MULTI-TEMPORAL, MULTI-SENSOR LAND USE / LAND COVER MAPPING: GOOGLE EARTH ENGINE AND RANDOM FOREST FOR THE CLASSIFICATION OF THE SCOTTISH FLOW COUNTRY

Name: NEIL ALEXANDER SUTHERLAND
Student Number: 20315184

Submitted in partial fulfilment of the requirements for the Degree of MRes Geospatial Data Science, Faculty of Engineering, University of Nottingham.

Nottingham Geospatial Institute, School of Engineering Surveying and Space Geodesy, 30 Triumph Road, Lenton, Nottingham, NG7 2TU, United Kingdom.

Date: Friday 20th August 2021

Declaration: "I hereby certify that this work is my own, except where otherwise acknowledged, and that it has not been submitted previously for a degree at this, or any other university."

NEIL ALEXANDER SUTHERLAND

Date: Friday 20th August 2021

Acknowledgements: I would like to thank to EPSRC and the Centre for Doctoral Training in Geospatial Systems for the opportunity to purse postgraduate research under a fully funded studentship.

I would also like to thank my supervisory team of Abbott Professor David Large and Professor Stuart Marsh for their counsel and support during this challenging project. They have provided invaluable advice and been calming influences during this challenging MRes.

MULTI-TEMPORAL, MULTI-SENSOR LAND USE/LAND COVER MAPPING: GOOGLE EARTH ENGINE AND RANDOM FOREST FOR THE CLASSIFICATION OF THE SCOTTISH FLOW COUNTRY

Abstract: Long-term monitoring of Land Use/Land Cover (LULC) dynamics is fundamental for implementing effective policy and mitigating the effects of climate change. In the UK, the Scottish Flow Country represents an area of ~4000km^2 spanning Caithness and Sutherland, encompassing 25% of global blanket bogs. There is a need to understand these peatland ecosystems in a broader context, appreciating their importance within evolving landscapes. Frequent advances in remote sensing (RS) have provided a means for large-scale LULC mapping to be executed with increasing temporal and spatial resolutions. In addition, cloud-computing services such as Google Earth Engine (GEE) have enabled the processing and analysis of geospatial data, allowing various stakeholders to address challenges with the assistance of "Geo Big Data". This study looks to assess how the LULC mapping can take advantage of geospatial data, cloud-computing and machine learning for the monitoring of peatland ecosystems within a broader economic and environmental policy-driven context. The following objectives were defined: (1) determine the optimal combination of optical, radar and topographic data for LULC mapping of the Scottish Land Use Strategy; (2) assess their application in GEE; and (3) evaluate Random Forest for classification of LULC classes. Results suggest a combination of optical, radar and topographic features is necessary for comprehensive LULC mapping (LUS_{TOR} OA=0.823 and KA=0.792), particularly when delineating ecologically, hydrologically and geomorphologically heterogenous landscapes. Finally, RF performance was evaluated, future improvements were outlined and the effectiveness of LULC mapping for policy assessments is discussed.

Keywords: Google Earth Engine (GEE), Random Forest (RF), Sentinel, Peatland, Land Use Land Cover Mapping (LULC), Pixel-Based Image Classification

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

- API – Application Programming Interface
- ASTER – Advanced Spaceborne Thermal Emission and Reflection Radiometer
- ASTER GDEM – ASTER Global Digital Elevation Model
- AWS – Amazon Web Services
- C – Carbon
- DEM – Digital Elevation Model
- DT – Decision Trees
- EO – Earth Observation
- ESA – European Space Agency
- EVI – Enhanced Vegetation Index
- GEE – Google Earth Engine
- GHGs – Greenhouse Gases
- GRD – Ground Range Detected
- Gt – Gigatonnes
- HPC – High Performance Computing
- IDE - Integrated Development Environment
- JHI – The James Hutton Institute
- KA – Kappa Accuracy
- LC – Land Cover
- LiDAR – Light Detection and Ranging
- LU – Land Use
- LULC – Land Use / Land Cover
- LUS – The Scottish Land Use Strategy
- LUS_{TO} – Land US Strategy Model (Topography and Optical)
- LUS_{TOR} – Land US Strategy Model (Topography, Optical and Radar)
- LUS_{TR} – Land US Strategy Model (Topography and Radar)
- ML – Machine Learning
- MSI – Multi-Spectral Imagery
- Mtry – Determination of Nodes
- NASA – The National Aeronautics and Space Administration
- NASADEM - The National Aeronautics and Space Administration Digital Elevation Model
- NDBI – Normalized Difference Built-Up Index
- NDVI - Normalized Differenced Vegetation Index
- NDWI – Normalized Difference Water Index
- NEX - NASA's Earth Exchange
- NIR – Near Infrared
- NormPol – Normalized Polarization
- Ntree – Number of Decision Trees
- OA – Overall Accuracy
- OSAVI – Optimized Soil-Adjusted Vegetation Index
- PA – Producer's Accuracy
- PC(A) – Principal Component (Analysis)
- PolDiff – Polarized Difference
- RF – Random Forest
- rfOOBe – Rand Forest Out-of-Bag Error
- RGB – Red, Green, Blue (in reference to visible optical bands)
- RS – Remote Sensing
- S1 – Sentinel-1 C-Band Synthetic Aperture Radar
- S2 – Sentinel-2 Multi-Spectral Imagery
- SAR – Synthetic Aperture Radar
- SAVI – Soil-Adjusted Vegetation Index
- SI – Spectral Indices
- SNIC – Simple Non-Iterative Clustering
- SRTM – Shuttle Radar Topography Mission
- SVM – Support-Vector Machines
- SWIR – Shortwave Infrared
- TPI – Topographic Position Index
- TWI – Topographic Wetness Index
- UA – User's Accuracy
- UKCEH – UK Centre for Ecology and Hydrology

1 – INTRODUCTION

The effective monitoring and management of peatland ecosystems is fundamental in the fight against climate change. An ever-expanding community is exploring global peatlands for their invaluable ecosystem services, notably their remarkable biodiversity, irreplaceable wildlife habitats, significance as carbon stocks and socio-economic value (Page and Baird, 2016). Contemporary research appreciates that comprehensive long-term monitoring projects are critical in the conservation and restoration of these ecosystems in the face of mounting pressures from environmental and anthropogenic changes (Hird *et al.*, 2017; DeLancey *et al.*, 2019). To assess these dynamic, often-transitional changes, research needs to simulate and project change in the context of the complex landscapes they reside within (Mitsch and Gossilink, 2000; Alonso *et al.*, 2016; Hird *et al.*, 2017). Peatlands should not be viewed as isolated features but spaces where complex 'human-natural systems' intertwine, with environmental, economic and social drivers balancing the costs and benefits of interventions. In the face of mounting demands, these processes need to be better understood so that long-term management can sustain the benefits of peatland ecosystems while maximising their responsible use for future generations (Alonso *et al.*, 2016).

1.1 – Peatlands: Definition, Formation and Carbon Storage

Peat can be broadly defined as a soil type consisting of sedentarily accumulated, partially decomposed organic material, derived from vegetation that has been preserved through anoxic waterlogged conditions induced by precipitation. Peat is formed in layers and builds up with increased accumulation and decomposition of surface plants and organic material (Joosten and Clarke, 2002; Page and Baird, 2016). Peat, which locks carbon sequestered by the vegetation its partially decomposes, builds up to result in incremental increases in soil formation and consolidation (Gorham, 1991). Peatland represents an area with or without surface vegetation within which naturally accumulated layers of peat have formed over sizeable expanses (Minasny *et al.*, 2019). Peatland can largely be found in the northern hemisphere, occurring in the boreal and temperate climates of North America, Europe and Russia. Northern peatlands have formed over millennia, varying in the rates of accumulation (averaging 18.6 g C m−2 yr) forming part of a global wetland carbon sink that equates to 90% of the global carbon pool (Yu *et al.*, 2010; DeLancey *et al.*, 2019).

Covering 3% of the Earth's surface, peat-forming systems play a pivotal role in mitigating the effects of greenhouse gases (GHGs) and represent a significant source of stored carbon (C) (Yu *et al.*, 2010). This can be equated to 113-612 Gigatonnes (Gt) of C, containing more C than tropical rainforests biomass (360 Gt)(Pan *et al.*, 2011) and representing half the C in the atmosphere (750 Gt)(Grace, 2004) (Köchy, Hiederer and Freibauer, 2015; Jackson *et al.*, 2017). Understanding the delicate hydrological dynamics of peat accumulation and degradation required for such ecosystems to continue to store carbon is of the upmost importance in determining their future effectiveness as mitigators of climate change (Gorham, 1991). The benefits of healthy peatland ecosystems are wide-ranging, including: improved water regulation; biodiversity protection; flood-risk mitigation; food production; fuel and socio-cultural experiences (Page and Baird, 2016; Sloan *et al.*, 2018; Lees *et al.*, 2021). The quantification of these benefits is vital for the appropriate management and targeted intervention within peatland landscapes, from which assessments of appropriate use and interaction can be determined.

Understanding the numerous actors involved in peatland monitoring and management is fundamental to assess appropriate intervention for their restoration and long-term conservation. The complex coupling of 'human-natural systems' has resulted in the marked degradation of global peatland areas, with agriculture, horticulture, forestry and energy sourcing significantly damaging peatland's long-term resilience to climate changes (Alonso *et al.*, 2016). However, research into the prolonged restoration of peatland has struggled to distinguish between climatic and anthropogenic factors responsible for their formation, highlighting the growing need to separate these two systems to best understand peatland dynamics (Tipping, 2008; Gallego-Sala *et al.*, 2016; Page and Baird, 2016; Pimm *et al.*, 2019). Lees et al. (2021) states the need for increasing research into the resilience of peatland by viewing climatic and anthropogenic stresses independently, assessing their ability to respond to climate change and our ability to accomplish sustainable use (Chambers, Allen and Cushman, 2019; Lees *et al.*, 2021). Future research trends are looking to document (1) how peatland will respond to rapid, short-term disturbances and to gradual, long-term changes in climate; (2) how scientists and policy makers can incorporate climatic and anthropogenic changes effectively into peatland monitoring ; and (3) how these changes can be modelled and mapped to aid conservation and restoration practices going forward (Page and Baird, 2016).

1.2 – Scottish Peatland: The Flow Country and Scotland's Land Use Strategy

In the UK, the Scottish Flow Country represents an area of ~4000km^2 spanning Caithness and Sutherland, containing 25% of the worlds blanket bog, peat-forming mires on undulating terrains with frequent precipitation, high atmospheric humidity, cool mean temperatures and minimal evapotranspiration (Lindsay *et al.*, 1988). Under such ombrotrophic conditions, Sphagnum mosses and acidophilous vegetation grow in near-natural conditions, providing the organic material from which peat can continue to form(Lees *et al.*, 2019). Within the Flow Country, this formation can ranges from 0.5m upwards of 10m in depth (Scottish Natural Heritage, 2015b). Subject to harmful commercial forestry policies from the Scottish government throughout from the 1950's-1980's, historical interventions within the Flow Country have altered the health of Scottish peatlands, including: drainage, peat cutting, afforestation, agriculture, sport estate management and burning (Lindsay *et al.*, 1988; Warren, 2000; Wallage, Holden and McDonald, 2006; Hambley *et al.*, 2018; Lees *et al.*, 2019). These interventions have dramatically scarred the Flow Country, with 80% of Scottish peatland now designated as degraded and in need of restoration (Scottish Government, 2020).

With greater appreciation for the significant role a green recovery and restoring biodiversity plays in the fight against climate change, policies supporting the consolidation and restoration of Scottish peatland and its accompanying biodiversity have materialised. The Scottish Government's ambitious *Update to the Climate Change Plan* looks to: reduce emissions 75% by 2030 resulting in net-zero by 2045; expand funding for Scottish Forestry to afforest 18,000 hectares of conifers annually by 2024; increase investment for agricultural technology to maximise efficient food production and reduce emissions; and restore 40% (over 250,000 hectares) of Scotland's degraded peatland by 2032 (Scottish Natural Heritage, 2015b; Scottish Government, 2020). In addition, the Scotland Land Use Strategy, in its third iteration entitled *Scotland's Third Land Use Strategy 2021-2026: Getting the Best from Our Land*, aims to takes a holistic approach to land use management by recognising the intrinsic links between Scotland's various industries and its natural land (Scottish Government, 2021). This novel approach relies on landowners, land users and government bodies working in collaboration to come to shape future transformations of peatlands and their surrounding landscapes (Fig. 1) (ClimateXChange, 2021; Scottish Government, 2021). This all revolves around coordinated research being undertaken to determine where strategic interventions are required and what necessary trade-offs are expected. For this to be undertaken, land use mapping must be employed to determine suitable areas of

change to meet the pioneering climate and socio-economic goals outlined by the Scottish Government.

Figure 1: 'Scotland Now' and 'Scotland Future' – Scotland's Third Land Use Strategy 2021-2026: Getting the Best from our Land

1.3 – Land Use Mapping: Challenges and Opportunities

Determining where strategic restoration in peatland can be undertaken relies on the accurate mapping and monitoring of land-cover (LC) and subsequent land use (LU), recording distribution, quality and quantity of peatland ecosystems (Adam, Mutanga and Rugege, 2010).

Land-cover mapping is vital in examining changes to surface vegetation, soil decomposition, water dynamics and geomorphology. In contrast, land use mapping looks to assess how human-derived alterations and modifications have impacted landscapes and resulted in discernible changes (Farda, 2017; Kolli *et al.*, 2020). From this, inventories and models can determine courses of action and resource allocation. It is also vital this information respects the spatial variability and heterogeneity of land use / land-cover (LULC) classes that are visible throughout the landscape (Mahdianpari, Brisco, *et al.*, 2020). LULC maps are vital in determining not only natural changes, but document anthropogenic disturbances in peatland landscapes, while providing a means of examining these at regional-to-global scales (Liu *et al.*, 2018; Chaves, Picoli and Sanches, 2020). LULC mapping utilising temporally- and spatially-rich information is vital to quantify land capability, land suitability and the optimisation of future policies for peatland conservation (Dóka, Kiss and Bárány-Kevei, 2019; De Feudis *et al.*, 2021).

Long-term monitoring is fundamental in tracking changes to peatland health, with land-cover and land use dynamics vital indicators in the condition of Flow Country blanket bogs. The need for dynamic seasonal and long-term processes to be understood requires the collection of decades of data that can examine the variability in both hydrological and ecological responses of peatland (Hancock *et al.*, 2018). In addition, regional and national spatial analyses of peat is vital in quantifying carbon storage and valuing socio-economic strategies (Hermosilla *et al.*, 2018; DeLancey *et al.*, 2019; Mahdianpari, Salehi, *et al.*, 2020). However, traditional mapping methods have largely relied on in-situ measurements and field surveys to document peatland conditions, resulting in significant challenges regarding scale and coverage. Limitations in cost, time and access have acted as barriers to mapping peatland ecosystems inventories, with traditional in-situ practices insufficient in providing necessary temporal coverage (Lees et al., 2019; Mahdianpari et al., 2019). In addition, such methods have struggled to contextualise peatlands within a spatially vast and varied setting, underappreciating peatland boundaries and proximities with its connected landscape. As stated by DeLancey et al. (2019), "Their large geographic extent, natural heterogeneity, and cultural and socio-economic value makes accurate identification and mapping of peatlands both critical and challenging". This necessitates the development of contemporary methods to advance the study and documentation of peatland ecosystems.

To combat these challenges, research has looked to exploit remote sensing (RS) techniques to provide improved spatial and temporal coverage of peatlands and their associated environments (Gallant, 2015; Mahdavi *et al.*, 2018). Through the exploitation of Earth Observation (EO),

largely optical, radar and LiDAR datasets, peatland ecosystems can be mapped effectively and efficiently. In comparison to the limitations of 'traditional' soil and LULC mapping techniques, RS is viewed as a more practical, economical and replicable method for documenting various ecosystems (Adam, Mutanga and Rugege, 2010; Mahdavi *et al.*, 2018). RS allows research to integrate vast amounts of multi-temporal satellite imagery into assessments of land use, allowing larger areas to be studied concurrently and for results to assess change over defined time periods (Amani *et al.*, 2019). In the context of LULC mapping, RS provides a means for changes to be documented within environmental, social and economic contexts, determining trends over decades and observing where policy has been pivotal in changing the landscape (Kolli *et al.*, 2020; Lees *et al.*, 2021). Such an ability is useful for understanding the changes within the Scottish Flow Country, as immediate policies look to drastically change the land composition of the highlands and implement a balanced Land Use Strategy to combat climate change.

1.4 – The advent of 'Geo Big Data', Cloud Computing and Machine Learning

LULC mapping has benefitted significantly from the increased availability of open-source EO data, allowing for peatland ecosystems to be monitored and managed with almost real-time, spatially comprehensive coverage. The term "Big Data", emerging from the tech-boom of the mid-1990's, describes data of significant volume, variety and velocity, while 'Geo Big Data' expands this principle to appreciate the scale of geospatial formats, characteristics and structures (Laney, 2001; Tamiminia *et al.*, 2020). 'Geo Big Data' includes data collected from terrestrial surveying, earth observation, and mobile-mapping procedures, dynamic in scale and contrasting in detail. The European Space Agency's (ESA) Copernicus series, NASA's Landsat series and other international government bodies have made vast volumes of geospatial data available through associated portals, allowing data to be downloaded and processed by a remarkably broader community of expert and non-expert users. However, this data cannot be viewed in isolation, but part of a growing field of computer science where technology has evolved to process and analyse this information seamlessly. Developments in cluster-based high performance computing (HPC), data storage and cloud computing have come to the forefront of RS and EO research (Ma *et al.*, 2015; Leinenkugel *et al.*, 2019). Advances in computation for peatland monitoring have resulted in reduced dependency on user expertise, less pressure on local systems, fewer storage limitations and increasing access to machine learning tools (Mahdianpari, Brisco, *et al.*, 2020).

The advances of cloud-computing services such as Google Earth Engine (GEE), NASA's Earth Exchange (NEX) and Amazon Web Service (AWS) has significantly increased the ability for geospatial research to be implemented for specific applications at regional and global scales (Hird *et al.*, 2017; Pimple *et al.*, 2018). With reference to GEE, computational barriers to the processing and automation of global remote sensing data are dramatically reduced, allowing various stakeholders to address challenges and determine solutions through 'Geo Big Data'. GEE, established in 2010, allows users to: access petabytes of pre-processed and georeferenced data; browse a public catalogue of geospatial and auxiliary datasets; extract and process functions through GEE's parallel processing; write code in an integrated JavaScript / Python IDE web-app; and apply algorithms at regional-to-global scales (Gorelick *et al.*, 2017; Kumar and Mutanga, 2018; Alonso, Muñoz-Carpena and Kaplan, 2020; Tamiminia *et al.*, 2020; Wu, 2020). Ultimately, GEE has enabled expert and non-expert users to apply machine learning on vast quantities of geospatial data to determine features, assess change and implement findings (DeLancey *et al.*, 2019).

Accompanying the vast capabilities of GEE are the numerous integrated machine learning algorithms that enable 'Geo Big Data' data to be processed, analysed and applied for specific purposes (Carrasco *et al.*, 2019; Tamiminia *et al.*, 2020). Still in its infancy for peatland and LULC mapping, GEE provides opportunities to combine various datasets to classify LULC, exploiting high-temporal and spatial resolutions to determine land classes and quantify changes within peatland ecosystems. Machine Learning (ML) in GEE supports numerous parametric and non-parametric supervised classification algorithms (i.e., maximum likelihood, k-Nearest Neighbors (kNN), decision trees (DT), support vector machines (SVM), and random forest (RF), all of which can determine LULC classes based on their detectable spectral and hydrological characteristics. However, the effectiveness of ML classification relies on a delicate balance between the appropriate selection of input features, the sourcing of representative training data and the configuration of a suitable supervised learning algorithm (Thanh Noi and Kappas, 2017; Yang *et al.*, 2021). With an ever-growing need for peatland policies to address current and future climate conditions, research must accurately map landscapes with a considerable degree of accuracy, from which policies can be implemented with confidence. Through detailed LULC mapping, the dynamics of land use can be traced to assess threats, priorities and opportunities for future climate change mitigation.

1.5 – Aims and Objectives

The Third Scottish Land Use Strategy (LUS) has outlined ambitious plans for the re-appropriation of Scottish landscapes, changing vast amounts of land to deliver economic and environmental targets. These changes need to be documented and quantified to understand the effectiveness of land use changes and quantify their benefits. However, this can only be done through the accurate and precise mapping of the current state of Scottish environment. LUS outlines seven classes within which land use has been grouped and policies have been defined: 'Settlements', 'Enclosed Farmland, 'Semi-Natural Land', 'Rivers and Waterbodies', 'Coastal', 'Islands' and 'Marine' (Scottish Government, 2021). By documenting the configuration and changes of these classes, broad assessments of policy attainment could be undertaken and targets could be evaluated. In the context of peatland, the LUS (in accordance with the National Peatland Action Plan and IUCN Peatland Code) determines; how peatlands will be restored; where investment is needed; and what environmental (carbon storage) and socio-economic value can be derived (Scottish Natural Heritage, 2015a; IUCN UK Peatland Programme, 2017). LULC policy mapping can determine which policies have ensured change, quantify results and act as a precursor for policy analysis and future amendments.

