Masters Program in Geospatial Technologies

Remote Sensing Based Multi-Criteria Blanket Bog Habitat Health Classification

An integrated comparative approach to assessing classification task performance of different blanket bog condition ranges using spatial cross validation and forward feature selection

Solenn Reeves Long

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Universität Münster

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Remote Sensing Based Multi-Criteria Blanket Bog Habitat Health Classification

An integrated comparative approach to assessing classification task performance of different blanket bog condition ranges using spatial cross validation and forward feature selection

by

Solenn Reeves Long

Master Dissertation presented as partial requirement for obtaining the Master's Degree in Geospatial Technologies

Supervised by

Christian Knoth, University of Münster

Marco Octávio Trindade Painho, NOVA Information Management School

Ana Cristina Costa, NOVA Information Management School

February, 2026

INDEX

INDEX OF FIGURES

INDEX OF TABLES

STATEMENT OF INTEGRITY

I declare that the work described in this document is my own and not from someone else. All the assistance I have received from other people is duly acknowledged and all the sources (published or not published) are referenced.

This work has not been previously evaluated or submitted to NOVA Information Management School or elsewhere. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 23/02/2026

Solenn Reeves Long

USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

Tasks	NO	YES	Generative Artificial Intelligence tools
Better understand issues related to the research		X	ChatGPT, Gemini
Summarizing text from bibliography / resources			
Summarizing the method(s) used			
Translating text			
Grammar check		X	ChatGPT
Paraphrase or rewriting text from other people / resources			
Coding in R, Python, etc.		X	ChatGPT, Gemini
Get help on a software		X	ChatGPT
Creating and editing images, maps, videos, etc.			
Data analysis			
Specify below other tasks not mentioned above:			

DEDICATION

This research is dedicated to my Grandad George who through his love of Irish boglands and membership with the Irish Peatland Conservation Council showed a respect for nature and ecosystems that are often over-looked.

ACKNOWLEDGMENTS

Thank you to the Wicklow Uplands Council and to Lorna Kelly for sharing the crucial ground-truth data that was needed for this research. Thank you also to Faith Wilson who carried out the surveys.

I would like to acknowledge Dr. John Connolly, my first GIS professor at Trinity College, and who's work on peatland monitoring has been very influential on this work. Also Dr. Wahaj Habib who was helpful in guiding the initial direction of this research.

I am extremely grateful to my support network who have helped, guided and provided me with excellent advice, my parents, my boyfriend, my friends and my sister who ventured into the Wicklow Mountains with me and happily went off trail to visit Liffey Head bog.

I would also like to acknowledge the work of Re-Wild Wicklow who have put tremendous effort into restoring the Wicklow Uplands with the support of the NPWS.

REMOTE SENSING BASED MULTI-CRITERIA BLANKET BOG HABITAT HEALTH CLASSIFICATION

An integrated comparative approach to assessing classification task performance of different blanket bog condition ranges using spatial cross validation and forward feature selection

ABSTRACT

Peatlands offer a unique opportunity for carbon sequestration so long as they remain in good condition. Today blanket bog in Ireland are degraded and in poor condition. Although much work has been done on applying machine learning and remote sensing to map conditions of peatlands such as raised bogs, blanket bog remain under researched in this area. This study undertakes a classification task of three different degradation classification systems applied to blanket bogs and using an experimental design explores how different sets of predictors impact the performance of a random forest classifier algorithm. Spatial cross validation and area of applicability were used to enhance interpretation of predicted maps and performance metrics. It was found that the binary classification of peat-forming and non-peat forming performed the best with an accuracy of 80%. Predicting the conservation status of blanket bogs remains challenging due to lack of representation in sample data between classes. A wide range of data types that represent vegetation, moisture and bare peat characteristics enhances classification performance. With better definitions of habitat extent and more representative sample data there is potential for blanket bog condition classification.

KEYWORDS

Habitat mapping; Remote sensing; Sentinel-1; Sentinel-2; Peatland degradation; Random Forest; Spatial cross validation; Area of applicability; Blanket bog

Sustainable Development Goals (SGD):

1. INTRODUCTION

1.1 MOTIVATION

Peatlands represent just 3% of global land surface, however they store up to 30% of the global soil carbon stock (Joosten et al, 2009), (Parish et al, 2008). The literature produced in recent decades has been unambiguous and clear on the need to restore peatlands (DeWaard et al, 2024), (UNEP, 2022), (Lindsay et al, 2010).

Peatlands offer unique environments that can support rare plant species such as Bog myrtle *Myrica gale* and Bog Orchid *hammarbya* (Conaghan, 2000). They also provide a wide array of ecosystem services that in recent years have been associated with economic value. This value, however, is dependent on the condition of the habitat (Stout et al, 2023), (Farrell et al, 2021). Ecosystem services are the goods and services provided by natural systems that provision for the fulfilment and well-being of human life (Daily, 1997). Peatlands specifically, provide clean water, climatic and flood regulation, pharmaceutical ingredients and more (Bonn et al, 2016). Under EU Legislation protection regimes must be in place to conserve certain habitats and species (Council Directive, 1992).

Peatlands form under very specific conditions where both the quantity and year-round distribution of rainfall are sufficiently high (UNEP, 2022). Specifically, blanket bogs in Ireland formed after tree clearance led to the formation of a hard pan, which altered the landscape's hydrology and created the waterlogged conditions necessary for the peat forming sphagnum moss, to thrive (Conaghan et al, 2000). The presence of sphagnum moss maintains the high-water table and creates an acidic environment, only allowing vegetation adapted to the waterlogged and acidic environment from growing. The high-water table, low oxygen and acidic environment prevent the decomposition of vegetation. Overtime organic material builds up and is buried thus storing large quantities of carbon in the form of peat. Due to this carbon storing ability (Kimmel & Mander, 2010) they play a crucial role in mitigating climate change and its effects.

Peatlands are distributed globally but are concentrated in northern regions where the necessary climatic conditions (high rainfall, moderate temperatures) prevail (UNEP, 2022). Ireland has a particularly high percentage of this land but in a study conducted by de Waard et al, (2024), the authors reported over 50% disturbed blanket bog in the Wicklow Mountains and estimated the national range to be up to 80% (de Waard et al, 2024). It was found that peatland disturbance on mountain blanket bogs was concentrated in areas of higher altitude and slope.

Peatland's ability to store carbon is intimately related to their condition. Disturbance in the form of peat extraction, grazing, burning or establishment of coniferous plantations affect the specific conditions that allow for the proper functioning of peatlands (Connolly, 2011). This results in greenhouse gas (GHG) emissions instead of sequestration (Renou-Wilson et al, 2025), (DeWaard et al, 2024), (UNEP, 2022).

Studies have found that 0.3% of degraded peatland can cause up to 5% of global CO2 emissions (Joosten, 2009). Restoring peatlands is crucial to preventing further GHG emissions and to increase carbon sequestration through the natural functioning of active peatland ecosystems. Mapping and quantifying the extent of degraded peatlands and improving our understanding of degradation dynamics is imperative to inform policy and steer data driven actions.

The National Parks and Wildlife Service (NPWS) are responsible for producing resources and documentation on the current state of habitats particularly those that fall under the EU's Habitat Directive. Blanket bog (7130) is one such priority habitat however, up until the 2025 report it was only considered a priority habitat if *actively peat forming (EUNIS, 2013). The 2025 report now states that blanket bog whether peat forming or not falls under the EU Habitats Directive Annex I habitats (NPWS, 2025). The distinction between peat-forming and not is important due to the GHG emissions/ sequestration dynamic dependent on the condition associated with this habitat.

Ingle et al, (2024) have mapped ecotopes on raised bogs using machine learning to measures methane fluxes associated with them. It remains challenging to define the GHG fluxes associated with varying degrees of condition for blanket bogs. However, it is nevertheless an important direction of research to investigate the extent to which the

condition of these habitats can be classified using machine learning for such purposes as GHG flux measurements. Improved understanding of landscape classification accuracy enables the development of temporal monitoring strategies and more robust estimates of GHG emissions and sequestration across habitats. These insights can support the integration of landscapes into carbon credit systems, creating potential income streams and inform targeted degradation prevention and restoration planning.

1.2 KNOWLEDGE GAPS

Pazúr et al (2021) highlight the gap in the literature that leverages global satellite imagery for use in landscape ecology mapping and monitoring. While de Waard et al. (2024) have identified the need for multidimensional approaches to peatland degradation mapping. Although there exists in the literature, studies that take into account hydrology, vegetation and peat exposure equally when assessing degradation, these were in the minority (de Waard et al. 2024). With the increasing availability and affordability of satellite imagery, there are growing opportunities to use these datasets to study environmental systems and landscapes.

Cruz et al. (2024) established methods for mapping temperate upland habitats using satellite imagery and machine learning. Their methods showed promising results and will be informative for this study. However, their study focused on habitat mapping rather than mapping peatland degradation. Karlson & Bastivken (2023) have used remote sensing and machine learning to classify wetland types in northern latitudes, however their classification system did not address degradation or condition but rather wetland type, e.g. swap, fen or bog. Other studies have used various aerial sensors to detect erosion gullies (signifiers of extensive degradation) and general degradation (Evans & Lindsay, 2010), (Carless et al, 2019) but so far there has not been a study conducted that focuses on mapping the extent of blanket bog degradation using satellite imagery in temperate regions.

Although much progress has been made on mapping blanket bog extents and to a certain degree their conditions (Minasny et al, 2024), often using areas of erosion as proxies for their condition (Evans & Lindsay, 2010), no study has so far attempted to classify and map upland blanket bogs using existing and official habitat classification systems. Additionally, there have been studies on classifying peatlands and their

condition using remote sensing and machine learning, however the majority of this work has been focused on raised bogs (Ingle et al, 2024), (Bhatnager et al, 2018).

1.4 Research questions

This study aims to address key gaps in the literature on the analysis of degradation in mountain blanket bogs, as outlined above. Using Sentinel-1 and Sentinel-2 datasets, and ancillary data derived from DEMs this study will examine the dynamic relationships between selected predictors and assess how they can be used in combination to effectively identify and map the extent of blanket bog degradation in the Wicklow Mountains.

We begin by asking: To what extent can different degradation classification systems be classified using traditional machine learning algorithms and a broad suite of optical, SAR and elevation derived predictors? Additionally, we address the sub-question: what challenges are associated with applying already existing methods of classifying raised bog degradation to upland blanket bogs?

Second, this study will look at what combination of predictors are most effective for classifying different peatland degradation classification systems? Within this we try to determine whether the three aspects of peatland degradation, hydrological processes, vegetation health, bare peat exposure, as defined by de Waard et al (2024) are reflected in the model's selected predictors.

1.5 Scope and Focus

This study applies a Random Forest (RF) classifier to a blanket bog landscape exhibiting varying degrees of degradation, with the aim of classifying the landscape according to three established and widely used peatland and peatland degradation classification systems: Fossitt (blanket bog, eroding bog and cutover bog), Annex I (intact blanket bog vs. non intact), Conservation Status (favourable, inadequate and bad). Rather than evaluating classification methodologies themselves, the focus of this work is on comparing model performance across different classification systems when applied to the same study area and predictor set. Within this scope, the study addresses the research questions outlined above and assesses the applicability of this

approach for improving the understanding and monitoring of blanket bog condition, an issue of clear relevance to peatland conservation and management.

To do this, this study will apply the CAST package in RStudio, an extension of the caret framework that allows for training, tuning, and evaluating machine-learning models. This package was chosen to create a spatially aware model that has been shown to reduce overfitting and over optimistic error assessment. Additionally, this package implements a forward feature selection algorithm that identifies predictors contributing most to improvements in model performance (Meyer et al, 2026). This model will be applied to four different experiments, testing model performance on different predictor sets. Each experiment will be applied to the three classification systems. Finally, evaluation of model performance using cartographic visualisations and area of applicability (AOA) maps will be conducted.

2. LITERATURE REVIEW

2.1 BLANKET BOGS

2.1.1 DEGRADATION CLASSIFICATION SYSTEMS

Throughout the literature different classification systems of peatland habitat degradation have been used to measure the health of bog ecosystems depending on the specific context of the study. Geerling et al (2002) used a five-tier degradation category system from “severely degraded” to “healthy sward”. Their classification system was based on specific characteristics observed in the field such as vegetation height and percentage of bare peat exposed (Geerling et al, 2002).

Still others have developed their own classification system of observable degradation on peatlands such as Cutler et al’s (2002) paper that used a supervised maximum likelihood classification using a neural network architecture to identify and quantify peat erosion. They used a four-tier classification system: well-humified, poorly humified burned and washed peat. These classes were developed from field work and based on observations in the regional context of the study. They were less tied to current policies that now inform condition monitoring. Their classification system is more contextually and regionally specific.

Blanket bogs and their condition can be been categorised under the Fossitt habitat classification system that distinguishes upland blanket bog (PB2) from eroding blanket bog (PB5). Under this classification system, cutover bog (PB4) is also a distinct category however, this class is reserved for manually extracted peatbogs and is more associated with raised bogs (Fossitt 2000). Eroding blanket bog is more commonly associated with upland and lowland blanket bogs and has a number of causes, some natural and some indirectly or directly anthropogenic. Upland blanket bog as compared to eroding bog is a useful categorization of degradation in the context of identifying blanket bog integrity though intermediate levels would add to a more dynamic understanding of degradation

The classification system and distinctions that are used in the EU Habitats Directive are more widely applicable. The Conservation Status (favourable, inadequate, bad) classification system is in fact used for all habitats and is not specific to peatlands (Council Directive, 1992). However, under the EU habitats directive the distinction between these classes for each habitat is defined and used by field ecologists when assessing area.

Another classification system that has been used is a binary classification system of actively peat forming or not. This distinction originates from the EU Habitats directive and is specific to peatlands. It describes active peat forming as “still supporting a significant area of vegetation that is normally peat forming”. It is less descriptive than the previous classification system but nevertheless an informative and useful classification when considering GHG fluxes (de Waard et al, 2024), (Ingle et al, 2023), (Lindsay et al, 2010). Although this is an important facet of blanket bog degradation, establishing whether a defined area is peat forming or not remains extremely challenging using remote sensing technology. Therefore, another way to approach this binary system is looking at the habitat as either intact blanket bog, which may or may not be peat forming, vs non intact which is definitely not actively forming peat. An area is considered to be intact blanket bog if the field surveyor deemed it appropriate to label it with the Annex I habitat code 7130. That is the case when enough indicators are present that the area falls under 7130 though it may or may not be peat-forming.

2.1.2 Threats and Pressures

In the literature, certain pressures emerged as having a significant impact on the ecological function of blanket bogs and there has been broader discussion on potential mitigation and monitoring efforts. Looking specifically at the Irish context the main drivers of degradation on blanket bogs are drainage for peat extraction, development of conifer plantations, overgrazing, and burning (NPWS, 2017b), (de Waard et al, 2024). Geerling et al (2002) make reference to disturbance of vegetation that can start a chain of events culminating in bare peat exposure and eventually erosion. Large & Hamilton (1991), along with the aforementioned pressures, also refer to slope and climate. A more recent publication by Flynn et al, additionally raises the issue of wind farms as a potential threat to blanket bog integrity (Flynn et al, 2021).

In summary pressures and threats to blanket bogs can fall into three categories, hydrological impacts, vegetation impacts and peat integrity impacts (de Waard et al, 2024). For instance, overgrazing would fall into vegetation impacts, peat extraction and burning would fall under peat integrity impacts and drainage for agricultural or extraction would fall under hydrological impacts. Threats can fall into more than one category but for the purposes of this study it is useful to use these categories when later we discuss sensors and indices that can provide insight into the dynamic relationship between each of these categories and observable degradation.

2.2 Remote Sensing of ecosystems

Within the literature, extensive work has been produced using a wide variety of instruments in the remote sensing category to monitor ecological processes at various scales and resolutions (Toca et al, 2023), (Knoth et al, 2013), (Geerling et al, 2002), (Lendzioch et al, 2021). De Waard et al, (2024) have published a review on current technologies and recent research of remotely sensing peatland degradation. They mention a broad array of tools used such as lidar, synthetic aperture radar (SAR), optical satellite imagery, hyperspectral imagery, unmanned aerial systems and thermal imagery. Each of the approaches used in these studies are designed for different scales of analysis. For instance, Knoth et al, (2013) used UAVs to map vegetation species on about 200m^2 raised bog with a 1.5cm resolution.

De Waard et al, (2024) found that optical remote sensing data has been the dominant data source in peatland degradation research. SAR imagery has also proven to be particularly useful in monitoring peatlands (Barrett et al, 2016), (Lees et al, 2021) due to the lack of influence from cloud cover which is a pertinent issue in northern regions where peatlands tend to be concentrated. Additionally, SAR is sensitive to landscape scale moisture dynamics which, as discussed above, is an important factor in peatland health. It is clear from the building research that a combination of data from different sensors improves model classification (Barrett et al, 2016), (Karlson & Bastviken, 2023).

2.2.1. Optical – Sentinel-2

Certain optically derived indices and band ratios emerge as being useful to train machine learning models to distinguish between ecologically distinct areas of interest. NDVI, EVI, GNDVI, NDRE, and RDVI have been used through-out the literature due to their sensitivity to vegetation (Barrett et al, 2016), (Bhatnager et al, 2018). Many studies use a combination of multiple indices related to vegetation to enhance classification performance (Barrett et al, 2016), (Bhatnager et al 2018), (Ingle et al, 2023), (Mahdianpari et al, 2021).

Another aspect of peatland degradation is moisture levels which are correlated with water table. Gao's (1996) NDWI is formulated to target sensitivity to vegetation moisture as opposed to open bodies of water. It has been used in in multiple studies to enhance machine learning and classification tasks (Karlson & Bastivken, 2023), (Mahdianpari et al, 2021), (Bhatnager et al, 2018).

Exposed peat is another vital factor in assessing blanket bog degradation (Trippier et al, 2020). Optical imagery can be sensitive to bare or exposed peat. NDVI will represent vegetated areas with high scores and conversely low scores for bare earth. Additionally, the NIR and SWIR bands are sensitive to bare earth and can be used on peatland condition monitoring (Reynolds et al, 2025).

2.2.2 Synthetic Aperture Radar (SAR, Sentinel-1)

Synthetic Aperture Radar's (SAR) capacity to penetrate the surface has led to growing research into its use for monitoring soil moisture and water table levels as well as vegetation structure and health (Lees et al. 2021), (Betbeder et al, 2015). Much of the

research on peatland monitoring has focused on producing accurate models that capture the landscape-scale temporal dynamics of water table depth and soil moisture (Räsänen et al, 2022), (Bechtold et al, 2018). Less attention has been given to using these variables as an additional layer of information for more general peatland classification. Some examples of the latter include Barrett et al's 2016 paper, though their focus was on vegetation habitat mapping, not peatland degradation classification. Also, Karlson & Bastviken (2023) used temporal aggregates of VV and VH polarisation and their ratio as predictors for their model.

Information derived from SAR imagery is highly relevant to peatland degradation as it is sensitive to moisture dynamics, surface roughness and vegetation structure. C-band SAR imagery shows sensitivity to soil moisture until the soil is inundated in the backscatter coefficient (Shultz et al, 2023). Hu et al (2024) provide a systemic overview of Radar Vegetation Indices that have been developed to date.

2.2.3 Ancillary Data

In addition to optical and radar data, slope, elevation and other ancillary data sources provide valuable information on peatland degradation and erosion risk. Many publications have made reference to the location of blanket bog erosion being located on or near mountain peaks. Intuitively slope has a strong impact on blanket bog integrity as gravity plays a role in encouraging erosion of peat through high water flow velocity as well as downward pull of matter. This makes certain topographical data important in identifying areas of degradation vulnerability (Harris & Baird 2019). Additionally in the literature the presence of wind farms, drainage and conifer plantation have shown to have an impact on ecological functioning of blanket bogs. (Large & Hamilton, 1991). Minasnay et al (2024) mention the importance of fire as an indicator of peatland condition.

2.3 Machine Learning for Habitat Classification

Significant advances have been made in recent years on the application of machine learning for ecological landcover mapping (Cruz et al, 2024), (Matyukira and Mhangara, 2024), (Gu & Zeng, 2023). A wide variety of machine learning models have shown to be useful such as artificial neural networks (ANN) (Grappiolo et al, 2024) support vector machines (SVM), bagged tree ensemble classifier and random forest.

Recent reviews of current trends in machine learning approaches to landscape ecology have found that RF models are used most frequently though with increasing access to higher compute power ANNs are gaining prominence in this application (Stupariu et al, 2022).

In the case of landscape ecology through remote sensing approaches, machine learning techniques are invaluable for time-series monitoring tasks (Schultz et al, 2023). Since the use of machine learning techniques have become more widely accessible remote sensing and machine learning are almost invariably used in tandem. In the literature reviewed for this study, Geerling et al's (2002) paper is the only example where remote sensing was used without the support of machine learning techniques.

Upscaling techniques, which enable landscape characteristics to be represented at larger scales than is possible with traditional survey methods, are especially relevant to the focus of this research project. For example, Ingle et al's (2024) paper used a weighted approach to be applied to different ecotopes present on raised bogs to account for methane fluxes. They used a RF model to map ecotopes of the area of interest. Similarly, Bhatnager (2020) conducted a study on mapping vegetation communities of bogs in Ireland comparing different machine learning models. They incorporated the bagged tree model into their Vegetation Mapping Vegetation Communities algorithm as it outperformed other machine learning models including RF (Bhatnager et al, 2020). Despite mixed performance across studies, RF remains widely used in ecological remote sensing due to its robustness, ability to model complex non-linear relationships, and strong performance with high-dimensional environmental datasets.

RF algorithms can produce either regression or classification output. Regression has been successfully used to model hydrological dynamics at landscape scales (Räsänen et al 2022), (Koch et al 2019), (Lendzioch et al, 2021). However, classification approaches are more common for mapping habitats and their condition (Bhatnager et al, 2020), (Ingle et al, 2023). This is largely due to ground truth data derived from ecological surveys being categorical as opposed to continuous in nature (Barrett et al, 2015), (Artz et al, 2018). Despite some papers finding marginal improvement using models other than RF (Bhatnager et al, 2020), for the most part RF algorithms are the

most used model in the literature. This can be attributed the lack of assumptions, the flexibility, non-parametric nature and applicability to many different data types (Stupariu et al, 2022), (Ghazaryan, 2024).

This algorithm has also performed well against the most widely used evaluation metrics (DeLancey et al, 2019), (Ingle et al, 2023), (Barrett et al, 2016). A confusion matrix to derive accuracy, precision, recall (sensitivity), F1 Score, precision and sensitivity are commonly used to evaluate the output of RF algorithm performance (DeLancey et al,2019), (Stupariu et al, 2022), (Räsänen, et al, 2019).

Recent advances have been made to incorporate spatial aspects to the application of RF algorithms in geographic disciplines. Georganos et al (2021) outline the limitations of applying ML algorithms without considering spatial heterogeneity and dependencies that may exist in the data. They put forward a method to incorporate spatial components into the learning algorithm to account for spatial heterogeneity that is likely to exist in the data (urban populations). They do this by building small local models using spatially weighted subsets of the data and then aggregating the local models to the global system. It is a modelling strategy that incorporates spatial heterogeneity that can reduce the dominance of spatial dependence during training. Although this is a strong methodology that enhances exploratory power of RF models in varying relationships between dependent and independent variable across space, this method performs best on spatially exhaustive data as is the case in Georganos et al (2021) and Yin et al (2026) and is less suited to highly clustered sample data that is derived from isolated polygons. This is because there may not been enough variance in the neighbourhoods to establish relationships and thus a valid local forest.

