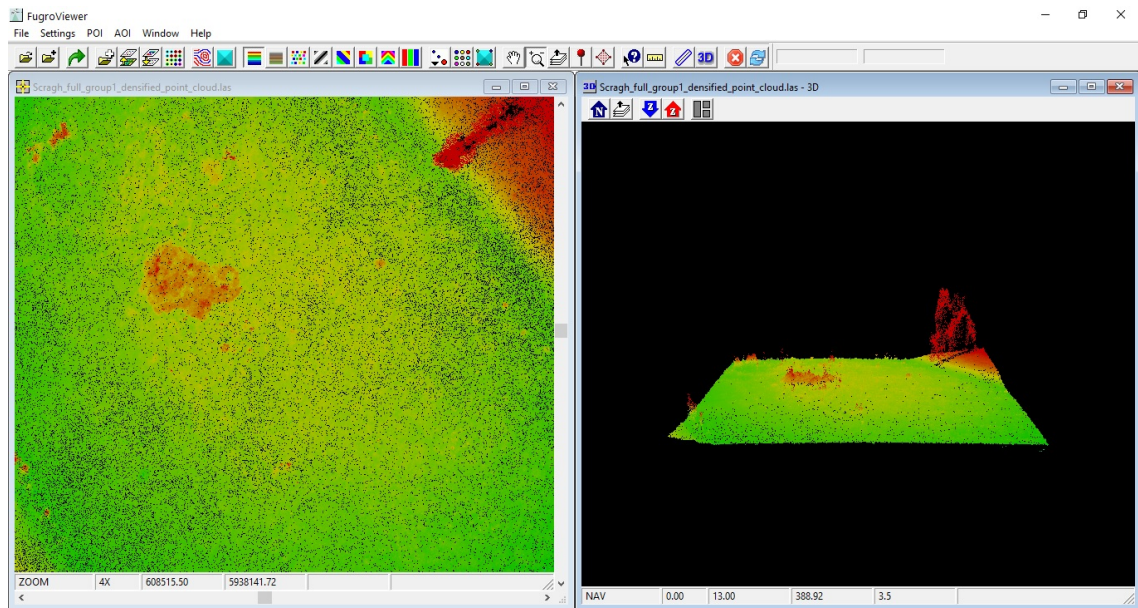
Figure 9.32: Alkaline fen (Scragh Bog) as point cloud

Part IV

Appendix D

9.4 Error Metric - Clara Bog (whole area)

The error metric is a systematic change detection metric, providing a comparison of spatial change between the images. The error metric was calculated between spring and summer maps of Clara bog, with the GT used as the reference against which other changes were detected. Hence, the approach is to attempt to use the changes detected between seasons as an additional metric by which boundaries of the different communities can be accurately determined. More details are given in Figure 7.23.
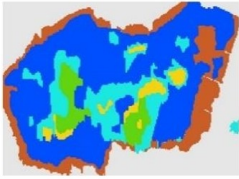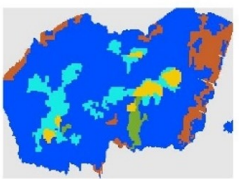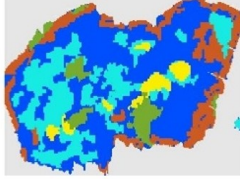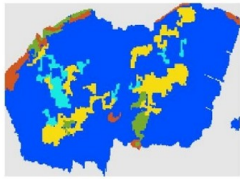
Date drone image acquired	18th May 2019	4th August 2019	14th November 2019
Date satellite image	21st April 2019	18th Sept 2019	12th November 2019
GT created using D images	D = 4	D = 5	D = 5
GT – created in 2017	Spring map (Z_1)	Summer map (Z_2)	Winter map (Z_3)

Figure 9.33: Date of drone and satellite images used, along with the number of images used to map 5 key ecotopes for the whole of Clara bog

9.4.1 Submarginal

The error metric of SM ecotope is presented below in Figure 7.24.

SUBMARGINAL		SPRING	SUMMER	WINTER
Jaccard Similarity (w.r.t. GT)		0.65	0.50	0.56
Area (m^2)	16568	18615	13986	18662
Centroid (*lat long;* XY map value)	*53..3194, -7.6599;* [115.41, 69.59]	*53.3195, -7.6600;* [114.90, 75.61]	*53.5181, -7.6600;* [129.08, 72.87]	*53.3187, -7.6600;* [123.37, 78.02]
Orientation	16.78	14.35	14.32	14.65
Extent	0.41	0.47	0.38	0.50
Change Image with centroid location				

Figure 9.34: Error metric for Submarginal (SM) ecotope for all the seasons

The submarginal ecotope appears to be the least variable ecotope in terms of identification between the seasons. Given the classification error and natural changes, a Jaccard index value of more than +0.5 for all the seasons was considered high. In reference to the GT, the area of SM has increased in the spring and winter and remains similar in summer. It is essential to understand that the ecotope does not increase/decrease in size in a couple of months – the change in identified area is linked to the ability of the methodology to identify the SM ecotope across the different seasons due to the difference in images used. This can be mainly due to misclassification/interference from the neighbouring wet communities like the subcentral (SC). This is due to growth and death of characteristic species mentioned in Figure 5.3. The SM and SC ecotopes are also highly homogenous, and hence, the misclassification rate is high (Bhatnagar et al. 2020a). During summer, the SM is comparatively dry with respect to the SC – leading to better differentiation identification. Also, the ecotope is more centred in summer as compared to spring and winter indicating its omnidirectional (extending in every direction) growth. From Figure 7.24, it can be seen that the major difference can be seen on the west side, where the submarginal was not appropriately identified in summer. From Figure 5.4, it can be seen that the spectral signature of this community for all seasons is overlapping. There is no significant shift in the centre of mass of the ecotope, and hence, based on the similarity index, it was concluded that this community is identified best in the spring season.

9.4.2 Subcentral

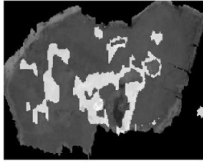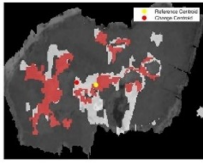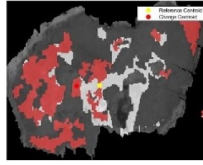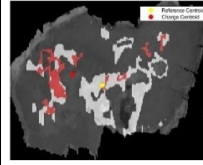The error metric of SC ecotope is presented in Figure 7.25.

SUBCENTRAL	SPRING	SUMMER	WINTER
Jaccard Similarity (w.r.t. GT)	0.28	0.20	0.16
Area (m²)	4416	2754	5532	1271
Centroid (*lat long;* XY map value)	*53.3197, -7.6600;* [112.73, 84.62]	*53.3216, -7.6600;* [90.051, 85.37]	*53.3218, -7.6600;* [88.85, 84.39]	*53.3215, -7.6600;* [92.40, 70.51]
Orientation	11.91	12.67	18.98	14.20
Extent	0.17	0.09	0.17	0.06
Change Image with centroid location				

Figure 9.35: Error metric for Subcentral (SC) ecotope for all the seasons

The Jaccard index of the SC ecotope keeps on changing every season – indicating the complex nature of the community. This ecotope is hard to identify using drone images, which leads to misclassification using the satellite image. However, when mapped directly using satellite imagery, this ecotope is mapped well (Bhatnagar et al. 2020a). There is a clear integration of the SM, SC ecotopes in summer. SC, just like SM, is relatively dry and hard to distinguish using an RGB sensor. From Figure 5.4, it can be seen that the spectral signatures of the SC ecotope are overlapping and remarkably similar to SM. A considerable reduction in SC is identified in winter, indicating the scattered presence of the ecotope. Also, due to homogeneity, misclassification, and pixel-mixing between SM, SC has led to a reduction in the SC area. Given that the ecotope itself will not have reduced significantly in this timeframe, the algorithm is only able to identify 30% of the total area of presence for this ecotope. Also, the vegetation doesn't have a hard boundary, and can vary in the maps generated manually. This directly affects the shift in the centroid of the community. The orientation in summer indicates the apparent growth of the community along the north-east direction. Therefore, it was concluded that the best time to map SC is spring, when the ecotopes SM, SC are most differentiable. It also means that for better identification of homogenous communities, a more detailed sensor (hyperspectral/LiDAR) is required.

9.4.3   Marginal

The error metric of M ecotope is presented in Figure 7.26.
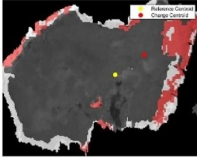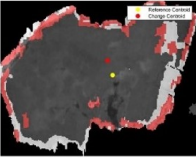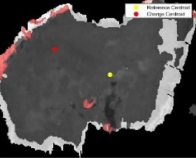
MARGINAL		SPRING	SUMMER	WINTER
Jaccard Similarity (w.r.t. GT)		0.35	0.34	0.14
Area (m^2)	5206	2946	3449	987
Centroid (*lat long; XY map value*)	53.3171, -7.6600; [140.08, 77.21]	53.3297, -7.6599; [161.00, 62.85]	53.3179, -7.6600; [131.12, 66.69]	53.3209, -7.6559; [98.10, 66.38]
Orientation	11.67	7.22	9.98	2.57
Extent	0.13	0.09	0.10	0.03
Change Image with centroid location				

Figure 9.36: Error metric for Marginal (M) ecotope for all the seasons

The Marginal ecotope has a mediocre similarity index for spring and summer but low in winter. This indicates the successful applicability of the algorithm for marginal ecotope in summer. The marginal ecotope is mainly dry, and depending on the season, green. Sentinel-2's vegetation index like NDVI can give a clear delineation of this community. From Figure 5.4, it can be seen that the spectral signature is not consistent for the Marginal ecotope. The appearance of this ecotope is similar to SM, SC in spring, and very similar to Active Flush (AF) in summer. This justifies the change in the level of identification of this ecotope in various seasons. Based on the area of the ecotope in all the seasons, the best time for identification of the ecotope was concluded to be summer.