Therefore, this study will assess how LULC mapping can take advantage of geospatial data, cloud-computing and machine learning for the monitoring of peatland ecosystems within a broader economic and environmental policy-driven context. The specific objectives of this study are to: (1) determine the optimal combination of optical, radar and topographic data for LULC mapping of the Scottish Land Use Strategy; (2) assess their application in GEE; and (3) evaluate Random Forest for classification of LULC. The resulting LULC maps will determine how effectively Scottish policymakers can implement LULC mapping and how frameworks can exploit these maps for future monitoring and management of peatland landscapes.

2 – LITERATURE REVIEW

An established consensus that RS provides an effective means of monitoring peatland ecosystems has been well-documented within wetland literature (Khatami, Mountrakis and Stehman, 2016; Tamiminia *et al.*, 2020), yet cloud-computing remains an emerging field for researchers, organisations and governments managing these ecosystems in broader contexts (DeLancey *et al.*, 2019). The determination of methodologies resulting in applicable LULC maps has looked to manipulate vast spatial and temporal resolutions to determine health and monitor progress through LULC change (Mahdavi *et al.*, 2018; Minasny *et al.*, 2019; Lees *et al.*, 2021). For LULC mapping, an appreciation for the spectral and spatial characteristics of differing classes is needed, understanding methods to discriminate these classes based on ecology, hydrology and geomorphology (Adam, Mutanga and Rugege, 2010). RS data, obtained through satellite imagery, such as the Sentinel and Landsat series, is vital for the representation of class vegetation, soil composition and water features.

2.1 – Optical Data and Spectral Indices

Research assessing peat ecosystems and peatland LULC mapping have exploited high-resolution optical data to determine effective bands that identify unique characteristics of both peat soils and its surface vegetation (Berhane *et al.*, 2018). This is done by determining the spectral reflectance of certain vegetation and soil types, delineating these from one another through visible, near- and mid-infrared wavelengths (Krankina *et al.*, 2008). In LULC mapping; urban, agricultural, forested and peatland classes will display markedly different characteristics, which can be detected and used to determine classes in optical imagery. Research into peatland mapping has exploit the unique spectral signature of peat and its associated vegetation, using red-edge, near- (NIR) and shortwave-infrared (SWIR) to determine shallow-watered wetlands, wetland soil moisture and sphagnum vegetation absorption (Harris and Bryant, 2009; Dvorett, Davis and Papeş, 2016; Mahdianpari *et al.*, 2019). In addition to peatland mapping, LU mapping has exploited optical data for the identification of LU features, determining areas and boundaries of urban settlements by targeting Red, Green and Blue bands (RGB) (Zhang *et al.*, 2021). To further delineate these classes, the combination of these band into targeted indices are instrumental in detecting vegetation greenness, moisture contents and soil exposure for LU mapping.

Spectral indices (SI's), calculated through the manipulation of optical bands, have proven critical for LULC mapping, targeting classes intrinsic sensitivities and maximising their detection through manipulating wavelengths. Notable SI's used in LULC mapping include: the Normalized Difference Vegetation Index (NDVI); Normalized Difference Water Index (NDWI); Enhanced Vegetation Index (EVI); and Soil Adjusted Vegetation Index (SAVI), modified and evolved for additional purposes. These SI's allow for rapid and long-term changes to be assessed comprehensively, with research demonstrating their effective use in LC change in heterogenous landscapes (Huang *et al.*, 2017; Zhang *et al.*, 2020), tidal inundations of mangrove wetlands (Chen *et al.*, 2017), phenological variation estimations (Zhang *et al.*, 2019) and monitoring deforested regions (Schultz *et al.*, 2016). Soil-specific SI's, which exploit sensitivities of soil types, have also proven critical in class delineation. With reference to peatland ecosystems, this derived spectral information can be used determine carbon fluxes (Lees *et al.*, 2020), Sphagnum moisture (Letendre, Poulin and Rochefort, 2008; Harris and Bryant, 2009) and peatland scale (Alonso *et al.*, 2016). The use of optical data for classification necessitates the balanced inclusion of uncorrelated variables, reflective of the sensitivities present within peatland landscapes.

2.2 – Synthetic Aperture Radar (SAR)

While optical data provides numerous benefits for LULC mapping, its inability to penetrate dense surface vegetation, cloud cover, atmospheric haze or night-time conditions means additional data is require to comprehensively map peatland landscapes with high-temporal and spatial resolution (Gallant, 2015; Dvorett, Davis and Papeş, 2016). This requires supplementary data that can complement existing datasets and provide complete coverage of defined research areas. Synthetic Aperture Radar (SAR) similarly plays an important role in peatland mapping, allowing for the determination of surface moisture and soil conditions. The use of SAR as a means of evaluating peatland processes allows not only for the assessment of hydrology, but the resilience of peatland ecosystems to changing land uses (Lees et al., 2021). Sentinel 1, equipped with dual-polarization C-band SAR sensors at 10-m resolution, has been used for "marine monitoring, shoreline detection, and mapping land cover, climate change, rice fields, and disasters such as flood monitoring" (Tamiminia *et al.*, 2020; European Space Agency, 2021a).

The application of SAR for peatland mapping has exploited these benefits to provide a data source that can combat the near-constant cloud cover experienced in northern peatland landscapes and to interpret the varying hydrological dynamics present with peatland ecosystems

(Mahdianpari *et al.*, 2019). Several studies have highlighted the significant improvements of SAR as a replacement/compliment to optical satellite imagery (Li and Chen, 2005; Mahdavi *et al.*, 2018). SAR provides opportunities to assess the hydrological and elastic processes experienced within peat soils, including appreciation of surface roughness, moisture, inundation and water-table depths (Adeli *et al.*, 2020; Lees *et al.*, 2021). When utilised for LULC mapping , C-Band SAR has shown capabilities to supplement optical data to provide additional features for classification at regional scales when incorporated into a comprehensive model that reflects all dynamic characteristics of peatland landscapes (Moreira *et al.*, 2013; Mahdavi *et al.*, 2018; Karlson *et al.*, 2019; Minasny *et al.*, 2019).

2.3 – Topography and Geomorphology

As evident from research examining peatland mapping, it is clear that SAR and optical data should not be used in isolation, but accompanied by topographic features which take into consideration peatland geomorphology (Karlson *et al.*, 2019). Global wetlands reside within markedly different topographies, which has in turn shaped the formation and function of peat-forming ecosystems (Mahdavi *et al.*, 2018; Belcore, Piras and Wozniak, 2020). For Scottish peatland, characterised by undulating blanket bogs, the subtle changes in elevation, slope and flow are important predictors in the separation of peatland classes from other LULC classes (Anderson *et al.*, 2010). (In)SAR-derived DEMs such as ASTER-GDEM and Shuttle Radar Topography Mission (SRTM), from which topographic metrics can be extracted, have proved vital for the delineation of waterbodies and vegetation. Several studies have determined topographic features to be the most important in wetland classification at regional scales, with the resolution of the generated Digital Elevation Model (DEM) determining its weight in subsequent classification models (Hird *et al.*, 2017; Minasny *et al.*, 2019; Liu *et al.*, 2020). In addition, research has determined increasing spatial resolution obtained through LiDAR-derived DEM's and topographic features provides markedly better classification results of wetland ecosystems situated in a dynamic anthropogenic and climatic areas (Mahdavi *et al.*, 2018).

Much like optical and radar data inputs, topographic datasets can generate indices that delineate peatlands based on proximity and hydrological impacts. The inclusion of topographic features and indices has shown to significantly improve classification of heterogenous wetland ecosystems, with the Topographic Wetness Index (TWI), Topographic Position Index (TPI) and Deviation from Mean Elevation (DEV) notable examples (De Reu *et al.*, 2013; Farda, 2017;

Berhane *et al.*, 2018; DeLancey *et al.*, 2019). When combined with both optical and radar datasets, a comprehensive assessment of peatland ecology, hydrology and geomorphology can be undertaken, with sufficient uncorrelated variables from which machine learning can be undertaken. LULC mapping relies on the accurate assessment of various anthropogenic and climatic conditions, which must appreciate subtle variations for effective classification and accurate results.

2.4 – Multi-Temporal, Multi-Sensor Analysis

For LULC mapping to determine strategies to mitigate climate change and increase socio-economic value, as outlined within the Land Use Strategy, an appreciation for multi-temporal analysis is fundamental. Research must incorporate multi-temporal, seasonally-adjusted input which can be maximise the differences between classes when conditions are optimal (Munyati, 2000; Mahdavi *et al.*, 2018; Karlson *et al.*, 2019). Assessment of multi-temporal or time-series imagery for classification has highlighted the accuracy obtained with composite stacks vs. single-date images, a process implemented by 96% of GEE research (Tamiminia *et al.*, 2020). Decisions related to image selection should reflect local vegetation, crop phenology and climate trends for classes to be accurately defined (Huang *et al.*, 2020). Temporal aggregation, using statistical methods to generate cohesive images, can combine vast quantities of geospatial data, reflecting seasonal and annual changes for LULC mapping whilst mitigating the various limitations of varying data sources (Carrasco *et al.*, 2019). Once generated in sufficient resolution, suitable for the scale of the targeted area, LULC mapping can accurately determine classes, quantify changes and assess future policies.

Research has dictated the need for peatland ecosystems and their associated landscapes must be assessed through multiple data types, using optical, radar and topographic products to comprehensively cover ecology, hydrology and geomorphology (Minasny *et al.*, 2019; Tamiminia *et al.*, 2020). This detail, coupled with model parameterization, allows for the creation of unique features suitable for supervised learning algorithms (DeLancey *et al.*, 2019). Multiple sensors provide a means of addressing limitations of singular inputs by assessing their collective implementation within climatically challenging and heterogenous regions (Mahdianpari *et al.*, 2019; Poggio, Lassauce and Gimona, 2019). However, the importance of each feature has been proven to vary depending on the area mapped, with variable importance and overall accuracy ranging for different models (Hird *et al.*, 2017; Karlson *et al.*, 2019). This necessitates models to

quantitively evaluate performance and weight features in resulting LULC outputs (Poggio, Lassauce and Gimona, 2019). Through this combination, classifiers can be fed with spatially-rich, temporally-ranging uncorrelated features that allow for LULC maps to be generated effectively (Dobrinić, Medak and Gašparović, 2020).

2.5 – Supervised Learning and Random Forests

For LULC maps to derive meaningful results from amalgamated data sources, the appropriate choice of machine learning (ML) classifier is of critical importance (Thanh Noi and Kappas, 2017). Research has shown the use of ML algorithms to be affective for capturing LULC changes and determining management policies, with different classifiers being compared for their individual strengths. Support Vector Machines (SVM), k-Nearest Neighbour (kNN) and Classification and Regression Tree-based (CART) algorithms have been frequently used for both peatland and LULC classification, demonstrating comparable overall accuracy with varying training datasets and parameter tuning (Thanh Noi and Kappas, 2017; Abdi, 2020; Li *et al.*, 2020). GEE provides a collection of integrated classification algorithms, scriptable within the IDE and customisable to classify specific climatic and anthropogenic changes for LULC mapping (Farda, 2017). Research denotes classification choice to be case-specific, determining the best suited algorithm based on the quality and quantity of input features, training data and resulting outputs (Maxwell, Warner and Fang, 2018; Shetty *et al.*, 2021).

Among the supervised learning algorithms used for LULC studies, Random Forest (RF), an ensemble CART-based parametric classifier, has proven the most popular (Breiman, 2001; Tamiminia *et al.*, 2020). RF has been recommended for: handling high-dimensionality data; accommodating heterogenous training data; modelling through user-defined parameters; and its integrated weighted variable importance (Maxwell, Warner and Fang, 2018; Berhane *et al.*, 2019; Diniz *et al.*, 2019). RF is an effective classifier for LULC mapping when dealing with a combination of input data sources, providing comparable accuracies to SVM, yet remains easier to implement, tune and provides resilience to noise and overtraining (Rodriguez-Galiano *et al.*, 2012; Mahdianpari *et al.*, 2017; Whyte, Ferentinos and Petropoulos, 2018). However, the use of RF must be accompanied with statistical error assessments to combat its inherent limitations, notably: dealing with outliers; determining appropriate training data size; distribution and class proportionality; and mitigating spatial autocorrelation of out-of-bag accuracy (Millard and Richardson, 2015; Belgiu and Drăgu, 2016; Berhane *et al.*, 2018). RF within GEE is well

established, and the exploration of integrating multiple sensory inputs for LULC mapping will contribute to the growing field determining approaches tuning algorithms for this unique landscapes.

3 – METHODOLOGY

Through the exploitation of 'Geo Big Data', cloud computing and machine learning, the following methodology has been designed to assess the optimal configuration of data sources for the monitoring and management of the Scottish Land Use Strategy. The objectives for this research, as stated in the Introduction, are repeated here: (1) determine the optimal combination of optical, radar and topographic data for LULC mapping of the Scottish Land Use Strategy; (2) assess their application in GEE; and (3) evaluate Random Forest for classification of LULC. The resulting LULC models will determine which combination of features can best map LULC dynamics and how effectively policies targeting peatland conservation and restoration can exploit RS as part of long-term monitoring and management practices.

3.1 – Study Area

The study area (Figure 2), located in the North-East of the Scottish Highlands, presents a perfect case study for the implementation of LULC mapping through GEE. All-bar-one Land Use Strategy class is represented in this region (excl. Islands), allowing for an effective assessment of policy mapping to be undertaken in the context of the LUS. The region follows the eastern coastline from Dunbeath up to Wick, heading inland to encapsulate a significant proportion of Scotland's Flow Country, an area of protected peatland amassing ~1,400km^2. The total area of this region is ~892km^2, centred at 58.4°N, -3.36°W. This area is characterised by a cool, oceanic climate, with annual rainfall of 971mm and mean temperatures in summer and winter averaging 10.7°C and 3.9°C respectively (Hancock *et al.*, 2018). The biodiversity of this region demonstrates some of the most natural blanket bogs and peat-forming systems on the globe, and provides numerous examples of healthy, deteriorated, and restored peatland. This will be critical for this classification and an important discussion of the classifier's success delineating these ecosystems. The region presents diverse vegetation and habitat types, including various grasslands, heathers, mosses and agricultural lands. In addition, several areas of deforestation are present around Camster, which has undergone forest-to-bog restoration or conversion for wind-farming.

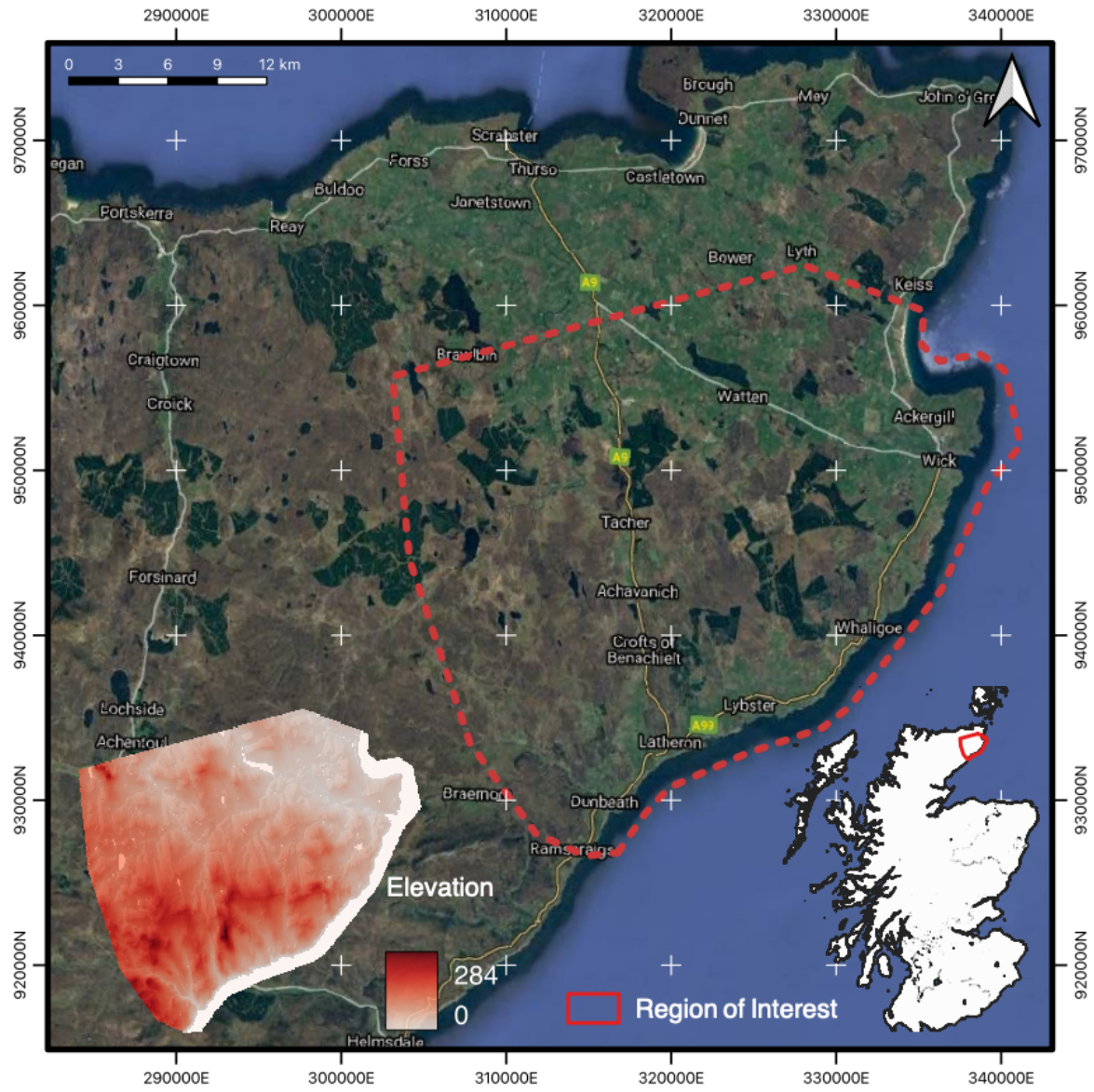

Figure 2 – Study Area

3.2 – Land Use Strategy Classes

The aim of this work is to classify a specific area of Scottish peatland into LU classes in accordance to the Scottish Land Use Strategy, which looks to balance ever-growing climate and anthropogenic needs to achieve sustainable land use practices. The LUS has separated Scotland into the following categories, within which key policies have been determined: 'Settlements', 'Enclosed Farmland', 'Semi-Natural Land', 'Rivers and Waterbodies', 'Coastal', 'Marine and Islands' (Scottish Government, 2021). These classes are compiled based on land productivity and potential, grouping of various land covers and soil types that share similar economic and environmental value. While some classes consist of narrowly defined purposes (e.g. Settlements, Rivers and Waterbodies, Marine), others demonstrate marked variation and compilation of

LULC subclasses. For example, Enclosed Farmland represents landscapes responsible for food production, livestock and "environmentally beneficial habitats" (Scottish Government, 2021, p. 21). Additionally, Semi-Natural Land, encompassing Scotland's "quintessential scenery", looks to represent mixed farmland, forest, hills, mountains and moors (Scottish Government, 2021, p. 22). Whilst LU class heterogeneity is likely present, their detection as a collection of subclasses is beneficial, with their similar spectral reflectance, moisture content and terrains enabling their grouping into these eight land use categories.

However, this study shall add two important classes to determine the effectiveness of the Scottish Land Use Strategy in meeting its targets as part of the ambitious climate goals set out by the *Update to the Climate Change Plan 2018-2032* (Scottish Government, 2021). Firstly, a class at the heart of this research, 'Peatland' is an important collection for the assessment of LULC change as part of a broader climate change plan. A significant proportion of the study area includes Flow Country peatland ecosystems, and the understanding of their interaction with surrounding LULC classes is of critical importance for future monitoring. Secondly, 'Forestry' represents an important class for the Scottish Governments *Update to the Climate Change Plan 2018-2032*. Hoping to average planting 18,000 hectares annually by 2024, this represents a significant land-cover that needs to be monitored to assess how effectively Scottish climate policies have been achieved. Furthermore, with significant areas undergoing forest-to-bog restoration, the changes in this LULC class will provide critical information for the conservation and restoration of Scottish peatland. A 'Forestry' class shall be determined by any land covered by coniferous trees, whilst forest-to-bog shall be classified as 'Peatland' upon removal of canopy cover (Scottish Natural Heritage, 2015b; Scottish Government, 2020). These classes shall be extracted from the 'Semi-Natural Land' LUS class and made into their own independent classes. As the study area includes no 'Islands', this class shall be removed leaving eight LUS classes.

