



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

**High-resolution remote sensing data and
machine learning approaches for mapping
and monitoring habitats**

Charmaine Cruz

A thesis submitted to Trinity College Dublin,
the University of Dublin, in candidature for the degree of
Doctor of Philosophy

Department of Geography,
School of Natural Sciences

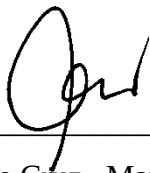
2024

Declaration

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

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

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

A handwritten signature in black ink, appearing to read 'Charmaine Cruz', is written over a horizontal line.

Charmaine Cruz, May 2024

Acknowledgements

Undertaking a PhD and navigating life abroad when everyone was trying to survive a global pandemic was tough. But throughout these challenging years, I have found the kindness of individuals who offered me their unwavering support and motivation.

I am immensely grateful to my supervisor, Dr John Connolly and co-supervisor, the late Dr Kevin McGuinness. Thank you for offering me this single iHabiMap PhD opportunity, always believing in my abilities and providing incredible support throughout my PhD journey. My very great thanks to other iHabiMap project experts, Dr Jerome O’Connell, Dr Philip Perrin and Dr James Martin, for generously sharing their extensive knowledge and expertise. Sincere thanks also to the iHabiMap Steering Committee for their valuable feedback. To my PhD thesis examiners, Prof Iris Möller and Dr Conor Cahalane, many thanks for taking the time to thoroughly discuss my thesis with me and providing insightful suggestions to improve it.

I want to thank EPA Ireland and E3 scholarships for funding my PhD research. Additional sponsorships for conference attendance, training, open-access publication fees and commercial satellite image access came from the TCD Department of Geography, ISPRS Foundation, Inc., IReL Consortium and the European Space Agency's Network of Resources sponsorship programme.

I am grateful to all the staff and fellow postgraduate students in the Department of Geography, especially Prof Iris Möller, for their support during my stay at Trinity College Dublin. A special mention goes to Orla for her constant motivation and empowerment. Many thanks to Dr Rory Hodd and Wahaj for their assistance with the field surveys. I am also thankful to my Dublin City University and Insight Centre colleagues for their support during the settling-in period and beyond.

Sincere thanks to the Filipino community in Ireland, particularly to the Rosin family, who embraced me as one of their own. To my Filipino friends who always have their empathetic ears ready when sadness and frustrations are eating me. Arnie and Michael, thank you for those small breaks during our library days.

To my support groups—ISPRS Student Consortium, Ladies of Landsat, Women in Geospatial, Sisters of SAR and SOC Youth Cluster—your “online” presence, especially during those pandemic lockdowns, has provided me with a much-needed source of social connection.

To Michael: I cannot thank you enough for helping me stay sane in the midst of insanity and uncertainties. Your company, understanding, constant encouragement and support have made my PhD years more bearable. We made it!

To my family and relatives—Papa Ariel, Mama Malou, sister Shey and brother Gah—who have always been my source of inspiration and strength, thank you for your understanding whenever I had to miss Christmas and other important family events. I have always been grateful to my aunt, Mamilen, for her constant support towards my education and aspirations in life. Your selfless love continues to inspire me as I reflect on my achievements. I hope I can make you all proud!

So, here is my PhD thesis. I am proud to have made it to this point. This PhD journey has been one of the best moments in my life.

Summary

The increasing decline in the status of habitats, mainly due to anthropogenic stressors, has spurred the development and implementation of many conservation-related legislation. This legislation involves mapping, a critical task that aims to gather important information about the habitats, including their locations, spatial extents and changes over time. In Ireland, mapping habitats is often based on field surveys. However, the methods of these surveys can be time-consuming and resource-intensive, especially if they need to be done frequently over large or remote areas. Advancements in mapping technologies, such as remote sensing and machine learning, can be integrated with field surveys to improve the efficiency and effectiveness of mapping habitats. However, they have not been well-examined in the context of mapping habitats in Ireland protected under the European Union's Habitats Directive. As a result, the potential of these mapping technologies for assessing Irish habitats remains underexplored. Motivated by this, the research has explored the use of remote sensing data and machine learning techniques in developing robust and repeatable methodologies for assessing Irish habitats. While the research was focused on mapping coastal and upland habitats at selected protected sites in Ireland, the developed methodologies could broadly apply to other habitat types or target plant species globally.

The initial task was the assessment of the effectiveness of remote sensing data acquired by Unoccupied Aerial Vehicles (UAVs) and machine learning as tools for accurate and detailed mapping of highly dynamic and fine-scale mosaics of habitats in a coastal dune environment (Chapter 2)¹. UAV images are characterised by high spatial resolution (centimetre-level), providing detailed habitat information. These images can also be flexibly acquired at specific periods of the year, aligning to key phenological stages of vegetation communities when spectral discrimination is best, for example, during the flowering or senescence period. In this study, UAV imagery and field data were acquired in a dune site in Co. Kerry during the growing season in 2020: 26 May (early), 28 July (mid), and 15 October (late). Topographic data representing the terrain of the study site were also generated during the photogrammetric processing of UAV images. These

¹ Cruz, C., O'Connell, J., McGuinness, K., Martin, J. R., Perrin, P. M., & Connolly, J. (2023). Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping. *European Journal of Remote Sensing*, 56(1), 2191870. <https://doi.org/10.1080/22797254.2023.2191870>

datasets were then processed and analysed using the Random Forest machine learning technique to classify dune habitats at this site. The results showed that using multiple UAV datasets acquired throughout the vegetation growing season achieved higher classification accuracy compared to using just a single dataset (92.37% vs. 84.09%, respectively). Also, including topographic data consistently improved the accuracy, regardless of the number of datasets. Comparing the three UAV-acquired datasets, the analysis suggested that the dataset acquired in the middle period of the growing season, i.e., the flowering period, was better than those acquired in the early or late periods for dune habitat mapping. The high-resolution maps produced from this study provide valuable information about the detailed distribution of dune habitats, as well as highlight areas that are vulnerable to the impacts of human activities, such as frequently used tracks and beach access. These maps can then be used to guide site managers in developing timely and appropriate mitigation strategies.

A critical aspect of habitat conservation is tracking the location and expansion of invasive species, which is considered a major threat to and pressures on Irish habitats. This study explored the potential of utilising UAV-acquired imagery and deep learning (DL) techniques for detailed mapping of invasive species, including those in the early stages of invasion, i.e., occurring in small patches (Chapter 3)². Creating a robust DL model is challenging due to the requirement for large and diverse labelled training data. In remote sensing, the available labelled data can be limited as it is time- and resource-intensive to generate. This limitation can restrict DL models to obtain better predictive performance when applied to new, unseen sets of data. The study implemented a DL semantic segmentation on UAV imagery, testing different model parameters—architectures, encoder backbones and input image patch sizes—to determine an effective network structure. Moreover, the potential of applying data augmentation and pseudo-labelling to increase the amount and diversity of labelled data was investigated. Results showed that the U-Net model architecture with Inception v3 as the backbone and trained on 128×128 image patches was the best model network structure based on its predictive performance: highest mean Intersection-Over-Union (*mIOU*) score of 0.832. Furthermore, the model trained on the augmented and pseudo-labelled data achieved an *mIOU* score of 0.712 on

² Cruz, C., McGuinness, K., Perrin, P. M., O’Connell, J., Martin, J. R., & Connolly, J. (2023). Improving the mapping of coastal invasive species using UAV imagery and deep learning. *International Journal of Remote Sensing*, 44(18), 5713–5735. <https://doi.org/10.1080/01431161.2023.2251186>

an independent dataset, while there was a decrease of 0.158 in model performance when only the original labelled data were used. This result suggests the potential of using data augmentation and pseudo-labelling techniques in creating more robust models. Automating invasive species mapping through deep learning, reducing the time and effort required, is a significant step towards a more efficient assessment of habitat conditions.

To date, a single UAV flight can only capture images over a small area—usually a few square kilometres—due to factors such as the limited battery life, the requirement of operating a UAV within the visual line-of-sight of the operator, and the specified maximum flight altitudes (120 metres in Ireland). These factors can make UAV surveys impractical for mapping habitats in an extensive, often rugged terrain, such as uplands. Therefore, high-resolution (2m) satellite imagery was used for mapping upland habitats (Chapter 4)³. The study specifically investigated the impact of varying spatial resolutions on classification accuracy. Moreover, a fuzzy classification technique based on Random Forest was implemented as upland habitats are typically heterogeneous and occur in complex mosaics. Results from the analysis indicated that using higher spatial resolution images generally led to a more accurate classification compared to using images of lower resolutions (2m-resolution image: 80.34% vs 10m-resolution image: 73.55%). The fuzzy classified maps provided information on the probability of each habitat being present in every pixel of the image, giving more insights into the habitat composition. The generated habitat probabilities were further used to estimate the spatial confidence of classification over an area. In this way, more targeted assessments of areas of high uncertainty can be recommended.

Overall, the combined use of high-resolution remote sensing data and machine learning techniques offers massive potential for repeatable and systematic approaches to the fine-scale characterisation of habitats. The resulting detailed maps from these approaches can provide critical information to guide and inform habitat conservation efforts. Moreover, these maps can support and contribute to the evidence-based implementation of Goal 15 of the Sustainable Development Goals, which focuses on “protecting, restoring and promoting sustainable use of terrestrial ecosystems and preventing biodiversity loss”.

³ Cruz, C., Perrin, P. M., Martin, J. R., O’Connell, J., McGuinness, K., & Connolly, J. Mapping of temperate upland habitats using high-resolution satellite imagery and machine learning. (Under review).

Contents

Acknowledgements.....	i
Summary.....	iii
Contents	vii
List of Appendices	xi
List of Figures.....	xii
List of Tables	xvi
List of Abbreviations	xviii
1. Introduction.....	1
1.1. Definition and importance of habitats.....	1
1.1.1. The decline of habitats	1
1.2. The need for habitat distribution maps.....	2
1.3. The role of remote sensing and machine learning	3
1.3.1. Principles of optical remote sensing	4
1.3.2. Principles of machine learning	11
1.4. Remote sensing and machine learning for habitat mapping and monitoring...	20
1.5. Aims of the thesis.....	24
1.6. Structure of the thesis.....	26
1.7. References	28
2. Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping	45
2.1. Abstract	47
2.2. Introduction.....	48
2.3. Materials and methods	51

2.3.1.	Study site description	51
2.3.2.	General workflow	55
2.4.	Results	67
2.4.1.	Acquisitions and variables used in the classification	67
2.4.2.	Temporal analysis of dune habitats identification.....	72
2.4.3.	Distribution and extent of dune habitat types.....	72
2.5.	Discussion.....	75
2.6.	Conclusion.....	79
2.7.	References	79
3.	Improving the mapping of coastal invasive species using UAV imagery and deep learning.....	91
3.1.	Abstract.....	93
3.2.	Introduction	94
3.3.	Materials and methods.....	98
3.3.1.	Study site description	99
3.3.2.	Data collection.....	101
3.3.3.	Data preparation	103
3.3.4.	Training parameters and evaluation metric	104
3.3.5.	Experiments.....	105
3.3.6.	Evaluation of the final model	109
3.4.	Results	109
3.4.1.	Model performance comparison between architectures and backbones using different patch sizes.....	109
3.4.2.	Data augmentation and pseudo-labelling	115
3.4.3.	Evaluation of the final model on test data.....	116
3.5.	Discussion.....	118
3.6.	Conclusion.....	121

3.7. References	122
4. Mapping of Irish upland habitats using high-resolution satellite imagery and machine learning.....	131
4.1. Abstract	133
4.2. Introduction	134
4.3. Materials and methods	136
4.3.1. Study site description.....	136
4.3.2. General workflow	139
4.4. Results	148
4.4.1. Effect of spatial resolution on mapping upland habitats.....	148
4.4.2. Model predictions	152
4.4.3. Distribution and extent of upland habitats	155
4.5. Discussion	158
4.6. Conclusion.....	162
4.7. References	162
5. Synthesis	173
5.1. Research aim	173
5.2. Summary of key findings	173
5.2.1. Mapping coastal dune habitats using UAV data and machine learning.	175
5.2.2. Deep learning-based mapping of invasive species using UAV images..	177
5.2.3. Mapping upland habitats using high-resolution satellite imagery and machine learning.....	179
5.3. Evaluation of the combined high-resolution imagery and machine learning approach for habitat monitoring	182
5.3.1. UAV surveys cover limited areas	182
5.3.2. Limited labelled remote sensing data	182

5.3.3.	Requirement of high computation resources.....	183
5.3.4.	Highly technical skills are needed for data collection, processing and analysis 183	
5.4.	Practical implications of the HRI-ML approach	184
5.5.	Overall summary, conclusions and recommendations for further work	190
5.5.1.	Recommendations for future work.....	192
5.6.	References	195
6.	Appendices	205
	Appendix A: Guidelines and best practices in data collection.....	206
	Appendix A1: UAV survey protocols and best practices	206
	Appendix A2: Field data collection protocols and best practices	211
	Appendix A3: Manual in using a GNSS receiver.	214
	Appendix A4: A selection of photographs taken during fieldwork.	235
	Appendix B: Supplementary materials for Chapter 2	240
	Appendix B1: Annex I dune habitat descriptions and plot data.....	240
	Appendix B2: Other supplementary materials	249
	Appendix C: Supplementary materials for Chapter 4	251
	Appendix D: List of publications, conference presentations, training and other outreach activities during PhD	262

List of Appendices

Appendix A: Guidelines and best practices in data collection

Appendix A1: UAV survey protocols and best practices

Appendix A2: Field data collection protocols and best practices

Appendix A3: Manual in using a GNSS receiver

Appendix A4: A selection of photographs taken during fieldwork

Appendix B: Supplementary Materials for Chapter 2

Appendix B1: Annex I dune habitat descriptions and plot data

Appendix B2: Other supplementary materials

Appendix C: Supplementary Materials for Chapter 4

Appendix D: List of publications, conference presentations, training and other outreach activities during PhD

List of Figures

Figure 1.1. The distribution of solar radiation, with most of it being in the visible (400-700 nm) and near-infrared (700-1300 nm) regions of the electromagnetic spectrum.	5
Figure 1.2. Incoming radiation – Earth’s surface interactions: reflected (a), absorbed (b) and/or transmitted (c). Source: Lillesand et al. (2015).....	6
Figure 1.3. Spectral reflectance curves for healthy vegetation (a), dry vegetation (b) and dry, bare soil (c).	7
Figure 1.4. A four-band image showing a single pixel with information on its size and location represented by row and column.....	8
Figure 1.5. Pixel values in a panchromatic image (left) and in a multi-spectral image (right).....	10
Figure 1.6. Examples of overfit (a), underfit (b) and good fit (c) models. Source: Burkov (2019, p. 11).	13
Figure 1.7. A decision tree diagram. The root node is represented by X. Decision (or nonterminal) nodes are represented by D_N and leaf (or terminal) nodes are represented by L_N . Figure modified from Breiman et al. (1984).....	14
Figure 1.8. Schematic diagram of Random Forest.....	16
Figure 1.9. Three main layers of a neural network – an input layer, one (or more) hidden layers, and an output layer. Source: Nielsen (2015)	18
Figure 1.10. Neural network architecture (Source: Castelluccio et al. (2015))	19
Figure 2.1. Inset map showing the location of the study site in County Kerry, Ireland (left); the boundary of the Tralee Bay and Magharees Peninsula, West to Cloghane SAC (upper right); and the boundary of the study site (lower right).	52
Figure 2.2. A typical dune profile (modified from Delaney et al., 2013) and sample field photos of Annex I and Fossitt habitats	54

Figure 2.3. Methodology workflow in using multi-temporal high-resolution UAV imagery to characterise different habitat types.	56
Figure 2.4. Diagram of the reference data collection and processing steps.....	58
Figure 2.5. Preparation of <i>Test 1</i> and <i>Test 2</i> datasets to be used for modelling scenarios. May (M), July (J) and October (O) were the three field campaign months in this study.	66
Figure 2.6. Confusion matrices obtained for each RF model. ‘Truth’ refers to the number of habitat samples from the reference data. ‘Predicted’ refers to the number of habitat samples identified from the classification. The colour scale along the major diagonal line indicates the proportion of reference samples classified correctly as a function of overall accuracy rounded between 0.0 and 1.0.	69
Figure 2.7. Supervised Random Forest classification map of coastal dune habitats in the Magharees, County Kerry and the computed area (in ha) for each habitat type classified using UAV images acquired at three different times within the growing season. The classified map is based on the July data.	74
Figure 3.1. <i>Spartina anglica</i> (at the back) displaces low-growing native <i>Salicornia</i> plants (in the foreground) (photo taken at North Bull Island, Ireland; August 2021).....	95
Figure 3.2. Workflow using a semantic segmentation deep learning algorithm.	99
Figure 3.3. Map location of the study area on North Bull Island, Ireland (left) and close-ups of the training and test study sites (right).....	100
Figure 3.4. Comparison between the UAV images taken at <i>Site_{train}</i> (left) and <i>Site_{test}</i> (right).	102
Figure 3.5. Sample RGB image and delineated polygons of <i>S. anglica</i> boundaries (left). Binary image representing the presence of <i>S. anglica</i> (white pixels) and its background (black pixels) (right).	103
Figure 3.6. Original image patch (a) and generated images after multiples of 90° rotation (b-d) and colour jitter (e-h) augmentations were applied.	107

Figure 3.7. The pseudo-labelling process.....	108
Figure 3.8. Training speed comparison of each model for different combinations of backbone, model architecture and patch size.	112
Figure 3.9. Inference speed (in milliseconds) on a single image patch for different combinations of backbone, model architecture and patch size.	114
Figure 3.10. Number of labelled patches to train a model and its corresponding performance represented by <i>mIOU</i> score.	115
Figure 3.11. Sample RGB images (left) and the equivalent probability images as training iteration progresses. The colour scale indicates the probability that <i>Spartina anglica</i> is present at the location of each pixel (1.0 = present, 0.0 = absent).	116
Figure 3.12. Sample RGB <i>test</i> images and their corresponding segmentation reference labels and model predictions, with an image indicating incorrectly classified pixels (red).	117
Figure 4.1. Inset maps showing the locations of the two study sites: Wicklow Mountains (top right) and Slieve Mish Mountains (bottom right), and the boundaries of the SAC covering each site.	137
Figure 4.2. Sample field pictures of the three main upland habitats in the study sites: wet heath, dry heath and blanket bog.....	138
Figure 4.3. Diagram of the workflow illustrating the processing performed in the study.	139
Figure 4.4. Sample images after resampling from the original 2-m resolution pixel size to 4-m, 6-m, 8-m and 10-m resolution pixel sizes.....	146
Figure 4.5. Variable importance scores for WM and SM models at different spatial resolutions. The x-axis displays the input variables analysed in the study, while the y-axis represents the variable importance score calculated using MDI. Variable descriptions can be found in Table 4.2.....	150

Figure 4.6. Crisp classification results over a portion of the WM site (top left) based on five different spatial resolutions (2 m, 4 m, 6 m, 8 m and 10 m). The encircled section illustrates the increasing size of mapped waterbodies with lower spatial resolution datasets.....	151
Figure 4.7. Areal proportion for each habitat computed from the model classification results for WM and SM sites	152
Figure 4.8. Basemap of a portion of the WM site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.....	154
Figure 4.9. Maps showing the basemap (a) and the DTM (b) of the WM site, and the corresponding crisp classification result (c) and entropy image (d) for the 2-m resolution model. The colours in (d) represent the level of certainty: lighter colours being more certain and darker colours being less certain.	156
Figure 4.10. Maps showing the basemap (a) and the DTM (b) of the SM site, and the corresponding crisp classification result (c) and entropy image (d) for the 2-m resolution model. The colours in (d) represent the level of certainty: lighter colours being more certain and darker colours being less certain.	157
Figure 4.11. Eroded peat surface in the WM site.	161
Figure 5.1. Overview of the key findings of this research.....	174
Figure 5.2. Sample output maps produced in this thesis. Refer to individual Chapters 2 and 4 for the complete habitat names for maps A and C, respectively.....	181
Figure 5.3. Sample UAV-derived elevation model (Source: Cruz, O’Connell, et al., 2023)	186
Figure 5.4. Code snippet on open-source Python	190

List of Tables

Table 2.1. Habitat types found within Irish dune systems and corresponding habitat codes (following European Commission, 2013; Fossitt, 2000).	53
Table 2.2. List of variables used in the study and their descriptions.	61
Table 2.3. Datasets used in training models.....	64
Table 2.4. 5-fold cross-validation mean accuracy and standard error (SE) values for models obtained for each dataset. The datasets considered are described in Table 2.3. Values are in percentage (%).	67
Table 2.5. Per-class precision (P) and recall (R) values and F1-scores (F1) obtained for models from each dataset. The habitat codes are described in Table 2.1.	71
Table 2.6. Summary table showing the mean and standard error (SE) values for each model evaluated on the <i>Test 1</i> dataset.	72
Table 2.7. Summary table showing the mean and standard error (SE) values for each model evaluated on the <i>Test 2</i> dataset.	72
Table 3.1. Information on the acquired UAV imagery.	103
Table 3.2. Information about the number of patches for the two patch sizes.	105
Table 3.3. List of encoder backbones used in this study.....	106
Table 3.4. <i>Mean IOU(mIOU)</i> scores when using different encoder backbones in U-Net, LinkNet and FPN applied on two different image patch sizes. The text in bold indicates the model with the highest performance and its corresponding <i>mIOU</i> score.....	110
Table 3.5. Model performance when data augmentation and pseudo-labelling techniques were removed from the process.	118
Table 4.1. List of habitats for each study site.....	141
Table 4.2. List of variables used in the study and their descriptions.	144

Table 4.3. List of RF-based models developed for this study. Each model of WM and SM sites was trained on different spatial resolutions (2 m, 4 m, 6 m, 8 m and 10 m)	147
Table 4.4. 5-fold cross-validation mean accuracy and standard error results for WM and SM models. Descriptions of the models are presented in Table 4.3.....	149

List of Abbreviations

The list defines abbreviations that will be later used within the body of the thesis.

ARVI	Atmospherically Resistant Vegetation Index
BVLOS	Beyond Visual Line of Sight
DL	Deep Learning
DSM	Digital Surface Model
EU	European Union
EVI	Enhanced Vegetation Index
EVI2	Enhanced Vegetation Index 2
GCP	Ground Control Point
GIS	Geographic Information System
GNDVI	Green Normalised Difference Vegetation Index
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HD	Habitats Directive
IAS	Invasive alien species
<i>IOU</i>	Intersection-over-Union
IVC	Irish Vegetation Classification
MGRVI	Modified Green Red Vegetation Index
<i>mIOU</i>	mean Intersection-over-Union
MSAVI2	Modified Soil-Adjusted Vegetation Index 2
NDRE	Normalised Difference Red Edge
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NIR	Near Infrared
NPWS	National Parks and Wildlife Service
OOB	Out-of-bag
OSAVI	Optimised Soil Adjusted Vegetation Index
PCA	Principal Component Analysis
RE	Red Edge
RF	Random Forest

RGB	Red Green Blue
RGBVI	Red Green Blue Vegetation Index
SAC	Special Area of Conservation
SE	Standard Error
SfM	Structure-from-Motion
SR	Simple Ratio
SSL	Semi-supervised Learning
UAV	Unoccupied Aerial Vehicle

“We cannot protect something we do not love, we cannot love what we do not know, and we cannot know what we do not see. Or hear. Or sense”

– Richard Louv

1. Introduction

1.1. Definition and importance of habitats

Habitats are natural environments where particular species of plants or animals normally live and grow. They are characterised by specific physical conditions, including topography, soil type, climate and water availability (Davies et al., 2004). Habitats provide many important benefits, such as biodiversity support, carbon capture and storage, water regulation, coastal defence and recreation (Millennium Ecosystem Assessment, 2003).

1.1.1. The decline of habitats

Although recognised for their importance, habitats are highly threatened by human activities, including unsustainable resource use, intensive agricultural practices, pollution and urbanisation (European Environment Agency, 2020; Scanes, 2018). Habitat loss and degradation due to these activities have been considered major drivers of biodiversity decline (Brooks et al., 2002; Sala et al., 2000). The urgent need to protect remaining intact habitats while restoring degraded habitats has become a global concern reflected by the numerous existing environmental legislations, such as the Habitats Directive (HD) (Vanden Borre, Paelinckx, et al., 2011). The HD is one of the European Union's (EU) legal frameworks for nature conservation. Under Article 11 of the HD, Member States are required to undertake surveillance of the conservation status of habitats—those listed in Annex I of the HD—protected within the EU territory (Commission of the European Communities, 1992). However, most of these habitats were recently assessed as having poor conservation status, with as much as 36% were recognised as deteriorating (European Environment Agency, 2020). Ireland is no exception; the latest report to the European Commission showed that only about 15% of the assessed habitats were in good conservation status (National Parks and Wildlife Service, 2019). One of the reported

major threats was the establishment and spread of invasive species (National Parks and Wildlife Service, 2019); hence, it is important to identify their locations and spatial extents. With these alarming assessment results, continuous habitat monitoring is crucial to ensure the development and implementation of timely and appropriate mitigation and restoration strategies to prevent further decline.

1.2. The need for habitat distribution maps

Habitat monitoring is an essential task for biodiversity conservation and protection. The term habitat monitoring is defined by Lengyel et al. (2008) as the repeated recording of the habitat condition to detect any changes or trends from a predetermined state of the habitat. Assessing the distribution of habitats is fundamental when evaluating their conditions. Monitoring, through repeated mapping, to document information about habitat locations, their spatial extent at specific sites, and any changes occurring over time is a crucial aspect of assessment (A. M. Schmidt & van Swaay, 2021).

To date, conventional field-based ecological surveys are the common method that is being utilised in Ireland to generate distribution maps for the purpose of monitoring habitats, including those protected under the HD (Brophy et al., 2019; Delaney et al., 2013; O'Neill et al., 2013; Perrin et al., 2014). However, there are limitations associated with field-based surveys. Firstly, manual methods alone can limit the size and temporal scope of the survey as they involve considerable time and effort (Roelfsema et al., 2021). Some areas are also under- or un-surveyed for safety considerations because they are remote or difficult to access, e.g., dangerously steep slopes (Perrin et al., 2009) or remote forests (Asner et al., 2015); this can lead to potential gaps in information. Secondly, repeated surveys of extensive areas are often impractical as they can be costly (Buchanan et al., 2005; Green et al., 1999; Rhodes et al., 2015), inhibiting periodic mapping of habitats, which is important and demanded by many nature legislations (e.g., Article 17 of the HD and Article 11 of the proposed Nature Restoration Law). Furthermore, there may be variations in the habitat mapping undertaken by diverse ecologists, resulting in a potential bias in habitat identification (Cherrill, 2016; Cherrill & McClean, 1995). In the context of this thesis, best practice guidance that is optimal for field data collection was developed to

mitigate potential inconsistencies between ecologists and ensure superior quality in the collected data (see Appendix A2).¹ Finally, field-based surveys can be particularly challenging when mapping individual habitat types within heterogeneous and complex mosaics, such as upland habitats (Perrin et al., 2009). Therefore, while field-based surveys are essential for gathering valuable habitat information, they may not be practical for repeated assessments of habitats over large areas at regional or national scales.

1.3. The role of remote sensing and machine learning

Remote sensing technology offers significant opportunities for mapping and monitoring habitats (Corbane et al., 2015; Nagendra et al., 2013). In general, this technology is based on studying reflected radiation by a feature of interest, measured and recorded by a sensor from a distance (Lillesand et al., 2015). It offers several advantages that complement traditional field-based observations (Vanden Borre, Haest, et al., 2011). First, remote sensing can provide spatially continuous information over broad areas, including dangerous or inaccessible sites (e.g., large wetlands or steep mountain slopes). This information can be used to produce a complete-coverage map of habitat distributions, as well as an estimation of many quantitative environmental parameters, such as soil moisture, chlorophyll content and biomass. Second, remote sensing allows for frequent, systematic observations over the same area, which is important for monitoring applications. Third, the data collected through remote sensing can be interpreted and analysed using advanced computer-assisted analytical techniques, such as machine learning. Machine learning is an efficient and effective technique for processing remotely sensed data (Maxwell et al., 2018). This technique can extract complex features from the data, which can be used to characterise habitats accurately. As such, the combined remote sensing data and machine learning techniques, together with field-based measurements, has been considered an alternative and promising tool to mapping habitats. This thesis develops on previous research on remote sensing-machine learning approaches for habitat mapping and monitoring.

¹ Ecologists were mainly involved in the field-based ecological surveys as their expertise is vital in habitat mapping research. Authors' contributions were explicitly stated at the beginning of Chapters 2, 3 and 4.

1.3.1. Principles of optical remote sensing

A broad and simple definition of remote sensing refers to collecting information about an object without being in physical contact with it (Barrett & Curtis, 1999; Elachi & Van Zyl, 2006). In this thesis, however, remote sensing specifically focuses on gathering information about the Earth's surface using airborne or space-borne platforms. Therefore, the definition of remote sensing was adopted from Campbell & Wynne (2011, p. 6), i.e.,

Remote sensing is the practice of deriving information about the Earth's land and water surfaces using images acquired from an overhead perspective, using electromagnetic radiation in one or more regions of the electromagnetic spectrum, reflected or emitted from the Earth's surface.

Using this definition, the sun is the most common energy source for remote sensing of the Earth's surface. It emits varying amounts of radiation at all wavelengths (denoted by the Greek letter lambda, λ) in the electromagnetic spectrum, with most of it in the visible (400–700 nm) and near-infrared (NIR) (700–1,300 nm) regions (Figure 1.1). Most common remote sensing systems for Earth observation operate in these regions. This type of remote sensing is called optical remote sensing. The narrow visible region is associated with the “colours” the human eyes can see. The NIR region is a portion of the spectrum that is invisible to the human eyes, yet it is used to measure important properties, such as vegetation health (Lillesand et al., 2015).

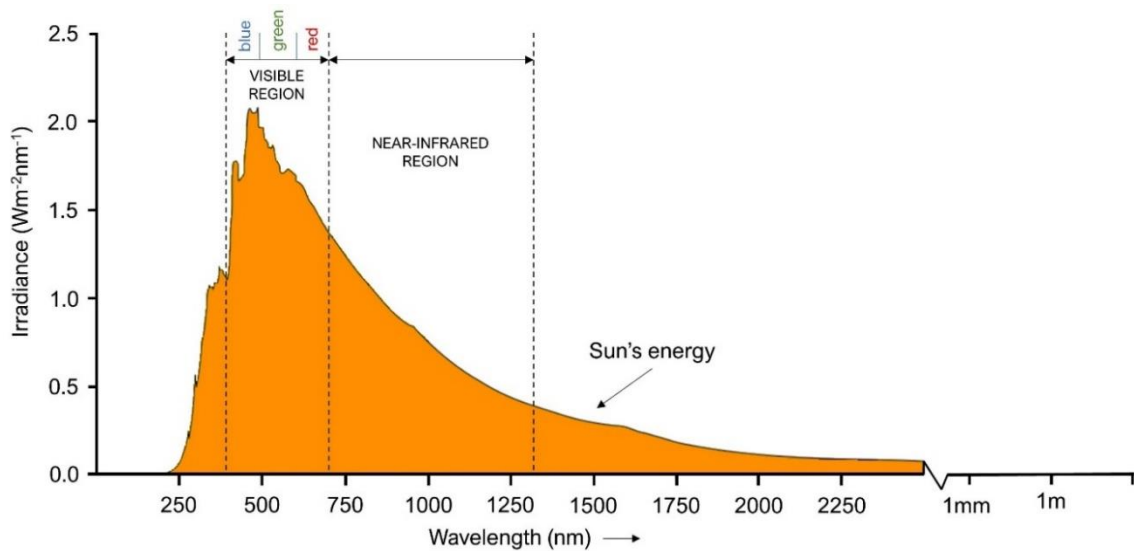


Figure 1.1. The distribution of solar radiation, with most of it being in the visible (400-700 nm) and near-infrared (700-1300 nm) regions of the electromagnetic spectrum.

The incoming radiation from the sun (irradiance), not influenced by atmospheric absorption, reaches and interacts with different features on the Earth's surface. Proportions of this are either reflected, absorbed and/or transmitted by a certain feature, depending on its properties (Figure 1.2) (Campbell & Wynne, 2011; Lillesand et al., 2015). Reflection occurs when the incoming radiation hitting a non-transparent surface bounces off and is then redirected into space at an angle (Figure 1.2a). Absorption is when the radiation is absorbed by the surface (Figure 1.2b), and transmission occurs when the radiation passes through a surface (Figure 1.2c) (Campbell & Wynne, 2011). In particular, analysing the amount of reflected radiation (i.e., reflectance) across a range of wavelengths is fundamental in remote sensing. This pattern of reflected radiation as a function of wavelength is called the spectral reflectance curve (Lillesand et al., 2015).

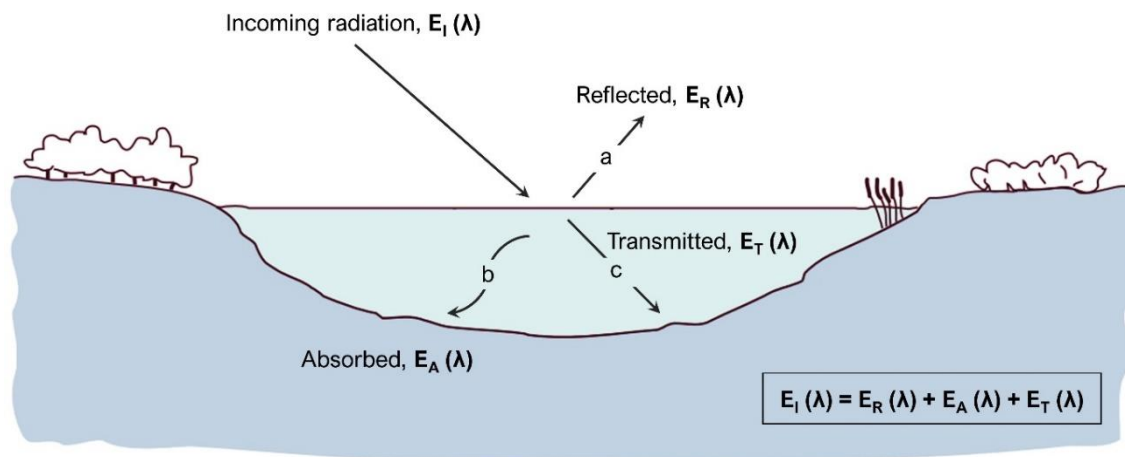


Figure 1.2. Incoming radiation – Earth’s surface interactions: reflected (a), absorbed (b) and/or transmitted (c). Source: Lillesand et al. (2015)

Earth’s surface features, such as vegetation and soil, exhibit different spectral reflectance curves (Figure 1.3). The information from these curves is useful in identifying different features from images acquired by remote sensing (Aggarwal, 2004; Campbell & Wynne, 2011; Lillesand et al., 2015). For example, a general spectral curve of healthy vegetation, as shown in Figure 1.3a, has high reflectance values in the NIR region due to the plant’s high chlorophyll concentration. In the visible region, chlorophyll reflects more energy in the green region (500-600 nm); this high reflectance gives vegetation its green colour. However, it absorbs as much as 70-90% radiation in the blue (400-500 nm) and red (600-700 nm) regions for use in photosynthesis (Campbell & Wynne, 2011). When the leaves dry out (i.e., no process of photosynthesis)—for example, due to a change in season—the red reflectance increases, causing the brown appearance of dried leaves. At the same time, the NIR reflectance decreases (Figure 1.3b). Similar to dry vegetation, the spectral curve of soil is less variable. In general, soil has an increasing reflectance towards the longer wavelength region (Figure 1.3c). Factors affecting soil reflectance include moisture condition, organic matter content, texture, roughness and iron oxide content (Lillesand et al., 2015).

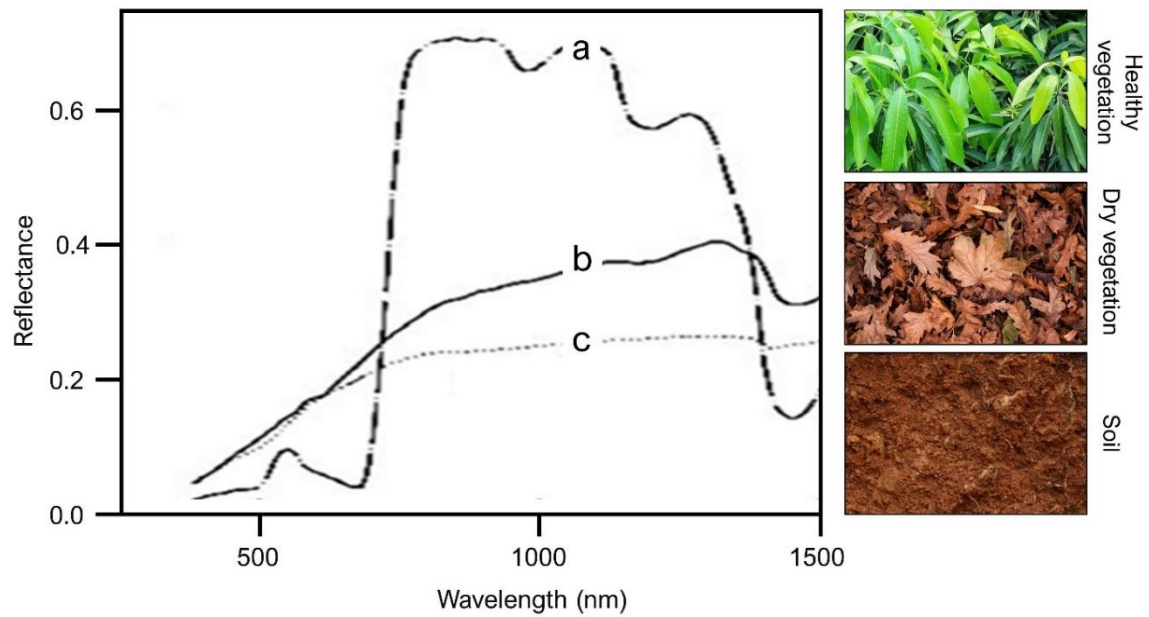


Figure 1.3. Spectral reflectance curves for healthy vegetation (a), dry vegetation (b) and dry, bare soil (c).

Optical remote sensing data can be represented by a two-dimensional array of discrete picture elements or pixels (Lillesand et al., 2015). In remote sensing, a pixel is defined as the smallest unit of an image representing an area on the Earth's surface at a specific location. The numerical value of each pixel corresponds to the amount of radiation reflected by the Earth's surface area at a particular wavelength, and its location is denoted by the row and column numbers within the image (Figure 1.4) (Campbell & Wynne, 2011). Generally, a separate image or 'band' is assigned for each recorded wavelength (Figure 1.4).

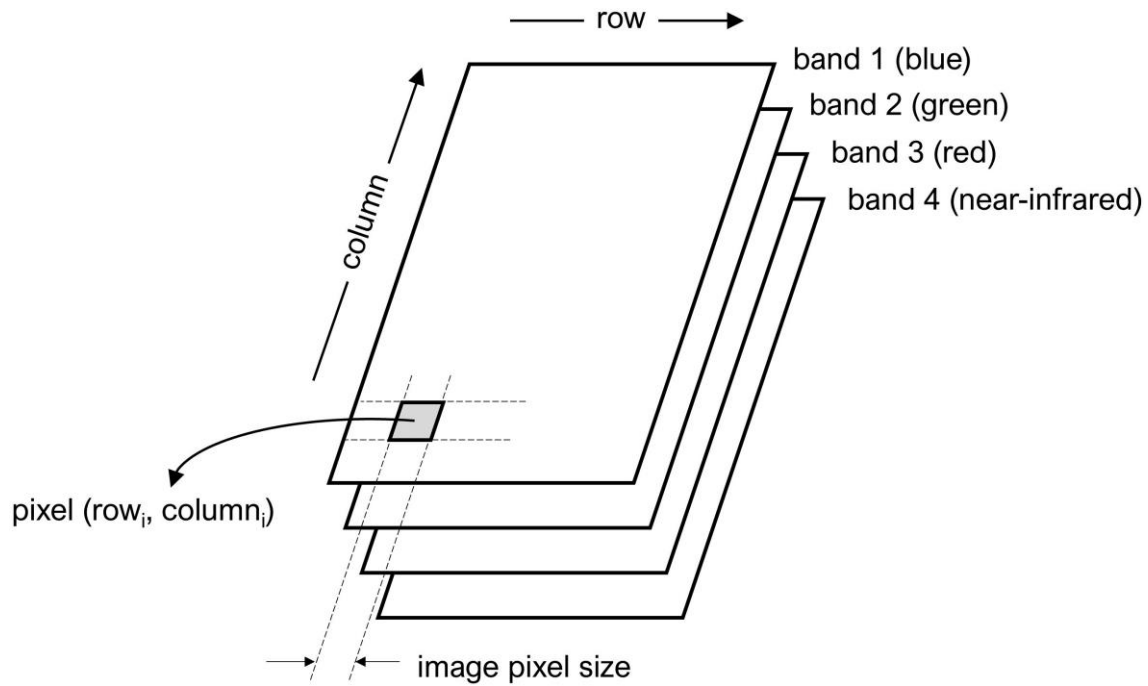


Figure 1.4. A four-band image showing a single pixel with information on its size and location represented by row and column.

1.3.1.1. Remote sensing platforms and sensors

A platform refers to the type of vehicle that supports or carries a sensor (Lillesand et al., 2015). A sensor is an instrument or device that measures electromagnetic radiation (de Jong et al., 2004; Jensen, 2014). Different combinations of platforms and sensors can determine the characteristics of the remotely sensed data. These include the ground area coverage (spatial extent), image pixel size (spatial resolution), number and width of wavelength intervals or bands where information is recorded (spectral resolution) and the amount of time needed to revisit and acquire data over the same location on the Earth's surface (temporal resolution) (Turner et al., 2003).

Different platforms, which can be operated at varying heights above the ground—from centimetres (ground-based: handheld devices) to metres (airborne: aircraft) or up to kilometres (space-borne: satellites)—can influence the size of the imaged area and the spatial resolution (Barrett & Curtis, 1999). In general, the higher a platform operates above the ground, the larger the area it can cover. A single satellite image typically covers an area of thousands of square kilometres on the ground (Drusch et al., 2012). An airborne

platform can cover an area of up to hundreds of kilometres on the ground in a single flight (Rowat et al., 2009).

The spatial resolution of the image corresponds to the size of the area on the ground that a single image pixel covers (Lillesand et al., 2015). For example, a 10-metre resolution image corresponds to a 10×10 -metre area on the ground. Many satellites acquire images with low (>30 metres/pixel) to medium (10-30 metres/pixel) spatial resolutions; most of these images are free to obtain (Kennedy et al., 2014; Wulder et al., 2022). Commercial satellites, however, can provide images with higher spatial resolutions (<10 metres/pixel). The revisit time of a satellite, which is often fixed, ranges from a few days to a few weeks. In contrast, airborne platforms offer flexible acquisition of submetre-level resolution images (Carbonneau & Lane, 2004; Nolan et al., 2015).

Depending on the sensor, a remote sensing image can be acquired in a single band, multiple bands or even hundreds of bands. A single-band image (panchromatic) has a single value for each pixel (Figure 1.5, left image). In a multi-band (multi-spectral) image, several bands—typically 3-10—represent reflected radiation measured at different wavelengths (Lillesand et al., 2015). In other words, each pixel on a multi-spectral image is associated with more than one value (Figure 1.5, right image).

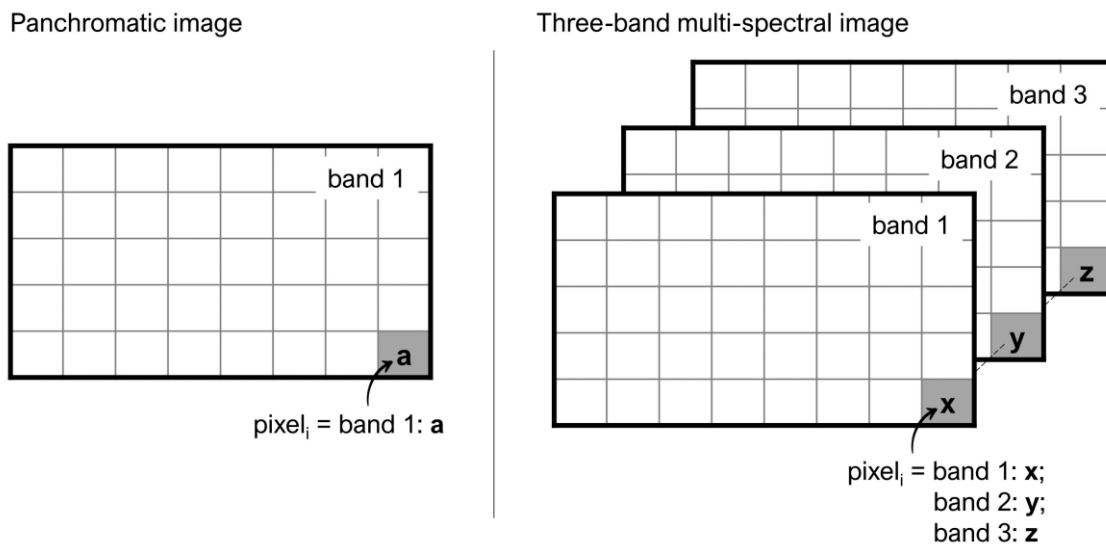


Figure 1.5. Pixel values in a panchromatic image (left) and in a multi-spectral image (right).

Overall, the ability to acquire repeated observations—hence, the potential to record changes in a feature’s spectral characteristics over time—makes remote sensing an important tool in monitoring applications.

1.3.1.2. Pixel-based and object-based image analyses

There are two approaches for analysing remote sensing images: pixel-based and object-based. In the traditional pixel-based image analysis, each pixel is assigned to a specific class, often based on its spectral properties (i.e., pixels with similar spectral characteristics are classified as a single class) (Lillesand et al., 2015). In contrast, an object-based image analysis (OBIA) involves dividing an image into groups of pixels—also known as objects—based on their spectral, spatial and contextual characteristics (segmentation) and then assigning these individual objects into classes (classification) (Blaschke, 2010). While OBIA has been widely used in several remote sensing studies (de Pinho et al., 2012; Silver et al., 2019; Walter, 2004), it has some limitations (Im et al., 2014; Liu & Xia, 2010; Zerrouki & Bouchaffra, 2014). Firstly, determining optimal segmentation parameters, such as scale, is crucial to achieving accurate OBIA classification. However, these parameters are set by an analyst and involve a trial-and-error method, which can be time-consuming, prone to error and impractical for large-scale applications. In other

words, selecting inappropriate scale values could impact the segmentation of objects, either through over- or under-segmentation, and hence, may lead to a decrease in the classification accuracy. Secondly, segmenting images of areas with complex habitat mosaics can be challenging. Different habitats may vary in size or have overlapping spectral signatures, which could lead to the inability of OBIA to accurately capture or detect gradual changes when conducting multi-temporal analysis. Finally, processing a large amount of high-resolution imagery for multi-temporal analysis using OBIA can be computationally intensive, requiring significant computing resources. For these reasons, pixel-based image analysis was used in this thesis.

1.3.2. Principles of machine learning

The ability to acquire remotely sensed data has been significantly improved in the last few decades. Different space-borne and airborne sensors now provide a massive amount of data characterised by increased spatial and spectral resolutions and higher revisit times (Chi et al., 2016; Liu, 2015; Ma et al., 2015; Waske et al., 2009).² This requires an efficient technique to fully exploit and extract useful information from these datasets.

Machine learning, a subset of artificial intelligence, offers the potential to effectively and efficiently process the large datasets provided by remote sensing (Maxwell et al., 2018). It refers to the development of methods that can improve performance by iteratively learning from data (Waske et al., 2009). Such methods can be broadly divided into two categories: unsupervised and supervised learning, based on whether they use manual annotation (i.e., data labelling) during the learning process or not. In unsupervised learning, only the input data is known, and no known output is provided. It deals with unlabelled data (i.e., no associated label) and extracts knowledge from the structure of this dataset (Muller & Guido, 2017; Waske et al., 2009). In supervised learning, both the input and desired outputs are provided. It deals with labelled data (i.e., associated with a label) to train a model and predicts the output from the input data. This type of learning

² NASA's Earth Science Data and Information System holds about 22 petabytes in 2017, almost three times the amount of data archived in 2013 (<https://www.earthdata.nasa.gov/learn/articles/earth-science-data-cloud-eosdis-cumulus-project>).

is further subdivided into two categories: regression and classification, based on the output they produce. Regression is used to predict a continuous quantitative variable, while classification is used to predict a discrete categorical variable (Muller & Guido, 2017). This thesis focuses on the classification type of learning.

Labelled data are divided into three non-overlapping subsets: 1. the training set; 2. the validation set; and 3. the test set.³ The training set is usually the largest one, comprising around 60% to 80% of the data, and is used for training the model. The validation set is used for monitoring the model performance, fine-tuning the parameters, and selecting the final model. The test data, a completely unseen dataset, is used to evaluate the generalisation performance of the final model (Burkov, 2019).

Generalisation is the ability of a model to perform well on new datasets. Two problems that can lead to poor model performance are overfitting and underfitting. Overfitting occurs when a model predicts well on the training data but exhibits poor generalisation on unseen test data (Figure 1.6a). It happens when the model is too complex for the data or when there are many features but few training samples (Burkov, 2019). On the other hand, underfitting occurs when the model does not adequately predict the labels of the data it was trained on (Figure 1.6b). It occurs when a model is too simple and does not capture the variability in the data (Muller & Guido, 2017). A good model that gives a good generalisation performance balances underfitting and overfitting (Figure 1.6c).

³ The usage of the terms here is different from the remote sensing literature. In this thesis, machine learning terminologies are used.