9.4.4 Central
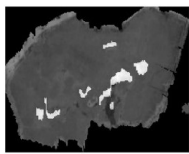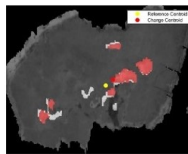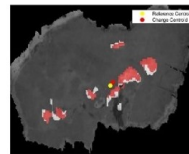
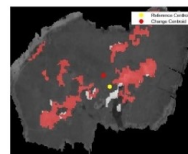The error metric of C ecotope is presented in Figure 7.27.

CENTRAL		SPRING	SUMMER	WINTER
Jaccard Similarity (w.r.t. GT)		0.45	0.48	0.17
Area (m^2)	1038	912	926	3116
Centroid (*lat long;* XY map value)	*53.3180, -7.6600;* [130.28, 86.69]	*53.3178, -7.6600;* [133.38, 82.64]	*53.3181, -7.6600;* [129.27, 82.09]	*53.3189, -7.6600;* [120,13, 73.99]
Orientation	22.86	20.57	12.47	23.53
Extent	0.08	0.04	0.04	0.11
Change Image with centroid location				

Figure 9.37: Error metric for Central (C) ecotope for all the seasons

The extent of Central ecotope is the most interesting when evaluating the ecological health of the wetland. The area of the Central ecotope for spring and summer is roughly equal to GT. This indicates that the community is mostly identified similarly. The centroid of the community is maintained between seasons indicating the high applicability of algorithm for this ecotope detection. The winter imagery was taken after a week of frequent rainfall (precipitation 19.6mm/week). The Central ecotope consists of moss species that hold water, though SC and AF also consist of water holding species, as indicated in Figure 5.3. There is also the presence of water pools, sphagnum lawns and extreme flushes near to the central ecotope complex. The water index, like NDWI, helps identify and distinguish such wet communities (Bhatnagar et al. 2020a). From Figure 5.4, it can be seen that the spectral signature of the spring and summer is overlapping, but winter has a distinct spectral signature. The winter spectral sign matches AF and SC, leading to an apparent overestimation of the central ecotope for winter. The central ecotope is the peat-forming ecotope, and its preservation is most important. Change detection techniques like this can confirm the status of the community. Since the overestimation in winter was due to the time of capture of the drone images, the best time to capture GT for this ecotope is spring and summer. By the area, summer was concluded to be the best time for the identification of the central ecotope.

9.4.5 Active flush

The error metric of AF ecotope is presented in Figure 7.28.
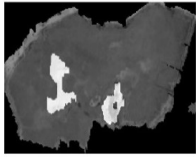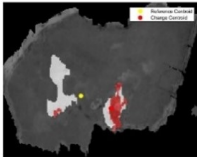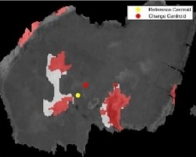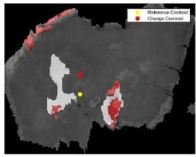
ACTIVE FLUSH		SPRING	SUMMER	WINTER
Jaccard Similarity (w.r.t. GT)		0.26	0.29	0.15
Area (m^2)	1730	750	1658	1096
Centroid (*lat long*; XY map value)	53.3210, -7.6600; [97.58, 96.88]	53.3184, -7.6600; [126.73, 93.50]	53.3199, -7.6600; [109.06, 85.19]	53.3196, -7.6599; [113.80, 68.88]
Orientation	-19.50	-13.14	-17.12	-11.45
Extent	0.24	0.03	0.05	0.04
Change Image with centroid location				

Figure 9.38: Error metric for Active Flush (AF) ecotope for all the seasons

The active flush ecotope existing in the east of the Clara bog is primarily identified correctly across all the seasons. The part of AF in the west appears to be merged and identified as central, SC leading to a low Jaccard index for the ecotope. The Active flush is best identified in summer image – this is due to a clear distinction using NDWI in Central and AF communities. AF almost dries up and merges with neighbouring communities in summer, but in winter, it is quite wet. Figure 5.4 depicts that the spectral sign of AF changes every season; it is notably different in summer (similar to SM, SC). A similarity of the spectral signature of AF with M in winter can also be seen (Figure 5.4), this justifies the identification of the AF ecotope on the boundary of the bog. Since the summer image gave the best identification of AF (taking area and Jaccard index in consideration), in a yearly map summer boundaries of the ecotope can be used.

Part V

Appendix E

9.5   First experience with Sentinel-5P data

9.5.1   S-5p for Ireland

- Correcting for in situ measurements
  The XCH4 in situ measurement for Ireland is done by Mace Head, the data is not available publicly. Total Carbon Column Observing Network (TCCON) data from Karlsruhe (Germany) located in same latitude range was used for adjusting the S-5p CH4 data. CH4 has an increasing trend both monthly and yearly. For effective comparison, this trend (rate of increase) is to be subtracted from all the images.

No. of days for which data is available	2018	2019
Karlsruhe	116	193
Sentinel 5P	55	44

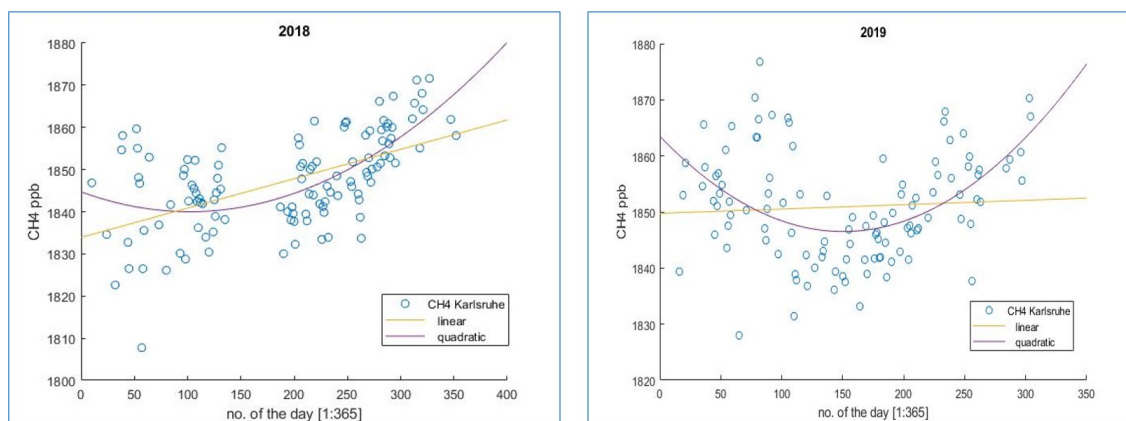| Rate of increase from 2018 to 2019	3. 72 ppb	
Rate of increase (ppb)	2018	2019
Linear	7.16	0.7763
Quadratic	7.20	4.8246

Figure 9.39: Rate of change calculation : TCCON, Kalshruhe

The linear annual increase in the XCH4 data as indicated by TCCON data at Karlsruhe is 3.3 ppb. CH4 signal varies from month to month and also yearly. If the variability is strong then the trend (rate of change) has to be subtracted from the satellite data.

Ireland will have a similar trend but 3.30 ppb of increase is not big, and will not cause any strong bias.

Therefore, no in situ corrections were required for the Irish data.

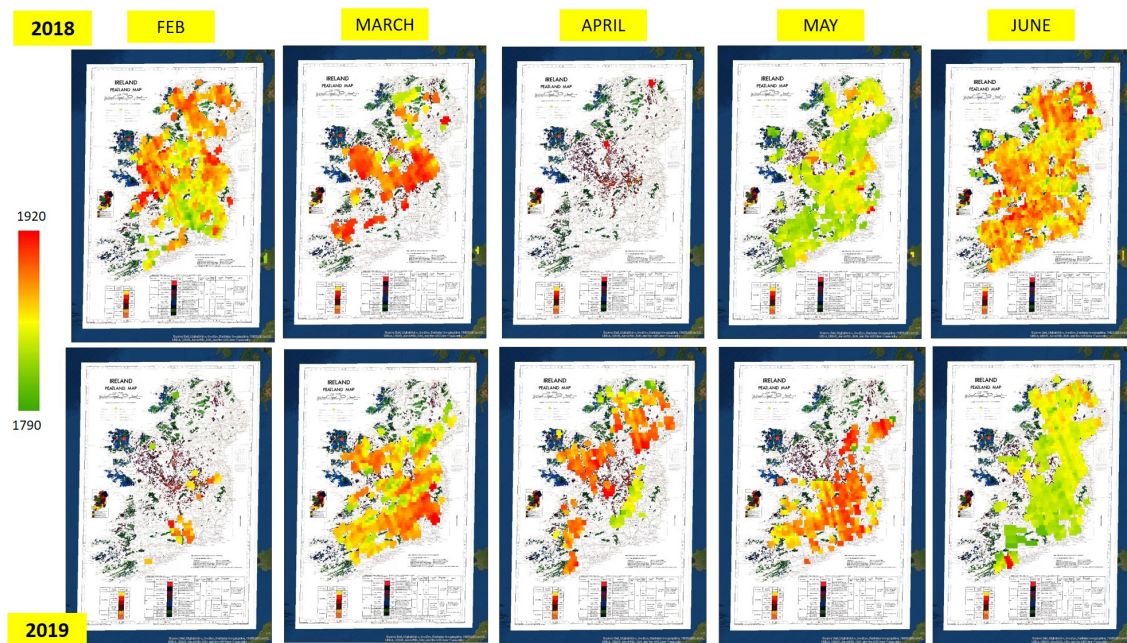- Monthly XCH4 maps, Ireland.