3.3 – Training Data: Collection, Pre-processing and Sampling

For these classes to be well represented, training data is needed to feed a Random Forest classifier, with the quality of this data determining the accuracy that pixels are allocated to the specified classes. Soil mapping collects information regarding numerous soil features: "parent material, major soil group, and soil sub-groups, drainage and (for phases), texture, stoniness, land use, rockiness, topography and organic matter" (The James Hutton Institute, 2021). LULC mapping can use soil mapping data for training, as this provides accurate assessments of land

capability (an agricultural metric) and land suitability (a sustainability metric) (De Feudis *et al.*, 2021; The James Hutton Institute, 2021). When in-situ data collection is impractical, sampling existing soil maps with sufficient accuracies and coverage has been adopted as a suitable approach for LULC mapping, yet must appreciate limitations of the derived data (Zhang and Roy, 2017; Hermosilla *et al.*, 2018; Abdi, 2020). Collected from the 1938 to 1986, the Macaulay Institute for Soil Research's Soil Survey of Scotland undertook a systematic survey of Scottish soils, from which 1:10,000 up to 1:250,000 scale maps were generated (Lilly *et al.*, 2015). Now organised by the James Hutton Institute (JHI), resulting land capability and peatland probability maps have been generated, with the 1:25,000 map being used to determine the 'Peatland' training data (The James Hutton Institute, 2021). This map denotes peatland of >50cm depth, which was chosen to define the 'Peatland' class.

In addition, the UK Centre for Ecology and Hydrology's (UKCEH) Land Cover Maps from 2017, 2018 and 2019, using the UK Biodiversity Action Plan Broad Habitats classes (Jackson, 2000), derived training data for the other seven classes of the LUS. The Broad Habitats report dictates 21 classes, which were grouped in accordance with the LUS classes. The UKCEH's maps generated results suitable for sampling of training data, determining overall accuracies of 78.6%, 79.6% and 79.4% for the 2017, 2018 and 2019 maps respectively, justifying their inclusion and combination with the JHI peatland set (UK Centre for Ecology and Hydrology, 2020). Both datasets were acquired in ESRI Shapefile format and imported into QGIS for processing (QGIS., 2021). While GEE allows for vector data to be imported as personal assets, the functionality of QGIS vector tools and manipulation functionalities required dedicated GIS software. An initial cleaning of the individual land-cover maps was undertaken, resulting in a single dissolved polygons from which points were extracted.

Multiple strategies exist for sampling LULC maps as training data proxies. Stratified Random Proportional Sampling has been shown effective for RF where singular classes dominate landscapes, but is less suited to classification with multiple balanced classes (Belgiu and Drăgu, 2016; Shetty *et al.*, 2021). As our eight classes appear fairly evenly distributed throughout out landscape, Stratified Random Sampling is an appropriate sampling technique to address both larger class (E.g. Peatland, Enclosed Farmland) and minority classes (e.g. Marine, Inland Water, Coastal) with RF (Maxwell, Warner and Fang, 2018; Shetty *et al.*, 2021). A common metric to determine the number of training data points equates to 0.25% of the total region area, (~400 points per class), therefore a collection of 500 points for each class was extracted from the JHI

and UKCEH LULC maps (Thanh Noi and Kappas, 2017). These points were collected with a minimum separation of 100m. This resulted in 4,000 training points representing the eight LUS classes. These points were then saved and imported into GEE as individual *FeatureCollections* in preparation for classification, then split 70:30 for training and validation purposes (Huang *et al.*, 2020; Li *et al.*, 2020).

3.4 – Google Earth Engine

GEE is a platform capable of planetary-scale analysis, exploiting the advances of cloud-computing to divide computational power over numerous servers. This allows geospatial analysis to be undertaken by expert and non-expert users within a fully integrated environment. GEE provides an extensive integrated data catalogue, storing vast amounts of geospatial and auxiliary data including the ESA's Copernicus Sentinel Series and NASA's Landsat series. In addition, GEE allows users to import/export data as assets in various geospatial formats, allowing personal data to be processed within the GEE Application Programming Interface (API). This API (accommodating both JavaScript and Python programming languages) allows users to undertake computation in its built-in web-based Interactive Development Environment (IDE) which provides the ability to interact with the numerous algorithms, datasets and visualisations. All data was accessed within (Senintel-1 and -2) or imported into (LULC training data and NASADEM topographic indices) GEE for analysis, using Google Drive for data storage and exporting (Gorelick *et al.*, 2017). Individual codes were written using JavaScript to handle individual processes of the classification workflow, compiled when the full workflow was executed. Figure 3 demonstrates the workflow undertaken to prepare data and create features in preparation for the Random Forest classification.

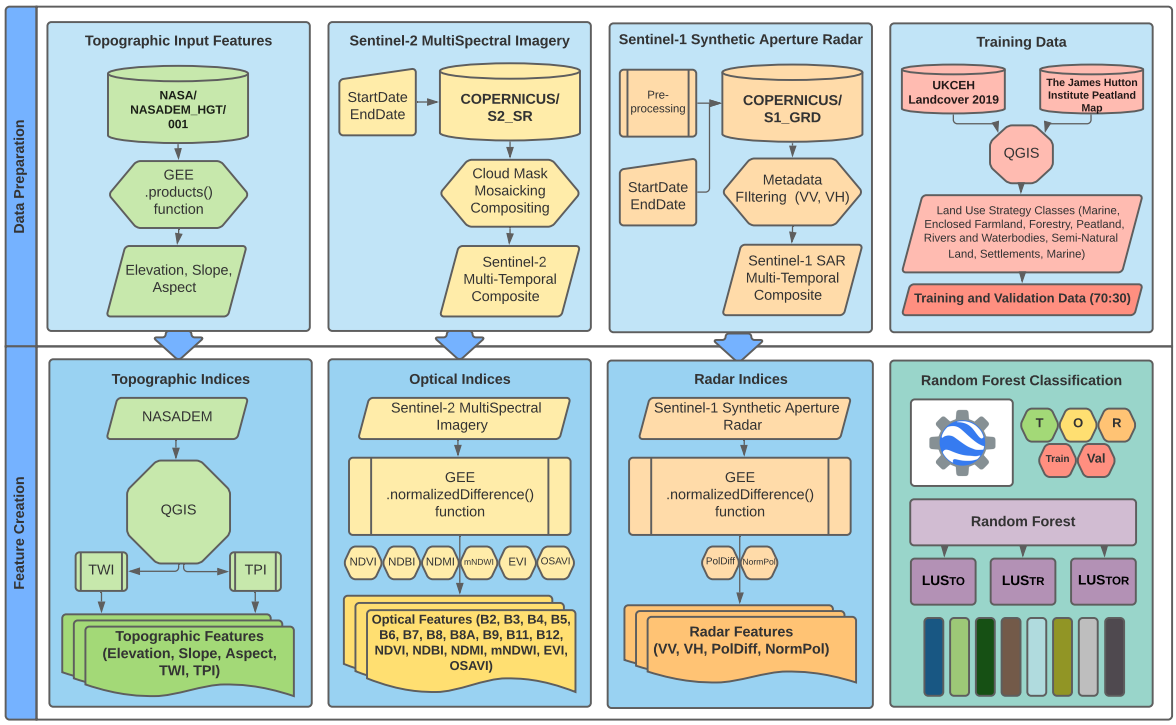

Figure 3 – Google Earth Engine Workflow

3.5 – Optical: Sentinel-2 Multi-Spectral Instrument (MSI)

The ESA's Sentinel-2 provides Multi-Spectral optical imagery obtained through two complimentary sensors (-2A and -2B), with global coverage every five days. The dual Multi-Spectral Instruments (MSI) record 13 optical bands, ranging from visible to Shortwave Infrared (SWIR). Sentinel-2 (S2) data can be provided in 10m, 20m and 60m resolutions, the best spectral, temporal and spatial resolutions available from current optical satellite data (Zhang *et al.*, 2020). Sentinel-2A imagery comes in both Top-of-Atmosphere (Level-1C) and Surface Reflectance (Level-2A) collections. The latter, chosen for this study due to its superiority in removing atmospheric haze, provides orthorectified atmospherically corrected imagery for the period assessed. Sentinel Level-2A data has coverage starting from March 2017, which allows for the assessment of monthly, seasonal and yearly analysis. To accurately reflect the study area, it was decided a year's worth of S2 data (1st January 2020 - 31st December 2020) would be used to assess the LUS classes. Using multiple optical scenes allows for the mitigation of inherent optical limitations, notably cloud cover, atmospheric haze, shadows and snow. Sentinel images for these date ranges were selected within GEE, with a resulting *ImageCollection* of 146 images being acquired for the defined study area.

As the study area experiences frequent cloud cover and snow during various seasons, a custom cloudmasking script was written and mapped over the *ImageCollection* to remove cloud-covered pixels. This involved the exploitation of S2's Coastal Aerosol (B1) and QA60 cloud mask quality bands, both of which allow for the identification and removal of cirrus, opaque and shadow cloud cover (Carrasco *et al.*, 2019; Belcore, Piras and Wozniak, 2020). A threshold of >1500 for B1 was used in accordance to existing peatland classification mapping (DeLancey *et al.*, 2019). For this data range to be evaluated collectively, compositing of the *ImageCollection* is needed to create a statistically coherent, singular cloud-free image from which band values can be extracted for classification. Compositing refers to the statistical process of combining spatially-overlapping images into a singular aggregated image. Through this process, a whole period can be represented, and individual image flaws can be mitigated. Once cloud and shadow pixels had been removed from each image, a singular temporal composite was obtained using the GEE *.median()* reducer, eliminating outlying pixels while maintaining the integrity and spatial variability of the 146 S2 images.

3.6 – Spectral Indices

As discussed, spectral indices derived from optical bands provide important information for the delineation of classes, enabling ecological, hydrological and geomorphological features to be identified. The selection of suitable indices should look to assist the classification of key LUS class characteristics, specifically greenness, wetness and soil-sensitive spectral properties. Numerous indices have been used for peatland and LULC mapping, but an identifiable pattern denotes using a balanced selection of uncorrelated indices that reflects the heterogeneity of the landscape in question. This typically includes a vegetation-based (e.g. NDVI, EVI), moisture-based (e.g. NDMI), water-based (e.g. mNDWI) and soil-based (e.g. OSAVI, NDBI) index (Rouse *et al.*, 1974; Rondeaux, Steven and Baret, 1996; Huete *et al.*, 2002; Gerard *et al.*, 2003; Zha, Gao and Ni, 2003; Xu, 2006). For the following study, Figure 4 shows the indices calculated and their calculations from the S2 Level-2A optical data.

While certain indices have been well-established within literature (e.g. NDVI and EVI), derivatives of established indices have been included here based on their effectiveness for detecting classes within the study area. OSAVI looks to appreciate the colour and reflectance of soil types, tuning of the X variable (to 0.16) which can best determine darker peat soils (Rondeaux, Steven and Baret, 1996). Furthermore, Xu (2006) determine a modified NDWI can

provide enhanced quality delineating water features, which has been an important alteration for wetland and water mapping (Xu, 2006; Li *et al.*, 2020). GEE provides numerous integrated features to calculate spectral indices, with the *.normalizedDifference()* function a convenient way of determining spectral indices. Each formula from Figure 4 was implemented into a single function and mapped over every image in the Sentinel-2 *ImageCollection* prior to compositing (Fig. 3). This allowed for the resulting composition to calculate medians of all the spectral indices in preparation for the final classification features.

Spectral Indices	Author(s)	Equation for Sentinel-2
NDVI	Rouse et al. (1974)	(B8 – B4)/(B8 + B4)
EVI	Huete et al. (2002)	2.5*(B8 − B4)/(B8 + 6*B4 − 7.5*B2 + 1)
NDMI	Gerard et al. (2003)	(B8 - B11)/(B8 + B11)
mNDWI	Xu et al. (2006)	(B3 - B11)/(B3 + B11)
OSAVI	Rondeaux et al. (1996)	1.16*B8 – B4/B8 + B4 + 0.16
NDBI	Zha et al. (2003)	(B11 – B8)/(B11 + B8)

Figure 4 – Spectral Indices derived from Sentinel-2 Level-2A Multi-Spectral Imagery

3.7 – Radar: Sentinel-1 GRD C-Band Synthetic Aperture Radar

To accompany Sentinel-2 and its derived indices, ESA's Sentinel-1 (S1) Ground Range Detected (GRD) C-Band Synthetic Aperture Radar (SAR) was utilised within GEE. S1 provides Interferometric Wide Swath data at a spatial resolution of 10m, with a coverage every six days over northern latitudes. S1 provides single-polarized vertically transmitted (VV) or dual-polarized vertically transmitted and horizontally received (VV/VH) data (European Space Agency, 2021c). S1 C-Band SAR, a formation of two side-looking sensors (-1A and -1B), is vital for the penetration of cloud cover, vegetation and soil to derive additional information for peatland and LULC mapping (Karlson *et al.*, 2019; Lees *et al.*, 2021). As part of its integrated pre-processing (Sentinel-1 Toolbox), data undergoes: thermal noise removal; radiometric calibration; terrain correction using SRTM30 or ASTER DEM; and conversion of backscatter coefficient to decibels (European Space Agency, 2021b). Accessing S1 SAR in GEE provides data after initial processing in a complete *ImageCollection* with documentation instructing how images can be filtered based on inherent metadata properties.

To prepare S1 SAR for classification, the same date range (1st January 2020 – 31st December 2020) was used to determine an *ImageCollection* of scenes within the study area. Metadata filters

determined scenes with dual-polarization of VV and VH bands, resulting in 330 total scenes. For each scene, two new bands were calculated to determine the Normalized Polarization (NormPol) and Polarization Difference (PolDiff)of the VV and VH bands. NormPol was created using the *.normalizedDifference()* function similar to the spectral indices obtained for S2 (Hird *et al.*, 2017; DeLancey *et al.*, 2019; Poggio, Lassauce and Gimona, 2019). Determined by the following equation, values should theoretically range between -1 and +1:

$$\frac{VH - VV}{VH + VV}$$

PolDiff defines the difference between the single co-polarised VV and dual-band co-polarised VH bands (Carrasco *et al.*, 2019), defined below:

$$VV - VH$$

Once bands had been added to each scene within the S1 *ImageCollection*, a per-pixel *.mean()* reducer was used to aggregate all bands into a single composite image, including the polarization equations as new bands. With identical spatial resolution, this was then combined with the S2 data by adding the S1 bands into a singular composite. This marked the completion of the optical and radar preparation within GEE, creating a pool of 34 bands in preparation for Random Forest classification.

3.8 – Topography: NASADEM and Topographic Indices

Topography is an important addition to peatland and LULC classification, particularly as wetland formation is highly dependent on local topographic character (Karlson *et al.*, 2019). Whilst terrain varies only slightly on the site, with a maximum elevation of 284m above sea-level, topographic derivates such as elevation, slope and aspect are vital for the metrics to determine flow, hydrology and geomorphology (Gallant and Wilson, 2000). Historically, Digital Elevation Model's (DEMs) such as NASA's Shuttle Radar Topography Mission (SRTM) and the Advanced Spaceborne Thermal Emission and Reflection Radiometer-Global Digital Elevation Model (ASTER-GDEM) allow for the analysis of vast geographies for LULC mapping. However, the recent release of NASADEM, a modernized DEM broadly based on the SRTM, has reprocessed the original SRTM signals with the Ice, Cloud, and Land Elevation Satellite (ICESat) Geoscience Laser Altimeter System (GLAS) and Advanced Spaceborne Thermal Emission and Reflection

Radiometer (ASTER) instruments. This has provided near-global elevation data (+80%) through interferometry SAR (InSAR) and LiDAR sensors to increase accuracy and fill gaps present within the original SRTM DEM (Buckley *et al.*, 2020). NASADEM provides a resolution of 1-arc second, equating to roughly 30 metres. This DEM was used to determine the elevation, slope, aspect and topographic indices of the ROI.

Derived from a base DEM, topographic indices provided considerable detail about a ROI's hydrological and geomorphological processes, notably its flow patterns, flow accumulation and catchment. The Topographic Wetness Index (TWI), based on slope and upslope contributing area, is considered a proxy for soil moisture within a particular vicinity, a significant feature for wetland and peat mapping (Hird *et al.*, 2017; Poggio, Lassauce and Gimona, 2019). TWI can be determined through the following equation:

$$TWI = ln\ \frac{\alpha}{\tan \beta}$$

Where α is the upslope contributing area and tanβ represents the local slope gradient (Beven and Kirkby, 1979). Areas with high TWI scores are expected to be wetter than those with lower scores, representing soils with greater water retention. The implementation of this equation followed the System for Automated Geoscientific Analyses (SAGA) workflow of Mattivi et al. (2019), executed in QGIS. This require the calculation of fill sink, flow accumulation, and flow width and specific catchment area, which were then used as proxies to calculate the TWI (Mattivi *et al.*, 2019). In addition, the Topographic Position Index (TPI) was calculated to accompany TWI as an additional derived topographic index. TPI measures the topographic position of a central cell in relation to the average elevation of its neighbouring cells, a pre-defined radius specified by the user (Gallant and Wilson, 2000). TPI can be calculated using the following equation:

$$TPI = Z_i - \bar{Z}_{R(i)}$$

Here, Z_i is the elevation of i^{th} cell and $Z_{R(I)}$ is the mean elevation with the specified radius (50m). Similarly, SAGA was used within QGIS to determine this calculation. TPI values ranged from -14.40 depicting cells significantly lower than their surroundings, to 10.65 representing those above their neighbouring cells. Once completed, both indices were exported from QGIS, imported into GEE as user assets and combined with the NASADEM derivatives.

3.9 – Random Forest: Variable Selection, Hyperparameter Tuning and Accuracy Assessment

Once optical, radar and topographic datasets were derived, preparation of three individual models was undertaken. This work looks to compare the significance of optical, radar and topographic features, and therefore three models were outline: LUS_{TO} (Topographic and Optical), LUS_{TR} (Topographic and Radar) and LUS_{TOR} (Topographic, Optical and Radar). The creation of three models will allow for the contrasting of input feature importance, the weight of individual bands/indices and the performance of RF. These combinations were based on the importance of topography within the literature, which justified its use as a constant within each model. Each model was determined suing the same RF classification scheme, individually tuned to obtain the strongest overall accuracy based on the input features selected. While the removal of correlated features is a useful step to improve RF classification, to maintain consistency grouped input features shall remain unchanged throughout each of the three models, meaning comparisons can be made between each model upon tuning completion. Figure 5. shows the combination of input features for the LUS_{TO}, LUS_{TR} and LUS_{TOR} models. Once the bands were selected, the RF classifier was run in GEE independently for each model, each using the same the training data *FeatureCollection*.

Model	Input Features
LUS_{TO}	**Topographic**: Elevation, Slope, Aspect, TWI, TPI **Optical**: B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI, NDMI, NDBI, mNDWI, OSAVI
LUS_{TR}	**Topographic**: Elevation, Slope, Aspect, TWI, TPI **Radar:** VV, VH, NormPol, PolDiff
LUS_{TOR}	**Topographic**: Elevation, Slope, Aspect, TWI, TPI **Optical**: B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, NDVI, NDMI, NDBI, mNDWI, OSAVI **Radar:** VV, VH, NormPol, PolDiff

Figure 5: LUS_{TO}, LUS_{TR} and LUS_{TOR} Model Input Features

The tuning of a Random Forest classifier has been shown as a significant prerequisite for its successful use. This tuning includes the selection of a suitable number of decision trees (Ntree) and the determination of nodes (Mtry). RF provides built-in measures to determine the success of these parameters and classification iterations are necessary to produce the best accuracy (Maxwell, Warner and Fang, 2018). Each model was run independently to determine the best parameters for each input feature combination. The choice of RF Ntree and Mtry parameters

were determined based on iterations of the RF classifier with each model. These iterations began with an Ntree value of 10 and iterated in 10 Ntree increments up to 500, from which the Ntree value producing the highest overall accuracy was selected. The determination of an appropriate Mtry value resulted in established use of the square root of the quantity of input variables (Belgiu and Drăgu, 2016). As noted previously, a significant strength of RF is its built-in variable importance metric, which determines the weighted importance of input features in a classification. For each model, LUS_{TO}, LUS_{TR} and LUS_{TOR}, variable importance, NTree iterations, Confusion Matrix and class area were obtained to evaluate the performance of input features.

4 – RESULTS

The following results depict the classification of three models, LUS_{TO}, LUS_{TR} and LUS_{TOR}, in accordance with the classes of the Scottish Land Use Strategy. To determine the accuracy of models, several accuracy metrics were calculated. This included Overall Accuracy (OA), Producer Accuracy (PA), User Accuracy (UA) and Kappa Accuracy (KA). OA dictates the overall accuracy of the model for predicting classes based on the input training data. Kappa denotes the overall accuracy of a classification appreciating that results can occur by chance, producing an accuracy slightly less than that of OA. PA and UA can be distinguished through the following statement: "User's accuracy is a measure of class reliability derived from the fraction of correctly classified pixels in relation to all pixels classified as this class in the map, [while] Producer's accuracy is the fraction of correctly classified pixels with regard to all pixels of a particular class in the validation dataset" (Karlson *et al.*, 2019, p.5).