Despite these limitations, other work has been done to address this exact issue of map accuracy when sample data is highly clustered, unevenly distributed or has high degrees of density variation. Meyer et al (2018) propose a method that incorporates spatial cross validation along with selecting predictor variables to remove unhelpful or counterproductive predictors to the prediction process. Their approach produces more realistic map accuracies and information on what specific predictors are useful for accurate predictions. This avenue of research has been further developed to also produce area of applicability (AOA) maps. This is calculated using the dissimilarity index (DI) and local point density (LPD) (Schumacher et al, 2024). These studies have

showed promising results in their application to ecological mapping tasks though limitations include under-optimisitic validation statistics estimates (Wadoux et al, 2020).

2.4 Applications to Peatland / Degradation Studies

Historical human interactions with peatlands vary from region to region and is reflected in the different management practices and policies developed across the globe (Parish et al, 2007). In the literature this has highlighted the need to address peatland management locally, particularly with regards to monitoring causes of degradation and their impact on peatland habitats. For example, in the UK there exists specific land management practices for the recreational sport of grouse hunting. These practices which involve controlled burning of vegetation have a negative impact on peatland habitat integrity and can culminate in gully erosion. Ireland on the other hand has had a long history of extracting peat as a fuel source and in more recent times on extraction on an industrial scale. This practice has required the creation of man-made drains on peatlands again disrupting the hydrology of the landscape and causing degradation.

Diversity in management practice as well as geomorphological characteristics necessitates a localized approach to monitoring of these habitats. Studies have addressed this need and used different methods and technologies in their research. For example, Habib et al (2024) used high resolution aerial imagery to work toward drain mapping on raised bogs in Ireland using a convolutional neural network and Robb et al (2023) conducted a similar study in the Scotland. These studies reported an accuracy of 80% and 79% respectively. Other studies have focused on gully erosion mapping on blanket bogs also using high resolution aerial imagery and deep learning methods (MacFarlane et al, 2024). These studies serve to highlight the context-specific nature of peatland studies that are highly valuable in the field of ecological monitoring.

A generalized mapping of vegetation communities in the uplands of Ireland has been conducted using a combination of optical and SAR imagery as input for an RF model (Barrett et al, 2016). Additionally, ecotope mapping (similar to degradation or condition assessment) techniques of raised bogs using satellite imagery have been explored (Ingle et al, 2023). Karlson & Bastviken (2023) used an RF algorithm and a mixture of S1, S2 and ancillary data to classify areas as fen, bog or swamp. However, to date there have not been any studies published that aim to map degradation of blanket bogs

along official degradation classification systems using a combination of optical and SAR imagery and derived variables as inputs to an RF model.

3. METHODOLOGY

The overall methodological workflow adopted in this study is outlined in Figure 1.

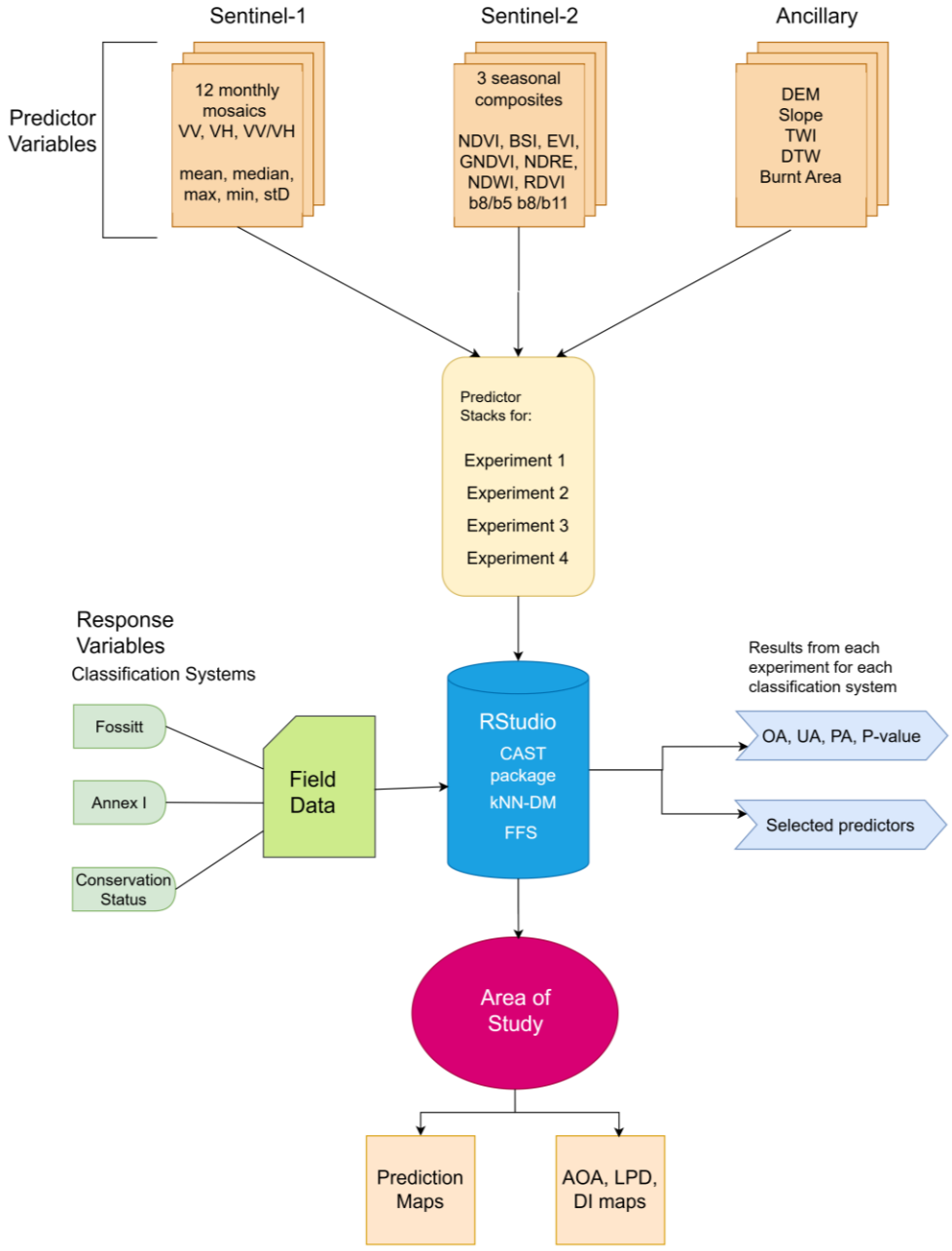


Figure 1. High level methodology diagram

3.1 AREA OF STUDY

This study is focused on areas of mountain blanket bogs in the Dublin and Wicklow Mountains located just West of Dublin in Ireland. The area of interest is 250 meters above sea-level, consistent with the definition of mountain upland blanket bog as opposed to lowland blanket bog. The boundary of the area of interest was taken from the CORINE Landcover data with minimum resolution of 25 hectares (Figure 2).Two large areas, defined under the class description “Peatlands” were selected and isolated to be used as the area of interest, they can be seen in Figure 2 in the context of Ireland. Together, they cover 13,007 hectares. This dataset was chosen as a boundary as it excludes similar habitats such as heathland, see Figure 3 showing different extents of blanket bog habitat.

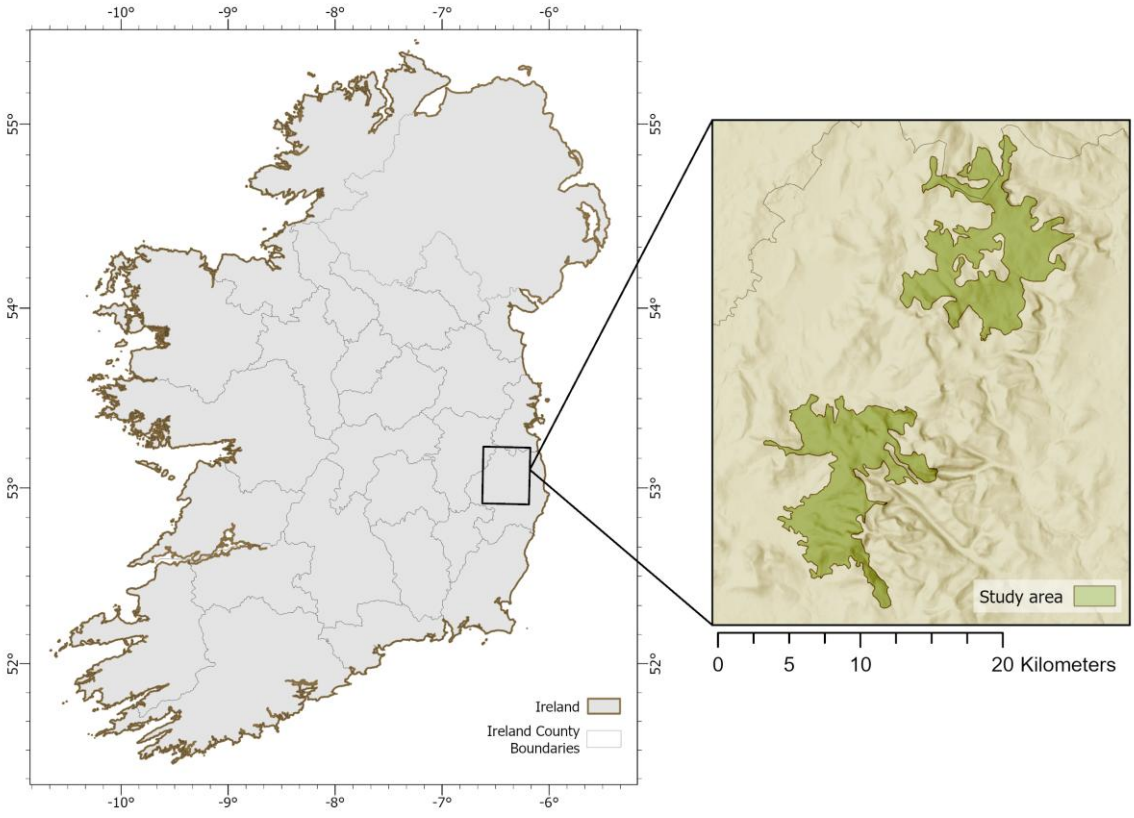


Figure 2. Location of area of study in Ireland

As will be discussed later under the Data section, the ground truth data contained other habitats that fall under the “Heathland” class description, habitats which were not to be used in this study. The CORINE Landcover dataset provided the clearest and most

discriminative blanket bog area that could be used for the predictive section of this study. Habitats are often difficult to distinguish and exist without clear boundaries. Different datasets will often draw boundaries between habitats with much variation. Figure 3 shows different extents of blanket bog. The purpose of this study is not to distinguish between habitats but rather between conditions within blanket bog habitats. This study used the most discriminative peatland boundary considered to be blanket bog as the area of interest.

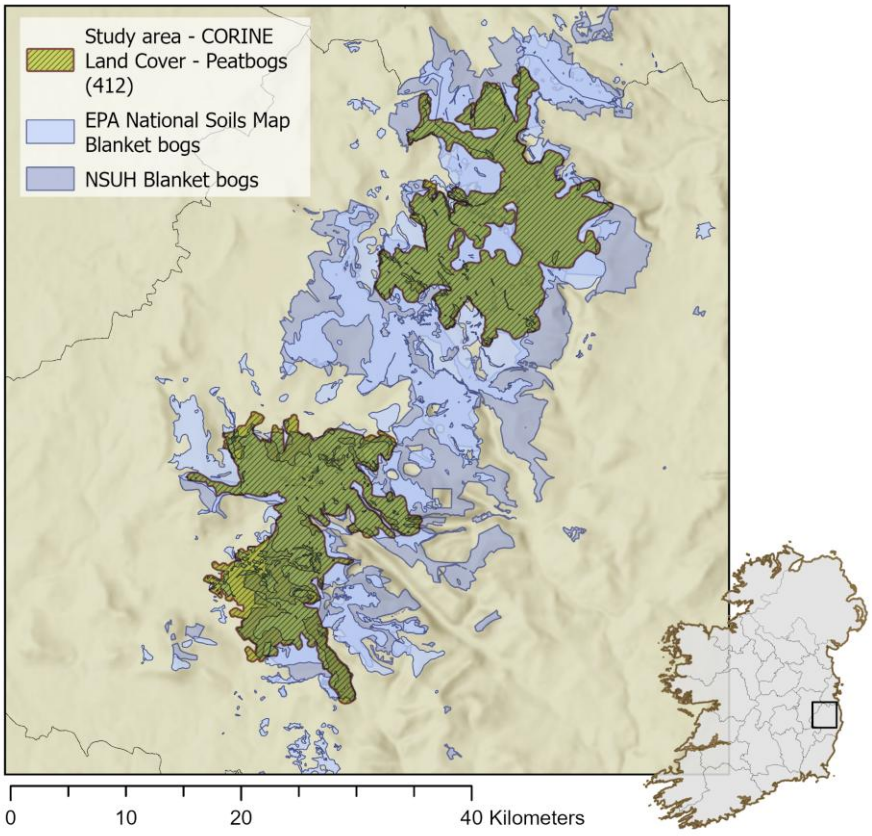


Figure 3. Map showing different extents for blanket bogs on the Wicklow Mountains SAC

The area of study exists within the Wicklow Mountains Special Area of Conservation (SAC). It has been identified as an area harbouring a significant number of habitats listed under the Annex I / II of the E.U. Habitats Directive. Blanket bog is one of the dominant habitats in the SAC along with heath and upland grassland (NPWS, 2017a). Some of the study area is also part of commonage land. Commonage in this area is mainly used for traditional grazing with a recent history of turf cutting (NPWS, 2005). These are areas shared between one or more parties. The Wicklow Mountains National Park is managed by the NPWS and also overlaps with the study area. The

study area thus has many different management systems with some conflicting interests between stakeholders such as farmers, private landowners, the NPWS and other administrative or land management bodies. **Error! Reference source not found.** shows the boundaries of different interest groups and how they overlap with the area of study.

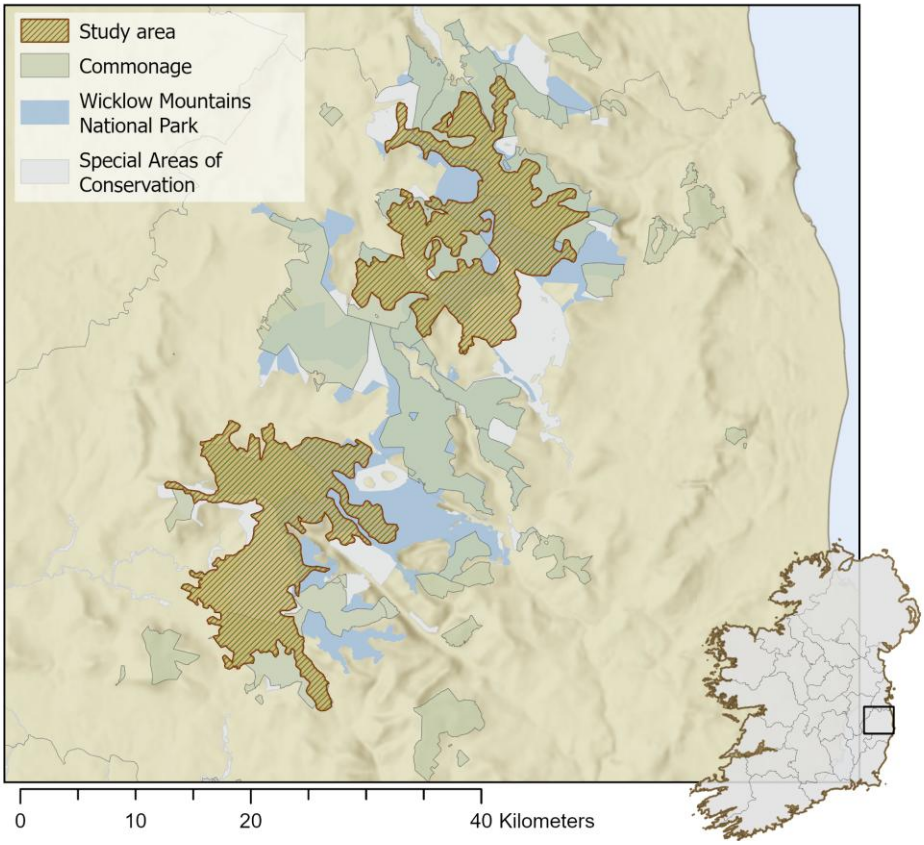


Figure 4. Map showing the jurisdiction of different interest groups that overlap with the area of study.

3.2 DATA

Table 1 below provides details on data characteristics and sources used in this study.

DATA	SOURCE	RESOLUTION	DATE	FORMAT	PRODUCT
Sentinel-2	ESA Copernicus	10m	2019	Geo-TIFF	12 band multi-spectral seasonal mosaics
Sentinel-1	ESA Copernicus	20m	2019	Geo-TIFF	monthly mosaic
DEM	OpenTopography DEM downloader	30m	2013	Geo-TIFF	NASA SRTM
Burnt areas	EFFIS MODIS RDA	250m	2012-2022	GeoJSON	Burnt area, IE
Surveys	Wicklow Uplands Council - SUAS	N/A	2019	shapefile	Habitat surveys conducted as part of the SUAS project conducted by the Wicklow Uplands Council commissioned by the NPWS.

Table 1. Data characteristics

3.2.1 RESPONSE VARIABLES

The ground truth data was gathered in 2019 under the Sustainable Uplands Agriculture-environment Scheme (SUAS) project, commissioned by the Wicklow Uplands Council. The aim of the project was to support sustainable management of farming activities in the Wicklow uplands region. As part of the project, land surveys establishing baseline ecological and hydrological conditions were conducted for the purpose of comparing before and after results of the project (Wicklow Uplands Council, 2018). These surveys provided the ground truth data to train and validate the model. Only sites involved in the project where Blanket Bog was present were selected. Of the ten sites that were part of the project seven sites were selected for this project covering an area of 497 hectares. Figure 5 shows the location of surveys alongside the study area for context.

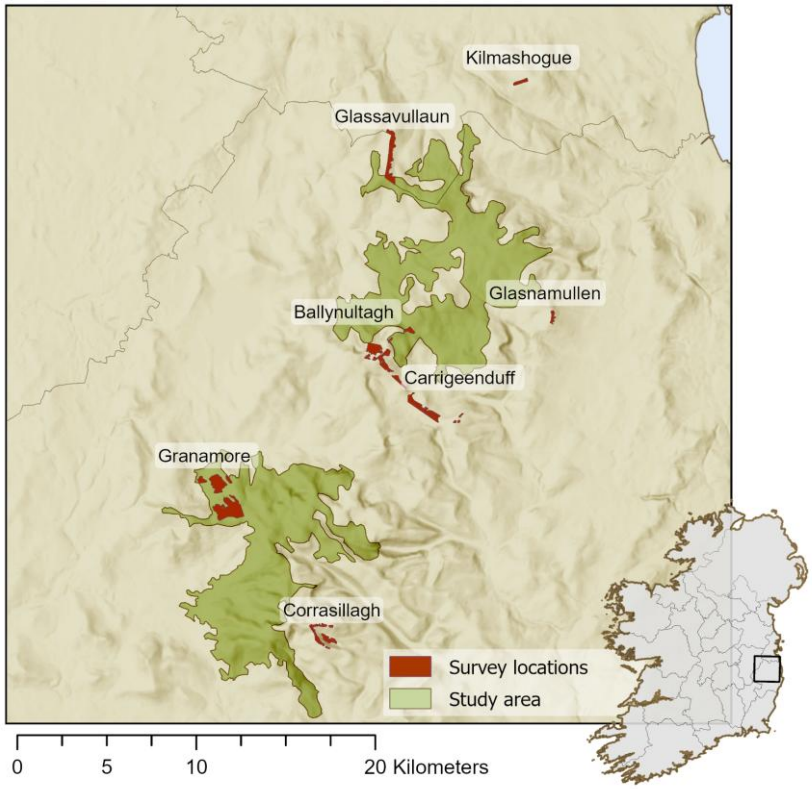


Figure 5. Map showing locations of surveys and their names with respect to the study area.

The areas that were surveyed were divided along ecological distinctiveness though some areas were difficult to distinguish and were labelled as mosaic habitats with the habitats present listed using their codes in the label, e.g. PB4/HH3. See section 3.2.1.1 for more details on how this was dealt with. The areas were defined based on background research and site visits. Detailed maps of the sites can be found on the SUAS website.

The Fossitt habitat classification covers all habitats within the sites, every area can be assigned to a specific habitat under the Fossitt system. This provided a baseline and information on all areas within the sites. Not all areas within the sites could be assigned to an EU Habitats Directive Annex I habitat type. This is because some habitats do not reach the criteria necessary to be assigned as providing enough services and ecological importance to be prioritised under this directive. This is the case for example with Fossitt habitat Eroding Blanket Bog (PB5), due to its degraded nature it is not assigned as a priority Annex I habitat type (Fossitt, 2000).

The Annex I habitat type that is relevant to this study is Active Blanket Bog, code 7130, and the use of an asterix indicates that the habitat is actively forming peat. This classification is based on vegetation cover, the type of vegetation coverage and the percentage of exposed peat (EUNIS, 2013). However, within this study, we will consider two classes of intact blanket bog and non-intact blanket bog, where intact encompasses areas of 7130 with or without the asterix meaning the area in question may or may not be forming peat.

The Conservation Status habitat classification system has three tiers: *Favourable, Inadequate and Bad*. Specific criteria are established in order to assess the condition of a habitat. These criteria are listed in the survey documentation published by the SUAS project. The criteria fall into three categories: Vegetation Composition, Vegetation Structure and Physical Structure. Other, broader criteria are also used to define the conservation status or condition of an area. *Area* takes into consideration change in area covered by a habitat over time. *Structure and function* is related to the physical attributes as well as ecological processes that shape it. *Future prospects* asks whether threats and pressures pose a significant risk to the existence of the habitat in question into the future.

3.2.1.1 DATA PRE-PROCESSING

The data that was used in this study contained some areas that were labelled as mosaic habitats, or it was unclear which habitat dominated. These were small areas in the dataset and were removed from the data that was subsequently used to train and validate the model. Due to the nature of the data that was used some processing needed to be carried out so that the prediction domain could be considered comprehensive. Table 2 and Table 3 show the data processing. This was carried out using the field calculator in QGIS if a polygon had a NULL Annex I attribute and either PB4 or PB5 label it would be assigned the attribute ‘7000’ under the Annex I field. Similarly of Conservation Status was NULL it would be assigned ‘unlisted’ if it was also attributed to PB4 or PB5. After cleaning and processing, random points were generated from the polygons, inheriting the attributes of their source polygons. Figure 6 provides a visual presentation of class imbalance among the classification system.