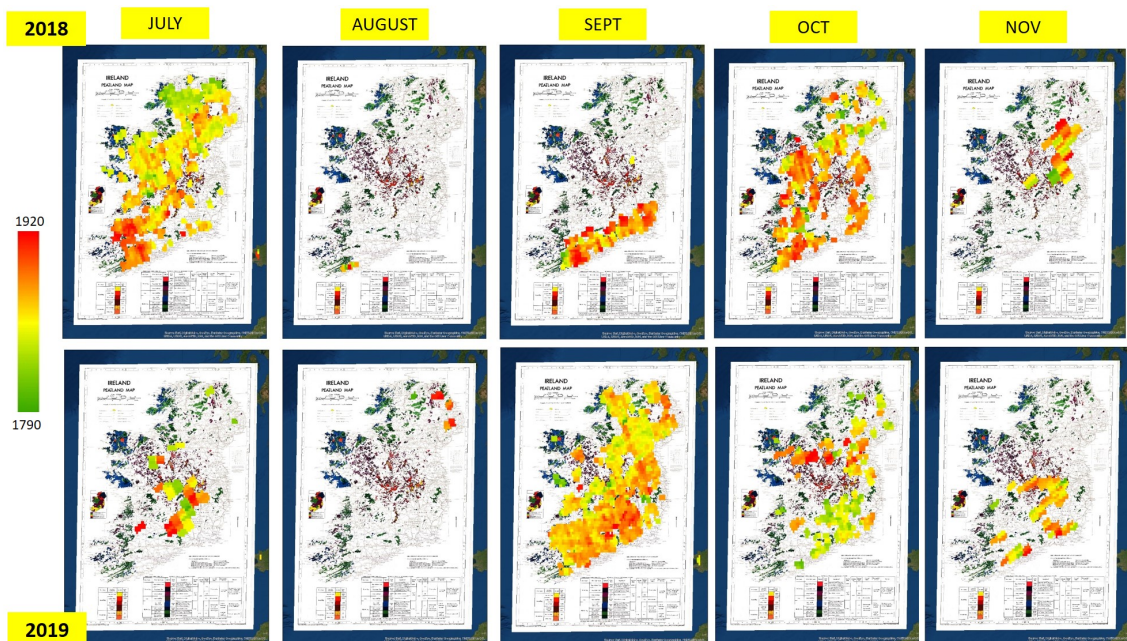Figure 9.40: Monthly maps (2018, 2019) - Part 1

Figure 9.41: Monthly maps (2018, 2019) - Part 2

- Yearly XCH4 Maps (2018, 2019)
  All the monthly maps were aggregated, and an averaged yearly map was created for both the years 2018, 2019.
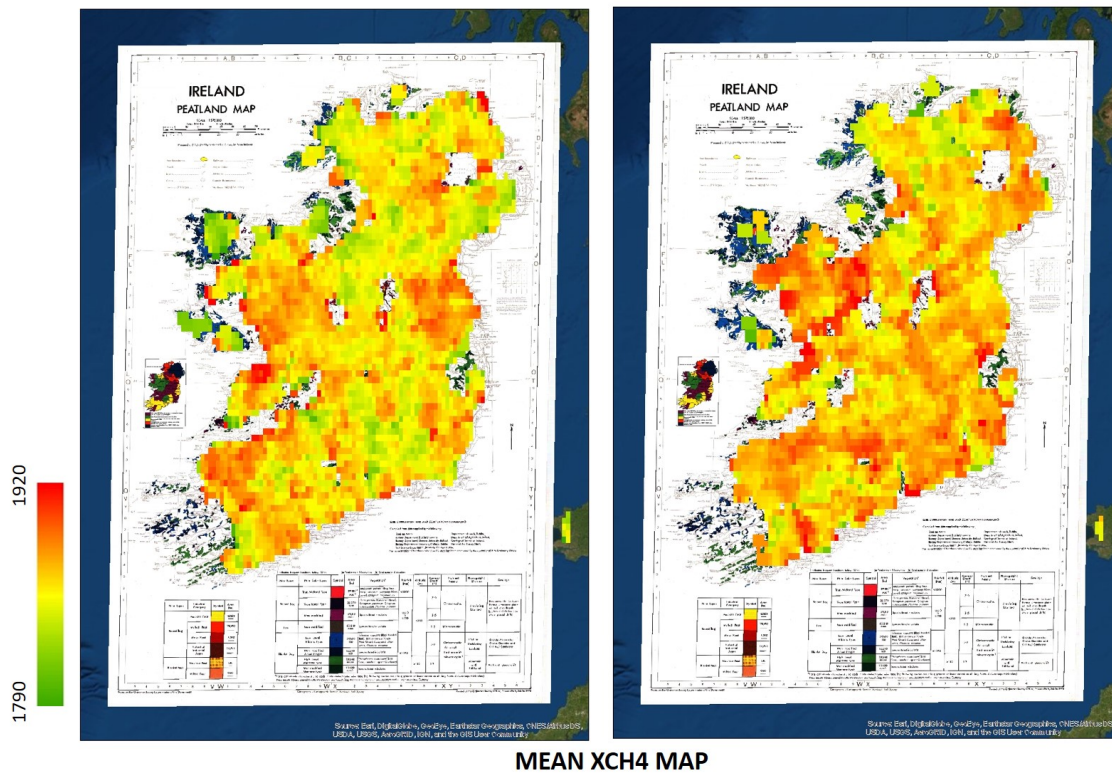
MEAN XCH4 MAP

Figure 9.42: Yearly Map (2018, 2019)

- Distribution of XCH4 in 2018, 2019
  Based on the yearly map, the XCH4 distribution for the entire country for the years
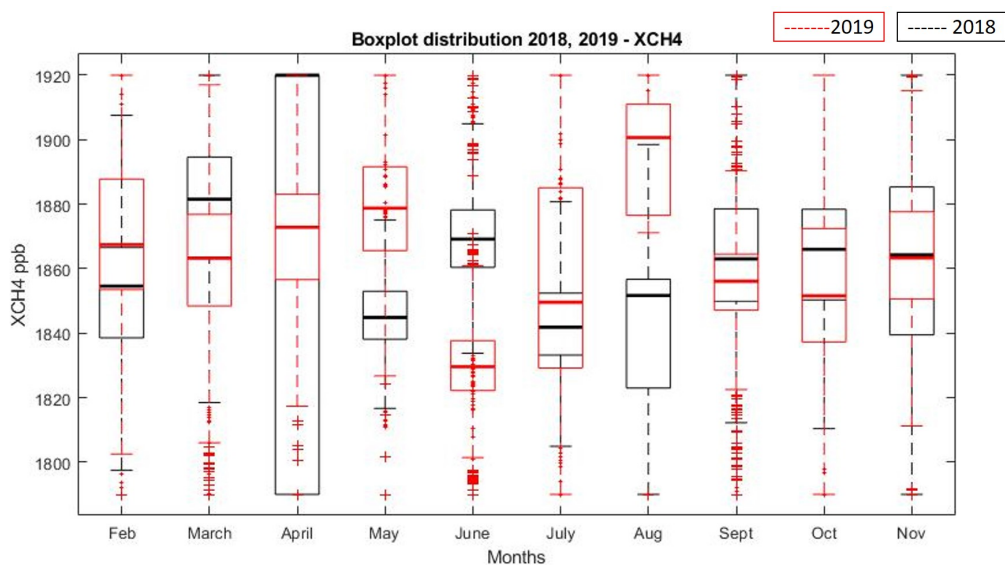 2018 and 2019 was calculated and analysed.

Figure 9.43: Box plot: XCH4 distribution (2018, 2019)

- Challenges with using XCH4 data in Ireland
  Very high influence of wind from neighbouring countries.
  Due to wind, high cloud cover, and presence of water - very less pixels are covered and available per year.
  Therefore, for wetland analysis, not enough XCH4 data is available for Ireland. Hence, other areas with high amount of wetlands and a bigger area (like Canada) was considered.

9.5.2 S-5p for Canada

Canada has 10 main land types: Bogs, Fens, Marshes, Swamps, Shallow water, Deep water, Forests, Grassland, Cropland and Barren land. Firstly, we have analysed the S-5p data associated with these land types. Due to the large size of the country, the initial analysis was done for a part of Canada (and not whole of Canada).
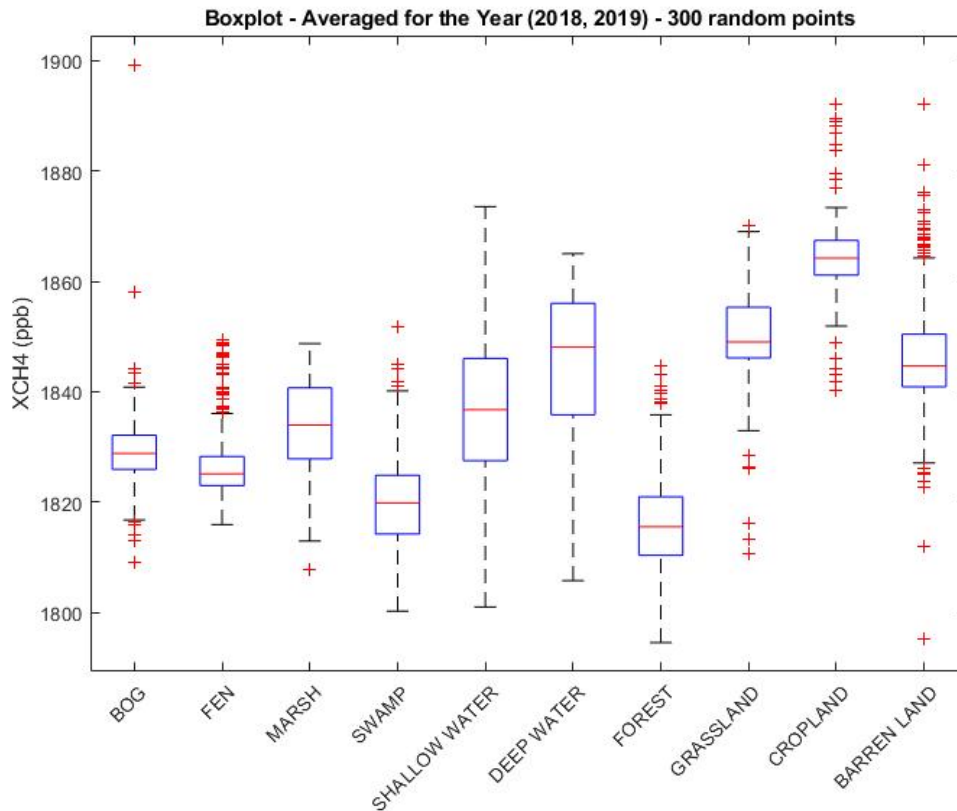
Data analysis of the communities