Additionally, for each model; a confusion matrix depicting misclassifications; RF hyperparameter tuning through Ntree iterations; and the variable importance of input features were determined. A confusion matrix has been used to depict the confusion experienced in the classifier, providing useful information to determine where there are misclassifications and which classes can be improved for future alterations. Furthermore, variable importance has been calculated to determine "out-of-bag" error (rfOOBe) (Breiman, 2001). This error is calculated by using withheld training data during classification to determine the mean decrease in accuracy when an input feature is not in the decision tree construction. The inclusion of this metric allows for the identification of variables that are significant in the classification, pointing to others that can be removed or amended for future models (Millard and Richardson, 2015).

4.1 – LUS_{TO} – Land Use Strategy (Topography and Optical)

The results for the first classification model incorporating the derived topographic and optical data are displayed in Figure 6a. This model incorporated the bands outline in Figure 5., creating a collection of 21 bands for the RF classifier. After initial iterations of the RF classifier were run, the hyperparameter tuning (Fig. 6c) determined 70 trees would provide the best overall accuracy. Once run with 70 trees, an overall accuracy of 0.789 was obtained. This suggests that almost 8/10 pixels were accurately classified when determined with the RF classifier. Upon visual inspection, the map has determined classes adequately, recognising the major features within the

landscape. A confusion matrix was used to determine the misclassification of the eight classes, including the UA and PA (Fig. 6b). This matrix showed 'Marine' (UA=0.987, PA=0.979), Forestry (UA=0.929, PA=0.917) and 'River and Waterbodies' (UA=0.978, PA=0.950) to be the strongest delineated classes.

Whilst other UA's and PA's were comparable, 'Semi-Natural Land' presenting markedly poorer PA (0.543). On assessment of variable importance (Fig. 6d), 'Elevation' was determined as the most important feature with the least rfOOBe. while remaining variables can be deemed comparable in importance. The mNDWI, used to delineate bodies of water and its reluctance properties, was the second most important feature, evident from the strong classification of the 'Marine' and 'Rivers and Waterbodies' classes in the confusion matrix. However, two of the five topographic inputs were deemed the least helpful for classification, which is unsurprising in this geomorphologically flat, undulating terrain. These classification results demonstrate the strength that optical data brings to LULC mapping, critical for the separation of vegetation and soil based on spectral reflectance properties.

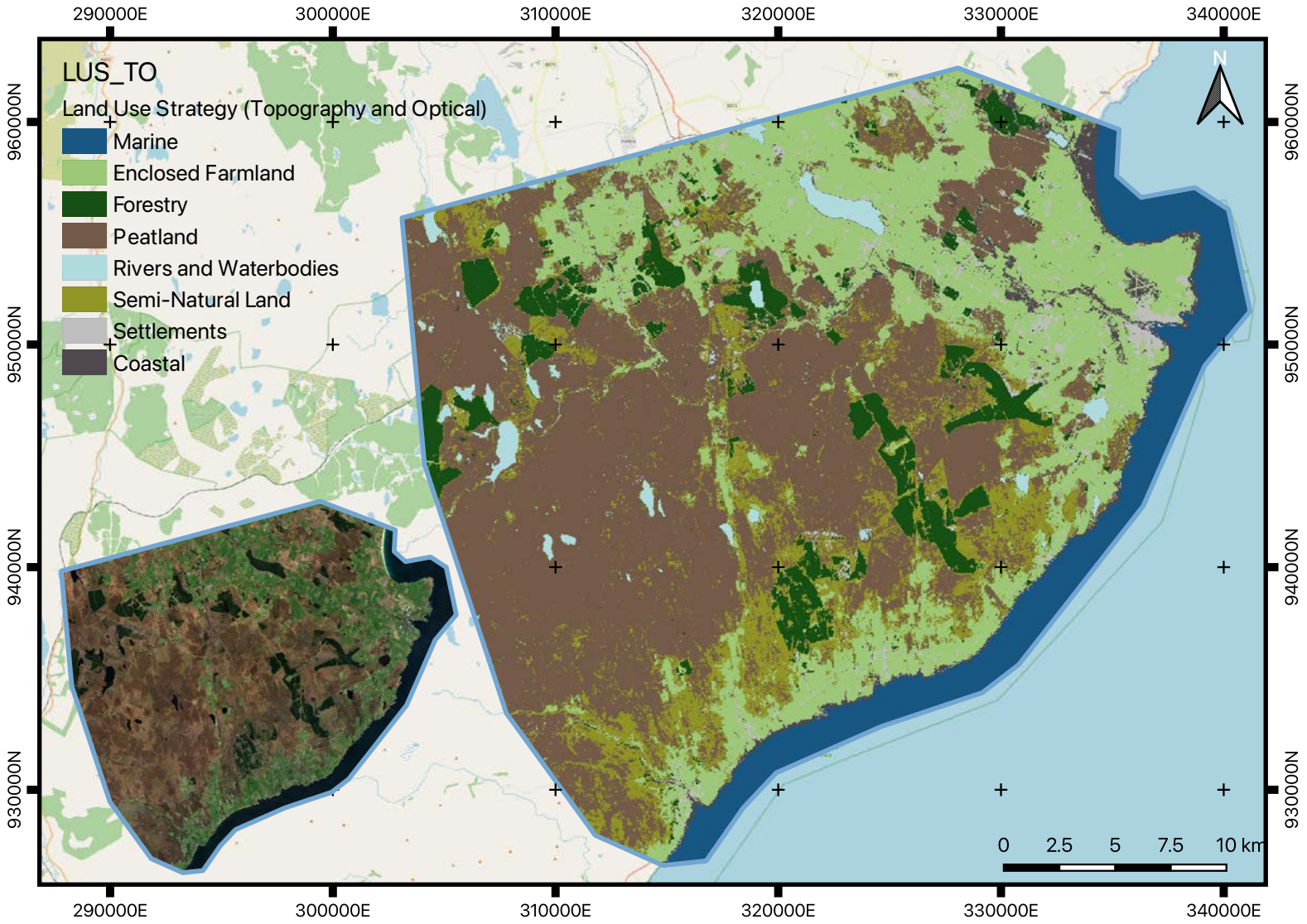

Figure 6b: LUS_{TO} Classification Results

LUSTO	Marine	Enclosed Farmland	Forestry	Peatland	Rivers and Water-bodies	Semi-Natural Land	Settlements	Coastal	Producer Accuracy
Marine	137	0	0	0	0	0	0	3	0.979
Enclosed Farmland	1	121	0	10	0	5	10	5	0.796
Forestry	0	1	143	2	0	10	0	0	0.917
Peatland	0	15	7	242	3	31	0	1	0.809
Rivers and Waterbodies	0	1	0	2	134	1	1	2	0.950
Semi-Natural Land	0	8	4	54	0	82	0	3	0.543
Settlements	0	9	0	0	0	1	50	7	0.746
Coastal	2	5	0	4	0	2	4	72	0.808
User Accuracy	0.987	0.756	0.929	0.771	0.978	0.621	0.770	0.774	**0.821**

Figure 6b: LUSTO Confusion Matrix

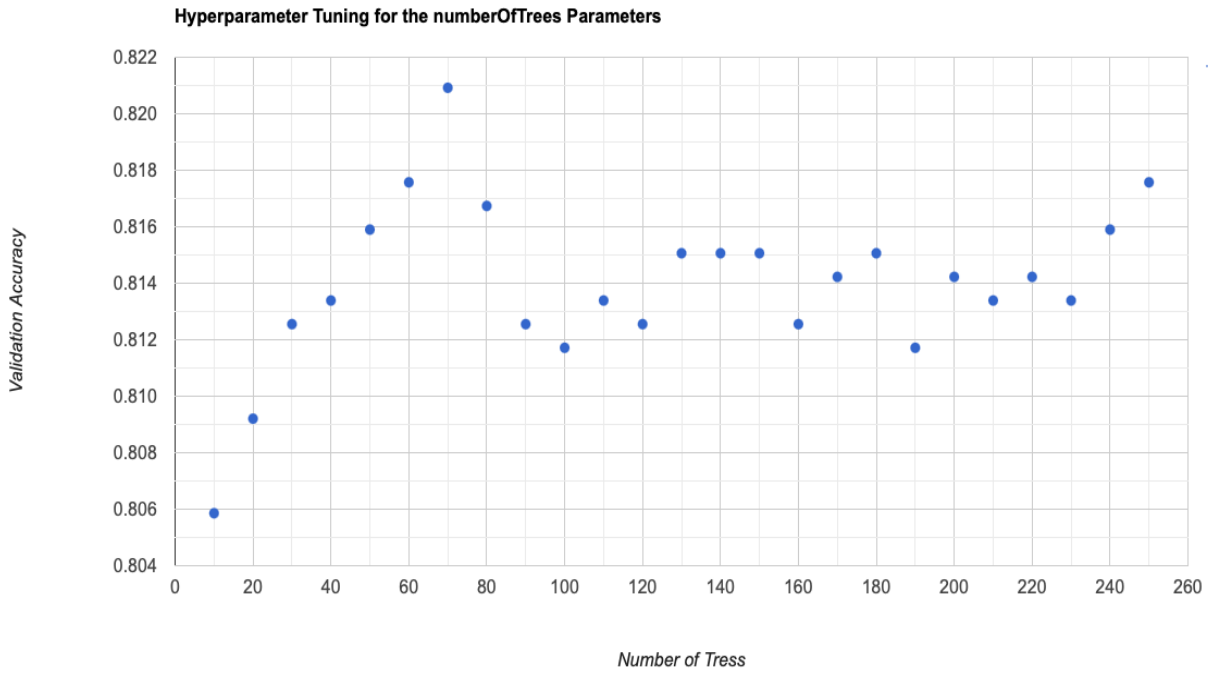

Fig 6c: LUSTO Hyperparameter Tuning for the number of Random Forest Trees (Ntree)

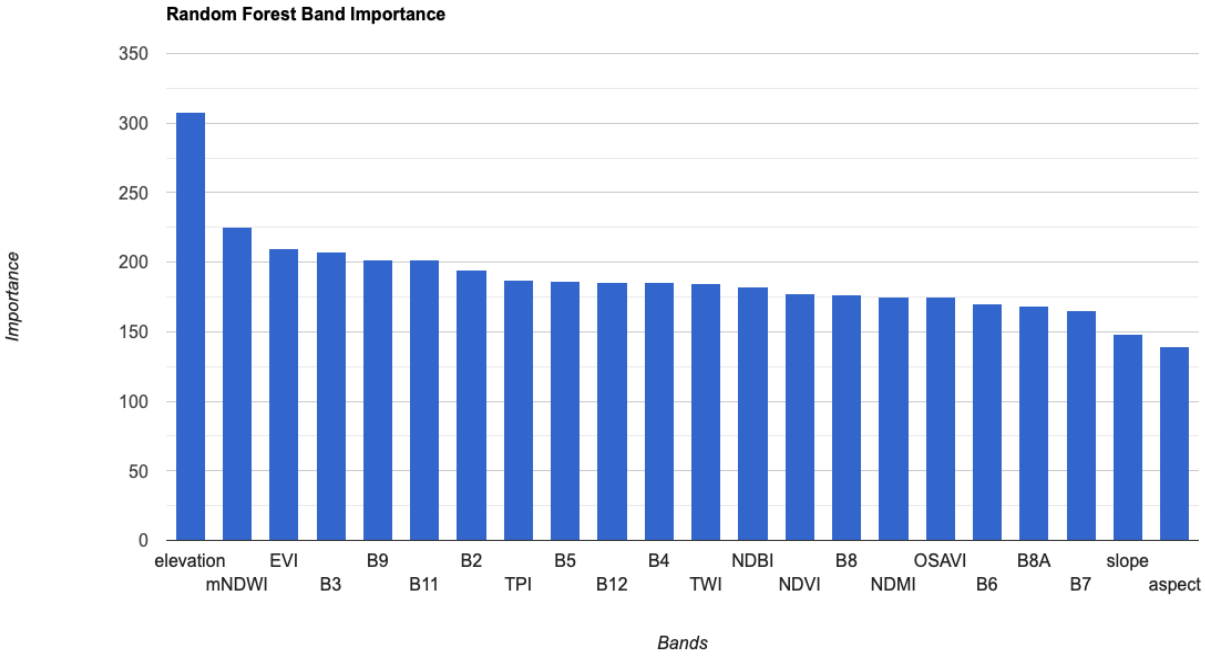

Figure 6d: LUS_{TO} Random Forest Variable Importance

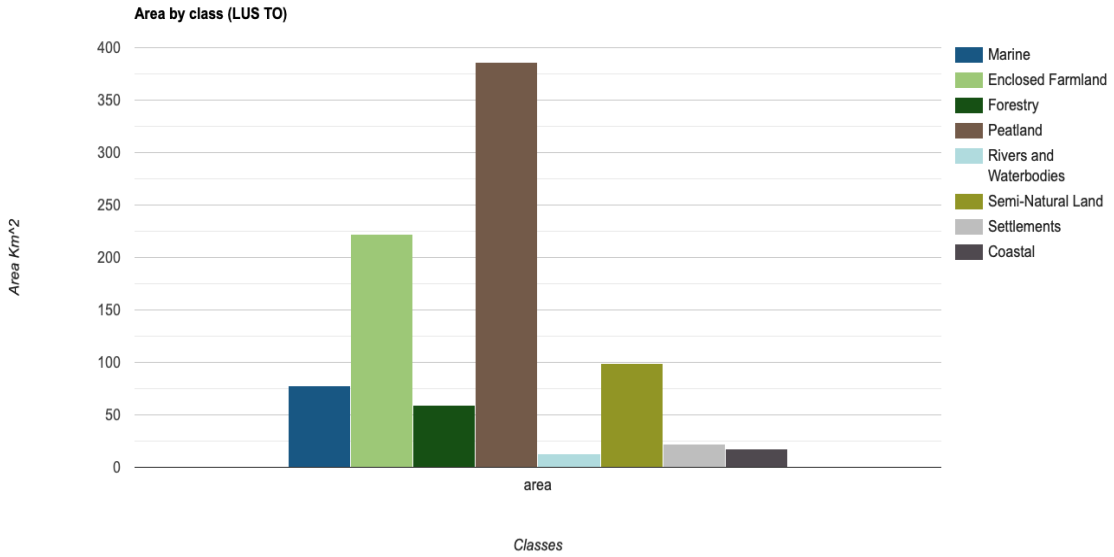

Figure 6e: LUS_{TO} Class Area (km^2)

4.2 – LUS_{TR} – Land Use Strategy (Topography and Radar)

The second classification model, LUS_{TR}, included the topographic and radar features as described in Table 1. This fed 9 bands, the least of the three models, into the RF classifier for initial assessment of classification performance. A preliminary test for the best hyperparameters dictated 210 Ntree would produce the best OA of 0.786 and KA of 0.748. This parameter was then set along with the Mtry (3) the determine the LULC classification results, confusion matrix and variable importance. Upon inspection of the confusion matrix, trends are evident in the classification of 'Marine', 'Forestry' and 'Rivers and Waterbodies', showing comparable UA and PA scores. Differences were identified in the poor UA experienced by LUS_{TR} in classifying 'Enclosed Farmland' and 'Semi-Natural Land', showing confusion with 'Peatland'. As SAR and its derivatives are used to detect soil saturation and surface moisture, there are likely explanations for these misclassifications. With a comparably smaller number of input features, for a supervised learning algorithm capable and suited to high-dimensionality analysis, it is expected that LUS_{TR} would produce reduced OA than LUS_{TO}.

Upon visual inspection and comparison to LUS_{TO}, the misclassifications become more visible. Firstly, LUS_{TR} has significantly more speckling than $LUS_{TO,}$ a product of the greater misclassifications seen in the confusion matrix. However, delineation between 'Peatland' and 'Semi-Natural Land' classes shows less definable boundaries when compared to $LUS_{TO,}$ which may be indicative of the presence of similar soil types. As 'Peatland' was defined as peat soils ≥ 50cm, radar may have identified areas as 'Semi-Natural Land' with 'peaty' soils with similar soil saturation and moisture content to that of the 'Peatland' class. However, LUS_{TR} has misclassified a significant portion of peatland in the north-east of the ROI identified by the other models. In terms of the LULC map's application for LUS policy assessments, the misclassification of significant areas of peatland demonstrates a problematic flaw, suggesting at the need for additional features to improve classification. Assessing variable importance, elevation is repeatedly the most important feature, while other variables presented comparable importance. Notably, slope and aspect were the least useful variables in both the LUS_{TO} and LUS_{TR} models. This model demonstrates the ability for radar derivatives to determine areas of peatland and determine LULC classes, importantly identifying soil characteristics and water features.

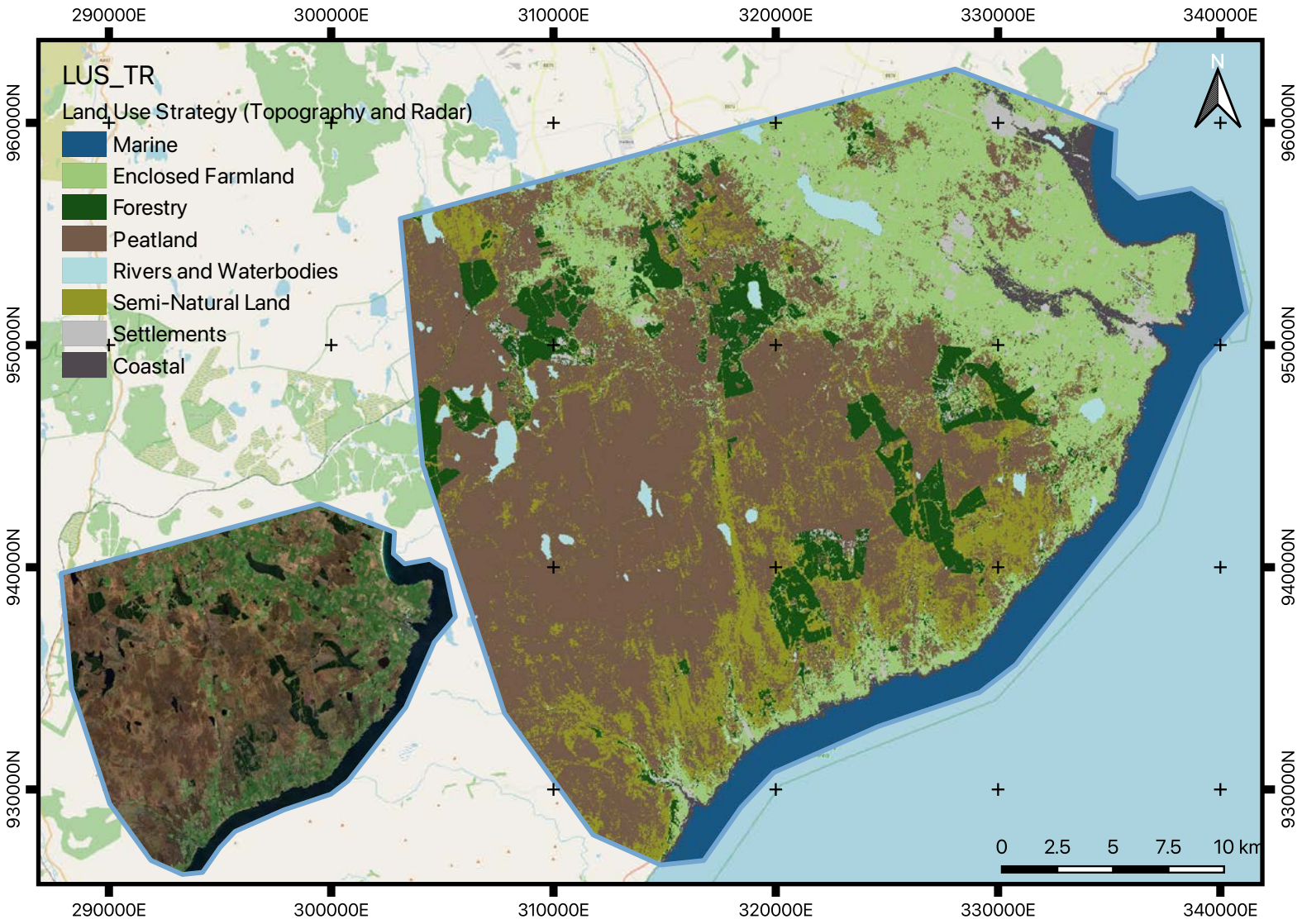

Figure 7a: LUS_{TR} Classification Results

LUS_{TR}	Marine	Enclosed Farmland	Forestry	Peatland	Rivers and Water-bodies	Semi-Natural Land	Settlements	Coastal	Producer's Accuracy
Marine	136	0	0	0	0	0	0	4	0.9714
Enclosed Farmland	0	93	5	25	1	15	9	4	0.612
Forestry	0	5	142	3	0	6	0	0	0.910
Peatland	0	28	7	236	3	24	1	0	0.789
Rivers and Waterbodies	0	1	0	1	139	0	0	0	0.986
Semi-Natural Land	0	13	2	56	1	77	1	1	0.510
Settlements	0	13	1	3	2	0	42	6	0.627
Coastal	3	8	0	0	1	0	3	74	0.831
Users Accuracy	0.978	0.577	0.904	0.728	0.946	0.631	0.750	0.831	**0.786**