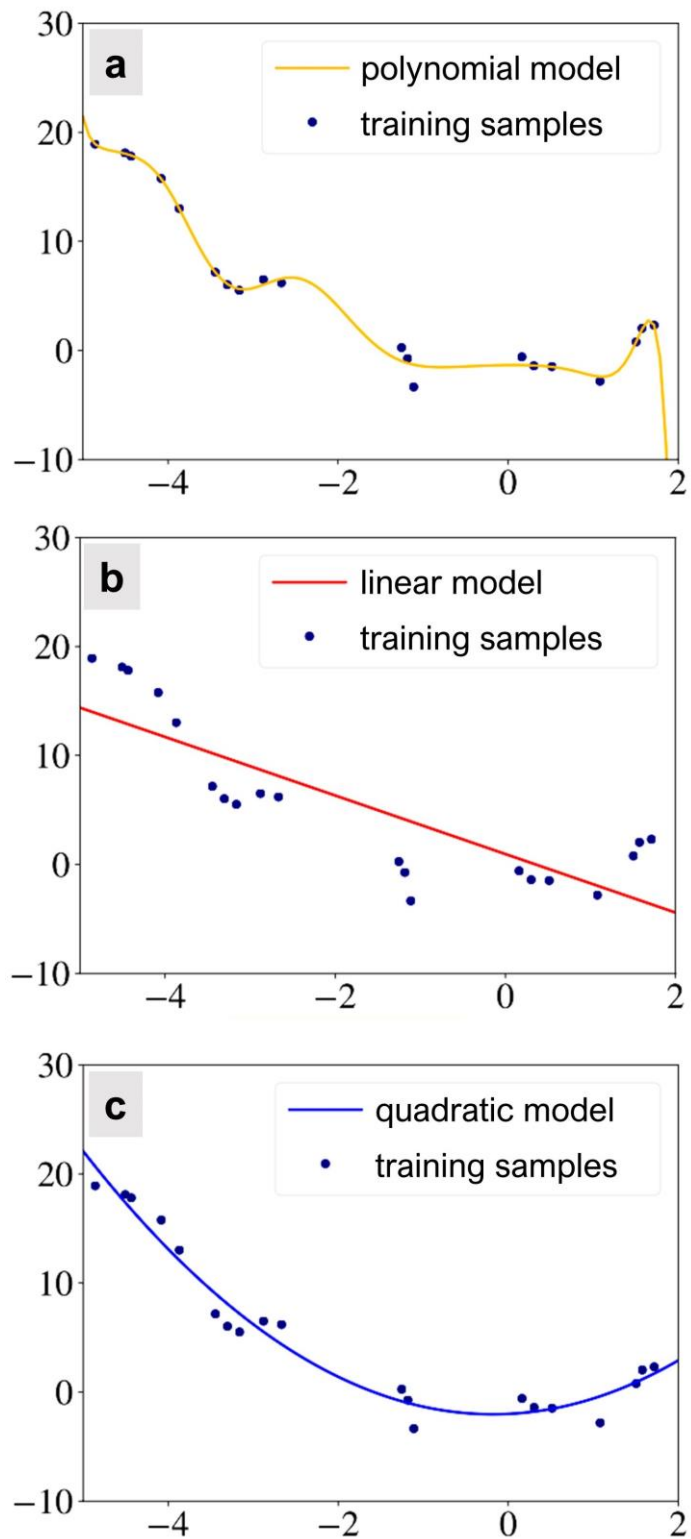


Figure 1.6. Examples of overfit (a), underfit (b) and good fit (c) models.

Source: Burkov (2019, p. 11).

The following subsections introduce the supervised machine learning methods used in this thesis to develop mapping methodologies with high-resolution remote sensing data.

a. Decision Trees

A decision tree is a hierarchical model used in supervised learning that follows a tree-like decision-making process (Figure 1.7) (Zhou, 2012). This process involves a series of recursive partitioning of input space, which means that the input data is divided into smaller groups based on certain conditions (Alpaydin, 2004; Dahan et al., 2014).

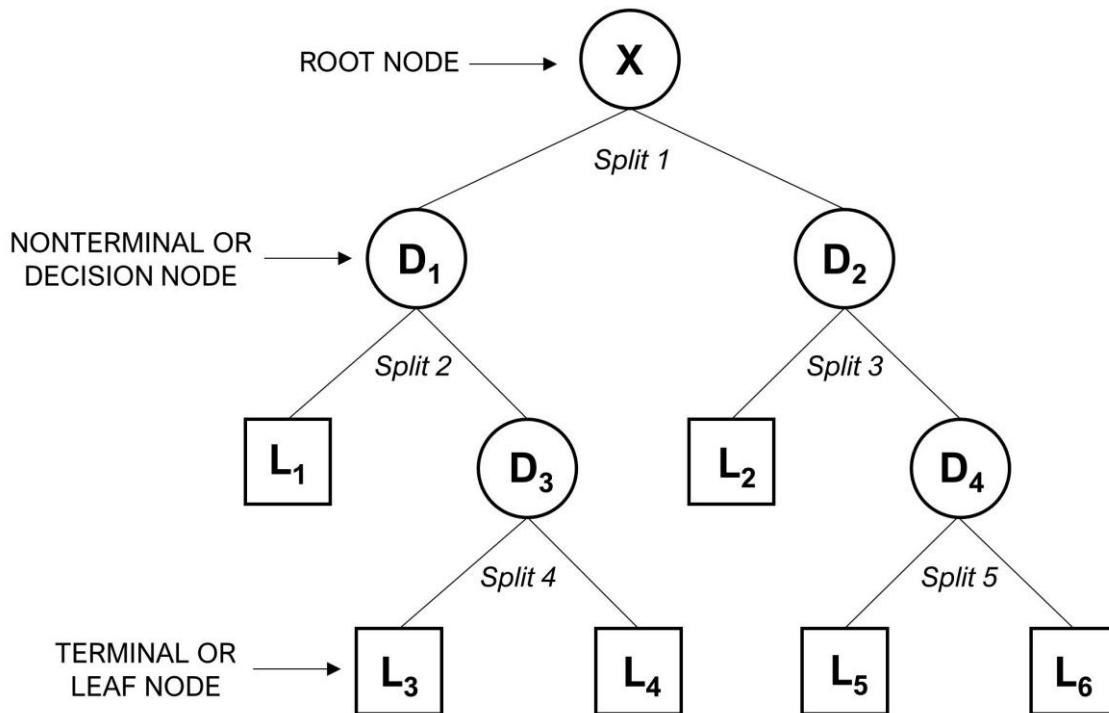


Figure 1.7. A decision tree diagram. The root node is represented by X. Decision (or nonterminal) nodes are represented by D_N and leaf (or terminal) nodes are represented by L_N . Figure modified from Breiman et al. (1984).

In a decision tree, as shown in Figure 1.7, all samples are initially included in the root node. The root node is then divided into two subnodes based on a condition. Each of these subnodes can be a nonterminal or a terminal node. A nonterminal (or decision) node

further splits into two daughter nodes if the condition is met, whereas a terminal (or leaf) node is a node that does not split further and is assigned a class label. To split a node, all possible ways of splitting are considered, and the split that generates the most pure subnodes is chosen (Breiman et al., 1984).

Generally, the process of growing a decision tree stops when all nodes become pure. However, it is necessary to stop the growing of a tree at some point because deep decision trees are complex and can lead to overfitting (Muller & Guido, 2017). Overfitting is when a tree is fully grown and cannot generalise well to new samples beyond the training data. This means that the tree has a low error rate when using the training data but a higher error rate when using new data. There are two ways to prevent overfitting: pre-pruning or post-pruning. Pre-pruning involves stopping the tree creation early by setting a minimum number of samples per node (node size), limiting the depth of the tree, or limiting the number of nodes (Muller & Guido, 2017). Post-pruning, sometimes referred to as pruning, is building the tree until all the leaf nodes are pure and then removing the nodes that provide little information (Muller & Guido, 2017).

b. Random Forest

The Random Forest (RF), introduced by Breiman (2001), is an effective method for constructing ensembles of decision trees (Figure 1.8). This method incorporates randomness in building the tree to ensure that each tree is unique, hence the name “random forest” (Muller & Guido, 2017). Two methods are used to construct each decision tree in a random forest: 1. drawing a subset of training samples with replacement; and, 2. randomly selecting input variables (Pal, 2005; Waske et al., 2009). Combining these two methods means that each tree trains a different set of data and looks at different sets of variables to split a node.

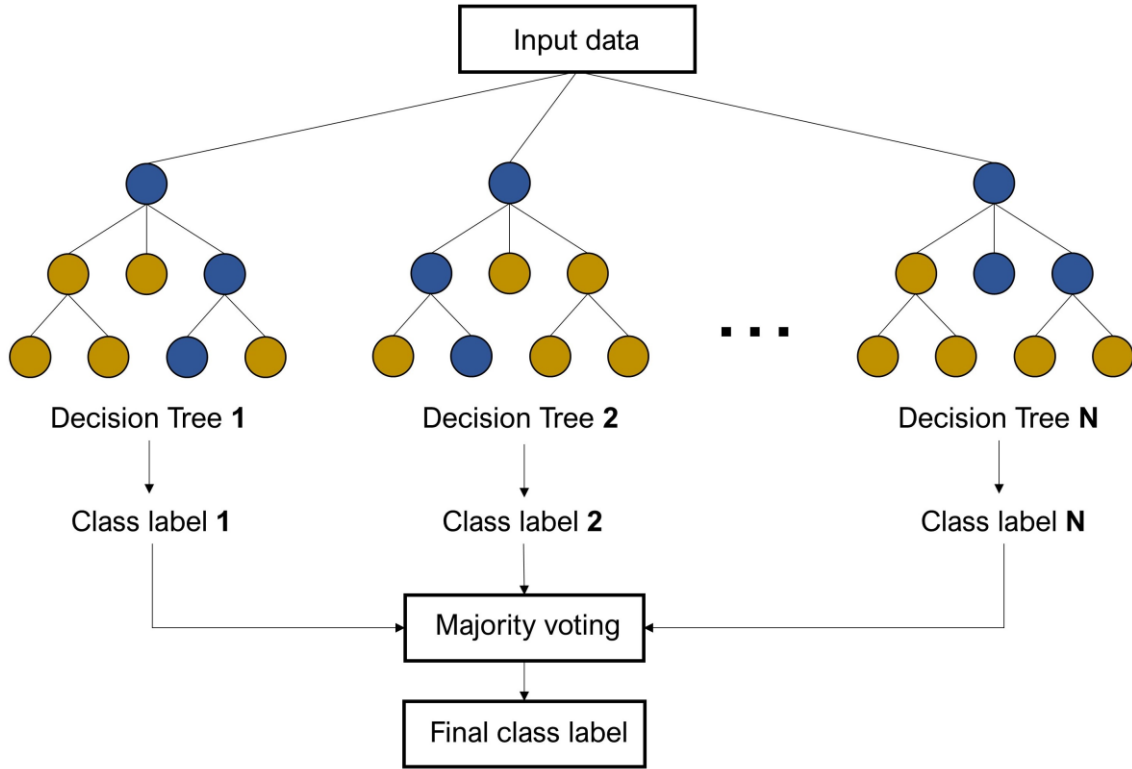


Figure 1.8. Schematic diagram of Random Forest

Each tree in RF is independently created by splitting each node using a randomly selected number of variables (denoted as M_{try}) and the forest is grown up to a certain number of trees (denoted as N_{tree}). Both M_{try} and N_{tree} are user-defined parameters required to create a random forest classifier (Pal, 2005). In classification, the default value is $M_{try} = \sqrt{M}$, where M is the total number of variables. In regression, the default value is $M_{try} = N/3$, where N is the sample size (Cutler et al., 2012). The value of N_{tree} can be selected as large as desired without increasing the error of the model, but rather resulted in more stable classifications (Rodriguez-Galiano et al., 2012). However, in some cases, a large N_{tree} only increases the computational cost without significant improvements in the performance (Oshiro et al., 2012).

About two-thirds of the original training samples are typically used to train the trees, and they are called in-bag samples. The remaining one-third of samples, which are not included in the in-bag samples, are known as out-of-the-bag (OOB) samples. The OOB samples are used to validate the performance of the model (Belgiu & Drăgut, 2016; Fawagreh et al., 2014). Once the RF has been trained and validated, it can be used to

predict the class labels of the unknown data. The prediction is made through majority voting. In a majority voting scheme, each tree in RF casts one vote for its predicted class label. Then, once all the trees have voted, the class with the most votes will be the final label of the unknown data (Figure 1.8) (Fawagreh et al., 2014).

The RF method can also provide the relative importance of the different variables used during the classification process (Breiman, 2001). Variable Importance (VI) indicates how each variable influences the accuracy of the model (Millard & Richardson, 2015). This aspect of RF is useful in determining the variables that contain the most information for the classification, particularly when using multi-source datasets (Rodriguez-Galiano et al., 2012). VI can be calculated using the Gini index or OOB samples (Belgiu & Drăgut, 2016). The Gini index is a measure of the impurity of the nodes in the tree.

c. Deep Learning and Neural Networks

The term “deep learning” is used here to refer to a subset of machine learning that uses multiple processing layers to automatically learn hierarchical representations of input data necessary for a classification or detection task (Lecun et al., 2015). Deep learning uses neural network architectures, often called deep neural networks.

The architecture of a neural network is composed of three main layers: an input layer, one or more hidden layers, and an output layer. It is graphically represented in Figure 1.9, wherein the leftmost layer (column) is the input layer, the rightmost layer is the output layer, and the middle layer is called a hidden layer (Nielsen, 2015). The network shown in Figure 1.9 has only one hidden layer, but most modern neural networks have more than one hidden layer (Qu et al., 2016; Siddique et al., 2019).

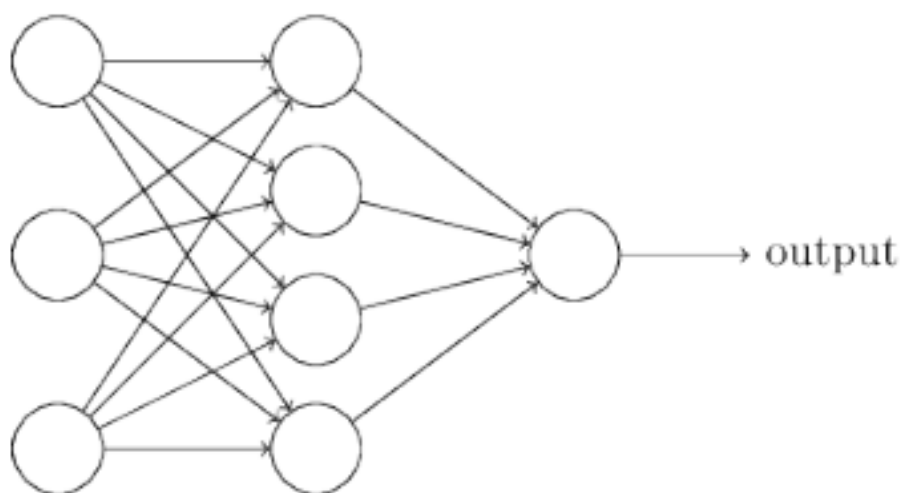


Figure 1.9. Three main layers of a neural network – an input layer, one (or more) hidden layers, and an output layer. Source: Nielsen (2015)

Each of these layers can have a different number of nodes, represented by small circles and arranged in a column representing that layer (Figure 1.9). The lines between the nodes indicate the flow of information from one node to the next. In this particular neural network example (Figure 1.9), the information flows from the input layer (left) to the output layer (right). In other words, the output from the previous layer is used as an input to the next layer. This network is called a feedforward neural network, i.e., information only moves forward, and no loops exist in the network. Neural networks that allow feedback loops—information can travel in both directions—are called recurrent neural networks (Nielsen, 2015).

Each input node in a layer is connected to every node in the next layer, wherein each connection has a particular weight (a trainable parameter) (Figure 1.10). This weight is multiplied by the input node value, and the products are summed to produce a single value. The node in the next layer then receives the sum of all its weighted inputs, $\sum x_i w_i$ where x_i is the input and w_i is the weight, plus a bias term θ (another trainable parameter). After computing the weighted sum, an activation function, usually nonlinear ϕ , is applied to the result. This function determines whether or not the node is activated (Burkov, 2019; Muller & Guido, 2017). This process is illustrated in Figure 1.10.

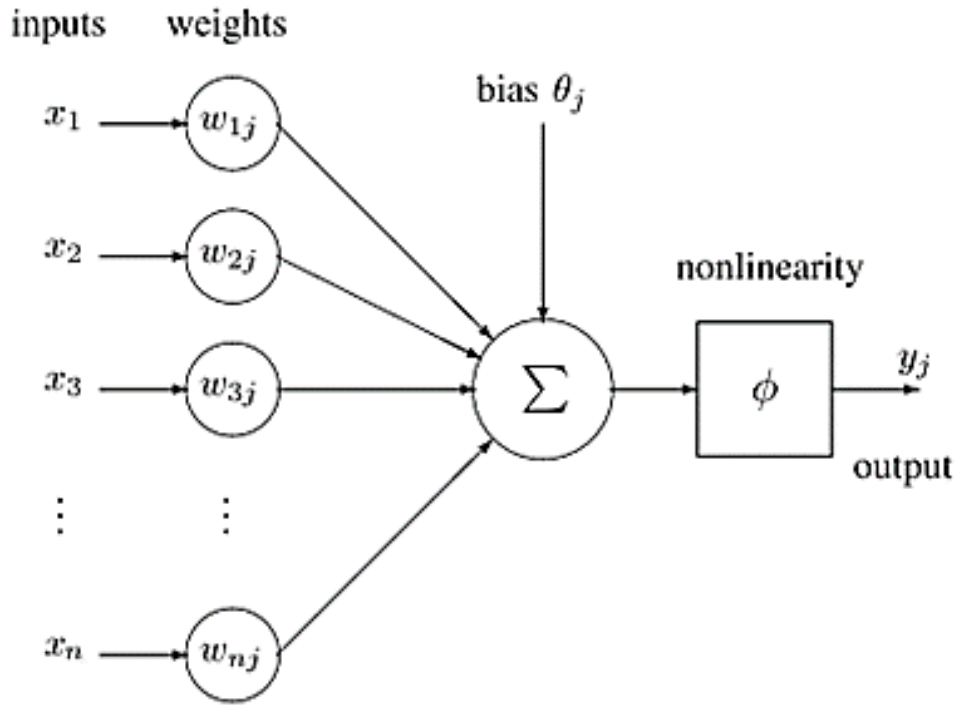


Figure 1.10. Neural network architecture (Source: Castelluccio et al. (2015))

In contrast to classical neural networks that can have only a few hidden layers, deep neural networks can have tens or even hundreds of hidden layers between the input and output layers, hence the term “deep”. The aim of a deep neural network is to solve a complicated problem by extracting increasingly abstract features, i.e., the initial layers learn to recognise simple visual features (low level), and the final layers build up a set of more complex and abstract concepts (high level). For instance, the first hidden layer of a deep neural network could recognise the edges. The second hidden layer could recognise basic shapes, while the third hidden layer could learn more complex shapes, and so on (Nielsen, 2015).

Developing a robust deep learning model usually requires a substantial and diverse set of labelled training samples to prevent overfitting and improve its generalisation ability (Gütter et al., 2022; Hu et al., 2015; Scott et al., 2017). However, this requirement presents a significant challenge, especially for remote sensing tasks, as there are limited labelled samples available (Li et al., 2018). This scarcity of samples could be due to the data being difficult or costly to collect and label or the data being not found very often. To address

this challenge, transfer learning and data augmentation are techniques that can be used to generate a well-generalised model with limited samples.

Transfer learning. Transfer learning is a technique that involves taking a deep learning model that has been previously trained to solve a specific task and then using it to solve other related tasks that have limited available data (Naushad & Kaur, 2021; Scott et al., 2017). This pre-trained model is developed using a huge amount of training data. Examples are models that are trained on the ImageNet dataset, which contains over 14 million annotated natural images with 1000 classes, such as animals, vehicles and fruits (Deng et al., 2009). Despite being trained on non-medical image data, these models have performed well in medical image analysis through transfer learning (Bar et al., 2015; Horry et al., 2020). Moreover, models that use transfer learning usually outperform models that are trained from scratch (de Lima & Marfurt, 2019; Yosinski et al., 2014).

Data augmentation. Data augmentation is a technique that can be used to generate additional and more diverse training data, which helps create a robust deep learning model (Yu et al., 2017). This technique involves applying random transformations, such as flips, translations, and rotations, to the original training data to increase the number and variability of data (Scott et al., 2017). Models trained on augmented data generally achieve higher accuracy than those trained on the original dataset (Yu et al., 2017).

1.4. Remote sensing and machine learning for habitat mapping and monitoring

Over the last few decades, remote sensing has contributed to nature conservation owing to its ability to serve as a data source for obtaining valuable habitat information (e.g., extent, location and size), ranging from local to regional scale (Cavender-Bares et al., 2022; Corbane et al., 2015; Kerr & Ostrovsky, 2003; Vanden Borre, Paelinckx, et al., 2011). Satellite remote sensing has revolutionised large-scale habitat monitoring by providing repeated spatially comprehensive data over broad geographical extents (Ali et al., 2016; Cawkwell et al., 2018; Lucas et al., 2007, 2011; Nagendra et al., 2013). These data, typically recorded in an image format, are used to delineate and record the extent

and conditions of habitats. Moreover, frequent observations are important to document how habitats change over time and to identify the drivers of these changes (Horning et al., 2010). However, acquiring usable satellite data can be problematic for countries with persistent cloud cover, such as Ireland, especially if the acquisition date is crucial for the analysis (e.g., monitoring vegetation phenology) (Horning et al., 2010; Wang & Moskovits, 2001). With the scheduled acquisitions of satellites, images may also not be available for certain locations or times of the year (Connolly, 2018). Hence, an alternative is needed, a more suitable platform capable of flexible image acquisition.

Freely available satellite images with coarse spatial resolution are often applied to mapping broad spatial extents and distributions of relatively homogeneous habitats (Corbane et al., 2015). However, the use of these images can become more challenging when mapping habitats in complex and heterogeneous environments, such as uplands (Cingolani et al., 2004; O’Connell et al., 2014). In most cases, small habitat patches and narrow linear features, such as waterways, are left unmapped; this limits the applicability of free satellite images in capturing fine-scale characterisation of habitats (Le Dez et al., 2021; Murray et al., 2022). Moreover, missed or underestimated detections of small-sized degradations or disturbances may lead to false conclusions about the actual status of habitats (Tian et al., 2017). Accurate and detailed mapping of habitats is crucial when conducting site-specific habitat assessments, as well as understanding many local-scale ecological investigations (Anderson & Gaston, 2013). Therefore, higher-resolution images are desired.

Technological advances in remote sensing platforms and sensors, providing higher spatial and temporal resolutions, have become integral in habitat monitoring (Boyle et al., 2014). Previous studies have used commercial satellite images—such as those from RapidEye, QuickBird and Ikonos missions—for mapping habitats, including grasslands (T. Schmidt et al., 2014), coastal dunes (Timm & McGarigal, 2012) and uplands (Mehner et al., 2004). While acquiring these images can often be cost-prohibitive, they are useful, as they can provide fine-scale information about the landscape. Nevertheless, future access to high-resolution satellite imagery ($< 1\text{m}$) at low (or no) cost would be highly beneficial to many environmental monitoring applications (Boyle et al., 2014). For example, the information derived from this imagery can be used for more accurate, precise and detailed assessments of habitat dynamics, ecosystem health and climate change impacts. Such assessments can

improve our understanding of various ecological and environmental processes at regional or global scales.

Another advancement in remote sensing is Unoccupied Aerial Vehicles (UAVs) or drones.⁴ A UAV is an aircraft with no pilot on board and which can either be flown based on autonomously programmed flight plans or is remotely controlled by a human (Pajares, 2015). While UAVs cannot compete with satellites in terms of the spatial coverage of the acquired images (a few square kilometres for UAVs vs up to thousands of square kilometres for satellites), there are multiple advantages to using UAVs when it comes to habitat mapping and monitoring. They can be flown more frequently based on user-defined parameters, even under overcast or cloudy conditions. Equipped with appropriate sensors, UAVs can be used to acquire high-resolution (centimetre level) images, providing valuable data for conservationists (Johnston, 2019; Koh & Wich, 2012). These data can be used in mapping heterogeneous and complex habitats (Fraser et al., 2016) and detecting small-sized and early stages of habitat degradation, such as the introduction of invasive species (Marzioletti et al., 2021). This increase in the spatial resolution of images can also help improve our ability to map vegetation and soil/substrate separately using various spectral indices. In contrast, separating these components can be challenging with coarser spatial resolution imagery (>1 m), as individual pixels may contain a mixture of soil, vegetation and other features. Nevertheless, advanced image analysis techniques like spectral unmixing can provide estimates of the proportions of each component in every pixel (e.g., He et al., 2020 and Laamrani et al., 2020).

UAVs can also be rapidly deployed to monitor and assess specific sites (e.g., remote areas in uplands and dynamic coastal areas) (Jayson-Quashigah et al., 2019; Scholefield et al., 2019). Moreover, collecting UAV images can be scheduled at the time corresponding to key phenological stages when there will be an increase in the likelihood of distinguishing different vegetation, such as during the flowering or senescence period (Cruzan et al.,

⁴ This thesis uses a gender-neutral term “unoccupied”.

Johnston, D. W. (2019). Unoccupied aircraft systems in marine science and conservation. *Annual Review of Marine Science*, 11, 439–463. <https://doi.org/10.1146/annurev-marine-010318-095323>.

Joyce, K., Anderson, K., & Bartolo, R. (2021). Of course we fly unmanned—we’re women! *Drones*, 5(1), 21. <https://doi.org/10.3390/drones5010021>

2016; Nagendra et al., 2013; Wood et al., 2022). These developments have undoubtedly offered ecologists and conservation site managers a promising avenue for detailed and timely monitoring of habitat conditions, improving the understanding of spatiotemporal variation of habitat distributions (Anderson & Gaston, 2013). The increasing spatial and temporal resolutions of data collected by satellite sensors, the growing use of UAVs and the amount of data they produce have driven developments in advanced image processing techniques, including machine learning (Chi et al., 2016).

In traditional pixel-based methods, remote sensing images are classified through thresholding, likelihood and clustering, which are generally based on the spectral information of individual pixels (Campbell & Wynne, 2011; Lillesand et al., 2015). However, this approach to classifying images may not be sufficient for discriminating between classes with similar spectral characteristics, such as vegetation types. Advanced image analysis methods, such as machine learning, are required. These methods facilitate the automation of extracting complex patterns, spatial context and relationships between neighbouring pixels from large datasets, allowing for efficient data analysis. Previous UAV studies have shown that machine-learning methods outperformed traditional pixel-based methods for vegetation mapping (Feng et al., 2015; Reis et al., 2019; X. Wang et al., 2021). They also improve previous techniques of manually analysing the collected data (e.g., digitising habitat extents on aerial photographs), which are often resource-intensive and may be subject to bias (Hamylton et al., 2020). With appropriate training data and generalisation techniques, machine learning-based models can also be applied to classify a new dataset acquired at different sites. In other words, machine learning methods offer significant improvements (i.e., higher accuracy, flexibility and efficiency) over the manual and traditional pixel-based methods in analysing remote sensing data. For these reasons, machine learning has been widely implemented in many applications, such as land cover mapping (Gavish et al., 2018; Rodriguez-Galiano et al., 2012), change detection (Zhu et al., 2018), urban monitoring (Schug et al., 2018) and wildlife conservation (Tuia et al., 2022).

Despite the considerable potential of combining high-resolution remote sensing data and machine learning techniques to support environmental conservation efforts, these advanced mapping technologies remain underexploited for monitoring protected habitats under the HD, including in Ireland (Vanden Borre, Haest, et al., 2011). These habitats—

listed in Annex I of HD—are considered threatened and endangered in the EU (National Parks and Wildlife Service, 2019). Therefore, it is crucial to obtain comprehensive information on the distributions and conditions of these habitats. Such information is an essential prerequisite in making data-informed decisions about which conservation measures can be applied (Vanden Borre, Haest, et al., 2011). The increasing need for timely, high-quality habitat maps raises the question of the potential utilisation of combined high-resolution remote sensing data and machine learning techniques in capturing fine-scale distributions of diverse habitat types over time. This question forms the focus of the thesis.

1.5. Aims of the thesis

The overarching aim of this research is to develop robust and repeatable methodologies for mapping and monitoring habitats using high-resolution remote sensing data and machine learning techniques. A particular focus lies on coastal (Aims 1 and 2) and upland habitats (Aim 3). These habitats represent Irish habitat types that are biodiverse and of high conservation importance. They also contain several examples of Annex I habitats (National Parks and Wildlife Service, 2019). However, their status shows strong deteriorating trends, as published in the recent habitat assessment reports in Ireland (National Parks and Wildlife Service, 2019) and in the EU (European Environment Agency, 2020). The habitat type names used in this thesis are consistent with the published *Interpretation manual of European Union habitats* (European Commission, 2013), which means that the results can be integrated into the existing monitoring initiatives of Ireland under the HD.

Monitoring habitats with remote sensing and machine learning has developed rapidly in the last decade, but some key challenges remain to which this thesis can contribute. First, monitoring dynamic coastal dune systems with freely available satellite imagery of coarse spatial resolutions (>10 m) can be difficult, e.g., when mapping very small areas of patchy dune habitats or identifying small-sized disturbed areas within these habitats caused by human tracks. This may result in inaccurate spatial extent and botanical information about the dune habitats. Additionally, since plant communities within a habitat can exhibit

varying colours during different periods of the year, determining the optimal timing for image acquisition to produce highly accurate maps of coastal dune habitats presents a challenge. Within this context, Chapter 2 of this thesis focuses on using high-resolution UAV data (6 cm) and a machine-learning technique for coastal dune habitat mapping. Specific notable advantages of UAV data, such as the flexibility in its image acquisition and the availability of UAV-based topographic information, were exploited to determine if and how they can improve classification accuracy.

Aim 1: To assess the effectiveness of UAV remote sensing and machine learning for mapping coastal dune habitats.

Second, monitoring the status of habitats entails the identification of threats to and pressures on habitats, which includes determining the location and spatial extent of invasive plant species. Early detection of these species, which typically occur in small patches, is crucial to prevent further spread. The utilisation of submetre-resolution UAV imagery is necessary to capture such information accurately. In that way, in areas which are monitored and invasive species are detected, mitigation measures can be applied at the earliest possible stage. *Spartina anglica*, commonly found in Irish saltmarsh and mudflat habitats, is considered one of the most problematic invasive species in the EU and globally; hence, identifying and tracking its expansion is crucial. However, there are currently very limited methods for detailed mapping of the extent of invasive species in Ireland. The second aim of this thesis is to explore the application of deep learning on UAV images to map the detailed spatial distributions of *Spartina anglica*.

Aim 2: To develop a deep learning model to automatically map the spatial distribution of a coastal invasive species *Spartina anglica* on UAV images.

Third, UAV surveying for mapping habitats in vast, rugged terrains of uplands may not be practical due to the limited area coverage of these surveys, usually a few square kilometres. Previous studies on mapping upland habitats have focused on using medium-resolution satellite imagery, typically ranging from 10 to 30 metres/pixel, and the conventional crisp classification technique, where a pixel is associated with a single class.

However, the spatial resolutions of these images as well as the processing technique used can create sharp boundaries that do not accurately represent the typical gradual transitions and complex mosaics between upland habitats. Therefore, a different mapping approach needs to be considered. The third aim is to explore the use of high-resolution (2 m) satellite imagery and machine learning for detailed mapping of upland habitats over large areas.

Aim 3: To utilise high-resolution satellite imagery and machine learning-based fuzzy classification techniques for mapping complex upland habitat types.

The knowledge gained on investigating different proposed methodologies to address the three research aims outlined above will be summarised and linked in the Synthesis chapter, thus creating a broader perspective on the use of advanced mapping technologies in monitoring habitats. It is important to note that while these methodologies focus on the coastal and upland habitats, they are broadly applicable to other habitats and other target plant species in both Ireland and elsewhere.

Overall, the methodologies to be presented in this thesis could contribute to the monitoring of protected habitats by providing valuable information on their spatiotemporal distributions, including on areas impacted by human activities. Such information can guide the conservation and management of these habitats. The maps produced from these methodologies can also be used to support monitoring initiatives, thereby achieving the high standards for nature conservation mandated by the EU.

1.6. Structure of the thesis

The thesis is comprised of five chapters. Enclosed by the introduction (Chapter 1) and synthesis (Chapter 5), Chapters 2, 3 and 4 present published and under review manuscripts covering methodologies for machine learning-based mapping and monitoring of habitats using data acquired from UAVs (Chapters 2 and 3) and satellites (Chapter 4). Each of these three core chapters (Chapters 2-4) is a standalone work with

its own introduction, objectives, materials and methods, results, discussion and conclusion sections. Hence, there are overlaps between these chapters in terms of the background of the study, related works and data collection methods. The authors' contributions are explicitly stated at the beginning of Chapters 2, 3 and 4. A brief description of each chapter is given below.

- Chapter 1** provides an overall background and motivation of the thesis, principles of remote sensing and machine learning technologies, followed by the three research aims to be covered in Chapters 2-4.
- Chapter 2** is a peer-reviewed article published in the *European Journal of Remote Sensing* (Cruz et al., 2023).⁵ This chapter assessed the effectiveness of using machine learning and UAV images acquired at different times during the growing season in mapping coastal dune habitats (Aim 1). The results presented in this chapter highlight the significant importance of using UAV-derived topography and multiple UAV image sets acquired throughout the growing season to improve accuracy in classifying dune habitats. Moreover, the results suggest an optimal time of the growing season for UAV data acquisition to differentiate dune habitats. The robust methodology in this chapter uses a low-cost multi-spectral UAV system and open-source tools (i.e., Python packages). Therefore, it can be adapted for use to map other habitat types.
- Chapter 3** is a peer-reviewed article published in the *International Journal of Remote Sensing* (Cruz et al., 2023).⁶ This chapter developed an effective deep learning-based segmentation model applied to UAV images for fine-scale mapping of coastal invasive species (Aim 2). The

⁵ Cruz, C., O'Connell, J., McGuinness, K., Martin, J. R., Perrin, P. M., & Connolly, J. (2023). Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping. *European Journal of Remote Sensing*, 56(1), 2191870. <https://doi.org/10.1080/22797254.2023.2191870>

⁶ Cruz, C., McGuinness, K., Perrin, P. M., O'Connell, J., Martin, J. R., & Connolly, J. (2023). Improving the mapping of coastal invasive species using UAV imagery and deep learning. *International Journal of Remote Sensing*, 44(18), 5713–5735. <https://doi.org/10.1080/01431161.2023.2251186>

results of this chapter proposed a method to improve the monitoring of invasive species. With this method, the spatiotemporal mapping of invasive species distribution could be substantially automated. Moreover, the method and Python code generated from this chapter could be used to further improve the generalisation ability of the developed model. It could also be adapted for mapping other target plant species.

Chapter 4 is an article currently under review in the *Environmental Monitoring and Assessment* journal (Cruz et al., under review).⁷ This chapter demonstrates the potential of high-resolution satellite imagery and the fuzzy classification technique in mapping the distribution of upland habitats, which are often heterogeneous and occur in mosaics (Aim 3). The results of this chapter contribute to improving the conventional mapping technique of upland habitats to better represent their complex distributions, as observed in the field.

Chapter 5 summarises the key findings in Chapters 2-4 and critically evaluates the use of the methodologies developed in this thesis. This chapter also provides implications of the findings in the broader context of habitat mapping and monitoring research. It concludes by outlining overall conclusions and recommendations for future work.

All supplementary information is presented in the Appendices section.

1.7. References

Aggarwal, S. (2004). Principles of remote sensing. In M. V. K. Sivakumar, P. S. Roy, K. Harmsen, & S. K. Saha (Eds.), *Satellite remote sensing and GIS applications in agricultural meteorology* (pp. 23–38).

⁷ Cruz, C., Perrin, P. M., Martin, J. R., O’Connell, J., McGuinness, K., & Connolly, J. Mapping of temperate upland habitats using high-resolution satellite imagery and machine learning. (Under review).

- Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., & Green, S. (2016). Satellite remote sensing of grasslands: From observation to management. *Journal of Plant Ecology*, 9(6), 649–671. <https://doi.org/10.1093/jpe/rtw005>
- Alpaydin, E. (2004). *Introduction to machine learning*. The MIT Press.
- Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, 11(3), 138–146. <https://doi.org/10.1890/120150>
- Asner, G. P., Martin, R. E., Anderson, C. B., & Knapp, D. E. (2015). Quantifying forest canopy traits: Imaging spectroscopy versus field survey. *Remote Sensing of Environment*, 158, 15–27. <https://doi.org/10.1016/j.rse.2014.11.011>
- Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., & Greenspan, H. (2015). Chest pathology detection using deep learning with non-medical training. *Proceedings - 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 294–297. <https://doi.org/10.1109/ISBI.2015.7163871>
- Barrett, E. C., & Curtis, L. F. (1999). *Introduction to environmental remote sensing* (4th ed.). Routledge. <https://doi.org/10.4324/9780203761038>
- Belgiu, M., & Drăgut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>
- Boyle, S. A., Kennedy, C. M., Torres, J., Colman, K., Pérez-Estigarribia, P. E., & De La Sancha, N. U. (2014). High-resolution satellite imagery is an important yet underutilized resource in conservation biology. *PLoS ONE*, 9(1), e86908. <https://doi.org/10.1371/journal.pone.0086908>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st ed.). CRC Press.
- Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., & Hilton-Taylor, C. (2002). Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, 16(4), 909–923. <https://doi.org/10.1046/j.1523-1739.2002.00530.x>
- Brophy, J. T., Perrin, P. M., Penk, M. R., Devaney, F. M., & Leyden, K. J. (2019). *Saltmarsh monitoring project 2017-2018 (Irish Wildlife Manuals No. 108)*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland. <http://hdl.handle.net/2262/89539>
- Buchanan, G., Pearce-Higgins, J., Grant, M., Robertson, D., & Waterhouse, T. (2005). Characterization of moorland vegetation and the prediction of bird abundance using remote sensing. *Journal of Biogeography*, 32(4), 697–707. <https://doi.org/10.1111/j.1365-2699.2004.01187.x>
- Burkov, A. (2019). *The hundred-page machine learning book*. Andriy Burkov.
- Campbell, J., & Wynne, R. (2011). *Introduction to remote sensing* (5th ed.). The Guilford Press.
- Carbonneau, P. E., & Lane, S. N. (2004). Catchment-scale mapping of surface grain size in gravel bed rivers using airborne digital imagery. *Water Resources Research*, 40(7), W07202. <https://doi.org/10.1029/2003WR002759>
- Castelluccio, M., Poggi, G., Sansone, C., & Verdoliva, L. (2015). *Land use classification in remote sensing images by Convolutional Neural Networks*. Available online: <http://arxiv.org/abs/1508.00092> (accessed on 13 May 2020)
- Cavender-Bares, J., Schneider, F. D., Santos, M. J., Armstrong, A., Carnaval, A., Dahlin, K. M., Fatoyinbo, L., Hurtt, G. C., Schimel, D., Townsend, P. A., Ustin, S. L., Wang, Z., & Wilson, A. M. (2022). Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology and Evolution*, 6(5), 506–519. <https://doi.org/10.1038/s41559-022-01702-5>

- Cawkwell, F., Raab, C., Barrett, B., Green, S., & Finn, J. (2018). *TaLAM: Mapping land cover in lowlands and uplands with satellite imagery*. Environmental Protection Agency, Ireland.
https://www.epa.ie/publications/research/waste/Research_Report_254.pdf
- Cherrill, A. (2016). Inter-observer variation in habitat survey data: investigating the consequences for professional practice. *Journal of Environmental Planning and Management*, 59(10), 1813–1832. <https://doi.org/10.1080/09640568.2015.1090961>
- Cherrill, A., & McClean, C. (1995). An investigation of uncertainty in field habitat mapping and the implications for detecting land cover change. *Landscape Ecology*, 10(1), 5–21. <https://doi.org/10.1007/BF00158550>
- Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207–2219. <https://doi.org/10.1109/JPROC.2016.2598228>
- Cingolani, A. M., Renison, D., Zak, M. R., & Cabido, M. R. (2004). Mapping vegetation in a heterogeneous mountain rangeland using Landsat data: An alternative method to define and classify land-cover units. *Remote Sensing of Environment*, 92(1), 84–97. <https://doi.org/10.1016/j.rse.2004.05.008>
- Commission of the European Communities. (1992). Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. *Official Journal of the European Union*. 206, 7-50.
- Connolly, J. (2018). Mapping land use on Irish peatlands using medium resolution satellite imagery. *Irish Geography*, 51(2), 187–204. <https://irishgeography.ie/index.php/irishgeography/article/view/1371>
- Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., Strasser, T., Vanden Borre, J., Toon, S., & Michael, F. (2015). Remote sensing for mapping natural habitats and their conservation status - New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 37(2015), 7–16. <https://doi.org/10.1016/j.jag.2014.11.005>
- Cruzan, M. B., Weinstein, B. G., Grasty, M. R., Kohn, B. F., Hendrickson, E. C.,

- Arredondo, T. M., & Thompson, P. G. (2016). Small unmanned aerial vehicles (micro-UAVs, drones) in plant ecology. *Applications in Plant Sciences*, 4(9), 1600041. <https://doi.org/10.3732/apps.1600041>
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 157–175). Springer Science+Business Media, LLC 2012. https://doi.org/10.1007/978-1-4419-9326-7_5
- Dahan, H., Cohen, S., Rokach, L., & Maimon, O. (2014). *Proactive data mining with decision trees*. SpringerBriefs in Electrical and Computer Engineering. https://doi.org/10.1007/978-1-4939-0539-3_1
- Davies, C. E., Moss, D., & Hill, M. O. (2004). *EUNIS Habitat Classification Revised 2004*. European Environment Agency and European Topic Centre on Nature Protection and Biodiversity. <https://www.eea.europa.eu/data-and-maps/data/eunis-habitat-classification-1/documentation/eunis-2004-report.pdf>
- de Jong, S. M., van der Meer, F. D., & Clevers, J. G. P. W. (2004). Basics of remote sensing. In S. M. De Jong & F. D. Van Der Meer (Eds.), *Remote sensing image analysis—including spatial domain* (pp. 1–15). Springer Science + Business Media, Inc. https://doi.org/10.1111/j.1477-9730.2006.00359_4.x
- de Lima, R., & Marfurt, K. (2019). Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing*, 12(86). <https://doi.org/10.3390/rs12010086>
- de Pinho, C. M. D., Fonseca, L. M. G., Korting, T. S., de Almeida, C. M., & Kux, H. J. H. (2012). Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis. *International Journal of Remote Sensing*, 33(19), 5973–5995. <https://doi.org/10.1080/01431161.2012.675451>
- Delaney, A., Devaney, F. M., Martin, J. M., & Barron, S. J. (2013). *Monitoring survey of Annex I sand dune habitats in Ireland (Irish Wildlife Manuals No. 75)*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland. <https://www.npws.ie/sites/default/files/publications/pdf/IWM75.pdf>

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.
<https://doi.org/10.1109/CVPR.2009.5206848>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36.
<https://doi.org/10.1016/j.rse.2011.11.026>
- Elachi, C., & Van Zyl, J. (2006). *Introduction to the physics and techniques of remote sensing* (2nd ed.). John Wiley & Sons.
- European Commission. (2013). *Interpretation manual of European Union habitats EUR 28*.
http://ec.europa.eu/environment/nature/legislation/habitatsdirective/docs/Int_Manual_EU28.pdf
- European Environment Agency. (2020). *State of nature in the EU: Results from reporting under the nature directives 2013-2018*. Publications Office,
<https://data.europa.eu/doi/10.2800/088178>
- Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: From early developments to recent advancements. *Systems Science and Control Engineering: An Open Access Journal*, 2(1), 602–609.
<https://doi.org/10.1080/21642583.2014.956265>
- Feng, Q., Liu, J., & Gong, J. (2015). UAV remote sensing for urban vegetation mapping using Random Forest and texture analysis. *Remote Sensing*, 7, 1074–1094.
<https://doi.org/10.3390/rs70101074>
- Fraser, R. H., Olthof, I., Lantz, T. C., & Schmitt, C. (2016). UAV photogrammetry for mapping vegetation in the low-Arctic. *Arctic Science*, 2(3), 79–102.
<https://doi.org/10.1139/as-2016-0008>
- Gavish, Y., O’Connell, J., Marsh, C. J., Tarantino, C., Blonda, P., Tomaselli, V., &

- Kunin, W. E. (2018). Comparing the performance of flat and hierarchical habitat/land-cover classification models in a NATURA 2000 site. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136, 1–12.
<https://doi.org/10.1016/j.isprsjprs.2017.12.002>
- Green, E. P., Edwards, A. J., Clark, C. D., & Mumby, P. J. (1999). The cost-effectiveness of remote sensing for tropical coastal resources assessment and management. *Journal of Environmental Management*, 55, 157–166.
<https://doi.org/10.1006/jema.1998.0255>
- Gütter, J., Kruspe, A., Zhu, X. X., & Niebling, J. (2022). Impact of training set size on the ability of deep neural networks to deal with omission noise. *Frontiers in Remote Sensing*, 3, 932431. <https://doi.org/10.3389/frsen.2022.932431>
- Hamylton, S. M., Morris, R. H., Carvalho, R. C., Roder, N., Barlow, P., Mills, K., & Wang, L. (2020). Evaluating techniques for mapping island vegetation from unmanned aerial vehicle (UAV) images: Pixel classification, visual interpretation and machine learning approaches. *International Journal of Applied Earth Observation and Geoinformation*, 89, 102085.
<https://doi.org/10.1016/j.jag.2020.102085>
- He, Y., Yang, J., & Guo, X. (2020). Green vegetation cover dynamics in a heterogeneous grassland: Spectral unmixing of Landsat time series from 1999 to 2014. *Remote Sensing*, 12(22), 3826. <https://doi.org/10.3390/rs12223826>
- Horning, N., Robinson, J. A., Sterling, E. J., Turner, W., & Spector, S. (2010). *Remote sensing for ecology and conservation: A handbook of techniques*. Oxford University Press Inc. <https://doi.org/10.1093/oso/9780199219940.001.0001>
- Horry, M. J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., & Shukla, N. (2020). COVID-19 detection through transfer learning using multimodal imaging data. *IEEE Access*, 8, 149808–149824.
<https://doi.org/10.1109/ACCESS.2020.3016780>
- Hu, F., Xia, G. S., Hu, J., & Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11), 14680–14707. <https://doi.org/10.3390/rs71114680>

- Im, J., Quackenbush, L. J., Li, M., & Fang, F. (2014). Optimum scale in object-based image analysis. In Q. Weng (Ed.), *Scale Issues in Remote Sensing* (First edition, pp. 197–214). John Wiley & Sons, Inc.
<https://doi.org/10.1002/9781118801628.ch10>
- Jayson-Quashigah, P. N., Appeaning Addo, K., Amisigo, B., & Wiafe, G. (2019). Assessment of short-term beach sediment change in the Volta Delta coast in Ghana using data from Unmanned Aerial Vehicles (Drone). *Ocean and Coastal Management*, 182, 104952. <https://doi.org/10.1016/j.ocecoaman.2019.104952>
- Jensen, J. R. (2014). *Remote sensing of the environment: an earth resource perspective* (2nd ed.). Pearson Education Limited.
- Johnston, D. W. (2019). Unoccupied aircraft systems in marine science and conservation. *Annual Review of Marine Science*, 11, 439–463.
<https://doi.org/10.1146/annurev-marine-010318-095323>
- Kennedy, R. E., Andréfouët, S., Cohen, W. B., Gómez, C., Griffiths, P., Hais, M., Healey, S. P., Helmer, E. H., Hostert, P., Lyons, M. B., Meigs, G. W., Pflugmacher, D., Phinn, S. R., Powell, S. L., Scarth, P., Sen, S., Schroeder, T. A., Schneider, A., Sonnenschein, R., ... Zhu, Z. (2014). Bringing an ecological view of change to Landsat-based remote sensing. *Frontiers in Ecology and the Environment*, 12(6), 339–346. <https://doi.org/10.1890/130066>
- Kerr, J. T., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends in Ecology and Evolution*, 18(6), 299–305.
[https://doi.org/10.1016/S0169-5347\(03\)00071-5](https://doi.org/10.1016/S0169-5347(03)00071-5)
- Koh, L. P., & Wich, S. A. (2012). Dawn of drone ecology: Low-cost autonomous aerial vehicles for conservation. *Tropical Conservation Science*, 5(2), 121–132.
<https://doi.org/10.1177/194008291200500202>
- Laamrani, A., Joosse, P., McNairn, H., Berg, A. A., Hagerman, J., Powell, K., & Berry, M. (2020). Assessing soil cover levels during the non-growing season using multitemporal satellite imagery and spectral unmixing techniques. *Remote Sensing*, 12(9), 1397. <https://doi.org/10.3390/RS12091397>

- Le Dez, M., Robin, M., & Launeau, P. (2021). Contribution of Sentinel-2 satellite images for habitat mapping of the Natura 2000 site ‘Estuaire de la Loire’ (France). *Remote Sensing Applications: Society and Environment*, 24, 100637. <https://doi.org/10.1016/j.rsase.2021.100637>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lengyel, S., Déri, E., Varga, Z., Horváth, R., Tóthmérész, B., Henry, P. Y., Kobler, A., Kutnar, L., Babij, V., Seliškar, A., Christia, C., Papastergiadou, E., Gruber, B., & Henle, K. (2008). Habitat monitoring in Europe: A description of current practices. *Biodiversity and Conservation*, 17, 3327–3339. <https://doi.org/10.1007/s10531-008-9395-3>
- Li, Y., Zhang, H., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6), 1–17. <https://doi.org/10.1002/widm.1264>
- Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). *Remote sensing and image interpretation* (7th ed.). John Wiley & Sons.
- Liu, D., & Xia, F. (2010). Assessing object-based classification: Advantages and limitations. *Remote Sensing Letters*, 1(4), 187–194. <https://doi.org/10.1080/01431161003743173>
- Liu, P. (2015). A survey of remote-sensing big data. *Frontiers in Environmental Science*, 3. <https://doi.org/10.3389/fenvs.2015.00045>
- Lucas, R., Medcalf, K., Brown, A., Bunting, P., Breyer, J., Clewley, D., Keyworth, S., & Blackmore, P. (2011). Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), 81–102. <https://doi.org/10.1016/j.isprsjprs.2010.09.004>
- Lucas, R., Rowlands, A., Brown, A., Keyworth, S., & Bunting, P. (2007). Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(3), 165–185. <https://doi.org/10.1016/j.isprsjprs.2007.03.003>

- Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51, 47–60.
<https://doi.org/10.1016/j.future.2014.10.029>
- Marzioletti, F., Frate, L., De Simone, W., Frattaroli, A. R., Acosta, A. T. R., & Carranza, M. L. (2021). Unmanned Aerial Vehicle (UAV)-based mapping of *Acacia Saligna* invasion in the Mediterranean Coast. *Remote Sensing*, 13(17), 3361. <https://doi.org/10.3390/rs13173361>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817.
<https://doi.org/10.1080/01431161.2018.1433343>
- Mehner, H., Cutler, M., Fairbairn, D., & Thompson, G. (2004). Remote sensing of upland vegetation: The potential of high spatial resolution satellite sensors. *Global Ecology and Biogeography*, 13(4), 359–369. <https://doi.org/10.1111/j.1466-822X.2004.00096.x>
- Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing*, 7(7), 8489–8515.
<https://doi.org/10.3390/rs70708489>
- Millennium Ecosystem Assessment. (2003). *Ecosystems and human well-being: a framework for assessment*. Island Press.
- Muller, A. C., & Guido, S. (2017). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media Inc.
- Murray, N. J., Worthington, T. A., Bunting, P., Duce, S., Hagger, V., Lovelock, C. E., Lucas, R., Saunders, M. I., Sheaves, M., Spalding, M., Waltham, N. J., & Lyons, M. B. (2022). High-resolution mapping of losses and gains of Earth's tidal wetlands. *Science*, 376(6594), 744–749. <https://doi.org/10.1126/science.abm9583>
- Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M.,