ID	FOSSITT	ANNEX_I	CONSERVATION STATUS
1	PB2 (Blanket Bog)	7130	Favourable
2	PB5 (Eroding Blanket bog)	N/A	N/A
3	PB2 (Blanket Bog)	N/A	N/A

Table 2. Example of data before editing fields

ID	FOSSITT	ANNEX_I	CONSERVATION STATUS
1	PB2 (Blanket bog)	7130	Favourable
2	PB5 (Eroding Blanket bog)	7000	unlisted
3	PB2 (Blanket Bog)	7000	unlisted

Table 3. . Example of data after editing fields

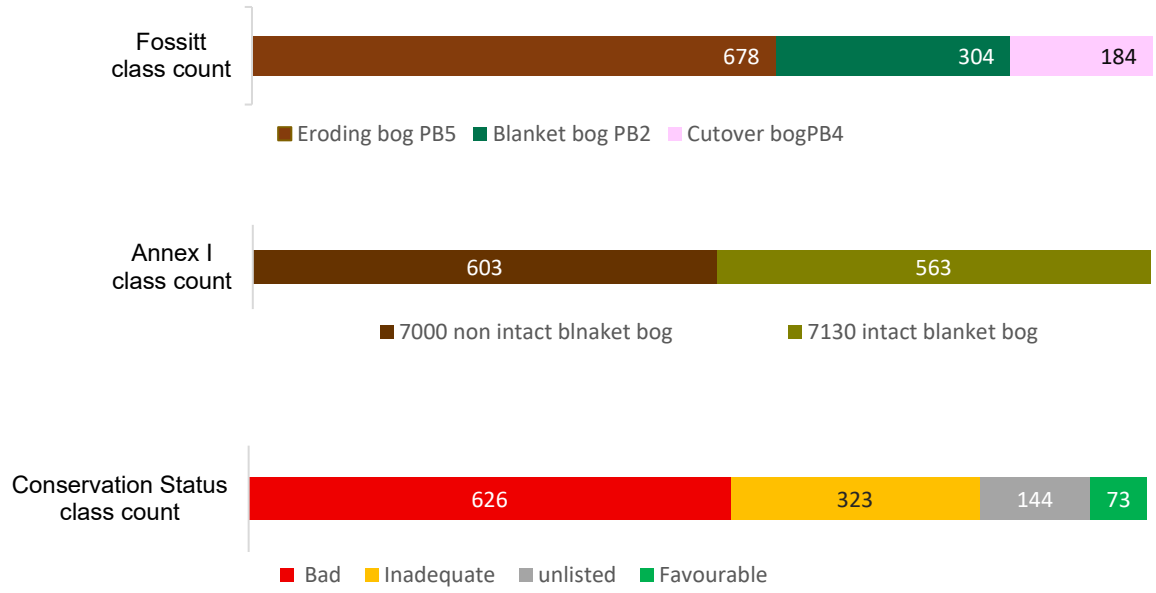


Figure 6. Class imbalance per classification system

3.2.2 PREDICTOR VARIABLES

Three separate 'families' of prediction variables were used in this study: optical imagery, SAR imagery and ancillary data. Table 4 presents the full stack of predictor variables.

SENTINEL-2			SENTINEL-1			ANCILLARY
Season 1	*Season 2*	*Season 3*	*VV*	*VH*	*VV/VH*	
Band 2	Band 2	Band 2	max	max	max	DEM
Band 3	Band 3	Band 3	min	min	min	Slope
Band 4	Band 4	Band 4	mean	mean	mean	TWI
Band 5	Band 5	Band 5	median	median	median	DTW
Band 6	Band 6	Band 6	std	std	std	Burnt areas
Band 7	Band 7	Band 7				
Band 8	Band 8	Band 8				
Band 11	Band 11	Band 11				
Band 12	Band 12	Band 12				
EVI	EVI	EVI				
NDVI	NDVI	NDVI				
GNDI	GNDI	GNDI				
NDRE	NDRE	NDRE				
RDVI	RDVI	RDVI				
BSI	BSI	BSI				
NDWI	NDWI	NDWI				
B8/B5	B8/B5	B8/B5				
B8/B11	B8/B11	B8/B11				

Table 4. Full list of predictor variables

3.2.2.1 OPTICAL IMAGERY AND DERIVED INDICES

2019 Sentinel-2 optical imagery from the Copernicus space telescope with 10-meter resolution was acquired via GEE (Gorelick et al, 2017). Four composites were initially created, however only three were used in the final predictor stack for this study. The composites were created to handle cloudy or otherwise unusable pixels. The S2Cloudless algorithm (Sinergise, 2019) was used to mask out cloud contaminated pixels with a strict 10% probability threshold (APPENDIX A). First a full year composite was calculated, second a pre-growth season from January to April, third a peak-growth season from May to August and fourth, a post-growth season from September to December were all calculated.

The full year composite was not used in the final model training. The decision to remove this composite took into consideration the significant differences in pixel colouration that were in close proximity. In other words, close areas were captured at very different times of the year (Figure 7). This caused issues with continuity within the landscape. Using three 'seasons' was a better approach that produced more seasonally coherent images.

Figure 7 Example of seasonal differences in the full year composite image

For each seasonal composite multiple indices and band ratios were calculated: NDVI, EVI, GNDVI, BSI, NDRE, NDWI, NDMI, RDVI, SR_B8_B5 and SR_B8_B11 (See APPENDIX for calculation script). These were selected to capture and enhance features in the landscape relevant to habitat classification and condition assessment. NDVI, EVI, B8/B5, GNDVI, NDRE, RDVI and B8/B11 all capture information related to plant health and stress. The condition and peat forming abilities of blanket bog habitat is related to vegetation cover and health. BSI is useful in identifying areas of bare soil, highly relevant to the condition of blanket bog as exposed peat is considered to be in poor condition and suffering from erosion (Fossitt, 2000), (IUCN, 2024), (MacFarlane et al, 2024). NDWI and NDMI are used to identify wet areas or areas of open water. An important feature of blanket bogs is the presence of bog pools as well as saturated ground and a high water table.

3.2.2.2 SAR IMAGERY AND TEMPORAL AGGREGATES

Sentinel-1 SAR imagery was acquired from the Copernicus website. Twelve monthly mosaics for each month of both VV and VH polarisations from 2019 were downloaded. Although these products are pre-processed (orbit correction, thermal noise removal, terrain correction, reprojection to a common spatial grid, and monthly temporal aggregation of all available acquisitions) they were downloaded in the raw format

(linear backscatter values without logarithmic scaling or visualisation) and subsequently converted to dB using a python script (see APPENDIX C).

Five temporal aggregates were calculated from the 12 monthly mosaics for VV and VH polarisation as well as the VV/VH ratio: median, mean, max, min and standard deviation. SAR imagery from the Copernicus data ecosystem come in a native 20-meter resolution. The SAR images were thus downsampled using nearest neighbour to the Sentinel-2 imagery resolution of 10 meters. These fifteen SAR images were used to train and validate the model in some of the experiments.

SAR imagery is sensitive to vegetation and moisture conditions as well as landscape structure. It shows great promise in habitat classification in northern regions as it is not hindered by the presence of clouds. Previous studies have shown improved results when combining optical and SAR imagery in remote sensing of ecological habitats (Barrett et al, 2016), (Ghazaryan et al, 2024), (Karlson & Bastivken, 2023).

3.2.2.3 ANCILLARY DATA

This category of data encompasses more static data forms. A DEM at 30-meter resolution was acquired using the OpenTopography DEM Downloader plugin in QGIS. From the DEM three derivatives were calculated: topographic wetness index (TWI), depth to water table index (DTW) and slope. The first two model hydrological flows indicating where water is more likely to accumulate and saturate the soil and slope can indicate where gravity may have a higher impact, increasing the risk of erosion. DTW was calculated in ArcGIS Pro using the Cost Distance tool with the tan of the slope and rasterised river and lake networks as inputs (Murphy et al, 2009). TWI was calculated by first calculating a flow accumulation layer then applying the TWI formula in QGIS Raster Calculator.

A vector layer of burnt areas was acquired from Copernicus' European Forest Fire Information System (EFFIS). This was subsequently rasterised to provide a binary raster where fires had historically occurred or not. See Figure 8 to see locations of historic fires and the overlap with the SUAS survey areas. All raster layers were aligned and downsampled to the resolution of the Sentinel-2 imagery.

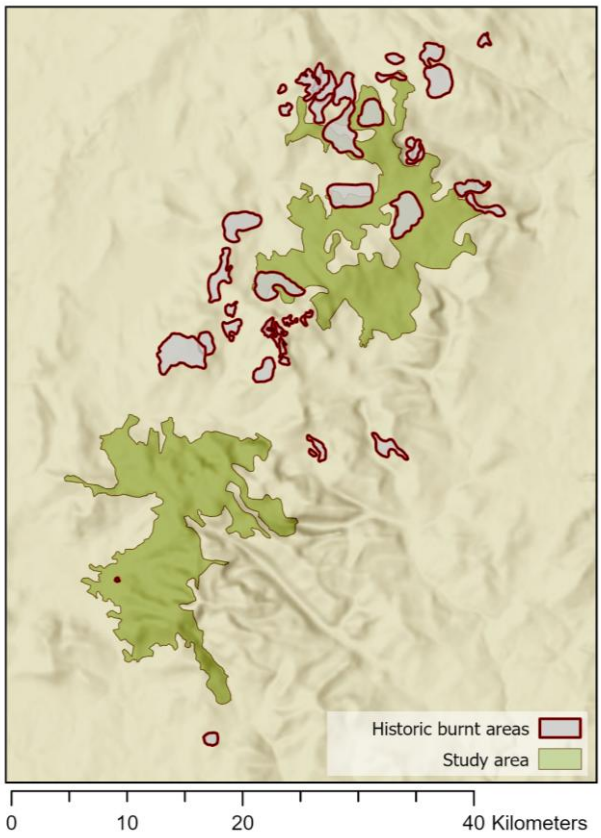


Figure 8. Location and extent of historic burnt area surrounding the study area

3.3 MACHINE LEARNING MODEL

The goal of this study is to explore the use of machine learning to support the monitoring of priority blanket bog habitats and their condition. To do this, RF machine learning algorithm was employed. Specifically, the Caret Applications for Spatio-Temporal (CAST) package was used and implemented through R. This package was chosen for the spatial cross validation strategy and incorporation of Forward Feature Selection (FFS) algorithm. These package features are particularly relevant to this study as it can address issues of over-fitting and over optimistic accuracy results and other performance metrics (Meyer et al, 2018). These issues are notable in mapping studies that use machine learning methods to predict landcover over large areas using small or concentrated training points.

3.3.1 SPATIAL CROSS VALIDATION

The training data available for this study is highly clustered. Leave one out (LOO) and random k-fold cross validation methods do not consider distance (geographic or feature space) between training and prediction points and therefore cannot reliably

represent the accuracy of the predicted map (Linnenbrink et al, 2024). The cross-validation strategy that was used in this study was the k-fold nearest-neighbour distance matching cross-validation (kNNDM CV). This method calculates the empirical cumulative distribution function (ECDF) of the prediction and training points and tries to create a similar ECDF of test and training points within each fold during the creation of folds. This method is computationally more efficient and has wider applicability than Nearest Neighbour Distance Matching Leave One Out Cross Validation (NNDM LOO CV) which preceded and was the foundation upon which kNNDM CV was developed. (Linnenbrink et al, 2024)

3.3.3 Forward Feature Selection

More traditional versions of RF models will often use a variable importance analysis to better understand what predictor variables are useful for a given prediction task. One way in which variable importance can be measured is by using the decrease in impurity that happens with each predictor variable. If a variable is used to split a node and this increases class purity, it increases in importance (Breiman, 2001). This method produces a list of predictor variables and their importance.

However, the FFS method was employed in this study to prevent spatial over-fitting sometimes associated with this type of predictive task. This method iteratively selects predictors based on marginal increases in accuracy measured during validation. This method excludes ‘noisy’ or redundant predictors that may show high correlation with the training data but fail to generalize to new geographic areas. By iteratively adding only those predictors that improve cross-validation scores, FFS ensures the model remains focused on robust spectral-predictor-response relationships (Meyer et al, 2019). Furthermore, this approach explicitly quantifies the contribution of each predictor to target-oriented performance metrics, generated here via k-NNDM CV, thereby ensuring that the final model is optimised for spatial transferability rather than localized memorization.

3.3.2 Area of Applicability

Included in this model are also assessments of AOA using two different approaches. First is the DI that calculates for any given location in the prediction domain (the real-world geographic location where landcover is being predicted), the distance to the nearest point in the multidimensional predictor space (Meyer and Pebesma, 2021).

This, however, does not consider how represented this point may be in the predictor space, i.e. how many points in the predictor space are similar to the given point. Thus, Local Point Density (LPD) can also be calculated using the number of points in the predictor space that are representative of a given location in the prediction domain (Schumacher et al, 2024).

3.4 EMPIRICAL STUDY

The classification process began with the extraction of predictor values at each sample point. To maintain data integrity, a quality filter was applied to remove any points where more than 20% of the predictor values were missing (e.g., due to cloud masking). The remaining data were partitioned into five folds for spatial cross-validation using the k-Nearest Neighbor Distance Matching (k-NNDM) method.

Using CAST::ffs the algorithm iteratively evaluated combinations of predictors, adding only those that improve model performance based on the spatial cross-validation metrics. For each iteration, the model performed hyperparameter tuning—specifically adjusting the number of variables at each split (*mtry*) and node size—to optimize accuracy. The forest size was fixed at 500 trees, and parallel processing was utilized to manage the computational load during the feature selection phase.

Following training and validation, the best-performing experiment model was used to predict across the full study domain. For this final output, the AOA was calculated to map the geographic extent where the model's training data was representative of the landscape, based on the DI and LPD within the multidimensional predictor space. Figure 9 provides a visualisation of the chain of process. The full script used train and test each experiment and classification system can be found in APPENDIX d.

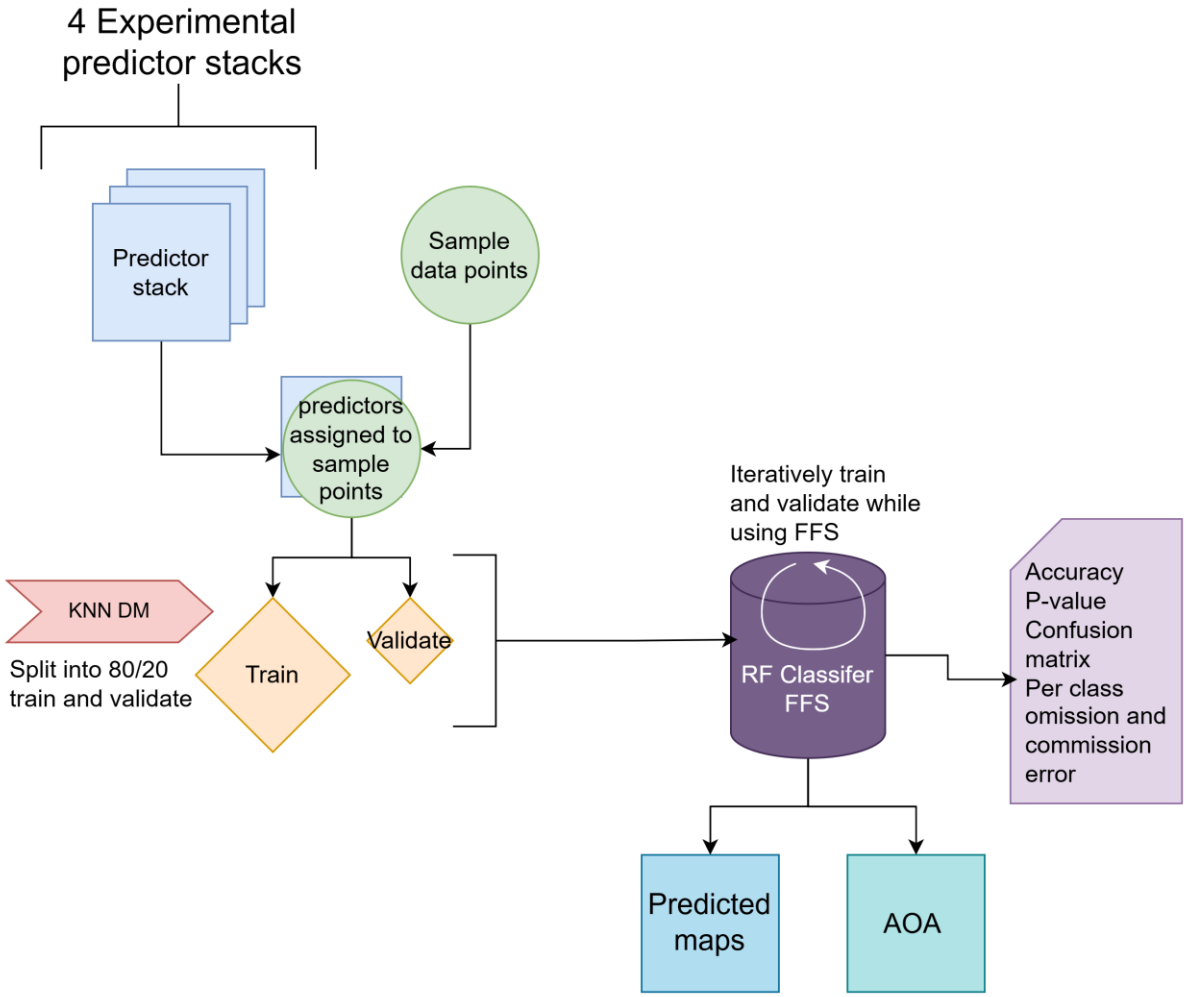


Figure 9. Flow chart of model

3.4.2. Experimental Design

The classification task was run a total of 12 times to compare performance across three classification systems (Fossitt, Annex I, and Conservation Status) and four experiments with unique predictor configurations (Table 5).

	SENTINEL - 2	SENTINEL - 1	ANCILLARY
EXPERIMENT 1	All Seasons	All	All
EXPERIMENT 2	Season 2	All	All
EXPERIMENT 3	All seasons	-	-
EXPERIMENT 4	-	All	All

Table 5. Experiments and tested predictor sets

For each experiment evaluation metrics were gathered to compare the performance of the model. The top predictors selected by the model were recorded along with overall accuracy, producer accuracy, precision and the confusion matrix. These performance metrics have been widely used (Karlson & Bastivken, 2023), (Ingle et al, 2023), (Barrett et al, 2016) in the literature and provide an adequate overview of the model performance that allows for some comparison between experiments. Additionally, the p-value measuring the accuracy vs no information rate (i.e. comparing the model to a 'dumb' model that always predicts the dominant class) was used to provide a benchmark estimation of model performance.

Due to varying results and high computation time, prediction maps and AOA maps were only generated for the best performing stack of predictor variables, from experiment 1. These maps could be overlayed high resolution satellite imagery base maps for visual evaluation as is recommended by (Meyer et al, 2021). This provided an additional layer of evaluation of the model's performance.

4. RESULTS

4.1 Experiment Results

4.1.1 Experiment 1 – All predictors available during training and validation

4.1.1.1 Fossitt

The overall accuracy for this experiment where the full stack of predictors was available to the classification algorithm was 63.81% and the P-Value 0.00004509. The sensitivity and precision were highest for the eroding bog (PB5) class at 80.09% and 80.08% respectively (Table 6). The lowest sensitivity and precision were produced by the cutover (PB4) at 23.91% for both. The blanket bog class produced slightly better metrics with a sensitivity and precision of 52.64% and 51.64% respectively.

The confusion matrix shows the model struggling to classify cutover bog correctly with a significant potion being incorrectly classified as eroding bog. The model distinguishes slightly better between blanket bog and cutover bog. Blanket bog was relatively well classified but there is significant over-lap with eroding bog.

The model found no increase in accuracy after 7 predictors were selected. An even number of predictors were selected between the optical and SAR data types, three variables each. Only DTW was selected from the ancillary data type. Of the three predictors selected under the SAR data type, two of them were derived from the VV/VH ratio.

FOSSITT	SENSITIVITY	PRECISION
PB2	0.5264	0.51644
PB4	0.23913	0.2391
PB5	0.8009	0.8008

Table 6. Experiment 1 - Fossitt per class sensitivity and precision results

PREDICTION	PB2	PB4	PB5
PB2	157	52	94
PB4	23	44	41
PB5	124	88	543

Table 7. Experiment 1 - Fossitt Confusion Matrix

4.1.1.2 ANNEX I

The overall accuracy from the first experiment for the Annex I classification task was recorded at 80.27%. The p-value for this experiment came out significantly low. The non-intact blanket bog class (7000) was selected as the 'positive' class.

The model demonstrated strong performance in identifying the positive class (7000), with a sensitivity of 82.92% (Table 9), indicating that the majority of true 7000 samples were correctly classified. The precision for class 7000 was 79.74%, suggesting that predictions of this class were generally reliable. Inspection of the confusion matrix shows (Table 8) a relatively low number of omission errors for class 7000 (103 cases), while commission errors were also limited (127 cases). Class 7130 was similarly well classified, with a high number of true negatives (436), indicating good overall class separability.

Seven predictors were selected by the model, and all of them were derived from the optical data type. Interestingly, NDVI was selected twice. The predictors selected

covered the three aspects of peatland degradation, vegetation health, cover of bare earth and hydrology.

PREDICTION	7000	7130
7000	500	127
7130	103	436

Table 8. Experiment 1 - Annex I confusion matrix

	SENSITIVITY	PRECISION
7000	0.8292	0.7974
7130	0.774	0.809

Table 9. Experiment 1 – Annex I per class sensitivity and precision results

4.1.1.3 CONSERVATION STATUS

The overall accuracy for this classification system was the lowest of the three at 57.68%. Additionally, the p-value was above the threshold of less than 0.05 indicating that the observed classification performance was not statistically significant and did not differ reliably from what would be expected by chance.

The model struggled significantly with the 'unlisted' which produced a sensitivity and precision of 0 (Table 10). The model was unable to identify this class. It was most confused with the 'bad' class with 58 points being classified as such (Table 11), 22 points classified as 'inadequate' and 7 points as 'favourable'. The class that was most correctly classified was the 'Bad' class but with considerable confusion with the 'inadequate' class. The 'favourable' class was somewhat well classified and was most misclassified with the adjacent 'inadequate' class. The 'inadequate' class was classified as 'bad' about three times more than 'favourable'. Indicating a more negative skew to the results of this classification prediction.

The model only selected three predictors, reporting no increase in accuracy with a fourth addition. Two of the three predictors were derived from the SAR imagery. Only band 2 of season three was selected form the optical data type.

CONDITION	SENSITIVITY	PRECISION
FAVOURABLE	0.2739	0.2597
BAD	0.7348	0.6424
INADEQUATE	40.25	0.4545
UNLISTED	0	0

Table 10. Experiment 1 - Condition per class sensitivity and precision results

	FAVOURABLE	BAD	INADEQUATE	UNLISTED
FAVOURABLE	20	14	31	12
BAD	20	460	140	96
INADEQUATE	26	94	130	36
UNLISTED	7	58	22	0

Table 11. Experiment 1 - Condition confusion matrix

4.1.2 EXPERIMENT 2 – ALL PREDICTOR GROUPS AVAILABLE – ONLY SEASON 2 OPTICAL IMAGERY AND DERIVED INDICES

4.1.2.1 FOSSITT

A slight increase in accuracy of about 0.5% was recorded for the second experiment where only season 2 of the optical data type and derived indices were available for model training. P-value remained significantly below the threshold suggesting the model classification performance has some statistically significance. For this experiment the sensitivity and precision (Table 12) of the blanket bog class reduced as compared to the first experiment while eroding and cutover bog classes increased slightly. Suggesting that the blanket bog class classification was more dependent on seasonal differences from the optical data type.