Figure 7b: LUS_{TR} Confusion Matrix

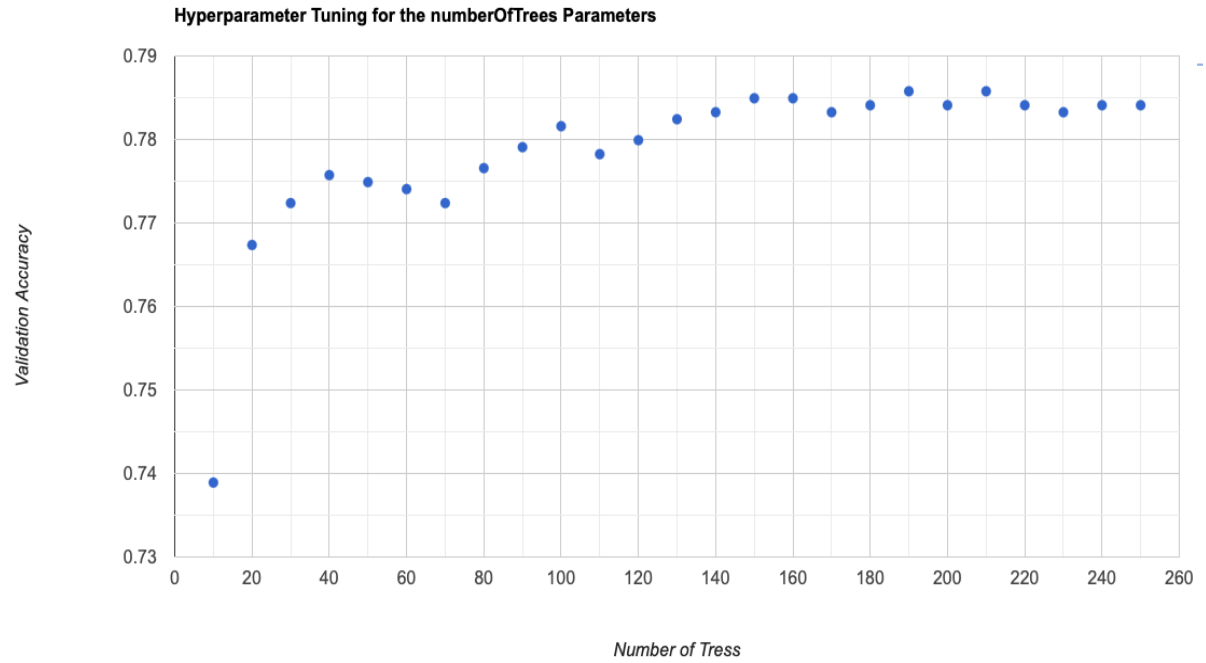

Fig 5: LUS_{TR} Hyperparameter Tuning for the number of Random Forest Trees (Ntree)

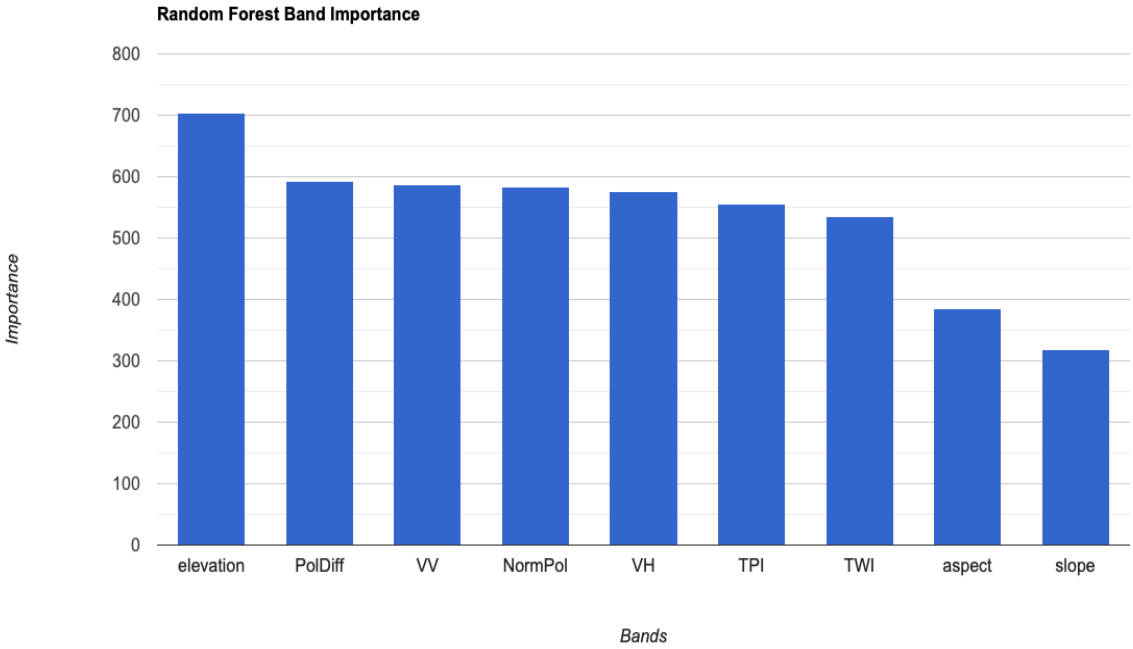

Figure 5: LUS_{TR} Random Forest Variable Importance

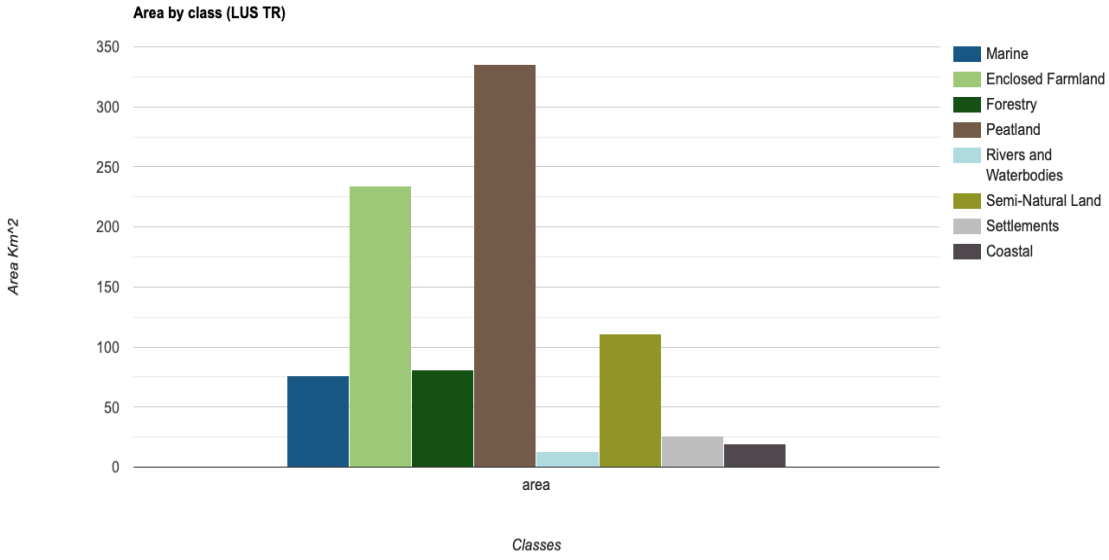

Figure 7e: LUS_{TR} Class Area (km^2)

4.3 – LUS_{TOR} - Land Use Strategy (Topography, Optical and Radar)

The final classification model, LUS_{TOR}, combined the topographic, optical and radar data, resulting in 25 bands input into the RF classifier (Table 1). As with the other classification models, an initial classifier was run to determine the best number of trees to run in the RF classifier. This resulted in an Ntree value of 350, representing 350 decisions trees. The obtained OA and KA were 0.823 and 0.792 respectively. This accuracy is almost identical to that of the LUS_{TO} model, suggesting the inability of radar to provide any significant advantages on top of the LUS_{TO} model when all combined without reduction. The derived confusion matrix (Fig. 8a) shows similar trends to that of LUS_{TR} and LUS_{TO}, with 'Marine', 'Forestry' and 'Rivers and Waterbodies' with strong PA and UA scores. Due to their unique spectral signatures, these features result in highly accurate identification and delineation in RF classification.

Similar to LUS_{TO}, 'Semi-Natural Land' produced the lowest PA and UA scores, suggesting at the spectral similarity between these features and 'Peatland' pixels. While peat soils and other soil types may have significantly different compositions, such differences are unlikely to be detected remotely, requiring accurate soil samples to separate these classes. This limitation needs to be explored to determine the effectiveness of LULC mapping of heterogenous landscapes. Upon visual inspection of the LUS_{TOR} map, evident strengths can be seen in the separation of urban settlements, including the identification of roads, houses and even windfarms. Furthermore, areas of forestry have not only been well classified, but areas of forest-to-bog restoration have suggested at the partial return of forested soils to 'Semi-Natural' soils (Fig. 8f). This identification process could be important in the assessment of Scottish forestry policies as part of the Land Use Strategy and *Update to the Climate Change Plan*.

Fig. 8d shows the variable importance of the LUS_{TOR} bands in the RF classifier. Much like the previous two models, an identifiable trend is present. Slope and aspect were the least useful bands in each of the three classifiers, likely due to the consistent topography of the ROI, while mNDWI was an important input feature for both LUS_{TO} and LUS_{TOR} models. As previously stated, this is evident from the strength of the water-based classes, with mNDWI a modified water index for delineating inland and oceanic water features. However, contrary to the OA determined for LUS_{TR}, VH, VV, PolDiff and NormPol were weighted higher than many of the optical bands and derived indices (Fig.12). This suggests that improvements to the processing and preparation of radar derivatives may lead to better classification.

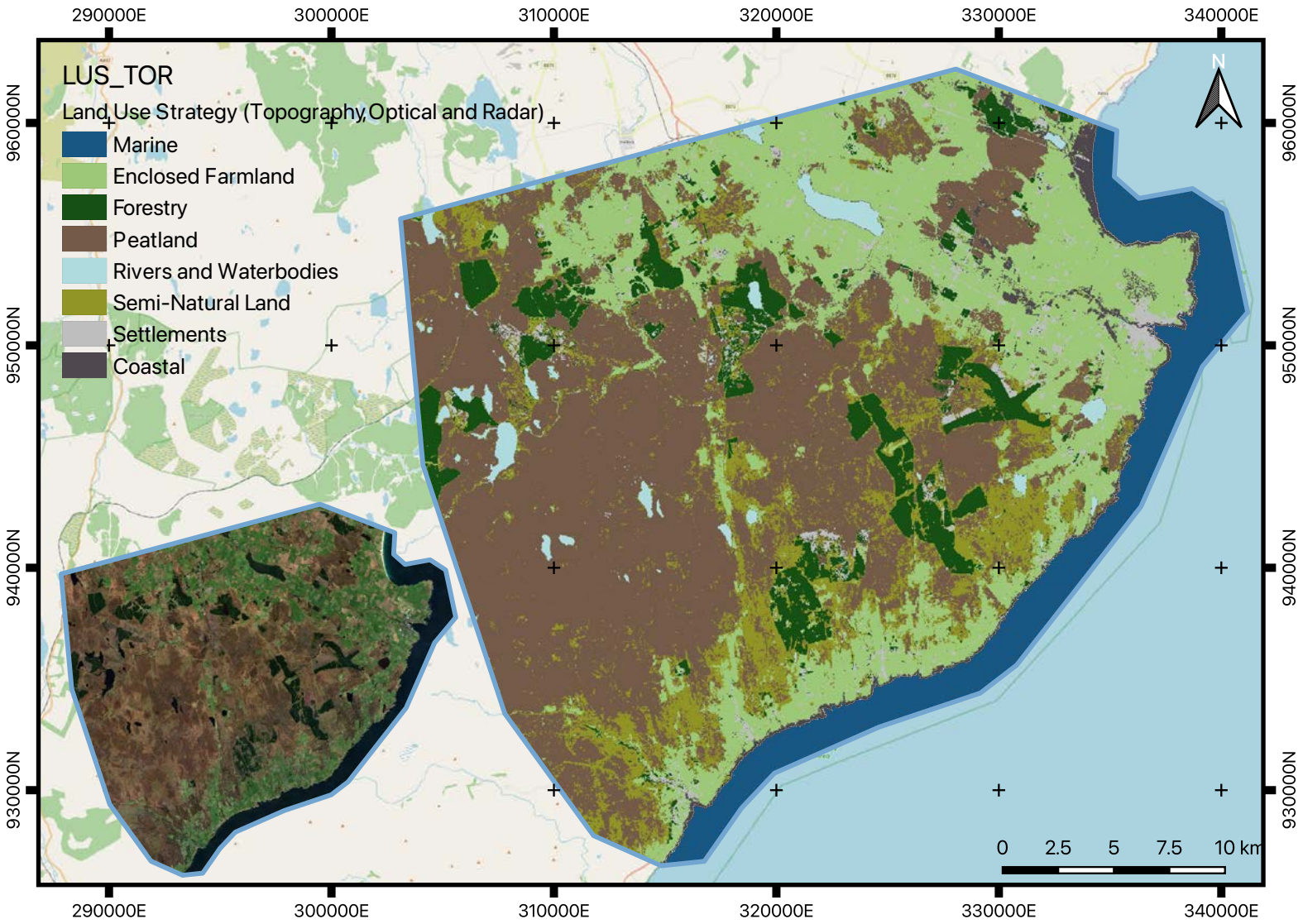

Figure 8a: LUS_{TOR} Classification Results

LUS_{TOR}	Marine	Enclosed Farmland	Forestry	Peatland	Rivers and Water-bodies	Semi-Natural Land	Settlements	Coastal	Producer's Accuracy
Marine	137	0	0	0	0	0	0	3	0.979
Enclosed Farmland	0	121		11	1	4	9	6	0.796
Forestry	0	1	144	2	0	8	1	0	0.923
Peatland	0	13	8	247	4	27	0	0	0.826
Rivers and Waterbodies	0	1	0	1	136	2	0	1	0.964
Semi-Natural Land	0	8	4	58	0	78	1	2	0.517
Settlements	0	11	0	0	0	0	48	8	0.716
Coastal	2	5	0	3	1	2	5	73	0.820
Users Accuracy	0.986	0.756	0.923	0.767	0.958	0.644	0.774	0.785	**0.823**

Table 8b: LUS_{TOR} Confusion Matrix

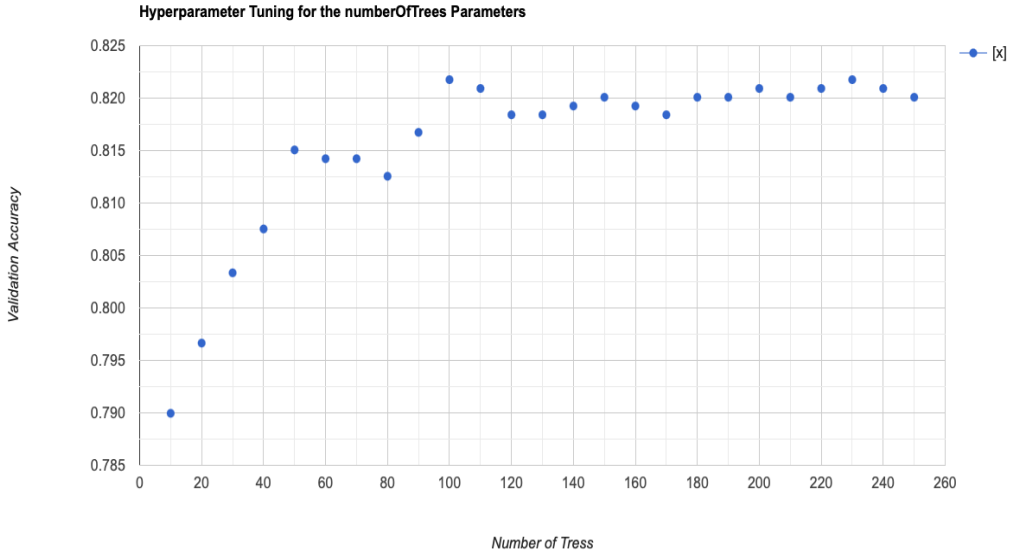

Fig 8c: LUS_{TR} Hyperparameter Tuning for the number of Random Forest Trees (Ntree)

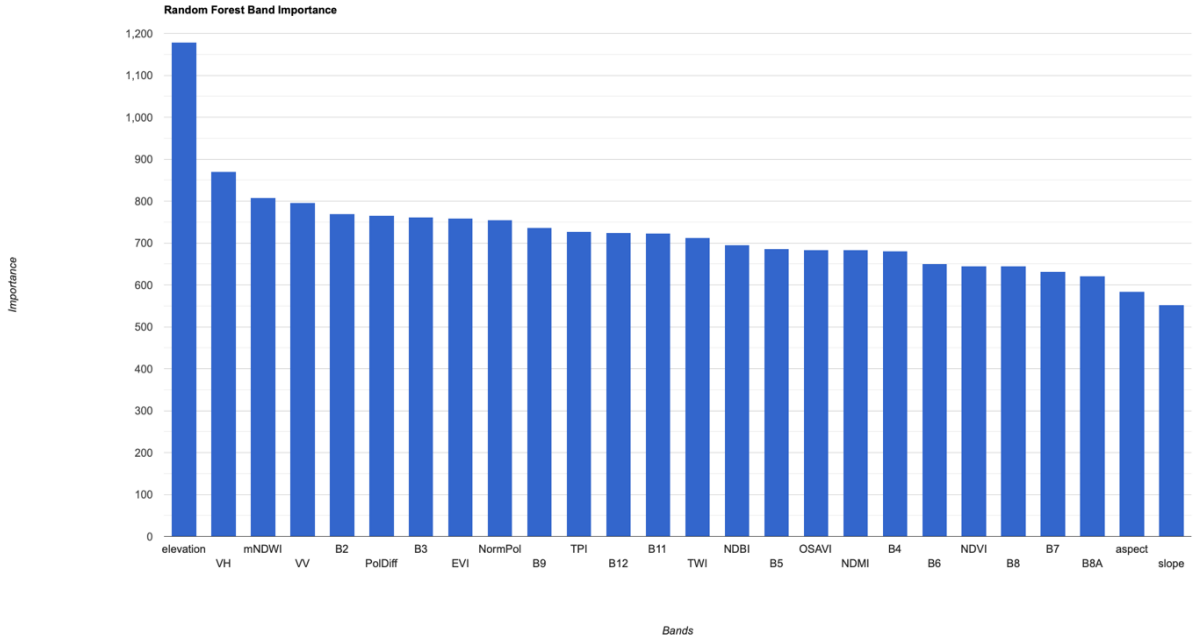

Figure 8d: LUS_{TR} Random Forest Variable Importance

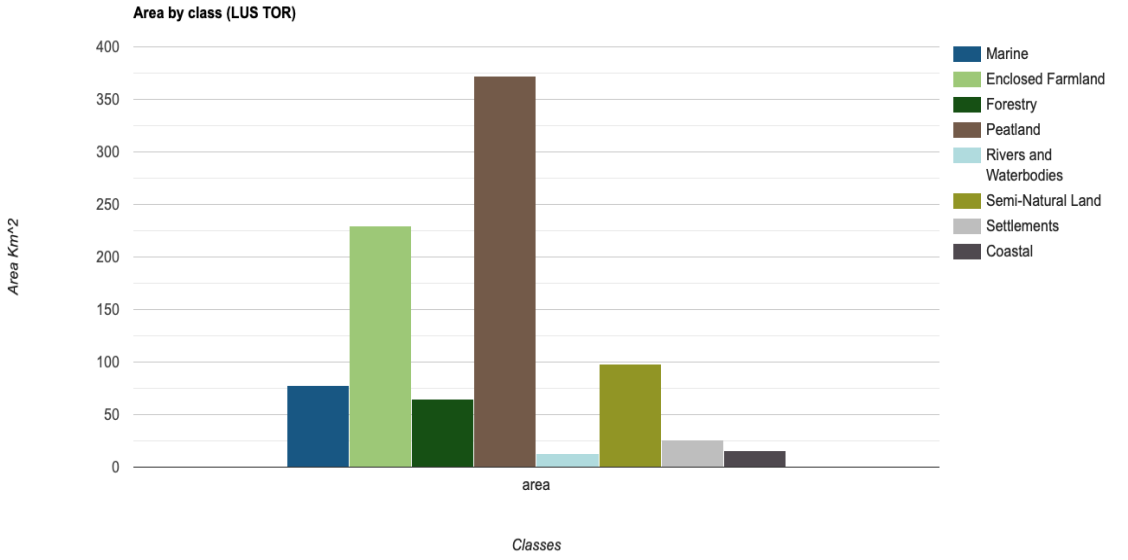

Figure 8e: LUS_{TOR} Class Area (km^2)

5 – DISCUSSION AND FUTURE DEVELOPMENTS

In this study, an evaluation of how LULC mapping could be informative for policy assessment was undertaken using Scotland's *Third Land Use Strategy 2021-2026: Getting the Best from our Land.* This policy outlines the specific targets set out by The Scottish Government to combat climate change, reverse biodiversity loss and provide a sustainable future. With reference to peatland, LUS looks to restore peatland ecosystems while balancing the uses of surrounding landscapes to meet both climate-driven and economic goals. The LUS classes were altered to specifically include 'Peatland' and 'Forestry', two important LULC classes for the attainment of Scottish climate goals. This research placed a particular emphasis on peatland delineation and the detection of features necessary for monitoring peatland restoration, a focus which needs to be explored to determine the effectiveness of LULC mapping for peatland management. The methodology used demonstrates the procedural use of cloud-computing, open-source geospatial data and machine learning algorithms for combining data sources for classification. Upon assessment of results, an evaluation of the methodology is needed, including: (1) the creation and implementation of the training data; (2) the importance of each input feature for the LUS classes; (3) the capability of RF with the outline tuning; and (4) the application of GEE for LULC policy mapping.