- & Mairota, P. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, 33, 45–59. <https://doi.org/10.1016/j.ecolind.2012.09.014>
- National Parks and Wildlife Service. (2019). *The status of EU protected habitats and species in Ireland. Volume 1: Summary overview*. Department of Culture, Heritage and the Gaeltacht, Ireland..
- Naushad, R., & Kaur, T. (2021). Deep transfer learning for land use and land cover classification: A comparative study. *Sensors*, 21(23), 8083. <https://doi.org/10.3390/s21238083>
- Nielsen, M. (2015). *Neural Networks and Deep Learning*. <http://neuralnetworksanddeeplearning.com>
- Nolan, M., Larsen, C., & Sturm, M. (2015). Mapping snow depth from manned aircraft on landscape scales at centimeter resolution using structure-from-motion photogrammetry. *The Cryosphere*, 9, 1445–1463. <https://doi.org/10.5194/tc-9-1445-2015>
- O’Connell, J., Connolly, J., & Holden, N. M. (2014). A monitoring protocol for vegetation change on Irish peatland and heath. *International Journal of Applied Earth Observation and Geoinformation*, 31, 130–142. <https://doi.org/10.1016/j.jag.2014.03.006>
- O’Neill, F. H., Martin, J. R., Devaney, F. M., & Perrin, P. M. (2013). *The Irish semi-natural grasslands survey 2007-2012 (Irish Wildlife Manuals No. 78)*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland. <https://www.npws.ie/sites/default/files/publications/pdf/IWM-78-Irish-semi-natural-grassland-survey.pdf>
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? In: Perner P. (Eds) *Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Lecture Notes in Computer Science*, 7376, 154–168. https://doi.org/10.1007/978-3-642-31537-4_13
- Pajares, G. (2015). Overview and current status of remote sensing applications based on

- unmanned aerial vehicles (UAVs). *Photogrammetric Engineering and Remote Sensing*, 81(4), 281–330. <https://doi.org/10.14358/PERS.81.4.281>
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222.
<https://doi.org/10.1080/01431160412331269698>
- Perrin, P. M., Barron, S. J., Roche, J. R., & O’Hanrahan, B. (2014). *Guidelines for a national survey and conservation assessment of upland vegetation and habitats in Ireland, Version 2.0 (Irish Wildlife Manuals, No. 79)*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland.
<https://www.npws.ie/sites/default/files/publications/pdf/IWM79.pdf>
- Perrin, P. M., O’Hanrahan, B., Roche, J. R., & Barron, S. J. (2009). *Scoping study and pilot survey for a national survey and conservation assessment of upland habitats and vegetation in Ireland*. National Parks and Wildlife Service, Department of Environment, Heritage & Local Government, Ireland.
http://www.botanicalenvironmental.com/wp-content/uploads/2010/03/Perrin_et_al_2009_Upland_Pilot.pdf
- Qu, B. Y., Lang, B. F., Liang, J. J., Qin, A. K., & Crisalle, O. D. (2016). Two-hidden-layer extreme learning machine for regression and classification. *Neurocomputing*, 175, 826–834. <https://doi.org/10.1016/j.neucom.2015.11.009>
- Reis, B. P., Martins, S. V., Fernandes Filho, E. I., Sarcinelli, T. S., Gleriani, J. M., Leite, H. G., & Halassy, M. (2019). Forest restoration monitoring through digital processing of high resolution images. *Ecological Engineering*, 127, 178–186.
<https://doi.org/10.1016/j.ecoleng.2018.11.022>
- Rhodes, C. J., Henrys, P., Siriwardena, G. M., Whittingham, M. J., & Norton, L. R. (2015). The relative value of field survey and remote sensing for biodiversity assessment. *Methods in Ecology and Evolution*, 6(7), 772–781.
<https://doi.org/10.1111/2041-210X.12385>
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67,

93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>

Roelfsema, C. M., Lyons, M., Murray, N., Kovacs, E. M., Kennedy, E., Markey, K., Borrego-Acevedo, R., Ordoñez Alvarez, A., Say, C., Tudman, P., Roe, M., Wolff, J., Traganos, D., Asner, G. P., Bambic, B., Free, B., Fox, H. E., Lieb, Z., & Phinn, S. R. (2021). Workflow for the generation of expert-derived training and validation data: A view to global scale habitat mapping. *Frontiers in Marine Science*, 8, 643381. <https://doi.org/10.3389/fmars.2021.643381>

Rowat, D., Gore, M., Meekan, M. G., Lawler, I. R., & Bradshaw, C. J. A. (2009). Aerial survey as a tool to estimate whale shark abundance trends. *Journal of Experimental Marine Biology and Ecology*, 368(1), 1–8. <https://doi.org/10.1016/j.jembe.2008.09.001>

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., Leemans, R., Lodge, D. M., Mooney, H. A., Oesterheld, M., Poff, N. L. R., Sykes, M. T., Walker, B. H., Walker, M., & Wall, D. H. (2000). Global biodiversity scenarios for the year 2100. *Science*, 287(5459), 1770–1774. <https://doi.org/10.1126/science.287.5459.1770>

Scanes, C. G. (2018). Human activity and habitat loss: destruction, fragmentation, and degradation. In C. G. Scanes & S. R. Toukhsati (Eds.), *Animals and Human Society* (pp. 451–482). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-805247-1.00026-5>

Schmidt, A. M., & van Swaay, C. A. M. (2021). Monitoring of species and habitats. In A. M. Schmidt & T. Van der Sluis (Eds.), *E-BIND Handbook (Part A): Improving the availability of data and information on species, habitats and sites* (pp. 4–39). https://www.ecologic.eu/sites/default/files/publication/2021/A1_Monitoring-of-species-and-habitats.pdf

Schmidt, T., Schuster, C., Kleinschmit, B., & Förster, M. (2014). Evaluating an intra-annual time series for grassland classification - How many acquisitions and what seasonal origin are optimal? *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8), 3428–3439. <https://doi.org/10.1109/JSTARS.2014.2347203>

- Scholefield, P., Morton, D., McShane, G., Carrasco, L., Whitfield, M. G., Rowland, C., Rose, R., Wood, C., Tebbs, E., Dodd, B., & Monteith, D. (2019). Estimating habitat extent and carbon loss from an eroded northern blanket bog using UAV derived imagery and topography. *Progress in Physical Geography: Earth and Environment*, 43(2), 282–298. <https://doi.org/10.1177/0309133319841300>
- Schug, F., Okujeni, A., Hauer, J., Hostert, P., Nielsen, J., & van der Linden, S. (2018). Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series. *Remote Sensing of Environment*, 210, 217–228. <https://doi.org/10.1016/j.rse.2018.03.022>
- Scott, G. J., England, M. R., Starns, W. A., Marcum, R. A., & Davis, C. H. (2017). Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(4), 549–553. <https://doi.org/10.1109/LGRS.2017.2657778>
- Siddique, F., Sakib, S., & Siddique, M. A. B. (2019). Recognition of handwritten digit using convolutional neural network in python with tensorflow and comparison of performance for various hidden layers. *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, 541–546. <https://doi.org/10.1109/ICAEE48663.2019.8975496>
- Silver, M., Tiwari, A., & Karnieli, A. (2019). Identifying vegetation in arid regions using object-based image analysis with RGB-only aerial imagery. *Remote Sensing*, 11(19), 2308. <https://doi.org/10.3390/rs11192308>
- Tian, Y., Luo, L., Mao, D., Wang, Z., Li, L., & Liang, J. (2017). Using Landsat images to quantify different human threats to the Shuangtai Estuary Ramsar site, China. *Ocean and Coastal Management*, 135, 56–64. <https://doi.org/10.1016/j.ocecoaman.2016.11.011>
- Timm, B. C., & McGarigal, K. (2012). Fine-scale remotely-sensed cover mapping of coastal dune and salt marsh ecosystems at Cape Cod National Seashore using Random Forests. *Remote Sensing of Environment*, 127, 106–117. <https://doi.org/10.1016/j.rse.2012.08.033>

- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1), 792. <https://doi.org/10.1038/s41467-022-27980-y>
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, 18(6), 306–314. [https://doi.org/10.1016/S0169-5347\(03\)00070-3](https://doi.org/10.1016/S0169-5347(03)00070-3)
- Vanden Borre, J., Haest, B., Lang, S., Spanhove, T., Förster, M., & Sifakis, N. I. (2011). Towards a wider uptake of remote sensing in Natura 2000 monitoring: Streamlining remote sensing products with users' needs and expectations. *2nd International Conference on Space Technology, ICST 2011*. <https://doi.org/10.1109/ICSpT.2011.6064686>
- Vanden Borre, J., Paelinckx, D., Múcher, C., Kooistra, L., Haest, B., De Blust, G., & Schmidt, A. (2011). Integrating remote sensing in Natura 2000 habitat monitoring: Prospects on the way forward. *Journal for Nature Conservation*, 19(2), 116–125. <https://doi.org/10.1016/j.jnc.2010.07.003>
- Walter, V. (2004). Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3–4), 225–238. <https://doi.org/10.1016/j.isprsjprs.2003.09.007>
- Wang, X., Wang, Y., Zhou, C., Yin, L., & Feng, X. (2021). Urban forest monitoring based on multiple features at the single tree scale by UAV. *Urban Forestry and Urban Greening*, 58, 126958. <https://doi.org/10.1016/j.ufug.2020.126958>
- Wang, Y., & Moskovits, D. K. (2001). Tracking fragmentation of natural communities and changes in land cover: Applications of landsat data for conservation in an urban landscape (Chicago Wilderness). *Conservation Biology*, 15(4), 835–843. <https://doi.org/10.1046/j.1523-1739.2001.015004835.x>
- Waske, B., Fauvel, M., Benediktsson, J. A., & Chanussot, J. (2009). Machine learning techniques in remote sensing data analysis. In G. Camps-Valls & L. Bruzzone

- (Eds.), *Kernel methods for remote sensing data analysis* (pp. 1–24). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470748992.ch1>
- Wood, D. J. A., Preston, T. M., Powell, S., & Stoy, P. C. (2022). Multiple UAV flights across the growing season can characterize fine scale phenological heterogeneity within and among vegetation functional groups. *Remote Sensing*, 14(5), 1290. <https://doi.org/10.3390/rs14051290>
- Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., ... Cook, B. D. (2022). Fifty years of Landsat science and impacts. *Remote Sensing of Environment*, 280, 113195. <https://doi.org/10.1016/j.rse.2022.113195>
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems 27 (NIPS '14)*, NIPS Foundation, 3320–3328.
- Yu, X., Wu, X., Luo, C., & Ren, P. (2017). Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience and Remote Sensing*, 54(5), 741–758. <https://doi.org/10.1080/15481603.2017.1323377>
- Zerrouki, N., & Bouchaffra, D. (2014). Pixel-based or object-based: Which approach is more appropriate for remote sensing image classification? *Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 864–869. <https://doi.org/10.1109/SMC.2014.6974020>
- Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*. CRC Press, Taylor & Francis. <https://doi.org/10.1201/b12207>
- Zhu, B., Gao, H., Wang, X., Xu, M., & Zhu, X. (2018). Change detection based on the combination of improved SegNet neural network and morphology. *2018 3rd IEEE International Conference on Image, Vision and Computing, ICIVC 2018*, 55–59. <https://doi.org/10.1109/ICIVC.2018.8492747>

2. Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping



Fixed dunes in the Magharees Peninsula, County Kerry. (Photo: Charmaine Cruz)

This chapter is based on the peer-reviewed article:

Cruz, C., O'Connell, J., McGuinness, K., Martin, J. R., Perrin, P. M. & Connolly, J. (2023). Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping, *European Journal of Remote Sensing*, 56(1), 2191870. <https://doi.org/10.1080/22797254.2023.2191870>

Description of authors' contributions:

C.C. designed the research and performed the analysis. C.C., J.O.C., J.M, P.M.P. and J.C. performed the data acquisition. C.C. prepared the original draft manuscript. J.C., K.M., P.M.P., J.M. and J.O.C. provided a critical review of the manuscript. C.C. led the submission of the manuscript to the journal. C.C. revised the manuscript based on four anonymous reviewers' comments and suggestions.

2.1. Abstract

Coastal dunes are considered some of the most threatened and vulnerable habitats in the European Union. Mapping the spatial distribution of these habitats is an essential task for their conservation. Advances in Unoccupied Aerial Vehicles (UAVs) facilitate the flexible acquisition of high-resolution imagery for identifying detailed spatial distributions of habitats within dune systems. This study aimed to assess the effectiveness of UAV remote sensing for mapping these habitat types. Specifically, we determined the impact of temporally acquired UAV-derived spectral and topographic information on classification accuracy. The work combined the multi-temporal UAV imagery with field observation data and used the Random Forest machine learning algorithm to classify dune habitats. Results showed that using multi-temporal UAV imagery increased classification accuracy compared to using uni-temporal UAV imagery (92.37% vs. 84.09%, respectively). Also, including topographic information consistently improved accuracy, regardless of the number of image sets used (the highest accuracy increased from 84.81% to 92.57% for a uni-temporal model). Temporal analyses showed that the data acquired in the middle period of the growing season were better than those acquired in the early or late periods. The methodology presented here demonstrates the potential of using UAV data for detailed mapping and monitoring of habitat types.

2.2. Introduction

The European coastline is frequently bordered by coastal dunes. They form a spatially narrow yet complex mosaic of dynamic habitats and are characterised by high floral diversity (Doody, 2013). They also provide important ecosystem services including coastal defence, maintenance of wildlife, water catchment, carbon sequestration and recreation (Barbier et al., 2011).

Coastal dunes are considered a highly dynamic environment. Formed by sand-sized sediment that is transported inland by onshore winds and trapped by vegetation, they exhibit a zonation pattern extending inland parallel to the beach (Bird, 2008, 2010; Hughes et al., 2014). Along beaches, the high tide line is marked by the deposition of a line of detritus, including seaweeds, and in this area vascular plants can establish. Around these drift lines, wind-blown sand can accumulate further, initiating the formation of the embryonic dunes which are unstable low mounds of sand, generally <1 m in height. The exposed sand of these mounds becomes vegetated by salt-tolerant species, such as sand couch (*Elytrigia juncea*) and lyme-grass (*Leymus arenarius*). As these dunes accumulate sand, their surface becomes less salty. This environment is where marram grass (*Ammophila arenaria*) begins to colonise. Marram grass has an extensive root system that stabilises the sand and allows the building of larger dunes. These marram dunes, or white dunes, are partially stabilised hills that are located farther inland than embryonic dunes and can reach 15-20 m in height (Fossitt, 2000). The next phase in succession is that of fixed dunes, otherwise known as grey dunes, which, compared with marram dunes, provide a more stabilised and sheltered environment. In this zone, the vegetation cover is usually characterised by herb-rich grassland and heath communities. Within low-lying, seasonally flooded areas in a dune system, dune slacks can also develop. These nutrient-enriched areas can support diverse floristic and plant communities. Dune scrub/woodland may be found in more generally stable parts of a system (Delaney et al., 2013).

In the European Union (EU), dune habitats are included in Annex I of the ‘Council Directive 92/43/EEC on the conservation of natural habitats and wild fauna and flora’, also known as the Habitats Directive (HD). The HD is an EU legal framework that aims to protect a wide range of rare, endangered, and vulnerable habitats (those listed in

Annex I) by requiring each Member State to monitor these habitats within their territories and report this to the European Commission on a six-yearly basis (Commission of the European Communities, 1992).

Within Ireland, eight Annex I coastal dune habitats are identified, and two other habitats are also often found in association with dune systems (Table 2.1). These dynamic habitats are vulnerable to natural hazards and human pressures, which can lead to habitat degradation and biodiversity loss (Delaney et al., 2013). In 2019, these ten Annex I coastal habitats were each assessed as having an Unfavourable-Inadequate or Unfavourable-Bad status (National Parks and Wildlife Service, 2019); these assessments were largely unchanged from the previous two HD reporting cycles in 2008 and 2013 (National Parks and Wildlife Service, 2008, 2013b). Two negative pressures associated with human activities in Irish dune systems are the use of vehicles and trampling, both of which can create networks of paths and tracks, resulting in damage to vegetation and increased dune instability (Delaney et al., 2013). Assessing and monitoring these coastal dunes is critical to identifying key conservation issues, measuring threat status, and developing and evaluating management and restoration programs (European Commission, 2022).

Habitat mapping is a fundamental task for generating the data required for monitoring habitats under the HD. Most habitat mapping is conducted by ecologists in the field. While these surveys can provide detailed information, manual methods alone can limit the size and temporal scope of the survey, primarily due to cost (Mack et al., 1997). Remote sensing offers an advantage over traditional field-based mapping techniques as it may capture data more frequently and across larger areas (Jensen, 2014; Vanden Borre et al., 2011). However, monitoring coastal dune systems with freely available medium-resolution (>10 m) satellite imagery can be difficult, for example when mapping very narrow areas of patchy mobile dunes or identifying small-sized disturbed areas within the habitats caused by human tracks and coastal erosion (Ettritch et al., 2018; Marzialesi et al., 2019). For this reason, the classification output from this imagery can result in inaccurate spatial and botanical information about dune status; higher resolution data is needed to capture this detailed information.

Recent advancements in remote sensing platforms and sensors, such as the development of Unoccupied Aerial Vehicles (UAVs), have enabled high-resolution mapping of habitats and vegetation (Gonçalves et al., 2016; Kattenborn, Eichel, et al., 2019; Michez et al., 2016; Müllerová et al., 2021; van Iersel et al., 2018). UAVs can acquire centimetre-level resolution imagery. They are also characterised by high efficiency and flexibility in image acquisition of small (< 20 ha) areas because they generally have low operational costs; this enables a high frequency of image acquisition that is suitable for environmental monitoring (Manfreda et al., 2018; Pajares, 2015; Yan et al., 2009).

In recent years, several UAV-based studies have implemented machine learning algorithms for image classification and habitat mapping due to the excellent classification results and fast processing time (Han et al., 2019; Lu & He, 2017; van Iersel et al., 2018). Random Forest (RF) is one of the most widely used algorithms (Belgiu & Drăgut, 2016; Pal, 2005). RF (Breiman, 2001) is a non-parametric, supervised machine learning algorithm that uses a collection of decision trees to solve a classification task. Each tree in an RF model generates a prediction and the final prediction is determined by majority voting (Fawagreh et al., 2014). It allows large volumes of highly correlated variables to be ingested and processed without data redundancy affecting the model performance. The efficiency of RF in processing large datasets makes it appropriate for multi-temporal analysis of UAV images. The RF algorithm was selected as the preferred classification method for the present study due to the reported high accuracy in vegetation mapping using various remote sensing data (Amani et al., 2017; Barrett et al., 2016; Raab et al., 2018; Villoslada et al., 2020).

In habitat mapping, image acquisition timing is an important factor since plant communities within a habitat may be more distinct from the surrounding vegetation at a particular time of the growing season (Marcinkowska-Ochtyra et al., 2019). The use of multi-temporal images in habitat mapping can increase classification accuracy compared to maps produced from an image collected at a single time (Marcinkowska-Ochtyra et al., 2019; Schmidt et al., 2014; van Iersel et al., 2018). Several studies have used multi-temporal UAV images in mapping habitats and vegetation (Belcore et al., 2021; Michez et al., 2016; van Iersel et al., 2018); however, very few studies have been carried out on coastal dune systems (Taddia et al., 2019).

This study aimed to assess the effectiveness of UAV remote sensing for mapping complex coastal dune habitat types. More specifically, we aimed to (1) determine how classification accuracy varies when using multi-temporal images acquired throughout the growing season, (2) identify if and how the inclusion of UAV-derived topography improves the classification accuracy, and (3) identify the optimum period of the growing season for UAV data acquisition for the purposes of differentiating coastal dune habitats. These objectives were achieved by integrating high-resolution UAV imagery with data from field surveys. The different habitat types were classified using the RF algorithm.

2.3. Materials and methods

2.3.1. Study site description

The study area is a 37-ha coastal dune site located on the Magharees Peninsula (52°16'21"N, 10°2'15"W) in County Kerry, Ireland (Figure 2.1). The Magharees Peninsula is a sand spit, about 5 km long, with an extensive area of fixed grey dunes containing several dune slacks (Delaney et al., 2013; National Parks and Wildlife Service, 2013a). This dune site is designated as part of the large Tralee Bay and Magharees Peninsula, West to Cloghane Special Area of Conservation (SAC code 002070, with an area of approximately 11,627 ha) (Figure 2.1), part of the Natura 2000 network (National Parks and Wildlife Service, 2018).

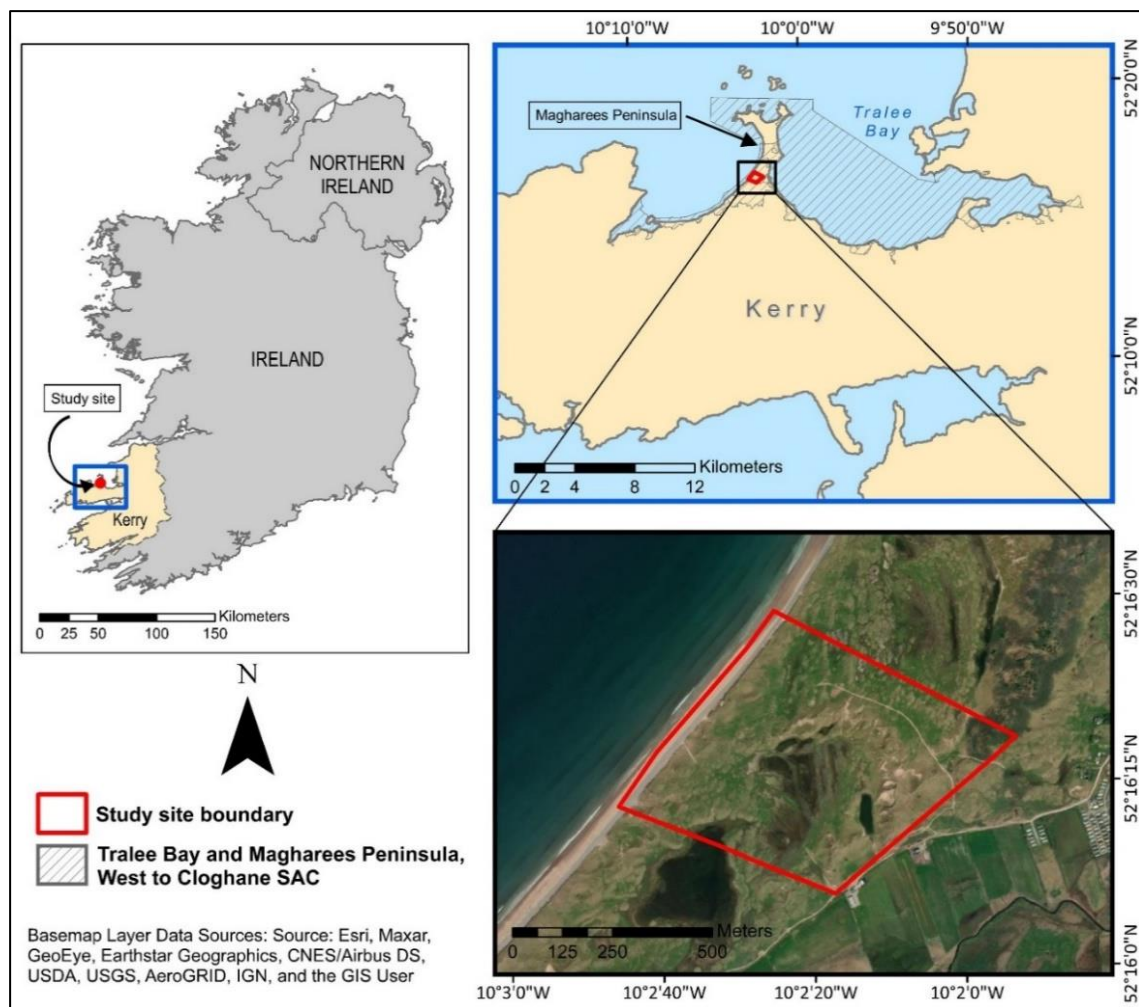


Figure 2.1. Inset map showing the location of the study site in County Kerry, Ireland (left); the boundary of the Tralee Bay and Magharees Peninsula, West to Cloghane SAC (upper right); and the boundary of the study site (lower right).

Six HD Annex I habitats are present within the study area (Table 2.1 and Figure 2.2) – annual vegetation of drift lines (1210), embryonic shifting dunes (2110), white dunes (2120), grey dunes (2130), dunes with *Salix repens* (2170) and dune slacks (2190). One of these, grey dunes, is a priority habitat that requires particular protection as its distribution is in danger of disappearance within the EU territory (National Parks and Wildlife Service, 2019). For botanical information on these six habitats at the study site, see Appendix B1.

Other common habitats present within the study area, which are not listed in Annex I of the HD, were categorised based on the Fossitt classification scheme (Table 2.1 and Figure

2.2). The Fossitt scheme is a standard guide for identifying and describing habitats in Ireland (Fossitt, 2000). It is important to determine the locations and sizes of these habitats as some represent disturbance to the dune system (e.g., exposed sand, recolonising bare ground and the presence of non-native species).

Table 2.1. Habitat types found within Irish dune systems and corresponding habitat codes (following European Commission, 2013; Fossitt, 2000).

Habitat type and code	Present within the study area
Annex I	
1210 - Annual vegetation of drift lines [†]	✓
1220 - Perennial vegetation of stony banks [†]	✗
2110 - Embryonic shifting dunes	✓
2120 - Shifting dunes along the shoreline with <i>Ammophila arenaria</i> (white dunes)	✓
2130 - Fixed coastal dunes with herbaceous vegetation (grey dunes)*	✓
2140 - Decalcified fixed dunes with <i>Empetrum nigrum</i> *	✗
2150 - Atlantic decalcified fixed dunes (<i>Calluno-Ulicetea</i>)*	✗
2170 - Dunes with <i>Salix repens</i> ssp. <i>argentea</i> (<i>Salicion arenariae</i>)	✓
21A0 - Machair*	✗
2190 - Humid dune slacks	✓
Fossitt	
CD4 - Dune scrub and woodland	✓
ED1 - Exposed sand	✓
ED3 - Recolonising bare ground	✓
LS1 - Shingle and gravel shores	✓
LS2 - Sandy shores	✓

* Priority habitats

† Habitats associated with dune system

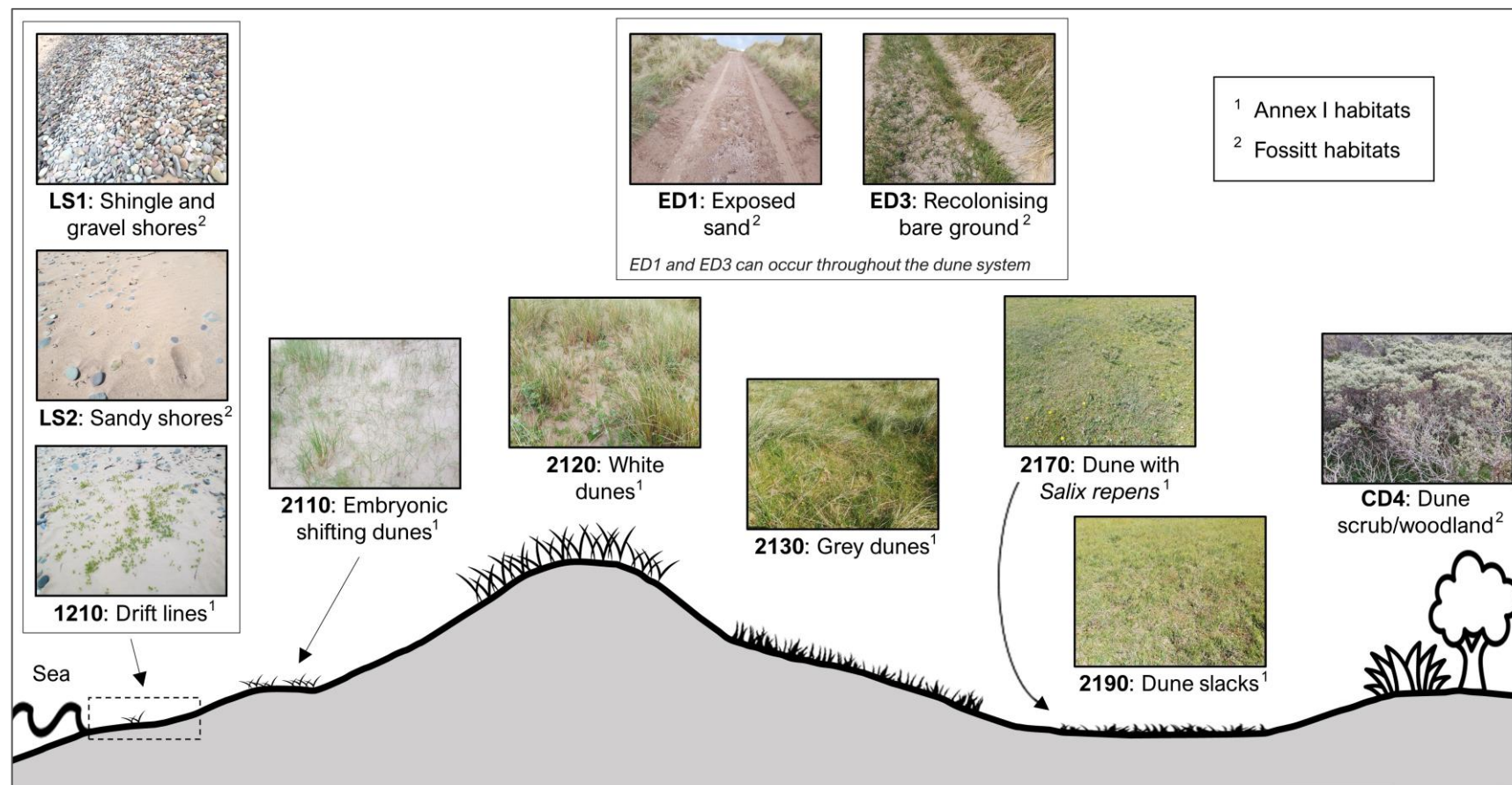


Figure 2.2. A typical dune profile (modified from Delaney et al., 2013) and sample field photos of Annex I and Fossitt habitats present within the study site.

2.3.2. General workflow

A replicable workflow using multi-temporal high-resolution UAV imagery and field data acquired during the vegetation growing season was used to characterise different habitat types. Several raster layers were generated from the UAV-derived photogrammetric products, including a raster orthomosaic and a Digital Surface Model (DSM). These layers, together with the field data as reference samples, were used to train the RF-based classification of each target habitat type. The trained model was then applied to the image mosaic to produce a spatially continuous map of habitats within the study area. These processes are illustrated in a flow diagram (Figure 2.3). Each step in the diagram is discussed in detail in the following subsections.

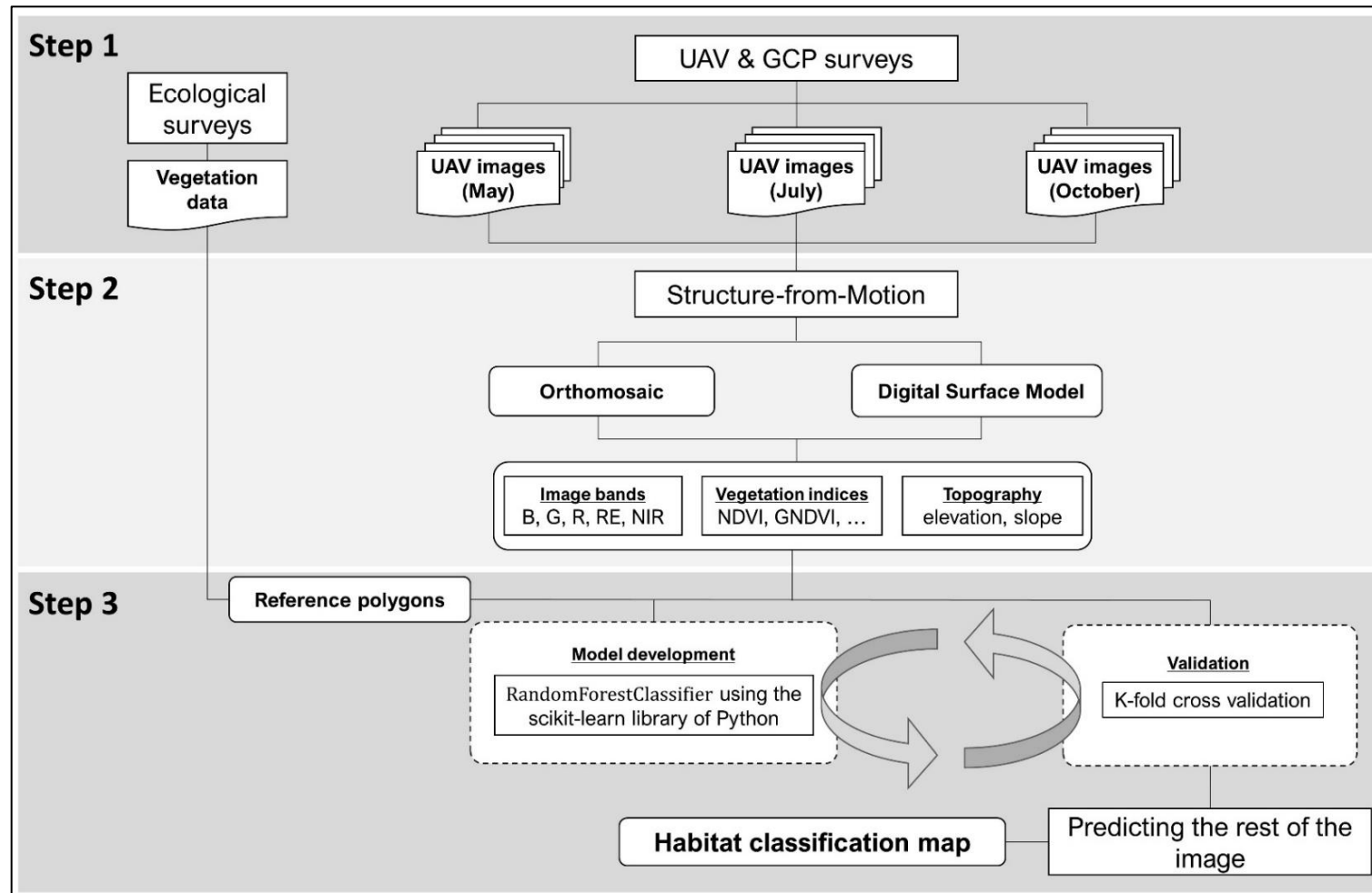


Figure 2.3. Methodology workflow in using multi-temporal high-resolution UAV imagery to characterise different habitat types.

Step 1. Field surveys

Image acquisition. Field campaigns were conducted during the growing season in 2020: 26 May (M, early), 28 July (J, mid), and 15 October (O, late), under suitable flying conditions (i.e., clear sky with low wind speed). High-resolution imagery (6 cm) was acquired using a DJI Phantom 4 Multispectral UAV sensor (<https://www.dji.com/p4-multispectral/specs>) with five spectral bands – blue (450 nm), green (560 nm), red (650 nm), red edge (RE) (730 nm) and near-infrared (NIR) (840 nm). Five Ground Control Points (GCPs) were evenly distributed throughout the study area to spatially reference the UAV-acquired images, following recommendations by Assmann et al. (2019) and Mestas-Carrascosa et al. (2015). The coordinates at the centre of each GCP were measured and recorded using an Emlid Reach RS+ Global Navigation Satellite System (GNSS) device (<https://emlid.com/reachrs/>) with the observation time set to two minutes. All three flights were operated with a side and forward overlap of 75% and 80% respectively, and a flight altitude of 60 m. Invariant panels were also placed within the study area and imaged during the flights. These were subsequently used for radiometric calibration. These panels were used to normalise for variations in atmosphere and solar illumination across the three image acquisition dates.

Reference data collection. Extensive ecological field surveys were carried out by experienced ecologists to collect ground truth data (i.e., reference points representing the actual location of a habitat type) to train the model and validate the results. The ecologists used the GNSS device to record the coordinates of the data points. Each point was located at the centre of an area of relatively homogeneous habitat cover. The ecologists recorded the radius of this area around each point, typically 1-10 m, based on their expert opinion of its spatial extent (Figure 2.4a). Each point also included a record of habitat codes based on the Annex I and Fossitt classification systems and was photographed in the north-view direction.

First, the team collected reference points in the week following the first UAV survey to prevent disturbance to the site. Further site visits that were coordinated with the second and third UAV surveys facilitated the addition of more points and assessment of any significant vegetation changes from the previous survey. After verifying that there was

no change in the habitat types that were recorded within the growing season (i.e., the habitat identification on the ground in the first visit remained correct in subsequent visits), the point data collected from all field campaigns were combined into a single set ($n = 418$) and used to classify all acquired UAV images. The point data were converted into polygons using the radii recorded in the field (Figure 2.4b). Both the remote sensing analyst and the ecologists visually inspected the collected field data (i.e., polygons) on the screen to ensure that there was no overlap and to verify that they were aligned with the correct habitats as seen on the UAV image (Figure 2.4c). Due to time constraints related to field-based data collection, 198 additional reference polygons were created in a GIS environment to supplement those classes with insufficient field-collected samples. This process comprised manual and subjective analysis of the UAV-acquired images by the ecologists who had conducted the field survey and were thus familiar with the site and habitats.

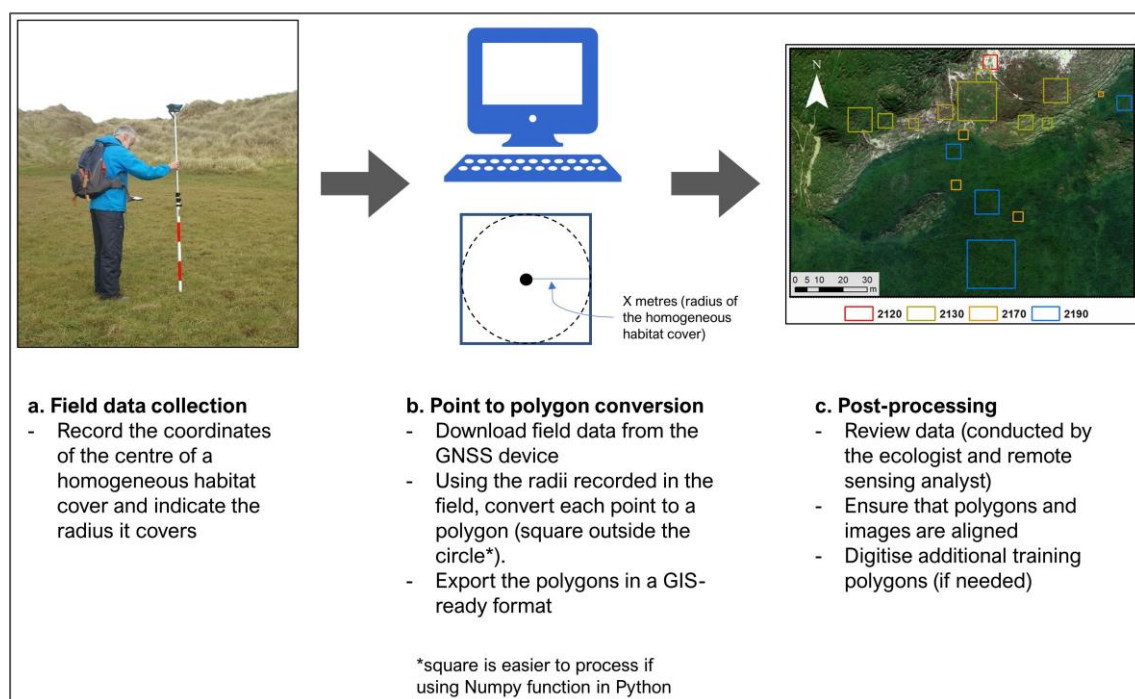


Figure 2.4. Diagram of the reference data collection and processing steps.

Step 2. Image Pre-processing

Structure-from-Motion (SfM). A DSM and an orthomosaic were created for each set of acquired images using the Structure-from-Motion (SfM)-based processing chain in Pix4DMapper (ver 4.3.33). The processing involves keypoint (i.e., distinct features in an image, usually of high contrast) extraction, keypoint matching between images, camera model correction, geolocation based on the GPS flight trajectories and the GCPs deployed onsite, and dense point cloud generation from highly overlapping (>70%) RGB image bands. The point cloud was then used to derive a DSM using Inverse Distance Weighting interpolation as it can provide good interpolation from a dense point cloud (Agüera-Vega et al., 2020). The parameter settings used in producing the DSM and orthomosaic are given in Table B.3 in Appendix B2.

Feature extraction. Feature extraction is a technique whereby raw raster images are used to generate new features or variables that provide more information to help discriminate between classes. Sixteen raster layers were generated from the photogrammetric products for each field campaign. These layers were grouped into two datasets comprising spectral and topographic variables (Table 2.2). The spectral variables were derived from the orthomosaic and consist of five radiometrically calibrated surface reflectance bands, the first Principal Component Analysis (PCA) band and eight vegetation indices. In contrast, topographic variables, such as elevation and slope, were derived through morphometric analysis of the DSM.

Reflectance bands were used as they can provide spectral information in the visible, RE and NIR regions of the image. The eight vegetation indices derived for this study have been found to have good performance in assessing vegetation cover, estimating chlorophyll content in densely and sparsely vegetated areas, estimating vegetation water content and providing evidence of correlation with biomass (Gitelson et al., 1996; Huete et al., 2002; Rouse et al., 1973; Xue et al., 2017). The PCA technique was utilised as it reduces data dimensionality by creating new uncorrelated variables (i.e., Principal Components) that contain the most important information from the data as represented by eigenvalues (Jolliffe, 2002). Here, the first PC band, having an eigenvalue of more than 75%, was extracted from the reflectance bands (see Table B.4, Appendix B2).

Finally, information on elevation and slope generally supports the dune habitat classification as they can be determinants for vegetation composition and distribution.

Table 2.2. List of variables used in the study and their descriptions.

Category	Variables	Descriptions	Reference
Colour image bands	Blue band (B)	Spectral reflectance value of the blue band	
	Green band (G)	Spectral reflectance value of the green band	
	Red band (R)	Spectral reflectance value of the red band	
	Red edge band (RE)	Spectral reflectance value of the red edge band	
	Near-infrared band (NIR)	Spectral reflectance value of the NIR band	
Spectral variables	Enhanced Vegetation Index (EVI)	$EVI = 2.5 * \frac{(NIR - R)}{[NIR + (6 * R) - (7.5 * B) + 1]}$	(Huete et al., 2002)
	Enhanced Vegetation Index 2 (EVI2)	$EVI2 = 2.5 * \frac{(NIR - R)}{[NIR + (2.4 * R) + 1]}$	(Jiang et al., 2007)
	Green Normalised Difference Vegetation Index (GNDVI)	$GNDVI = \frac{(NIR - G)}{(NIR + G)}$	(Gitelson et al., 1996)
	Modified Soil-Adjusted Vegetation Index 2 (MSAVI2)	$MSAVI2 = \frac{(2 * NIR) + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - R)}}{2}$	(Qi et al., 1994)
	Normalised Difference Red Edge (NDRE)	$NDRE = \frac{(NIR - RE)}{(NIR + RE)}$	(Barnes et al., 2000)

		Normalised Difference Vegetation Index (NDVI)	$NDVI = \frac{(NIR - R)}{(NIR + R)}$	(Rouse et al., 1973)
		Normalised Difference Water Index (NDWI)	$NDWI = \frac{(G - NIR)}{(G + NIR)}$	(McFeeters, 1996)
		Simple Ratio Index (SR)	$SR = \frac{NIR}{R}$	(Jordan, 1969)
Image bands transformation	Principal Component 1 (PC1)	First principal component that accounts for the greatest variance in the dataset		(Jolliffe, 2002)
Topographic variables	Elevation	Represents the height of the terrain.		
	Slope	Represents the rate of change of elevation.		

The reference data polygons (various sizes) were used to extract pixels from each variable in Table 2.2. This study used a 17-pixel non-overlapping sliding kernel (equivalent to 1×1 m on the ground) to divide each reference polygon into several equal-sized sub-polygons. The mean of the pixel values within each sub-polygon was calculated for every raster layer. The resulting mean pixel values and the corresponding polygon labels constituted the final reference dataset. This approach of a sliding averaging kernel reduces issues around reflectance-based adjacency effects and erroneous values due to geometric misalignment when using individual pixel values.

Step 3. Modelling and Validation

Several classification models were generated using the supervised RF algorithm to investigate the effect of vegetation phenology and topographic information on classification accuracy. The input for each model varied depending on combinations of variables (spectral variables only - RFs, spectral and topographic variables - RF_{ST}) and the number of field campaigns. In total, there were fourteen data combinations subjected to the RF algorithm (Table 2.3). In this study, the RF was implemented in Python 3.7 using the scikit-learn library (Pedregosa et al., 2011). The parameters used for each model were $M_{var} = \sqrt{M}$ for the number of variables at each split, where M is the total number of variables, and $N_{tree} = 500$ trees (Gislason et al., 2006). The relative importance of variables was analysed based on Mean Decrease in Impurity (Breiman et al., 1984).

Table 2.3. Datasets used in training models.

Dataset	Field campaign			Variables	
	May (M)	July (J)	October (O)	Spectral only (RF _s)	Spectral and topographic (RF _{ST})
M_RF _s	✓			✓	
J_RF _s		✓		✓	
O_RF _s			✓	✓	
MJ_RF _s	✓	✓		✓	
JO_RF _s		✓	✓	✓	
MO_RF _s	✓		✓	✓	
MJO_RF _s	✓	✓	✓	✓	
M_RF _{ST}	✓				✓
J_RF _{ST}		✓			✓
O_RF _{ST}			✓		✓
MJ_RF _{ST}	✓	✓			✓
JO_RF _{ST}		✓	✓		✓
MO_RF _{ST}	✓		✓		✓
MJO_RF _{ST}	✓	✓	✓		✓

M, J, O – field campaign month

s – spectral variables

ST – spectral and topographic variables

The RF technique was applied to each dataset to classify four of the coastal dune Annex I habitats—white dunes (2120), grey dunes (2130), dunes with *Salix repens* (2170) and dune slacks (2190)—and five Fossitt habitats. The other two Annex I habitats recorded—annual vegetation of drift lines (1210) and embryonic shifting dunes (2110)—were not included in the classification as they occurred in small, isolated patches and only 13 reference points were collected. Including them would have resulted in a severely imbalanced training set and, therefore, a biased model (Khoshgoftaar et al., 2007).

To evaluate the model performance, the accuracy was estimated using *k*-fold cross-validation. This method estimates model performance by randomly splitting up the data into *k* number of divisions or folds, wherein each fold is held out as the test set, and the other *k* – 1 folds are used as the training sets (Berrar, 2019). Five folds were used here,

i.e., the model training and validation process was repeated five times. The average and standard error values were calculated from each 5-fold model. The final model to predict the class labels of the entire study area used all the training data.

The model performance for each class was also evaluated using confusion matrices and estimating derived evaluation metrics (Equations 2.1–2.3) known as precision, recall, and F1-score (Muller & Guido, 2017). Precision is the fraction of correct predictions for each class and recall is the fraction of true labels that were correctly classified. F1-score is a single value that reports the overall accuracy, considering both recall and precision. It is the harmonic mean (weighted) of recall and precision values. These three metrics were calculated as follows:

$$Precision = TP_i / (TP_i + FP_i), \quad (\text{Equation 2.1})$$

$$Recall = TP_i / (TP_i + FN_i), \quad (\text{Equation 2.2})$$

$$F1\text{-score} = 2 \times (Recall \times Precision) / (Recall + Precision), \quad (\text{Equation 2.3})$$

where TP_i (True Positives) is the number of examples correctly labelled as class i , FP_i (False Positives) is the number of examples incorrectly labelled as class i and FN_i (False Negatives) is the number of examples incorrectly rejected as class i (Muller & Guido, 2017).

Temporal analysis of coastal dune habitats.

Two tests were prepared to determine which time of year is optimal, in terms of classification accuracy, for image acquisition for coastal dune habitat mapping (Figure 2.5). The first test (*Test 1*) was applied to the M_RF_{ST} , J_RF_{ST} and O_RF_{ST} training models produced above. For each of these models, data from the image dates that were not used in training were combined to generate the test dataset. So, the M_RF_{ST} model was evaluated with the combined July and October data. The generated test dataset was randomly resampled 1000 times to create 1000 different test datasets. In other words, the M_RF_{ST} model was evaluated using 1000 test datasets, which generated 1000

accuracy values. Next, the J_RF_{ST} model was evaluated using the combination of the May and October data. Finally, the O_RF_{ST} model was evaluated using the combined May and July data. Therefore, each of these models has 1000 accuracy values. For the second test (*Test 2*), each model was reproduced but this time with only 80% of the data (sub-polygons). The remaining 20% of the data from each of the three campaigns were pooled and used to generate the test dataset. In this way, each model was assessed using the same test dataset. Similar to *Test 1*, the test dataset was randomly resampled 1000 times to create 1000 different test datasets. These 1000 test datasets were used to evaluate each *Test 2* model. The mean and standard error were computed for each of the models from *Test 1* and *Test 2*.

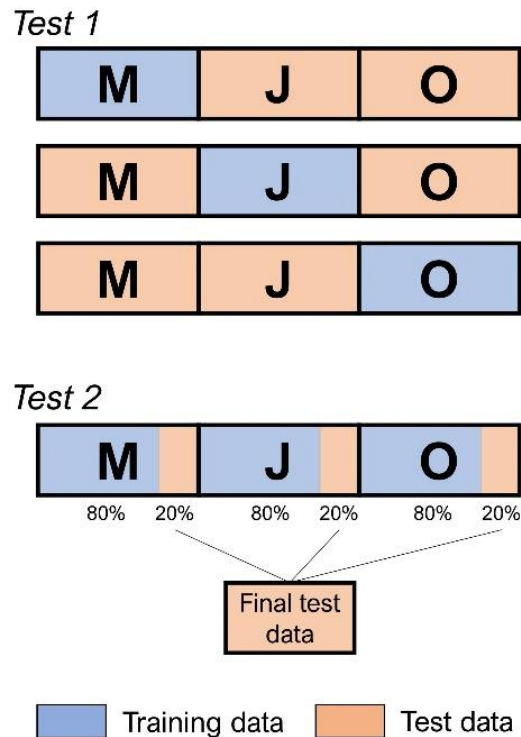


Figure 2.5. Preparation of *Test 1* and *Test 2* datasets to be used for modelling scenarios. May (M), July (J) and October (O) were the three field campaign months in this study.

2.4. Results

2.4.1. Acquisitions and variables used in the classification

The results obtained from 5-fold validation are presented in Table 2.4. The mean accuracy values of all models were above 80%, with MJO_RF_{ST} showing the best performance (94.26%) and O_RF_S the least good (84.09%). The computed standard error (SE) showed that mean accuracy based on $k = 5$ had low variability across the data.

Table 2.4. 5-fold cross-validation mean accuracy and standard error (SE) values for models obtained for each dataset. The datasets considered are described in Table 2.3. Values are in percentage (%).

Dataset	Accuracy based on 5-fold cross-validation method		Dataset	Accuracy based on 5-fold cross-validation method	
	Mean	SE		Mean	SE
M_RF _S	84.58	0.58	M_RF _{ST}	91.35	0.45
J_RF _S	84.81	0.42	J_RF _{ST}	92.57	0.52
O_RF _S	84.09	1.22	O_RF _{ST}	91.68	1.06
MJ_RF _S	88.48	1.17	MJ_RF _{ST}	93.49	0.52
MO_RF _S	90.85	0.48	MO_RF _{ST}	93.93	0.33
JO_RF _S	89.76	1.14	JO_RF _{ST}	92.54	1.13
MJO_RF _S	92.37	0.26	MJO_RF _{ST}	94.26	0.44

The models that used only spectral variables (RF_S) achieved higher accuracy when combining data from two or more field campaigns. This pattern was also seen in the results from the models using both spectral and topographic variables (RF_{ST}) with just one exception: JO_RF_{ST}, which had a slightly lower accuracy (92.54%) than J_RF_{ST} (92.57%).