Figure 9.44: Averaged Box plot for all communities, Canada
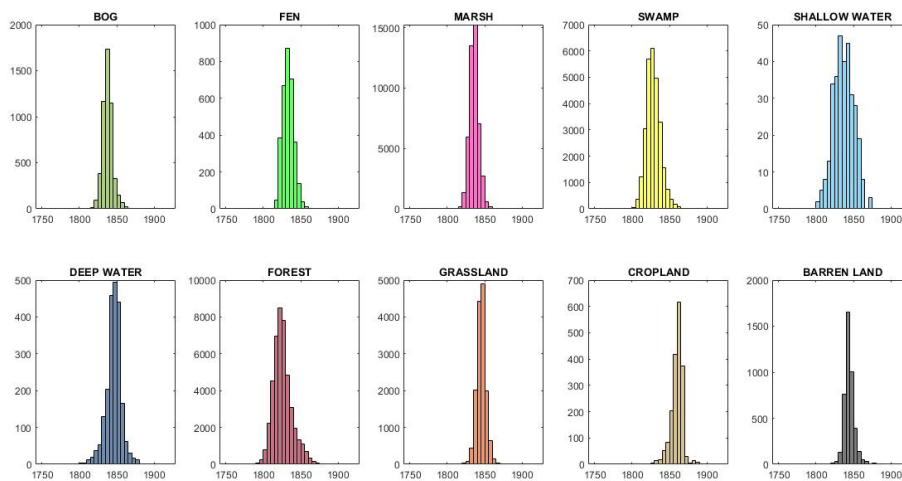
Figure 9.45: Histogram for all the XCH4 S-5p data
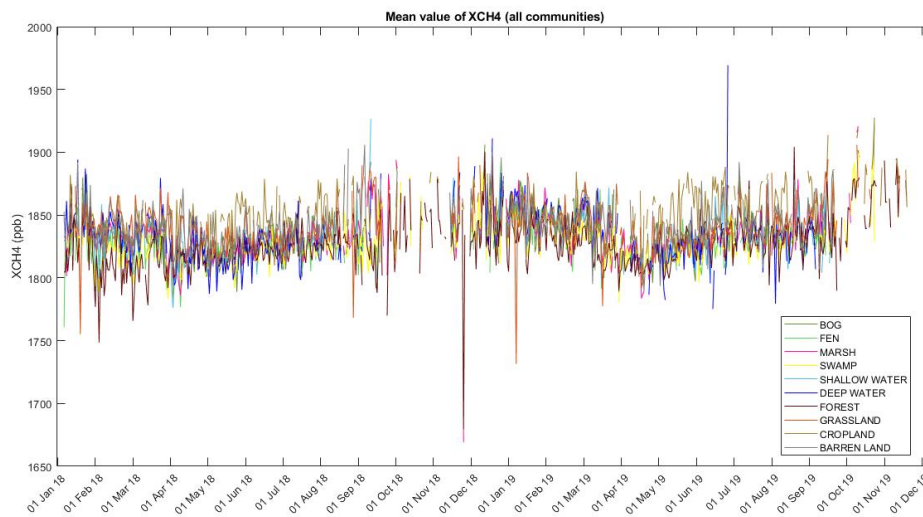
Figure 9.46: Mean value plots for all the XCH4 data, Canada

Part VI

Appendix F

9.6 Hydrometer metric for Turloughs

This section contains all the supplementary data from Chapter 7.
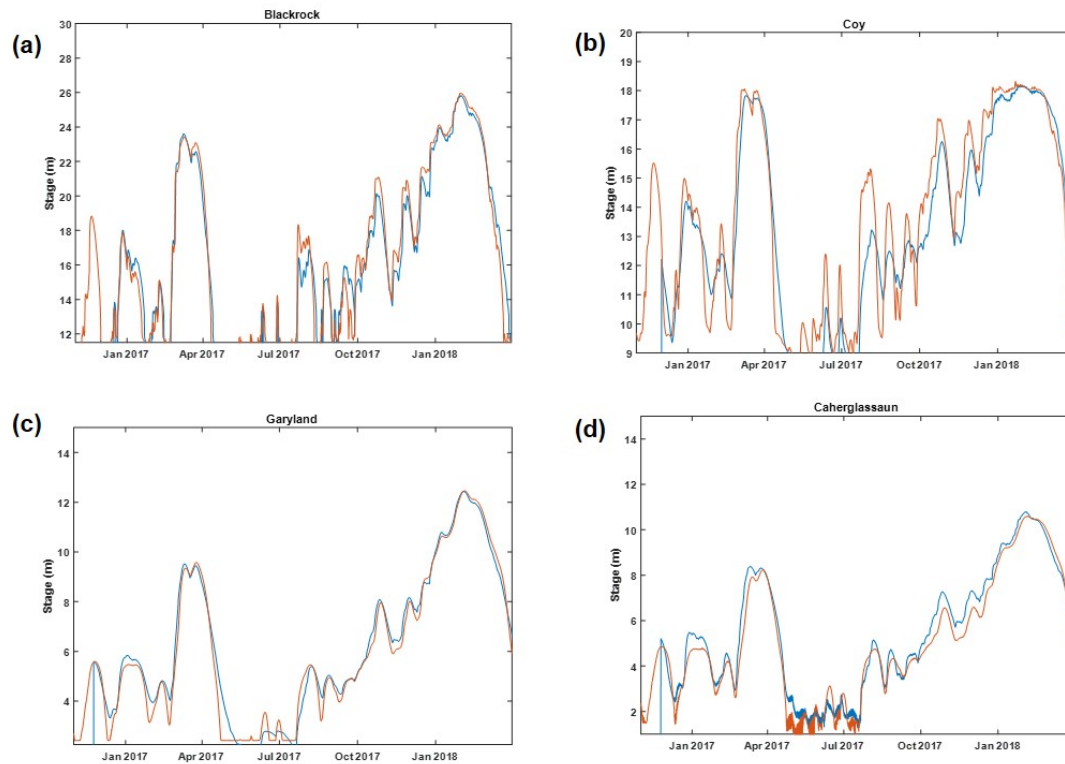
Figure 9.47: Calibration plots for the 1D/2D Infoworks ICM karst pipe network model calibration period between 01/11/2016 and 31/03/2018 at the 4 turlough locations: (a) Blackrock, (b) Coy, (c) Garryland (d) Caherglassaun.
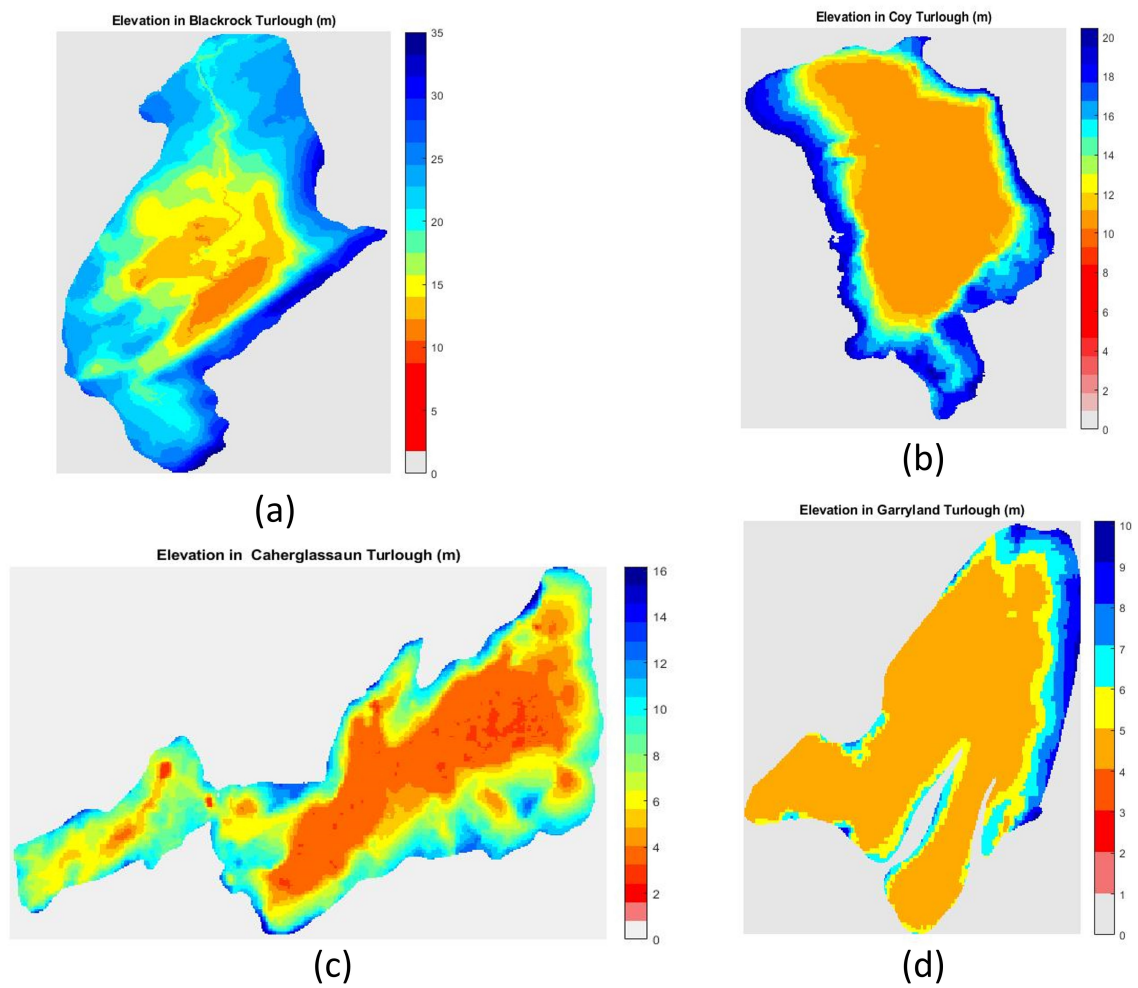
Figure 9.48: Actual topography of turloughs (a) BL (b) CH (c) CY (d) GL