5.1 – Training Data: Quantity and Quality

A well-established census within machine learning research dictates that the quality and quantity of training data to be one of the most important factors in classification accuracy. This includes the determination of appropriate sample size, sampling method (e.g. random or proportional), spatial autocorrelation and class heterogeneity (Millard and Richardson, 2015; Shetty *et al.*, 2021). The increasing of training data quantity has shown to result in lower rfOOBe scores and greater classification accuracy, highlighting a potential area for improvement of the current methodology (Maxwell, Warner and Fang, 2018). Using 500 points per LUS class, 4,000 training data points were used for an area of ~892 km^2. This number was chosen in accordance to the 0.25% of the study area per class (~400 points) and also the limitations experienced within GEE, with computation timing out at 5,000 points. Undoubtedly, increasing the number of quality training points would have increased the overall classifications for each of the three models. In addition, while stratified random sampling was used to extract training data points, other sampling methods may have been appropriate. The clear dominance of predicted 'Peatland' and 'Enclosed

Farmland' within the ROI (Fig. 6e, 7e, 8e), coupled with the input data's strength delineating 'Marine', 'Forestry' and 'Rivers and Waterbodies' classes, may justify the use of proportional sampling (Shetty *et al.*, 2021).

To accompany increases in training data quantity, an assessment of training data quality demonstrates a clear need for this research. Whilst acquiring training data from existing LULC maps has been undertaken extensively (Zhang and Roy, 2017; Hermosilla *et al.*, 2018; Shetty *et al.*, 2021), it is important to note that errors from the original classification will be inherited (Abdi, 2020). Training data was sampled from the UKCEH 2019 LULC map, which had an overall accuracy of 79.4%. However, this map inherited its training data from the UKCEH 2018 LULC map (OA 79.6%), which had sampled the UKCEH 2017 LULC map (78.6%). A process the UKCEH termed 'bootstrap sampling', this will have likely instilled errors in the training data used for the three model's classification (UK Centre for Ecology and Hydrology, 2020). With the Scottish Land Use Strategy looking to make ambitious changes to the landscape in a short period of time, including drastic landcover changes (notably afforestation and forest-to-bog restoration), the use of training data from a previous year may not provide the necessary quality for training. Therefore, a need for in-situ sampling is necessary either as replacement training data or a quality control metric to determine the accuracy of the derived training data.

Furthermore, whilst training data quality must reflect the heterogeneity of individual classes, inherent similarities between LUS classes has led to observed misclassifications. The notable confusion in all three models between 'Semi-Natural Land' and 'Peatland', resulted in the lowest UA (LUS_{TO}, LUS_{TOR}) and PA (LUS_{TO}, LUS_{TR}, LUS_{TOR}) scores for each model. As the 'Peatland' class was originally derived from the 'Semi-Natural Land' class to assess dynamic peatland changes within the ROI, it is evident similarities between the classes lead to misclassification. This is likely due to the similarities of peat (>50cm) compared to peaty (<50cm) soils derived from the UKCEH 2019 map. This demonstrates a 'fine-line' between class heterogeneity and class similarity that must be clarified when undertaking peatland classification (Berhane *et al.*, 2019). In addition, the visible confusion in areas of forest-to-bog restoration (Fig. 9) reaffirms the need for accurate training data that can delineate not only peatland from other classes, but the properties, condition or stage of peatland restoration (DeLancey *et al.*, 2019; Mahdianpari *et al.*, 2019). For Scottish peatland to be monitored effectively using LULC mapping, it is clear training data must determine: (1) the type of peatland (e.g. bog, fen, marsh, mire etc.) ; (2) the condition of the peat (e.g. drained, cut, burnt); and (3) the restoration process it is/has

undergone (e.g. forest-to-bog, rewetting). These tasks undoubtedly require a greater focus on the peatland and tailored models to target peat ecosystems, moving away from a general LULC assessment such as the one outlined in this study.

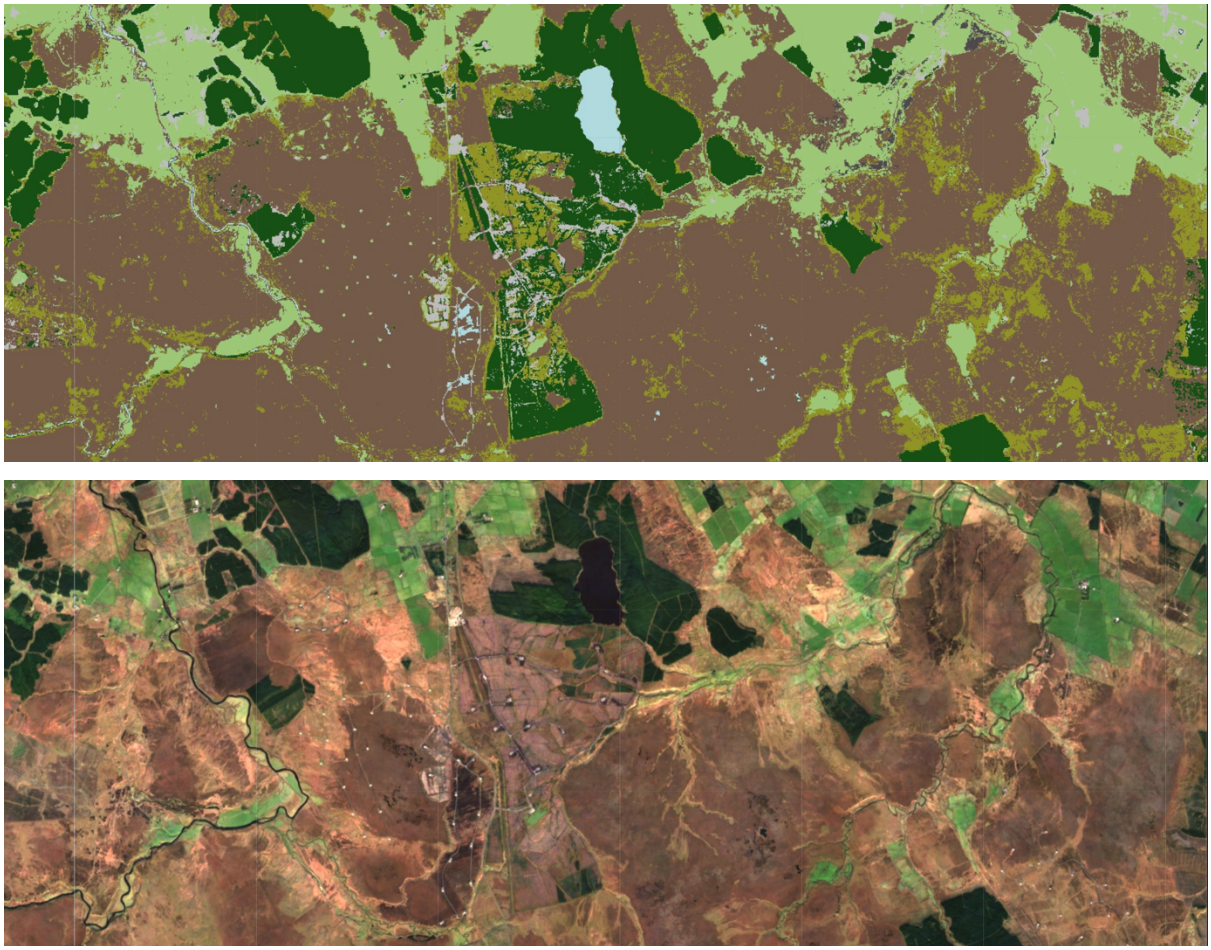

Figure 9: LUS_{TOR} Visual Inspection of Forest-to-Bog Windfarm at Camster

5.2 – Evaluating Input Features: Optical, Radar and Topographic

The strength of optical inputs in the LUS_{TO} and LUS_{TOR} models mirrors results obtained within LULC mapping and peatland mapping research (Minasny *et al.*, 2019; Chaves, Picoli and Sanches, 2020). LUS_{TO} (OA= 0.821) outperformed LUS_{TR} (OA=0.786), suggesting the superiority of optical inputs over radar, a trend identified by Mahdianpari et al. (2019). Sentinel-2 provided data at a 10m resolution, the highest spatial resolution open-source optical data available, which performed remarkably well when delineating important features (e.g. airfields, field boundaries, roads, windfarms, streams, pools)(Fig.10). S2 also provided the most derived input features, including 11 bands and 5 indices. Examining the importance of the optical bands (Fig. 6d, 7d and 8d) indicates no discernible pattern to their weight, suggesting at the possible need for a

reduction or removal of uncorrelated features for better classification. Similarly, spectral indices were dispersed throughout the variable importance graphs for both LUS_{TO} and LUS_{TOR} models, similar to results obtained by other wetland classification studies (Berhane *et al.*, 2018). However, the mNDWI index presented high variable importance scores in both LUS_{TO} and LUS_{TOR} models. The influence of this metric for inland water, wetland and LULC mapping is well established, but further work is needed to confirm its ability to detect peat-specific ecosystems (Xu, 2006; Abdi, 2020; Yang *et al.*, 2021).

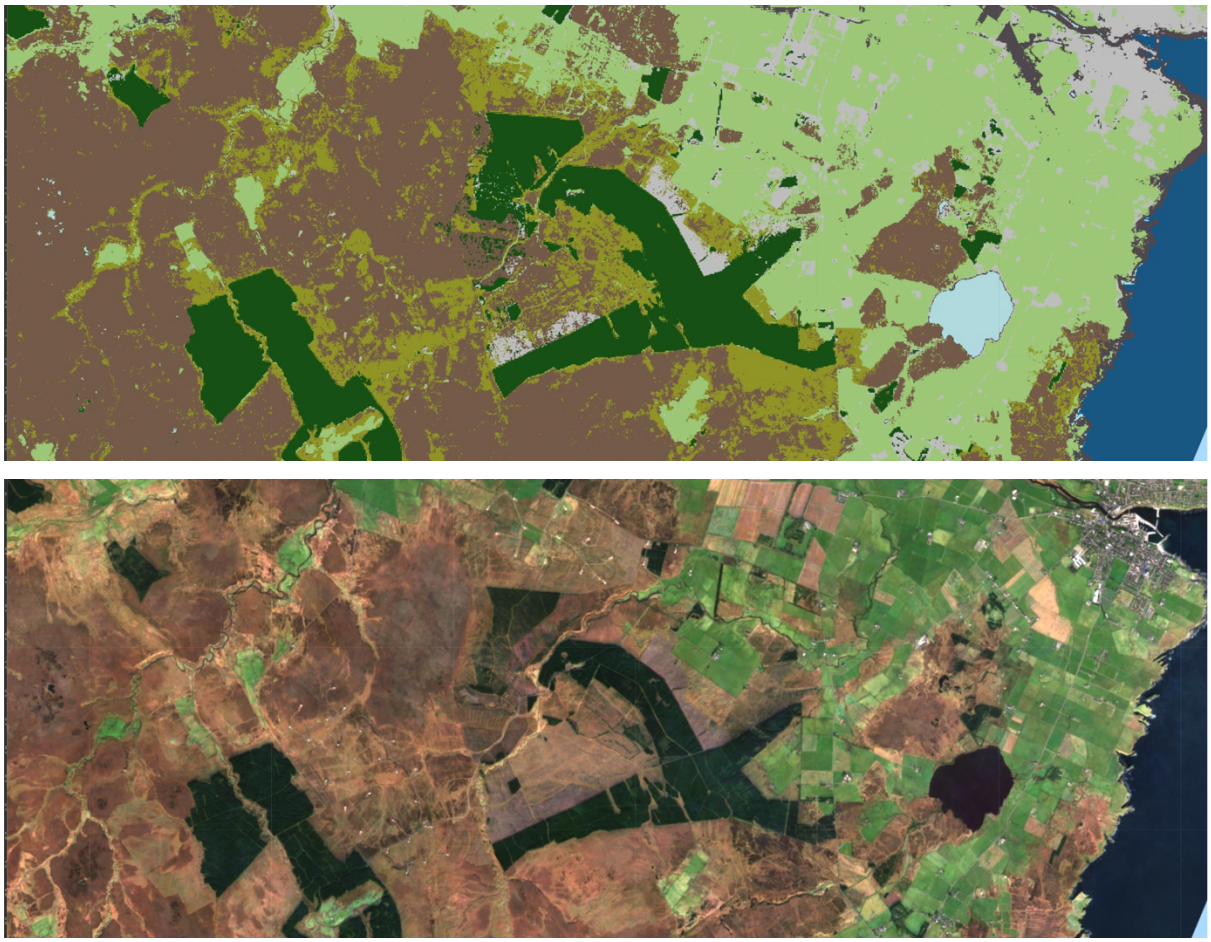

Figure 10: LUS_{TOR} Visual Inspection of Afforestation outside Wick

The inclusion of Sentinel-1 SAR data into a multi-sensor classification model marks an emerging trend within LULC mapping, exploiting the increasing availability of open-source SAR datasets (Dobrinić, Medak and Gašparović, 2020). While established in wetland mapping, the exploration of additional SAR derivatives could prove useful to increase its effectiveness when combined with optical and topographic data (Adeli *et al.*, 2020). This might include the refinement of SAR through polarisation, orbit selection (e.g. ascending or descending), instrument (IW, EW or SM), or decomposition mechanisms (Hird *et al.*, 2017). The SAR features included in both the LUS_{TR}

and LUS_{TOR} models, customisable through metadata filters within GEE, are limited however by the polarizations obtained in northern latitudes. With regards to peatland mapping, the emergence of Interferometric SAR (InSAR) techniques for the monitoring and classification of Scottish peatland point to the increasing possibilities of mapping landscapes based on surface motion and seasonal subsidence (Alshammari *et al.*, 2018, 2020). Applicable to both Sentinel-1 and -2 datasets, investigation of seasonal composites is necessary to determine whether the LUS classes can be separated with greater accuracy during drier or wetter seasons. Furthermore, the addition of textural features (e.g. the Gray Level Co-occurrence Matrices) has shown to drastically improve OA, effective in wetland classification and the delineation of vegetation surface roughness (Khatami, Mountrakis and Stehman, 2016; Berhane *et al.*, 2018; Belcore, Piras and Wozniak, 2020).

The importance of topographic features within each classifier cannot be understated, with elevation the most important variable in each of the three models (Fig. 6d, 7d and 8d). The use of the NASADEM provides a suitable base from which additional datasets can be added, however refinement of the derived features needs to be undertaken. Both TWI and TPI allow for parameters to be changed and processes to be defined to improve the performance of the derived index. Changing the TWI depression, flow accumulation and specific catchment algorithms will determine different TWI scores, whilst changing the neighbourhood size (100m) of the TPI will similarly change results (Hird *et al.*, 2017; Mattivi *et al.*, 2019). Both these processes were areas of uncertainty within this research, and further investigation into each model is needed to tune the indices and derive enhanced results. Furthermore, the inclusion of slope and aspect were consistently the least important features within each classifier, justifying their exclusion in future iterations.

Significant improvements in topographic accuracies have been observed with the substitution of SAR-based DEM's with Light Detection and Ranging (LiDAR) DEM's, particularly for mapping boreal wetland and northern peatlands (Lang *et al.*, 2013; Karlson *et al.*, 2019; Wu *et al.*, 2019). LiDAR provides finer spatial resolution for topography, wetland inundation and surface vegetation, critical for our study area with: low topographic variation; peatland ecosystems which exhibit 'bog-breathing'; and forested peatlands experiencing expansion and subsidence (Lindsay *et al.*, 1988; Lang *et al.*, 2013). Conversely, Minasny et al. (2019) have argued the superiority of LiDAR is yet to be comprehensively supported for peatland mapping, suggesting flatter areas like this area benefit from coarser resolution DEMs depicting local relief (Minasny *et al.*, 2019).

NASADEM was chosen as the topographic data source due to its improvements in accuracy when compared to ASTER GDEM and SRTM, and due to the lack of Scottish LiDAR data covering the defined region. However, as open-source LiDAR products become increasingly available through 'Open Data' policies, their application for peatland and LULC mapping can be examined and trialled (Lang *et al.*, 2013; Hird *et al.*, 2017).

5.3 – Random Forest: Performance and Optimization

The justification for using Random Forest over comparable machine learning algorithms was based on the following: (1) its ability to deal with the high dimensionality input; (2) its customisable parameters (Ntree and Mtry); (3) its comparable insensitivity to noise; and (4) its integrated variable importance metrics. Sensitive to the quality and quantity of the training data, RF still provided reasonably strong OA's (LUS_{TO}=0.821, LUS_{TR}=0.786, LUS_{TOR}=0.823) even with the highlighted limitations of the UKCEH LULC 2019 map. For each model, iterations of the classifier were run with incremental Ntree values, with an additional 10 trees added per classification starting from 10 up to 500. This determined the best number of decision trees for each model. As denoted by the research, this allows for the optimisation of each RF classifier resulting in varying classification accuracies and variable importance scores (Millard and Richardson, 2015). However, this process was only run for Ntree and did not take into consideration the Mtry value. Whilst the square root of the total number of input features is standard (LUS_{TO}=√22, LUS_{TR}=√9, LUS_{TOR}=√26), optimisation of this metric could improve classification performance. Zhang et al. (2021) defined a systematic method to tune the Mtry parameter using Bayesian optimization, which outperformed both SVM and RF with default parameters (Zhang *et al.*, 2021). This suggests the current RF Mtry methodology can be improved and accuracy can be increased.

Variable importance allows for the identification of input dependency when Ntree's are run within the RF classifier, a unique feature of RF and one that can highlight where variables are under-performing. Each model produced three contrasting variable importance charts (Fig. 6d, 7d and 8d), suggesting no identifiable pattern in most of the input features (excl. Elevation, Slope and Aspect). When run through the multiple iterations necessary to determine the Ntree parameter, these variable importance rankings varied upon each classification. This significantly limited the ability to determine feature strength and prevented feature reduction. Millard and Richardson (2015) determined variable importance to be a highly unstable metric when run

through multiple iterations, creating a McNemar test to statistically map importance over an iterative classification model. This allowed for the identification of important variables based on ranked appearances in the top five positions. This suggests variable importance can be determined effectively when multiple iterations are part of the RF classification process. Exploration of how iterations of Ntree, Mtry and variable importance can be implemented into GEE is a notable gap in the research.

Variable reduction was not undertaken for the any of the RF models to maintain consistency of input features over each model. However, research suggests the need for variable reduction to obtain the best classification scores within RF classification (Millard and Richardson, 2015). Principal Component Analysis (PCA), which transforms correlated features into un-correlated features, has been utilised within GEE to improve LULC classification and probabilistic mapping wetland ecosystems (Maxwell, Warner and Fang, 2018; DeLancey *et al.*, 2019). An assessment of its effect on the included datasets is required to determine if classification accuracy would be improved in any of the models. To summarise, RF performed well for the classification of the LUS_{TO}, LUS_{TR} and LUS_{TOR} models. There is considerable room for the improvement of the RF classifier model for this LULC mapping application, and future comparisons with other machine learning algorithms should be undertaken to statistically justify selection.