All the models that used both the spectral and topographic variables (RF_{ST}) had a mean accuracy that exceeded 90%, with the highest accuracy acquired, which was for MJO_RF_{ST}, being 94.26% and the lowest, which was for M_RF_{ST}, being 91.35%. When comparing the RF_{ST} models with the corresponding RF_S models, the addition of topographic data always improved mean accuracy, which was also supported by the

variable importance ranking (see Table B.5, Appendix B2). The accuracy increase ranged from 2-8%, depending on the number of image sets considered in the model. The increase in accuracy for single image models was greater (M, +6.77%; J, +7.76%; O, +7.59%), than for the multiple image models (MJ, +5.01%; MO, +3.08%; JO, +2.78%; MJO, +1.89%). In other words, the more image sets considered in the model, the smaller the benefit of including the topographic variables. Moreover, analysis of the variable importance of uni-temporal RF_{ST} models indicates that, aside from the topographic variables, NDWI and GNDVI were the two most important variables common to these models. In contrast, NIR and RE bands were among the variables of low importance (see Table B.5, Appendix B2).

The model performance for each habitat differed between datasets, as presented by the confusion matrices, recall, precision and F1-score values (Figure 2.6 and Table 2.5). Overall, all models were able to predict most of the four Annex I dune habitats correctly (F1-score >0.75). Misclassifications among Annex I habitats primarily occurred between dunes with *Salix repens* (2170) and dune slacks (2190) (Figure 2.6). White dunes (2120) were also misclassified as ED1 and ED3, probably due to large areas of exposed sand in all three habitats. This result for white dunes (2120) was consistent with its slightly lower precision values (Table 2.5).

When topographic variables are included in the models, the F1-scores increase for most of the habitat types, with some classes reaching F1-scores >0.90 (Table 2.5). In general, both marine littoral sediments, LS1 and LS2, showed the highest F1-score values at 0.96 and above (Table 2.5). In contrast, degraded versions of Annex I habitats—bare sand (ED1) and recolonising bare ground (ED3)—had the lowest F1-scores for all the models, specifically due to low recall values (Table 2.5).

During the growing season, different months were identified as optimal for discriminating between the different habitats. For example, using the RF_{ST} models, both white dunes (2120) and dunes with *Salix repens* (2170) can be best differentiated in July as the F1-scores of 0.91 and 0.90, respectively, are slightly higher than the F1-scores for May and October (Table 2.5). In contrast, October was the best time to identify grey dunes (2130). Finally, dune slacks (2190) achieved the same F1-score (0.92) across all acquisition dates (Table 2.5).

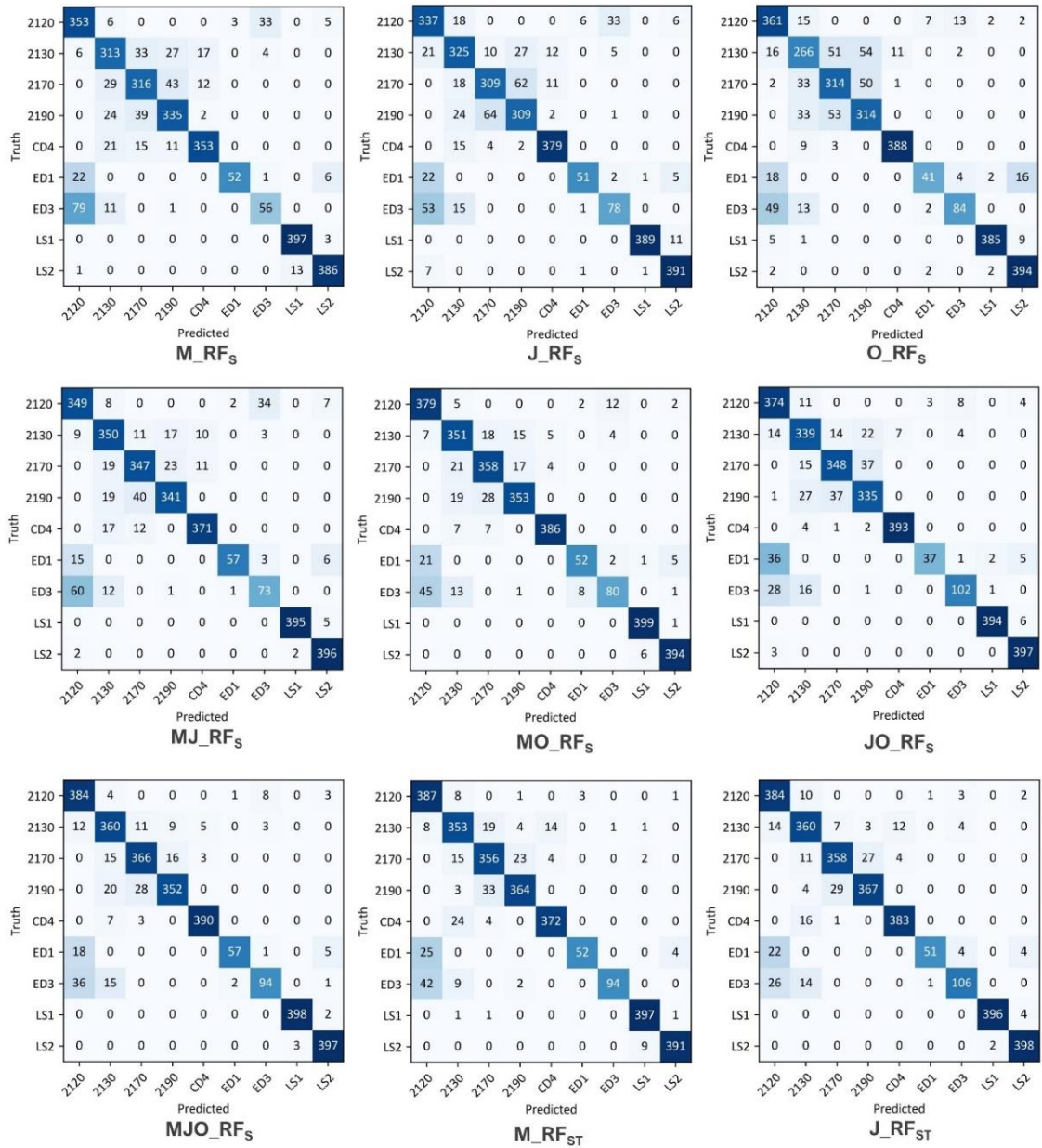


Figure 2.6. Confusion matrices obtained for each RF model. ‘Truth’ refers to the number of habitat samples from the reference data. ‘Predicted’ refers to the number of habitat samples identified from the classification. The colour scale along the major diagonal line indicates the proportion of reference samples classified correctly as a function of overall accuracy rounded between 0.0 and 1.0.

Figure continued on the next page.

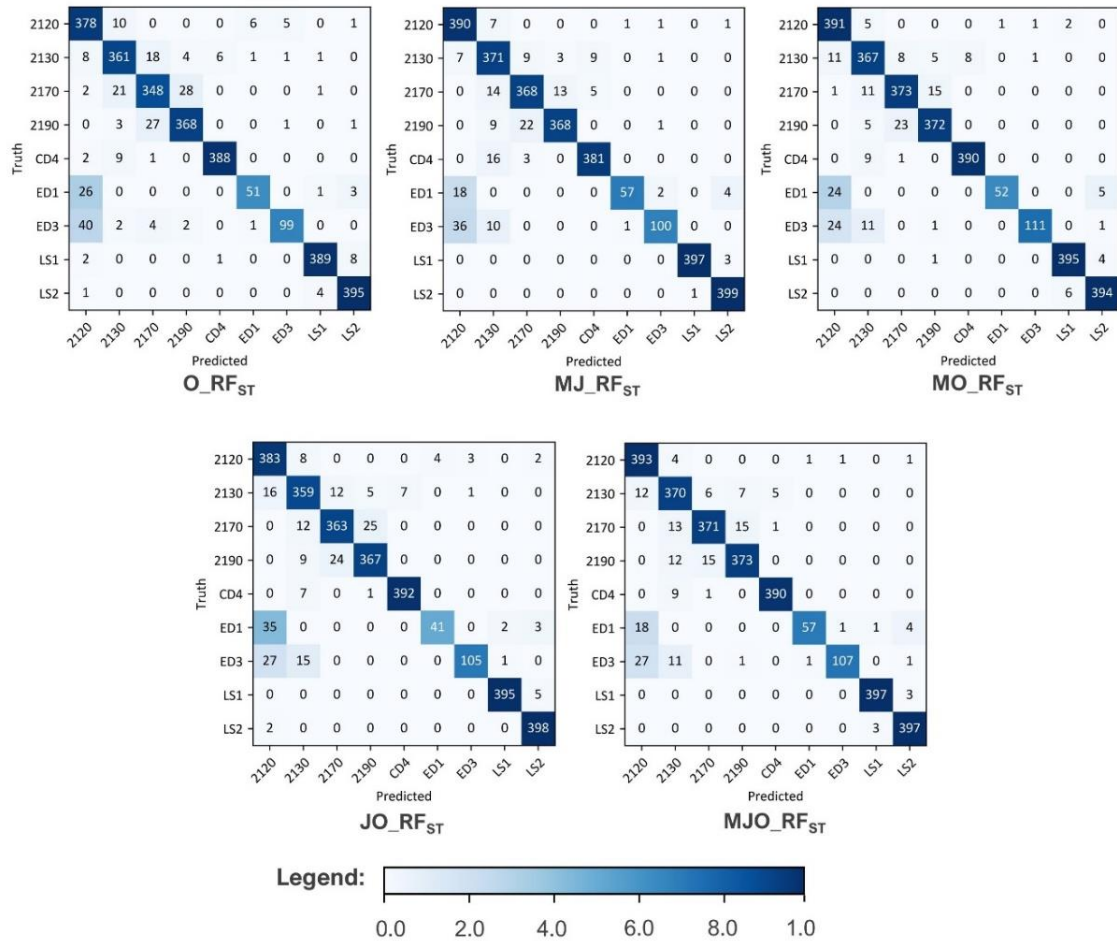


Figure 2.6. (continued)

Table 2.5. Per-class precision (P) and recall (R) values and F1-scores (F1) obtained for models from each dataset. The habitat codes are described in Table 2.1.

Habitat	May			July			October			May-July			May-Oct			July-Oct			May-July-Oct		
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1
Spectral variables only (RF _s)																					
2120	0.88	0.77	0.82	0.84	0.77	0.80	0.90	0.80	0.85	0.87	0.80	0.84	0.95	0.84	0.89	0.94	0.82	0.87	0.96	0.85	0.90
2130	0.78	0.77	0.78	0.81	0.78	0.80	0.67	0.72	0.69	0.88	0.82	0.85	0.88	0.84	0.86	0.85	0.82	0.83	0.90	0.86	0.88
2170	0.79	0.78	0.79	0.77	0.80	0.79	0.79	0.75	0.76	0.87	0.85	0.86	0.90	0.87	0.88	0.87	0.87	0.87	0.92	0.90	0.91
2190	0.84	0.80	0.82	0.77	0.77	0.77	0.79	0.75	0.77	0.85	0.89	0.87	0.88	0.91	0.90	0.84	0.84	0.84	0.88	0.93	0.91
CD4	0.88	0.92	0.90	0.95	0.94	0.94	0.97	0.97	0.97	0.93	0.95	0.94	0.97	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98
ED1	0.64	0.95	0.76	0.63	0.86	0.73	0.51	0.79	0.62	0.70	0.95	0.81	0.64	0.84	0.73	0.46	0.93	0.61	0.70	0.95	0.81
ED3	0.38	0.60	0.46	0.53	0.66	0.59	0.57	0.82	0.67	0.50	0.65	0.56	0.54	0.82	0.65	0.69	0.89	0.78	0.64	0.89	0.74
LS1	0.99	0.97	0.98	0.97	0.99	0.98	0.96	0.98	0.97	0.99	0.99	0.99	1.00	0.98	0.99	0.99	0.99	0.99	1.00	0.99	0.99
LS2	0.97	0.97	0.97	0.98	0.95	0.96	0.99	0.94	0.96	0.99	0.96	0.97	0.99	0.98	0.98	0.99	0.96	0.98	0.99	0.97	0.98
Spectral and topographic variables (RF _{ST})																					
2120	0.97	0.84	0.90	0.96	0.86	0.91	0.95	0.82	0.88	0.98	0.86	0.92	0.98	0.87	0.92	0.96	0.83	0.89	0.98	0.87	0.92
2130	0.88	0.85	0.87	0.90	0.87	0.88	0.90	0.89	0.90	0.93	0.87	0.90	0.92	0.90	0.91	0.90	0.88	0.89	0.93	0.88	0.90
2170	0.89	0.86	0.88	0.90	0.91	0.90	0.87	0.87	0.87	0.92	0.92	0.92	0.93	0.92	0.93	0.91	0.91	0.91	0.93	0.94	0.94
2190	0.91	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.96	0.94	0.93	0.94	0.94	0.92	0.92	0.92	0.93	0.94	0.94
CD4	0.93	0.95	0.94	0.96	0.96	0.96	0.97	0.98	0.98	0.95	0.96	0.96	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
ED1	0.64	0.95	0.76	0.63	0.96	0.76	0.63	0.86	0.73	0.70	0.97	0.81	0.64	0.98	0.78	0.51	0.91	0.65	0.70	0.97	0.81
ED3	0.64	0.99	0.78	0.72	0.91	0.80	0.67	0.93	0.78	0.68	0.95	0.79	0.75	0.98	0.85	0.71	0.96	0.82	0.72	0.98	0.83
LS1	0.99	0.97	0.98	0.99	0.99	0.99	0.97	0.98	0.98	0.99	1.00	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99
LS2	0.98	0.98	0.98	1.00	0.98	0.99	0.99	0.97	0.98	1.00	0.98	0.99	0.99	0.98	0.98	1.00	0.98	0.99	0.99	0.98	0.99

2.4.2. Temporal analysis of dune habitats identification

The RF-based models trained with data from different months to the test data (*Test 1*) each achieved mean accuracies of above 70% (Table 2.6). However, the accuracy of the model trained on the July data was higher than that of the models trained on the May or October data. The results of *Test 2* followed the same trend as those of *Test 1*. The J_RF_{ST} model had the highest performance, followed by O_RF_{ST} and M_RF_{ST} (Table 2.7). These results suggest that the best month for image acquisition was found to be the middle of the growing season for coastal dune vegetation. The mean accuracies obtained from all models evaluated using *Test 1* data ranged from 0.74 to 0.81, which was slightly lower than the accuracy range when *Test 2* was used (from 0.78 to 0.86) (Table 2.6 and Table 2.7).

Table 2.6. Summary table showing the mean and standard error (SE) values for each model evaluated on the *Test 1* dataset.

Model	Mean	SE
M_RF _{ST}	0.742	0.00026
J_RF _{ST}	0.809	0.00021
O_RF _{ST}	0.747	0.00024

Table 2.7. Summary table showing the mean and standard error (SE) values for each model evaluated on the *Test 2* dataset.

Model	Mean	SE
M_RF _{ST}	0.788	0.00026
J_RF _{ST}	0.860	0.00023
O_RF _{ST}	0.829	0.00025

2.4.3. Distribution and extent of dune habitat types

The computed area of the different habitat types, as classified by RF models based on three single image acquisition times, and the spatial classification output for July are shown in Figure 2.7. The study site was dominated by grey dunes (2130), which covered almost 60% of the total area. This was followed by dunes with *Salix repens* (2170) and

dune slacks (2190), which had about 12% - 15% each of the area. White dunes (2120) covered around 5% - 6%. All Fossitt habitats (i.e., dune scrub and woodland, exposed sand, recolonising bare ground, shingle and gravel shores, and sandy shores) corresponded to the smallest area proportions within the study area and collectively accounted for less than 10% of the total area. The proportions of habitats did not differ markedly between models.

Dune slacks (2190) were surrounded mainly by dunes with *Salix repens* (2170). White dunes (2120) were generally located parallel to the shore, alongside LS1 and LS2 habitats. Scrub composed of the invasive species sea buckthorn (*Hippophae rhamnoides*) was classified as habitat CD4 on the east side of the study site. The presence of human disturbance (beach access paths) can also be observed in the coastal dune environment.

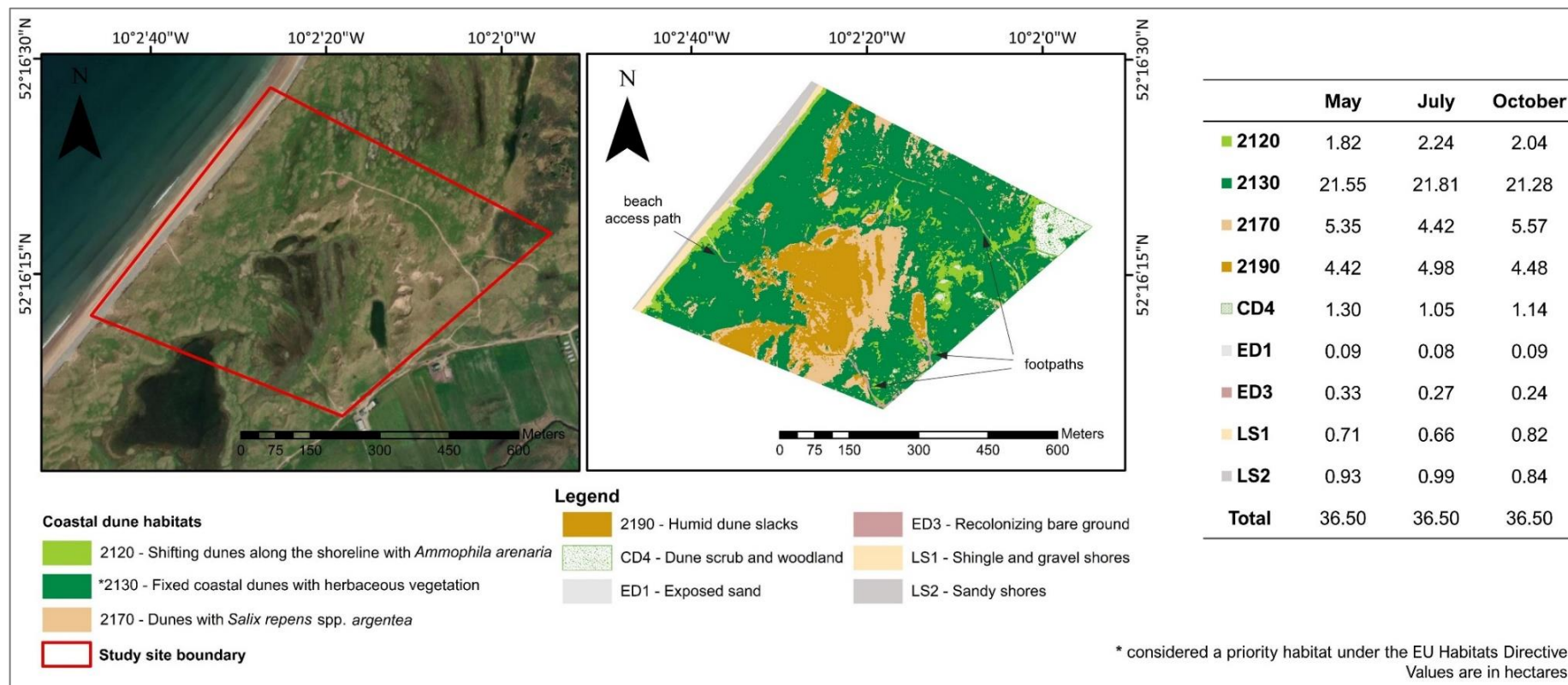


Figure 2.7. Supervised Random Forest classification map of coastal dune habitats in the Magharees, County Kerry and the computed area (in ha) for each habitat type classified using UAV images acquired at three different times within the growing season. The classified map is based on the July data.

2.5. Discussion

This study focused on the development of an automatic classification method using high-resolution UAV imagery and the Random Forest classifier. The result was a highly accurate (94.26%) map of coastal dune habitats. Laporte-Fauret et al. (2020) achieved a similar level of accuracy for the classification of dune vegetation using RF. Our study, using RF, yielded more accurate results than Suo et al. (2019) achieved when mapping another Irish dune system using supervised Maximum Likelihood. This result confirms the high potential of using RF to map complex habitats, such as coastal dunes, using UAV (Feng et al., 2015; Sotille et al., 2022; van Iersel et al., 2018).

A major strength of RF is that it is an ensemble classifier, i.e., it can efficiently handle a large number of variables with a lower risk of overfitting than models using single classifiers (Miao et al., 2012). This means that several variables can be incorporated into the model, providing more information about the target classes and improving the classification. In this study, elevation and two spectral variables, NDWI and GNDVI, appear to be the most discriminating variables for dune habitats (see Table B.5, Appendix B2). NDWI, an index related to plant water content, can distinguish dune slacks (2190), as they are generally characterised by saturated soil and moisture-loving plants (National Parks and Wildlife Service, 2019), whereas, GNDVI is highly sensitive to changes in chlorophyll content among different vegetation types (Gitelson et al., 1996).

The methodology was also designed to assess the impact of temporally acquired UAV-derived spectral and topographic data on classification accuracy. The inclusion of the UAV-derived topographic variables (elevation and slope) increased classification accuracy (Table 2.4), and this has been demonstrated in other vegetation and habitat-related studies (Fraser et al., 2016; Scholefield et al., 2019). In the present study, the inclusion of these topographic variables helped discriminate between dune habitats, especially dunes with *Salix repens* (2170) and dune slacks (2190), thus increasing the overall accuracy of the model. While habitats 2170 and 2190 contain similar plant communities and are even categorised as a single habitat type in the Irish Fossitt classification system (Fossitt, 2000), their distinct topographic characteristics enable their discrimination. Dune slacks (2190) are predominantly located at the lowest elevation in

the site and on flat terrain, whereas dunes with *Salix repens* (2170) can be found within or adjacent to dune slacks but at slightly higher elevations (National Parks and Wildlife Service, 2019). Moreover, the misclassifications were reduced between spectrally similar white dunes (2120) and those degraded areas in a dune system (ED1 and ED3) after adding topographic data. This was shown by the improved precision values in those three classes, indicating that more predictions in the RF_{ST} models were correct (compare RF_S and RF_{ST}, Table 2.5). White dunes (2120) are characterised by partially stabilised hills of sand with an incomplete vegetation cover. In contrast, ED1 and ED3 include areas on generally flat terrain that are regularly trampled or driven over. Hence, these areas have less vegetation cover (Fossitt, 2000). However, discriminating between white dunes (2120) and grey dunes (2130) remains a challenge even after the inclusion of topographic data (Figure 2.6). This is probably because, within the study site, both habitats are often characterised by hills of sand with the presence of *Ammophila arenaria* (see Appendix B1).

Other vegetation mapping studies have used separate sensors to acquire optical and topographic data; for example, some have used LiDAR (Bazzichetto et al., 2016; Marcinkowska-Ochtyra et al., 2019; Ward et al., 2013; Yousefi Lalimi et al., 2017). However, combining optical and topographic data acquired by different sensors can be costly and challenging as it requires geometric alignment between the datasets. In contrast, UAV-based topographic data can be obtained free of charge during the photogrammetric processing with a high level (± 3 cm) of vertical accuracy (Cohen & Héquette, 2020).

Previous studies have shown that using multi-temporal satellite or UAV images can, in many cases, improve classification accuracy compared with using uni-temporal images (Aghababaei et al., 2021; van Iersel et al., 2018; Vilar et al., 2020; Wood et al., 2022). This is confirmed by our results (Table 2.4). The improved overall accuracy was due to the increased correct classifications of the four Annex I habitats. As these habitats are characterised by different vegetation, their classification can benefit from the acquisition of multi-temporal images which provide spectral response variations throughout the growing season.

Interestingly, the models using both spectral and topographic variables (the RF_{ST} models) show only a slight increase of 3% in accuracy when more temporal data were included (91.35% vs 94.26%). In the models that only use spectral variables (RF_S models), the accuracy improved by up to 8% when data from two or more acquisition times were used (84.09% vs 92.37%). Therefore, from a practical point of view, using the RF_{ST} model, with UAV data acquired from a single time, is deemed sufficient for accurately classifying dune habitats as it can have a similar level of accuracy to data from multiple times.

The best acquisition time for each Annex I dune habitat varied as indicated by the F1-scores in Table 2.5. Hence, it can be assumed that there was no single optimum acquisition period for classifying these individual habitats. However, our analysis showed that the model trained on July data achieved the highest overall accuracy on two test datasets (*Test 1* = 0.81, *Test 2* = 0.86, see Tables 2.6 and 2.7). The difference between the two accuracy values could be due to *Test 2* models being tested with a dataset that contained samples from the same month as the data used to train the model.

This higher overall accuracy obtained in July may be explained by the fact that this period of the growing season is a time of maximum photosynthetic and floristic activity in the dune vegetation. Hence, spectral differences between the dune vegetation types are enhanced. For example, the dune-binding species *Ammophila arenaria* changes its colour in July and August as the golden flowers appear. This result is consistent with other studies that use spectral differences to discriminate between species. Michez et al. (2016) and Hill et al. (2010) found that the late season was a more appropriate period for image acquisition for forest areas because phenological differences between forest species are more enhanced during the senescence period. In other words, different habitats may exhibit greater spectral variations at different times of the year, and these can be used to devise an optimal period for monitoring. Despite this, misclassifications still occurred in the July model when evaluated using data from other times of the year. The water table in dune slacks (2190) remains at or close to the surface for most of the year (Fossitt, 2000). However, in 2020, they were flooded in May and dry in July and October. This water table fluctuation could account for the misclassifications in this area.

The use of UAV surveys to acquire centimetre-resolution images is currently limited to small areas—typically a few square kilometres—due to factors such as the flight speed, legally required proximity of operators to UAVs and battery life. This limitation might not be practical for operational use in regular habitat or vegetation mapping at regional or national scales. Our method is a practical approach for monitoring areas locally, and it could be extended to a larger scale when combined with freely available satellite imagery, such as that from the Sentinel-2 mission (e.g., Kattenborn et al., 2019, Riihimäki et al., 2019). While the UAV sensor used in this study operates in five spectral bands that are also present in Sentinel-2 sensors, understanding sensors' spectral characteristics (e.g., bandwidth, signal-to-noise ratio) is crucial for optimising multi-sensor synergies (Alvarez-Vanhard et al., 2020; Jiang et al., 2022). Moreover, aside from having lower spatial resolution compared to UAV imagery, the issue of cloud cover is another notable limitation when using satellite imagery. This means that countries with high levels of cloud cover, such as Ireland, can have less benefit from using optical satellite imagery for habitat monitoring (Connolly, 2018). Nevertheless, UAV technology is continuously developing and, in particular, the flight range of UAVs is increasing. This could address the current challenges in terms of spatial areal coverage without compromising the spatial image resolution.

With increased technological advances, in both computer processing power and remote sensing platforms, there are opportunities to explore the application of deep learning algorithms in habitat mapping. Moreover, the calculation of different texture features from the imagery could be used to represent vegetation structure within habitats. These data, combined with image bands, vegetation indices, and three-dimensional information, could further improve habitat classification accuracy. The addition of texture features in the classification was demonstrated by previous habitat and vegetation mapping studies (Gonçalves et al., 2016; Wood et al., 2012). It is also worth emphasising that future analysis in a coastal dune environment can take advantage of the UAV-derived data in other applications, such as in examining high-resolution morphodynamic processes (Laporte-Fauret et al., 2019).

The robust methodology presented here uses a low-cost multi-spectral UAV system to produce a highly accurate (>94%) high-resolution Annex I dune habitat map. It also highlights degradation caused by both natural and human disturbance (e.g., frequently

used tracks) and the presence of invasive species. This methodology is transferable and could be used for the accurate mapping of other habitats. Accurate maps are invaluable resources that provide information about habitat extent and condition. The maps produced from this methodology are useful for policymakers as well as site managers and can be used to monitor changes from baseline data as well as progress in restoration activities. Moreover, ecologists and other monitoring experts can integrate UAV remote sensing with the traditional field methods for the targeted surveillance and monitoring of habitats (e.g., detection of the extent and direction of change of the habitat location and status) (Anderson & Gaston, 2013; Cruzan et al., 2016).

2.6. Conclusion

Under the Habitats Directive, one key requirement for each EU Member State is the spatiotemporal monitoring of its natural habitats, including coastal dunes. The emergence of mapping technologies and UAV platforms providing high-resolution spatial data is a tool that can support habitat monitoring.

This study aimed to improve the method for monitoring complex dune habitats by integrating UAV data and using the Random Forest classifier to automatically map these habitats with a high level of accuracy. The result was a series of models and maps with accuracy ranging from 84% to 94%. The integration of both spectral and topographic data derived from three sets of images gave the highest accuracy map (94.26%). The classification results also suggest that the middle period of the growing season is better than the early or late seasons for image acquisitions for coastal dune systems in Ireland. The high accuracies obtained in this study highlight the application of this robust methodology to aid the spatiotemporal monitoring of Annex I habitats.

2.7. References

Aghababaei, M., Ebrahimi, A., Naghipour, A. A., Asadi, E., & Verrelst, J. (2021).
Vegetation types mapping using multi-temporal Landsat images in the Google

- Earth Engine platform. *Remote Sensing*, 13, 4683.
<https://doi.org/10.3390/rs13224683>
- Agüera-Vega, F., Agüera-Puntas, M., Martínez-Carricondo, P., Mancini, F., & Carvajal, F. (2020). Effects of point cloud density, interpolation method and grid size on derived Digital Terrain Model accuracy at micro topography level. *International Journal of Remote Sensing*, 41(21), 8281–8299.
<https://doi.org/10.1080/01431161.2020.1771788>
- Alvarez-Vanhard, E., Houet, T., Mony, C., Lecoq, L., & Corpetti, T. (2020). Can UAVs fill the gap between in situ surveys and satellites for habitat mapping? *Remote Sensing of Environment*, 243, 111780. <https://doi.org/10.1016/j.rse.2020.111780>
- Amani, M., Salehi, B., Mahdavi, S., Granger, J. E., Brisco, B., & Hanson, A. (2017). Wetland classification using multi-source and multi-temporal optical remote sensing data in Newfoundland and Labrador, Canada. *Canadian Journal of Remote Sensing*, 43(4), 360–373. <https://doi.org/10.1080/07038992.2017.1346468>
- Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, 11(3), 138–146. <https://doi.org/10.1890/120150>
- Assmann, J. J., Kerby, J. T., Cunliffe, A. M., & Myers-Smith, I. H. (2019). Vegetation monitoring using multispectral sensors — Best practices and lessons learned from high latitudes. *Journal of Unmanned Vehicle Systems*, 7(1), 54–75.
<https://doi.org/10.1139/juvs-2018-0018>
- Barnes, E. M., Clarke, T. R., Richards, S. E., Collaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., & Moran, M. S. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. *Proceedings of the 5th International Conference on Precision Agriculture and other resource management*. Precision Agriculture Center.
- Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C., & Silliman, B. R. (2011). The value of estuarine and coastal ecosystem services. *Ecological Monographs*, 81(2), 169–193. <https://doi.org/10.1890/10-1510.1>

- Barrett, B., Raab, C., Cawkwell, F., & Green, S. (2016). Upland vegetation mapping using Random Forests with optical and radar satellite data. *Remote Sensing in Ecology and Conservation*, 2(4), 212–231. <https://doi.org/10.1002/rse2.32>
- Bazzichetto, M., Malavasi, M., Acosta, A. T. R., & Carranza, M. L. (2016). How does dune morphology shape coastal EC habitats occurrence? A remote sensing approach using airborne LiDAR on the Mediterranean coast. *Ecological Indicators*, 71, 618–626. <https://doi.org/10.1016/j.ecolind.2016.07.044>
- Belcore, E., Pittarello, M., Lingua, A. M., & Lonati, M. (2021). Mapping riparian habitats of Natura 2000 network (91E0*, 3240) at individual tree level using UAV multi-temporal and multi-spectral data. *Remote Sensing*, 13, 1756. <https://doi.org/10.3390/rs13091756>
- Belgiu, M., & Drăgut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Berrar, D. (2019). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Bird, E. (2008). *Coastal geomorphology: An introduction* (2nd ed.). John Wiley & Sons Ltd.
- Bird, E. (2010). *Encyclopedia of the world's coastal landforms*. Springer Science & Business Media.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st ed.). CRC Press.
- Cohen, O., & Héquette, A. (2020). Recent advances in coastal survey techniques: From GNSS to LiDAR and digital photogrammetry - Examples on the northern coast of France. In J. P. Tiefenbacher & D. Poreh (Eds.), *Spatial variability in environmental science - patterns, processes, and analyses* (pp. 1–22). IntechOpen.

- Commission of the European Communities. (1992). Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. *Official Journal of the European Union*, 206, 7-50.
- Connolly, J. (2018). Mapping land use on Irish peatlands using medium resolution satellite imagery. *Irish Geography*, 51(2), 187–204. <https://irishgeography.ie/index.php/irishgeography/article/view/1371>
- Cruzan, M. B., Weinstein, B. G., Grasty, M. R., Kohrn, B. F., Hendrickson, E. C., Arredondo, T. M., & Thompson, P. G. (2016). Small unmanned aerial vehicles (micro-UAVs, drones) in plant ecology. *Applications in Plant Sciences*, 4(9), 1600041. <https://doi.org/10.3732/apps.1600041>
- Delaney, A., Devaney, F. M., Martin, J. M., & Barron, S. J. (2013). *Monitoring survey of Annex I sand dune habitats in Ireland (Irish Wildlife Manuals No. 75)*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland. <https://www.npws.ie/sites/default/files/publications/pdf/IWM75.pdf>
- Doody, J. P. (2013). *Sand dune conservation , management and restoration*. Springer. <https://doi.org/10.1007/978-94-007-4731-9>
- Ettritch, G., Bunting, P., Jones, G., & Hardy, A. (2018). Monitoring the coastal zone using earth observation: Application of linear spectral unmixing to coastal dune systems in Wales. *Remote Sensing in Ecology and Conservation*, 4(4), 303–319. <https://doi.org/10.1002/rse2.79>
- European Commission. (2013). *Interpretation manual of European Union habitats EUR 28*. http://ec.europa.eu/environment/nature/legislation/habitatsdirective/docs/Int_Manual_EU28.pdf
- European Commission. (2022). *Proposal for a regulation of the European Parliament and of the Council on nature restoration*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0304&from=EN>
- Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: From early developments to recent advancements. *Systems Science and Control Engineering:*

- An Open Access Journal*, 2(1), 602–609.
<https://doi.org/10.1080/21642583.2014.956265>
- Feng, Q., Liu, J., & Gong, J. (2015). UAV remote sensing for urban vegetation mapping using Random Forest and texture analysis. *Remote Sensing*, 7, 1074–1094.
<https://doi.org/10.3390/rs70101074>
- Fossitt, J. A. (2000). *A guide to habitats in Ireland*. The Heritage Council.
- Fraser, R. H., Olthof, I., Lantz, T. C., & Schmitt, C. (2016). UAV photogrammetry for mapping vegetation in the low-Arctic. *Arctic Science*, 2(3), 79–102.
<https://doi.org/10.1139/as-2016-0008>
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300.
<https://doi.org/10.1016/j.patrec.2005.08.011>
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment*, 58(3), 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Gonçalves, J., Henriques, R., Alves, P., Sousa-Silva, R., Monteiro, A. T., Lomba, Â., Marcos, B., & Honrado, J. (2016). Evaluating an unmanned aerial vehicle-based approach for assessing habitat extent and condition in fine-scale early successional mountain mosaics. *Applied Vegetation Science*, 19, 132–146.
<https://doi.org/10.1111/avsc.12204>
- Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., Li, Z., & Yang, X. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods*, 15, 10.
<https://doi.org/10.1186/s13007-019-0394-z>
- Hill, R. A., Wilson, A. K., George, M., & Hinsley, S. A. (2010). Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science*, 13(1), 86–99. <https://doi.org/10.1111/J.1654-109X.2009.01053.X>

- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Hughes, M. G., Knight, J., & Masselink, G. (2014). *Introduction to coastal processes and geomorphology* (2nd ed.). Routledge.
- Jensen, J. R. (2014). *Remote sensing of the environment: an earth resource perspective* (2nd ed.). Pearson Education Limited.
- Jiang, Z., Huete, A. R., Kim, Y., & Didan, K. (2007). 2-band enhanced vegetation index without a blue band and its application to AVHRR data. *Proc. SPIE 6679, Remote Sensing and Modeling of Ecosystems for Sustainability IV*, 667905. <https://doi.org/10.1117/12.734933>
- Jiang, J., Johansen, K., Tu, Y. H., & McCabe, M. F. (2022). Multi-sensor and multi-platform consistency and interoperability between UAV, Planet CubeSat, Sentinel-2, and Landsat reflectance data. *GIScience and Remote Sensing*, 59(1), 936–958. <https://doi.org/10.1080/15481603.2022.2083791>
- Jolliffe, I. T. (2002). *Principal Components Analysis* (2nd ed.). Springer. <https://doi.org/10.1007/b98835>
- Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, 50(4), 663–666. <https://doi.org/10.2307/1936256>
- Kattenborn, T., Eichel, J., & Fassnacht, F. E. (2019). Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific Reports*, 9, 17656. <https://doi.org/10.1038/s41598-019-53797-9>
- Kattenborn, T., Lopatin, J., Förster, M., Braun, A. C., & Fassnacht, F. E. (2019). UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. *Remote Sensing of Environment*, 227, 61–73. <https://doi.org/10.1016/j.rse.2019.03.025>

- Khoshgoftaar, T. M., Golawala, M., & Van Hulse, J. (2007). An empirical study of learning from imbalanced data using Random Forest. *19th IEEE International Conference on Tools with Artificial Intelligence*, 310–317.
<https://doi.org/10.1109/ICTAI.2007.46>
- Laporte-Fauret, Q., Lubac, B., Castelle, B., Michalet, R., Marieu, V., Bombrun, L., Launeau, P., Giraud, M., Normandin, C., & Rosebery, D. (2020). Classification of Atlantic coastal sand dune vegetation using in situ, UAV, and airborne hyperspectral data. *Remote Sensing*, 12(14), 2222.
<https://doi.org/10.3390/rs12142222>
- Laporte-Fauret, Q., Marieu, V., Castelle, B., Michalet, R., Bujan, S., & Rosebery, D. (2019). Low-cost UAV for high-resolution and large-scale coastal dune change monitoring using photogrammetry. *Journal of Marine Science and Engineering*, 7(3), 63. <https://doi.org/10.3390/jmse7030063>
- Lu, B., & He, Y. (2017). Species classification using Unmanned Aerial Vehicle (UAV)-acquired high spatial resolution imagery in a heterogeneous grassland. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128, 73–85.
<https://doi.org/10.1016/j.isprsjprs.2017.03.011>
- Mack, E. L., Firbank, L. G., Bellamy, P. E., Hinsley, S. A., & Veitch, N. (1997). The comparison of remotely sensed and ground-based habitat area data using species-area models. *Journal of Applied Ecology*, 34(5), 1222–1228.
<https://doi.org/10.2307/2405233>
- Manfreda, S., McCabe, M. F., Miller, P. E., Lucas, R., Madrigal, V. P., Mallinis, G., Dor, E. Ben, Helman, D., Estes, L., Ciraolo, G., Müllerová, J., Tauro, F., de Lima, M. I., de Lima, J. L. M. P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., ... Toth, B. (2018). On the use of unmanned aerial systems for environmental monitoring. *Remote Sensing*, 10(4), 641. <https://doi.org/10.3390/rs10040641>
- Marcinkowska-Ochtyra, A., Gryguc, K., Ochtyra, A., Kopeć, D., Jarocińska, A., & Sławik, Ł. (2019). Multitemporal hyperspectral data fusion with topographic indices—Improving classification of Natura 2000 grassland habitats. *Remote Sensing*, 11(19), 2264. <https://doi.org/10.3390/rs11192264>

- Marzioletti, F., Giulio, S., Malavasi, M., Sperandii, M. G., Acosta, A. T. R., & Carranza, M. L. (2019). Capturing coastal dune natural vegetation types using a phenology-based mapping approach: The potential of Sentinel-2. *Remote Sensing*, 11(12), 1506. <https://doi.org/10.3390/rs11121506>
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Mesas-Carrascosa, F. J., Torres-Sánchez, J., Clavero-Rumbao, I., García-Ferrer, A., Peña, J. M., Borra-Serrano, I., & López-Granados, F. (2015). Assessing optimal flight parameters for generating accurate multispectral orthomosaicks by UAV to support site-specific crop management. *Remote Sensing*, 7(10), 12793–12814. <https://doi.org/10.3390/rs71012793>
- Miao, X., Heaton, J. S., Zheng, S., Charlet, D. A., & Liu, H. (2012). Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data. *International Journal of Remote Sensing*, 33(6), 1823–1849. <https://doi.org/10.1080/01431161.2011.602651>
- Michez, A., Piégay, H., Lisein, J., Claessens, H., & Lejeune, P. (2016). Classification of riparian forest species and health condition using multi-temporal and hyperspatial imagery from unmanned aerial system. *Environmental Monitoring and Assessment*, 188, 146. <https://doi.org/10.1007/s10661-015-4996-2>
- Muller, A. C., & Guido, S. (2017). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media Inc.
- Müllerová, J., Gago, X., Bučas, M., Company, J., Estrany, J., Fortesa, J., Manfreda, S., Michez, A., Mokroš, M., Paulus, G., Tiškus, E., Tsiafouli, M. A., & Kent, R. (2021). Characterizing vegetation complexity with unmanned aerial systems (UAS) – A framework and synthesis. *Ecological Indicators*, 131, 108156. <https://doi.org/10.1016/j.ecolind.2021.108156>
- National Parks and Wildlife Service. (2008). *The status of EU protected habitats and species in Ireland*. Department of the Environment, Heritage and Local Government, Ireland.

- National Parks and Wildlife Service. (2013a). *Site synopsis: Tralee Bay and Magharees Peninsula, West to Cloghane SAC* (pp. 1–6). Department of Arts, Heritage and the Gaeltacht, Ireland.
- National Parks and Wildlife Service. (2013b). *The status of EU protected habitats and species in Ireland*. Department of Arts, Heritage and the Gaeltacht, Ireland.
- National Parks and Wildlife Service. (2018). *Natura 2000 Standard Data Form: Tralee Bay and Magharees Peninsula, West to Cloghane SAC*. Department of Arts, Heritage and the Gaeltacht, Ireland.
- National Parks and Wildlife Service. (2019). *The status of EU protected habitats and species in Ireland. Volume 1: Summary overview*. Department of Culture, Heritage and the Gaeltacht, Ireland.
- Pajares, G. (2015). Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogrammetric Engineering and Remote Sensing*, 81(4), 281–330. <https://doi.org/10.14358/PERS.81.4.281>
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. <https://doi.org/10.1080/01431160412331269698>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119–126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- Raab, C., Stroh, H. G., Tonn, B., Meißner, M., Rohwer, N., Balkenhol, N., & Isselstein, J. (2018). Mapping semi-natural grassland communities using multi-temporal RapidEye remote sensing data. *International Journal of Remote Sensing*, 39(17), 5638–5659. <https://doi.org/10.1080/01431161.2018.1504344>

- Riihimäki, H., Luoto, M., & Heiskanen, J. (2019). Estimating fractional cover of tundra vegetation at multiple scales using unmanned aerial systems and optical satellite data. *Remote Sensing of Environment*, 224, 119–132.
<https://doi.org/10.1016/j.rse.2019.01.030>
- Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. In S. C. Freden, E. P. Mercanti, & M. A. Becker (Eds.), *Third Earth Resources Technology Satellite-1 Symposium: The proceedings of a symposium held by Goddard Space Flight Center at Washington, D.C. on December 10-14, 1973* (pp. 309–317). National Aeronautics and Space Administration
- Schmidt, T., Schuster, C., Kleinschmit, B., & Förster, M. (2014). Evaluating an intra-annual time series for grassland classification - How many acquisitions and what seasonal origin are optimal? *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8), 3428–3439.
<https://doi.org/10.1109/JSTARS.2014.2347203>
- Scholefield, P., Morton, D., McShane, G., Carrasco, L., Whitfield, M. G., Rowland, C., Rose, R., Wood, C., Tebbs, E., Dodd, B., & Monteith, D. (2019). Estimating habitat extent and carbon loss from an eroded northern blanket bog using UAV derived imagery and topography. *Progress in Physical Geography: Earth and Environment*, 43(2), 282–298. <https://doi.org/10.1177/0309133319841300>
- Sotille, M. E., Bremer, U. F., Vieira, G., Velho, L. F., Petsch, C., Auger, J. D., & Simões, J. C. (2022). UAV-based classification of maritime Antarctic vegetation types using GEOBIA and random forest. *Ecological Informatics*, 71, 101768.
<https://doi.org/10.1016/j.ecoinf.2022.101768>
- Suo, C., McGovern, E., & Gilmer, A. (2019). Coastal dune vegetation mapping using a multispectral sensor mounted on an UAS. *Remote Sensing*, 11(15), 1814.
<https://doi.org/10.3390/rs11151814>
- Taddia, Y., Corbau, C., Zambello, E., & Pellegrinelli, A. (2019). UAVs for structure-from-motion coastal monitoring: A case study to assess the evolution of embryo dunes over a two-year time frame in the Po River Delta, Italy. *Sensors*, 19(7),

1717. <https://doi.org/10.3390/s19071717>
- van Iersel, W., Straatsma, M., Middelkoop, H., & Addink, E. (2018). Multitemporal classification of river floodplain vegetation using time series of UAV images. *Remote Sensing*, 10(7), 1144. <https://doi.org/10.3390/rs10071144>
- Vanden Borre, J., Haest, B., Lang, S., Spanhove, T., Förster, M., & Sifakis, N. I. (2011). Towards a wider uptake of remote sensing in Natura 2000 monitoring: Streamlining remote sensing products with users' needs and expectations. *2nd International Conference on Space Technology, ICST 2011*, 1–4. <https://doi.org/10.1109/ICSpT.2011.6064686>
- Vilar, P., Morais, T. G., Rodrigues, N. R., Gama, I., Monteiro, M. L., Domingos, T., & Teixeira, R. F. M. (2020). Object-based classification approaches for multitemporal identification and monitoring of pastures in agroforestry regions using multispectral unmanned aerial vehicle products. *Remote Sensing*, 12(5), 814. <https://doi.org/10.3390/rs12050814>
- Villoslada, M., Bergamo, T. F., Ward, R. D., Burnside, N. G., Joyce, C. B., Bunce, R. G. H., & Sepp, K. (2020). Fine scale plant community assessment in coastal meadows using UAV based multispectral data. *Ecological Indicators*, 111, 105979. <https://doi.org/10.1016/j.ecolind.2019.105979>
- Ward, R. D., Burnside, N. G., Joyce, C. B., & Sepp, K. (2013). The use of medium point density LiDAR elevation data to determine plant community types in Baltic coastal wetlands. *Ecological Indicators*, 33, 96–104. <https://doi.org/10.1016/j.ecolind.2012.08.016>
- Wood, D. J. A., Preston, T. M., Powell, S., & Stoy, P. C. (2022). Multiple UAV flights across the growing season can characterize fine scale phenological heterogeneity within and among vegetation functional groups. *Remote Sensing*, 14(5), 1290. <https://doi.org/10.3390/rs14051290>
- Wood, E. M., Pidgeon, A. M., Radeloff, V. C., & Keuler, N. S. (2012). Image texture as a remotely sensed measure of vegetation structure. *Remote Sensing of Environment*, 121, 516–526. <https://doi.org/10.1016/j.rse.2012.01.003>

- Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, 2017, 1353691. <https://doi.org/10.1155/2017/1353691>
- Yan, L., Gou, Z., & Duan, Y. (2009). A UAV remote sensing system: Design and tests. In D. Li, J. Shan, & J. Gong (Eds.), *Geospatial technology for earth observation* (pp. 27–44). Springer Science & Business Media. <https://doi.org/10.1007/978-1-4419-0050-0>
- Yousefi Lalimi, F., Silvestri, S., Moore, L. J., & Marani, M. (2017). Coupled topographic and vegetation patterns in coastal dunes: Remote sensing observations and ecomorphodynamic implications. *Journal of Geophysical Research: Biogeosciences*, 122(1), 119–130. <https://doi.org/10.1002/2016JG003540>

3. Improving the mapping of coastal invasive species using UAV imagery and deep learning



UAV survey on North Bull Island (left) and sample UAV-acquired image of a saltmarsh site. (Photos: Charmaine Cruz)

This chapter is based on the peer-reviewed article:

Cruz, C., McGuinness, K., Perrin, P. M., O’Connell, J., Martin, J. R., Connolly, J. (2023). Improving the mapping of coastal invasive species using UAV imagery and deep learning, *International Journal of Remote Sensing*, 44(18), 5713–5735. <https://doi.org/10.1080/01431161.2023.2251186>

Description of authors' contributions:

C.C. designed the research and performed the analysis. C.C. and P.M.P. performed the data acquisition. C.C. prepared the original draft manuscript. J.C., K.M., P.M.P., J.M. and J.O.C. provided a critical review of the manuscript. C.C. led the submission of the manuscript to the journal. C.C. revised the manuscript based on anonymous reviewers' comments and suggestions.

3.1. Abstract

Spartina anglica C. E. Hubb. is an invasive species of saltmarsh and mudflats in Ireland. It spreads quickly and can form extensive meadows, which can lead to the extinction of native plant and animal species. Traditional ground-based field surveys would be expensive for the scale of monitoring needed to formulate site-specific management strategies. Advanced mapping techniques, such as Unoccupied Aerial Vehicle (UAV) remote sensing and deep learning (DL), offer an opportunity to automatically map invasive species' occurrence. However, training a DL model requires lots of labelled data which can be prohibitively difficult and expensive to obtain. We implemented a DL semantic segmentation technique on UAV imagery of *Spartina*-invaded habitats and tested a range of hyperparameters—model architectures, encoder backbones and input image patch sizes—to determine an effective segmentation network structure. We also investigated applying data augmentation and pseudo-labelling techniques to increase the size of the labelled dataset. U-Net architecture with Inception-v3 as its backbone trained on 128×128 -pixel image patches offered the best model performance: mean Intersection-Over-Union (*mIOU*) score = 0.832. The model trained on the combined augmented and pseudo-labelled data achieved an *mIOU* score of 0.712 on a test dataset, while there was a decrease of 0.158 in model performance when only the original labelled data were used. This result suggests the potential for using these techniques in creating more robust models. The proposed methodology demonstrates that the combination of UAV imagery and deep learning could be a promising tool for mapping the distribution of invasive plant species.

3.2. Introduction

Invasive alien species (IAS) are organisms found in an environment outside their natural range and cause negative impacts on this new environment (IUCN, 2000). They are considered amongst the major drivers of global biodiversity loss as they can alter ecosystem functions and services and cause species richness decline (Mollot et al., 2017). The adverse impacts of IAS have been recognised by international organisations, such as the International Union for Conservation of Nature (IUCN, 2000) and the European Union (EU) through the *EU regulation on the prevention and management of the introduction and spread of IAS* (European Commission, 2014). In Ireland's most recent EU Habitats Directive report, the presence of IAS is one of the major threats to and pressures on habitats, affecting more than 40% of all reported habitats (National Parks and Wildlife Service, 2019). Furthermore, the estimated cost of IAS to the Irish economy was €202 million in 2013 (Kelly et al., 2013).