The mean duration and depth under flood for all the turloughs are as follows:
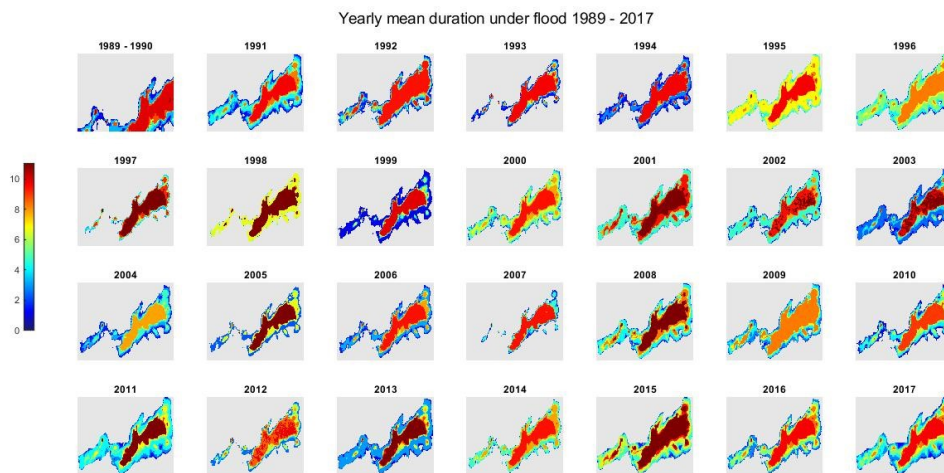
Figure 9.49: Mean duration under flood for all 28 years - CH
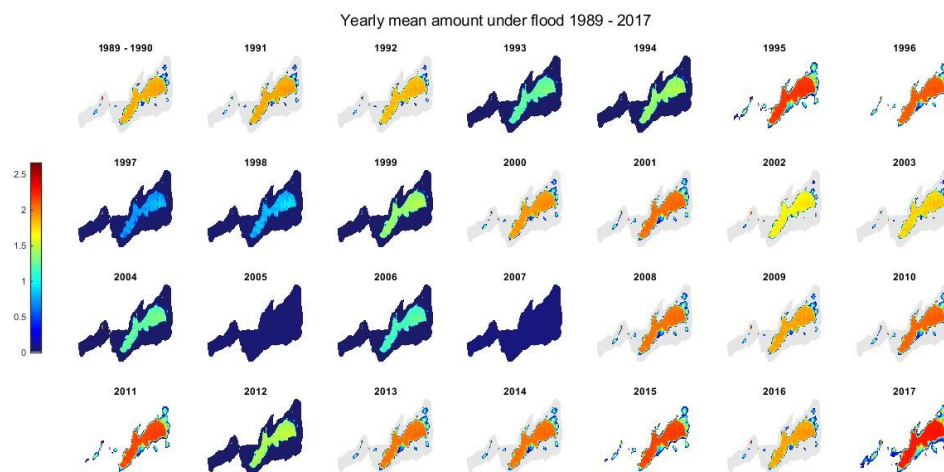
Figure 9.50: Mean deth under flood for all 28 years - CH
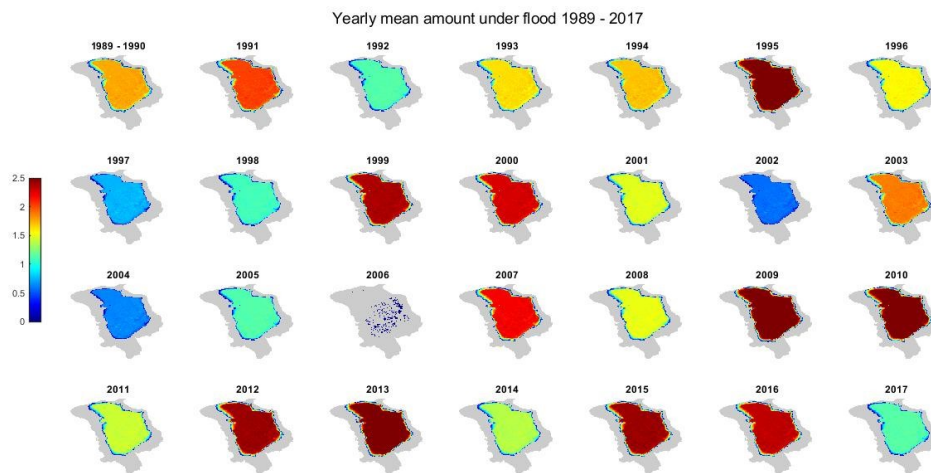
Figure 9.51: Mean duration under flood for all 28 years - CY
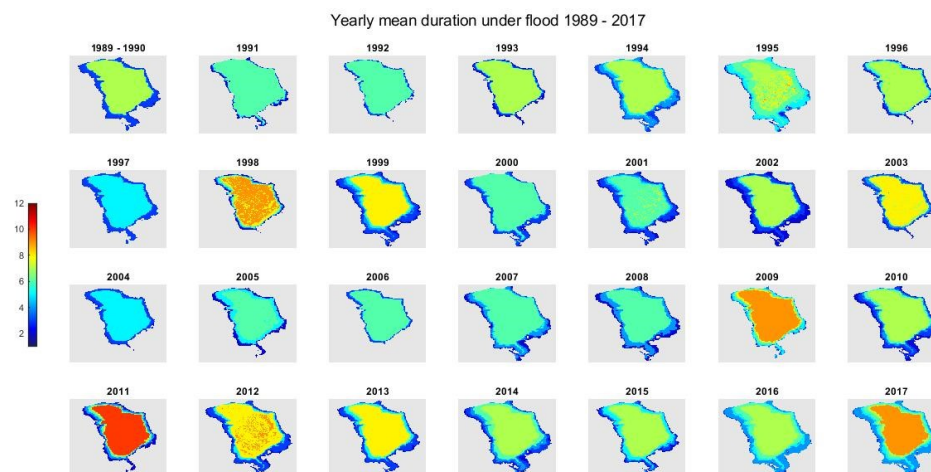
Figure 9.52: Mean depth under flood for all 28 years - CH
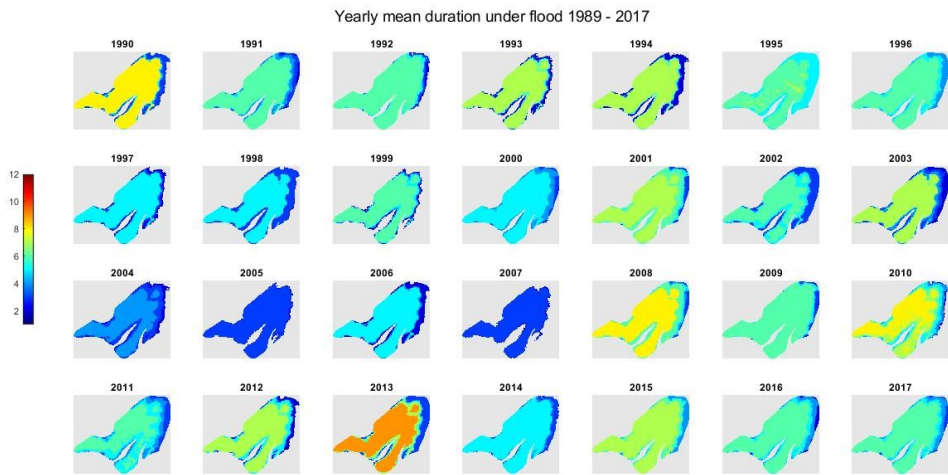
Figure 9.53: Mean duration under flood for all 28 years - GL
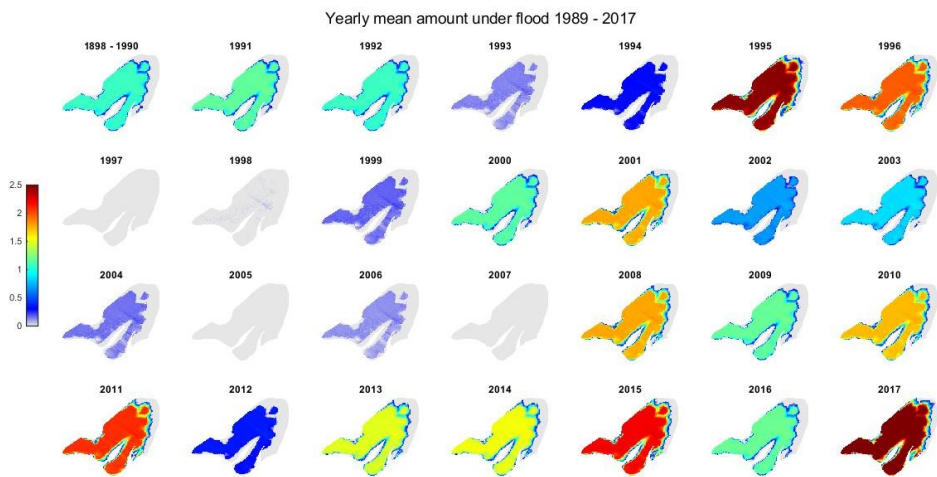
Figure 9.54: Mean depth under flood for all 28 years - GL
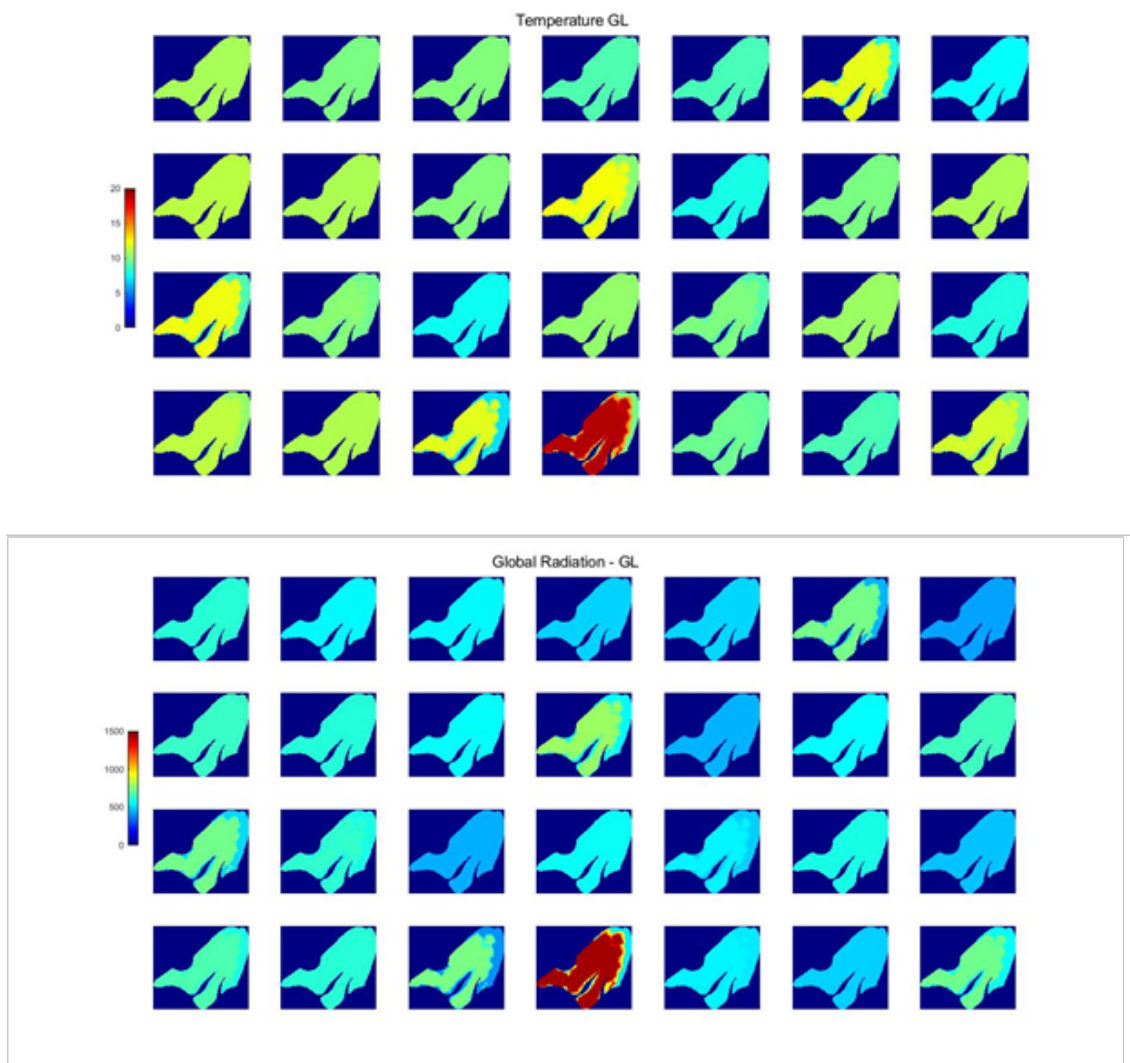
Figure 9.55: Mean global radiation and temperature for all 28 years - GL