The confusion matrix (Table 13) shows again that PB5 was the most accurately classified class, with a high number of correctly predicted samples (564), indicating strong separability from the other classes. In contrast, PB2 showed substantial confusion with both PB4 and PB5, with a large proportion of PB2 reference samples misclassified as PB5 (139). PB4 was again the least well classified class, with fewer correct predictions (47) and frequent misclassification as both PB2 and PB5. Overall,

misclassification patterns indicate that confusion was most pronounced between PB2 and PB5, while PB4 exhibited widespread confusion across all classes.

The selected predictors show similar distribution across the different data types but with the addition of 'burnt area' as a predictor. The SAR derived predictors remained the same but season 2's B8/B5 predictor was selected and B2 was dropped even though it remained available as a predictor under this experiment.

FOSSITT	SENSITIVITY	PRECISION
PB2	0.4572	0.4572
PB4	0.2554	0.2554
PB5	0.8319	0.8318

Table 12. Experiment 2 - Fossitt sensitivity and precision results

PREDICTION	PB2	PB4	PB5
PB2	139	51	84
PB4	26	47	30
PB5	139	86	564

Table 13. Experiment 2 Fossitt confusion matrix

4.1.2.2 ANNEX I

The accuracy under experiment 2 dropped markedly by over 10% though p-value remained satisfactorily low. The sensitivity (Table 14) dropped by a larger percentage than precision with sensitivity dropping from 82% to 64% compared to precision that dropped from 79% to 73%. Looking at the confusion matrix the biggest difference between experiments 1 and 2 was the increase in omission error for the intact blanket bog class (7130) as compared to the previous experiment with the number of 7130 points being classified as 7000 doubling (Table 15).

In terms of selected predictors, none of those selected for the first experiment were selected for this experiment. Only four total were chosen, bands 2 and 3 as well as elevation and burnt area. Similarly to the previous experiment no SAR derived predictors were selected, instead the model selected predictors form the ancillary data type.

	SENSITIVITY	PRECISION
7000	0.6467	0.7373
7130	0.753	0.666

Table 14. Experiment 2 - Annex I producer and precision

PREDICTION	7000	7130
7000	390	139
7130	213	424

Table 15. Experiment 2 - Annex I confusion matrix

4.1.2.3 CONSERVATION STATUS

Accuracy also dropped for this classification task and p-value remains high above the threshold indicating no statistically significant prediction power of this classification task. Again, the sensitivity and precision for the unlisted class was 0 highlighting the model's difficulty in identifying this class at all. Only the 'bad' class presented any meaningful results in the reported sensitivity and precision, at 69% and 61% respectively (Table 16). Notably there was a marked decrease in sensitivity for the 'favourable' class, down to 12%, comparatively the precision only decreased slightly at 24%. The confusion matrix shows poor classification performance for the 'favourable' class while the 'bad' class and the 'inadequate' class show stronger classification performance. More inadequate labelled points were classified as 'bad' rather than 'favourable', repeating the observation form the previous experiment of a skewedness towards negative classes (Table 17).

Similarly to the previous experiment, only three predictors were selected with a similar propensity for SAR derived predictors with a VH aggregate being chosen this time instead of VV and with the addition of a VV/VH aggregate. Burnt area was also selected for this experiment.

CONSERVATION STATUS	SENSITIVITY	PRECISION
FAVOURABLE	0.1232	0.2432
BAD	0.6853	0.6172
INADEQUATE	0.4551	0.4288
UNLISTED	0	0

Table 16. Experiment 2 - Condition sensitivity and precision

CONSERVATION STATUS	FAVOURABLE	BAD	INADEQUATE	UNLISTED
FAVOURABLE	9	10	12	6
BAD	29	429	145	92
INADEQUATE	22	136	147	46
UNLISTED	13	51	19	0

Table 17. Experiment 2 - Condition confusion matrix

4.1.3 Experiment 3 – Only Optical Imagery and Derived Indices Available (all three seasons)

4.1.3.1 Fossitt

Under the third experiment where only optical imagery and derived indices (from all three seasons) was available to the model, the accuracy for this classification task dropped significantly by around 10%. P-value was much higher from this experiment at 0.9822, above the p-value threshold that implies no statistically significant prediction power. The sensitivity and precision (Table 18) for the eroding bog class remained relatively high for this experiment at 74%, though this was still a decrease from the previous two experiments. Notably, this experiment produced the highest sensitivity and precision for the cutover bog class at 29.34% for both. The confusion matrix (Table 19) shows similar patterns to previous experiments with eroding bog being classified the best, cutover the worst with much confusion between the other two classes, and blanket bog being most confused with eroding bog.

With only optical imagery available, the model relied more heavily on indices that are sensitive to vegetation health such as ‘EVI’, ‘GNDVI’ and ‘NDVI’. There seems to be no preference toward one season with predictors being selected from all three seasons.

FOSSITT	SENSITIVITY	PRECISION
PB2	0.3651	0.3651
PB4	0.2934	0.2934
PB5	0.705	0.705

Table 18. Experiment 3 – Fossitt sensitivity and precision

PREDICTION	PB2	PB4	PB5
PB2	111	31	145
PB4	39	54	55
PB5	154	99	478

Table 19. Experiment 3 – Fossitt confusion matrix

4.1.3.2 ANNEX I

The accuracy from this experiment reflects that of the first experiment where all predictors were available at 80.27% and p-value remains low. The sensitivity and precision (Table 20) are the same as the first experiment change by only small fractions of a percent. The confusion matrix (Table 21) reflects these metrics, with omission and commission errors changing only slightly which is reflected in the fractional differences in sensitivity and precision between the first and third experiment.

The similarities in results between experiments 1 and 3 are also seen in the predictors selected, the only difference being the addition of season 2's in place of season 2's NDWI.

	SENSITIVITY	PRECISION
7000	0.8259	0.7993
7130	0.778	0.807

Table 20. Experiment 3 - Annex I per class sensitivity and precision

PREDICTION	7000	7130
7000	498	125
7130	105	438

Table 21. Experiment 3 - Annex I confusion matrix

4.1.3.3 CONSERVATION STATUS

This experiment resulted in the lowest recorded accuracy at 48.97% and similarly high p-value. Values for sensitivity and precision (Table 22) show very little change compared to experiment 2. The sensitivity and precision showed the greatest change for the 'inadequate' class, dropping by around 4% for both. The 'bad' class produced the highest sensitivity and precision with the 'unlisted' class achieving 0 for both. The CM (Table 23) continues to show that only the 'bad' class has any predictive power compared to the others in this classification system and that the 'inadequate' class is more likely to be confused with 'bad' rather that 'favourable'. The predictors selected, similarly to the Fossitt classification system, are related to vegetation health.

CONSERVATION STATUS	SENSITIVITY	PRECISION
FAVOURABLE	0.1232	0.2307
BAD	0.6821	0.6022
INADEQUATE	0.418	0.3879
UNLISTED	0	0

Table 22. Experiment 3 – Conservation status sensitivity and precision

	FAVOURABLE	BAD	INADEQUATE	UNLISTED
FAVOURABLE	9	12	13	5
BAD	28	427	155	99
INADEQUATE	36	137	135	40
UNLISTED	0	50	20	0

Table 23. Experiment 3 – Conservation Status confusion matrix

4.1.4 EXPERIMENT 4 – ONLY SAR DERIVED AGGREGATES AND ANCILLARY DATA TYPES AVAILABLE

4.1.4.1 FOSSITT

The accuracy from this experiment where the model only had available SAR derived and ancillary predictors came out at a modest 58% but with a p-value above the

threshold for predictive statistical significance at 0.44. Similar trends persist with regards to per class performance with no significant differences in results from experiments 1 to 3, the sensitivity and precision from this experiment are quite average in comparison to the previous experiments, with only slight increase in performance metrics compared to experiment 3. The confusion matrix (Table 25) shows the persistent tendency for this model to over predict eroding bog, little to no predictive power for cutover bog and confusion of blanket bog with eroding bog.

In experiment 1 where the model had all predictors available, the model selected a VH derived aggregate and this trend continues with this limited set of predictors. Similarly, DTW was selected again from this more limited set of predictors. The VV/VH ratio aggregates were also selected again building on the strength of their predictive power from the first experiment.

FOSSITT	SENSITIVITY	PRECISION
PB2	0.4342	0.4342
PB4	0.2554	0.2554
PB5	0.7404	0.7404

Table 24. Experiment 4 - Fossitt sensitivity and precision

PREDICTION	PB2	PB4	PB5
PB2	132	53	125
PB4	37	47	51
PB5	135	84	502

Table 25. Experiment 4 - Fossitt confusion matrix

4.1.4.2 ANNEX I

This experimented resulted in the lowest accuracy metrics for this classification task at 60% though the p-values remains well below the threshold. The sensitivity and precision (Table 26) reflect this modest accuracy at 54% and 64% respectively. In this instance the model seems to struggle more with distinguishing 7130 from 7000 with 275 points being incorrectly classified as 7000 in the confusion matrix (Table 27). This

behaviour is similar to experiment 2 where optical imagery was limited. Once again, similar to experiment 2, elevation and burnt area are selected when optical imagery is limited or unavailable.

	SENSITIVITY	PRECISION
7000	0.5439	0.6406
7130	0.673	0.58

Table 26. Experiment 4 - Annex I per class sensitivity and precision

PREDICTION	7000	7130
7000	328	184
7130	275	379

Table 27. Confusion Matrix for Annex I experiment 4

4.1.4.3 CONSERVATION STATUS

Accuracy was only marginally increased from experiment 3 at 49% though p-value remains stubbornly high for this classification task at 0.99. The sensitivity and precision (Table 28) remain similarly low to the previous experiments with no stand-out results in any class. The confusion matrix (Table 29) reflects the general trend of this classification task across the experiments that has been mentioned above. This experiment resulted in only two predictors being selected, the maximum aggregate of VH and the VV/VH mean. This classification task has had a stronger propensity toward the SAR derived predictors that the other classification tasks though this has not resulted in a usable predictive model.

CONSERVATION STATUS	SENSITIVITY	PRECISION
FAVOURABLE	0.1369	0.2631
BAD	0.6821	0.617
INADEQUATE	0.4551	0.4152
UNLISTED	0	0

Table 28. Experiment 4 – Conservation status sensitivity and precision

	FAVOURABLE	BAD	INADEQUATE	UNLISTED
FAVOURABLE	10	10	12	6
BAD	28	427	146	91
INADEQUATE	23	137	147	47
UNLISTED	12	52	18	0

Table 29. Experiment 4 – Conservation status confusion matrix

4.1.5 Experiments Summary

Overall experiment 1 with all predictor variables available performed the best with the exception of the Fossitt classification improving marginally with the more limited set of optical data type predictors. Each classification system established preferred sets of predictors, and when the availability of the preferred set of predictors was reduced this was reflected in the resulting metrics. Fossitt classification system relied on predictors form the broader suite of data types. Annex I heavily relied on optically derived data that reflected the three components to blanket bog degradation: vegetation health, bare earth exposure and hydrological processes. Conservation Status classification relied most on SAR derived predictors with its lowest accuracy resulting from the experiment where these were not available in training.

The Annex classification task consistently outperformed the other two classification tasks with p-value remaining low implying strong predictive power. The Conservation Status classification task significantly under performed with p-value never coming near to falling below the threshold required to indicate statistically significant predictive power. The Fossitt classification task had some ability to distinguish between classes but suffered significant when the wider suite of predictors was not available with accuracy dropping and p-value raising above the threshold of 0.05.

4.2 Visual Inspection of Predicted maps

Although it is challenging to visually compare the model predictions to ground truth using satellite imagery, there are some clear indicators of habitat condition such as eroded areas, a clear and visually obvious indicator of non-peat forming and non-intact blanket bog areas. See Figure 10 for an example of eroding blanket bog, the dark brown colour indicates severe erosion.

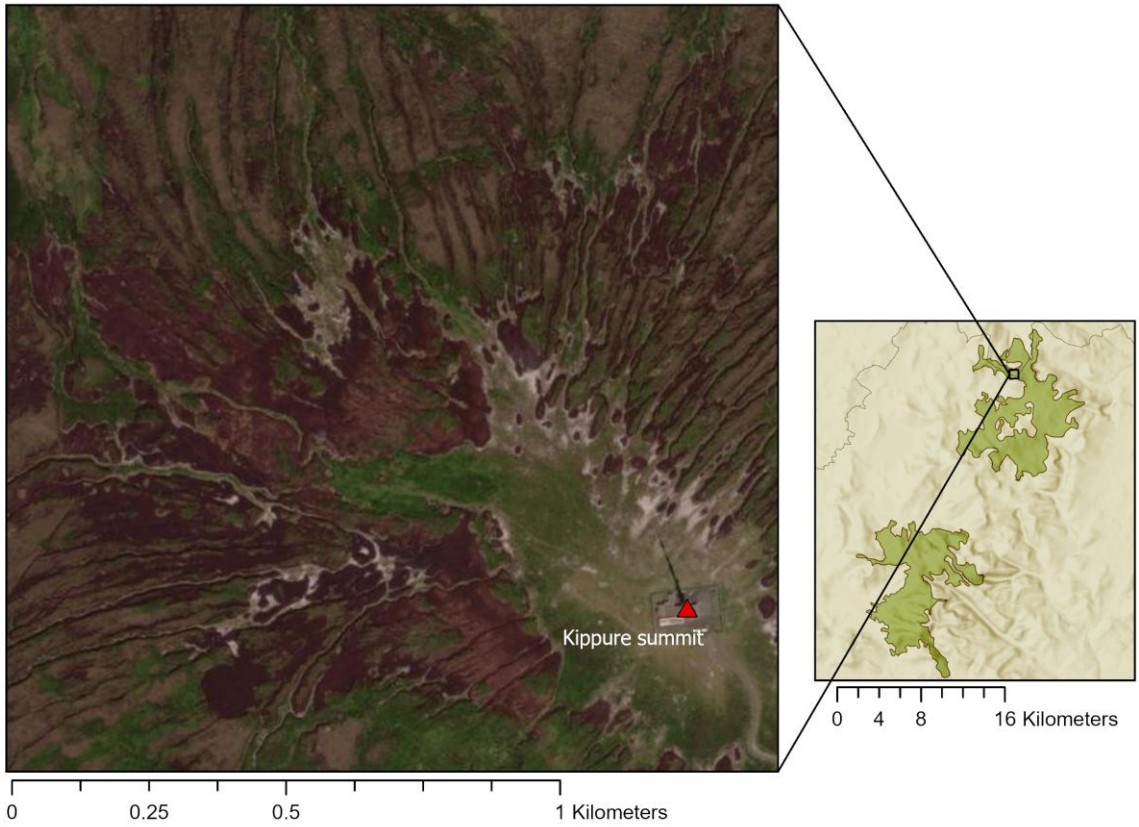


Figure 10. Example of eroding blanket bog, non-peat forming around Kippure

In addition to using such visual indicators to assess model's prediction performance it is also useful to compare how the model classified areas under the Fossitt system and the Annex I system. Intuitively, the model should classify areas of cutover bog (PB4) and eroding bog (PB5) as non-intact blanket bog areas (7000). Similarly, areas classified under the Fossitt system as blanket bog (PB2) should have some classification agreement with intact blanket bog (7130). Looking from a wider view at the landscape classification prediction of the Fossitt in comparison to the Annex I system, the Fossitt system shows a more negative picture of the habitat health, though this may be due to the fact that there are two classes that would be considered 'bad', eroding and cutover bog whereas only one class represents 'good' health, blanket bog. Looking at the confusion matrix across the experiments, PB4 was more likely to be confused with PB5 as opposed to PB2. Nevertheless, comparing the overall classification of the Fossitt and Annex systems, there is relatively good alignment, particularly in the northern portion of the area of study.

Under the Fossitt classification system there are clear signs in that the model relied on hydrological predictor variables. Figure 11 shows how the classification follows morphological features related to a river, and these morphological features are represented by the TWI and DTW; notably DTW was a predictor selected by the model.

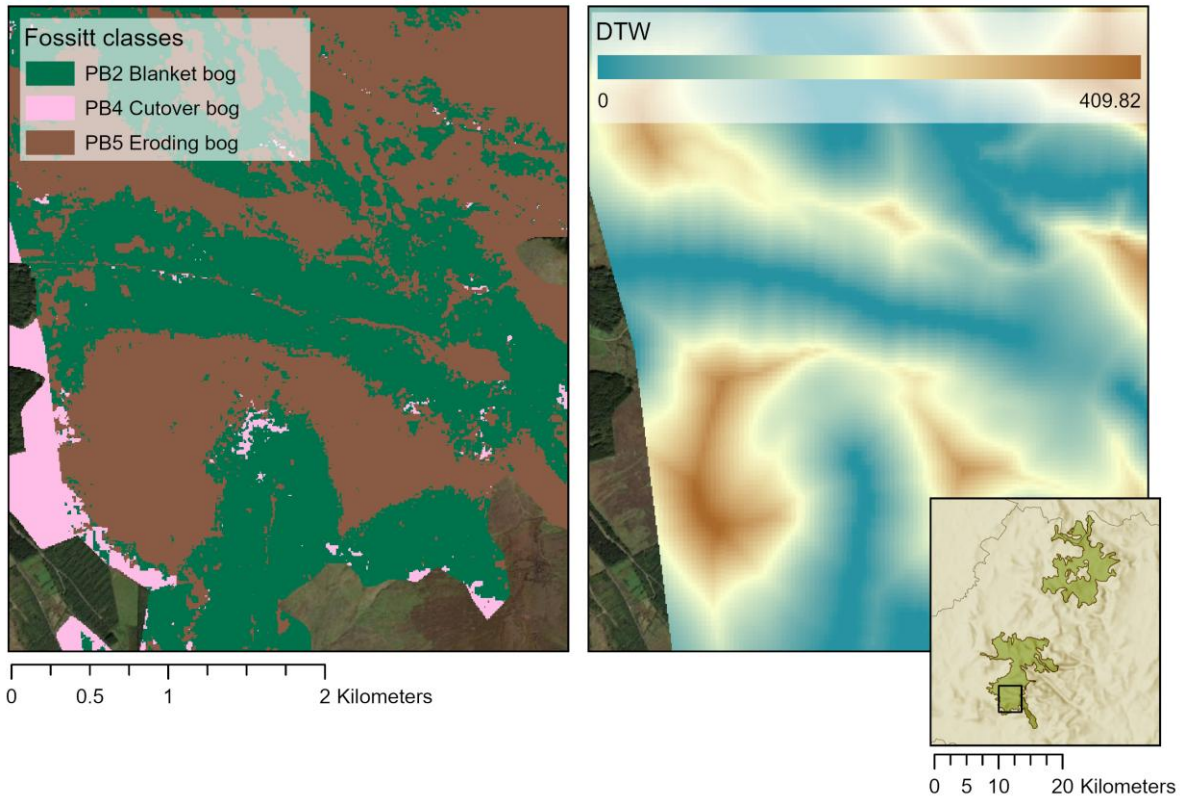


Figure 11. Right: Fossit classifiction. Left: DTW predictor. This snapshot shows how the model used DTW to assign classes in some cases.

Unfortunately, based on visual inspection the model was unable to distinguish areas of cutover bog (PB4) satisfactorily. Cutover bog is relatively easy to visually identify using high resolution satellite imagery basemaps due to the linear scars that result from industrial peat extraction. Visually examining examples of cutover bog and comparing to the Fossitt classification prediction, the model did, for the most part, correctly identify areas of cutover bog, however it clearly over classified areas as cutover bog. This is also reflected in the confusion matrix as mentioned above.

4. 3. AOA, DI AND LPD

Prediction maps and AOA maps were only generated from the first experiment. Despite efforts to use an area that should be just blanket bog habitat (see section 3.1 AREA OF STUDY), the AOA for both the Fossitt and Annex I classification tasks highlighted

areas within the area of study that were categorically not areas of mountain blanket bog. For example, a small coniferous plantation in the southern portion of the area of study, water features such as lakes and some farmland. Figure 12 shows an area of coniferous plantation that fell outside the AOA.

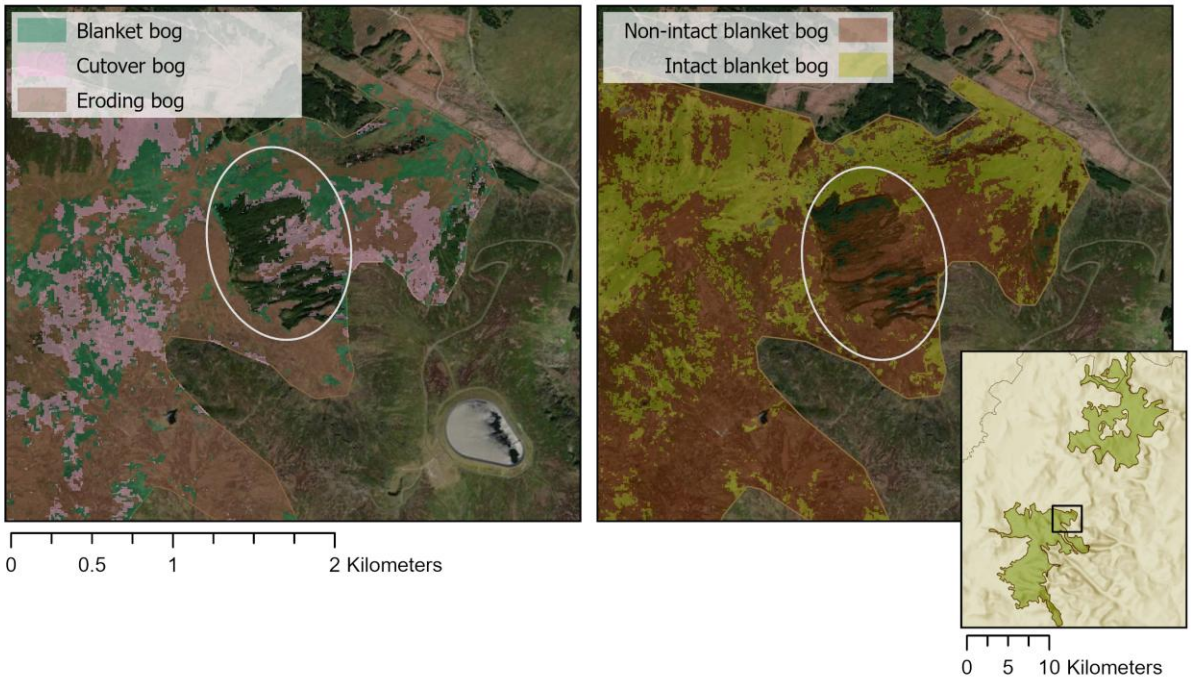


Figure 12. AOA masked Fossitt and Annex I predictions with coniferous plantation indicated

The Fossitt AOA was more conservative than the Annex I AOA, particularly in the southern portion of the study area. Figure 13 shows the full study classification prediction for Fossitt and Annex I classification systems and Figure 13 shows the breakdown of pixels that were identified as being within the AOA.