In general, assessment of IAS requires accurate maps to monitor their location and expansion over time. These maps can also be used to identify hotspots where management efforts can be focused, as well as measure the success of control programmes. The information is commonly obtained by traditional ground-based field surveys, which can be costly, spatially restricted and involve considerable time and effort. Furthermore, the rate of IAS introductions is increasing globally, which can limit the potential of manual methods alone for large-scale monitoring (Seebens et al., 2017). Consequently, an alternative is needed, a more efficient and repeatable technique to improve the way of capturing and monitoring the spatial distribution of IAS, especially those with the highest environmental and socioeconomic impact (Nentwig et al., 2018).

Spartina anglica C.E. Hubb (= *Sporobolus anglicus* (C.E. Hubb.) P. M. Peterson & Saarela), also known as common cordgrass, is on the list of the most problematic IAS in the EU (Nentwig et al., 2018) and globally (Lowe et al., 2000). This perennial grass, commonly found in saltmarsh and mudflat habitats, spreads by seed dispersal or vegetatively by rhizomes and can form extensive meadows (Figure 3.1) (Nehring and Adsersen, 2006). In Ireland, *S. anglica* was first introduced in Cork Harbour to bind and stabilise sediment (McCorry and Ryle, 2009); it has then subsequently spread

to other locations on the Irish coast. It invades and displaces EU Habitats Directive Annex I habitats, including tidal mudflats and sandflats, Atlantic salt meadows and *Salicornia* beds (National Parks and Wildlife Service, 2019). This reduces feeding habitats for birds (Stokes et al., 2006). Monitoring the location and extent of *S. anglica* is an important step in assessing its rate of expansion (Brophy et al., 2019).



Figure 3.1. *Spartina anglica* (at the back) displaces low-growing native *Salicornia* plants (in the foreground) (photo taken at North Bull Island, Ireland; August 2021).

Advances in mapping technology, such as remote sensing, offer a potential solution to monitoring *S. anglica* (Huang and Asner, 2009). Medium-resolution satellite imagery, typically > 10 m/pixel, has been used to detect *S. anglica* covering large areas (J. Li et al., 2010; Zuo et al., 2012). However, it is too coarse for detailed mapping, such as delineating *S. anglica* when mixed with native species or detecting expansion in the form of small clonal patches. Early detection of *S. anglica* expansion is considered the best approach to prevent further spread, to minimise the cost of controlling the plant and to

increase the success of eradicating it (Brophy et al., 2019). Mapping small patches of *S. anglica* requires submetre spatial resolution imagery.

In recent years, Unoccupied Aerial Vehicles (UAVs) have become an invaluable remote sensing platform for mapping and monitoring vegetation (Belcore et al., 2021; Beyer et al., 2019; Kattenborn et al., 2019). UAVs enable the acquisition of spatially continuous data with centimetre-level resolution. They can also be operated at low altitudes (i.e., below clouds). This means that countries with high levels of cloud cover, such as Ireland, can benefit from using UAVs for monitoring (Manfreda et al., 2018). Hence, UAVs can be considered a suitable platform for mapping *S. anglica* patches.

Previous remote sensing studies for mapping *S. anglica* used traditional machine-learning classifiers like Random Forest (Proença et al., 2019; Van Beijma et al., 2014). This classifier can process a large number of explanatory features that have been manually extracted (handcrafted) (Belgiu and Drăgut, 2016). However, choosing relevant features can be time-consuming and involves expert knowledge of the characteristics of the plant species. Deep learning (DL) can eliminate the need for these pre-specified handcrafted features (Y. Li et al., 2018). DL is a subfield of machine learning that aims to learn complex patterns and automatically extract high-level features from data rather than manually selecting them (Goodfellow et al., 2016; Y. Li et al., 2018). In many cases, DL requires large amounts of labelled training data to produce a robust model. The acquisition and building of manually labelled datasets can require considerable effort, costs and time. These requirements can be prohibitive, therefore, a DL approach may not always appear feasible.

Data augmentation and semi-supervised learning (SSL) are techniques that can reduce the reliance on large amounts of labelled data. Data augmentation does this by applying transformations to the existing data, thus increasing the amount and diversity of labelled data. Common augmentation techniques involve geometric transformations, including flips, rotations, translations and cropping. More advanced techniques to augment data, including colour space modifications, have recently been developed (Shorten and Khoshgoftaar, 2019). In SSL, both labelled and unlabelled data are used to train a model instead of solely using labelled data. Pseudo-labelling (Lee, 2013) is an approach to SSL that deals with unlabelled data. It uses a model initially trained on labelled data to make

predictions for unlabelled data. These newly labelled data are integrated with those previously labelled, and the model is retrained. This process is then iterated. Since it is tedious and expensive, in terms of cost and time, to label images for DL tasks but cheap to gather unlabelled data, the model takes advantage of both the labelled and unlabelled data. However, the DL studies that have investigated the benefits of augmentation and pseudo-labelling techniques have mostly been related to medical images (Chlap et al., 2021; Mao et al., 2022) and other natural images of common objects (Perez and Wang, 2017; Zheng et al., 2020). The effectiveness of applying these techniques to UAV-acquired images for mapping invasive species with an aim to improve accuracy and prevent overfitting, hence improving transferability, has not yet been fully explored.

Semantic segmentation is an image-processing technique that aims to assign a semantic label to every pixel in the image (Yu et al., 2018). This technique has gained significant attention due to its application of DL in many computer vision tasks, such as x-ray image segmentation for medical diagnosis (Ahmed et al., 2020) and road segmentation for self-driving cars (Sharma et al., 2019). It is also becoming popular for processing remote sensing data for many environmental applications, such as vegetation mapping (Balado et al., 2021; Kattenborn et al., 2019; Osco et al., 2021). In the context of IAS, semantic segmentation could be a foundation for applications such as food security and habitat degradation monitoring. However, due to the need to account for factors such as variable field conditions and plant growth stages, accurate segmentation of plant species presents a challenge in DL.

This study develops an automatic segmentation model for mapping *S. anglica* using UAV images. The objectives of this study are: (1) to select an effective segmentation network structure by comparing the performance differences of three network hyperparameters: model architecture, encoder backbone and input image patch size, (2) to propose a method of combining data augmentation and pseudo-labelling techniques for segmentation of UAV-acquired images, (3) to assess the performance of the proposed method, and (4) to evaluate the performance of the final segmentation model on a dataset acquired using a different sensor. The aim is that by using this approach, the mapping and monitoring of *S. anglica* and other invasive species can be substantially automated.

3.3. Materials and methods

The methodology is divided into four stages: data collection, data preparation, model development and selection and model evaluation (Figure 3.2). The first stage consists of acquiring UAV images and field data, and the second stage is transforming these datasets into a format suitable for training a deep learning model. The third stage includes examining the effects of applying different modifications (patch size, architecture, pre-trained model, augmentation technique and pseudo-labelling) to the model to facilitate the model selection. The fourth stage evaluates the final model to see its generalisation performance when applied to new and unseen data.

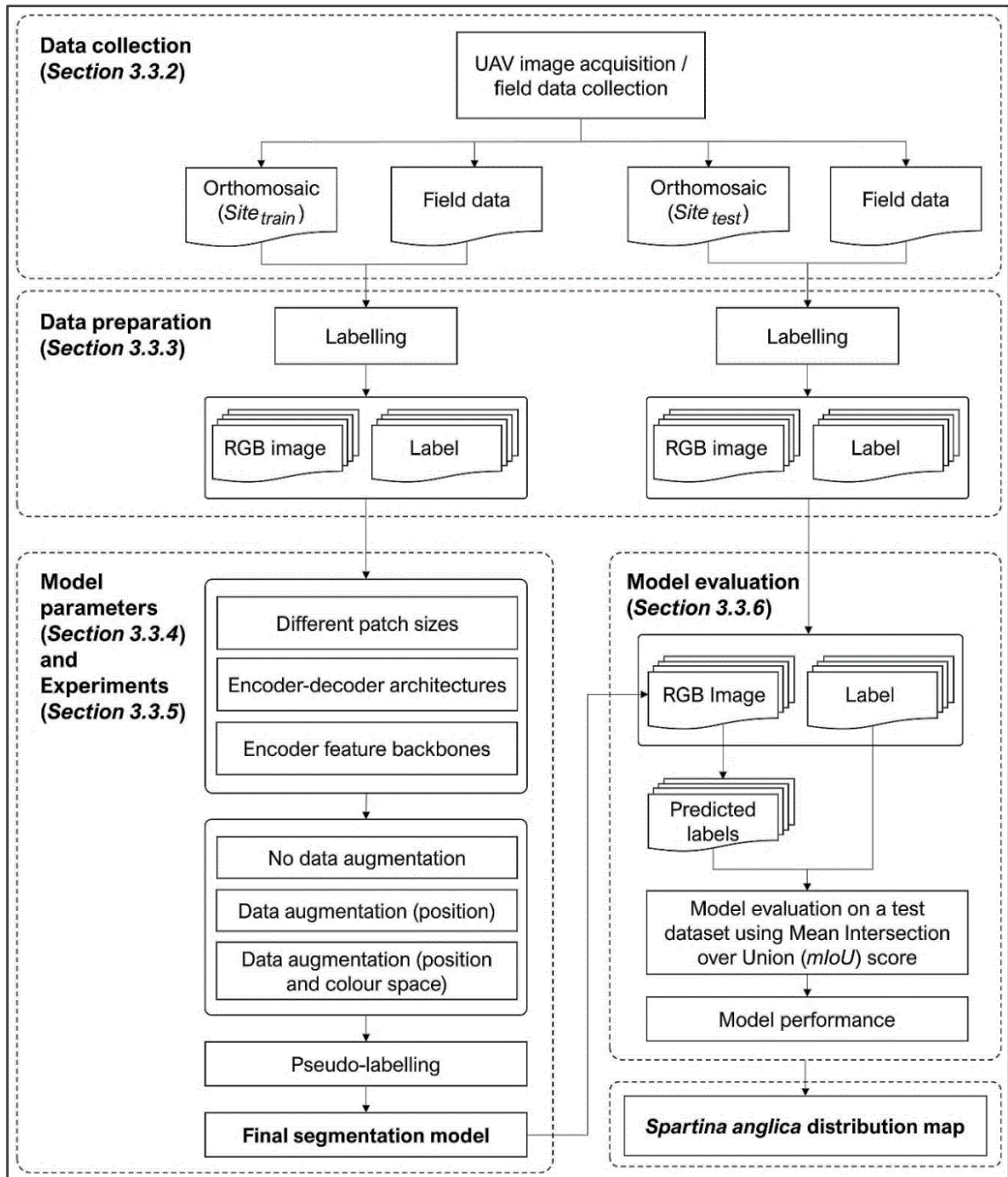


Figure 3.2. Workflow using a semantic segmentation deep learning algorithm.

3.3.1. Study site description

North Bull Island is a wedge-shaped, low-lying coastal ‘barrier island’ located in the northern part of Dublin Bay on the east coast of Ireland (Figure 3.3) (Mathew et al., 2019). It is part of the UNESCO-designated Dublin Bay Biosphere Reserve and the North Dublin Bay Special Area of Conservation site (SAC code 000206, approximately 1,474 ha in

area) of the Natura 2000 network. This SAC includes four saltmarsh habitats listed under Annex I of the EU Habitats Directive (National Parks & Wildlife Service, 2013). However, *S. anglica* is also present, forming extensive meadows and smaller patches.

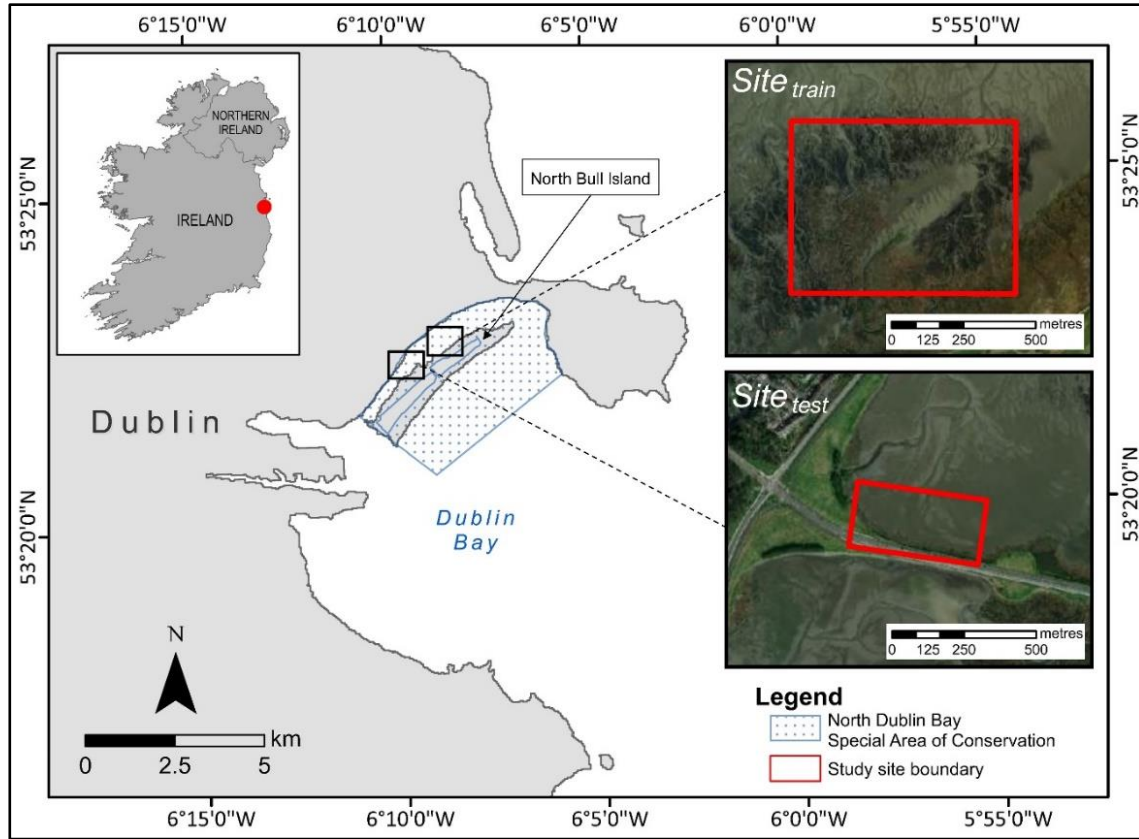


Figure 3.3. Map location of the study area on North Bull Island, Ireland (left) and close-ups of the training and test study sites (right).

In this study, two sites were chosen to assess how well the method would perform in delineating *S. anglica*. These sites were located on the sheltered (western) side of North Bull Island within the saltmarsh habitat. The first site (*Site_{train}*) was chosen for model training and selection, while the second site (*Site_{test}*) was for model evaluation. This way, there was no overlap between model training and evaluation datasets to avoid any bias in predictions.

3.3.2. Data collection

Field data

Extensive field surveys were conducted by an experienced ecologist to record locations of *S. anglica* point samples using an Emlid Reach RS+ Global Navigation Satellite System (GNSS) device (<https://emlid.com/reachrs/>). Each point was located at the centre of a homogeneous area of *S. anglica* cover. The radius of the area around each point, typically 1-10 m, was estimated and recorded by the ecologist. The point data were then converted into polygons using the recorded radii. These polygons were used to guide the preparation of reference data (labels) in Section 3.3.3.

UAV images

Three field campaigns were conducted to acquire high-resolution (5 cm) UAV imagery, two for *Site_{train}* and one for *Site_{test}*. For *Site_{train}*, both flights (August 2019 and October 2021) used a Micasense Altum multispectral sensor with five spectral bands: blue, green, red, red edge and near-infrared (MicaSense, 2019). For *Site_{test}*, a digital camera DJI Zenmuse P1 ([dji.com/ie/zenmuse-p1/specs](https://www.dji.com/ie/zenmuse-p1/specs)) was used to acquire imagery in October 2022. Due to different acquisition dates and different camera sensors used, there were variations in the captured images in terms of the appearance and colour of the vegetation (Figure 3.4).

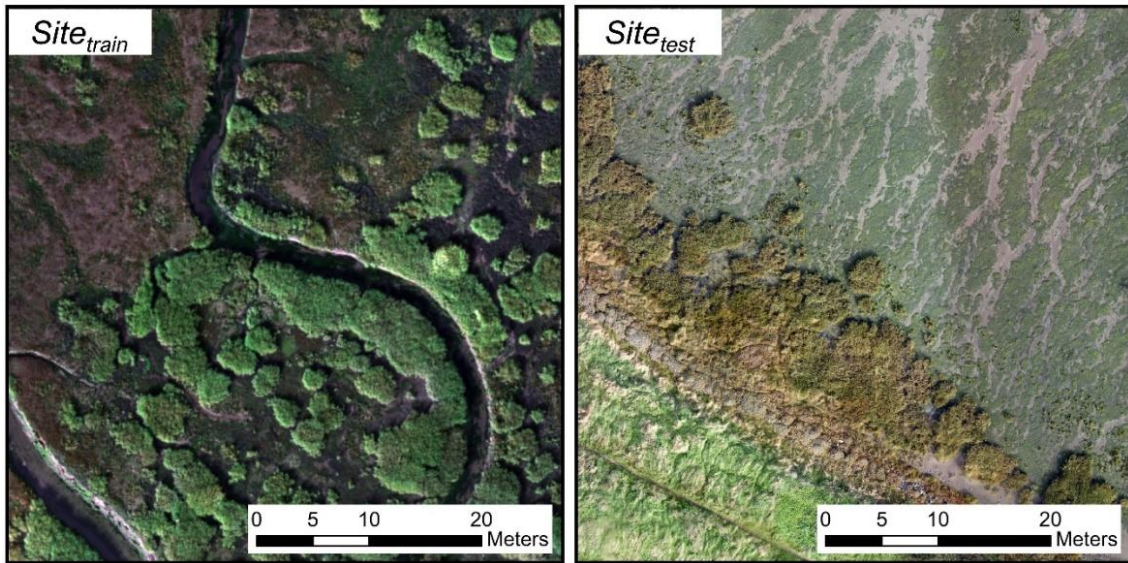


Figure 3.4. Comparison between the UAV images taken at *Site_{train}* (left) and *Site_{test}* (right).

Each set of UAV imagery was processed using the Structure-from-Motion technique in Pix4D software (Pix4D, Switzerland, version 4.3.33) to generate an orthomosaic. The processing involves keypoint extraction, keypoint matching between images, camera model optimisation, geolocation based on GPS flight trajectories and dense point cloud generation. For this study, only red, green and blue (RGB) image bands were utilised in the modelling because RGB images are relatively easy to acquire using cheap hardware and would therefore represent a more financially viable option for invasive plant managers when applying this technique.

The *Site_{train}* orthomosaics were subdivided into three sets: *train*, *validation* and *unlabelled*. The *train* and *validation* sets were used to facilitate the DL training process, while the *unlabelled* set was for the pseudo-labelling (see Section 3.3.5). On the other hand, the *Site_{test}* orthomosaic was used to evaluate the generalisation performance of the final model (*test* set). Table 3.1 summarises the information on the imagery used in this study. The study used different sensors, hence the ability of the model to generalise across sensors was tested.

Table 3.1. Information on the acquired UAV imagery.

Site	Sensor	Acquisition date	Purpose
$Site_{train}$	Micasense Altum	18 August 2019 4 October 2021	for model training, selection and semi-supervised learning
$Site_{test}$	DJI Zenmuse P1	12 October 2022	for model evaluation

3.3.3. Data preparation

The reference data (labels) were prepared by manually delineating *S. anglica* boundaries (in the form of polygons) in a GIS environment (Figure 3.5 left). This delineation was based on the acquired field data and used image interpretation. A binary image was created where pixels belonging to *S. anglica* were assigned a value of 1, whereas pixels representing the background (i.e., exposed mud, macroalgae and other plant communities) were assigned a value of 0. Thus, the input for the modelling was a labelled image with two classes: *S. anglica* and the background (Figure 3.5 right). This labelling was done for the *train*, *validation* and *test* sets.

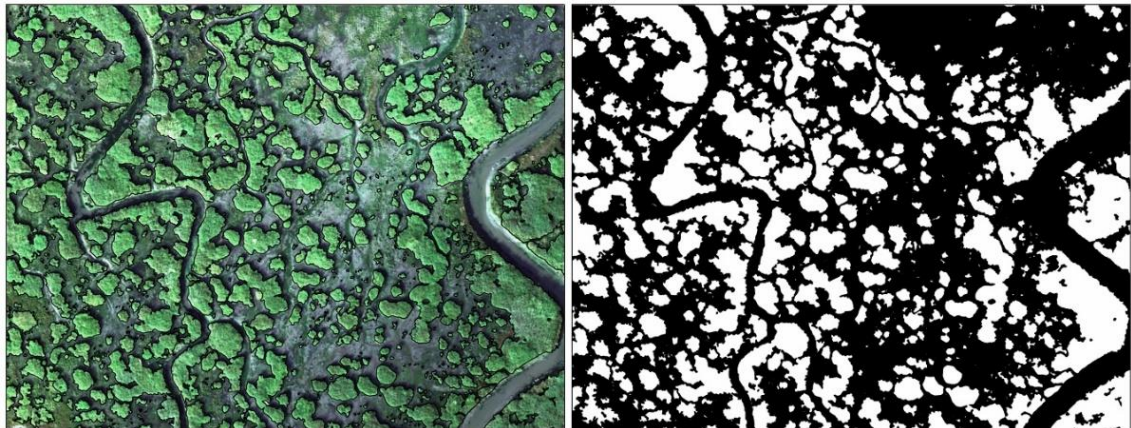


Figure 3.5. Sample RGB image and delineated polygons of *S. anglica* boundaries (left). Binary image representing the presence of *S. anglica* (white pixels) and its background (black pixels) (right).

3.3.4. Training parameters and evaluation metric

Semantic segmentation was implemented using the Segmentation models library of Python 3.7 based on Keras 2.9 and TensorFlow 2.9 (Yakubovskiy, 2019). The training phase involves passing the image patches into a deep learning network over a certain number of model iterations (epochs). Here, all models were trained in 100 epochs. The goal is to improve the result per epoch by optimising model weights. The model weights were only considered and saved if the loss value was lower than the value derived in the previous epoch. The Adam algorithm (Kingma and Ba, 2015), with a learning rate of 10^{-4} , was used in optimisation. Binary focal loss (Lin et al., 2018) was selected as the loss function. Given the binary classification problem (absence of *S. anglica* = 0, presence of *S. anglica* = 1), the final layer activation was set to a sigmoid function. We used 50% as the threshold to transform the results into a binary classification image (0-50%: class 0, 50.01-100%: class 1). All computations were performed on a local workstation using a CUDA-compatible GPU with 12 GB of GDDR6 memory (NVIDIA GeForce RTX 3060).

For model comparison and selection, we used the Jaccard index, also known as the Intersection-Over-Union (*IOU*) metric (Equation 3.1). The Jaccard index (Jaccard, 1908) is a well-known statistic used to measure the spatial similarity between two sample sets. This index is computed by dividing the area of overlap between the model-predicted image and the labelled image (predicted AND labelled) by the area of union between them (predicted OR labelled):

$$IOU = (Area\ of\ overlap)/(Area\ of\ union). \quad (Equation\ 3.1)$$

The value ranges from 0 to 1, where 0 indicates no overlap between the two, and 1 indicates a complete overlap. This means that the higher the *IOU* score, the better the model is. The *mean IOU (mIOU)* score was computed for each class by averaging all *IOU* scores of that class.

3.3.5. Experiments

Image patch size and network structure

Image patch size. The labelled *train* image was split into sub-images called patches, using a sliding kernel with 50% overlap. In this way, we created more training patches. Similarly, the labelled *validation* image was divided into patches but with no overlap. This experiment aimed to test the effect of image patch size on the model performance. Hence, we prepared two sets of patches of sizes: 128×128 and 256×256 pixels. As expected, fewer patches were generated with the larger patch size (Table 3.2).

Table 3.2. Information about the number of patches for the two patch sizes.

Image patch size	<i>train</i>	<i>validation</i>
128×128 pixels	4092	828
256×256 pixels	920	207

Model architecture and backbone. A general semantic segmentation network structure consists of an encoder and a decoder. An encoder is used to extract features from an image by downsampling, while a decoder is used to upsample the extracted features to their original image dimension. In this study, three encoder-decoder model architectures were implemented: U-Net (Ronneberger et al., 2015), LinkNet (Chaurasia and Culurciello, 2018) and Feature Pyramid Network (FPN) (Lin et al., 2017), to determine the model network structure to be used in segmentation. For the feature encoder backbones, we considered ten existing pre-trained models initialised with ImageNet weights (Deng et al., 2009) (Table 3.3). Further, each network structure was trained on the two image patch sizes— 128×128 and 256×256 pixels. Hence, 60 models were trained and compared in this section. Details of each pre-trained model can be found in the cited references (Table 3.3).

Table 3.3. List of encoder backbones used in this study.

Encoder backbone	Reference
VGG (VGG-16, VGG-19)	(Simonyan & Zisserman, 2015)
ResNet (ResNet-50, ResNet-101, ResNet-152)	(He et al., 2016)
ResNeXt (ResNeXt-50, ResNeXt-101)	(Xie et al., 2017)
DenseNet-121	(G. Huang et al., 2017)
Inception-v3	(Szegedy et al., 2016)
Inception-ResNet-v2	(Szegedy et al., 2017)

Using *mIOU* as the evaluation metric, this experiment showed that the U-Net architecture with Inception-v3 as its backbone (U-Net_Inception-v3) and trained on 128×128 -pixel image patches offered the best model performance. This will be further detailed in Section 3.4.1. Consequently, we used U-Net_Inception-v3 in all the subsequent processing.

Data augmentation and pseudo-labelling

Data augmentation. The *train* image patches were subjected to augmentation techniques using the Albumentations library (Buslaev et al., 2020) in Python. These techniques were grouped into two categories: position and colour space augmentations. In the first group, position augmentation, the orientation of the features on the image was altered. In this case, a rotation scheme was used. Rotation is a suitable augmentation technique for UAV images, as they can be viewed from any orientation above. In order to avoid border and resampling artefacts due to arbitrary rotation angles, an image patch (Figure 3.6a) was rotated around its centre using multiples of 90° (90° , 180° or 270°) (Figure 3.6b-d). This transformation also requires corresponding rotational changes to the labels. The second group, colour space augmentation, allows random changes in the brightness, hue and saturation of an image by applying the colour jitter technique (Figure 3.6e-h). The purpose of colour jitter is to simulate images acquired in different natural lighting conditions and at different acquisition dates, which will affect the appearance and colour of the vegetation. In contrast to the first group, this technique does not require any changes to the labels as there is no geometric change as there would be with a rotation.

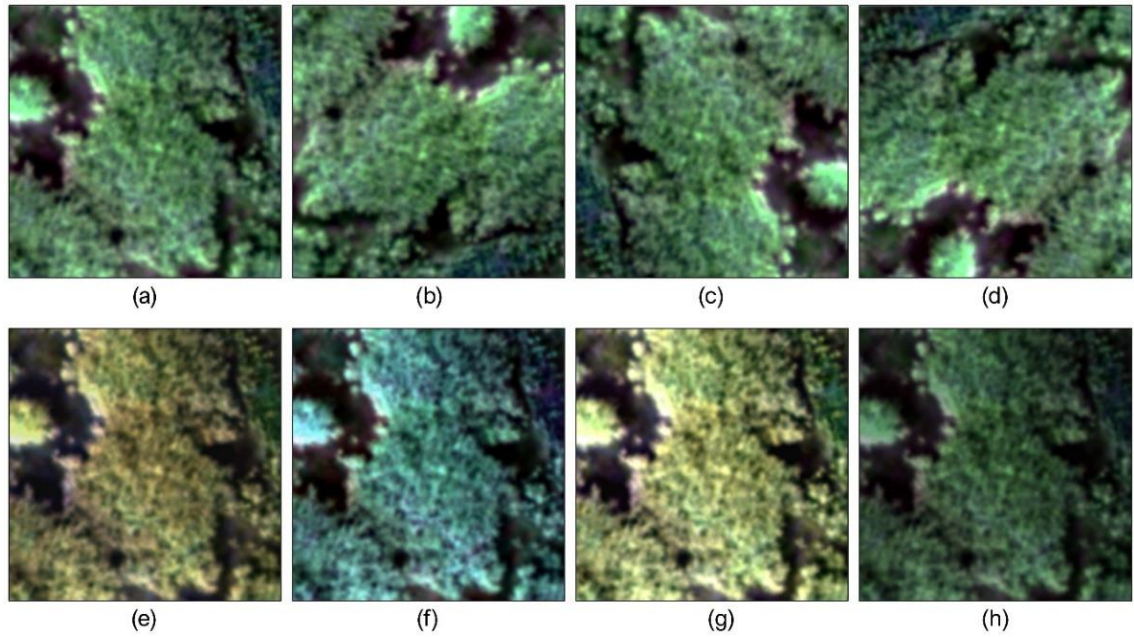


Figure 3.6. Original image patch (a) and generated images after multiples of 90° rotation (b-d) and colour jitter (e-h) augmentations were applied.

Pseudo-labelling. There were 28,564 unlabelled patches of 128×128 pixels generated for this study. The model trained using the augmented data was used to predict the labels (pseudo-labels) of all pixels of each unlabelled patch (Figure 3.7). The labelled dataset was then expanded based on the confidence of these pseudo-labels. To improve the overall performance of the final model, we used a threshold to select only those pseudo-labels that the model was very confident about. Since the model output prediction probabilities from 0 to 1 of *S. anglica* presence, all predictions with $Pr(y > 0.95) + Pr(y < 0.05) > 0.90$ were selected and added to the new training set, and the remaining were designated as unlabelled data. The model was then retrained with the new training set consisting of the labelled data and the selected subset of pseudo-labelled data. This new model was then used to predict the remaining unlabelled dataset. This process was repeated until 95% of the unlabelled samples passed the threshold and was used to retrain the model.

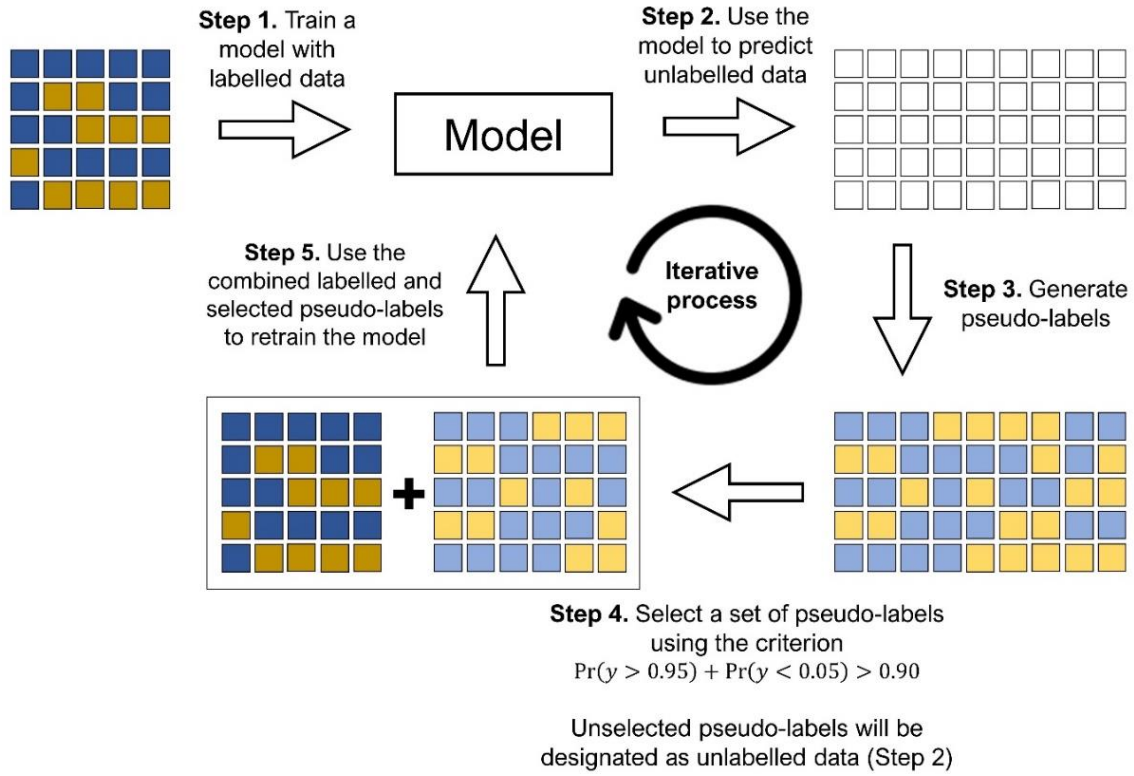


Figure 3.7. The pseudo-labelling process.

For the experiments involving pseudo-labels, we slightly modified the *validation* set by splitting off samples ($n = 100$) from the *test* set and adding them to the *validation* set, resulting in a new set of validation data (*validation_{aug_pseudo}* set). This was done because the domain gap (different sensor, site and seasonal variation) between the original *validation* set and the *test* set meant that optimising hyperparameters for the original *validation* set may result in a model that does not generalise well to domain variations. Including validation samples from the target domain allows for a model with better generalisation characteristics to be selected. Note that the remaining *test* set did not contain any of the samples that were split off for validation. The effects of augmentation and pseudo-labelling on the accuracy of the model will be discussed in Section 3.4.2.

3.3.6. Evaluation of the final model

The final model derived from the previous section was evaluated on the *test* set. To investigate the contribution of data augmentation and pseudo-labelling on the overall model performance on the *test* set, we ran the same experiment, but this time there was no data augmentation or pseudo-labelling implemented on the original *train* data.

3.4. Results

3.4.1. Model performance comparison between architectures and backbones using different patch sizes

Comparing all the models produced in Section 3.3.5, U-Net_Inception-v3 applied on the 128×128 -pixel image patches yielded the highest *mIOU* score of 0.832 (Table 3.4). This was followed by the LinkNet_Inception-v3 model applied on the 128×128 -pixel image patches, which had a slightly lower *mIOU* score of 0.830. In general, no pattern was observed in which a specific image patch size led to better model performance. This means that patch sizes of 128×128 and 256×256 pixels, where 1 pixel = 5 cm, can both be used to recognise and capture contextual information of *S. anglica* on the images. Instead, the model performance varied depending on the architecture and encoder backbone used.

Table 3.4. *Mean IOU (mIOU)* scores when using different encoder backbones in U-Net, LinkNet and FPN applied on two different image patch sizes. The text in bold indicates the model with the highest performance and its corresponding *mIOU* score.

Backbone	128 × 128-pixel image patches		
	U-Net	LinkNet	FPN
VGG-16	0.692	0.690	0.788
VGG-19	0.647	0.724	0.760
ResNet-50	0.712	0.797	0.742
ResNet-101	0.742	0.734	0.726
ResNet-152	0.675	0.565	0.732
ResNeXt-50	0.802	0.724	0.705
ResNeXt-101	0.772	0.784	0.697
DenseNet-121	0.567	0.763	0.764
Inception-v3	0.832	0.830	0.754
Inception-ResNet-v2	0.797	0.803	0.803
Backbone	256 × 256-pixel image patches		
	U-Net	LinkNet	FPN
VGG-16	0.785	0.538	0.786
VGG-19	0.688	0.612	0.747
ResNet-50	0.701	0.652	0.669
ResNet-101	0.807	0.736	0.800
ResNet-152	0.704	0.677	0.724
ResNeXt-50	0.774	0.766	0.802
ResNeXt-101	0.721	0.776	0.771
DenseNet-121	0.631	0.672	0.703
Inception-v3	0.799	0.795	0.783
Inception-ResNet-v2	0.812	0.728	0.795

Aside from *mIOU* scores, training speed varied among models, ranging from approximately 25 minutes to almost 3 hours (Figure 3.8). Overall, in the cases of both patch sizes, the U-Net and LinkNet models were trained at similar speeds, whereas the FPN models were the slowest to be trained. Regarding the backbones used, ResNeXt-101 was the slowest to train, followed by ResNeXt-50, ResNet-152 and Inception-ResNet-v2. The remaining backbones (VGG-16, VGG-19, ResNet-50, ResNet-101, DenseNet-121, Inception-v3) were trained at similar speeds, each approximately three times faster than the ResNeXt-101 training speed. Furthermore, models trained on 128 × 128-pixel image patches were consistently slower to train than those trained on the 256 × 256-pixel image

patches, possibly due to the difference in the total number of resulting image patches, as described in Section 3.3.5.

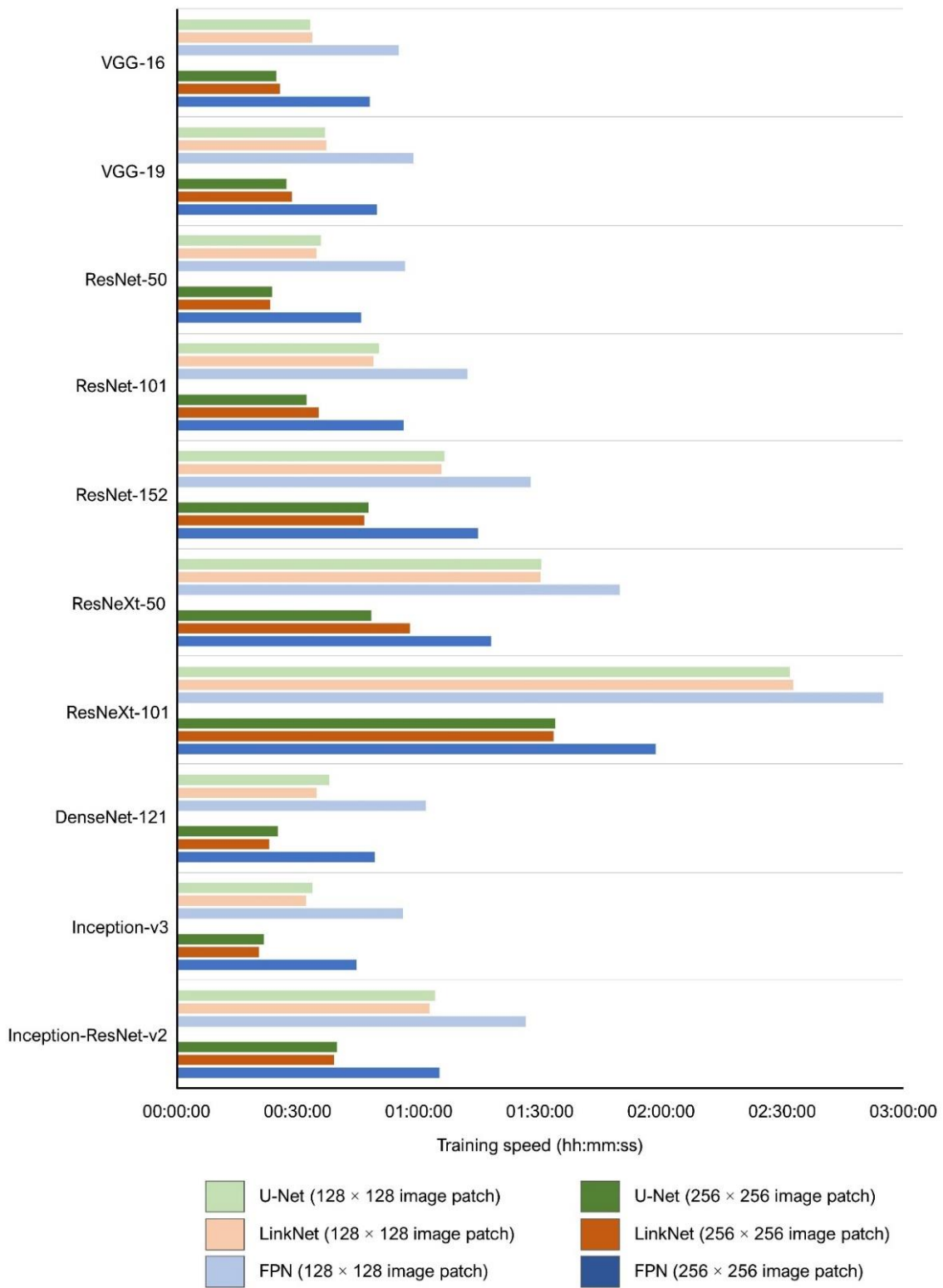


Figure 3.8. Training speed comparison of each model for different combinations of backbone, model architecture and patch size.

Applying the trained model to a new set of patch-based images was extremely fast and could produce segmented results in less than a second. Figure 3.9 shows the average inference speed, or the time it took for each model to make predictions on a single image patch, computed over 100 image patches. Comparing all the models, the two VGG backbones exhibited the fastest inference speeds. The U-Net and LinkNet models usually offered faster inference speeds than the FPN models, although the difference was marginal. Despite the 128×128 -pixel image patches comprising only a quarter as many pixels as the 256×256 -pixel image patches, they were less than four times as fast.

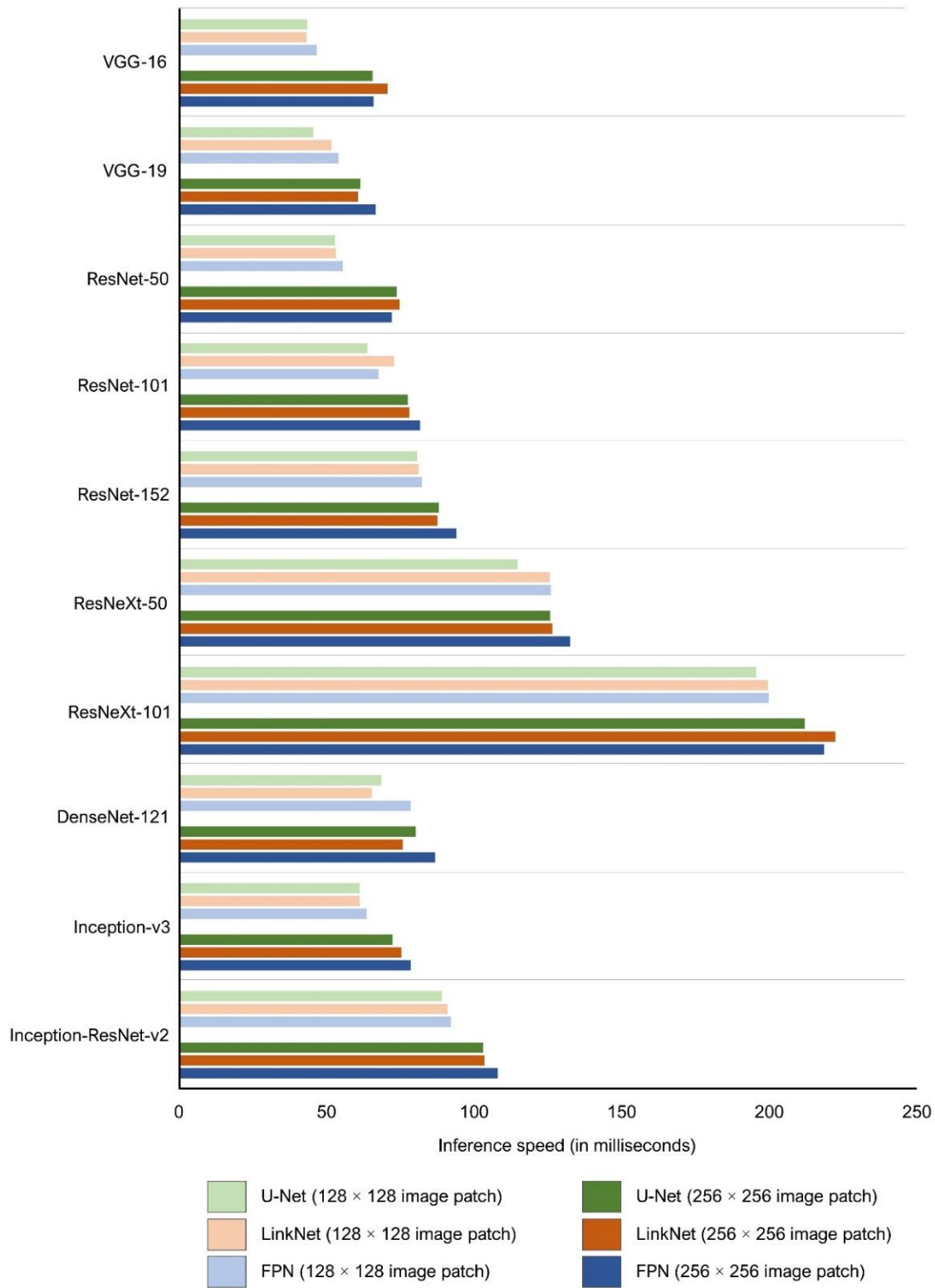


Figure 3.9. Inference speed (in milliseconds) on a single image patch for different combinations of backbone, model architecture and patch size.

3.4.2. Data augmentation and pseudo-labelling

Figure 3.10 shows the performance comparison between models with and without applying data augmentation and pseudo-labelling techniques. The baseline model (without data augmentation and pseudo-labels) achieved a 0.761 *mIOU* score on *validation_{aug_pseudo}* data. Applying data augmentation techniques to the initially labelled data improved the model performance, with an increase of 0.070 when using rotations only and a further increase of 0.010 when colour jitter augmentation was applied to the rotated images. Augmentation increased the number of labelled patches fourfold, from 4,092 to 16,368 (+12,276). Pseudo-labelling also contributed to the performance improvement of the model. After iteratively retraining the model with augmented labelled data and unlabelled data with pseudo-labels, its performance slightly improved from 0.841 to 0.851 (Figure 3.10). However, this modest improvement required the addition of a further 27,570 pseudo-labelled training patches. Figure 3.11 shows sample predictions on unlabelled data after the first and second iterations. It can be observed that the predictions became more accurate after the second iteration.

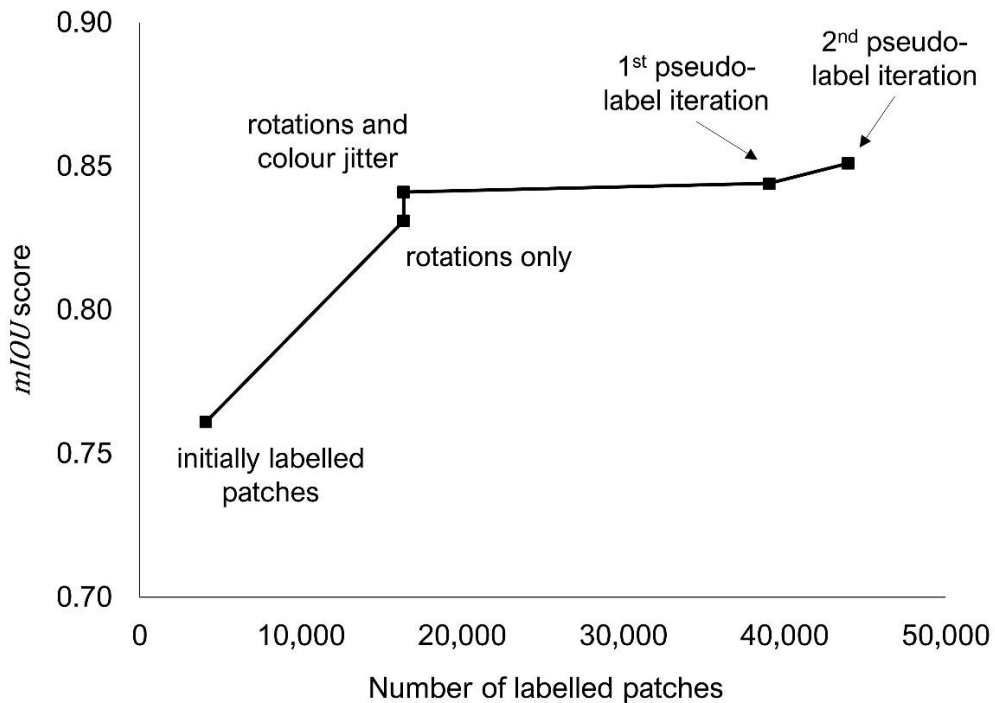


Figure 3.10. Number of labelled patches to train a model and its corresponding performance represented by *mIOU* score.

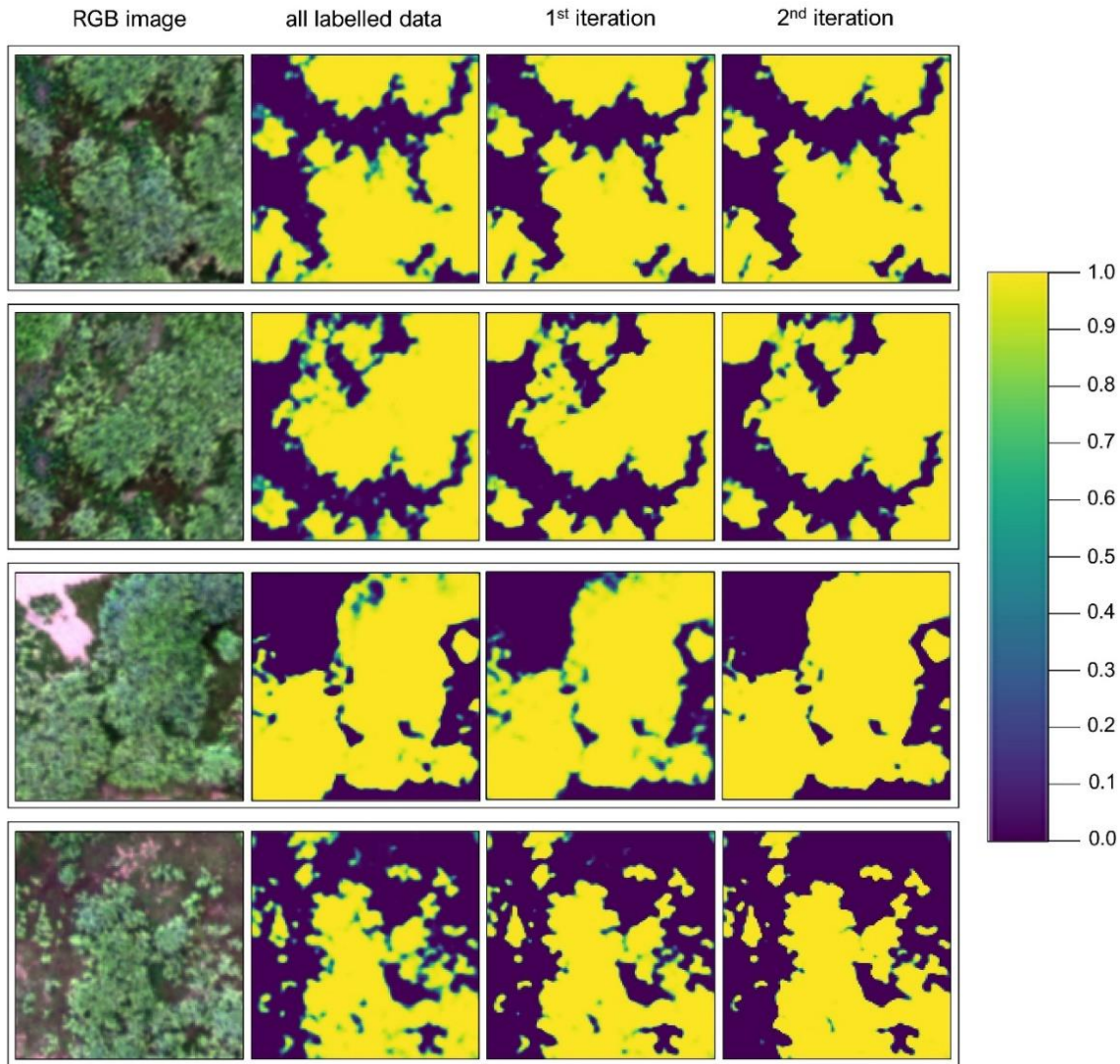


Figure 3.11. Sample RGB images (left) and the equivalent probability images as training iteration progresses. The colour scale indicates the probability that *Spartina anglica* is present at the location of each pixel (1.0 = present, 0.0 = absent).