The mean global radiation and temperature when the turlough comes out of the flood, for each of the turloughs is presented as follows:
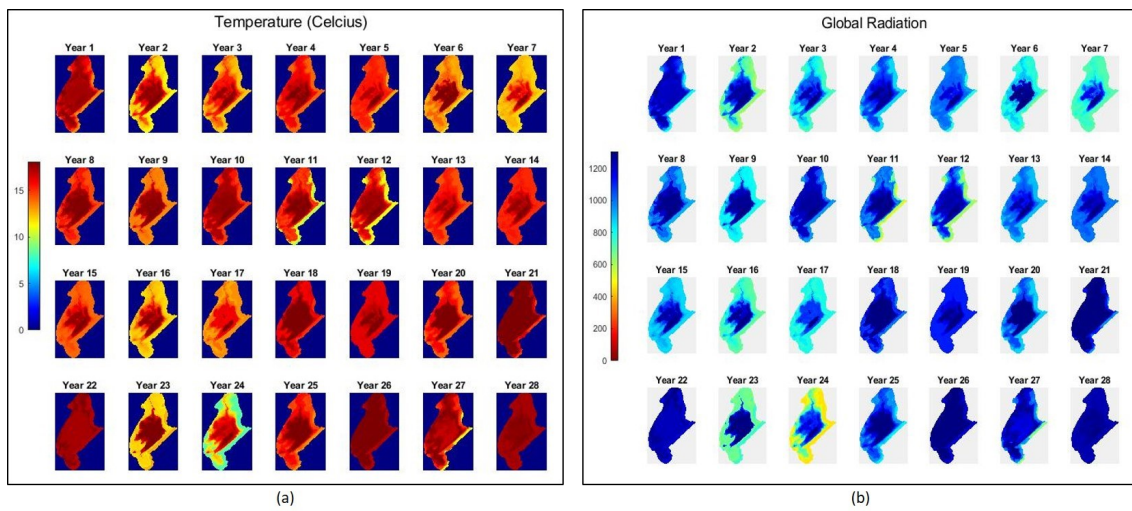
Figure 9.56: Mean global radiation and temperature for all 28 years - BL
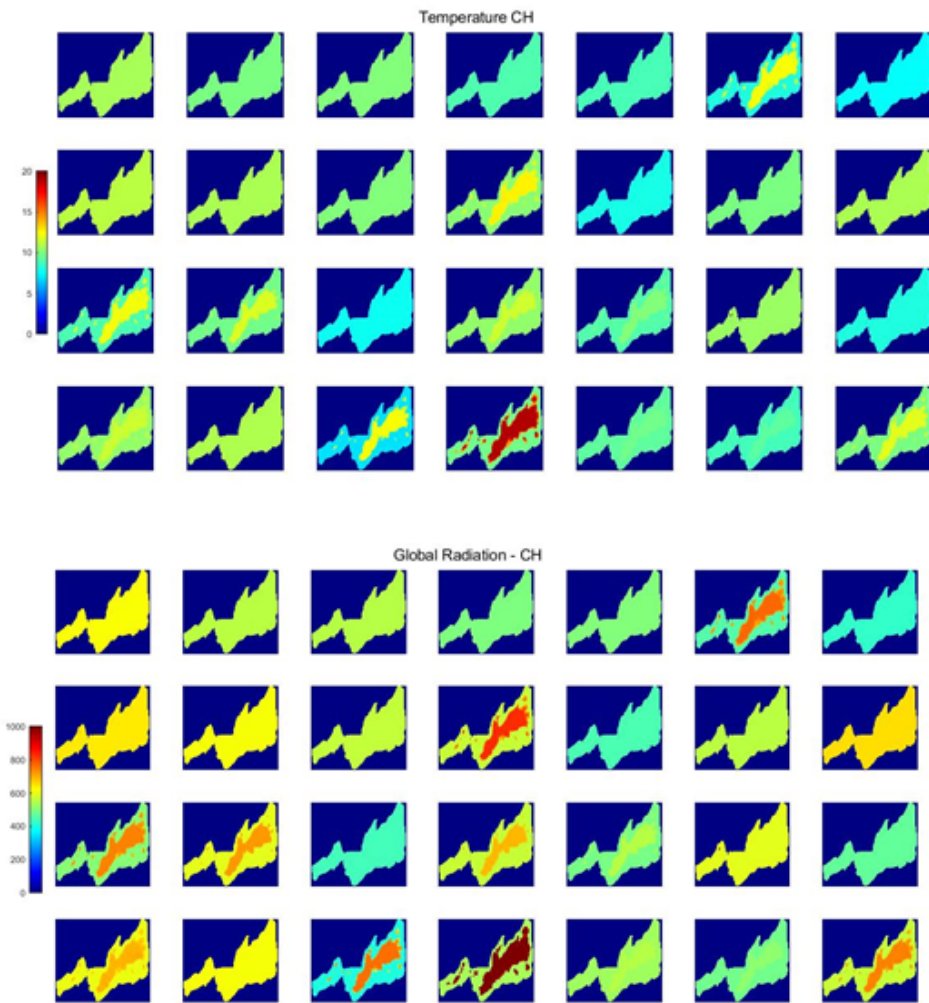
Figure 9.57: Mean global radiation and temperature for all 28 years - CH
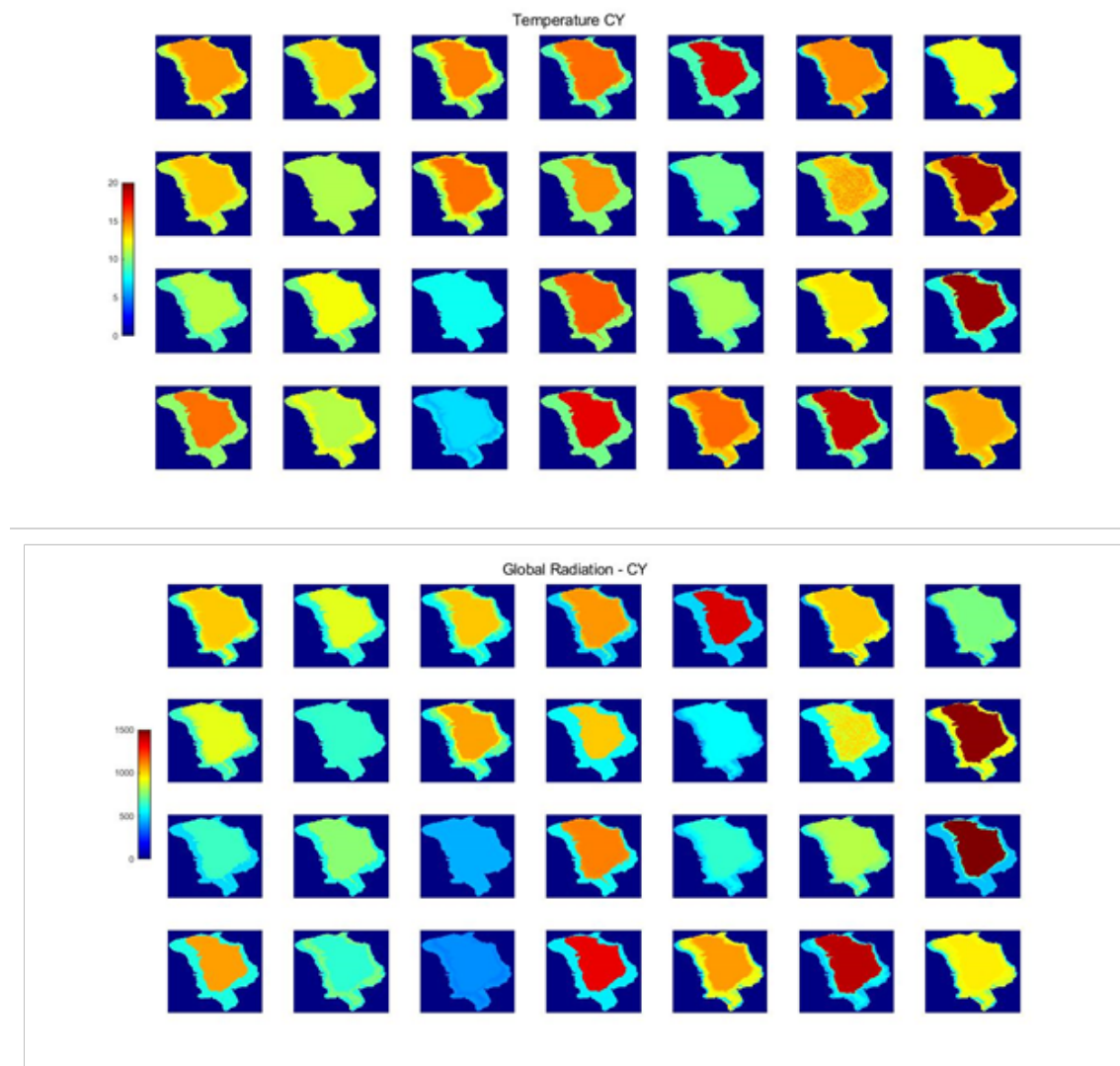
Figure 9.58: Mean global radiation and temperature for all 28 years - CY