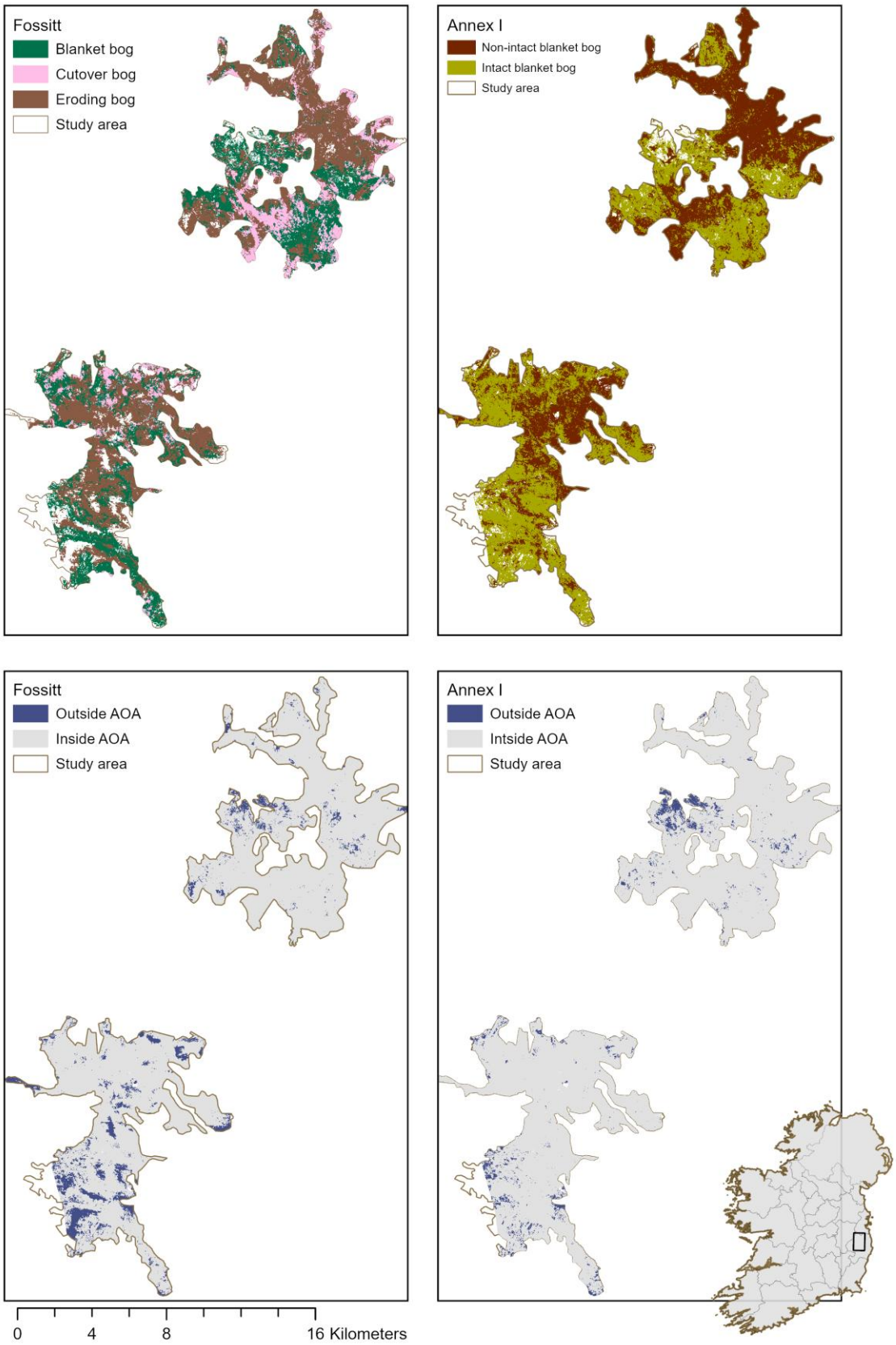


Figure 13. Top: classification prediction masked by AOA for Fossitt and Annex I. Bottom: AOA with areas outside AOA depicted in blue.

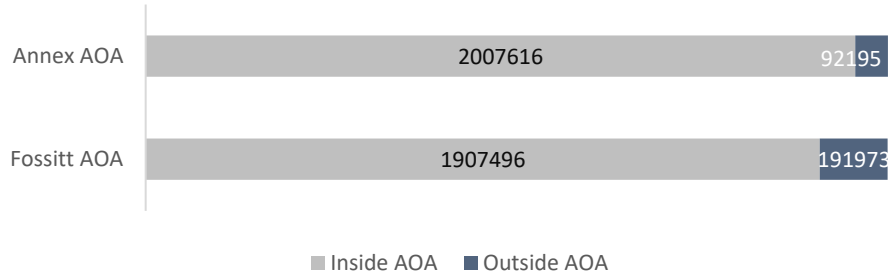


Figure 14. Number of pixels that fall within and outside the AOA for Fossitt and Annex I

DI and LPD plots for the Fossitt and Annex I prediction domains are shown in Figure 15. The DI values exhibited a more skewed distribution, with higher values concentrated within a relatively small number of localised areas. In contrast, LPD values were more evenly distributed across the range, indicating greater variability and a broader dynamic range.

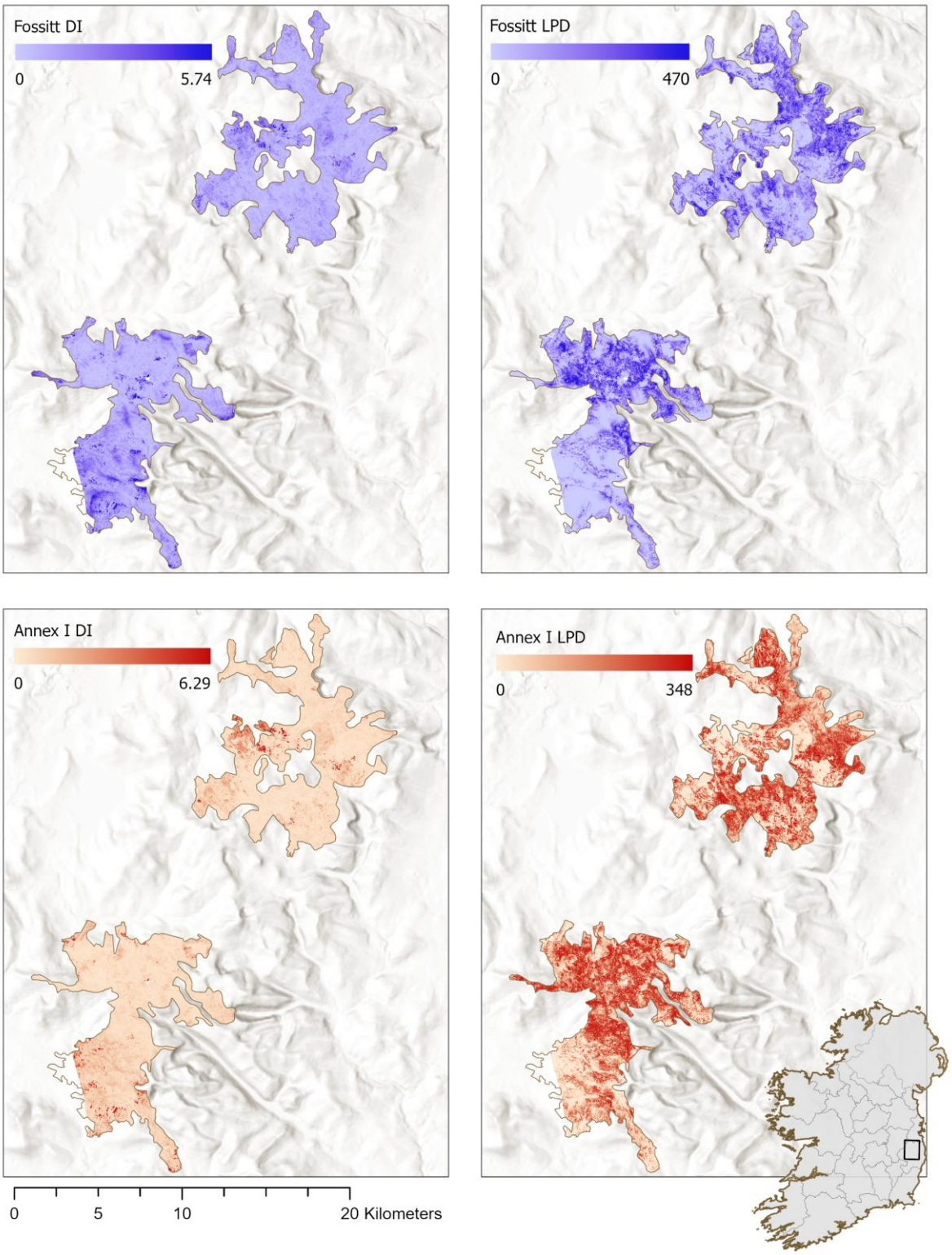


Figure 15. Plotted in blue, Fossitt DI and LPD, and in red Annex I DI and LPD

5. DISCUSSION

5. 1. COMPARISON WITH RESULTS REPORTED IN THE LITERATURE

The best performing classification task (Annex I, exp. 1) was on par with similar studies in the literature with a sensitivity of 82% for the non-intact blanket bog class (7000) and an overall accuracy of 80%. Barrett et al (2016) reported sensitivity and precision for upland blanket bog between 83% and 92%, however they used random subsets of the sample points to calculate these accuracies, which as has been discussed can lead to over optimistic accuracy estimation (Meyer et al, 2018), (Linnenbrink et al, 2024). Similarly, Ingle et al (2023) reported classification accuracies between 83% and 99% for ecotypes existing within raised bogs using a random sampling technique. Furthermore, Karlson & Bastviken (2023) reported sensitivities between 71% and 90%, and precisions between 65% and 92% for classifying different wetland types in regions around the world. They note that the global classifier, trained using reference data aggregated across all five study sites, consistently underperformed as compared to the regional classifiers, highlighting the challenges of achieving reliable global generalization. They also used random sub-sampling during testing which could explain the discrepancy in accuracies reported from the regional models and the global model.

5.2. DISCUSSION OF RESULTS FOR EACH CLASSIFICATION TASK FROM THE EXPERIMENTS

The first research question of this study is to ask: to what extent can different degradation classification systems be classified using traditional machine learning algorithms and a broad suite of optical, SAR and elevation derived predictors? The following section aims to answer this question by examining each classification task and how it performed under each experiment. Table 30 provides a high-level overview of the results from each classification task for each experiment

	CLASSIFICATION TASK	OVERALL ACCURACY	P-VALUE
EXPERIMENT 1	Fossitt	63.81	0.00004509
	Annex I	80.27	2E-16
	Condition	57.68	0.8337516
EXPERIMENT 2	Fossitt	64.32	0.000009433
	Annex I	69.81	2E-16
	Condition	50.17	0.9925
EXPERIMENT 3	Fossitt	55.15	0.982243
	Annex I	80.27	2E-16
	Condition	48.97	0.9994
EXPERIMENT 4	Fossitt	58.4	0.44164
	Annex I	60.63	5.52E-10
	Condition	49.49	0.998154

Table 30. Overall accuracy and p-value for each classification task under each experiment

5.2.1 FOSSITT CLASSIFICATION

The model performed significantly better than a majority-class baseline when optical information was retained (Bicego & Mensi, 2023) but the 10% reduction in accuracy following the removal of Sentinel-1 and ancillary predictors highlights the added value of multi-source data (Table 31). This can be seen from the consistent DTW selection as well as cross-polarised backscatter (VH, VH/VV) suggesting these predictors provide some discriminative information during training. The marginal improvement observed when a single optical season was used instead of three suggests potential redundancy within multi-season optical composites. Overall, classification performance was primarily driven by optical data but was enhanced through the inclusion of SAR and ancillary predictors.

Eroding bog (PB5) was the best classified response variable and the best evaluation metrics resulted from experiment 2, when only season two was available in the optical imagery data type. This is possibly because there is greater contrast between bare soil and unhealthy vegetation, and healthy vegetation during the peak growing season as compared to pre-growth and post-growth seasons. The selection of the B8/B5 (NIR / RedEdge) band ratio, which is sensitive to vegetation health and chlorophyll content (Bandyopadhyay et al, 2017), further supports this.

The cutover bog consistently underperformed the other two classes in this Fossitt classification task. It consistently produced the lowest sensitivity and precision for this

task indicating that the model struggled to distinguish this class from eroding and blanket bog. The confusion matrices shows that cutover bog was consistently confused with eroding bog, across all four experiments. This is unsurprising considering the similarities of these two classes. The model also had difficulty distinguishing the blanket bog (PB2) class from the eroding bog (PB4) class, though to a lesser degree.

The results of this classification task suggests that there is some potential for distinguishing between different Fossitt habitat classes within the upland blanket bog habitat when a wide suite of data is available during training. Notably, the cutover bog was not well classified during any of the experiments. This habitat is usually associated with raised bogs, but it does occur in the uplands as well. This points to the challenge of classifying land in such dynamic mosaiced landscapes.

EXPERIMENT	CLASSIFICATION TASK	INPUT PREDICTORS	SELECTED PREDICTORS	OVERALL ACCURACY	P-VALUE
1	Fossitt	All predictor variables used	mean_VH | ratio_median | 2_B11 | DTW | 2_NDRE | 1_B2 | ratio_stD	63.81	4.51E-05
2	Fossitt	All families only season 2	mean_VH | ratio_median | 2_B11 | DTW | 2_SR_B8_B5 | burntarea_historic | ratio_stD | TWI	64.32	9.43E-06
3	Fossitt	Just optical	1_B3 | 2_NDVI | 2_B12 | 2_B3 | 3_EVI | 1_GNDVI	55.15	0.982243
4	Fossitt	No optical	mean_VH | ratio_median | DTW | max_VH | ratio_max	58.4	0.44164

Table 31. Fossitt results from experiments 1-4

5.2.2 ANNEX CLASSIFICATION

This classification task was the most reliant on optical predictors which is reflected in the highest overall accuracy when provided with all three optical seasons and lowest performance when no optical data was available (Table 32). This is reinforced by no observed reduction in performance when SAR and ancillary data types were removed. The P-value indicated a strong classification ability compared to no-information baseline which speaks to the balanced classes in this data set and indicates spectral differences between classes.

Intact blanket bog is strongly linked to vegetation cover, and this is reflected in the model's selection of predictors related to vegetation cover and condition (EVI, NDRE, NDVI, VV/VH see **Table 32**). Additionally, due to the impact of fires on vegetation, the burnt area predictor was able to support model classification when optical predictors were either missing or less abundant as was the case in experiments 2 and 4. Interestingly the Annex I classification task with no optical data was the only model run that selected for DEM pointing to a reliance on topography when vegetation-based indicators were lacking. This reflects what has been noticed in the field, that areas of bogs in poor condition are often concentrated on mountain peaks or areas of higher elevation (NPWS, 2017a).

EXPERIMENT	INPUT PREDICTORS	SELECTED PREDICTORS	OVERALL ACCURACY	P-VALUE
1	All predictor variables used	1_EVI | 2_NDRE | 3_B12 | 2_NDVI | 3_BSI | 1_NDVI | 2_NDWI	80.27	2E-16
2	All families only season 2	2_B2 | 2_B3 | elevation | burntarea_historic	69.81	2E-16
3	Just optical	1_EVI | 2_NDRE | 3_B12 | 2_NDVI | 3_BSI | 1_NDVI | 2_BSI	80.27	2E-16
4	No optical	Elevation | ratio_max | ratio_min | ratio_stD | burntarea_historic	60.63	5.52E-10

Table 32. Annex I results from experiments 1-4

This classification task presents the greatest potential in terms of monitoring changes to the landscape from non-intact blanket bog to intact blanket bog areas over time. The binary nature of this task certainly makes this relatively simple to compute. Nevertheless, this classification task is highly relevant to ongoing efforts to restore peatlands and enhance their function as carbon sinks in climate change mitigation strategies. Further study is needed, however, to explicitly evaluate the potential for applying this technique for monitoring temporal changes as well as the important distinction of peat forming or not.

5.2.3 CONSERVATION STATUS CLASSIFICATION

The Conservation Status classification task performed the most poorly throughout all four experiments (Table 33). This is most likely due to high class imbalance and the higher number of classes to be predicted. This classification task was most reliant on SAR data as compared to the other two classification tasks. The reliance on this data

type is also reflected in the lowest overall accuracy (by a small margin) when this data type was removed from the predictor stack. Although this reliance on SAR data is observed, conclusions about the usefulness of this data type in classifying Conservation Status remain limited due to the poor performance of this classification task. This assessment is most supported by the high p-value that was recorded for every experiment. This indicates that this model was systematically unable to produce statistically significant classification predictions.

Curiously, this classification did not select many predictor variables, the most being five and this resulted in the lowest overall accuracy. This classification task was the only one to select for VV polarization and this was selected for when the model produced the highest overall accuracy indicating it has some differentiating information for this classification task. The methods used in this study did not produce satisfactory results for this classification task of distinguishing between 'favourable', 'inadequate' and 'bad' conservation statuses. The addition of the fourth 'unlisted' class to represent areas that were not assigned a conservation status by the surveyor negatively impacted the predictive power of this classification task. Although it had only the second smallest representation in the sample data, it vastly under-performed the least represented favourable class. This is possibly because these areas did not represent a coherent ecological or spectral category but instead encompassed a heterogeneous mixture of land cover conditions and transitional states.

EXPERIMENT	INPUT PREDICTORS	SELECTED PREDICTORS	OVERALL ACCURACY	P-VALUE
1	All predictor variables used	max_VV | mean_VH | 3_B2	57.68	0.8337516
2	All families only season 2	max_VH | ratio_mean | burntarea_historic	50.17	0.9925
3	Just optical	2_B2 | 2_SR_B8_B5 | 2_NDVI | 2_EVI | 3_EVI	48.97	0.9994
4	No optical	max_VH | ratio_mean	49.49	0.998154

Table 33. Conservation Status results from experiments 1-4

5.3 PREDICTOR SELECTION

The second research question of this study revolves around the usefulness of different predictors for these classification tasks. Certain patterns emerged from the experiments that can inform future work in this area. As mentioned above, due to the

low performance of the Conservation Status classification task, we are unable to draw conclusions around the usefulness of SAR data in distinguishing Conservation Status, though the model's propensity for selecting of SAR data is noted.

Moving forward, the FFS produced some insightful patterns for the Fossitt and Annex I classification tasks. The Fossitt classification tasks benefitted from a wider range of data supporting the hypothesis that classification can be improved using richer, more dynamic data sets as was observed by Karlson & Bastviken (2023) in their study. However, this was not the case in the Annex I classification tasks, where a clear preference for optically derived predictors was observed.

Although these findings are noted, it remains challenging to compare these patterns and draw strong conclusions due to the imbalance in the number and type of predictors derived from each data type. The optical predictor set had three seasonal composites each comprising of 9 bands and 9 derived indices. The full optical predictor set had 19 predictors total that highlighted a variety of landscape features that are relevant to blanket bog degradation. In contrast the SAR predictor set only had temporal aggregates of the two polarizations and their ratio and no specific indices. Furthermore, the ancillary predictor set was comparatively weak with only 5 total predictors. Therefore, the strength of the optical predictor set, compared to the SAR and ancillary predictor sets should be interpreted with caution. That being said, in Karlson & Bastviken, optical data did present better classification results than SAR data while the best result was achieved with mixed sensor approach (Karlson & Bastviken, 2023).

Despite these limitations, looking at each predictor and what they are sensitive to, the model did select predictors that represent, hydrological processes, bare peat exposure and vegetation health, see **Table 32** and **Table 31**. Notably the model for the Conservation Status classification tasks did not select predictors representing these three components of peatland degradation. When the model selected the most predictors, they were all related to vegetation health. Returning to the second research question of this study and looking at how the model performed when selecting different predictors, these observations could imply that for classification to be successful, that the three components of peatland degradation should be incorporated into a classifier's predictors.

5.4 PREDICTED MAPS

Predicted maps were generated only for the Fossitt and Annex I classification tasks from Experiment 1. Although the Fossitt classification achieved a marginally higher accuracy in Experiment 2, the improvement was small. Restricting prediction comparisons to a single experiment therefore improves comparability across classification tasks, while little additional insight would be gained from using the marginally better-performing experiment. The Conservation Status prediction map is provided in APPENDIX E. Due to the low predictive performance of this classification task, it is not discussed further in this section.

Under the Fossitt classification prediction, the area of study according to this degradation classification system is in bad condition (See Figure 16) This is in agreement with current reporting on the conservation status for blanket bogs as published in The Status of EU Protected Habitats and Species in Ireland 2025 (NPWS, 2025). This publication reports the current threats to peat extraction, overgrazing, burning, afforestation, and agricultural activities causing nitrogen deposition (NPWS, 2025). Of these threats, only peat extraction (BSI), overgrazing (vegetation indices) and burning (burnt area) were represented in the predictors. A raster based on distance to afforested sites, agricultural activities and historic peat extraction sites could be incorporated into the predictor set to represent these threats.

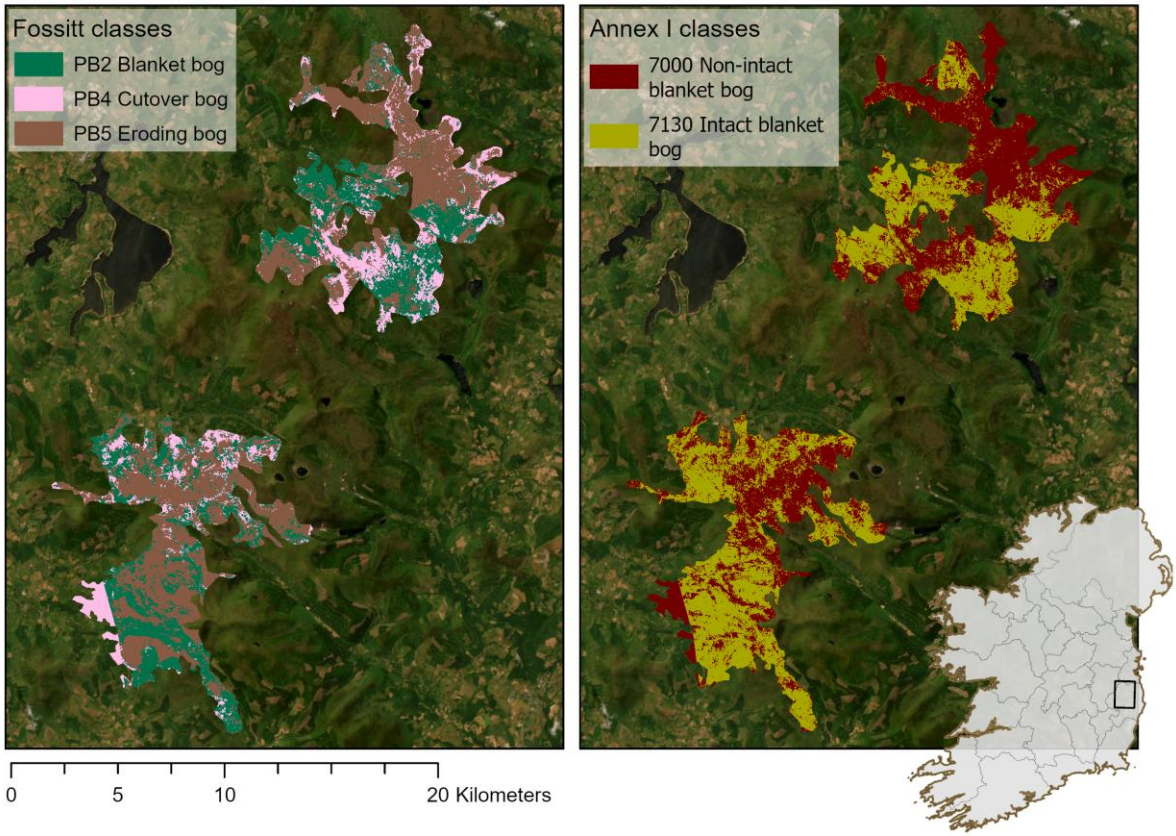


Figure 16. Prediction map under Fossitt classification system, experiment 1. Green: blanket bog, brown: eroding bog, purple: cutover bog.