3.4.3. Evaluation of the final model on test data

The *mIOU* score when the final model was applied to *test* data was 0.712 (Table 3.5). Results also showed a total *mIOU* score decrease of 0.158 in the model performance when we eliminated both the augmentation and pseudo-labelling techniques from the process (Table 3.5). Figure 3.12 shows sample *test* images and the comparison between their reference labels and the labels predicted by the model.

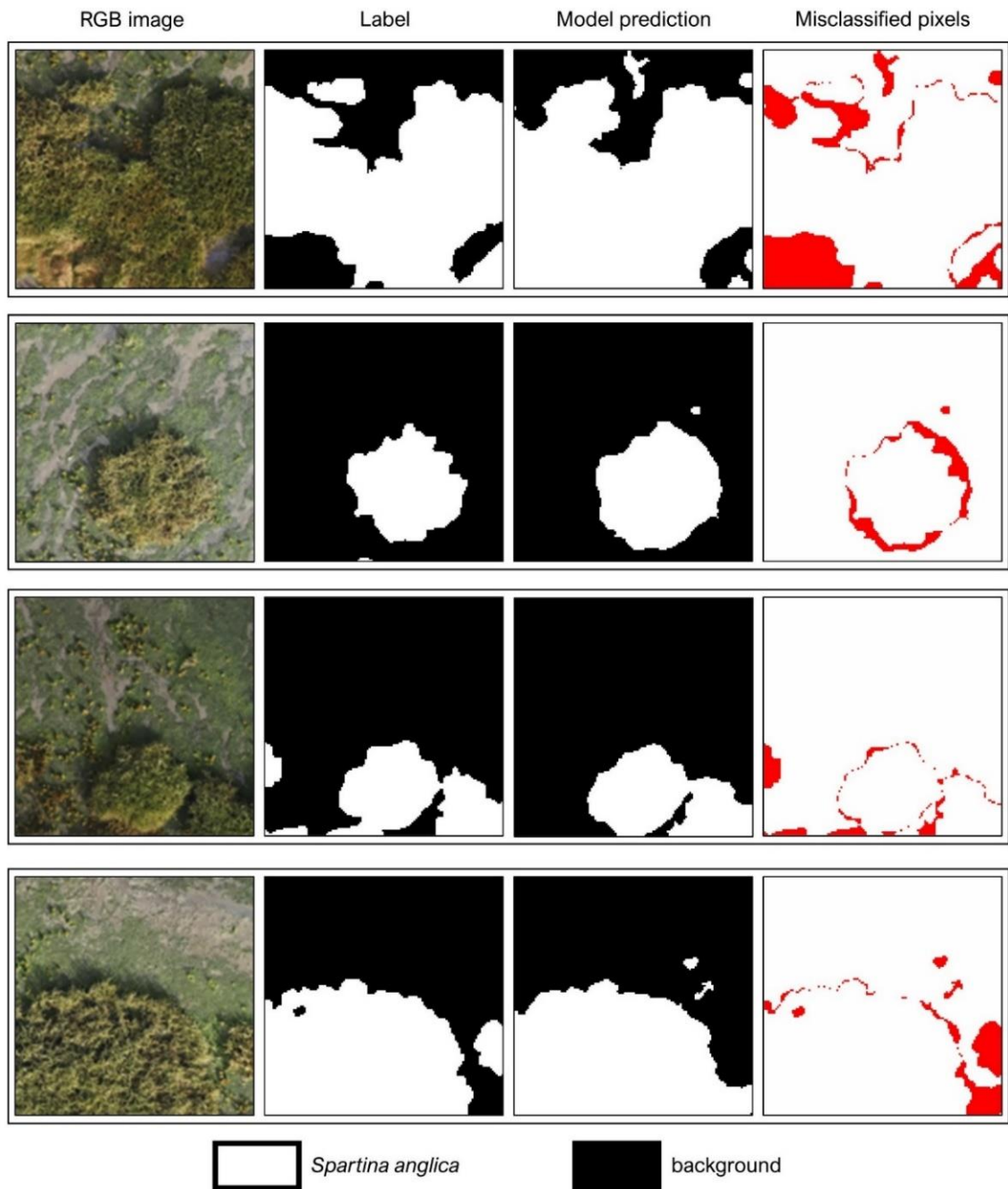


Figure 3.12. Sample RGB *test* images and their corresponding segmentation reference labels and model predictions, with an image indicating incorrectly classified pixels (red).

Table 3.5. Model performance when data augmentation and pseudo-labelling techniques were removed from the process.

Scenario	<i>test data</i>	
	<i>mIOU</i>	<i>mIOU</i> Δ
Final model on the <i>test</i> set	0.712	-
Model with data augmentation but without pseudo-labelled data	0.698	-0.014
Model without data augmentation and pseudo-labelled data	0.554	-0.158

3.5. Discussion

Our study demonstrates that deep learning-based segmentation can accurately map the distribution of *S. anglica* in UAV imagery. The results are in line with previous studies that have shown the potential of deep learning for high-resolution mapping of invasive species using UAV imagery (Gonçalves et al., 2022; James and Bradshaw, 2020; Qian et al., 2020).

The U-Net_Inception-v3 model was selected based on its performance (highest *mIOU* score of 0.832). It also offered the best balance between training speed (fastest to train with approximately 34 minutes for 100 epochs on a GPU) and accuracy. However, this model exhibited an average inference speed of 61.19 ms per input image—17.94 ms slower than the model with the fastest inference speed (LinkNet_VGG-16). The difference in inference speed between the two models is very small; this may not be a problem in most monitoring programmes. While LinkNet_VGG-16 had the fastest inference speed of 43.25 ms, its *mIOU* score (0.692) was 0.14 lower than U-Net_Inception-v3 (Table 3.4). Monitoring of an invasive species requires a model with good segmentation results so that the model can accurately determine the spatial extent and rate of expansion of that species.

The application of augmentation and pseudo-labelling techniques helped improve model robustness by increasing the diversity of the training data. The integration of these

techniques by James and Bradshaw (2020) enhanced vegetation mapping. These techniques are advantageous as manual delineation of irregular plant canopies on the imagery is time-consuming and tedious. According to our analysis, after applying data augmentation, there was a substantial increase in the *mIOU* score, however, when a large amount of unlabelled data was subsequently included, there was only a slight improvement in the model performance (Figure 3.10). This result could be due to a large amount of variance added to the training data by applying data augmentation techniques. For remote sensing, however, acquiring RGB images using low-cost UAVs to generate unlabelled data is relatively simple. Hence, pseudo-labelling can have an important practical significance in improving the segmentation performance as domain-specific labelled samples may be limited.

In general, the time of the year when the image was acquired can affect the model performance. Plant colouration and morphology will change with different seasons of the year. *S. anglica* changes from mid green during the growing season to pale brown during the winter as it dies back. Variations in sunlight levels at the time of UAV surveys will also affect the brightness of the captured images. Aiming to assess the transferability of the model, we acquired a new set of images and performed the prediction under different conditions, i.e., different cameras, different sites and different times of the year. Hence, variations in illumination and differences in vegetation appearance were observed when comparing the images used in model training and testing. Despite this different scenario, the final model still achieved good results with an *mIOU* score of 0.712.

Inaccuracies in the segmentation could be due to the location of the training and test datasets and potential errors in the delineation of *S. anglica* boundaries. The *train* images were acquired from a site where areas dominated by *S. anglica* were intermixed with *Salicornia* beds. The *test* images were acquired from a site where areas dominated by *S. anglica* adjoined Atlantic salt meadows dominated by common saltmarsh-grass (*Puccinellia maritima*) and sea-purslane (*Atriplex portulacoides*). It is likely that some *S. anglica* patches were not correctly segmented due to variations in species composition surrounding *S. anglica* between the training and test sites. Even though the features of *S. anglica* on UAV imagery are quite distinct, performing manual delineation by image interpretation can be subjective and introduce positional errors. This can happen along the edges of plant canopies where image pixels are spectrally mixed with other plants or

other environmental background noise, such as mud. In this study, the chance of incorrect segmentation increased along the edges of *S. anglica* (see Figures 3.11 and 3.12). These inaccuracies might affect the training process, thus lowering the predictive accuracy.

A requirement of the presented methodology is a computer that supports the workload associated with deep learning. Computational resources and dataset size both impact the time it takes to train a model. The increasing amount of available remote sensing data means that the demand for computing resources to train models in a reasonable time is also increasing (Chi et al., 2016). Therefore, there is a need for the use of GPUs, which can process multiple computations simultaneously, as these can significantly speed up deep learning operations. As these models are retrained with more datasets and applied to larger spatial extents, there may be a need for either cloud-based services (e.g., Google Colaboratory (Carneiro et al., 2018)) or national centres for high-performance computing (e.g., European High-Performance Computing), which can provide access to computing resources for deep learning projects to accelerate training and inference speeds with minimal computing resources from the users. Moreover, recent deep learning architectures proposed for the segmentation of large, high-resolution images could be explored. An example is MFVNet (Li et al., 2023) which utilises multiple fields of view to efficiently extract information on images of large spatial extents. In this way, the mapping of *S. anglica* or other target species could be extended to a regional or national scale.

In Ireland, *S. anglica* is found in many coastal counties and can grow on mudflats and saltmarshes. Saltmarshes occur in a range of geomorphological contexts: estuaries, bays, fringes, sand flats and lagoons (Curtis and Sheehy Skeffington, 1998). Considering this variation in situations where *S. anglica* can establish, there are opportunities to explore updating our model to generalise well with other locations when more data becomes available.

The techniques used here provided robust results; however, the model accuracy could be improved further. Future studies can consider using more extensive combinations of augmentation techniques and more domain adaptation techniques, such as CycleGAN (Zhu et al., 2017) and Contrastive Unpaired Translation (Park et al., 2020), which can potentially improve the generalisation ability of a deep learning model. Cross-domain

segmentation techniques that specifically deal with variations in remote sensing data (e.g., when acquired by different sensors or at different geographic locations) could also be explored (Li et al., 2021). Furthermore, future studies can consider using weighted cross entropy with weights reducing for image pixels nearer to the edges of plant canopies (James and Bradshaw, 2019). In this way, potential errors caused by manual delineation of the indistinct nature of plant canopy edges can be minimised.

3.6. Conclusion

Developing baseline data on the current extent of invasive species, as well as timely and effective monitoring of their expansion, is crucial for biodiversity conservation and the sustainable management of habitats. Advances in UAV remote sensing and deep learning have provided a huge potential for accurately mapping these invasive species. The results of this study show that the choice of model network structure and the use of techniques which can enhance the size and quality of training data are important decisions when creating robust deep learning models for mapping invasive species.

This study aimed to improve the method for mapping the spatial distribution of *S. anglica* invading saltmarsh habitats by using deep learning-based semantic segmentation applied to high-resolution UAV imagery. The results indicated that the U-Net architecture with Inception-v3 as the encoder backbone trained on 128×128 -pixel image patches was the best model in terms of model performance. Applying data augmentation to the initially labelled data increased the *mIOU* score by 0.08, with a further small improvement of 0.01 after adding pseudo-labels. These techniques improved model robustness. The final model evaluated on a separate test dataset achieved an *mIOU* score of 0.712, indicating good generalisation ability.

For practical purposes, the segmentation model developed in the study can be utilised for spatiotemporal mapping of *S. anglica* distribution. These high-resolution maps are useful to understand the pattern and rate of *S. anglica* invasion and potentially to evaluate the effectiveness of control efforts. The proposed methodology is transferable and can be adapted for mapping other invasive species.

3.7. References

- Ahmed, L., Iqbal, M. M., Aldabbas, H., Khalid, S., Saleem, Y., & Saeed, S. (2020). Images data practices for semantic segmentation of breast cancer using deep neural network. *Journal of Ambient Intelligence and Humanized Computing*.
<https://doi.org/10.1007/s12652-020-01680-1>
- Balado, J., Olabarria, C., Martínez-Sánchez, J., Rodríguez-Pérez, J. R., & Pedro, A. (2021). Semantic segmentation of major macroalgae in coastal environments using high-resolution ground imagery and deep learning. *International Journal of Remote Sensing*, 42(5), 1785–1800.
<https://doi.org/10.1080/01431161.2020.1842543>
- Belcore, E., Pittarello, M., Lingua, A. M., & Lonati, M. (2021). Mapping riparian habitats of Natura 2000 network (91E0*, 3240) at individual tree level using UAV multi-temporal and multi-spectral data. *Remote Sensing*, 13, 1756.
<https://doi.org/10.3390/rs13091756>
- Belgiu, M., & Drăgut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Beyer, F., Jurasinski, G., Couwenberg, J., & Grenzdörffer, G. (2019). Multisensor data to derive peatland vegetation communities using a fixed wing unmanned aerial vehicle. *International Journal of Remote Sensing*, 40(24), 9103–9125.
<https://doi.org/10.1080/01431161.2019.1580825>
- Brophy, J. T., Perrin, P. M., Penk, M. R., Devaney, F. M., & Leyden, K. J. (2019). *Saltmarsh monitoring project 2017- 2018 (Irish Wildlife Manuals No. 108)*. National Parks and Wildlife Service, Department of Culture, Heritage and the Gaeltacht, Ireland. <http://hdl.handle.net/2262/89539>
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Alumentations: Fast and flexible image augmentations. *Information*, 11(2), 125. <https://doi.org/10.3390/info11020125>

- Carneiro, T., Da Nobrega, R. V. M., Nepomuceno, T., Bian, G. Bin, De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance analysis of Google Colaboratory as a tool for accelerating deep learning applications. *IEEE Access*, 6, 61677–61685. <https://doi.org/10.1109/ACCESS.2018.2874767>
- Chaurasia, A., & Culurciello, E. (2018). LinkNet: Exploiting encoder representations for efficient semantic segmentation. *Proceedings of 2017 IEEE Visual Communications and Image Processing*, 1–4. <https://doi.org/10.1109/VCIP.2017.8305148>
- Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207–2219. <https://doi.org/10.1109/JPROC.2016.2598228>
- Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5), 545–563. <https://doi.org/10.1111/1754-9485.13261>
- Curtis, T. G. F., & Sheehy Skeffington, M. J. (1998). The salt marshes of Ireland: An inventory and account of their geographical variation. *Biology and Environment: Proceedings of the Royal Irish Academy*, 98B(2), 87–104. <http://www.jstor.org/stable/20500023>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- European Commission. (2014). Regulation (EU) No 1143/2014 of the European Parliament and the Council of 22 October 2014 on the prevention and management of the introduction and spread of invasive alien species. *Official Journal of the European Union*, Vol. 317, 35–55. <http://data.europa.eu/eli/reg/2014/1143/oj>
- Gonçalves, C., Santana, P., Brandão, T., & Guedes, M. (2022). Automatic detection of *Acacia longifolia* invasive species based on UAV-acquired aerial imagery. *Information Processing in Agriculture*, 9(2), 276–287.

<https://doi.org/10.1016/j.inpa.2021.04.007>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

<http://www.deeplearningbook.org/>

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2016.90>

Huang, C., & Asner, G. P. (2009). Applications of remote sensing to alien invasive plant studies. *Sensors*, 9(6), 4869–4889. <https://doi.org/10.3390/s90604869>

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>

IUCN. (2000). *IUCN guidelines for the prevention of biodiversity loss caused by alien invasive species*. <https://portals.iucn.org/library/efiles/documents/Rep-2000-052.pdf>

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de La Société Vaudoise Des Sciences Naturelles*, 44, 223–270. <https://doi.org/10.5169/seals-268384>

James, K., & Bradshaw, K. (2020). Detecting plant species in the field with deep learning and drone technology. *Methods in Ecology and Evolution*, 11(11), 1509–1519. <https://doi.org/10.1111/2041-210X.13473>

James, K., & Bradshaw, K. (2019). Segmenting objects with indistinct edges, with application to aerial imagery of vegetation. *Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019*. <https://doi.org/10.1145/3351108.3351124>

Kattenborn, T., Eichel, J., & Fassnacht, F. E. (2019). Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific Reports*, 9, 17656.

<https://doi.org/10.1038/s41598-019-53797-9>

- Kelly, J., Tosh, D., Dale, K., & Jackson, A. (2013). *The economic cost of invasive and non-native species in Ireland and Northern Ireland*. A report prepared for the Northern Ireland Environment Agency and National Parks and Wildlife Service as part of Invasive Species Ireland.
- Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015*. <https://doi.org/10.48550/arXiv.1412.6980>
- Lee, D. H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. *ICML 2013 Workshop: Challenges in Representation Learning Atlanta, Georgia, USA*.
- Li, J., Gao, S., & Wang, Y. (2010). Invading cord grass vegetation changes analyzed from Landsat-TM imageries: A case study from the Wanggang area, Jiangsu coast, eastern China. *Acta Oceanologica Sinica*, 29(3), 26–37.
<https://doi.org/10.1007/s13131-010-0034-y>
- Li, Y., Chen, W., Huang, X., Gao, Z., Li, S., He, T., & Zhang, Y. (2023). MFVNet: a deep adaptive fusion network with multiple field-of-views for remote sensing image semantic segmentation. *Science China Information Sciences*, 66(4), 140305.
<https://doi.org/10.1007/s11432-022-3599-y>
- Li, Y., Shi, T., Zhang, Y., Chen, W., Wang, Z., & Li, H. (2021). Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, 20–33.
<https://doi.org/10.1016/j.isprsjprs.2021.02.009>
- Li, Y., Zhang, H., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6), e1264. <https://doi.org/10.1002/widm.1264>
- Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. *Proceedings of the 2017 IEEE*

- Conference on Computer Vision and Pattern Recognition*, 936–944.
<https://doi.org/10.1109/CVPR.2017.106>
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2018). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 2980-2988. <https://doi.org/10.1109/TPAMI.2018.2858826>
- Lowe, S., Browne, M., Boudjelas, S., & Poorter, M. De. (2000). *100 of the world's worst invasive alien species: a selection from the global invasive species database*. The Invasive Species Specialist Group (ISSG), Auckland.
- Mao, J., Yin, X., Zhang, G., Chen, B., Chang, Y., Chen, W., Yu, J., & Wang, Y. (2022). Pseudo-labeling generative adversarial networks for medical image classification. *Computers in Biology and Medicine*, 147, 105729.
<https://doi.org/10.1016/j.compbimed.2022.105729>
- Manfreda, S., McCabe, M. F., Miller, P. E., Lucas, R., Madrigal, V. P., Mallinis, G., Dor, E. Ben, Helman, D., Estes, L., Ciruolo, G., Müllerová, J., Tauro, F., de Lima, M. I., de Lima, J. L. M. P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., ... Toth, B. (2018). On the use of unmanned aerial systems for environmental monitoring. *Remote Sensing*, 10(4), 641. <https://doi.org/10.3390/rs10040641>
- Mathew, S., Pellicer, X. M., Caloca, S., Monteys, X., Zarroca, M., & Jiménez-Martín, D. (2019). Bull Island: Characterisation and development of a modern barrier island triggered by human activity in Dublin Bay, Ireland. *Irish Geography*, 52(1), 75–100. <https://irishgeography.ie/index.php/irishgeography/article/view/1378>
- McCorry, M., & Ryle, T. (2009). *Saltmarsh Monitoring Project 2007-2008*. National Parks and Wildlife Service, Department of Environment, Heritage & Local Government, Ireland.
https://www.npws.ie/sites/default/files/publications/pdf/McCorry_%26_Ryle_2009_Saltmarsh_survey_V1.pdf
- MicaSense. (2019). *Altum-PT specifications*, <https://ageagle.com/drone-sensors/altum-pt-camera/>. Accessed: 29 June 2022.
- Mollot, G., Pantel, J. H., & Romanuk, T. N. (2017). The effects of invasive species on

- the decline in species richness: A global meta-analysis. In D. A. Bohan, A. J. Dumbrell, & F. Massol (Eds.), *Advances in Ecological Research* (Vol. 56, pp. 61–83). Elsevier Ltd. <https://doi.org/10.1016/bs.aecr.2016.10.002>
- National Parks and Wildlife Service. (2013). *Site Synopsis: North Dublin Bay SAC*. Department of Arts, Heritage and the Gaeltacht, Ireland. <https://www.npws.ie/sites/default/files/protected-sites/synopsis/SY000206.pdf>
- National Parks and Wildlife Service. (2019). *The status of EU protected habitats and species in Ireland. Volume 1: Summary overview*. Department of Culture, Heritage and the Gaeltacht, Ireland.
- Nehring, S., & Adersen, H. (2006). NOBANIS – Invasive alien species fact sheet *Spartina anglica*. In *Online Database of the European Network on Invasive Alien Species*. Accessed 29 June 2022. www.nobanis.org
- Nentwig, W., Bacher, S., Kumschick, S., Pyšek, P., & Vilà, M. (2018). More than “100 worst” alien species in Europe. *Biological Invasions*, 20(6), 1611–1621. <https://doi.org/10.1007/s10530-017-1651-6>
- Osco, L. P., Nogueira, K., Marques Ramos, A. P., Fanta Pinheiro, M. M., Furuya, D. E. G., Gonçalves, W. N., de Castro Jorge, L. A., Marcato Junior, J., & dos Santos, J. A. (2021). Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. *Precision Agriculture*, 22(4), 1171–1188. <https://doi.org/10.1007/s11119-020-09777-5>
- Park, T., Efros, A. A., Zhang, R., & Zhu, J. Y. (2020). Contrastive learning for unpaired image-to-image translation. *European Conference on Computer Vision*, 319–345. https://doi.org/10.1007/978-3-030-58545-7_19
- Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit (2017)*. <https://doi.org/10.48550/arXiv.1712.04621>
- Proença, B., Frappart, F., Lubac, B., Marieu, V., Ygorra, B., Bombrun, L., Michalet, R., & Sottolichio, A. (2019). Potential of high-resolution Pléiades imagery to monitor salt marsh evolution after *Spartina* invasion. *Remote Sensing*, 11(8), 968.

<https://doi.org/10.3390/rs11080968>

- Qian, W., Huang, Y., Liu, Q., Fan, W., Sun, Z., Dong, H., Wan, F., & Qiao, X. (2020). UAV and a deep convolutional neural network for monitoring invasive alien plants in the wild. *Computers and Electronics in Agriculture*, 174, 105519. <https://doi.org/10.1016/j.compag.2020.105519>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In A. Navab, N., Hornegger, J., Wells, W., Frangi (Ed.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
- Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., Pagad, S., Pyšek, P., Winter, M., Arianoutsou, M., Bacher, S., Blasius, B., Brundu, G., Capinha, C., Celesti-Gradow, L., Dawson, W., Dullinger, S., Fuentes, N., Jäger, H., ... Essl, F. (2017). No saturation in the accumulation of alien species worldwide. *Nature Communications*, 8, 14435. <https://doi.org/10.1038/ncomms14435>
- Sharma, S., Ball, J. E., Tang, B., Carruth, D. W., Doude, M., & Islam, M. A. (2019). Semantic segmentation with transfer learning for off-road autonomous driving. *Sensors*, 19(11), 2577. <https://doi.org/10.3390/s19112577>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.1409.1556>
- Stokes, K., O'Neill, K., & McDonald, R. A. (2006). *Invasive species in Ireland*. Report to Environment & Heritage Service and National Parks & Wildlife Service by Quercus, Queens University. Environment & Heritage Service, Belfast and National Parks & Wildlife Service, Dublin.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-

- ResNet and the impact of residual connections on learning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, 4278–4284.
<https://doi.org/10.1609/aaai.v31i1.11231>
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826.
<https://doi.org/10.1109/CVPR.2016.308>
- Van Beijma, S., Comber, A., & Lamb, A. (2014). Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sensing of Environment*, 149, 118–129.
<https://doi.org/10.1016/j.rse.2014.04.010>
- Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 5987–5995.
<https://doi.org/10.1109/CVPR.2017.634>
- Yakubovskiy, P. (2019). *Segmentation models*. GitHub.
https://github.com/qubvel/segmentation_models
- Yu, H., Yang, Z., Tan, L., Wang, Y., Sun, W., Sun, M., & Tang, Y. (2018). Methods and datasets on semantic segmentation: A review. *Neurocomputing*, 304, 82–103.
<https://doi.org/10.1016/j.neucom.2018.03.037>
- Zheng, Q., Yang, M., Tian, X., Jiang, N., & Wang, D. (2020). A full stage data augmentation method in deep convolutional neural network for natural image classification. *Discrete Dynamics in Nature and Society*, 4706576.
<https://doi.org/10.1155/2020/4706576>
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2242–2251.
<https://doi.org/10.1109/ICCV.2017.244>
- Zuo, P., Zhao, S., Liu, C., Wang, C., & Liang, Y. (2012). Distribution of *Spartina* spp.

along China's coast. *Ecological Engineering*, 40, 160–166.
<https://doi.org/10.1016/j.ecoleng.2011.12.014>

4. Mapping of Irish upland habitats using high-resolution satellite imagery and machine learning



Upland habitats in Slieve Mish, Co. Kerry (Photo: Charmaine Cruz)

This chapter is currently under review in *Environmental Monitoring and Assessment*:

Cruz, C., Perrin, P. M., Martin, J. R., O’Connell, J., McGuinness, K., & Connolly, J. Mapping of temperate upland habitats using high-resolution satellite imagery and machine learning. (Under review)

Description of authors' contributions:

C.C. designed the research and performed the analysis. C.C., J.M. and P.M.P. performed the data acquisition. C.C. prepared the original draft manuscript. J.C., K.M., P.M.P., J.M. and J.O.C. provided a critical review of the manuscript. C.C. led the submission of the manuscript to the journal.

4.1. Abstract

Upland habitats provide vital ecological services, yet they are highly threatened by natural and anthropogenic stressors. Monitoring these vulnerable habitats is fundamental for conservation and involves determining information about their spatial locations and conditions. Remote sensing has evolved as a promising tool to map the distribution of upland habitats in space and time. However, the resolutions of most freely available satellite images (e.g., 10-m resolution for Sentinel-2) may not be sufficient for mapping relatively small features, especially in the heterogeneous landscape—in terms of habitat composition—of uplands. Moreover, the use of traditional remote sensing methods, imposing discrete boundaries between habitats, may not accurately represent upland habitats as they often occur in mosaics and merge with each other. In this context, we used high-resolution (2 m) Pleiades satellite imagery and Random Forest (RF) machine learning to map habitats at two Irish upland sites. Specifically, we investigated the impact of varying spatial resolutions on classification accuracy and proposed a complementary approach to traditional methods for mapping complex upland habitats. Results showed that the accuracy generally improved with finer spatial resolution data, with the highest accuracy values (80.34% and 79.64%) achieved for both sites using the 2-m resolution datasets. The probability maps derived from the RF-based fuzzy classification technique can represent complex mosaics and gradual transitions occurring in upland habitats. The presented approach can potentially enhance our understanding of the spatiotemporal dynamics of habitats over large areas.

4.2. Introduction

Uplands comprise a range of extensive, mostly semi-natural habitats, including blanket bogs, heaths, fens, grasslands and those associated with exposed rocks and scree (Perrin et al., 2009). These habitats are protected in the European Union (EU) under the Habitats Directive (HD) (Commission of the European Communities, 1992). They provide important services, such as carbon sequestration and storage, biodiversity support, flood mitigation and water quality regulation (Bonn et al., 2008). However, they are also highly vulnerable to climate change and to increasing pressures and threats from anthropogenic stressors, mainly by land-use changes (Connolly, 2018; Crowle & McCormack, 2009; Perrin et al., 2017; Young et al., 2005). Examples include unsustainable agricultural practices, extensive afforestation, wind farm development and scrub encroachment. These stressors can lead to habitat fragmentation and widespread degradation, causing these habitats to lose their capacity to deliver such services (Connolly, 2018; Perrin et al., 2009). In the recent ‘State of Nature in the EU’ report, only 15% of the habitats in the EU were in favourable condition, with blanket bogs showing a deteriorating trend (European Environment Agency, 2020).

The alarming decline in habitat conditions within the EU has been recognised by the European Commission (EC) through the proposed Nature Restoration Law (European Commission, 2022). This law will require EU member states to develop and implement restoration measures for degraded ecosystems and habitats, such as drained bogs. Each member state must also monitor and report its progress in implementing the law (European Council, 2023). Comprehensive mapping is fundamental for executing this legal requirement as it can provide baseline data, such as the location and extent of habitats, and can be used to monitor and track their condition over time as well as to support restoration programmes. Common methods to map habitats include ground-based field surveys or interpretation from aerial imagery (Smith et al., 2011). However, the vast, remote and rugged terrain of upland habitats may limit the frequency of manual field surveys to conduct mapping of these habitats as there are issues with resource availability (i.e., limited time and cost) (Buchanan et al., 2005; Müller & Brandl, 2009). Using aerial imagery to delineate habitat boundaries—in the form of polygons—can also be challenging as several upland habitats typically occur together as mosaics, making it difficult to represent them as separate polygons (O’Connell et al., 2014; Perrin et al.,

2009). A complementary approach to address these mapping challenges should be considered, particularly with the increased need for accurate, timely and broad-scale spatial information on these protected habitats.

Remote sensing is a promising approach to mapping and monitoring habitats and vegetation communities within them (Corbane et al., 2015; Mùcher & Hazeu, 2021; Nagendra et al., 2013). Specifically, optical imagery acquired by satellites has been used for mapping upland habitats (Barrett et al., 2016; Connolly, 2018; Mehner et al., 2004; O’Connell et al., 2014). The spatial extent that a single satellite image can cover (e.g., $100 \times 100 \text{ km}^2$ for a Sentinel-2 image (Drusch et al., 2012)) is a significant advantage, particularly in upland habitats, as they can be extensive (JNCC, 2015). Moreover, satellite remote sensing technologies have significantly improved over the last decade (Cantrell et al., 2021; Gleyzes et al., 2012; Kim et al., 2022). They now provide imagery at higher spatial, spectral and temporal resolutions than previously available (i.e., spatial resolution of a few metres with less than a week revisit time), creating opportunities for detailed botanical mapping of upland habitats (Ingle et al., 2023). Previous studies have demonstrated that higher spatial resolution imagery (<5 m) can improve classification accuracy. However, these studies were mainly focused on general land cover classes wherein spectral characteristics can be quite distinct (Boyle et al., 2014; J. Fisher et al., 2017). The impact of image spatial resolution on classifying spectrally similar habitats, such as in uplands, has not been fully explored.

Most upland habitat mapping studies have used the conventional hard or crisp classification technique (Barrett et al., 2016; Mehner et al., 2004). This technique involves a classifier making a definitive decision on which class a pixel will be assigned to, typically to the class with the highest proportion in that pixel. Each pixel is assumed to be pure; hence, it is only associated with a single class (P. Fisher, 1997). However, information in a satellite image pixel often consists of multiple classes. As some habitats are often intermixed, the crisp classification technique may mean that information about other classes will be omitted (Lucas et al., 2007). Moreover, the technique can result in abrupt transitions between habitats in the maps, which poorly describe the continuous nature of upland habitats. Upland habitats rarely have sharp boundaries. They are mainly heterogeneous, occurring in complex mosaics (Gatis et al., 2022; O’Connell et al., 2014). A soft or fuzzy classification technique can be used to address the challenge of mapping

transitional boundaries in the uplands. In contrast to the crisp classification technique, fuzzy classification allows each image pixel to be described by the probabilities of occurrence of all classes being considered (Foody, 1996). In other words, instead of classifying individual pixels into mutually exclusive classes, fuzzy classification provides information on the probability that a certain class is present in every pixel. This technique can allow the representation of mixed and gradual transitions that often exist between habitats in complex environments, such as in uplands (Feilhauer et al., 2021). Moreover, the area computed from class probabilities can also be more accurate and closer to the actual area of that class (Sales et al., 2022).

Random Forest (RF) is a widely used machine-learning algorithm for classifying habitat and vegetation because of its robustness and excellent results (Amani et al., 2017; Cruz et al., 2023; van Iersel et al., 2018). RF (Breiman, 2001) is an ensemble learning made up of decision trees, wherein each tree makes its own class prediction. The final class prediction has the most votes over all these trees (i.e., majority voting). RF can also provide information on class probabilities (Malley et al., 2012). This attribute of RF, which has been used by previous satellite-based remote sensing studies (O’Connell et al., 2015; Raab et al., 2018; Sales et al., 2022), could be explored for classifying upland habitats.

This study aimed to utilise a combination of machine learning and high-resolution satellite imagery to map complex upland habitats. Specifically, our objectives are the following: (1) to determine the impact of spatial resolution on classification accuracy, (2) to compare crisp and fuzzy classification approaches using the RF algorithm for mapping upland habitats, and (3) to apply RF-based models to map and describe the spatial distributions of habitats at two studied Irish upland sites.

4.3. Materials and methods

4.3.1. Study site description

Two Irish upland sites, one in the Wicklow Mountains and one in the Slieve Mish Mountains, were selected for this study (Figure 4.1). Both sites are designated as Special

Areas of Conservations (SACs), part of the Natura 2000 network (National Parks and Wildlife Service, 2016, 2017), the world's largest network of protected sites covering the most valuable yet threatened habitats and species in the EU territory (Evans, 2012). The two sites are dominated by three upland habitats protected under the HD (those listed under Annex I): wet heath, dry heath and blanket bog (Figure 4.2) (Perrin, Roche, et al., 2014; Perrin et al., 2017). The study focused on classifying these three Annex I upland habitats. Other common habitats present at each site were also classified as some of them represent disturbance on uplands, such as the eroding blanket bog and the presence of dense bracken.

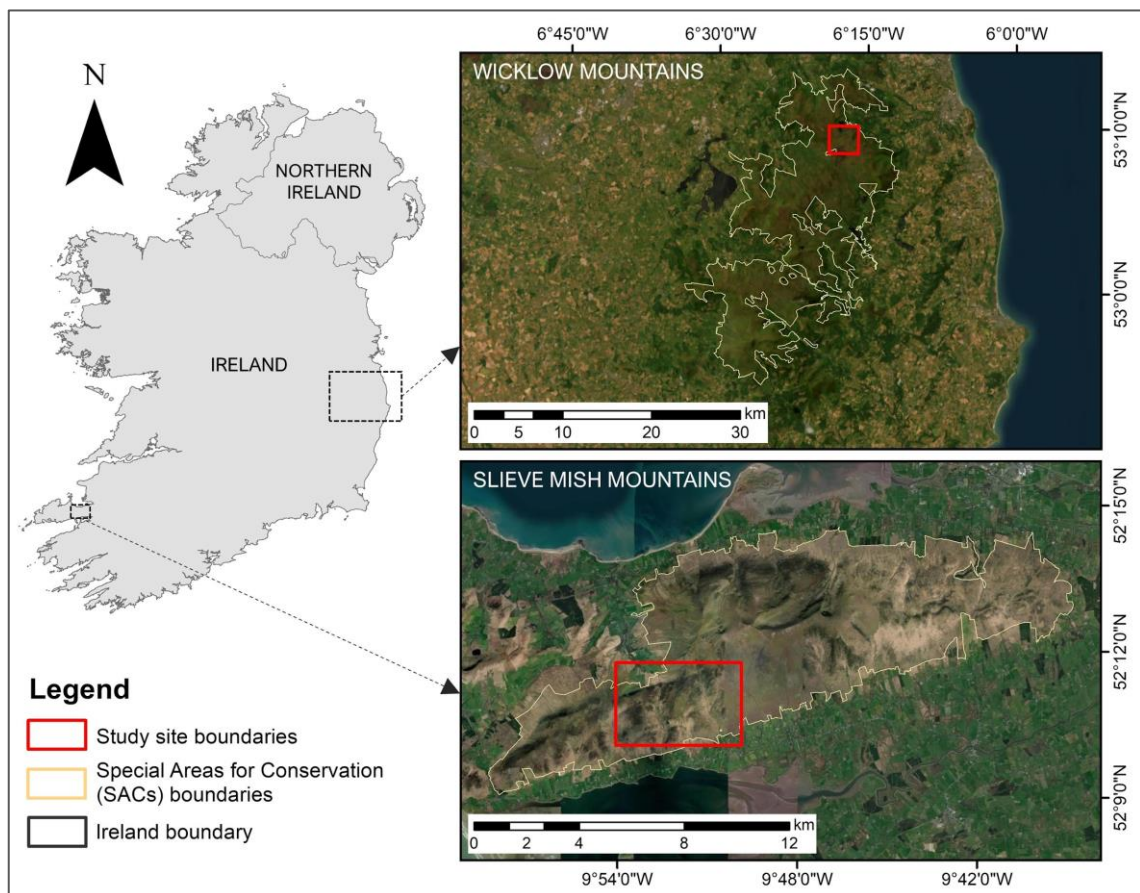


Figure 4.1. Inset maps showing the locations of the two study sites: Wicklow Mountains (top right) and Slieve Mish Mountains (bottom right), and the boundaries of the SAC covering each site.

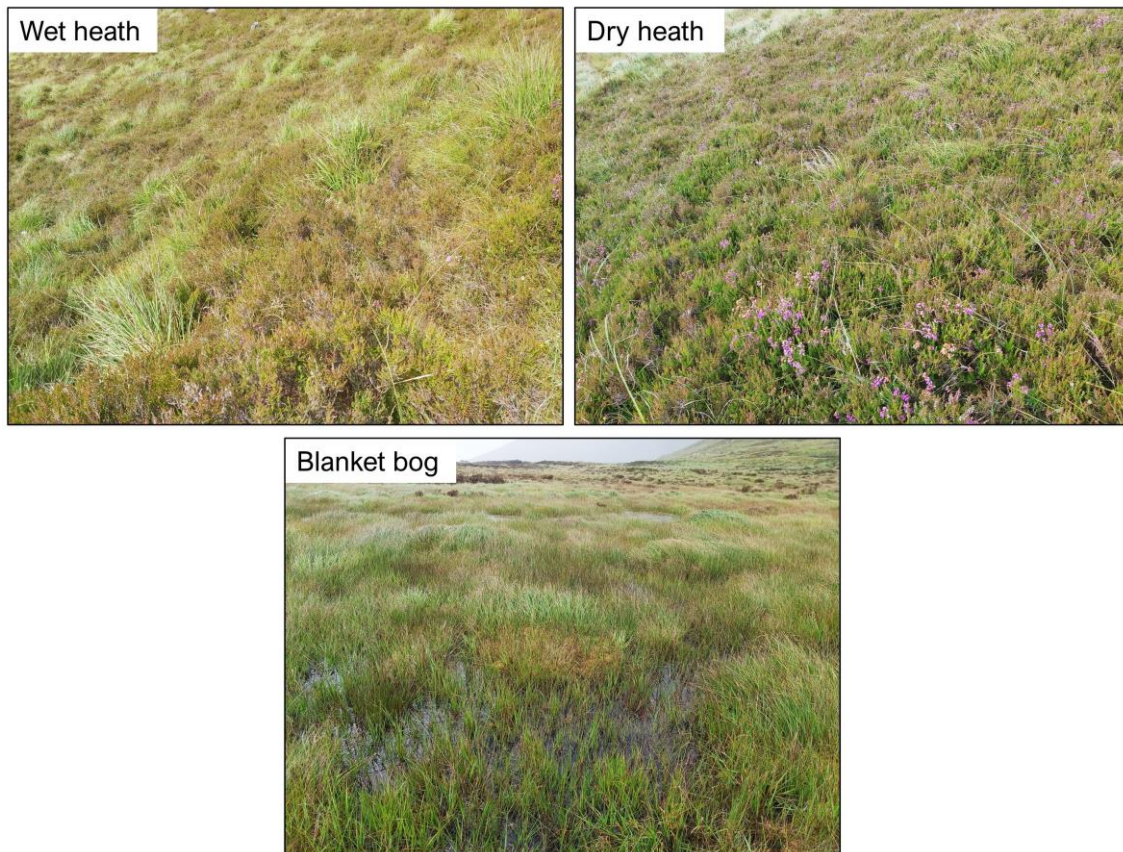


Figure 4.2. Sample field pictures of the three main upland habitats in the study sites: wet heath, dry heath and blanket bog.

The first site (WM) is in the Wicklow Mountains, located south of County Dublin. WM covers an area of ~1,020 hectares, which is just over 3% of the entire Wicklow Mountains SAC extent. This site is a good example of a mountain blanket bog in eastern Ireland, consisting of deep peat with depths ranging from one to six metres (Holden & Connolly, 2011).

The second studied site (SM) is in Slieve Mish Mountains SAC, located on the eastern side of the Dingle Peninsula in County Kerry. SM extends to an area of approximately 1,500 hectares, about 15% of the Slieve Mish Mountains SAC extent, which is 9,790 hectares (Perrin, Roche, et al., 2014).

4.3.2. General workflow

High-resolution (2 m) Pleiades satellite imagery and Digital Terrain Models (DTMs) were obtained for both sites. Several raster layers, or variables, were generated for each site using the obtained datasets to describe the characteristics of each habitat. These variables were resampled from the original 2 m resolution to 4 m, 6 m, 8 m and 10 m resolutions, generating five sets of variables. A reference dataset was used to train and evaluate an RF-based classification model for each set. The trained model was then applied to produce spatially continuous crisp and fuzzy-classified images within the study site. Figure 4.3 provides an overview of the workflow.

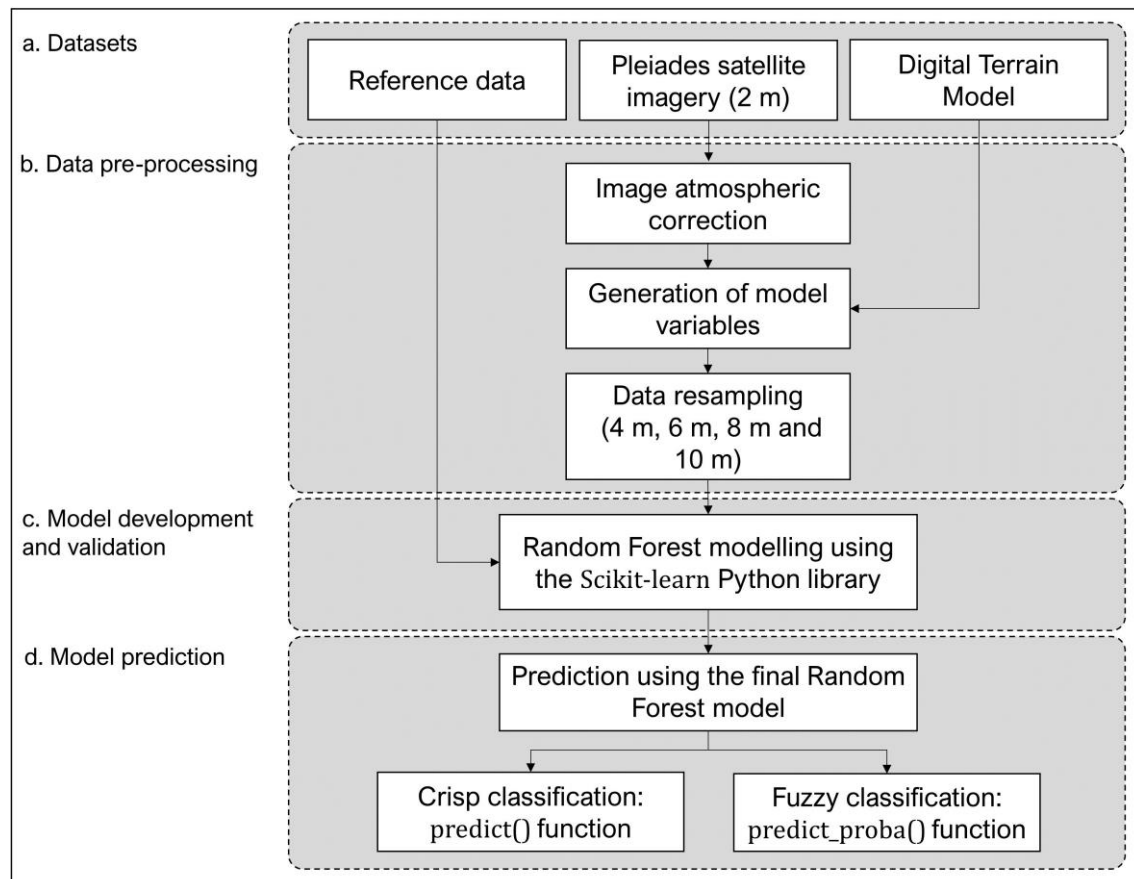


Figure 4.3. Diagram of the workflow illustrating the processing performed in the study.

a. Datasets

Pleiades satellite imagery and Digital Terrain Model. Cloud-free Pleiades images for each site were downloaded from the Sentinel Hub EO browser through the European Space Agency sponsorship program (sentinel-hub.com/Network-of-Resources/). The downloaded images for the SM and WM sites were captured on 15 April 2020 and 8 May 2022, respectively, which closely aligns with the collection period of field datasets. Both images have a spatial resolution of 2 m and consist of four spectral bands on the following wavelengths: blue (430-550 nm), green (490-610 nm), red (600-720 nm) and near-infrared (750-950 nm). The downloaded images were already orthorectified and projected on the Universal Transverse Mercator Zone 29 North coordinate system. DTMs for both sites were also obtained and reprojected to the same coordinate system as the Pleiades images.

To ensure the classification only focused on the upland area, the image extents for both sites were modified to align with the boundary used in the national survey of Irish upland habitats (Perrin, Barron, et al., 2014). This boundary was based on the definition of upland habitats by Perrin et al. (2009), i.e., any unenclosed land areas found at altitudes above 150 m and contiguous areas of related habitats below this value. The downloaded image for the SM site was clipped using the boundary. No clipping was necessary for the WM site image as its extent was already within the boundary.

Reference dataset. For this study, a reference dataset was used that consisted of data collected both in the field and through on-screen digitisation. In the field, the ecologists used Global Navigation Satellite System receivers (centimetre-level accuracy) to record the location of a habitat, represented by a point. To ensure the recording of the correct coordinates of a point within a particular habitat, each point was measured at the centre of an area of a relatively homogeneous habitat cover. Ecologists also recorded the corresponding Annex I habitat name and photographed it in the north-view direction. If a point was not an Annex I habitat, the Fossitt habitat classification scheme was used to label it. Fossitt (2000) provides a standard guide for habitat identification in Ireland. The field data points were then visually inspected on the screen by the remote sensing analyst to ensure they aligned with the correct habitats, as seen on the satellite image. A similar

process was applied to the additional points added in a GIS environment to supplement the field data. This process consisted of manual and subjective interpretation of the satellite image guided by the data from previous field surveys (Perrin et al., 2017; Perrin, Roche, et al., 2014). These previous survey data consist of polygons with records of approximate percentages of the constituent habitat types. Hence, the additional points were based on those polygons mostly comprising a single habitat (>98%). Table 4.1 provides the list of habitats considered for each study site. To account for the shaded areas in the image caused by tree canopy and steep slopes, we added a pseudo-habitat ‘shadow’ to the lists.

Table 4.1. List of habitats for each study site.

Study site	Habitats
WM	Wet heath
	Dry heath
	Blanket bog
	Grasslands/fens
	Bracken
	Eroding blanket bog
	Artificial surfaces
	Shadow
	Watercourses/bodies
SM	Wet heath
	Dry heath
	Blanket bog
	Grasslands/fens
	Artificial surfaces (including exposed rocks)
	Shadow
	Watercourses/bodies

b. Data pre-processing

Generation of atmospherically corrected surface reflectance data. Each pixel in a single band of the satellite imagery was represented by the top-of-atmosphere reflectance scaled by 10,000. Hence, we divided each pixel value by 10,000 to rescale them to floating-point values 0.0 to 1.0, consistent with the range of reflectance. To correct the image for the

influence of the atmosphere, we undertook the Dark Object Subtraction (DOS) method. DOS is a simple atmospheric correction method that has been used in many vegetation-related remote sensing applications (Gilmore et al., 2015; Hadjimitsis et al., 2010; Melnikova et al., 2018; Wicaksono, 2017). This method works on the assumption that the darkest pixels in the image (e.g., deep water, shadow) would have zero reflectance if not for the effects of the atmosphere on the image (Chavez, 1988). However, dark pixel targets are not always present in the image. Wicaksono & Hafizt (2018) showed that selecting non-ideal dark targets, such as turbid water pixels, resulted in many negative pixel values in the corrected images, which reduces the effectiveness of the DOS method. In this study, the pixel values within the 10th percentile in a single band were averaged, and the result was then subtracted from all pixel values from that band. This process was repeated for all the bands of the satellite imagery, resulting in an atmospherically corrected image.

Variable preparation. For each study site, variables were prepared to be used as input for the modelling (Table 4.2). Individual atmospherically corrected image bands were included as they can provide information in the visible and NIR regions of the image. Various vegetation indices were generated as they have been found to perform well in assessing vegetation cover and are extensively used for mapping habitats and vegetation communities (Bendig et al., 2015; Bhatnagar et al., 2020; Cruz et al., 2023; Suo et al., 2019). A further variable related to Principal Component Analysis (PCA) was also developed. PCA is a technique that reduces data dimensionality by creating new, uncorrelated variables known as the Principal Components (Jolliffe, 2002). Here, the first Principal Component band, having an eigenvalue of more than 75%, was extracted as it contains most of the information from the original data (see Table C.1, Appendix C). Additionally, four textural variables (mean, contrast, variance and correlation) were derived based on the grey-level co-occurrence matrix (Haralick et al., 1973) of each image band. Specifically, these variables were computed over a 5×5 neighbourhood. Previous remote sensing studies have shown that the inclusion of textural variables in the classification improves accuracy, as they can describe spatial patterns and variations of the features within a band (Barrett et al., 2016; Kattenborn, Lopatin, et al., 2019). Finally, elevation and slope were also generated as they describe the site topography, which can be determinants of habitat distribution (Cruz et al., 2023; Scholefield et al., 2019; Zuleta

et al., 2018). Overall, 33 variables were considered in the study. The description of each variable is shown in Table 4.2.

Table 4.2. List of variables used in the study and their descriptions.