5.4 – Google Earth Engine: Application and Possibilities

Google Earth Engine has been proven an interactive and capable tool for the analysis of LULC mapping, providing seamless access to vast amounts of geospatial data, computational power and integrated algorithms. In addition, GEE allowed data to be imported and exported in various file formats (GeoTIFF and ESRI Shapefile) to complete computation in other tools. Furthermore, GEE's accompanying web-based IDE enabled JavaScript code to be organised, accessed and run instantly with no limitations. For non-expert users, GEE's accompanying documentation and data catalogue makes the identification of algorithms and examples simple, providing numerous scripts that can be amended for specific purposes. Still in its relative infancy as a GIS tool, certain capabilities required additional software (notably QGIS) for the processing of training data, topographic indices creation and visualisation, a limitation found within other LULC studies (Zhang and Zhang, 2020). In addition, regular memory issues were experienced when algorithms were run on vast amounts of data, with scaling and reducing of datasets necessary for tasks to be executed sufficiently (Gorelick *et al.*, 2017; Tamiminia *et al.*, 2020). The

definable benefit of cloud-based computing is access to a wealth of computational power, shielding the user from the resource allocation, parallelism, and data distribution. As stated by Gorelick et al. (2016): "The price of liberation from these details is that the user is unable to influence them: the system is entirely responsible for deciding how to run a computation" (Gorelick *et al.*, 2017, p. 25).

The growing body research exploiting GEE for LULC mapping, and its use for peatland mapping, highlights the capabilities of cloud-computing for supervised classification. To meet this demand, increases in the number of GEE algorithms and auxiliary datasets are needed to accommodate improvements in research deliverables (Huang *et al.*, 2017). Specifically for image classification, advances in GEE object-based analysis can allow for machine learning results better interpret and present LULC mapping products (Dronova, 2015). Object-based analysis allows for; the removal of 'salt-and-pepper' speckling of resulting classification images; the contextualisation of pixels on spectral, textural and spatial features; and tailored segmentation and clustering (Amani *et al.*, 2019; Mahdianpari *et al.*, 2019). Tamiminia et al. (2019) highlight the significant lack of object-based (2%) vs. pixel-based (98%) studies within GE due to the current lack of functionality (Tamiminia *et al.*, 2020). While options for object-based analysis within GEE remain limited, improved accuracies for LULC and wetland mapping using RF have been achieved using object-based algorithms such as Simple Non-Iterative Clustering (SNIC)(Amani *et al.*, 2019; Mahdianpari *et al.*, 2019). In addition, tile-based segmentation has shown its use in delineating spectrally and spatially heterogenous LULC classes, which might address the misclassification of confused classes in the current research (Zhang *et al.*, 2020). Although the computational complicity of the proposed methods should always be considered prior to selecting a classification method, the implementation of object-based analyses is a logical next step for this research, allowing for areas to be more accurately segmented and for features within the landscape to be better understood.

Finally, GEE's data catalogue is ever-expanding, which can provide meaningful datasets for LULC mapping and analysis. The integration of supplementary datasets such as climate, socio-economic and population density have aided the detection of peatlands, identification of land use change drivers and been incorporated into policy decisions (Connolly *et al.*, 2011; Minasny *et al.*, 2019; Liu *et al.*, 2020). In the context of The Flow Country and the Land Use Strategy, socio-economic datasets reflecting population, agricultural land capability and carbon storage could dramatically enhance the derived LULC mapping applications. For the LULC classification maps

to be effective for policy assessment, a framework within which additional data can be updated and simulated to address prospective policies and trending dynamics is desirable. Future developments of possible datasets that could supplement the existing models to improve application as policy tools should be undertaken, from which an assessment of feasibility can be assumed.

5.5 – LULC Policy Mapping: Function and Implementation

The execution of this study looked to assess how LULC models could aid the implementation and future progress of the Scottish Land Use Strategy. Outlining ambitious alterations to the structure of Scottish landscapes, the need for tools to quantify these changes and determine actions were identified, with the resulting model acting as indicators of LULC dynamics. The obtained results have calculated areas allocated to each of the eight defined LUS classes, which can in turn be used to determine the feasibility of policies and the attainability of yearly targets (Fig. 6e, 7e and 8e). For example, an assessment of Scottish Forestry objectives, looking to afforest 18,000 hectares of conifers annually by 2024, can use the obtained model to determine suitable LUS classes for transformation based on spatial proximities, available area and likely trade-offs. It has been noted that the inclusion of auxiliary data would drastically improve the classification and application of the defined LULC model, notably soil mapping, socio-economic and stored carbon datasets. From this, assessments of land capability and land suitability could be derived, determining the productivity of lands for agriculture and horticultural use.

The procedure outlined for the creation of LULC maps demonstrates a functional workflow for monitoring Scottish policy, with future versions benefitting from model refinement, supplementary data and enhanced features. It is important to note that the obtained accuracies (LUS_{TO}=0.821, LUS_{TR}=0.786, LUS_{TOR}=0.823) are currently insufficient for meaningful policy assessment, however, once the model workflow is amended to correct existing limitations, several approaches can be undertaken to improve its application. Firstly, once the classifier has been tuned to provide sufficient accuracy, classification of multiple historical composites can determine historical assessments of LULC change and provide projections based on statistical trends. This process has been used to determine failed historical LULC changes and where policy corrections were successful (Kolli *et al.*, 2020). In relation to peatland mapping, statistical modelling could amend the existing model workflow to obtain peatland occurrence probability maps (DeLancey *et al.*, 2019). Applications of LULC maps for policy mapping are wide-ranging,

yet require detailed assessment to provide meaningful results. This methodology has provided a broad example of what is capable using cloud-computing and GEE, but it is evident a narrower focus on specific LULC classes could result in great application. For mapping Scottish peatland, the spatial variability of peatland ecosystems requires particular attention, with greater focus on the unique hydrological and ecological properties needed for meaningful monitoring and management.

6 – CONCLUSION

Long term monitoring of LULC change is vital for implementing effective policy and mitigating climate change. In the Scottish Flow Country, peatland ecosystems present significant benefits such as rich biodiversity, carbon storage and water regulation. There is a need to understand these systems in a broader context, appreciating peatland within an evolving landscape. This has necessitated the outlined methodology, taking advantage of cloud-computing, 'Geo Big Data' and machine learning to automate the detection of LULC changes with high-temporal and spatial resolution. This study assessed the combination of multiple sensors for accurate LULC mapping based of the Scottish Land Use Strategy, using optical, radar and topographic input features to determine their strength within a RF classifier. Results suggest a combination of optical, radar and topographic features is best for accurate LULC mapping (LUS_{TOR} OA=0.823 and KA=0.792), particularly when delineating ecologically, hydrologically and geomorphologically challenging landscapes classes such as peatland.

In addition, the use of Random Forest has shown its capability in handling high dimensionality data and integrating heterogenous training data. However, more research is needed to determine its how increases in quality training data, enhanced tuning of hyperparameters (Ntree and Mtry) and determination of variable importance could further improve the defined LUS classification model. Furthermore, comparisons between the outlined methodology and other machine learning algorithms are needed to justify RF selection over similar supervised classification tools. The advent of cloud-computing for expert and non-expert users marks a significant opportunity in LULC mapping, with GEE providing numerous datasets, tools and functionality for geospatial analysis. With ever increasing access to new datasets (e.g. LiDAR, climate, economic) and image processing algorithms, the capability for LULC mapping outputs to be more intuitive and helpful in policy assessments is an exciting future direction of this work. Finally, this work has determined where LULC mapping can be implemented for future policy assessment, acting as a tool form which preliminary results can be determined. The methodology outlined, once improved, can provide a useful model for specific class analysis (e.g. land capability or peatland probability map) or append data to assist general investigations into LULC monitoring and management.

7 – BIBLIOGRAPHY

Abdi, A. M. (2020) 'Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data', *GIScience and Remote Sensing*, 57(1), pp. 1–20. doi: 10.1080/15481603.2019.1650447.

Adam, E., Mutanga, O. and Rugege, D. (2010) 'Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review', *Wetlands Ecology and Management*, 18(3), pp. 281–296. doi: 10.1007/s11273-009-9169-z.

Adeli, S. *et al.* (2020) 'Wetland monitoring using SAR Data: A meta-analysis and comprehensive review', *Remote Sensing*, 12(14). doi: 10.3390/rs12142190.

Alonso, A. *et al.* (2016) 'Wetland landscape spatio-temporal degradation dynamics using the new google earth engine cloud-based platform: Opportunities for non-specialists in remote sensing', *Transactions of the ASABE*, 59(5), pp. 1333–1344. doi: 10.13031/trans.59.11608.

Alonso, A., Muñoz-Carpena, R. and Kaplan, D. (2020) 'Coupling high-resolution field monitoring and MODIS for reconstructing wetland historical hydroperiod at a high temporal frequency', *Remote Sensing of Environment*, 247(May), p. 111807. doi: 10.1016/j.rse.2020.111807.

Alshammari, L. *et al.* (2018) 'Long-term peatland condition assessment via surface motion monitoring using the ISBAS DInSAR technique over the Flow Country, Scotland', *Remote Sensing*, 10(7). doi: 10.3390/rs10071103.

Alshammari, L. *et al.* (2020) 'Use of Surface Motion Characteristics Determined by InSAR to Assess Peatland Condition', *Journal of Geophysical Research: Biogeosciences*, 125(1), pp. 1–15. doi: 10.1029/2018JG004953.

Amani, M. *et al.* (2019) 'Canadian wetland inventory using Google Earth Engine: The first map and preliminary results', *Remote Sensing*, 11(7). doi: 10.3390/RS11070842.

Anderson, K. *et al.* (2010) 'Combining LiDAR and IKONOS Data for Eco-Hydrological Classification of an Ombrotrophic Peatland', *Journal of Environmental Quality*, 39(1), pp. 260–273. doi: https://doi.org/10.2134/jeq2009.0093.

Belcore, E., Piras, M. and Wozniak, E. (2020) 'Specific alpine environment land cover classification methodology: Google earth engine processing for sentinel-2 data', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 43(B3), pp. 663–670. doi: 10.5194/isprs-archives-XLIII-B3-2020-663-2020.

Belgiu, M. and Drăgu, L. (2016) 'Random forest in remote sensing: A review of applications and future directions', *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24–31. doi: 10.1016/j.isprsjprs.2016.01.011.

Berhane, T. M. *et al.* (2018) 'Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory', *Remote Sensing*, 10(4). doi: 10.3390/rs10040580.

Berhane, T. M. *et al.* (2019) 'The influence of region of interest heterogeneity on classification accuracy in wetland systems', *Remote Sensing*, 11(5). doi: 10.3390/rs11050551.

Beven, K. J. and Kirkby, M. J. (1979) 'A physically based, variable contributing area model of basin hydrology', *Hydrological Sciences Bulletin*, 24(1), pp. 43–69. doi: 10.1080/02626667909491834.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45, pp. 5–32. doi: 10.1023/A:1010950718922.

Buckley, S. M. *et al.* (2020) 'NASADEM: User Guide', *Nasa/Jpl*, (January), p. 48. Available at: https://lpdaac.usgs.gov/documents/592/NASADEM_User_Guide_V1.pdf.

Carrasco, L. *et al.* (2019) 'Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine', *Remote Sensing*, 11(3). doi: 10.3390/rs11030288.

Chambers, J. C., Allen, C. R. and Cushman, S. A. (2019) 'Operationalizing Ecological Resilience

Concepts for Managing Species and Ecosystems at Risk', *Frontiers in Ecology and Evolution*, 7(July), pp. 1–27. doi: 10.3389/fevo.2019.00241.
Chaves, M. E. D., Picoli, M. C. A. and Sanches, I. D. (2020) 'Recent applications of Landsat 8/OLI and Sentinel-2/MSI for land use and land cover mapping: A systematic review', *Remote Sensing*, 12(18). doi: 10.3390/rs12183062.
Chen, B. *et al.* (2017) 'A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform', *ISPRS Journal of Photogrammetry and Remote Sensing*, 131, pp. 104–120. doi: 10.1016/j.isprsjprs.2017.07.011.
ClimateXChange (2021) *ClimateXChange: Scotland's Centre of Expertise Connecting Climate Change Research and Policy*. Edinburgh, Scotland.
Connolly, J. *et al.* (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), pp. 2377–2393. doi: 10.1080/01431161003698310.
DeLancey, E. R. *et al.* (2019) 'Large-scale probabilistic identification of peatlands in the boreal natural region of Alberta, Canada using Google Earth Engine, open-access satellite data, and machine learning', *PLoS ONE*, 14(6), p. e0218165.
Diniz, C. *et al.* (2019) 'Brazilian mangrove status: Three decades of satellite data analysis', *Remote Sensing*, 11(7). doi: 10.3390/rs11070808.
Dobrinić, D., Medak, D. and Gašparović, M. (2020) 'Integration of multitemporal sentinel-1 and sentinel-2 imagery for land-cover classification using machine learning methods', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 43(B1), pp. 91–98. doi: 10.5194/isprs-archives-XLIII-B1-2020-91-2020.
Dóka, R., Kiss, M. and Bárány-Kevei, I. (2019) 'Land use anomalies on wetlands in different time horizons - A case study from Hungary', *Carpathian Journal of Earth and Environmental Sciences*, 14(2), pp. 287–300. doi: 10.26471/CJEES/2019/014/080.
Dronova, I. (2015) 'Object-based image analysis in wetland research: A review', *Remote Sensing*, 7(5), pp. 6380–6413. doi: 10.3390/rs70506380.
Dvorett, D., Davis, C. and Papeş, M. (2016) 'Mapping and Hydrologic Attribution of Temporary Wetlands Using Recurrent Landsat Imagery', *Wetlands*, 36(3), pp. 431–443. doi: 10.1007/s13157-016-0752-9.
European Space Agency (2021a) *Sentinel-1 Observation Scenarios: Planned Acquisitions*. Available at: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/observation-scenario (Accessed: 17 August 2021).
European Space Agency (2021b) *Sentinel-1 SAR User Guide*. Available at: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar (Accessed: 17 August 2021).
European Space Agency (2021c) *Sentinel-2-Observation Scenario—Planned Acquisitions*. Available at: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/observation-scenario (Accessed: 17 August 2021).
Farda, N. M. (2017) 'Multi-temporal Land Use Mapping of Coastal Wetlands Area using Machine Learning in Google Earth Engine', *IOP Conference Series: Earth and Environmental Science*, 98(1), pp. 0–12. doi: 10.1088/1755-1315/98/1/012042.
De Feudis, M. *et al.* (2021) 'GIS-based soil maps as tools to evaluate land capability and suitability in a coastal reclaimed area (Ravenna, northern Italy)', *International Soil and Water Conservation Research*, 9(2), pp. 167–179. doi: 10.1016/j.iswcr.2020.11.007.
Gallant, A. L. (2015) 'The challenges of remote monitoring of wetlands', *Remote Sensing*, 7(8), pp. 10938–10950. doi: 10.3390/rs70810938.
Gallant, J. C. and Wilson, J. P. (2000) 'Primary Topographic Attributes', *Terrain Analysis: Principles and Applications*, (October), pp. 51–85.
Gallego-Sala, A. V. *et al.* (2016) 'Climate-driven expansion of blanket bogs in Britain during the Holocene', *Climate of the Past*, 12(1), pp. 129–136. doi: 10.5194/cp-12-129-2016.

Gerard, F. *et al.* (2003) 'Forest fire scar detection in the boreal forest with multitemporal SPOT-VEGETATION data', *IEEE Transactions on Geoscience and Remote Sensing*, 41(11), pp. 2575–2585. doi: doi:10.1109/tgrs.2003.819190.

Gorelick, N. *et al.* (2017) 'Google Earth Engine: Planetary-scale geospatial analysis for everyone', *Remote Sensing of Environment*, 202(2016), pp. 18–27. doi: 10.1016/j.rse.2017.06.031.

Gorham, E. (1991) 'Northern Peatlands : Role in the Carbon Cycle and Probable Responses to Climatic Warming Author (s): Eville Gorham Published by : Wiley on behalf of the Ecological Society of America Stable URL : http://www.jstor.org/stable/1941811 JSTOR is a not-for-pr', 1(2), pp. 182–195.

Grace, J. (2004) 'Understanding and managing the global carbon cycle', *Journal of Ecology*, 92(2), pp. 189–202. doi: 10.1111/j.0022-0477.2004.00874.x.

Hambley, G. *et al.* (2018) 'Net ecosystem exchange from two formerly afforested peatlands undergoing restoration in the flow country of northern Scotland', *Mires and Peat*, 23, pp. 1–14. doi: 10.19189/MaP.2018.DW.346.

Hancock, M. H. *et al.* (2018) 'Vegetation response to restoration management of a blanket bog damaged by drainage and afforestation', *Applied Vegetation Science*, 21(2), pp. 167–178. doi: 10.1111/avsc.12367.

Harris, A. and Bryant, R. G. (2009) 'A multi-scale remote sensing approach for monitoring northern peatland hydrology: Present possibilities and future challenges', *Journal of Environmental Management*, 90(7), pp. 2178–2188. doi: https://doi.org/10.1016/j.jenvman.2007.06.025.

Hermosilla, T. *et al.* (2018) 'Disturbance-Informed Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series', *Canadian Journal of Remote Sensing*, 44(1), pp. 67–87. doi: 10.1080/07038992.2018.1437719.

Hird, J. N. *et al.* (2017) 'Google earth engine, open-access satellite data, and machine learning in support of large-area probabilisticwetland mapping', *Remote Sensing*, 9(12). doi: 10.3390/rs9121315.

Huang, C. *et al.* (2020) 'Land cover mapping in cloud-prone tropical areas using Sentinel-2 data: Integrating spectral features with Ndvi temporal dynamics', *Remote Sensing*, 12(7). doi: 10.3390/rs12071163.

Huang, H. *et al.* (2017) 'Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine', *Remote Sensing of Environment*, 202, pp. 166–176. doi: 10.1016/j.rse.2017.02.021.

Huete, A. *et al.* (2002) 'Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices', *Remote Sensing of Environment*, 83, pp. 195–213. doi: 10.1016/S0034-4257(02)00096-2.

IUCN UK Peatland Programme (2017) *Peatland Code (Version 1.1 - March 2017)*. Available at: https://www.iucn-uk-peatlandprogramme.org/funding-finance/introduction-peatland-code.

Jackson, D. L. (2000) *Guidance on the interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater types): Definitions and the relationship with other classifications*. Available at: http://www.jncc.gov.uk/page-2433 accessed 2020-06-01.

Jackson, R. B. *et al.* (2017) 'The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls', *Annual Review of Ecology, Evolution, and Systematics*, 48(1), pp. 419–445. doi: 10.1146/annurev-ecolsys-112414-054234.

Joosten, H. and Clarke, D. (2002) *Wise use of mires and peatlands -Background and principles including a framework for decision-making*, *International Mire Conservation Group and International Peat Society*. Available at: http://www.gret-perg.ulaval.ca/fileadmin/fichiers/fichiersGRET/pdf/Doc_generale/WUMP_Wise_Use_of_Mires_and_Peatlands_book.pdf.

Karlson, M. *et al.* (2019) 'Delineating northern peatlands using Sentinel-1 time series and terrain indices from local and regional digital elevation models', *Remote Sensing of Environment*,

231(August 2018), p. 111252. doi: 10.1016/j.rse.2019.111252.
Khatami, R., Mountrakis, G. and Stehman, S. V. (2016) 'A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research', *Remote Sensing of Environment*, 177(May), pp. 89–100. doi: 10.1016/j.rse.2016.02.028.
Köchy, M., Hiederer, R. and Freibauer, A. (2015) 'Global distribution of soil organic carbon – Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world', *SOIL*, 1(1), pp. 351–365. doi: 10.5194/soil-1-351-2015.
Kolli, M. K. *et al.* (2020) 'Mapping of major land-use changes in the Kolleru Lake freshwater ecosystem by using landsat satellite images in google earth engine', *Water (Switzerland)*, 12(9), pp. 1–15. doi: 10.3390/w12092493.
Krankina, O. N. *et al.* (2008) 'Meeting the challenge of mapping peatlands with remotely sensed data', *Biogeosciences*, 5(6), pp. 1809–1820. doi: 10.5194/bg-5-1809-2008.
Kumar, L. and Mutanga, O. (2018) 'Google Earth Engine applications since inception: Usage, trends, and potential', *Remote Sensing*, 10(10), pp. 1–15. doi: 10.3390/rs10101509.
Laney, D. (2001) 'META Delta', *Application Delivery Strategies*, 949(February 2001), p. 4. Available at: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.
Lang, M. *et al.* (2013) 'Topographic metrics for improved mapping of forested wetlands', *Wetlands*, 33(1), pp. 141–155. doi: 10.1007/s13157-012-0359-8.
Lees, K. J. *et al.* (2019) 'A model of gross primary productivity based on satellite data suggests formerly afforested peatlands undergoing restoration regain full photosynthesis capacity after five to ten years', *Journal of Environmental Management*, 246(February), pp. 594–604. doi: 10.1016/j.jenvman.2019.03.040.
Lees, K. J. *et al.* (2020) 'Using Spectral Indices to Estimate Water Content and GPP in <italic>Sphagnum</italic> Moss and Other Peatland Vegetation', *IEEE Transactions on Geoscience and Remote Sensing*, 58(7), pp. 4547–4557. doi: 10.1109/TGRS.2019.2961479.
Lees, K. J. *et al.* (2021) 'Using remote sensing to assess peatland resilience by estimating soil surface moisture and drought recovery', *Science of the Total Environment*, 761, p. 143312. doi: 10.1016/j.scitotenv.2020.143312.
Leinenkugel, P. *et al.* (2019) 'The potential of open geodata for automated large-scale land use and land cover classification', *Remote Sensing*, 11(19). doi: 10.3390/rs11192249.
Letendre, J., Poulin, M. and Rochefort, L. (2008) 'Sensitivity of spectral indices to CO2 fluxes for several plant communities in a Sphagnum-dominated peatland', *Canadian Journal of Remote Sensing*, 34(November), pp. S414–S425. doi: 10.5589/m08-053.
Li, J. and Chen, W. (2005) 'A rule-based method for mapping Canada's wetlands using optical, radar and DEM data', *International Journal of Remote Sensing*, 26(22), pp. 5051–5069. doi: 10.1080/01431160500166516.
Li, W. *et al.* (2020) 'Monitoring and landscape dynamic analysis of alpine wetland area based on multiple algorithms: A case study of Zoige plateau', *Sensors (Switzerland)*, 20(24), pp. 1–19. doi: 10.3390/s20247315.
Lilly, A. *et al.* (2015) 'MAPPING SCOTLAND ' S SOIL RESOURCES Allan Lilly , David Miller , Willie Towers , David Donnelly , Laura Poggio and Pat Carnegie', 48(Figure 2), pp. 35–46.
Lindsay, R. A. *et al.* (1988) 'Part I Peatland Ecology', *The Flow Country - The peatlands of Caithness and Sutherland*, pp. 9–32.
Liu, C. *et al.* (2018) 'Improving large-scale moso bamboo mapping based on dense Landsat time series and auxiliary data: a case study in Fujian Province, China', *Remote Sensing Letters*, 9(1), pp. 1–10. doi: 10.1080/2150704X.2017.1378454.
Liu, C. *et al.* (2020) 'Land use/land cover changes and their driving factors in the northeastern tibetan plateau based on geographical detectors and google earth engine: A case study in gannan prefecture', *Remote Sensing*, 12(19). doi: 10.3390/RS12193139.