The distribution of temperature across the vegetation communities in Blackrock turlough across all 28 years is shown as:
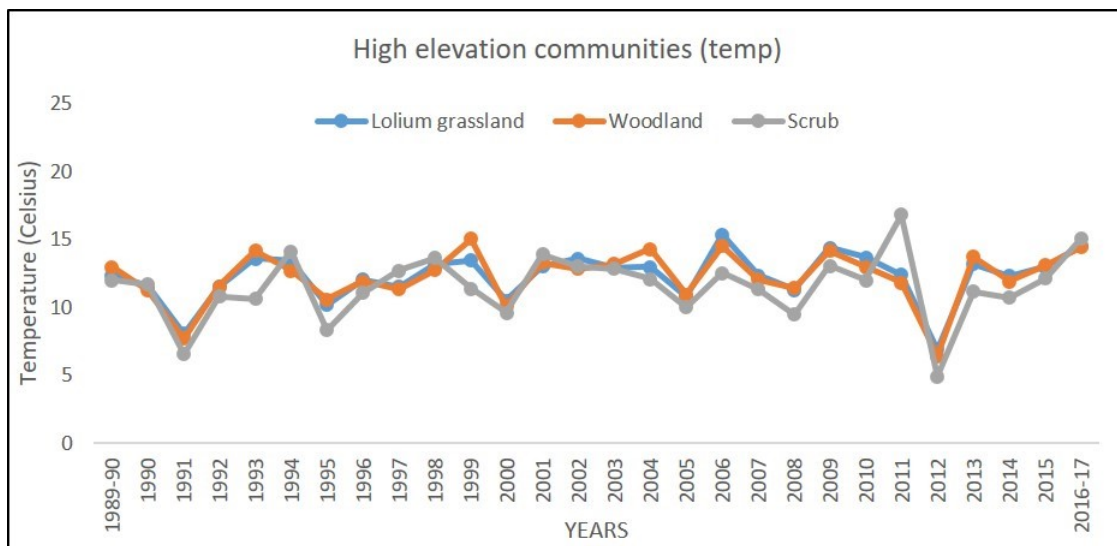
(a)
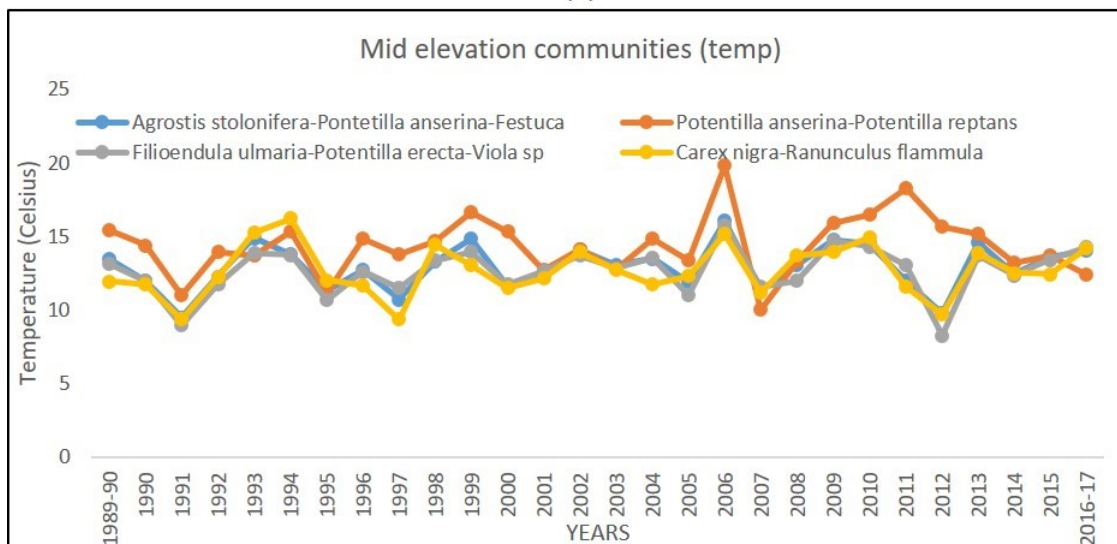
(b)
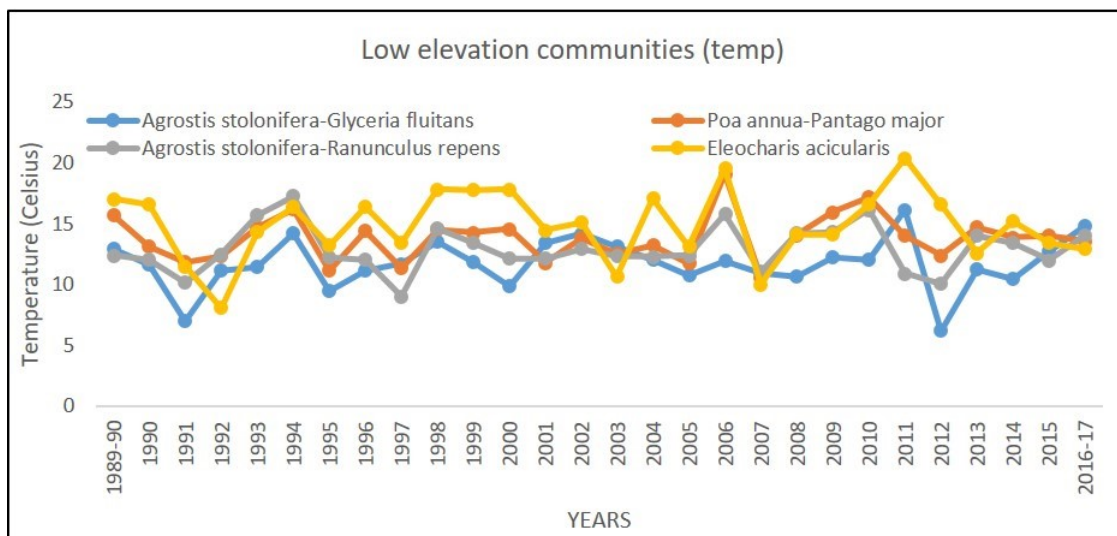
(c)

Figure 9.59: Mean temperature for the week/month after which the different communities came out of flood for the year

The relationship between the variables is shown as:

	DURATION	DEPTH	GLOBAL RADIATION	TEMPERATURE
DURATION	Correlation Coefficient = 1.00 $R^2 = 1.00$	Correlation Coefficient = 0.962 $R^2 = 0.96$	Correlation Coefficient = 0.862 $R^2 = 0.91$	Correlation Coefficient = 0.789 $R^2 = 0.91$
DEPTH	Correlation Coefficient = 0.962 $R^2 = 0.97$	Correlation Coefficient = 1.00 $R^2 = 1.00$	Correlation Coefficient = 0.938 $R^2 = 0.95$	Correlation Coefficient = 0.937 $R^2 = 0.96$
GLOBAL RADIATION	Correlation Coefficient = 0.829 $R^2 = 0.50$	Correlation Coefficient = 0.938 $R^2 = 0.90$	Correlation Coefficient = 1.00 $R^2 = 1.00$	Correlation Coefficient = 0.997 $R^2 = 0.99$
TEMPERATURE	Correlation Coefficient = 0.789 $R^2 = 0.45$	Correlation Coefficient = -0.937 $R^2 = 0.95$	Correlation Coefficient = 0.997 $R^2 = 0.99$	Correlation Coefficient = 1.00 $R^2 = 1.00$