	Variables	Description	Reference
Spectral	Blue band (B)	Spectral reflectance value of the blue band	
	Green band (G)	Spectral reflectance value of the green band	
	Red band (R)	Spectral reflectance value of the red band	
	Near-infrared band (NIR)	Spectral reflectance value of the near-infrared band	
	Simple Ratio (SR)	$SR = \frac{NIR}{R}$	(Jordan, 1969)
	Normalised Difference Vegetation Index (NDVI)	$NDVI = \frac{(NIR - R)}{(NIR + R)}$	(Rouse et al., 1973)
	Enhanced Vegetation Index (EVI)	$EVI = 2.5 * \frac{(NIR - R)}{[NIR + (6 * R) - (7.5 * B) + 1]}$	(Huete et al., 2002)
	Normalised Difference Water Index (NDWI)	$NDWI = \frac{(G - NIR)}{(G + NIR)}$	(McFeeters, 1996)
	Modified Soil Adjusted Vegetation Index 2 (MSAVI2)	$MSAVI2 = \frac{(2 * NIR) + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - R)}}{2}$	(Qi et al., 1994)
	Atmospherically Resistant Vegetation Index (ARVI)	$ARVI = \frac{(NIR - RB)}{(NIR + RB)}$	(Kaufman & Tanre., 1992)

	Green Normalised Difference Vegetation Index (GNDVI)	$GNDVI = \frac{(NIR - G)}{(NIR + G)}$	(Gitelson et al., 1996)
	Red, Green Blue Vegetation Index (RGBVI)	$RGBVI = \frac{G^2 - (B \times R)}{G^2 + (B \times R)}$	(Bendig et al., 2015)
	Modified Green Red Vegetation Index (MGRVI)	$MGRVI = \frac{G^2 - R^2}{G^2 + R^2}$	(Bendig et al., 2015)
	Optimised Soil Adjusted Vegetation Index (OSAVI)	$OSAVI = \frac{(NIR - R)}{(NIR + R + 0.16)}$	(Rondeaux et al., 1996)
	First Principal Component band (PC1)	First principal component that accounts for the greatest variance in the dataset	(Jolliffe, 2002)
Textural	Textural variables	Represent spatial variations surrounding the pixel (mean, contrast, variance and correlation)	(Haralick et al., 1973)
Topographic	Elevation	Represents the height of the terrain.	
	Slope	Represents the rate of change of elevation.	

Image resampling. To determine the impact of spatial resolution on the classification accuracy, all variables were resampled from the original 2-m spatial resolution to 4-, 6-, 8- and 10-m resolutions (Figure 4.4). This created five datasets for each site. Ten metres was selected as the lowest resampled resolution because freely accessible Sentinel-2 satellite images have this spatial resolution for the visible and NIR bands. This way, the accuracy difference between a high spatial resolution commercial satellite image and a free medium-resolution satellite image can be assessed.

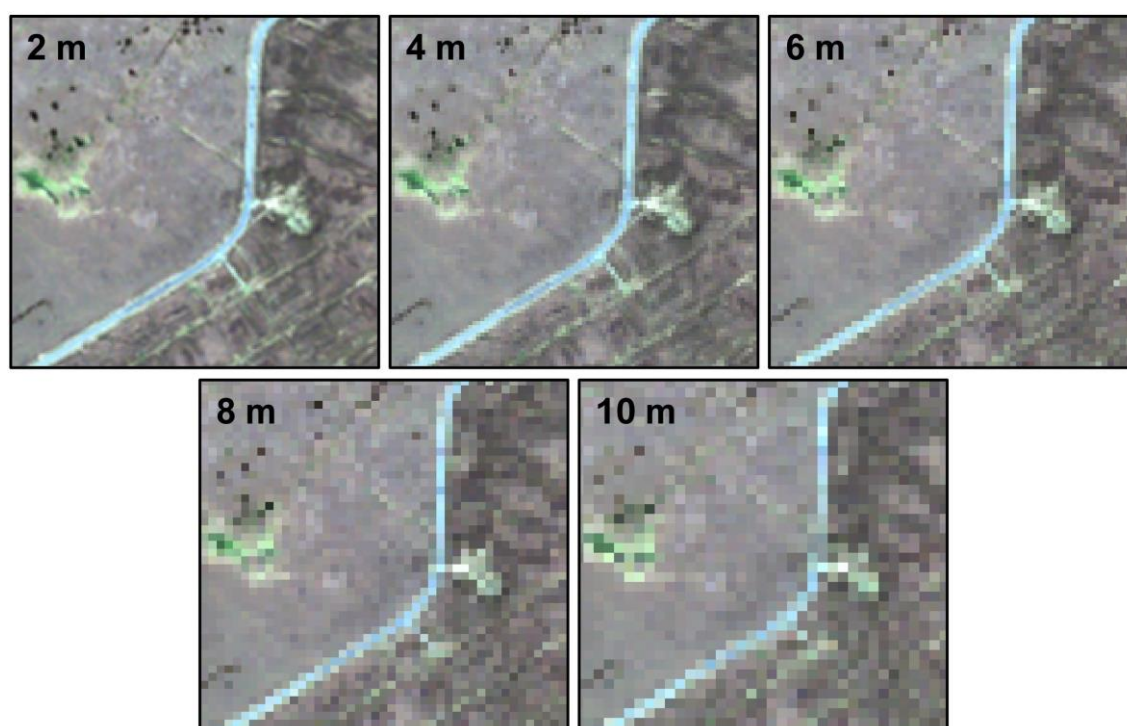


Figure 4.4. Sample images after resampling from the original 2-m resolution pixel size to 4-m, 6-m, 8-m and 10-m resolution pixel sizes

The resampling technique was based on averaging the pixel values within the new pixel size. This technique, however, could result in multiple field data points falling within a single pixel of the coarser image. Hence, we reviewed their locations on the 10-m resolution image to ensure that there was no more than one point within a pixel.

c. Model development and validation

The Scikit-learn Python library (Pedregosa et al., 2011) was used to implement the RF technique. This technique was applied to each dataset (i.e., variables of different spatial resolutions) to create a habitat classification model. Thus, five models were developed for each site (Table 4.3). Each model was evaluated using the k -fold cross-validation method. In this method, the points were randomly divided into k subsets or folds of approximately equal size. Each fold was used to evaluate the model, which was trained on the remaining $k - 1$ folds; this process was repeated until all folds were used in the validation (Berrar, 2019). Five folds were used in this study, which means that the model training and validation process was repeated five times. The average and standard error of the accuracy scores were then computed across these five folds. The final model to predict the pixel labels for the entire study site used all the points. Additionally, the relative importance of variables for each model was analysed based on Mean Decrease in Impurity (MDI) (Breiman et al., 1984).

Table 4.3. List of RF-based models developed for this study. Each model of WM and SM sites was trained on different spatial resolutions (2 m, 4 m, 6 m, 8 m and 10 m)

Spatial resolution	Models	
	WM site	SM site
2 m	WM_2m	SM_2m
4 m	WM_4m	SM_4m
6 m	WM_6m	SM_6m
8 m	WM_8m	SM_8m
10 m	WM_10m	SM_10m

d. Model prediction

Using the final model developed for each site, we generated two predictions: habitat predictions (crisp) and habitat probability predictions (fuzzy). The `predict()` function in Scikit-learn was used to return the predicted class for each pixel, i.e., the class with the highest probability across all the decision trees, generating a single image with row and column dimensions similar to the input satellite image. The `predict_proba()` function

was used to return a vector of class probabilities within the interval [0,1] for each pixel that were computed by averaging the class-predicted probabilities of all the decision trees. This latter function generates multiple images where each image represents the probabilities for a different habitat considered by the model (i.e., one image for each habitat).

To help visualise the spatial confidence of the classification associated with probabilities, we used entropy. In machine learning, entropy is defined as a measure of purity in a data set (Shannon, 1948). It is computed by:

$$E(p) = -\sum_{n=1}^c p_i \log_2 p_i,$$

where p is the probabilities obtained from the fuzzy classification. The value of entropy is inversely proportional to the confidence in the prediction. In other words, the entropy value is low for high-confidence predictions (peaky probability distribution), and the entropy value is high for low-confidence predictions (flat probability distribution). For example, if two pixels have probability predictions of (0.05, 0.90, 0.05) and (0.35, 0.40, 0.25), it can be considered that the first pixel was predicted with higher confidence, i.e., having a lower entropy value, than the second pixel.

4.4. Results

4.4.1. Effect of spatial resolution on mapping upland habitats

Table 4.4 summarises the results obtained by the 5-fold cross-validation method for WM and SM models trained at five different spatial resolutions. Overall, the models trained using a higher spatial resolution dataset generally achieved better performance. For both sites, the models trained on 2-m resolution datasets had the highest mean accuracy (WM_2m: 80.34% and SM_2m: 79.64%), followed by models trained on 4-m resolution datasets (WM_4m and SM_4m) with an accuracy of 77-79%. Furthermore, there was a decrease in accuracy of about 4-7% when the coarser 10-m resolution datasets were used for model training compared to when the 2-m resolution datasets were used. However, this trend of higher spatial resolution datasets leading to better classification accuracy was

not always the case. For example, the coarser resolution-based WM_8m model achieved a higher accuracy (75.94%) than the WM_6m model (73.71%). Similarly, the SM_10m model performed better than the SM_8m model (75.44% vs. 73.83%).

Table 4.4. 5-fold cross-validation mean accuracy and standard error results for WM and SM models. Descriptions of the models are presented in Table 4.3.

Wicklow Mountains site			Slieve Mish Mountains site		
Models	Mean accuracy (%)	Standard error	Models	Mean accuracy (%)	Standard error
WM_2m	80.34	3.78	SM_2m	79.64	5.43
WM_4m	77.47	3.21	SM_4m	78.78	5.15
WM_6m	73.71	3.92	SM_6m	77.81	6.34
WM_8m	75.94	2.10	SM_8m	73.83	6.25
WM_10m	73.55	1.95	SM_10m	75.44	7.61

The contributions of each input variable to the model predictions are summarised in Figure 4.5. In general, spectral variables consistently showed higher importance for both sites, while textural variables presented lower importance values. It can be observed that the importance of variables varied depending on the spatial resolution of the datasets used in the modelling. For example, the importance of textural variables increased gradually as the spatial resolution coarsened. Whereas for spectral variables, their importance values slightly decreased with lower spatial resolutions. Elevation and slope variables showed almost similar importance values across spatial resolutions.

Models for each site exhibited different important variables. The elevation and slope, representing the topography of the site, were often ranked as the most important variables in relation to the WM models. Conversely, the MGRVI vegetation index and NIR_correlation textural variable were the most important variables in the majority of the SM models (Figure 4.5).

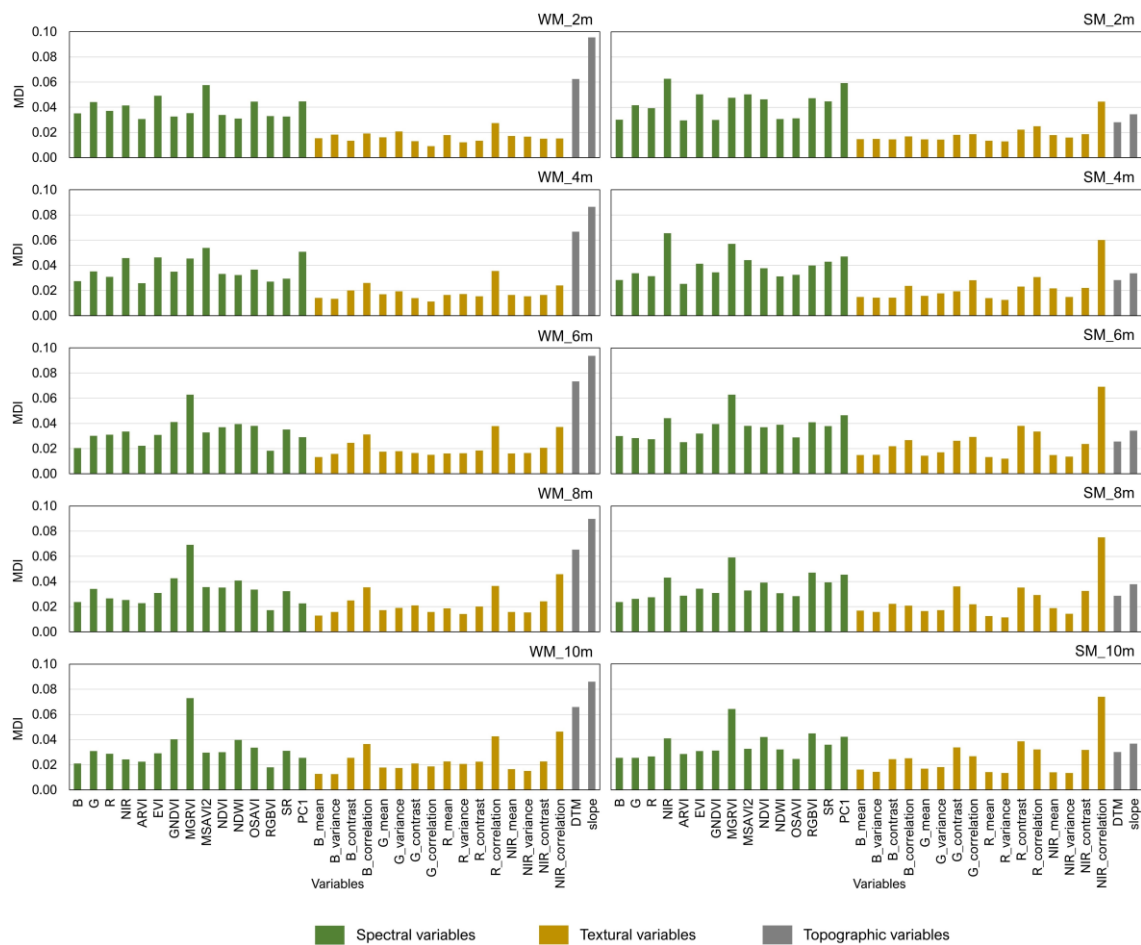


Figure 4.5. Variable importance scores for WM and SM models at different spatial resolutions. The x-axis displays the input variables analysed in the study, while the y-axis represents the variable importance score calculated using MDI. Variable descriptions can be found in Table 4.2.

Figure 4.6 shows the crisp RF-based classified images of a part of the WM site to compare the impact of using different spatial resolution datasets on model output. Overall, classification results across the five spatial resolutions demonstrated comparable patterns of habitat distributions with greater spatial detail in the 2-m classified image. These details, however, differ depending on the spatial resolution used. For example, the size of mapped waterbodies within the encircled section of Figure 4.6 increased when using lower spatial resolutions. A similar trend was also observed for patches of eroding blanket bogs.

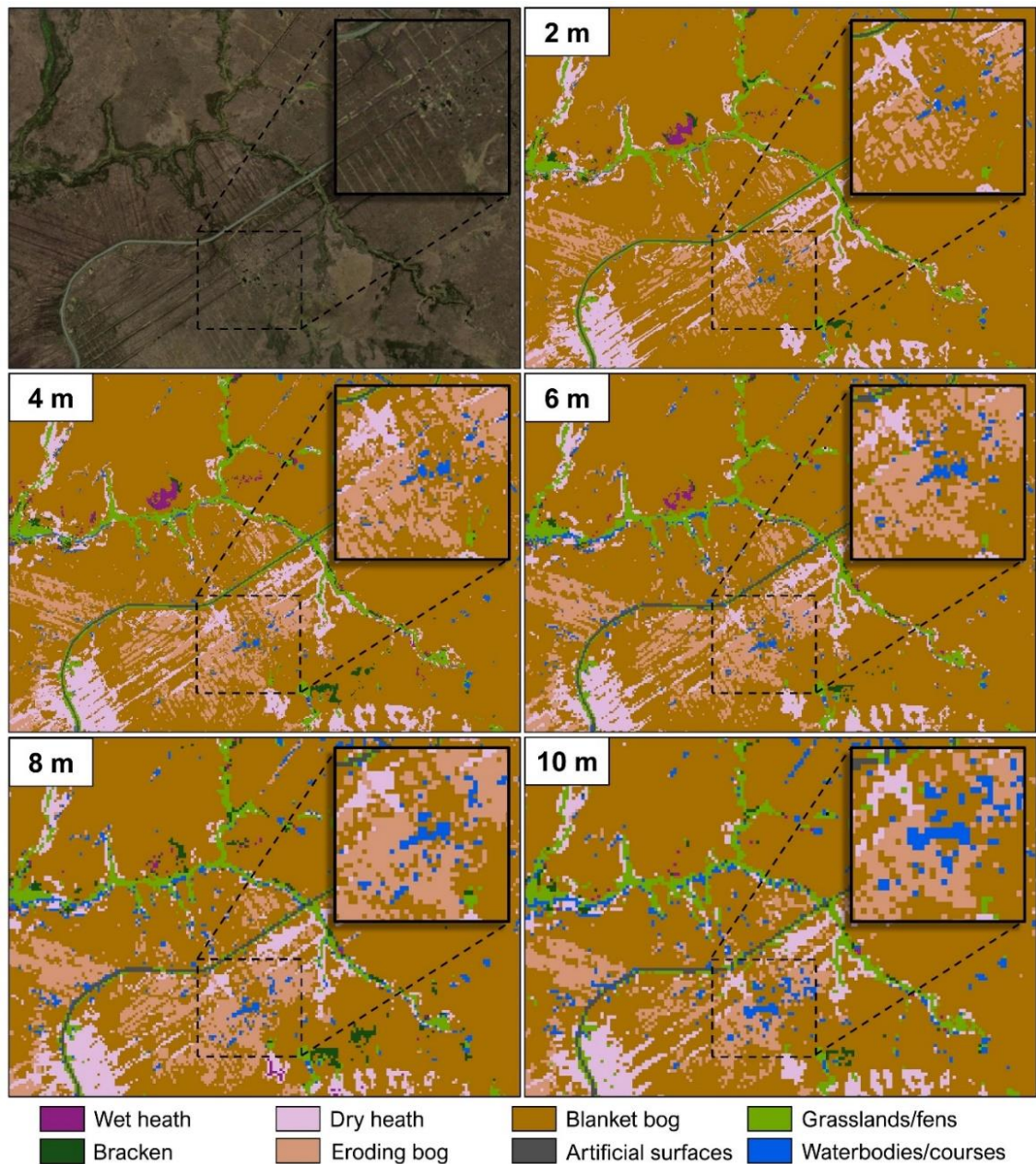


Figure 4.6. Crisp classification results over a portion of the WM site (top left) based on five different spatial resolutions (2 m, 4 m, 6 m, 8 m and 10 m). The encircled section illustrates the increasing size of mapped waterbodies with lower spatial resolution datasets.

Figure 4.7 displays the areal proportions of habitats as classified using different resolutions for the WM and SM sites. No significant differences were observed in the

areal proportions between the different habitats. However, it can be observed that habitats with the smallest proportions generally increased in size with lower spatial resolutions.

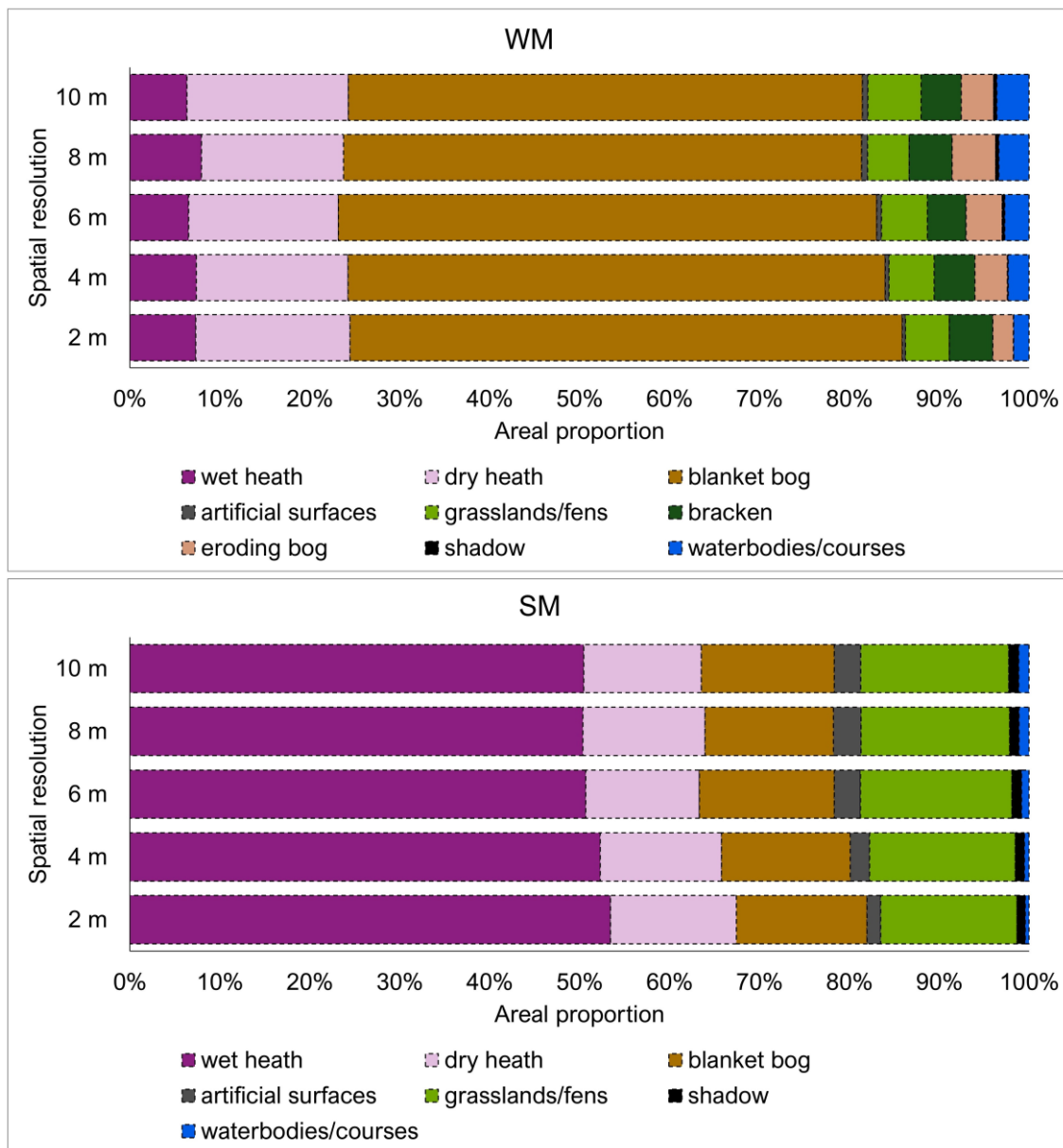


Figure 4.7. Areal proportion for each habitat computed from the model classification results for WM and SM sites.

4.4.2. Model predictions

Figure 4.8 shows the crisp and fuzzy classification results over a portion of the WM site (Figure 4.8a). In the crisp classification result (Figure 4.8b), each pixel was characterised

by a single habitat type. In contrast, the fuzzy classification results (Figure 4.8c – Figure 4.8k) represented the probability of each habitat being present on each pixel, i.e., there was one greyscale image for every habitat type. Hence, the number of output fuzzy images corresponds to the number of habitats considered in the modelling. The white pixels in the image indicate areas where a particular habitat was predicted to be 100% present in that pixel. For example, blanket bog (Figure 4.8e) and grasslands/fens (Figure 4.8f) were some of the habitats that were distinctly separated from their surroundings. In contrast, black pixels represent those areas where a habitat was predicted with 0% probability, indicating its absence in that pixel. Grey pixels, therefore, denote predictions that fall between 0.0 and 1.0. These predictions could suggest areas of mixed habitats. The total of all the probabilities in a pixel adds up to 1.0. The complete crisp and fuzzy classification results for WM and SM sites can be found in Figures C.1-C.10, Appendix C.

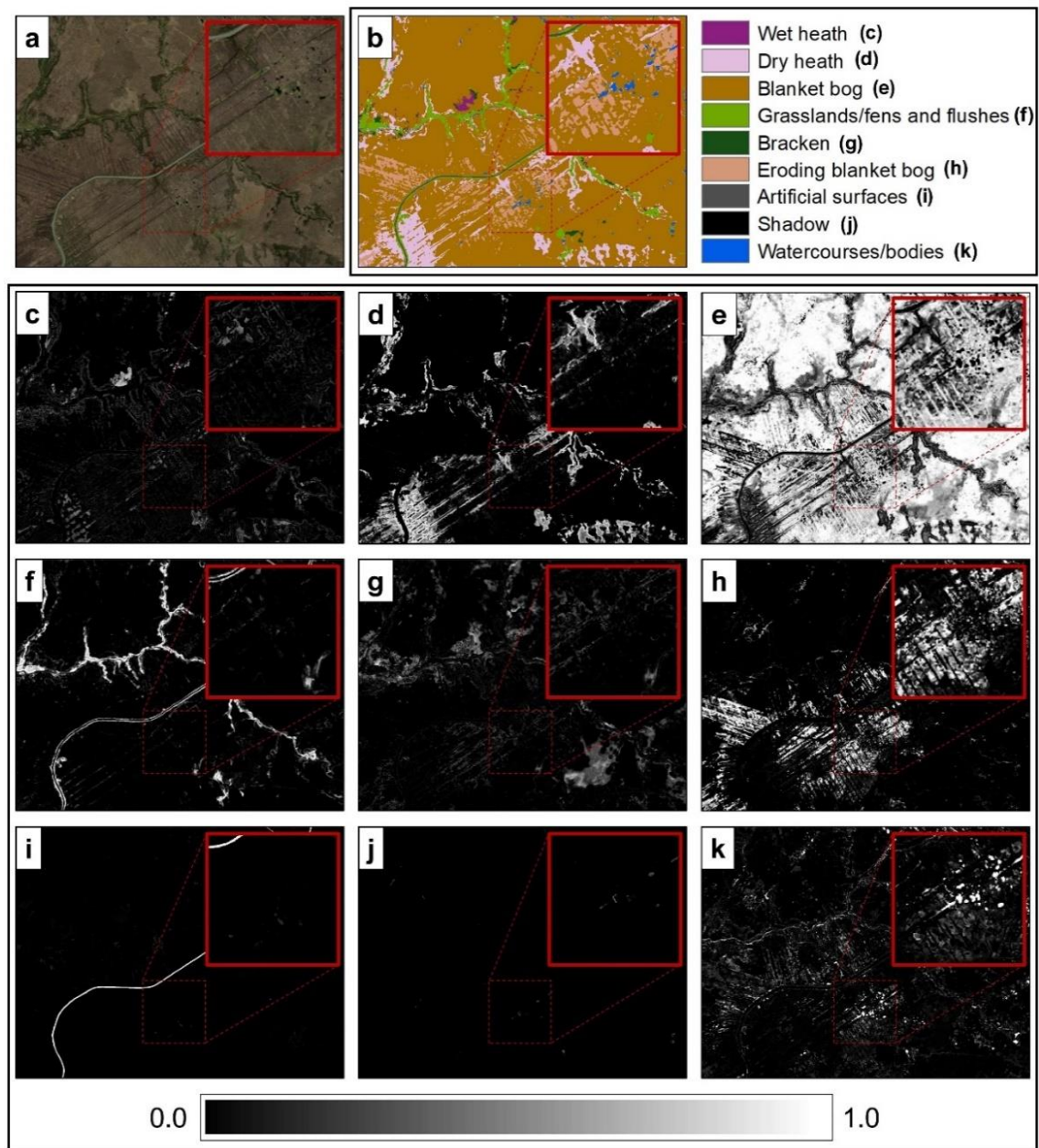


Figure 4.8. Basemap of a portion of the WM site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

4.4.3. Distribution and extent of upland habitats

The WM site (Figure 4.9a) was dominated by blanket bogs, covering almost 60% of the site. This habitat was most extensive on a relatively flat or gently sloping higher altitude part of the site (Figure 4.9b and Figure 4.9c). Large areas of wet heath and dry heath were found on the hillsides, where the terrain was characterised by steep slopes (Figure 4.9b and Figure 4.9c). In these areas, mosaics of dense bracken and grasslands/fens were also present. Eroding blanket bog areas were located near the roads (i.e., classified as artificial surfaces) and often occurred in straight parallel lines, intermixed with dry heath. The site was also characterised by a big lake in the north and several small bog pools around the centre of the map; both were classified as waterbodies/courses (Figure 4.9c).

The colour scale of the entropy image (Figure 4.9d) indicates the level of certainty in the classification, i.e., light pixels being more confident and dark pixels being less confident. The areas mapped as blanket bogs and as other common features that are spectrally distinct, such as waterbodies and roads, were classified with high certainty, as observed by the light colour on the map (Figure 4.9d). However, most wet heath or dry heath areas were classified with less certainty.

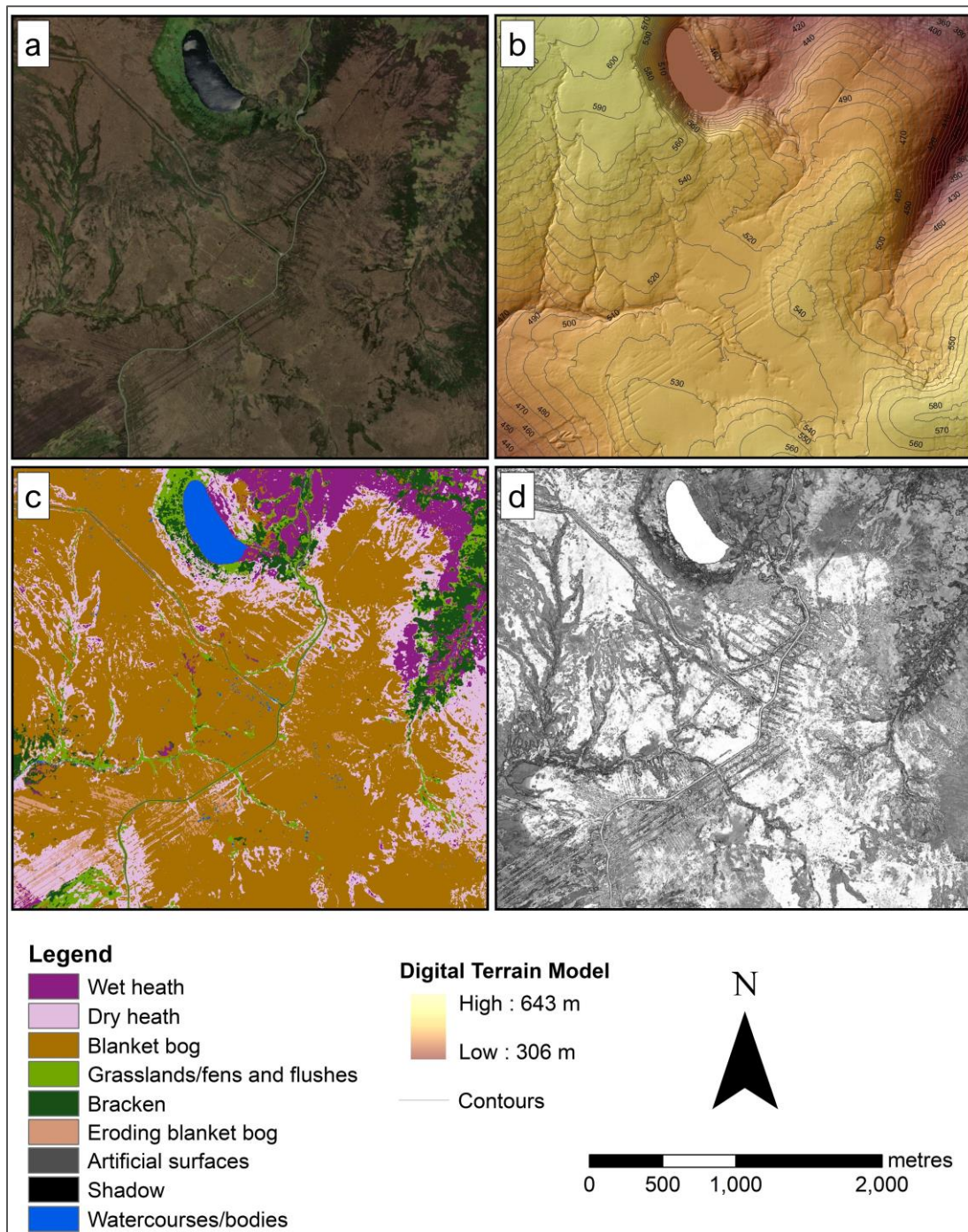


Figure 4.9. Maps showing the basemap (a) and the DTM (b) of the WM site, and the corresponding crisp classification result (c) and entropy image (d) for the 2-m resolution model. The colours in (d) represent the level of certainty: lighter colours being more certain and darker colours being less certain.

The SM site (Figure 4.10a) was covered by about 50% of wet heath. This habitat was mainly located at lower altitudes of the site and where the slope was less steep. As the

altitude increases and the slope gradually becomes steep, wet heath habitat transitioned into mosaics of blanket bogs, dry heath and grasslands/fens (Figure 4.10b and Figure 4.10c). Patches of exposed rocks/scree were scattered across the site. The site was also crossed by narrow watercourses. Shadow was recorded in the north-western part of the SM site due to the quite steep slope obscuring the actual habitat (Figure 4.10b and Figure 4.10c). The entropy image (Figure 4.10d) showed that the dominant cover wet heath areas were generally classified with high certainty, i.e., low entropy. In contrast, dry heath and blanket bog were the areas mapped with less certainty, illustrating an ambiguity in the prediction, possibly due to these habitats occurring in mosaics.

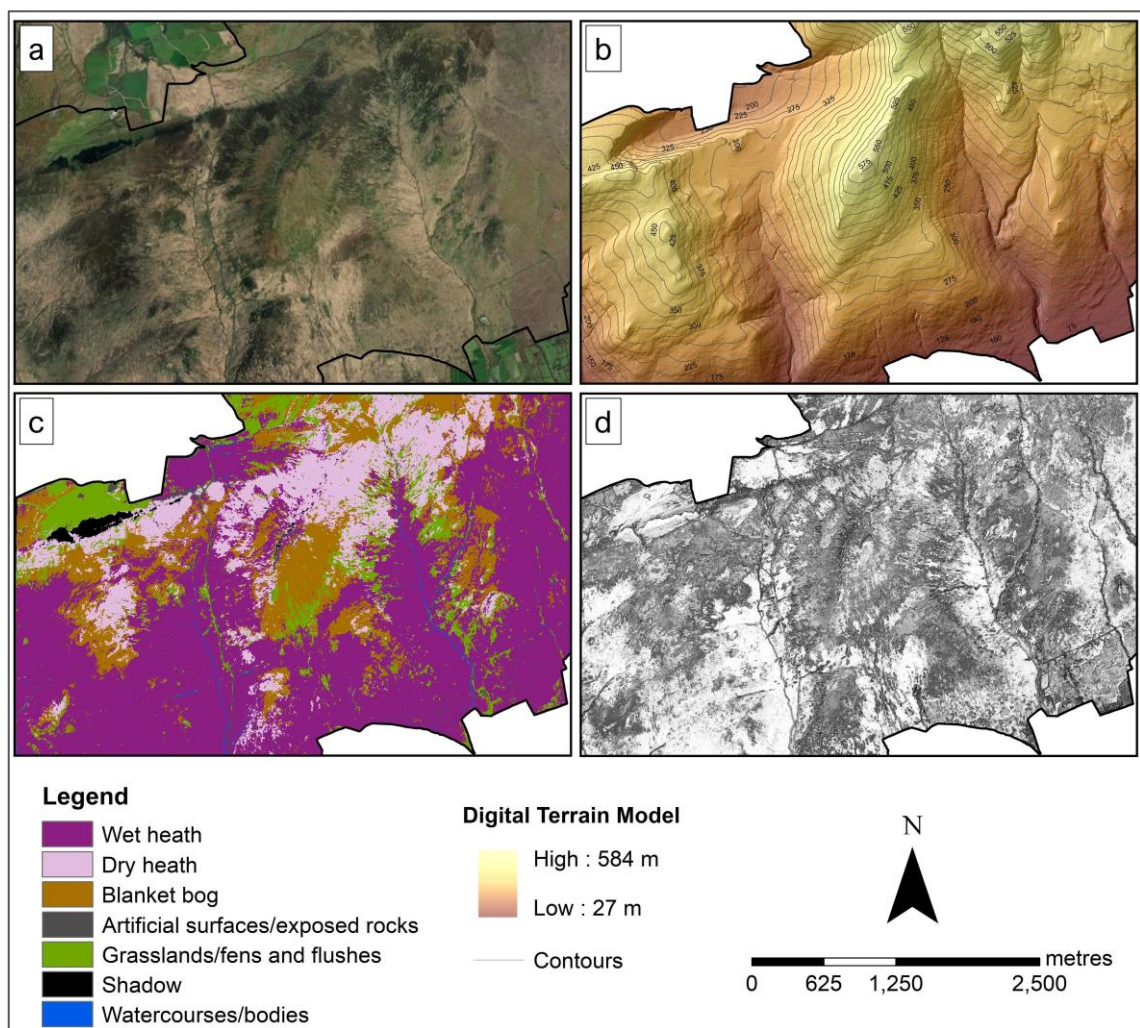


Figure 4.10. Maps showing the basemap (a) and the DTM (b) of the SM site, and the corresponding crisp classification result (c) and entropy image (d) for the 2-m resolution model. The colours in (d) represent the level of certainty: lighter colours being more certain and darker colours being less certain.

4.5. Discussion

This study demonstrates the potential of using high-resolution Pleiades satellite imagery and RF machine learning to accurately map upland habitats in two designated upland sites. The spatial distribution maps produced using the developed RF models provide essential information on the extent and location of protected habitats, as well as on the degraded versions of blanket bogs.

The methodology was designed to evaluate the impact of varying image spatial resolutions on classification accuracy. Our results showed that using finer spatial resolution data generally leads to improved classification accuracy, consistent with other satellite-based remote sensing studies (Boyle et al., 2014; Dorji & Fearn, 2017; J. Fisher et al., 2017). The observed differences between the area of each habitat, depending on what spatial resolution was used in the modelling (see Figure 4.7), could be explained by the lower spatial resolution imagery overestimating and underestimating different habitats. In our study, habitats that occurred as small patches, such as bog pools, were highly influenced by an overestimation of area, while dominant habitats, such as blanket bogs and heath, were affected by an underestimation of area when a lower spatial resolution image was used (see Figure 4.6 and Figure 4.7). This finding has implications when using freely available satellite imagery, such as Sentinel-2 and Landsat-8, for assessing highly heterogeneous or fragmented landscapes. A coarse-resolution image may under- or overestimate the actual size of different habitats.

While our results showed only slight differences in the computed area when mapping with different spatial resolutions ($\Delta_{ave} = 14.7$ hectares between the 2-m and 10-m resolution images), this may lead to a more significant difference if the mapping is conducted on a regional scale. This variation in areas may lead to inaccurate assessments and reporting of the actual habitat status. For example, Hernando et al. (2017) demonstrated that the observed spatial patterns (i.e., connectivity and fragmentation) in forest cover maps varied based on the spatial resolutions of the data used, leading to different conservation status assessments of those maps. Therefore, it is important to consider using an image with a fixed spatial resolution when monitoring habitats with remote sensing to ensure reliability in reporting spatial extents and accurately measuring

the change over time. Reporting the spatial resolution of the data used in the analysis is also crucial.

Our study also compared two RF mapping techniques in classifying upland habitats: crisp and fuzzy. Both techniques can map the general distribution of upland habitats at WM and SM sites. In the crisp classification technique, the distribution of upland habitats is presented as a single map. However, this technique (i.e., assigning one habitat per pixel) may oversimplify the typical complex distributions of upland habitats observed in the field, especially when using images of lower spatial resolutions (Figure 4.6). In the fuzzy classification technique, the distribution of habitats is presented in multiple maps, providing rich information (probabilities of habitat occurrence per pixel) about the habitat composition and patterns (Figure 4.8c-k). The technique can also be applied to detect transition zones between habitats. These zones are observed as the spatial transition in the probability of a habitat from high to low, and in the same area, an inverse pattern of probability change (i.e., from low to high) is exhibited by another habitat (Feilhauer et al., 2021). The study of de Klerk et al. (2018) also used probability maps to further describe and categorise transition zones—depending on their widths—as sharp, moderate and slow transitions. While these zones can be mapped as separate classes using the crisp technique, this may result in many classes, making it difficult to interpret the map (Feilhauer et al., 2021).

Our study also demonstrated the use of entropy images as a measure to estimate confidence predictions. Previous studies have also proposed other measures, such as confusion indices and RGB colour blend images, for estimating confidence in predictions generated from fuzzy classification (Duff et al., 2014; Feilhauer et al., 2021; Zlinszky & Kania, 2016). Understanding a set of fuzzy maps may not be straightforward, especially for non-specialist users; therefore, most fuzzy maps are converted to a single crisp classified map (Zlinszky & Kania, 2016). While conservation managers may prefer using crisp classification maps because they are generally easy to interpret, they can also benefit from entropy images. These images can be used to identify areas on the crisp classified map that are most likely occupied by mosaics of habitats and not just by a single habitat that the crisp classified map indicates. This could suggest that using entropy images can have a more practical significance when mapping with freely accessible satellite images (e.g., Landsat-8 and Sentinel-2), as their coarse spatial resolutions would mean that a

single pixel most likely covers more than one habitat. Also, areas on the map with higher entropy values are considered less accurate than areas with lower entropy values; further assessment of these less accurate areas can be recommended especially if they were previously assessed with lower entropy values, i.e., considered highly accurate. Therefore, this finding suggests that presenting habitat distribution with a crisp classified map can be supplemented with its corresponding entropy image as the latter provides additional information about the confidence in predictions.

Our results highlighted that small or narrow patches were clearly mapped in higher spatial resolution imagery. Detecting small but important features on the map may help better assess the habitat condition. For example, eroding blanket bogs (Figure 4.11), which are often characterised by a network of channels occurring in linear patterns (Fossitt, 2000), were visibly delineated on the WM_2m map (Figure 4.9c). Missed detection of small-sized degradation may lead to false conclusions on the habitat status. Moreover, the information on locations of these eroding bogs (i.e., mapped closer to the road) may indicate that bogs closer to the road are more vulnerable to anthropogenic activities, such as drainage and turf-cutting, than those located far from the road. In the report of Perrin et al. (2017), a lot of the eroding bogs at the WM site is the result of peat extraction by hand and using sausage machines. Therefore, creating detailed and accurate maps have an important implication for conservation and can be used to propose recommendations; for example, bogs closer to the road require special attention for protection as the soil carbon loss in these areas may be significant.



Figure 4.11. Eroded peat surface in the WM site.

Overall, while there is a high cost involved in acquiring higher-resolution satellite data, the detailed information it can provide is of utmost importance in habitat conservation, particularly in spatially extensive, heterogeneously complex landscapes. The crisp and fuzzy mapping approaches presented here can also be used to represent the natural complexity of the spatial distribution of upland habitats. Future studies can explore applying these approaches for mapping uplands more extensively as well as mapping dynamic habitats, such as coastal areas, and those with spectrally similar vegetation communities, such as grasslands. Moreover, since high-resolution DSMs are usually expensive to obtain, future studies could investigate using publicly available global topographic data, such as the Copernicus DSMs (spacedata.copernicus.eu/en/web/guest/collections/copernicus-digital-elevation-model), in the modelling process. Alternatively, the methodology could be modified by excluding DSM-derived variables from the model (i.e., develop a habitat classification model using information from spectral variables only).

4.6. Conclusion

Our study used high-resolution satellite imagery and RF machine learning to map upland habitats. Using higher spatial resolution imagery generally improves mapping accuracy. In the two upland sites we studied, the highest accuracy maps were obtained (WM site – 80.34% and SM site – 79.64%) with the 2-m resolution datasets. These maps provide information on the spatial distribution of habitats in great detail. Coarser spatial resolution datasets, however, resulted in a reduction of the accuracy and a slight overestimation of area for narrow and small-sized habitats. The total percentage area differences between the 2-m and 10-m resolution images are 8% and 11% for the SM and WM sites, respectively. Although these differences may appear small, they could have significant implications in monitoring small-sized habitats, particularly when tracking gradual and subtle changes in these habitats over time. Therefore, a higher spatial resolution dataset is preferred if mapping habitats in a more heterogeneous and diverse landscape. Our study also demonstrated the use of crisp and fuzzy classification techniques in mapping upland habitats. Crisp classification results in a single habitat map, which is relatively easy to interpret. Fuzzy classification delivers probability maps for each habitat considered in the modelling. While these probability maps may be more difficult to interpret, they can represent the typical complex mosaics and gradual transitions of upland habitats as observed in the field. They can also be used to describe spatial confidence in the classification through computing the entropy. Using fuzzy classified maps has the potential to improve our understanding of nature's fuzzy patterns.