Ma, Y. *et al.* (2015) 'Remote sensing big data computing: Challenges and opportunities', *Future Generation Computer Systems*, 51, pp. 47–60. doi: https://doi.org/10.1016/j.future.2014.10.029.
Mahdavi, S. *et al.* (2018) 'Remote sensing for wetland classification: a comprehensive review', *GIScience and Remote Sensing*, 55(5), pp. 623–658. doi: 10.1080/15481603.2017.1419602.
Mahdianpari, M. *et al.* (2017) 'Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery', *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, pp. 13–31. doi: https://doi.org/10.1016/j.isprsjprs.2017.05.010.
Mahdianpari, M. *et al.* (2019) 'The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the Google Earth Engine cloud computing platform', *Remote Sensing*, 11(1). doi: 10.3390/rs11010043.
Mahdianpari, M., Salehi, B., *et al.* (2020) 'Big Data for a Big Country: The First Generation of Canadian Wetland Inventory Map at a Spatial Resolution of 10-m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform', *Canadian Journal of Remote Sensing*, 46(1), pp. 15–33. doi: 10.1080/07038992.2019.1711366.
Mahdianpari, M., Brisco, B., *et al.* (2020) 'The Second Generation Canadian Wetland Inventory Map at 10 Meters Resolution Using Google Earth Engine', *Canadian Journal of Remote Sensing*, 46(3), pp. 360–375. doi: 10.1080/07038992.2020.1802584.
Mattivi, P. *et al.* (2019) 'TWI computation: a comparison of different open source GISs', *Open Geospatial Data, Software and Standards*, 4(1). doi: 10.1186/s40965-019-0066-y.
Maxwell, A. E., Warner, T. A. and Fang, F. (2018) 'Implementation of machine-learning classification in remote sensing: An applied review', *International Journal of Remote Sensing*, 39(9), pp. 2784–2817. doi: 10.1080/01431161.2018.1433343.
Millard, K. and Richardson, M. (2015) 'On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping', *Remote Sensing*, 7(7), pp. 8489–8515. doi: 10.3390/rs70708489.
Minasny, B. *et al.* (2019) 'Digital mapping of peatlands – A critical review', *Earth-Science Reviews*, 196(January). doi: 10.1016/j.earscirev.2019.05.014.
Mitsch, W. J. and Gossilink, J. G. (2000) 'The value of wetlands: Importance of scale and landscape setting', *Ecological Economics*, 35(1), pp. 25–33. doi: 10.1016/S0921-8009(00)00165-8.
Moreira, A. *et al.* (2013) 'SAR-Tutorial-March-2013', *ieee Geoscience and remote sensinG maGazine*, (march).
Munyati, C. (2000) 'Wetland change detection on the Kafue Flats, Zambia, by classification of a multitemporal remote sensing image dataset', *International Journal of Remote Sensing*, 21(9), pp. 1787–1806. doi: 10.1080/014311600209742.
Page, S. E. and Baird, A. J. (2016) 'Peatlands and Global Change: Response and Resilience', *Annual Review of Environment and Resources*, 41, pp. 35–57. doi: 10.1146/annurev-environ-110615-085520.
Pan, Y. *et al.* (2011) 'A large and persistent carbon sink in the world's forests', *Science*, 333(6045), pp. 988–993. doi: 10.1126/science.1201609.
Pimm, S. L. *et al.* (2019) 'Measuring resilience is essential to understand it', *Nature Sustainability*, 2(10), pp. 895–897. doi: 10.1038/s41893-019-0399-7.
Pimple, U. *et al.* (2018) 'Google Earth Engine Based Three Decadal Landsat Imagery Analysis for Mapping of Mangrove Forests and Its Surroundings in the Trat Province of Thailand', *Journal of Computer and Communications*, 06(01), pp. 247–264. doi: 10.4236/jcc.2018.61025.
Poggio, L., Lassauce, A. and Gimona, A. (2019) 'Modelling the extent of northern peat soil and its uncertainty with Sentinel: Scotland as example of highly cloudy region', *Geoderma*, 346(March), pp. 63–74. doi: 10.1016/j.geoderma.2019.03.017.
QGIS. (2021) 'QGIS Geographic Information System'. QGIS Association. Available at: http://www.qgis.org.
De Reu, J. *et al.* (2013) 'Application of the topographic position index to heterogeneous landscapes', *Geomorphology*, 186, pp. 39–49. doi: 10.1016/j.geomorph.2012.12.015.

Rodriguez-Galiano, V. F. *et al.* (2012) 'An assessment of the effectiveness of a random forest classifier for land-cover classification', *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, pp. 93–104. doi: https://doi.org/10.1016/j.isprsjprs.2011.11.002.
Rondeaux, G., Steven, M. and Baret, F. (1996) 'Optimization of soil-adjusted vegetation indices', *Remote Sensing of Environment*, 55(2), pp. 95–107. doi: 10.1016/0034-4257(95)00186-7.
Rouse, J. . *et al.* (1974) 'Monitoring vegetation systems in the Great Plains with ERTS', *NASA special publication*, 351 (1974), p. 309.
Schultz, M. *et al.* (2016) 'Performance of vegetation indices from Landsat time series in deforestation monitoring', *International Journal of Applied Earth Observation and Geoinformation*, 52(May 2012), pp. 318–327. doi: 10.1016/j.jag.2016.06.020.
Scottish Government (2020) 'Update to the Climate Change Plan 2018 – 2032: Securing a Green Recovery on a Path to Net Zero', *Scottish Government*, p. 255.
Scottish Government (2021) 'Scotland ' s Third Land Use Strategy Getting the best from our land', (March).
Scottish Natural Heritage (2015a) *Scotland's National Peatland Plan: Working for our Future*. Edinburgh, Scotland.
Scottish Natural Heritage (2015b) 'Scotland's National Peatland Plan', p. 52.
Shetty, S. *et al.* (2021) 'Assessing the effect of training sampling design on the performance of machine learning classifiers for land cover mapping using multi-temporal remote sensing data and google earth engine', *Remote Sensing*, 13(8). doi: 10.3390/rs13081433.
Sloan, T. J. *et al.* (2018) 'Peatland afforestation in the UK and consequences for carbon storage', *Mires and Peat*, 23, pp. 1–17. doi: 10.19189/MaP.2017.OMB.315.
Tamiminia, H. *et al.* (2020) 'Google Earth Engine for geo-big data applications: A meta-analysis and systematic review', *ISPRS Journal of Photogrammetry and Remote Sensing*, 164(May), pp. 152–170. doi: 10.1016/j.isprsjprs.2020.04.001.
Thanh Noi, P. and Kappas, M. (2017) 'Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery', *Sensors (Basel, Switzerland)*, 18(1). doi: 10.3390/s18010018.
The James Hutton Institute (2021) *Scotland's Soil Data*. Available at: https://www.hutton.ac.uk/learning/natural-resource-datasets/soilshutton/soils-maps-scotland/download (Accessed: 7 August 2021).
Tipping, R. (2008) 'Blanket peat in the Scottish Highlands: Timing, cause, spread and the myth of environmental determinism', *Biodiversity and Conservation*, 17. doi: 10.1007/s10531-007-9220-4.
UK Centre for Ecology and Hydrology (2020) *The UKCEH Land Cover Maps for 2017, 2018 and 2019*. Bailrigg, England. Available at: https://www.ceh.ac.uk/ukceh-land-cover-maps (Accessed: 20 July 2021).
Wallage, Z. E., Holden, J. and McDonald, A. T. (2006) 'Drain blocking: An effective treatment for reducing dissolved organic carbon loss and water discolouration in a drained peatland', *Science of the Total Environment*, 367(2–3), pp. 811–821. doi: 10.1016/j.scitotenv.2006.02.010.
Warren, C. (2000) '"Birds, bogs and forestry" revisited: The significance of the flow country controversy', *Scottish Geographical Journal*, 116(4), pp. 315–337. doi: 10.1080/00369220018737103.
Whyte, A., Ferentinos, K. P. and Petropoulos, G. P. (2018) 'A new synergistic approach for monitoring wetlands using Sentinels -1 and 2 data with object-based machine learning algorithms', *Environmental Modelling & Software*, 104, pp. 40–54. doi: https://doi.org/10.1016/j.envsoft.2018.01.023.
Wu, Q. *et al.* (2019) 'Integrating LiDAR data and multi-temporal aerial imagery to map wetland inundation dynamics using Google Earth Engine', *Remote Sensing of Environment*, 228(March), pp. 1–13. doi: 10.1016/j.rse.2019.04.015.
Wu, Q. (2020) 'geemap: A Python package for interactive mapping with Google Earth Engine',

Journal of Open Source Software, 5(51), p. 2305. doi: 10.21105/joss.02305.
Xu, H. (2006) 'Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery', *International Journal of Remote Sensing*, 27(14), pp. 3025–3033. doi: 10.1080/01431160600589179.
Yang, Y. *et al.* (2021) 'The spatial distribution and expansion of subsided wetlands induced by underground coal mining in eastern China', *Environmental Earth Sciences*, 80(3), pp. 1–14. doi: 10.1007/s12665-021-09422-y.
Yu, Z. *et al.* (2010) 'Global peatland dynamics since the Last Glacial Maximum', *Geophysical Research Letters*, 37(13). doi: 10.1029/2010GL043584.
Zha, Y., Gao, J. and Ni, S. (2003) 'Use of normalized difference built-up index in automatically mapping urban areas from TM imagery', *International Journal of Remote Sensing*, 24(3), pp. 583–594. doi: 10.1080/01431160304987.
Zhang, D. D. and Zhang, L. (2020) 'Land cover change in the central region of the lower yangtze river based on landsat imagery and the google earth engine: A case study in Nanjing, China', *Sensors (Switzerland)*, 20(7), pp. 1–20. doi: 10.3390/s20072091.
Zhang, H. and Roy, D. (2017) 'Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)', *Remote Sensing of Environment*, 197, pp. 15–34. doi: 10.1016/j.rse.2017.05.024.
Zhang, M. *et al.* (2019) 'Mapping bamboo with regional phenological characteristics derived from dense Landsat time series using Google Earth Engine', *International Journal of Remote Sensing*, 40(24), pp. 9541–9555. doi: 10.1080/01431161.2019.1633702.
Zhang, M. *et al.* (2020) 'Automatic high-resolution land cover production in madagascar using sentinel-2 time series, tile-based image classification and google earth engine', *Remote Sensing*, 12(21), pp. 1–15. doi: 10.3390/rs12213663.
Zhang, T. *et al.* (2021) 'Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier', *Applied Sciences (Switzerland)*, 11(2), pp. 1–17. doi: 10.3390/app11020543.

Data Management Plan

Version: 1.1

Friday 19th August 2021

Version History:

Version Number	Date	Author	Description of Changes	Approved By
1.0	7th May 2021	NS	First Draft	-
1.1	19th August 2021	NS	Amendments	-

Project Information

Project Name: MRes in Geospatial Data Science: Classification of Flow Country Peatland: Multi-Temporal, Multi Sensor Land Use Land Cover Classification Land using Google Earth Engine (GEE)

Grant Reference: EP/S023577/1

Start and End Dates: September 2020 – September 2021

Principal Investigator: Neil Sutherland
neil.sutherland@nottingham.ac.uk

Organisation

Funder / Institution(s): EPSRC Geospatial Systems CDT / University of Nottingham, University of Newcastle

Any Other Team Members: David Large (Department of Chemical and Environmental Engineering)
Stuart Marsh (Nottingham Geospatial Institute)

Introduction and Scope

The purpose of this data management plan (DMP) is to set up a coherent approach to data issues pertaining to the aforementioned MRes in Geospatial Data Science. This document shall act as a record of data management requirements going forward and outlines a structure for regular amendments as the project develops. The data management objectives are to ensure that:

- A framework for detailed explanation of input and generated data is outlined
- Data production, archiving, and quality processes are understood and documented
- Resulting code and output data can be validated and rigorously tested
- Datasets are made available to users where and when appropriate for further use
- Iterations of the DMP can build upon previous versions as the project develops

Roles and Responsibilities

The data created and shared within this project will be implemented, managed and updated by the principal investigator as specified above. The principal investigator will be responsible for making data available when necessary and organising future uses of the resulting products. The execution of these actions shall seek consultation from both supervisors listed and advice shall be reflected in updated versions of the DMP.

Data Generation Activities

The project outlined uses Google Earth Engine, a cloud-computing platform for scientific analysis and valuation of geospatial data. Within this platform, numerous geospatial datasets are provided in open access formats, including LANDSAT and Copernicus Sentinel satellite imagery. These datasets are stored in Google Cloud as part of the Google Cloud public data program. GEE also supports the upload of proprietary / auxillary data for analysis in various file formats (including GeoTIFF and Shapefiles). In addition, the results of any analysis (source / algorithmic) code created through the GEE API is owned by the user and can be shared appropriately. Result or output data can be freely downloaded and shared through open-access standards. The following table outlines the current datasets to be used within GEE:

Datasets

Name	Research Data Type / Classification	Data Type / Format	Date Generated	Owner	Data Size	Re-Use Plan	Preservation Plan
Sentinel-2 MultiSpectral Imagery	Secondary / Derived-Complied	GeoTIFF	19th August 2021	European Space Agency (ESA)	N/A	Open-source and stored within GEE Cloud Storage	N/A
Sentinel-1 Synthetic Aperture Radar	Secondary / Derived-Complied	GeoTIFF	19th August 2021	European Space Agency (ESA)	N/A	Open-source and stored within GEE Cloud Storage	N/A
NASADEM	Secondary / Derived-Complied	GeoTIFF	19th August 2021	NASA	N/A	Open-source and stored within GEE Cloud Storage. Additional processing completed on local server and stored in designated hard drive	Open-source data to be stored on personal computer in accordance to open-license

In addition, training data has been derived by several open-source soil maps, specified below:

Name	Research Data Type / Classification	Data Type / Format	Date Generated	Owner	Data Size	Re-Use Plan	Preservation Plan
Carbon and Peatland map 2016	Secondary / Derived-Complied	ESRI Shapefile	19th August 2021	The James Hutton Institute	7.5MB	Open-source and stored within GEE Cloud Storage	Open-source data to be stored on personal computer in accordance to open-license
UKCEH Landcover map 2017, 2018 and 2019	Secondary / Derived-Complied	ESRI Shapefile	19th August 2021	UK Centre for Ecology and	N/A	Open-source and stored within GEE Cloud Storage	N/A

				Hydrology			
NASADEM	Secondary / Derived-Complied	GeoTIFF	19th August 2021	NASA	N/A	Open-source and stored within GEE Cloud Storage. Additional processing completed on local server and stored in designated hard drive	Open-source data to be stored on personal computer in accordance to open-license

Data shall be collected and created using the following approaches:

- Software Code - This will be created primarily using the Google Earth Engine Code Editor, utilising the JavaScript programming language.
- Data can be exported from GEE into various formats (.shp or GeoTIFF), which were both utilised

In-Project Data Management Approach

A complete dataset catalogue shall be recorded and amended regularly throughout the Research Data lifecycle. This will be the gateway to all project data and metadata, and to all relevant information links. This shall be stored within a designated University of Nottingham or funder-specific data centre upon completion of the project. The data centre will be the primary source of reference regarding the data archive and will be updated as new information is available.

Metadata and Documentation

The ISO 19115-2:2019 Geographic Information metadata standards shall be adopted for the following project. This document outlines the necessary information required for each dataset, which shall be stored in a .txt or .csv file to encourage open-access. The metadata shall look to include:

- The research questions, project outline and designated aims of the research
- Any environmental conditions relevant to the data including positioning and timing
- The instruments used to obtain the data, any specification, calibration and error details
- The methodology undertaken in sufficient detail for replication of processes
- A data dictionary to define variables including units, formats, definitions, abbreviations and codes
- PI contact information

Data Quality Assurances

To determine data quality, two processes shall be carried out on any generated data. Firstly, a 'checklist' to determine confidence within the datasets used. Secondly, an 'audit trail' will determine the processes undertaken with the data to determine any potential problems and document the results of an executed checks. These shall both be appended to this DMP and featured within the metadata for each dataset.

Ethics and Legal Compliance

Currently no ethical issues exist regarding personal or commercial data. In the event personal information is required, data shall be stored in accordance with GDPR / Data Protection Act 2018. While GEE is open-source software, all software used from proprietary sources shall hold appropriate licenses and guidelines.

Storage and Backup

The University of Nottingham provides 5TB of free-at-point-of-use Research Data storage for postgraduate students through Microsoft OneDrive. This storage shall be sufficient for the outlined project, as data is largely stored within the Google Cloud Storage platform and therefore Microsoft OneDrive shall be able to provide adequate storage for personal primary and secondary datasets.

In addition, several backups shall be taken during the project's lifecycle, following the recommended 3-point backup guidelines. This shall include two separate hard-drive back-ups each with 1TB of storage, one of which shall provide automatic daily backups, the other providing weekly backups. This proposed storage and backup plan shall safeguard data against hardware and software failures. Access shall be restricted, and password protected to maintain file-naming conventions and structure within the DDS.

Data Preservation, Access and Publication

The aim of this project is that data created data will provide long-term value, and therefore actions must be outlined to determine the preservation of data throughout the defined project. This DMP is focussed on the FAIR principles for Research Data, encouraging transparency of data practices through the Research Data lifecycle. This requires ease of access with persistent identifiers; links to necessary metadata; appropriate formats for open data access and exchange; and licensing covering future use.

The data will be made available at the end of the funded project period in September 2021. Data shall be stored in a suitable data repository, either through the University of Nottingham or a funder-specific data centre. This process shall also assess what can be shared / restricted, how this data shall be stored (format requirements) and a determination of timescale. To document the software code acquired through GEE, GitHub shall be used as a repository to document updates and versions created throughout the project. Furthermore, this will allow for the sharing of GEE source code and processes for other researchers to explore.

Version Control

The DMP is a dynamic process and is expected to be regularly amended during the Research Data lifecycle. This is a first amendment of the DMP, with additional versions being stored in the DDS for reference and documented for alterations. The next amendment shall be on submission of the project, after feedback has been received on the project and this accompanying DMP.

Gantt Chart

Month	April				May				June					July				August		
Week (of year)	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Identify Research Area																				
Clarification and Refinement																				
Initial Literature Review																				
Formulate Research Questions, Aims and Objectives																				
Formulate Research Strategy, Design and Methodology																				
Research Proposal																				
Data Management, Archiving and Storage																				
Extended Literature Review																				
Data Sourcing																				
GEE - Learning, Trials and Familiarisation																				
GEE Spectral Indices creation and testing																				
GEE Image Mosaicking, Compositing and Quality Checking																				
GEE Land-Cover Classification Supervised Learning Creation and Testing																				
GEE Classification Testing and Statistical Analysis of Output																				
Validation and Accuracy Assessment																				
Writing																				
Drafting (Multiple)																				
Formatting and Condensing (PDF)																				