The confusion matrix showed the model struggling to accurately predict cutover bog (PB4) comparing the predicted map for this class to satellite imagery basemap, this class was badly predicted and over predicted this class in many cases. The confusion matrix showed the model struggling to accurately predict cutover bog (PB4) comparing the predicted map for this class to satellite imagery basemap, this class was badly predicted and over predicted this class in many cases. Cutover bogs are more likely to appear in easier to access areas for example in proximity to roads. It is plausible that cutover bog is over-represented in the sample data as applied to this specific area of study. Despite this the model was more likely to classify areas as eroding blanket bog rather than blanket bog. This implies a potential bias towards over-estimating degradation within the mapped area.

Visually inspecting different areas of the map, the model seemed to be able to identify areas where the blanket bog was eroding or had weaker structural integrity classifying

such areas as non-intact blanket bog (7000) under the Annex I classification system and eroding bog (PB5) under the Fossitt classification system (see Figure 17).

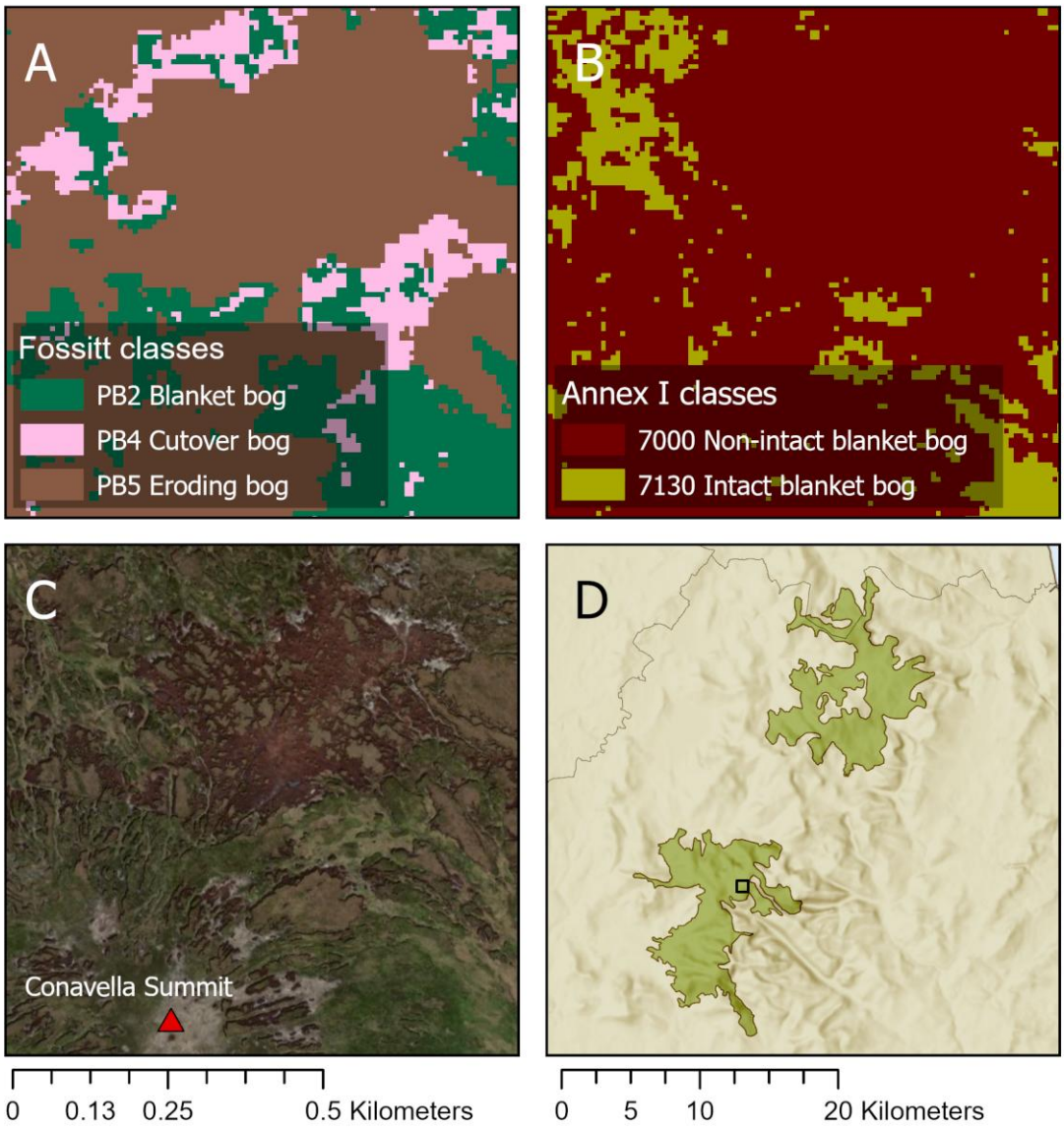


Figure 17. Area of degraded bog. A: Fossitt classification prediction. B: Annex I classification prediction. C: High resolution satellite imagery basemap showing degraded, eroding area neat the summit of Conavalla. D: Location and extent of maps A to C.

It remains challenging to visually assess whether the model was able to predict blanket bog (PB2) correctly. One area that has been reported as 'well developed' is Liffey head bog, located in the northern portion of the area of study (NPWS, 2017b). This site was visited in August 2025, and while the summit of Kippure exhibited signs of severe erosion (see Figure 10) the area adjacent to the R115 road showed signs of recovery from restoration work that had been done since 2021. This area was flagged for requiring restoration work as it is the source of the River Liffey and was severely

degraded. This site was classified as eroding bog (PB5) by the model (see Figure 18). While this area was not 'eroding', it was degraded at the time when the data was gathered for this model (2019) and this was correctly identified by the model. This area was mostly classified as non- intact blanket bog (7000).

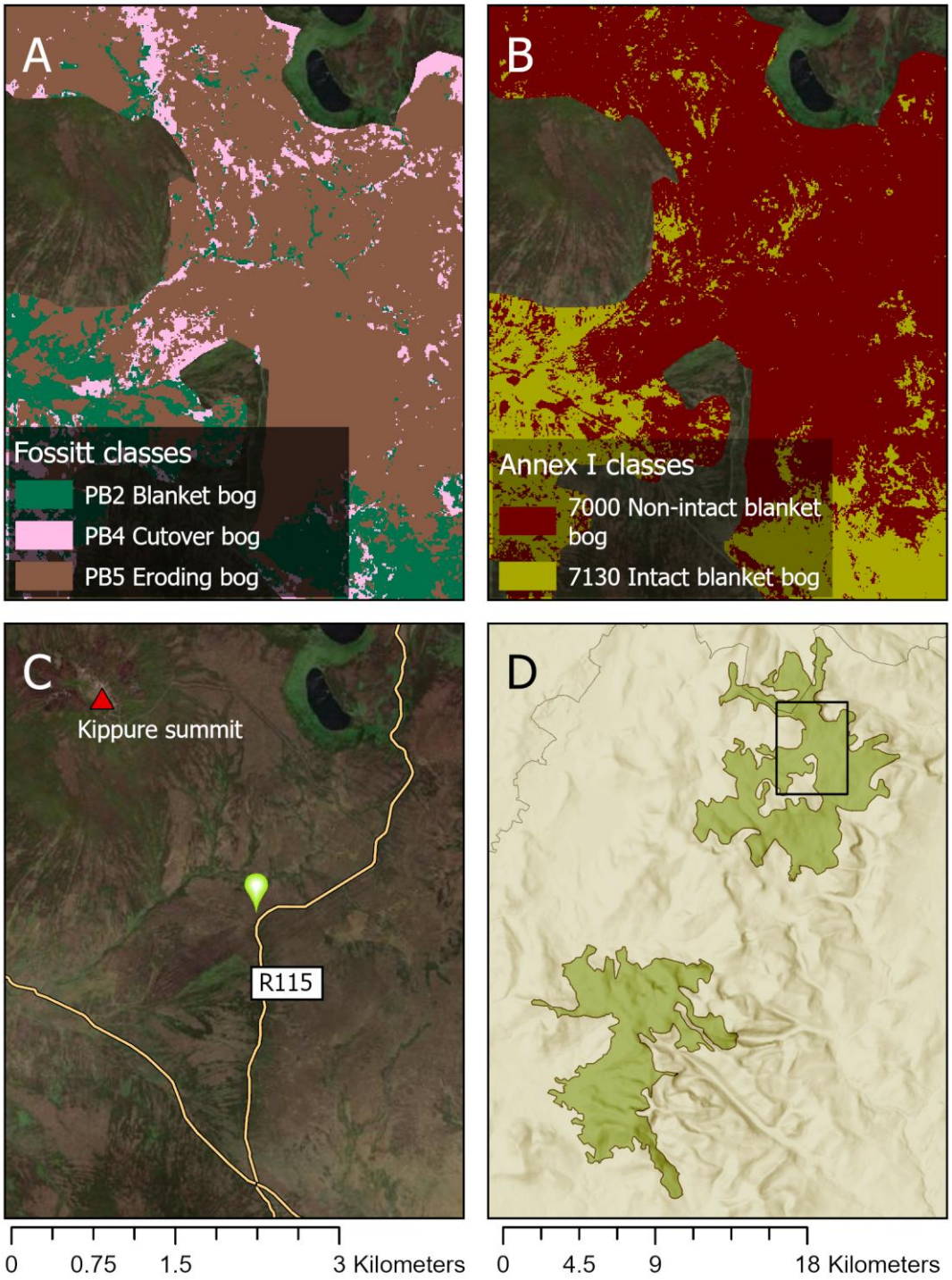


Figure 18. Area along the R115. A: Fossitt classification. B: Annex I classification. C. High resoltion satellite imagery basemap showing location of restoration work. D: Location of maps within area of study

5.4 Prediction maps and AOA, LPD and DI

The AOA, DI, and LPD are valuable tools for interpreting prediction maps, as they provide an explicit indication of model certainty and uncertainty across the mapped area. Figure 19 and Figure 20 shows the prediction where transparency of pixels are based on LPD values, darker and more saturate areas are more certain and paler more transparent areas are less certain providing a visual indicator of how densely represented a pixel is in the model predictor space.

These calculated rasters enable the identification of regions that are poorly represented in the training data and therefore lie outside the model's reliable domain. The benefits of this approach are two-fold: first, predictions in areas lacking adequate representation can be treated with caution or excluded from interpretation; and second, such areas can be used to guide future data collection efforts to ensure that environmental conditions within the area of interest are more comprehensively represented in the training dataset.

The LPD gave an interesting and very dynamic view of the landscape to be predicted. Visually inspecting the LPD and comparing to basemap satellite imagery we can interrogate which areas the model might have more difficulty predicting. The LPD is more difficult to directly compare to predictor variables as can be done with the DI as it represents a more computationally complex calculation. However, by assigning a transparency value based on the LPD value we can use this additional level of information to enhance map accuracy interpretation.

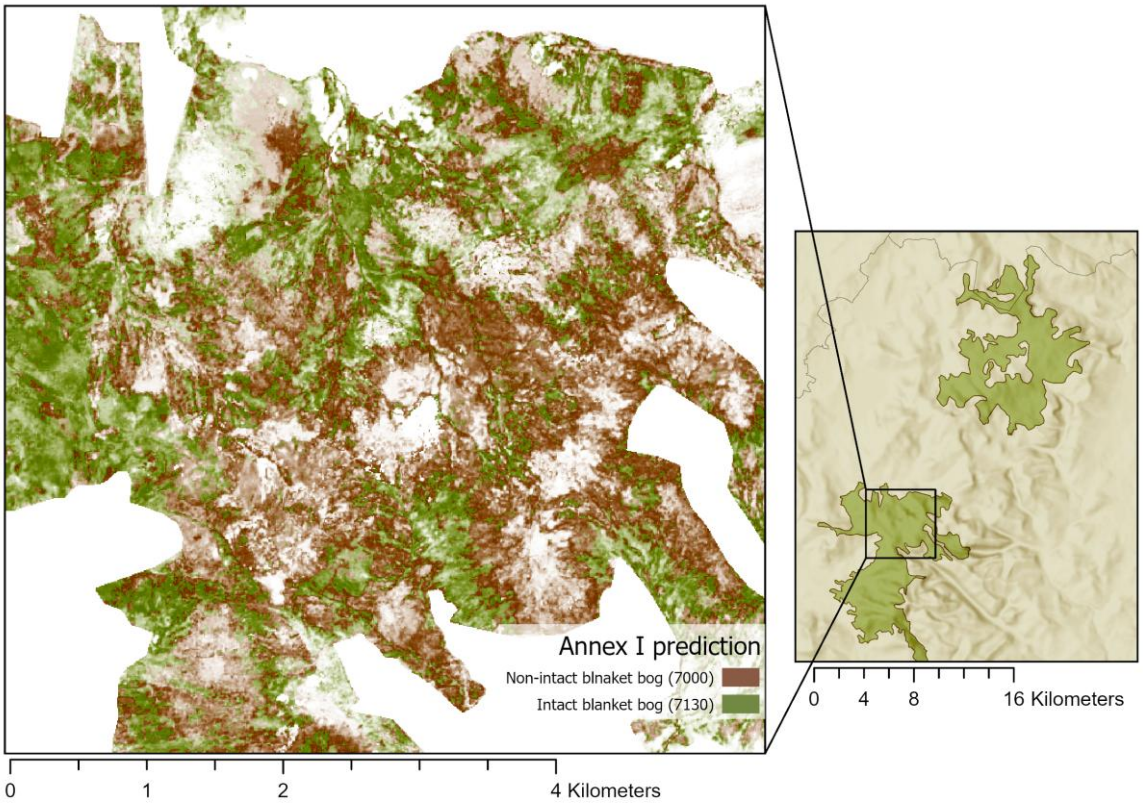


Figure 19. Annex classification with pixel transparency dependent on level of LPD

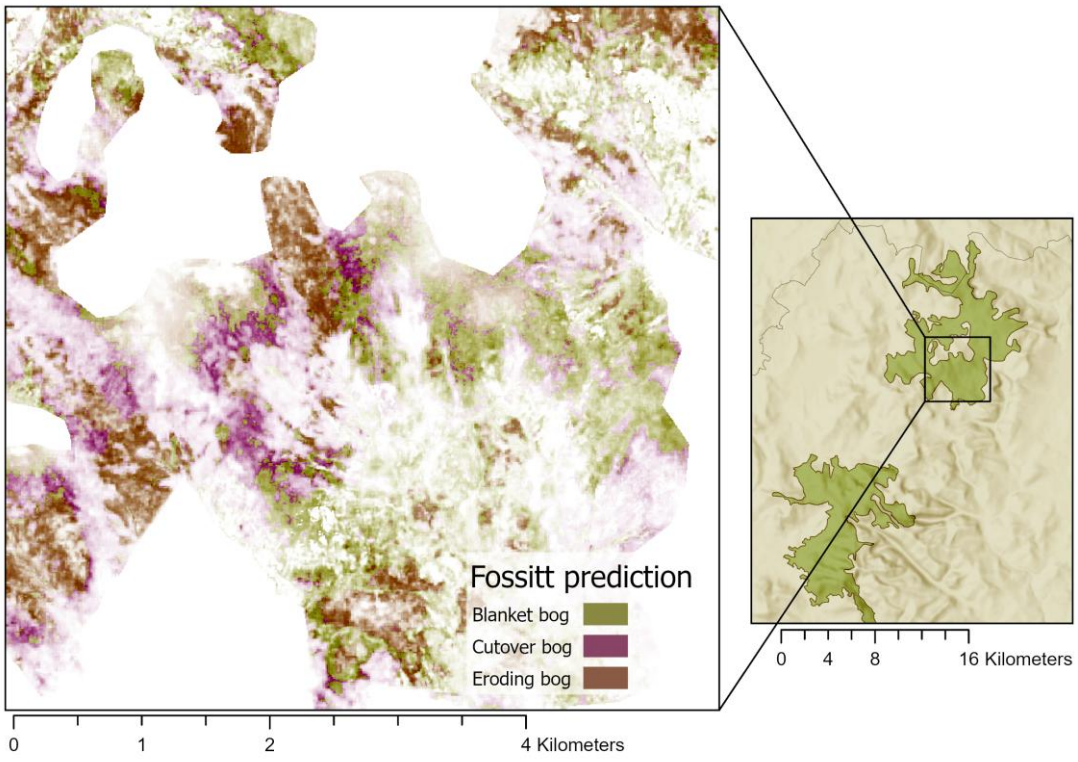


Figure 20. Fossitt classification with pixel transparency dependent on level of LPD

The DI was particularly useful for examining the predictors and identifying unusual patterns. For example, the DI highlighted certain areas that were far outside the range of the training data. Inspecting these areas on the satellite imagery basemap revealed that they should not have been so strongly flagged. However, by combining this information with the predictors selected by the model, it was possible to pinpoint specific predicotors that contained irregularities, such as cloudy pixels (Figure 21).

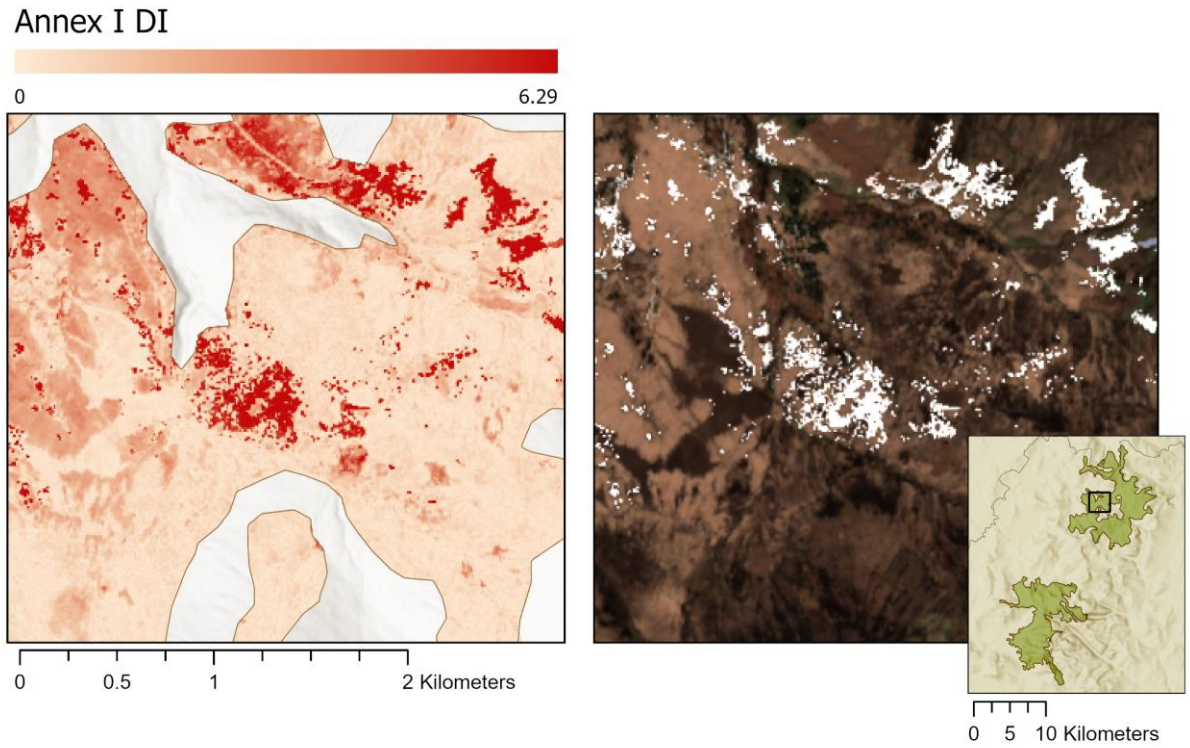


Figure 21. Left: DI Calculated for the Annex I classification task from experiment 1. Right: RGB colour composite from season 1 Sentinel-2 imagery depicting masked cloud contamination.

6. CONCLUSION

The first research question this study aimed to answer is the extent to which different classification systems can be predicted in specific landscapes. The results of the model and experiments suggests that intra-habitat classification is possible if sample data is well distributed and number of classes to be distinguished remain low. This is evidenced in the high performance of the Annex I classification tasks, a binary classification. However, sample data with better coverage and representativeness of the landscape would improve classification.

In addition, we asked: what challenges are associated with applying already existing methods of classifying raised bog degradation to upland blanket bogs? This question in many ways defines the previous question in terms of limitations and challenges to this method. Nevertheless, this study found these methods are applicable to upland blanket bog despite certain challenges. Due to accessibility constraints and the mosaiced nature of blanket bogs (Fossitt, 2000) it is difficult to draw clear boundaries as to their extent. The mosaiced nature of upland areas makes intra-habitat condition classification using machine learning and remote sensing particularly difficult.

However, this specific challenge of the mosaiced landscape and consequent blurry boundaries between habitats in the uplands was to some degree been addressed with the incorporation of AOA calculations. They provided visual indications of how represented areas are in the sample data, there is potential for these calculations and their use in analyses to identify areas that are not the habitat of interest.

Although accessibility and financial constraints means that current ground surveys constrain machine learning capabilities in terms of coverage and representativeness of sample data, clear understanding on the state of blanket bogs is vital for future monitoring of restoration activities. The potential for carbon sequestration in these landscapes depending on their condition is a significant incentive to develop better strategies for mapping their condition. Better information on the location and extent of degraded areas could also help steer decision making around restoration efforts.

In order to develop these strategies and to plan for this type of work, understanding how our prediction models function and what data is needed is essential. The second question this study undertook was to establish whether the three components of peatland condition, hydrology, exposed peat and vegetation, which are used by field surveyors to assess peatland condition, is also used by the model in classifying degradation classification systems. As was discussed in section 5.3 Predictor selection, this was found for the most part to be the case. The selected predictors often represented the three facets for the best performing classifications task though not necessarily evenly.