Figure 9.60: Regression analysis and correlation between all four parameters averaged for all turloughs

	BL		CH		CY		GL		
DURATION (months)	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	AVERAGED RANGE
Lolium grassland	2	5	1	4	1	3	N/A	N/A	[1 -5]
Agrostis stolonifera-Pontetilla anserina-Festuca	2	5	1	5	0	4	N/A	N/A	[1 -5]
Carex nigra-Ranunculus flammula	1	5	N/A	N/A	N/A	N/A	N/A	N/A	[1 -5]
Agrostis stolonifera-Glyceria fluitans	1	5	N/A	N/A	N/A	N/A	N/A	N/A	[1 -5]
Poa annua-Pantago major	2	6	3	8	N/A	N/A	N/A	N/A	[2 -8]
Agrostis stolonifera-Ranunculus repens	1	5	N/A	N/A	4	8	3	9	[3 -8]
Potentilla anserina-Potentilla reptans	4	7	1	5	1	5	N/A	N/A	[2 -6]
Filioendula ulmaria-Potentilla erecta-Viola sp	2	5	1	5	1	5	2	6	[2 -6]
Eleocharis acicularis	6	11	5	11	5	10	3	9	[5 -11]
Carex nigra-Carex panicea	N/A	N/A	N/A	N/A	N/A	N/A	1	5	[1 -5]
Woodland	1	4	1	4	1	5	1	5	[1 -5]
Scrub	1	4	1	4	1	5	1	5	[1 -5]
	BL		CH		CY		GL		
DEPTH (in meters)	MINIMUM	MAXIMUM	MINIMUM	MAXIMUM	MINIMUM	MAXIMUM	MINIMUM	MAXIMUM	AVERAGED RANGE
Lolium grassland	0.43	1.38	0.06	1.12	0.18	0.74	N/A	N/A	[0.22 - 1.08]
Agrostis stolonifera-Pontetilla anserina-Festuca	0.47	2.72	0.06	0.70	0.00	0.00	N/A	N/A	[0.27 - 1.71]
Carex nigra-Ranunculus flammula	0.00	0.00	N/A	N/A	N/A	N/A	N/A	N/A	0.00
Agrostis stolonifera-Glyceria fluitans	0.00	0.00	N/A	N/A	N/A	N/A	N/A	N/A	0.00
Poa annua-Pantago major	0.45	2.07	0.01	2.15	N/A	N/A	N/A	N/A	[0.23 - 2.11]
Agrostis stolonifera-Ranunculus repens	0.00	0.00	N/A	N/A	0.02	2.15	0.01	2.70	0.00
Potentilla anserina-Potentilla reptans	0.47	4.72	0.11	1.86	0.09	1.59	N/A	N/A	[0.22 - 2.72]
Filioendula ulmaria-Potentilla erecta-Viola sp	0.43	1.38	0.09	0.73	0.34	1.08	0.02	1.56	[0.22 - 1.19]
Eleocharis acicularis	2.72	3.72	0.20	3.24	0.01	3.01	0.01	2.87	[0.74 - 3.21]
Carex nigra-Carex panicea	N/A	N/A	N/A	N/A	N/A	N/A	0.04	0.42	[0.04 - 0.42]
Woodland	0.00	0.00	0.32	0.71	0.00	0.55	0.05	1.58	[0.13 - 0.95]
Scrub	0.00	0.00	0.32	0.71	0.32	1.30	0.08	0.80	[0.24 - 0.94]

Table 9.1: Summary table of metrics for different communities

	last 10 years - Vegetation Survey												last 10 years - Sentinel-2 survey												
	AMT (meters)				DUR (months)				FREQUENCY				AMT (meters)				DUR (months)				FREQUENCY			
	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE	MIN	MEAN	MAX	RANGE
Open water	1.55	2.01	2.41	[1.55 - 2.41]	6.49	7.73	9.46	[6.4 - 9.4]	1	1.54	2.99	[1 - 2.99]	1.55	1.99	2.4	[1.55 - 2.40]	6.48	7.71	9.42	[6.4 - 9.4]	1	1.54	2.99	[1 - 2.99]
Eleocharis acicularis	1.68	2.46	3.1	[1.68 - 3.10]	6.18	7.55	9.67	[6.1 - 9.6]	1	1.58	3.00	[1 - 3.00]	1.42	2.2	2.84	[1.42 - 2.84]	5.98	7.29	9.19	[5.9 - 9.1]	1	1.58	3.00	[1 - 3.00]
Potentilla anserina-Potentilla reptans	1	1.54	2.11	[1.00 - 2.11]	4.17	5.11	5.69	[4.1 - 5.6]	1	1.43	2.24	[1 - 2.24]	0.44	0.9	1.93	[0.44 - 1.93]	3.66	4.66	5.05	[3.6 - 5.0]	1	1.43	2.24	[1 - 2.24]
Agrostis stolonifera-Ranunculus repens	1.01	1.49	1.86	[1.01 - 1.86]	4.62	5.37	5.91	[4.6 - 5.9]	1	1.44	2.80	[1 - 2.80]	0.77	1.11	1.48	[0.77 - 1.48]	4.4	5.27	5.67	[4.4 - 5.6]	1	1.44	2.80	[1 - 2.80]
Poa annua-Pantago major	0.3	0.98	1.7	[0.30 - 1.70]	3.63	5.14	5.72	[3.6 - 5.7]	1	1.47	2.39	[1 - 2.39]	0.06	0.69	1.52	[0.06 - 1.52]	3.82	5.37	5.86	[3.8 - 5.8]	1	1.47	2.39	[1 - 2.39]
Agrostis stolonifera-Pontetilla anserina-Festuca	0.75	1.01	1.28	[0.75 - 1.28]	2.16	2.95	3.74	[2.1 - 3.7]	1	1.20	1.78	[1 - 1.78]	0	0.03	0.09	[0.00 - 0.09]	1.31	2.37	3.57	[1.3 - 3.5]	1	1.20	1.78	[1 - 1.78]
Filipendula ulmaria-Potentilla erecta-Viola sp	0.44	0.68	0.91	[0.44 - 0.91]	3.06	4.01	4.46	[3.0 - 4.4]	1	1.35	2.00	[1 - 2.00]	0.03	0.21	0.47	[0.03 - 0.47]	2.34	3.38	4.15	[2.3 - 4.1]	1	1.35	2.00	[1 - 2.00]
Agrostis stolonifera-Glyceria fluitans	0.38	0.92	1.46	[0.38 - 1.46]	1	2.47	3.78	[1.0 - 3.7]	1	1.17	2.66	[1 - 2.66]	NaN	NaN	NaN	[NaN - NaN]	NaN	NaN	NaN	[NaN - NaN]	NaN	NaN	NaN	[NaN - NaN]
Carex nigra-Ranunculus flammula	0.41	1.05	1.68	[0.41 - 1.68]	1.7	2.94	3.99	[1.7 - 3.9]	1	1.27	2.89	[1 - 2.89]	0.01	0.01	0.01	[0.01 - 0.01]	0.43	1.62	3.07	[0.4 - 3.0]	1	1.27	2.89	[1 - 2.89]
Carex nigra-Carex panicea	0.04	0.21	0.42	[0.04 - 0.42]	2.13	3.52	4.44	[2.1 - 4.4]	1	1.40	2.15	[1 - 2.15]	0	0.04	0.23	[0.00 - 0.23]	1.9	3.75	4.82	[1.9 - 4.8]	1	1.40	2.15	[1 - 2.15]
Lolium grassland	0.45	0.76	1	[0.45 - 1.00]	1.98	2.54	3.31	[1.9 - 3.3]	1	1.14	1.88	[1 - 1.88]	0.02	0.04	0.1	[0.02 - 0.10]	1.12	1.77	2.9	[1.1 - 2.9]	1	1.14	1.88	[1 - 1.88]
Woodland	0.28	0.7	1.14	[0.28 - 1.14]	2.46	3.22	3.79	[2.4 - 3.7]	1	1.24	1.86	[1 - 1.86]	0.01	0.2	0.5	[0.01 - 0.50]	1.79	2.61	3.5	[1.7 - 3.5]	1	1.24	1.88	[1 - 1.88]
Scrub	0.52	0.67	0.76	[0.52 - 0.76]	2.63	3.2	3.74	[2.6 - 3.7]	1	1.27	2.02	[1 - 2.02]	0.03	0.07	0.18	[0.03 - 0.18]	1.72	2.48	3.28	[1.7 - 3.2]	1	1.28	2.04	[1 - 1.88]
Limestone Grassland	0.31	0.44	0.73	[0.31 - 0.73]	2.26	2.68	3.3	[2.2 - 3.3]	1	1.17	1.35	[1 - 1.35]	0.01	0.02	0.05	[0.01 - 0.05]	0.21	1.67	3.18	[0.2 - 3.1]	1	1.17	1.35	[1 - 1.35]
Flooded Pavement	0.43	0.43	0.43	[0.43 - 0.43]	1.36	2.22	2.92	[1.3 - 2.9]	1	1.06	1.49	[1 - 1.49]	0	0.01	0.01	[0.00 - 0.01]	0.06	1.53	3.01	[0.0 - 3.0]	1	1.06	1.49	[1 - 1.49]

Table 9.2: Metric describing averaged duration and depth for last 10 years using S2 maps for all communities over 4 turloughs.