6.1 LIMITATIONS

The data available for this study presented some limitations. It was not collected for the purposes of the methodologies used in this study. The inclusion of mosaiced areas necessitated the removal of some areas, reducing the coverage of sample data. Due to lack of representation of 7130* as opposed to 7130, this study was not able to distinguish between these two classes and instead referred to areas in this classification system as 7130 or 7000 meaning the important peat forming attribute of some areas was not considered

A key driver to blanket bog degradation is overgrazing by sheep and deer. This is particularly acute in the Wicklow Mountains, and it is very challenging to account for in such studies. It is nevertheless important to acknowledge this missing component from the study.

As discussed in section 5.3 Predictor selection, the predictors were unevenly distributed between SAR, optical and ancillary data as well as between the three aspects of peatland degradation. Therefore, strong conclusions as to how useful data types are when applied to these methods are limited. Without control predictors (unrelated to vegetation, hydrology, bare peat) and uneven distribution of predictors between these three facets, degradation it is difficult to procure strong statements as to their importance in classifying areas of blanket bog according to their condition. Though it is noted that vegetation was the most useful facet while hydrology and bare peat exposure provided support.

6.2 FUTURE WORK

Building upon the limitations outlined above, future work could address these challenges through the derivation of additional predictors, such as distance to conifer plantations, proximity to historic industrial peat extraction sites, and distance to farmland under the ancillary data type. Also, additional indices could have been derived from the SAR imagery.

Furthermore, the optical predictor stack could have been refined by removing similar indices or unnecessary bands. This could be challenging as the model did select

similar predictors (e.g. season 1 NDVI and season 2 NDVI in experiment 1, Annex I classification task).

In order to test the generalisability of the model it would be pertinent to collect data in two different areas of blanket bog, preferably in geographically distant regions and to test the applicability of a model trained in one region but applied to another, similar to Karlson & Bastivken's (2023) approach.

Using remote sensing and machine learning to make the distinction of peat forming or not is a significant area of future work but would require the careful collection of data to clearly represent area of peat forming blanket bogs though these areas are rare. Furthermore, data collection with the purpose of employing remote sensing technology and machine learning methodology in mind would also have improved the quality and useability of the sample data. Greater input from field experts would have provided improved evaluation of model prediction i.e. relying on more general, in field knowledge of blanket bog condition and their location to validate predictions.

BIBLIOGRAPHICAL REFERENCES

Bandyopadhyay, D., Bhavsar, D., Pandey, K., Gupta, S., & Roy, A. (2017). Red edge index as an indicator of vegetation growth and vigor using hyperspectral remote sensing data. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *87*(4), 879-888. https://doi.org/10.1007/s40010-017-0456-4

Barrett, B., Raab, C., Cawkwell, F., & Green, S. (2016). Upland vegetation mapping using Random Forests with optical and radar satellite data. *Remote Sensing in Ecology and Conservation*, *2*(4), 212–231. https://doi.org/10.1002/rse2.32

Bechtold, M., Schlaffer, S., Tiemeyer, B., & De Lannoy, G. (2018). Inferring Water Table Depth Dynamics from ENVISAT-ASAR C-Band Backscatter over a Range of Peatlands from Deeply-Drained to Natural Conditions. *Remote Sensing*, *10*(4), 536. https://doi.org/10.3390/rs10040536

Betbeder, J., Hubert-Moy, L., Burel, F., Corgne, S., & Baudry, J. (2015). Assessing ecological habitat structure from local to landscape scales using synthetic aperture radar. *Ecological Indicators*, *52*, 545-557.

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. https://doi.org/10.1016/j.jag.2020.102083

Bicego, M., & Mensi, A. (2023). Null/No Information Rate (NIR): a statistical test to assess if a classification accuracy is significant for a given problem. *arXiv preprint arXiv:2306.06140*.

Bonn, A., Allott, T., Evans, M., Joosten, H., & Stoneman, R. (Eds.). (2016). Peatland restoration and ecosystem services: science, policy and practice. *Cambridge University Press.*

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Carless, D., Luscombe, D. J., Gatis, N., Anderson, K., & Brazier, R. E. (2019). Mapping landscape-scale peatland degradation using airborne lidar and multispectral data. *Landscape Ecology*, *34*, 1329-1345.

Conaghan, J., Douglas, C., Grogan, H., O' Sullivan, A., Kelly, L., Garvey, L., Van Doorslaer, L., Scally, L., Dunnells, D., & Wyse Jackson, M., Goodwillie, R., Mooney, E. (2000). The Distribution, Ecology and Conservation of Blanket Bog in the Republic of Ireland. *Enviroscope Environmental Consultancy.*

Connolly, J., Holden, N. M., Seaquist, J. W., & Ward, S. M. (2011). Detecting recent disturbance on Montane blanket bogs in the Wicklow Mountains, Ireland using the MODIS enhanced vegetation index. *International journal of remote sensing*, *32*(9), 2377-2393.

Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (Habitats Directive) and Directive 2009/147/EC (codified version of Directive 79/409/EEC as amended) (Birds Directive) – transposed into Irish law as European Communities (Birds and Natural Habitats) Regulations 2011 (SI 477/2011).

Cruz, C., Perrin, P. M., Martin, J. R., O'Connell, J., McGuinness, K., & Connolly, J. (2024). Mapping of temperate upland habitats using high-resolution satellite imagery and machine learning. *Environmental Monitoring and Assessment*, *196*(9), 869. https://doi.org/10.1007/s10661-024-12998-0

Cutler, M. E. J., McMorrow, J., & Evans, M. (2002). Remote sensing peat erosion in the southern Pennines. *North West Geography*, *2*(1), 21-30.

Daily, G. (1997). Introduction: What Are Ecosystem Services. *Nature's services: Societal dependence on natural ecosystems.* Island Press, Washington, DC.

DeLancey, E. R., Kariyeva, J., Bried, J. T., & Hird, J. N. (2019). Large-scale probabilistic identification of boreal peatlands using Google Earth Engine, open-access satellite data, and machine learning. *PloS one*, *14*(6), e0218165. https://doi.org/10.1371/journal.pone.0218165

De Waard, F., Connolly, J., Barthelmes, A., Joosten, H., & Van Der Linden, S. (2024). Remote sensing of peatland degradation in temperate and boreal climate zones – A review of the potentials, gaps, and challenges. *Ecological Indicators*, *166*, 112437. https://doi.org/10.1016/j.ecolind.2024.112437

European Nature Information System [EUNIS], (2013). Habitats Annex I Directive, Blanket bogs (* if active bog) Fact Sheet. *Interpretation Manual of European Union Habitat version EUR 28.* European Commission, DG-ENV. URL: https://eunis.eea.europa.eu/habitats/10144

Evans, M., & Lindsay, J. (2010). High resolution quantification of gully erosion in upland peatlands at the landscape scale. *Earth Surface Processes and Landforms*, *35*(8), 876-886.

Farrell, C.A., Coleman, L., Norton, D., Kelly-Quinn, M., Obst, C., Eigenraam, M., O'Donoghue, C., Kinsella, S., Smith, F., Sheehy, I. and Stout, J.C., (2021). Developing peatland ecosystem accounts to guide targets for restoration. One Ecosystem 6: e76838. https://doi.org/10.3897/oneeco.6.e76838

Flynn, R., Mackin, F., & Renou-Wilson, F. (with Environmental Protection Agency). (2021). *Towards the quantification of blanket bog ecosystem services to water (2015-NC-MS-5)* (Online version). Environmental Protection Agency.

Fossitt, J. A. (2000). *A guide to habitats in Ireland*. Heritage Council/Chomhairle Oidhreachta

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. https://doi.org/10.1016/S0034-4257(96)00067-3

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., … Kalogirou, S. (2021). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, *36*(2), 121–136. https://doi.org/10.1080/10106049.2019.1595177

Ghazaryan, G., Krupp, L., Seyfried, S., Landgraf, N., & Nendel, C. (2024). Enhancing peatland monitoring through multisource remote sensing: Optical and radar data

applications. *International Journal of Remote Sensing*, *45*(18), 6372–6394. https://doi.org/10.1080/01431161.2024.2387133

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, *202*, 18-27.

Grappiolo, C., Gurusiddappa, V., Regan, S., Boydell, O., & Holohan, E. (2024, April). A Satellite-derived Peatland Ecotype Classification Method Using Artificial Neural Network Hierarchical Ensembles. In *EGU General Assembly Conference Abstracts* (p. 19643).

Gu, Z., & Zeng, M. (2024). The Use of Artificial Intelligence and Satellite Remote Sensing in Land Cover Change Detection: Review and Perspectives. *Sustainability*, *16*(1), 274. https://doi.org/10.3390/su16010274

G.W. Geerlingl, C. B. L. van Gestel, M. Sheehy Skeffington, M. G. C. Schouten P. H. Nienhuis & R. S. E. W. Leuvenl (2002). Blanket Bog Degradation in River Catchments in the West of Ireland. *Application of Geographic and Remote Sensing in River Studies* (pp. 25-40). Backhuys Publishers.

Habib, W., Cresson, R., McGuinness, K., & Connolly, J. (2024). Mapping artificial drains in peatlands—A national-scale assessment of Irish raised bogs using sub-meter aerial imagery and deep learning methods. *Remote Sensing in Ecology and Conservation*, *10*(4), 551-562. https://doi.org/10.1002/rse2.387

Harris, A., & Baird, A. J. (2019). Microtopographic Drivers of Vegetation Patterning in Blanket Peatlands Recovering from Erosion. *Ecosystems*, *22*(5), 1035–1054. https://doi.org/10.1007/s10021-018-0321-6

Hu, X., Li, L., Huang, J., Zeng, Y., Zhang, S., Su, Y., Hong, Y., & Hong, Z. (2024). Radar vegetation indices for monitoring surface vegetation: Developments, challenges, and trends. *Science of The Total Environment*, *945*, 173974. https://doi.org/10.1016/j.scitotenv.2024.173974

Ingle, R., Habib, W., Connolly, J., McCorry, M., Barry, S., & Saunders, M. (2023). Upscaling methane fluxes from peatlands across a drainage gradient in Ireland using PlanetScope imagery and machine learning tools. *Scientific Reports*, *13*(1), 11997. https://doi.org/10.1038/s41598-023-38470-6

IUCN UK Weathering, Erosion and Mass Movement of Blanket Bog, Briefing Note No. 9. (2014) *Committee Peatland Programme*

Joosten, H., 2009. The global peatland CO2 picture. Peatland status and drainage related emissions in all countries of the world. https://www.wetlands.org/publications/the-global-peatland-co2-picture/. Accessed 24 September 2021.

Karlson, M., & Bastviken, D. (2023). Multi-Source Mapping of Peatland Types Using Sentinel-1, Sentinel-2, and Terrain Derivatives—A Comparison Between Five High-Latitude Landscapes. *Journal of Geophysical Research: Biogeosciences*, *128*(4), e2022JG007195. https://doi.org/10.1029/2022JG007195

Kimmel, K., & Mander, Ü. (2010). Ecosystem services of peatlands: Implications for restoration. *Progress in Physical Geography: Earth and Environment*, *34*(4), 491–514. https://doi.org/10.1177/0309133310365595

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. https://doi.org/10.1111/avsc.12024

Large, A. R. G., & Hamilton, A. C. (1991). The distribution, extent and causes of peat loss in central and northwest Ireland. *Applied Geography*, *11*(4), 309–326. https://doi.org/10.1016/0143-6228(91)90020-A

Lees, K. J., Artz, R. R. E., Chandler, D., Aspinall, T., Boulton, C. A., Buxton, J., Cowie, N. R., & Lenton, T. M. (2021). Using remote sensing to assess peatland resilience by estimating soil surface moisture and drought recovery. *Science of The Total Environment*, *761*, 143312. https://doi.org/10.1016/j.scitotenv.2020.143312

Lendzioch, T., Langhammer, J., Vlček, L., & Minařík, R. (2021). Mapping the Groundwater Level and Soil Moisture of a Montane Peat Bog Using UAV Monitoring and Machine Learning. *Remote Sensing*, *13*(5), 907. https://doi.org/10.3390/rs13050907

Lindsay, R., Campus, S., & Lane, W. (2010). Peatbogs and carbon: a critical synthesis to inform policy development in oceanic peat bog conservation and restoration in the context of climate change. *RSPB Scotland*, *315*.

Linnenbrink, J., Milà, C., Ludwig, M., & Meyer, H. (2024). kNNDM CV: k-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. *Geoscientific Model Development*, *17*(15), 5897-5912. https://doi.org/10.5194/gmd-17-5897-2024

Macfarlane, F., Robb, C., Coull, M., McKeen, M., Wardell-Johnson, D., Miller, D., Parker, T. C., Artz, R. R. E., Matthews, K., & Aitkenhead, M. J. (2024). A deep learning approach for high-resolution mapping of Scottish peatland degradation. *European Journal of Soil Science*, *75*(4), e13538. https://doi.org/10.1111/ejss.13538

Mahdianpari, M., Brisco, B., Granger, J., Mohammadimanesh, F., Salehi, B., Homayouni, S., & Bourgeau-Chavez, L. (2021). The third generation of pan-Canadian wetland map at 10 m resolution using multisource earth observation data on cloud computing platform. *IEEE journal of selected topics in applied earth observations and remote sensing*, *14*, 8789-8803. https://doi.org/10.1109/JSTARS.2021.3105645

Matyukira, C., & Mhangara, P. (2024). Advances in vegetation mapping through remote sensing and machine learning techniques: A scientometric review. *European Journal of Remote Sensing*, *57*(1), 2422330. https://doi.org/10.1080/22797254.2024.2422330

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, *101*, 1-9. https://doi.org/10.1016/j.envsoft.2017.12.001

Meyer H, Milà C, Ludwig M, Linnenbrink J, Schumacher F (2026). *CAST: 'caret' Applications for Spatial-Temporal Models*. R package version 1.0.4, https://github.com/hannameyer/cast.

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications–Moving from data reproduction to

spatial prediction. *Ecological Modelling*, *411*, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*(9), 1620-1633. https://doi.org/10.48550/arXiv.2005.07939

Minasny, B., Adetsu, D. V., Aitkenhead, M., Artz, R. R. E., Baggaley, N., Barthelmes, A., Beucher, A., Caron, J., Conchedda, G., Connolly, J., Deragon, R., Evans, C., Fadnes, K., Fiantis, D., Gagkas, Z., Gilet, L., Gimona, A., Glatzel, S., Greve, M. H., … Zak, D. (2024). Mapping and monitoring peatland conditions from global to field scale. *Biogeochemistry*, *167*(4), 383–425. https://doi.org/10.1007/s10533-023-01084-1

Murphy, P. N. C., Ogilvie, J., & Arp, P. (2009). Topographic modelling of soil moisture conditions: a comparison and verification of two models. *European journal of soil science*, *60*(1), 94-109.

NASA Shuttle Radar Topography Mission (SRTM) Global,(2013). Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF.

National Parks and Wildlife Service [NPWS] (2025). The Status of EU Protected Habitats and Species in Ireland. Volume 1: Summary Overview. Unpublished NPWS report. URL: https://www.npws.ie/sites/default/files/publications/pdf/article-17-report-2025-volume-1.pdf

National Parks and Wildlife Service [NPWS] (2017a). Wicklow Mountains SAC (002122: Site synopsis. *Department of Culture, Heritage and the Gaeltacht.*URL: https://www.npws.ie/protected-sites/search/by-code?code=002122

National Parks and Wildlife Service [NPWS] (2017b). Wicklow Mountains SAC (site code 002122) Conservation objectives supporting document - blanket bogs and associated habitats NPWS. *Department of Culture, Heritage and the Gaeltacht.* URL: https://www.npws.ie/protected-sites/search/by-code?code=002122

National Parks and Wildlife Service [NPWS] (2005). *Management Plan for Wicklow Mountains National Park 2005–2009*. Department of the Environment, Heritage and Local Government, Ireland. URL: https://www.npws.ie/sites/default/files/publications/pdf/WMNP.pdf

Parish, F., Sirin, A., Charman, D., Joosten, H., Minayeva, T., Silvius, M., Stringer, L. (Eds.), (2008). Assessment on peatlands, biodiversity and climate change: Main report. Global Environment Centre & Wetlands International Wageningen, Kuala Lumpur, 2 volumes.

Parish, F., Silvius, M., Reed, M. S., Stringer, L. C., Joosten, H., Suryadiputra, N., & Lin, C. K. (2007). Management of peatlands for biodiversity and climate change. In *Assessment on peatlands, biodiversity and climate change* (pp. pp-9). Global Environment Centre, Kuala Lumpur & Wetlands International Wageningen.

Pazúr, R., Price, B., & Atkinson, P. M. (2021). Fine temporal resolution satellite sensors with global coverage: an opportunity for landscape ecologists. *Landscape Ecology*, *36*, 2199-2213. https://doi.org/10.1007/s10980-021-01303-w

Räsänen, A., Tolvanen, A., & Kareksela, S. (2022). Monitoring peatland water table depth with optical and radar satellite imagery. *International Journal of Applied Earth*

Observation and Geoinformation, *112*, 102866. https://doi.org/10.1016/j.jag.2022.102866

Räsänen, A., Elsakov, V., & Virtanen, T. (2019). Usability of one-class classification in mapping and detecting changes in bare peat surfaces in the tundra. *International Journal of Remote Sensing*, *40*(11), 4083-4103. https://doi.org/10.1080/01431161.2018.1558376

Renou-Wilson, Florence, et al. *Peat Hub Ireland: 2022-NE-1129*. With Environmental Protection Agency, Online version, Environmental Protection Agency, (2025). Research Report, no. 494

Reynolds, N., Mota, B., & Nightingale, J. M. (2025). Open-access satellite data for peatland condition and restoration monitoring in the UK: a review. *Frontiers in Environmental Science*, *13*, 1685165. https://doi.org/10.3389/fenvs.2025.1685165

Robb, C., Pickard, A., Williamson, J. L., Fitch, A., & Evans, C. (2023). Peat drainage ditch mapping from aerial imagery using a convolutional neural network. *Remote Sensing*, *15*(2), 499. https://doi.org/10.3390/rs15020499

Schultz, S., Millard, K., Darling, S., & Chénier, R. (2023). Investigating the Use of Sentinel-1 for Improved Mapping of Small Peatland Water Bodies: Towards Wildfire Susceptibility Monitoring in Canada's Boreal Forest. *Hydrology*, *10*(5), 102. https://doi.org/10.3390/hydrology10050102

Schumacher, F., Knoth, C., Ludwig, M., & Meyer, H. (2024). Assessing the area of applicability of spatial prediction models through a local data point density approach. *European Geosciences Union General Assembly 2024 (EGU24)*, 10278. https://ui.adsabs.harvard.edu/link_gateway/2024EGUGA..2610278S/doi:10.5194/egusphere-egu24-10278

Sinergise. (2019). *Sentinel-2 Cloud Detector*. Sentinel Hub. https://github.com/sentinel-hub/sentinel2-cloud-detector

Stout, J. C., Farrell, C. A., Kelly-Quinn, M., Coleman, L., Kinsella, S., O'Donoghue, C., Norton, D., Obst, C., Eigenraam, M., Smith, F., Sheehy, I., & Zimmermann, S. (with Environmental Protection Agency). (2023). *Irish Natural Capital Accounting for Sustainable Environments (INCASE): 2018-NC-LS-2* (Online version). Environmental Protection Agency. https://www.epa.ie/publications/research/epa-research-2030-reports/Research_Report-441.pdf

Stupariu, M.-S., Cushman, S. A., Pleşoianu, A.-I., Pătru-Stupariu, I., & Fürst, C. (2022). Machine learning in landscape ecological analysis: A review of recent approaches. *Landscape Ecology*, *37*(5), 1227–1250. https://doi.org/10.1007/s10980-021-01366-9

Toca, L., Artz, R. R. E., Smart, C., Quaife, T., Morrison, K., Gimona, A., Hughes, R., Hancock, M. H., & Klein, D. (2023). Potential for Peatland Water Table Depth Monitoring Using Sentinel-1 SAR Backscatter: Case Study of Forsinard Flows, Scotland, UK. *Remote Sensing*, *15*(7), 1900. https://doi.org/10.3390/rs15071900

Trippier, B., Robinson, P., Colson, D. & Hutchison, J. 2020. Developing a framework for using Earth Observation imagery to monitor peatland condition. JNCC Report No. 667, JNCC, Peterborough, ISSN 0963-8091.

UNEP (2022). Global Peatlands Assessment – The State of the World's Peatlands: Evidence for action toward the conservation, restoration, and sustainable management of peatlands. Main Report. Global Peatlands Initiative. *United Nations Environment Programme, Nairobi.* https://wedocs.unep.org/handle/20.500.11822/41222

Wadoux, A. M. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, *210*, 103359. https://doi.org/10.1016/j.earscirev.2020.103359

Wicklow Uplands Council. (2018). *SUAS pilot project plan: Sustainable Uplands Agriculture-environment Scheme*. URL:
https://wicklowuplands.ie/suasproject/suas-documents/

Yin, L., Wei, W., Li, H., & Bo, L. (2026). Hidden risks in greening: Unveiling the impact of bare land changes on landscape ecological risks in arid and semi-arid regions of China. *Environmental Impact Assessment Review*, *117*, 108244. https://doi.org/10.1016/j.eiar.2025.108244

APPENDIX A

GEE Cloudless Sentinel-2 imagery for 2019

```
1
2   // ------------------------------------------------------------------
3   // 1. INPUTS & PARAMETERS
4   // ------------------------------------------------------------------
5
6   var roi = ee.FeatureCollection("users/your_username/path_to_your_asset");
7
8   // Year for analysis
9   var YEAR = 2019;
10  var YEAR_STR = YEAR.toString();
11
12  // Strict cloud probability threshold
13  var CLOUD_MASKING_THRESHOLD = 10;
14
15  // Initial filter for collection (discard images > 50% cloud cover early)
16  var CLOUD_FILTERING_THRESHOLD = 50;
17
18  // Define the bands to export (All relevant 10m/20m bands)
19  var BANDS = ['B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B8', 'B8A', 'B11', 'B12'];
20
21  // Define seasonal time windows (Astronomical Seasons)
22  var SEASONS = {
23    'Season_1': {start: YEAR_STR + '-01-01', end: YEAR_STR + '-04-30'},
24    'Season_2': {start: YEAR_STR + '-05-01', end: YEAR_STR + '-08-31'},
25    'Season_3': {start: YEAR_STR + '-09-01', end: YEAR_STR + '-12-31'}
26  };
27
28  // Load the cloud probability collection (Only filter by bounds globally)
29  var s2Clouds = ee.ImageCollection('COPERNICUS/S2_CLOUD_PROBABILITY')
30      .filterBounds(roi);
31
32  // ------------------------------------------------------------------
33  // 2. DATA PREPARATION FUNCTIONS
34  // ------------------------------------------------------------------
35
36  // --- Cloud and Shadow Masking Function ---
37  var maskCloudsAndShadows = function(image) {
38    // Get the time stamp of the current S2 image
39    var imageDate = image.date();
40
41    // Look for the corresponding cloud probability image by matching time and index
42    var cloud_img = ee.Image(s2Clouds
43      .filter(ee.Filter.date(imageDate.getRange('day'))) // Filter cloud collection to the S2 image's day
44      .filter(ee.Filter.eq('system:index', image.get('system:index')))
45      .first()
46    );
47
48    // If no cloud image is found, return the original image unmasked.
49    if (cloud_img === null) {
50        return image;
51    }
52
```

```
  // 1. Cloud Mask
  var cloud_probability = cloud_img.select('probability');
  var cloud_mask = cloud_probability.gt(CLOUD_MASKING_THRESHOLD);

  // 2. Cloud Shadow Mask
  var T_DARK = 0.2;
  var shadow_cdd = 10;

  var ir = image.select('B8').gt(T_DARK).rename('ir_test');
  var swir = image.select('B12').gt(T_DARK).rename('swir_test');
  var shadow_mask = ir.and(swir).not();

  // Simple morphological dilation to expand shadow
  shadow_mask = shadow_mask.focal_min(shadow_cdd);

  // 3. Combine and Apply
  var combined_mask = cloud_mask.or(shadow_mask).eq(0);

  // Update mask and return
  return image.updateMask(combined_mask);
};

// ---------------------------------------------------------------------
// 3. SEASONAL COMPOSITING FUNCTION
// ---------------------------------------------------------------------

var createSeasonalComposite = function(seasonName, dates) {
  var startDate = dates.start;
  var endDate = dates.end;

  // Load and filter S2 Surface Reflectance (Harmonized)
  var s2Sr = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED')
      .filterDate(startDate, endDate)
      .filterBounds(roi)
      .filterMetadata('CLOUDY_PIXEL_PERCENTAGE', 'less_than',
CLOUD_FILTERING_THRESHOLD);

  // Apply the cloud/shadow mask
  var s2Clean = s2Sr.map(maskCloudsAndShadows);

  // Create the median composite
  var composite = s2Clean.median()
                    .select(BANDS)
                    .clip(roi)
                    .set('system:index', seasonName);

  return composite;
};

// ---------------------------------------------------------------------
// 4. GENERATE AND EXPORT ALL COMPOSITES
// ---------------------------------------------------------------------

// Use the map function on the season dictionary to generate the four composites
var seasonalComposites = ee.Dictionary(SEASONS).map(createSeasonalComposite);

// --- Visualization (Optional) ---
var visParams = {min: 0, max: 3000, bands: ['B4', 'B3', 'B2']};
Map.centerObject(roi, 10);

// Add two composites to the map for inspection
```

```
Map.addLayer(seasonalComposites.get('Summer'), visParams, 'Summer Composite (Peak)');
Map.addLayer(seasonalComposites.get('Spring'), visParams, 'Spring Composite');

// --- Export Loop ---
// This loop iterates through the seasonal dictionary and exports each image.

ee.List(ee.Dictionary(SEASONS).keys()).getInfo().forEach(function(season) {
  var seasonalImage = ee.Image(seasonalComposites.get(season));

  // Export to Google Drive
  Export.image.toDrive({
    image: seasonalImage,
    description: 'S2_2019_' + season + '_Median',
    folder: 'earthengine',
    fileNamePrefix: 'S2_2019_' + season + '_Median',
    region: roi.geometry().bounds(),
    scale: 10,
    maxPixels: 1e13
  });
});

```

APPENDIX B

Python script for calculating indices

```
1
2   # 1. CONFIGURATION
3
4   # List your four seasonal composite files here
5   INPUT_FILES = [
6       r"C:\Users\msi fast\Desktop\Thesis work\Data\S2\BlanketBog_S2_2019_Clean.tif",
7       r"C:\Users\msi fast\Desktop\Thesis work\Data\S2\BlanketBog_S2_2019_Clean_1.tif",
8       r"C:\Users\msi fast\Desktop\Thesis work\Data\S2\BlanketBog_S2_2019_Clean_2.tif",
9       r"C:\Users\msi fast\Desktop\Thesis work\Data\S2\BlanketBog_S2_2019_Clean_3.tif"
10  ]
11
12  # Output folder for the calculated indices
13  OUTPUT_FOLDER = r"C:\Users\msi fast\Desktop\Thesis work\Data\S2\indeces_2"
14
15  # BAND MAPPING (0-based index)
16  # Based on your export: [B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12]
17  BANDS = {
18      'BLUE': 0,  # B2
19      'GREEN': 1, # B3
20      'RED': 2,   # B4
21      'RE1': 3,   # B5 (Red Edge 1) - Critical for NDRE
22      'NIR': 6,   # B8
23      'SWIR1': 8, # B11
24      'SWIR2': 9  # B12
25  }
26
27  # 2. CALCULATION FUNCTIONS
28
29  def calculate_index(name, bands, formula_func):
30      """
31      Helper to calculate an index safely ignoring divide by zero.
32      """
33      print(f"  Calculating {name}...")
34      try:
35          # Allow division by zero (results in inf/nan) to be handled later
36          with np.errstate(divide='ignore', invalid='ignore'):
37              result = formula_func(bands)
38
39          # Clean up infinite or NaN values (set to -9999)
40          result[~np.isfinite(result)] = -9999
41          return result
42      except Exception as e:
43          print(f"Error calculating {name}: {e}")
44          return None
45
46  def save_raster(output_path, data, reference_ds):
47      """
48      Saves the numpy array as a GeoTIFF using the reference dataset for projection.
49      """
50      driver = gdal.GetDriverByName('GTiff')
51      rows, cols = data.shape
52      out_ds = driver.Create(output_path, cols, rows, 1, gdal.GDT_Float32)
53
54      # Set GeoTransform and Projection from source
55      out_ds.SetGeoTransform(reference_ds.GetGeoTransform())
```

```
    out_ds.SetProjection(reference_ds.GetProjection())

    # Write data
    out_band = out_ds.GetRasterBand(1)
    out_band.WriteArray(data)
    out_band.SetNoDataValue(-9999)

    # Close/Flush
    out_ds = None
    print(f"    Saved: {os.path.basename(output_path)}")

# 3. MAIN PROCESSING LOOP

if not os.path.exists(OUTPUT_FOLDER):
    os.makedirs(OUTPUT_FOLDER)

for file_path in INPUT_FILES:
    if not os.path.exists(file_path):
        print(f"Skipping missing file: {file_path}")
        continue

    filename = os.path.splitext(os.path.basename(file_path))[0]
    print(f"\nProcessing: {filename}")

    # Open dataset
    ds = gdal.Open(file_path)
    if ds is None:
        print("Could not open raster.")
        continue

    # READ BANDS INTO MEMORY (Float32 for precision)
    b_blue  = ds.GetRasterBand(BANDS['BLUE'] + 1).ReadAsArray().astype(np.float32)
    b_red   = ds.GetRasterBand(BANDS['RED'] + 1).ReadAsArray().astype(np.float32)
    b_re1   = ds.GetRasterBand(BANDS['RE1'] + 1).ReadAsArray().astype(np.float32)
    b_nir   = ds.GetRasterBand(BANDS['NIR'] + 1).ReadAsArray().astype(np.float32)
    b_swir1 = ds.GetRasterBand(BANDS['SWIR1'] + 1).ReadAsArray().astype(np.float32)

    band_data = {
        'B': b_blue, 'R': b_red, 'G': b_green,
        'RE1': b_re1, 'N': b_nir, 'S1': b_swir1
    }

    # 3 DEFINE FORMULAS

    # 1. EVI (With Outlier Clamping)
    evi = calculate_index('EVI', band_data,
        lambda b: 2.5 * (b['N'] - b['R']) / (b['N'] + 6 * b['R'] - 7.5 * b['B'] + 1))

    if evi is not None:
        # Clamp values to valid biological range to remove math artifacts
        valid_mask = (evi != -9999)
        np.clip(evi, -1.5, 1.5, out=evi, where=valid_mask)
        save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_EVI.tif"), evi, ds)

    # 2. NDRE (Normalized Difference Red Edge)
    # Formula: (NIR - RedEdge) / (NIR + RedEdge)
    ndre = calculate_index('NDRE', band_data,
        lambda b: (b['N'] - b['RE1']) / (b['N'] + b['RE1']))
    if ndre is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_NDRE.tif"), ndre, ds)

    # 3. NDWI  (Moisture Index using SWIR)
```

```
    # Formula: (NIR - SWIR1) / (NIR + SWIR1)
    ndwi = calculate_index('NDWI', band_data,
        lambda b: (b['N'] - b['S1']) / (b['N'] + b['S1']))
    if ndwi is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_NDWI_MSI.tif"),
ndwi, ds)

    # 4. BSI (Bare Soil Index)
    # Formula: ((SWIR + Red) - (NIR + Blue)) / ((SWIR + Red) + (NIR + Blue))
    bsi = calculate_index('BSI', band_data,
        lambda b: ((b['S1'] + b['R']) - (b['N'] + b['B'])) / ((b['S1'] + b['R']) + (b['N'] + b['B'])))
    if bsi is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_BSI.tif"), bsi, ds)

    # 5. SR_B8_B5 (Red Edge Simple Ratio)
    # Formula: NIR / RedEdge
    sr_re = calculate_index('SR_B8_B5', band_data,
        lambda b: b['N'] / b['RE1'])
    if sr_re is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_SR_B8_B5.tif"),
sr_re, ds)

    # 6. SR_B8_B11 (SWIR Simple Ratio)
    # Formula: NIR / SWIR1
    sr_swir = calculate_index('SR_B8_B11', band_data,
        lambda b: b['N'] / b['S1'])
    if sr_swir is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_SR_B8_B11.tif"),
sr_swir, ds)

    # 7. NDVI
    ndvi = calculate_index('NDVI', band_data,
        lambda b: (b['N'] - b['R']) / (b['N'] + b['R']))
    if ndvi is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_NDVI.tif"), ndvi, ds)

    # 8. GNDVI
    gndvi = calculate_index('GNDVI', band_data,
        lambda b: (b['N'] - b['G']) / (b['N'] + b['G']))
    if gndvi is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_GNDVI.tif"), gndvi,
ds)

    # 9. RDVI
    rdvi = calculate_index('RDVI', band_data,
        lambda b: (b['N'] - b['R']) / np.sqrt(b['N'] + b['R']))
    if rdvi is not None: save_raster(os.path.join(OUTPUT_FOLDER, f"{filename}_RDVI.tif"), rdvi, ds)

    Clean up memory
    ds = None
    band_data = None

print("\nAll indices processed.")
```

APPENDIX C

Conversion from linear to dB for SAR

```
1
2   input_folder = r"C:\Users\msi fast\Desktop\Thesis work\Data\SAR_2019\2019_VV_raw"
3   output_folder = r"C:\Users\msi fast\Desktop\Thesis work\Data\SAR_2019\2019_VV_dB"
4
5   # Create output folder if it doesn't exist
6   os.makedirs(output_folder, exist_ok=True)
7
8
9   # PROCESS ALL TIFF FILES
10
11  for file in os.listdir(input_folder):
12      if file.lower().endswith(".tif") or file.lower().endswith(".tiff"):
13
14          input_path = os.path.join(input_folder, file)
15          output_path = os.path.join(output_folder, file)
16
17          print("Processing:", file)
18
19          # Open image
20          ds = gdal.Open(input_path)
21          band = ds.GetRasterBand(1)
22          arr = band.ReadAsArray().astype(np.float32)
23
24          # Convert to dB
25          arr_db = 10 * np.log10(arr)
26          arr_db[arr == 0] = -9999  # handle log(0)
27
28          # Write output
29          driver = gdal.GetDriverByName("GTiff")
30          out_ds = driver.Create(
31              output_path,
32              ds.RasterXSize,
33              ds.RasterYSize,
34              1,
35              gdal.GDT_Float32
36          )
37
38          out_ds.SetGeoTransform(ds.GetGeoTransform())
39          out_ds.SetProjection(ds.GetProjection())
40          out_band = out_ds.GetRasterBand(1)
41          out_band.WriteArray(arr_db)
42          out_band.SetNoDataValue(-9999)
43
44          out_ds.FlushCache()
45          out_ds = None
46
47          print("Saved:", output_path)
48
49  print("All done!")
50
51
```

APPENDIX D

R Script for training model, predicting in model domain and producing AOA, LPD and DI

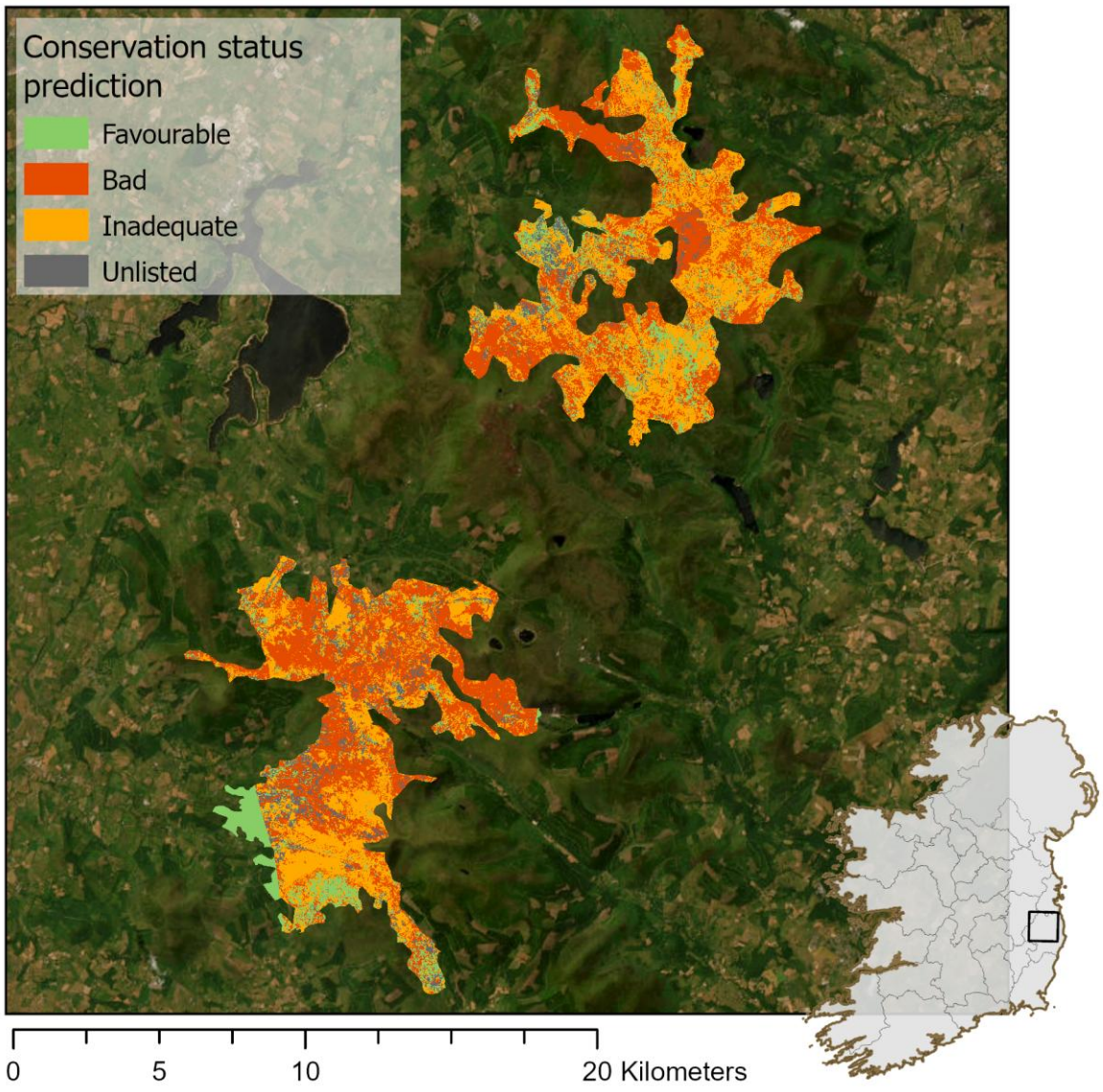
```
#------------------------------------------------------------
# 1. LOAD AND CLEAN TRAINING POINTS
#------------------------------------------------------------
SUAS_PB_points <- st_read("SUAS_pure_points.gpkg", quiet = TRUE)

habitat_field <- "Conservati"

SUAS_PB_points$habitat <- SUAS_PB_points[[habitat_field]]

SUAS_PB_points <- SUAS_PB_points |>
  filter(!is.na(habitat), habitat != "")

SUAS_PB_points$habitat <- droplevels(factor(SUAS_PB_points$habitat))

#------------------------------------------------------------
# 2. LOAD PREDICTORS (RASTER STACK)
#------------------------------------------------------------

raster_files <- list.files(
  "raster_stack_minimised_masked",
  pattern = "\\.tif$",
  full.names = TRUE
)

predictors <- terra::rast(raster_files)
predictor_names <- names(predictors)

SUAS_PB_points <- st_transform(SUAS_PB_points, crs(predictors))

#------------------------------------------------------------
# 3. EXTRACT RASTER VALUES TO POINTS
#------------------------------------------------------------

training_data_extracted <- terra::extract(predictors, SUAS_PB_points)
training_data_extracted$habitat <- SUAS_PB_points$habitat

training_data <- training_data_extracted |> select(-ID)

#------------------------------------------------------------
# 4. REMOVE POINTS WITH TOO MANY NA PREDICTORS
#------------------------------------------------------------

threshold <- 0.8 * length(predictor_names)

keep_rows <- rowSums(!is.na(training_data[predictor_names])) > threshold

training_data <- training_data[keep_rows, ]
SUAS_PB_points_filt <- SUAS_PB_points[keep_rows, ]

training_data$habitat <- droplevels(training_data$habitat)
SUAS_PB_points_filt$habitat <- training_data$habitat

#------------------------------------------------------------
# 5. SET UP SPATIAL CROSS-VALIDATION (kNNDM)
#------------------------------------------------------------
```

```
56
57  knndm_folds <- CAST::knndm(
58    tpoints = SUAS_PB_points_filt,
59    modeldomain = predictors[[1]],
60    k = 5
61  )
62
63  tr_control <- caret::trainControl(
64    method = "cv",
65    index = knndm_folds$indx_train,
66    indexOut = knndm_folds$indx_test,
67    savePredictions = TRUE
68  )
69
70  tunegrid <- expand.grid(
71    mtry = c(5, 10, 15),
72    splitrule = "gini",
73    min.node.size = c(1, 3, 5)
74  )
75
76  #------------------------------------------------------------
77  # 6. RUN FEATURE FORWARD SELECTION (FFS)
78  #------------------------------------------------------------
79
80  n_cores <- parallel::detectCores() - 1
81  cl <- makeCluster(n_cores)
82  registerDoParallel(cl)
83
84  rfmodel_ffs <- CAST::ffs(
85    predictors = training_data[, predictor_names],
86    response   = training_data$habitat,
87    method     = "ranger",
88    trControl  = tr_control,
89    tuneGrid   = tunegrid,
90    num.trees  = 500,
91    importance = "impurity"
92  )
93
94  stopCluster(cl)
95  registerDoSEQ()
96
97  #------------------------------------------------------------
98  # SET UP MODEL DOMAIN
99  #------------------------------------------------------------
100
101 modeldomain <- st_read("Corine_BlanketBog.gpkg", quiet = TRUE)
102 modeldomain <- st_transform(modeldomain, st_crs(SUAS_PB_points))
103 modeldomain <- st_transform(modeldomain, crs(predictors))
104
105 modeldomain_vect <- vect(modeldomain)
106
107 #------------------------------------------------------------
108 # PREDICTION
109 #------------------------------------------------------------
110
111 predictors_aoi <- crop(predictors, modeldomain_vect)
112 predictors_aoi <- mask(predictors_aoi, modeldomain_vect)
113
114 prediction_ffs <- terra::predict(
115   predictors_aoi,
116   rfmodel_ffs,
```

```
  type = "raw"
)

prediction_ffs_masked <- mask(prediction_ffs, modeldomain_vect)

#------------------------------------------------------------
# VISUALISATION
#------------------------------------------------------------

fossitt_colors <- c(
  "PB2" = "#238823",
  "PB4" = "#B284BE",
  "PB5" = "#987654"
)

annex_colors <- c(
  "7130" = "#238823",
  "7000" = "#987654",
)

condition_colors <- c(
  "Favourable" = "#238823",
  "Unfavourable - Inadequate" = "#FFBF00",
  "Unfavourable - Bad" = "#D2222D",
  "unlisted" = "#D3D3D3"
)

tm_shape(prediction_ffs_masked) +
  tm_raster(
    title = "Predicted Condition Classes",
    palette = condition_colors,
    legend.show = TRUE
  ) +
  tm_shape(modeldomain_vect) +
  tm_borders(col = "black", lwd = 2) +
  tm_layout(
    legend.position = c("right", "bottom"),
    frame = FALSE,
    outer.margins = 0
  )

terra::writeRaster(
  prediction_ffs_masked,
  filename = "Predicted_Condition_Full.tif",
  overwrite = TRUE
)

plot(
  prediction_ffs_masked,
  col = annex_colors,
  legend = TRUE,
  main = "Prediction (FFS)"
)
levels(prediction_ffs_masked)
#------------------------------------------------------------
# 7. RESULTS
#------------------------------------------------------------

print("Selected Variables:")
print(rfmodel_ffs$selectedvars)

```

```
plot(rfmodel_ffs)

preds <- rfmodel_ffs$pred
levels_all <- levels(training_data$habitat)

best_preds <- preds |>
  dplyr::filter
(
    mtry == rfmodel_ffs$bestTune$mtry,
    min.node.size == rfmodel_ffs$bestTune$min.node.size
  )

best_preds$pred <- factor(best_preds$pred, levels = levels_all)
best_preds$obs  <- factor(best_preds$obs,  levels = levels_all)

cm <- caret::confusionMatrix(best_preds$pred, best_preds$obs)
cm
cm$byClass
rfmodel_ffs$bestTune

#------------------------------------------------------------
# 8. AOA / DI / LPD CALCULATION (FFS-CONSISTENT)
#------------------------------------------------------------
# 1. FFS-selected variables
sel_vars <- rfmodel_ffs$selectedvars

# 2. Identify non-zero-variance predictors in training data
vars_var <- apply(
  training_data[, sel_vars, drop = FALSE],
  2,
  var,
  na.rm = TRUE
)
sel_vars_aoa <- names(vars_var[vars_var > 0])

# 3. Subset training data
train_aoa <- training_data[, sel_vars_aoa, drop = FALSE]
train_aoa <- train_aoa[complete.cases(train_aoa), ]

# 4. Subset raster predictors
predictors_sel_aoi <- predictors_aoi[[sel_vars_aoa]]

# 5. AOA
AOA <- CAST::aoa(
  newdata   = predictors_sel_aoi,
  train     = train_aoa,
  variables = sel_vars_aoa,
  LPD       = TRUE,
  useWeight = FALSE
)
#------------------------------------------------
# 9. Export files
#----------------------------

writeRaster(AOA$AOA, "Model_AOA_Condition_full.tif", overwrite=TRUE)
writeRaster(AOA$DI, "Model_DI_Condition_full.tif", overwrite=TRUE)
writeRaster(AOA$LPD, "Model_LPD_Condition_full.tif", overwrite=TRUE)
```

APPENDIX E


Conservation status prediction
Favourable
Bad
Inadequate
Unlisted
0
5
10
20 Kilometers

Masters Program in **Geospatial Technologies**

Supported by:
Education and Culture
ERASMUS MUNDUS