4.7. References

- Amani, M., Salehi, B., Mahdavi, S., Granger, J. E., Brisco, B., & Hanson, A. (2017). Wetland classification using multi-source and multi-temporal optical remote sensing data in Newfoundland and Labrador, Canada. *Canadian Journal of Remote Sensing*, 43(4), 360–373. <https://doi.org/10.1080/07038992.2017.1346468>
- Barrett, B., Raab, C., Cawkwell, F., & Green, S. (2016). Upland vegetation mapping using Random Forests with optical and radar satellite data. *Remote Sensing in Ecology and Conservation*, 2(4), 212–231. <https://doi.org/10.1002/rse2.32>

- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, 39, 79–87. <https://doi.org/10.1016/j.jag.2015.02.012>
- Berrar, D. (2019). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Bhatnagar, S., Gill, L., Regan, S., Naughton, O., Johnston, P., Waldren, S., & Ghosh, B. (2020). Mapping vegetation communities inside wetlands using Sentinel-2 imagery in Ireland. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102083. <https://doi.org/10.1016/j.jag.2020.102083>
- Bonn, A., Rebane, M., & Reid, C. (2008). Ecosystem services: a new rationale for conservation of upland environments. In A. Bonn, T. Allott, K. Hubacek, & J. Stewart (Eds.), *Drivers of environmental change in uplands* (1st Ed.). Routledge. <https://doi.org/10.4324/9780203886724>
- Boyle, S. A., Kennedy, C. M., Torres, J., Colman, K., Pérez-Estigarribia, P. E., & De La Sancha, N. U. (2014). High-resolution satellite imagery is an important yet underutilized resource in conservation biology. *PLoS ONE*, 9(1), e86908. <https://doi.org/10.1371/journal.pone.0086908>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st ed.). CRC Press.
- Buchanan, G., Pearce-Higgins, J., Grant, M., Robertson, D., & Waterhouse, T. (2005). Characterization of moorland vegetation and the prediction of bird abundance using remote sensing. *Journal of Biogeography*, 32(4), 697–707. <https://doi.org/10.1111/j.1365-2699.2004.01187.x>
- Cantrell, S. J., Christopherson, J. B., Anderson, C., Stensaas, G. L., Ramaseri Chandra, S. N., Kim, M., & Park, S. (2021). System characterization report on the

- WorldView-3 Imager (ver. 1.1, October 2021). In S. N. Ramaseri Chandra (Ed.), *System characterization of Earth observation sensors: U.S. Geological Survey Open-File Report 2021–1030* (Issue October).
<https://doi.org/10.3133/ofr20211030I>
- Chavez, P. S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24(3), 459–479. [https://doi.org/10.1016/0034-4257\(88\)90019-3](https://doi.org/10.1016/0034-4257(88)90019-3)
- Commission of the European Communities. (1992). Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. *Official Journal of the European Union*, 206, 7–50.
- Connolly, J. (2018). Mapping land use on Irish peatlands using medium resolution satellite imagery. *Irish Geography*, 51(2), 187–204. <https://irishgeography.ie/index.php/irishgeography/article/view/1371>
- Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., Strasser, T., Vanden Borre, J., Toon, S., & Michael, F. (2015). Remote sensing for mapping natural habitats and their conservation status - New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 37(2015), 7–16. <https://doi.org/10.1016/j.jag.2014.11.005>
- Crowle, A., & McCormack, F. (2009). Condition of upland terrestrial habitats. In A. Bonn, T. Allott, K. Hubacek, & J. Stewart (Eds.), *Drivers of environmental change in uplands* (pp. 156–170). Routledge.
- Cruz, C., O’Connell, J., McGuinness, K., Martin, J. R., Perrin, P. M., & Connolly, J. (2023). Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping. *European Journal of Remote Sensing*, 56(1), 2191870. <https://doi.org/10.1080/22797254.2023.2191870>
- de Klerk, H. M., Burgess, N. D., & Visser, V. (2018). Probabilistic description of vegetation ecotones using remote sensing. *Ecological Informatics*, 46, 125–132. <https://doi.org/10.1016/j.ecoinf.2018.06.001>
- Dorji, P., & Fearn, P. (2017). Impact of the spatial resolution of satellite remote

- sensing sensors in the quantification of total suspended sediment concentration: A case study in turbid waters of Northern Western Australia. *PLoS ONE*, 12(4), e0175042. <https://doi.org/10.1371/journal.pone.0175042>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- Duff, T. J., Bell, T. L., & York, A. (2014). Recognising fuzzy vegetation pattern: the spatial prediction of floristically defined fuzzy communities using species distribution modelling methods. *Journal of Vegetation Science*, 25(2), 323–337. <https://doi.org/10.1111/jvs.12092>
- European Commission. (2022). *Proposal for a regulation of the European Parliament and of the Council on nature restoration*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0304&from=EN>
- European Council. (2023). *Nature restoration: Council and Parliament reach agreement on new rules to restore and preserve degraded habitats in the EU*. <https://www.consilium.europa.eu/en/press/press-releases/2023/11/09/nature-restoration-council-and-parliament-reach-agreement-on-new-rules-to-restore-and-preserve-degraded-habitats-in-the-eu/>
- European Environment Agency. (2020). *State of nature in the EU – Results from reporting under the nature directives 2013-2018*, Publications Office, <https://data.europa.eu/doi/10.2800/088178>
- Evans, D. (2012). Building the European Union's Natura 2000 network. *Nature Conservation*, 1, 11–26. <https://doi.org/10.3897/natureconservation.1.1808>
- Feilhauer, H., Zlinszky, A., Kania, A., Foody, G. M., Doktor, D., Lausch, A., & Schmidtlein, S. (2021). Let your maps be fuzzy!—Class probabilities and floristic gradients as alternatives to crisp mapping for remote sensing of vegetation. *Remote Sensing in Ecology and Conservation*, 7(2), 292–305. <https://doi.org/10.1002/rse2.188>

- Fisher, J., Acosta, E. A., Dennedy-Frank, P. J., Kroeger, T., & Boucher, T. M. (2017). Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Remote Sensing in Ecology and Conservation*, 4(2), 137–149. <https://doi.org/10.1002/rse2.61>
- Fisher, P. (1997). The pixel: A snare and a delusion. *International Journal of Remote Sensing*, 18(3), 679–685. <https://doi.org/10.1080/014311697219015>
- Foody, G. M. (1996). Fuzzy modelling of vegetation from remotely sensed imagery. *Ecological Modelling*, 85, 3–12. [https://doi.org/10.1016/0304-3800\(95\)00012-7](https://doi.org/10.1016/0304-3800(95)00012-7)
- Fossitt, J. A. (2000). *A guide to habitats in Ireland*. The Heritage Council.
- Gatis, N., Carless, D., Luscombe, D. J., Brazier, R. E., & Anderson, K. (2022). An operational land cover and land cover change toolbox processing open-source data. *Ecological Solutions and Evidence*, 3. <https://doi.org/10.1002/2688-8319.12162>
- Gilmore, S., Saleem, A., & Dewan, A. (2015). Effectiveness of DOS (Dark-Object Subtraction) method and water index techniques to map wetlands in a rapidly urbanising megacity with Landsat 8 data. *CEUR Workshop Proceedings*, 1323, 100–108. <http://hdl.handle.net/20.500.11937/43918>
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment*, 58(3), 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Gleyzes, M. A., Perret, L., & Kubik, P. (2012). Pleiades system architecture and main performances. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B1, 537–542. <https://doi.org/10.5194/isprsarchives-xxxix-b1-537-2012>
- Hadjimitsis, D. G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M. G., Retalis, A., Michaelides, S., Chrysoulakis, N., Toullos, L., & I. Clayton, C. R. (2010). Atmospheric correction for satellite remotely sensed data intended for agricultural applications: Impact on vegetation indices. *Natural Hazards and Earth System Science*, 10(1), 89–95. <https://doi.org/10.5194/nhess-10-89-2010>

- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 610–621. <https://doi.org/10.1190/segam2015-5927230.1>
- Hernando, A., Velázquez, J., Valbuena, R., Legrand, M., & García-Abril, A. (2017). Influence of the resolution of forest cover maps in evaluating fragmentation and connectivity to assess habitat conservation status. *Ecological Indicators*, 79, 295–302. <https://doi.org/10.1016/j.ecolind.2017.04.031>
- Holden, N. M., & Connolly, J. (2011). Estimating the carbon stock of a blanket peat region using a peat depth inference model. *Catena*, 86(2), 75–85. <https://doi.org/10.1016/j.catena.2011.02.002>
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Ingle, R., Habib, W., Connolly, J., Mccorrey, M., Barry, S., & Saunders, M. (2023). Upscaling methane fluxes from peatlands across a drainage gradient in Ireland using PlanetScope imagery and machine learning tools. *Scientific Reports*, 13(11997), 1–11. <https://doi.org/10.1038/s41598-023-38470-6>
- JNCC. (2015). *UK terrestrial and freshwater habitat types: Upland habitat descriptions*. <https://data.jncc.gov.uk/data/b0b5e833-7300-4234-8ae5-bdbf326e854c/habitat-types-uplands.pdf>
- Jolliffe, I. T. (2002). *Principal Components Analysis* (2nd ed.). Springer. <https://doi.org/10.1007/b98835>
- Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 50(4), 663–666. <https://doi.org/10.1155/2012/651039>
- Kattenborn, T., Lopatin, J., Förster, M., Braun, A. C., & Fassnacht, F. E. (2019). UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. *Remote Sensing of Environment*, 227, 61–73. <https://doi.org/10.1016/j.rse.2019.03.025>

- Kaufman, Y., & Tanre., D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 260–271.
- Kim, M., Park, S., Anderson, C., & Stensaas, G. L. (2022). System characterization report on Planet’s SuperDove. In S. N. Ramaseri Chandra (Ed.), *System characterization of Earth observation sensors: U.S. Geological Survey Open-File Report 2021–1030*. <https://doi.org/10.3133/ofr20211030F>
- Lucas, R., Rowlands, A., Brown, A., Keyworth, S., & Bunting, P. (2007). Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(3), 165–185. <https://doi.org/10.1016/j.isprsjprs.2007.03.003>
- Malley, J. D., Kruppa, J., Dasgupta, A., Malley, K. G., & Ziegler, A. (2012). Probability machines: Consistent probability estimation using nonparametric learning machines. *Methods of Information in Medicine*, 51(1), 74–81. <https://doi.org/10.3414/ME00-01-0052>
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432.
- Mehner, H., Cutler, M., Fairbairn, D., & Thompson, G. (2004). Remote sensing of upland vegetation: The potential of high spatial resolution satellite sensors. *Global Ecology and Biogeography*, 13(4), 359–369. <https://doi.org/10.1111/j.1466-822X.2004.00096.x>
- Melnikova, I., Awaya, Y., Saitoh, T. M., Muraoka, H., & Sasai, T. (2018). Estimation of leaf area index in a mountain forest of central Japan with a 30-m spatial resolution based on Landsat Operational Land Imager Imagery: An application of a simple model for seasonal monitoring. *Remote Sensing*, 10(2), 179. <https://doi.org/10.3390/rs10020179>
- Mücher, I. S., & Hazeu, G. (2021). Contribution of remote sensing techniques for monitoring Natura 2000 sites. In A. M. Schmidt & T. Van der Sluis (Eds.), *Improving the availability of data and information on species, habitats and sites*

- (pp. 40–67). Wageningen Environmental Research/ Ecologic Institute /Milieu Ltd. Wageningen, The Netherlands.
- Müller, J., & Brandl, R. (2009). Assessing biodiversity by remote sensing in mountainous terrain: the potential of LiDAR to predict forest beetle assemblages. *Journal of Applied Ecology*, 46(4), 897–905. <https://doi.org/10.1111/j.1365-2664.2009.01677.x>
- Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M., & Mairota, P. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators*, 33, 45–59. <https://doi.org/10.1016/j.ecolind.2012.09.014>
- National Parks and Wildlife Service. (2016). *Site synopsis: Slieve Mish Mountains SAC*. <https://www.npws.ie/protected-sites/sac/002185>. Department of Arts, Heritage and the Gaeltacht, Ireland.
- National Parks and Wildlife Service. (2017). *Site synopsis: Wicklow Mountains SAC*. <https://www.npws.ie/protected-sites/sac/002122>. Department of Arts, Heritage and the Gaeltacht.
- O’Connell, J., Bradter, U., & Benton, T. G. (2015). Wide-area mapping of small-scale features in agricultural landscapes using airborne remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 165–177. <https://doi.org/10.1016/j.isprsjprs.2015.09.007>
- O’Connell, J., Connolly, J., & Holden, N. M. (2014). A monitoring protocol for vegetation change on Irish peatland and heath. *International Journal of Applied Earth Observation and Geoinformation*, 31, 130–142. <https://doi.org/10.1016/j.jag.2014.03.006>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

- Perrin, P. M., Barron, S. J., Roche, J. R., & O'Hanrahan, B. (2014). *Guidelines for a national survey and conservation assessment of upland vegetation and habitats in Ireland, Version 2.0. (Irish Wildlife Manuals, No. 79)*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland.
<https://www.npws.ie/sites/default/files/publications/pdf/IWM79.pdf>
- Perrin, P. M., O'Hanrahan, B., Roche, J. R., & Barron, S. J. (2009). *Scoping study and pilot survey for a national survey and conservation assessment of upland habitats and vegetation in Ireland*. National Parks and Wildlife Service, Department of Environment, Heritage & Local Government, Ireland.
http://www.botanicalenvironmental.com/wp-content/uploads/2010/03/Perrin_et_al_2009_Upland_Pilot.pdf
- Perrin, P. M., Roche, J. R., Barron, S. J., Daly, O. H., Hodd, R. L., & Devaney, F. M. (2014). *National survey of upland habitats (Phase 4, 2013-2014). Site report no. 15: Slieve Mish Mountains cSAC (002185), Co. Kerry*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland.
https://www.npws.ie/sites/default/files/publications/pdf/NSUH13_Slieve_Mish_Mountains_01b_M.pdf
- Perrin, P. M., Roche, J. R., Daly, O. H., & Simon, J. B. (2017). *Blanket bog turf-cutting management system: Site Report No. 1 Wicklow Mountains SAC (002122)*. National Parks and Wildlife Service, Department of Arts, Heritage and the Gaeltacht, Ireland.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119–126.
[https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- Raab, C., Stroh, H. G., Tonn, B., Meißner, M., Rohwer, N., Balkenhol, N., & Isselstein, J. (2018). Mapping semi-natural grassland communities using multi-temporal RapidEye remote sensing data. *International Journal of Remote Sensing*, 39(17), 5638–5659. <https://doi.org/10.1080/01431161.2018.1504344>
- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2), 95–107.

[https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7)

- Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. In S. C. Freden, E. P. Mercanti, & M. A. Becker (Eds.), *Third Earth Resources Technology Satellite-1 Symposium: The proceedings of a symposium held by Goddard Space Flight Center at Washington, D.C. on December 10-14, 1973* (pp. 309–317).
- Sales, M. H. R., De Bruin, S., Souza, C., & Herold, M. (2022). Land use and land cover area estimates from class membership probability of a Random Forest classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 4402711. <https://doi.org/10.1109/TGRS.2021.3080083>
- Scholefield, P., Morton, D., McShane, G., Carrasco, L., Whitfield, M. G., Rowland, C., Rose, R., Wood, C., Tebbs, E., Dodd, B., & Monteith, D. (2019). Estimating habitat extent and carbon loss from an eroded northern blanket bog using UAV derived imagery and topography. *Progress in Physical Geography: Earth and Environment*, 43(2), 282–298. <https://doi.org/10.1177/0309133319841300>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- Smith, G. F., O'Donoghue, P., O'Hara, K., & Delaney, E. (2011). Best practice guidance for habitat survey and mapping. In *The Heritage Council*. https://www.heritagecouncil.ie/content/files/best_practice_guidance_habitat_survey_mapping_onscreen_version_2011_8mb.pdf
- Suo, C., McGovern, E., & Gilmer, A. (2019). Coastal dune vegetation mapping using a multispectral sensor mounted on an UAS. *Remote Sensing*, 11(15), 1814. <https://doi.org/10.3390/rs11151814>
- van Iersel, W., Straatsma, M., Middelkoop, H., & Addink, E. (2018). Multitemporal classification of river floodplain vegetation using time series of UAV images. *Remote Sensing*, 10(7), 1144. <https://doi.org/10.3390/rs10071144>
- Wicaksono, P. (2017). Mangrove above-ground carbon stock mapping of multi-

- resolution passive remote-sensing systems. *International Journal of Remote Sensing*, 38(6), 1551–1578. <https://doi.org/10.1080/01431161.2017.1283072>
- Wicaksono, P., & Hafizt, M. (2018). Dark target effectiveness for dark-object subtraction atmospheric correction. *IET Image Processing*, 12(4), 582–587. <https://doi.org/https://doi.org/10.1049/iet-ipr.2017.0295>
- Young, J., Watt, A., Nowicki, P., Alard, D., Clitherow, J., Henle, K., Johnson, R., Laczko, E., McCracken, D., Matouch, S., Niemela, J., & Richards, C. (2005). Towards sustainable land use: Identifying and managing the conflicts between human activities and biodiversity conservation in Europe. *Biodiversity and Conservation*, 14(7), 1641–1661. <https://doi.org/10.1007/s10531-004-0536-z>
- Zlinszky, A., & Kania, A. (2016). Will it blend? Visualization and accuracy evaluation of high-resolution fuzzy vegetation maps. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 41, 335–342. <https://doi.org/10.5194/isprsarchives-XLI-B2-335-2016>
- Zuleta, D., Russo, S. E., Barona, A., Barreto-Silva, J. S., Cardenas, D., Castaño, N., Davies, S. J., Detto, M., Sua, S., Turner, B. L., & Duque, A. (2018). Importance of topography for tree species habitat distributions in a terra firme forest in the Colombian Amazon. *Plant and Soil*, 450, 133–149. <https://doi.org/10.1007/s11104-018-3878>

5. Synthesis

In this chapter, a research overview and a summary of the key findings related to the use of high-resolution remote sensing data and machine learning techniques, referred to as HRI-ML, as an approach for habitat monitoring is presented. It is then followed by a general evaluation of the limitations of using the HRI-ML approach. Additionally, the practical implications of the research will be highlighted. The chapter concludes by outlining overall conclusions and suggested topics for future research.

5.1. Research aim

The overarching aim of this research was to develop robust and repeatable methodologies for mapping and monitoring habitats using advanced remote sensing technology and machine learning techniques. The research was motivated by the increasing need for accurate, detailed and timely habitat distribution maps, which are crucial for thorough habitat assessments. Specifically, this thesis has focused on the potential of using HRI-ML technologies to map and monitor protected habitats in Ireland. These cutting-edge mapping technologies have not been fully explored in Ireland despite the excitement surrounding them.

5.2. Summary of key findings

The findings from this research highlight the substantial potential of HRI-ML for generating high-quality, up-to-date, spatially continuous maps of habitat distribution, encompassing areas of degradation within these habitats. This research is one of the first to assess the effectiveness and applicability of state-of-the-art HRI-ML technologies for a detailed characterisation of high-nature-value habitats in Ireland.

The objective and key findings from each of the core chapters of this thesis are summarised in the following subsections and presented in Figure 5.1.

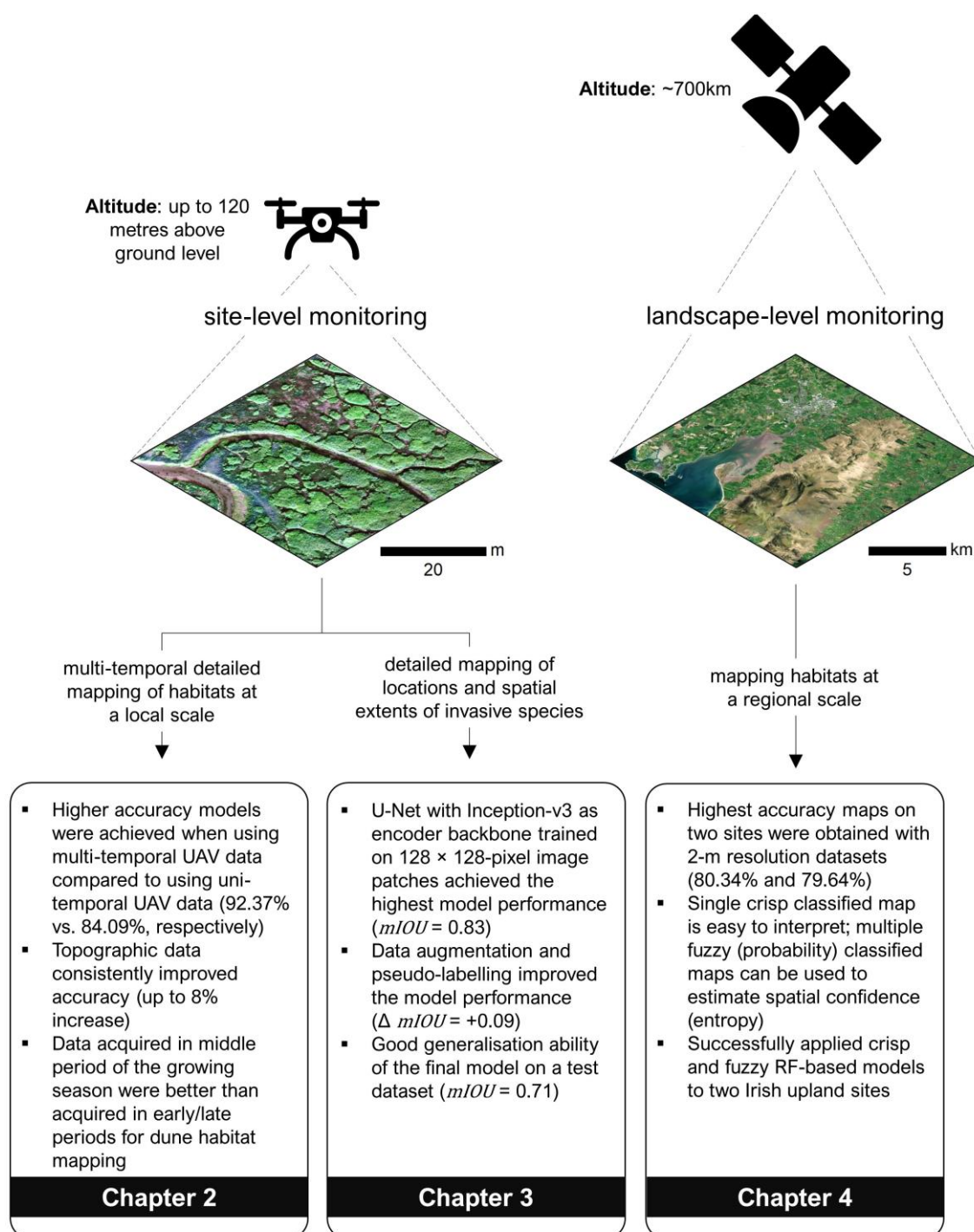


Figure 5.1. Overview of the key findings of this research.

5.2.1. Mapping coastal dune habitats using UAV data and machine learning.

Monitoring fine-scale mosaic and highly dynamic habitats in coastal dune systems with freely available medium-resolution (>10 m) satellite imagery can be difficult, for example, when mapping very narrow areas of patchy mobile dunes or identifying small-scale disturbed areas within the habitats caused by human tracks and coastal erosion (Marzioletti et al., 2019). Chapter 2 assessed the effectiveness of using multiple sets of centimetre-resolution UAV data (acquired throughout the vegetation growing season: early, middle and late) and the Random Forest machine learning algorithm as a tool for accurate and detailed mapping of coastal dune habitats. This study produced higher resolution maps of dune habitats than previous dune studies using UAV multispectral data (e.g., 9.6 cm for Suo et al., 2019, 14 cm for Madurapperuma et al., 2018 and 10-m resampled map for Agrillo et al., 2023). The results also showed that models trained on multiple sets of UAV data acquired throughout the growing season achieved higher accuracy than models trained on just one set of data. This indicates that the variations observed in spectral information (i.e., reflectance) of vegetation communities within habitats throughout the growing season, are useful to discriminate vegetative differences. When considering the use of multiple UAV datasets in modelling, processing several model variables, such as vegetation indices, is not problematic for Random Forest. This algorithm adeptly manages many variables, reducing the risk of overfitting.

In multi-temporal mapping using the Random Forest algorithm, the images acquired at different times must be radiometrically calibrated (i.e., adjusting the image pixel values to account for variations caused by solar and atmospheric effects, such as absorption and scattering). This calibration is crucial to ensure the consistency of measurements between images, allowing for more accurate and reliable comparison and analysis of changes over time.

Furthermore, the results highlighted the importance of UAV-derived topographic data (elevation and slope) in the modelling. Its inclusion increased classification accuracy, suggesting that topographic information is relevant and crucial for mapping the spatial distribution of habitats, especially those with similar vegetation communities, e.g., dunes with *Salix repens* (Annex I habitat 2170) and dune slacks (Annex I habitat 2190). This

finding demonstrates an advantage of using UAVs in mapping habitats, i.e., topographic data can be obtained free of charge, as this is an intermediate output from the photogrammetric processing of UAV images.

Identifying the optimal period for image acquisition during the growing season, which yields the highest differentiation between habitats, is crucial for practical reasons. For example, conducting a single UAV flight can be more cost-efficient and less complicated in terms of logistics than carrying out multiple flights in a year. Results of this study showed that the middle period of the growing season, i.e., the flowering period, provides the highest overall accuracy on two test datasets (*test 1* = 0.81, *test 2* = 0.86) for dune habitat classification. During this period, the spectral differences between habitats are enhanced. This finding can be used as a guide for scheduling future UAV image acquisitions for dune systems.

When planning a UAV survey, it is crucial to set the parameters correctly to ensure that the images captured have the desired level of detail. In this study, the spatial resolution was chosen to be sufficient in distinguishing between different vegetation communities within habitats. The required spatial resolution, recommended flight line overlaps (> 75%) and the size of the study site led to setting the UAV flight altitude at 60 meters for the DJI Phantom 4 Pro. Although a higher spatial resolution is sometimes desirable, achieving it may require multiple flights to cover the entire site due to the limitation of the UAV's battery life. This scenario could cause variations in the brightness levels of images across the different flights.

It is equally important to note that errors and uncertainties may have occurred during the modelling. One potential source of error could be in the data collection stage. Specifically, the spatial inaccuracy of the UAV and the GNSS systems may result in up to a metre of error. This spatial error can propagate throughout the workflow by causing misalignments between the GIS-based polygon field data and the habitats captured by the UAV, leading to incorrect labels for each habitat type on the image. In this study, this potential error was minimised by ensuring that both data acquired by UAV and GNSS systems were geometrically corrected and used the same reference system, which is the WGS 1984 Universal Transverse Mercator Zone 29N. Moreover, field data collectors (i.e., ecologists) and remote sensing analysts performed on-screen checks to ensure the field

data were positioned correctly with the correct habitats in the image. Other potential errors related to data collection include the influence of adverse weather conditions (e.g., rain, fog and high winds) on the quality of the acquired images, the different levels of competency of the ecologists in identifying habitat types correctly, human error in recording habitat information and the complexity of the habitats. Appendices A1 and A2 provide detailed UAV survey and field data collection protocols and strategies, which can be used as a guide to mitigate the occurrence of these errors.

The study also considered the potential effects of spatial autocorrelation when classifying habitats. One approach to mitigate these effects was the implementation of a random sampling design method during field data collection. This sampling method helped minimise spatial bias by ensuring the collected samples were not clustered based on proximity. The RF algorithm, as an ensemble learning technique, also helped address spatial autocorrelation. By introducing randomness into the training process (i.e., each tree in an RF model was independently trained on different subsets of samples and predictor variables), an RF model could reduce spatial dependence within the data. While this study used cross-validation to randomly select training and validation datasets multiple times during the modelling, it can be beneficial to apply spatially stratified cross-validation to further ensure that these datasets are divided into distinct geographical subsets.

This work addressed the specific objective of “assessing the effectiveness of UAV remote sensing and machine learning for mapping coastal dune habitats”.

5.2.2. Deep learning-based mapping of invasive species using UAV images.

Deep learning techniques have been previously applied to aerial RGB imagery to delineate building and road features. These features are typically regular in shape with smooth edges and textures, and their appearance can remain stationary over time (Alshehhi et al., 2017; Ji et al., 2019; Li et al., 2021). Chapter 3 of this thesis explored deep learning techniques on RGB imagery to delineate irregularly shaped canopies of plant species, which can be challenging to map. Moreover, some plant species are highly

dynamic, and their colours and density can change with different seasons throughout the year, making it even more challenging to map them accurately.

Tracking the location and expansion of invasive species is fundamental for effectively managing habitats (Pyšek & Richardson, 2010). In this study, an application of deep learning and UAV images for mapping *Spartina anglica*, an invasive species in Ireland. The choice of model network structure and the use of techniques to increase the size and quality of training data are important considerations when creating a robust model for mapping invasive species. When training a deep learning model, splitting the entire UAV orthomosaic into smaller image patches is essential, as using the entire orthomosaic can require significant computational resources. This splitting approach allows for more efficient processing as only a subset of data is loaded at any given time, reducing computer memory usage. However, an image patch size should be chosen carefully to ensure it is large enough to provide sufficient spatial context for a deep learning network to recognise. The minimum image patch size usually depends on the spatial resolution of the image (i.e., higher spatial resolution requires a bigger image patch size) as well as how many pixels are needed to define the spatial context of individual clonal patches of the plant species. Generally, at least four pixels are required to detect the smallest object or at least two pixels to represent the narrowest object in an image (Hengl, 2006). In this study, sixty models were trained and compared. The U-Net architecture model with Inception-v3 as the encoder backbone, trained on 128×128 -pixel image patches, performed the best with an *mIOU* score of 0.832.

The study also highlighted that increasing the amount and diversity of training data by applying augmentation and pseudo-labelling techniques improved the predictive performance of the model (increased *mIOU* score by 0.09). The final model was tested on a new set of UAV images that was acquired from a different location at a different time of the year. Despite the variations present in these images—in terms of image brightness, colour of *S. anglica* and species composition surrounding *S. anglica*—the model achieved an *mIOU* score of 0.712. This score was 0.158 higher than the score obtained when only the original labelled data were used to train the model. The result indicates that the final model exhibits good generalisation ability, suggesting that the applied techniques helped the model handle spatial and temporal variations in the test data.

This work addressed the specific objective of “developing a deep learning model to automatically map the spatial distribution of a coastal invasive species *Spartina anglica* on UAV images”.

5.2.3. Mapping upland habitats using high-resolution satellite imagery and machine learning.

The spatial extent limitation of UAV surveys, as well as various environmental challenges (e.g., high winds, rain, poor visibility and steep slopes) that are often present when surveying the upland, motivated the use of satellite imagery. Chapter 4 utilised high-resolution satellite imagery for mapping upland habitat types. The impact of spatial resolution on the accuracy and classification details was investigated. In general, models trained on higher spatial resolution images performed better than those trained on lower resolution images. These higher resolution (2 metres) images are beneficial for mapping in the heterogeneous landscape of uplands, which may be difficult when using the freely available but coarser resolution (10 metres) Sentinel-2 satellite imagery.

Crisp and fuzzy classification techniques based on Random Forest were implemented on the images, and the results were compared. The crisp technique assigns a single habitat to every pixel on the image, which is then displayed in different colours based on the habitats they represent. This produces a map that is simple and straightforward to interpret. In relation to the spatial resolution of the image, the coarse resolution did not match the size of small-scale habitats in a diverse landscape of uplands. An example is when the resampled 10-m resolution image was used to map small-scale bog pools or patches of eroding blanket bog. This mismatch between the image resolution and the size of the habitats being mapped could lead to an inaccurate assessment of the actual habitat status. Similarly, when mapping different plant species that often occur in mosaic patterns, using images with coarser spatial resolutions can result in overestimating or underestimating areas. This issue arises when multiple species are present within a single pixel. In such instances, the crisp classification model assigns the pixel to its most abundant species, leading to the misrepresentation of the actual spatial distribution of each species in a landscape.

On the other hand, using the fuzzy classification technique to generate probability maps is a promising alternative to assess the spatial distribution of complex mosaics and the heterogeneous nature of many upland areas. The maps produced from this technique highlighted areas where gradual transitions and mosaics of habitats occur. They were also used to estimate the spatial confidence in classification by generating an entropy map. A high entropy value suggests lower confidence in the classification, usually corresponding to areas of mixed habitats. A lower entropy value indicates higher confidence in the classification, generally corresponding to areas of dominant habitats. Both techniques were successfully used to map habitats at two Irish upland sites.

This work addressed the specific objective of “utilising high-resolution satellite imagery and machine learning-based fuzzy classification techniques for mapping complex upland habitat types”.

Overall, it can be concluded that the work presented in this thesis achieved the overall research aim, as defined in Chapter 1.5. While specific habitats (Chapter 2: coastal dunes, Chapter 4: uplands) and plant species (Chapter 3: *Spartina anglica*) at selected sites were used as case studies, the developed methodologies have wider applications beyond the studied habitats and geographical areas. With the wide-scale availability of UAV and machine learning technologies, these methodologies can be applied to map different habitat types in other regions. With this, the potential of these technologies can be effectively utilised to help achieve high standards for nature conservation in the EU and globally (e.g., EU Biodiversity Strategy 2030, Nature Restoration Law and Ramsar Convention on Wetlands). Figure 5.2 displays sample maps produced from the methodologies presented in this thesis.

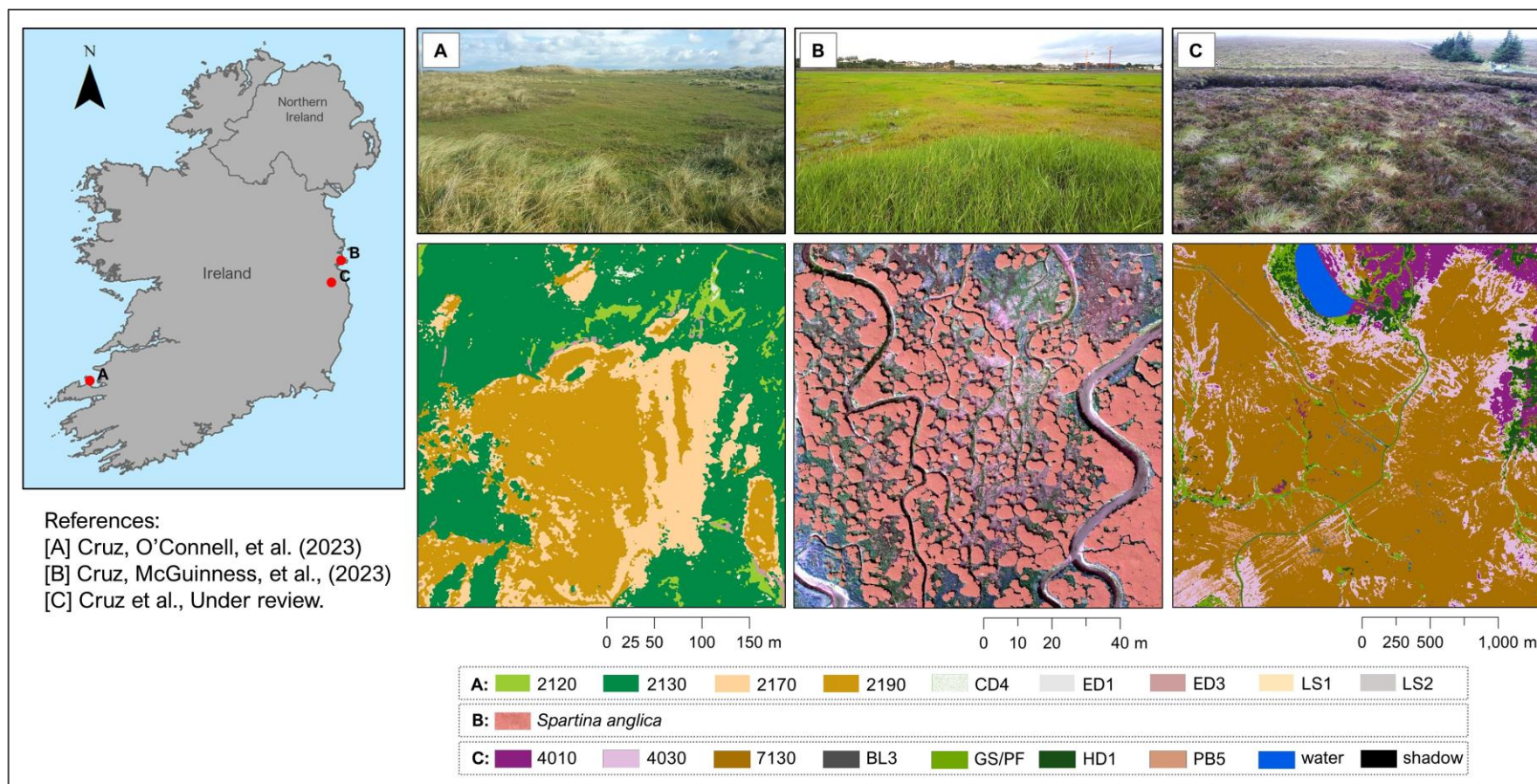


Figure 5.2. Sample output maps produced in this thesis. Refer to individual Chapters 2 and 4 for the complete habitat names for maps A and C, respectively.

5.3. Evaluation of the combined high-resolution imagery and machine learning approach for habitat monitoring

This subsection discusses a number of challenges associated with the use of the HRI-ML approach in habitat monitoring. Although the approach can generally offer accurate, detailed and faster generation of habitat maps, continued developments are essential to address these challenges, thereby further enhancing its value for operational use. Some of these challenges are also highlighted in Chapters 2, 3 and 4.

5.3.1. UAV surveys cover limited areas

Limited battery life and current UAV flight regulations restrict the area that can be covered by a UAV flight (Cruzan et al., 2016; Fraser et al., 2016). A single battery of most consumer and professional UAV systems typically lasts between 20 to 40 minutes of flight time; this can hinder surveying large areas (Gonçalves & Henriques, 2015). Moreover, flying against strong winds—typical in the upland and coastal regions—can significantly drain the batteries, further decreasing the area covered. Flight regulations related to always maintaining a visual line of sight of operators with the UAV when it is in the air, airspace restrictions, as well as the maximum flight altitude (e.g., 120 metres above ground level in the case of Ireland) also reduce the area that a survey can cover (Duffy et al., 2018; Manfreda et al., 2018). For these reasons, multiple flights are often required to survey large areas exceeding a few square kilometres. However, this requirement increases the resources (time and cost) needed to complete the survey (Gonçalves & Henriques, 2015). Also, data acquired from multiple flights may contain illumination variations, which can introduce additional challenges in relation to processing, such as mosaicking and calibration of images (Whitehead & Hugenholtz, 2014).

5.3.2. Limited labelled remote sensing data

In most cases, large amounts of labelled training data are needed to create a well-generalised (robust) machine learning model. A robust model is characterised by its

ability to accurately predict the distributions of similar habitats on imagery acquired from different geographic sites (spatial generalisation) or at different periods of the year (temporal generalisation) (Tuanmu et al., 2011). In remote sensing, however, there is only very limited available labelled data to train a machine learning model. The study proposed methods for increasing the amount and diversity of labelled data for the classification of UAV images: sliding kernel dividing reference data polygons (Chapter 2) and data augmentation and pseudo-labelling techniques (Chapter 3). Nevertheless, more high-quality labelled data is still needed due to a wide range of scenarios and variability in the real world, such as habitat complexity, seasonal changes and environmental conditions.

5.3.3. Requirement of high computation resources

Generally, training a deep learning model can be very computationally demanding due to the network complexity, which can have up to millions of parameters to train, and the need for a large set of labelled data. For example, the processing of about 44,000 128×128 -pixel image patches in Chapter 3 (*Spartina anglica* mapping using deep learning) was performed using a GPU with 12 GB of GDDR6 memory. It is expected that higher computational power is needed to retrain the model with more datasets before application to a regional scale; this can be logistically prohibitive. Nevertheless, ongoing developments are focused on increasing the computational efficiency of high-performing deep learning techniques (e.g., Y. Li et al., 2023).

5.3.4. Highly technical skills are needed for data collection, processing and analysis

Highly specialised technical skills are critical for incorporating advanced remote sensing and machine learning technologies into operational monitoring tasks. Experienced and trained ecologists are crucial for collecting high-quality field data in order to produce accurate machine learning-based habitat maps. A skilled and licensed UAV operator is also vital to ensure the acquisition of high-quality UAV images. The majority of UAV operators do not have the necessary equipment for habitat and vegetation mapping (i.e., UAV-based multispectral cameras), as these sensors are more expensive than the standard RGB sensors. Moreover, conducting GCP surveys to provide accurate positioning of the images is necessary for multi-temporal mapping applications. Technical skills and

knowledge are required to operate modern GNSS devices, accurately measure the coordinates of GCPs and, if needed, post-process the data (i.e., post-processing kinematic).¹

Data processing and analysis can either be implemented on a code-based platform or open-source/commercial software. Extensive training is necessary to replicate or adapt the methods for use in the future. Continuous learning is also fundamental to stay up to date with the technological advancements in this field.

5.4. Practical implications of the HRI-ML approach

There is no one-size-fits-all methodological solution for habitat monitoring (Sparrow et al., 2020; Vanden Borre, Haest, et al., 2011). The choice of processing techniques and remote sensing platforms for data collection depends on several factors, including but not limited to the spatial extent of the area to be covered and the spatiotemporal dynamics of habitats being monitored (Alvarez-Vanhard et al., 2020). In this thesis, pixel-based Random Forest analysis of data acquired using UAVs showed promising results for detailed mapping of complex dune habitats (Chapter 2). These high-resolution habitat distribution maps can provide critical information on areas that are vulnerable to the impacts of human activities, such as frequently used tracks and beach access (Cruz, O’Connell, et al., 2023). In addition, applying deep learning techniques on UAV imagery to delineate the spatial distributions of invasive species resulted in a robust segmentation model (Chapter 3). This high-resolution mapping of invasive species is crucial for the early detection of its expansion before it becomes widespread and difficult to control. However, acquiring data using UAVs to produce high-resolution maps, such as in Chapters 2 and 3, is only suitable for small areas, usually a few square kilometres. For monitoring habitats in an extensive and rugged terrain of uplands, where access by UAV operators may be difficult and can be associated with adverse weather conditions (poor visibility, strong winds and rain), satellite imagery is more appropriate (Chapter 4). The fuzzy classification technique was used to map areas in the uplands, as they usually

¹ Post-processing kinematic (PPK) is a surveying technique used to accurately determine the position of a point on the Earth’s surface by applying the corrections after data collection. Refer to Appendix A3 for the processing steps involved in PPK.

comprise complex mosaics and highly diverse habitats. Therefore, different applications require careful consideration of the data and techniques to ensure the relevance of the mapping output to the actual needs of the users.

Equally important is the integration of methods ranging from satellite-based to UAV-based mapping to field-based assessments for more effective and holistic monitoring of habitats (Mao et al., 2022). Satellite imagery can cover regional-extent areas; however, they often have coarse spatial resolutions (e.g., Landsat with 30-metre resolution and Sentinel-2 with 10-metre resolution), which may not be appropriate for more detailed site-level monitoring (Anderson & Gaston, 2013). Nevertheless, analysis of multi-temporal satellite images can be highly beneficial in detecting any changes (loss or degradation) that may have occurred in habitats (Xian et al., 2012). This process involves using change detection methods, comparing multiple images taken at different times and identifying specific areas facing changes (e.g., land cover) over time. Frequent, more detailed and targeted spatial assessments of these identified small areas can then be conducted using high-resolution UAV images and ground observations. UAVs can be flexibly deployed for repeat surveys, following similar flight paths and data processing methods (Yi, 2017). This consistent approach makes the use of UAVs applicable for monitoring purposes. The high-quality information on the distribution of habitats from the analysis of UAV data allows for a closer examination of the locations and trajectory of habitat changes. This way, intensive field surveys can focus on assessing these hotspot areas, allowing for efficient utilisation of the field survey time and financial resources.

With the growing range of remote sensing platforms and sensors, the future of habitat monitoring with these technologies is an exciting and interesting field. As UAVs are rapidly developing, the ability to fly for longer times and “Beyond the Visual Line of Sight” (BVLOS) can provide a massive opportunity for habitat monitoring (van der Sluijs et al., 2023). These allow UAVs to cover longer distances and acquire data on distant areas that are either not accessible or unsafe to reach by the operators—for example, surveying remote areas or islands where access by boat or foot is challenging (Dabski et al., 2020; Zmarz et al., 2018, 2023). Moreover, not only changes in spectral information can be analysed through multi-temporal UAV-acquired data, but also topographic changes, such as monitoring deposition/erosion in beach environments caused by storms and other human activities (e.g., Casella et al., 2016; Chen et al., 2018; Scarelli et al.,

2017). The 3D information derived from UAV images (Figure 5.3) can also be used to better understand habitat distribution, as certain vegetation communities are likely to occur at specific ranges of elevation or slope (Cruz, O’Connell, et al., 2023). Furthermore, with the decreasing revisit time between satellite observations and increasing spatial resolution of the images, spatiotemporal habitat dynamics can now be observed in real-time, provided cloud-free images are available.

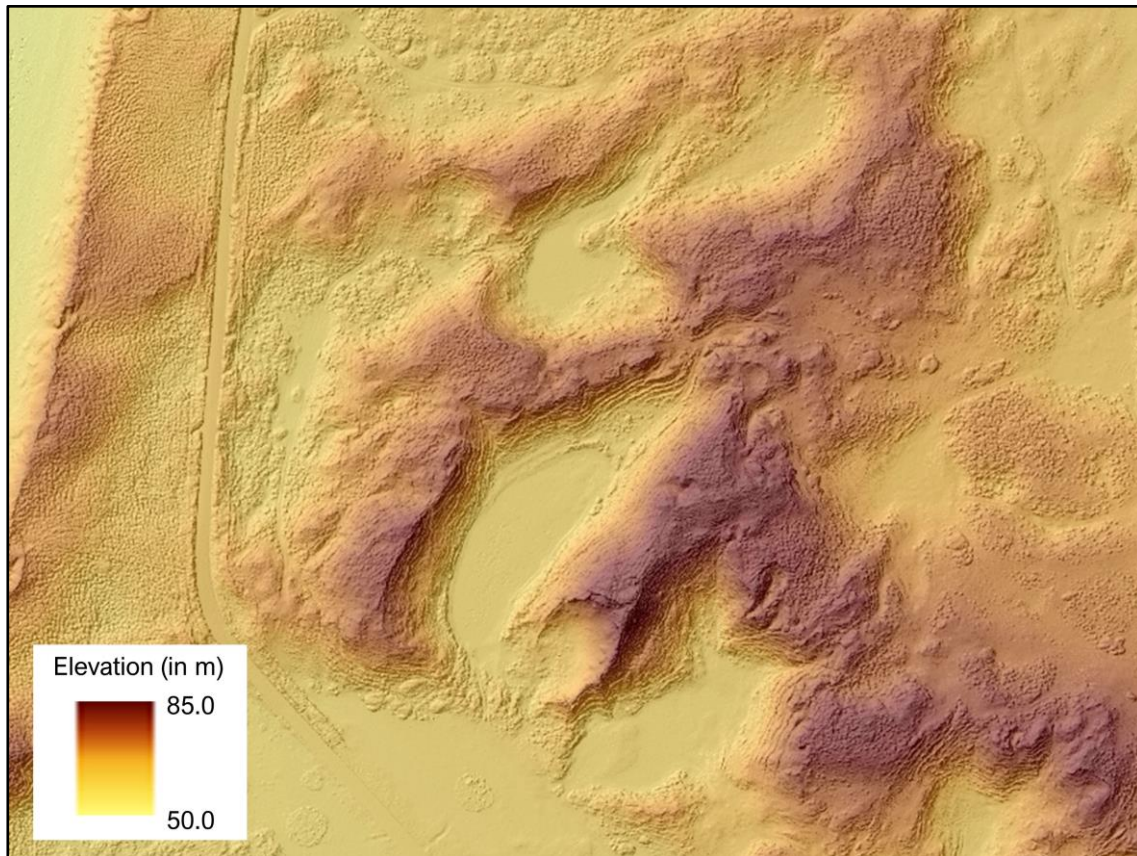


Figure 5.3. Sample UAV-derived elevation model (Source: Cruz, O’Connell, et al., 2023)

Conducting UAV surveys under clear sky conditions generally produces high-quality images, facilitating accurate classifications of different habitat types. However, overcast skies can result in diffuse lighting conditions, reducing the image quality and, therefore, may lead to a lower classification accuracy. It is, however, important that illumination conditions during the surveys are consistent when comparing vegetation conditions over time. Previous studies have developed methods to address this challenge, such as the use of reference panels during a UAV survey to normalise pixel values in images captured at different periods (Y. Wang et al., 2023). Moreover, deep learning is a recent development

in remote sensing classification of RGB imagery wherein a neural network automatically learns patterns on images acquired under various light conditions.

Machine/deep learning algorithms (e.g., Random Forest and Neural Networks) have revolutionised the conservation and management of habitats (Abrams et al., 2019; Barrett et al., 2016; Cruz, McGuinness, et al., 2023; Cruz, O’Connell, et al., 2023; Kattenborn et al., 2019). These algorithms can automate the processing and analysis of remotely sensed datasets and produce accurate results. Such automation is particularly important as vast amounts and increased volumes of datasets from various sources are becoming accessible; this speeds up the mapping process (Parente et al., 2019). However, the performance of a machine/deep learning model is only as good as the data it is trained on (Impollonia et al., 2022). This means that a model trained on a particular habitat in a specific location might not classify similar habitats accurately in different geographic locations (i.e., an overfitting model). For instance, the model developed for mapping dune habitats in the Magharees (Chapter 2) may result in misclassifications when applied to map similar habitats at another dune site in North Bull Island. These misclassifications may be due to the differences in the characteristics of certain habitats from these sites. The fixed dune habitat (Annex I habitat 2130) in North Bull Island is often characterised by rank and grass-dominated, with a high cover of Marram grass due to lack of grazing in the area, as compared to some of the grazed species-rich 2130 habitats recorded within the Magharees site. Therefore, training data should include examples from various scenarios on which the model is expected to generalise. In other words, if a model is intended to be used for vegetation mapping, it is important to include examples from different geographical locations and seasons (e.g., Ge et al., 2021; Wang et al., 2022; Yang et al., 2019).

Similarly, generating machine-labelled data using techniques, such as augmentation or pseudo-labelling, can provide diverse training data necessary for developing a robust model. Chapter 3 confirms that using these machine-labelled data resulted in higher model accuracy (i.e., an increase from 0.554 to 0.712) on a test dataset. Selecting appropriate and relevant data augmentation techniques is crucial for improving the model’s predictive performance in remote sensing tasks. These techniques may include rotations, as UAVs can acquire images from various angles, and colour transformations, which add variations to the images caused by different lighting conditions. Furthermore,

the computation of confidence of the generated pseudo-labels in Chapter 3 allowed the inclusion of only those predicted with higher confidence to be used in subsequent training rounds. In the future, combining all available labelled data generated for different geographical locations can create a large number of more diverse data, allowing the development of models with spatial robustness.

The choice of remote sensing technology, including the type of sensor used, can influence the accuracy of the maps. While an RGB sensor is relatively cheaper to get and its data are easier to process than those of a multispectral, this sensor lacks specific spectral bands useful for habitat and vegetation mapping (e.g., NIR and red edge). Without these bands, the ability to accurately differentiate vegetation types can be limited. These bands are essential in generating various indices for characterising different vegetation properties, such as growth, cover, biomass and chlorophyll content (Xue & Su, 2017). As demonstrated in Chapter 2, the most important variables in the models were NDWI and GNDVI, both requiring an NIR band (refer to Supplementary Table B.4). This result highlights the significance of using a multispectral sensor for vegetation-related mapping tasks. In other words, using a sensor that provides spectral information about the features to be mapped has implications on the accuracy of the output. Nevertheless, deep learning techniques can enhance the capabilities of RGB imagery for vegetation mapping by extracting valuable information from spatial patterns (Chapter 3).

The time requirement of the entire workflow discussed in this thesis should be considered before adopting the methodologies. A UAV survey to acquire imagery typically requires a few hours in the field. Ecological surveys for collecting training and validation datasets depend on the size of the study site and the habitat complexity. For instance, a 36-ha dune site in the Magharees with eleven habitats was surveyed for almost a week by one ecologist. Most of the data processing steps (i.e., generation of model variables, classification and model prediction) in Chapters 2 and 4 were automated. Adopting the developed segmentation methods in Chapter 3 would take about two months of work, including acquiring UAV images, collecting field data, labelling images, training/validating a machine learning model and mapping using the trained model. While developing a robust model initially requires a huge amount of time and effort, once successfully trained, this model can automatically analyse large volumes of data, reducing the need for retraining and constant updates. This automation represents a significant

improvement in the detailed mapping of habitat types and target plant species, which can be suitable for long-term monitoring programs.

Overall, the methodologies presented in Chapters 2-4 are robust and repeatable, meeting the overall research aim, as defined in Chapter 1: “... to develop robust and repeatable methodologies for mapping and monitoring habitats using high-resolution remote sensing images and machine learning techniques”. These methodologies are built using open-source Python (Figure 5.4), ensuring other users with the necessary knowledge and skills in basic remote sensing, machine learning and Python programming can replicate and incorporate them into existing workflows. This means the methodologies can be broadly applied to systematic mapping of different habitat types or target plant species with high accuracy. The required inputs are a table of habitat names and their corresponding coordinates saved as a CSV file (for a classification task) or labelled images (for a segmentation task), a remote sensing image and a digital elevation model (optional). The pre- and post-processing steps—such as manually labelling of datasets, clipping images to site boundary extents and creating map layouts—can be carried out using other open-source software like QGIS.²

² QGIS.org, 2023. QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>

```

1  # -*- coding: utf-8 -*-
2  """
3  Created on Tue Nov 10 08:33:30 2020
4
5  @author: Charmaine
6  """
7
8  import geopandas as gpd
9  import pandas as pd
10 import rasterio
11 import os
12 from shapely.geometry import mapping
13 from rasterio.mask import mask as mk
14 import numpy as np
15 import argparse
16 import sys
17
18 def import_image(rasterpath):
19     try:
20         image = rasterio.open(rasterpath)
21     except OSError:
22         print ("WARNING: could not open raster image %s " % rasterpath)
23     return image
24
25 def resample_by_kernel (shapefile, rasterpath, class_system, kernel_size):
26     class_distribution(shapefile, class_system)
27     final_df = pd.DataFrame()
28     for num, row in shapefile.iterrows():
29         df = image_clip_from_shapefile(num, row, rasterpath, row[0], kernel_size, class_system)
30         final_df = final_df.append(df, ignore_index = True)
31     return final_df
32
33 def image_clip_from_shapefile(num, row, rasterpath, class_code, kernel_size, class_system):
34     stride = kernel_size
35     field_df = pd.DataFrame()
36     for filename in os.listdir(rasterpath):
37         if filename.endswith(".tif"):
38             prefix_colname = os.path.splitext(filename)[0]
39             i = os.path.join(rasterpath, filename)
40             image = import_image(i)
41             try: # use the mask function to mask the original image using the polygon from each row
42                 masked_image, out_transform = mk(image, [mapping(row['geometry'])],
43                                                  crop=True, nodata= 0, all_touched = True)
44                 df = sliding_window(masked_image, kernel_size, stride, prefix_colname)
45                 field_df = pd.concat([field_df, df], axis=1)
46             except Exception as e:
47                 print (e)
48     field_df.insert(0, class_system, class_code)
49     return field_df
50

```

Figure 5.4. Code snippet on open-source Python

5.5. Overall summary, conclusions and recommendations for further work

Mapping habitats is a crucial step considering the implementation of numerous conservation laws at a national or European level. An example is the EU Habitats Directive, which relies on accurate and timely habitat distribution maps to help achieve the biodiversity conservation goals in Europe (Vanden Borre, Paelinckx, et al., 2011). These maps provide valuable information on the location, area and trend of each protected habitat, as well as information on specific areas that are vulnerable to factors, such as human disturbances and invasive species. In Ireland, field surveys and manual visual interpretation of aerial photographs are currently the standard methods for acquiring this

information (Barron et al., 2022; Martin et al., 2018; O'Neill et al., 2023); however, they are often time- and labour-intensive for repeated mapping of large areas. The emergence of more advanced mapping technologies, including high-resolution imagery and automated image interpretation through machine learning techniques, has been recognised as promising complementary tools for habitat conservation (Anderson & Gaston, 2013; Hamylton et al., 2020; Kattenborn et al., 2019). However, these innovative technologies are underexplored in monitoring highly threatened Irish habitats through an accurate and detailed mapping of their distributions. The research presented within this thesis aimed to investigate the potential utilisation of combined high-resolution remote sensing data and machine learning in capturing fine-scale distributions of diverse habitat types. These remote sensing data, together with field-based data, were analysed through various machine learning techniques to develop novel approaches for monitoring Irish habitats. The following are the main conclusions that can be drawn from this research.

- Analysing high-resolution UAV data using the Random Forest machine learning algorithm produced an accurate and comprehensive distribution map of Annex I habitats and the degradations within them.
- Applying methods to deal with changes in vegetation colours with the seasons is crucial in producing accurate habitat distribution maps. This can be done by modelling to determine which period of the growing season is the optimum period for habitat classification, i.e., when spectral differences between vegetation are more enhanced (Chapter 2), or by using data augmentation techniques of deep learning to simulate the colour of the target plant species at any period of the year or time of the day (Chapter 3).
- Implementing data augmentation and pseudo-labelling techniques improved model robustness to spatiotemporal changes in the test data.
- Deep learning is a powerful tool to harness high-resolution data acquired by UAVs for automatically mapping invasive species. This cutting-edge technique can offer a more efficient way of tracking these species, allowing for the timely and targeted implementation of intervention measures.

- The combination of satellite imagery and fuzzy classification technique based on the Random Forest algorithm showed promising results for mapping complex mosaics of habitats in difficult terrain of uplands, where UAVs may be challenging to deploy.
- There is no ‘one-size-fits-all’ mapping methodology for habitat monitoring. Instead, it is a combination of various data and methods that can deal with different scenarios and diverse application areas.
- The HRI-ML approach can provide quantitative information for assessing highly valued habitats. Although the research focuses on specific habitats at selected sites, the approaches presented here—with an appropriate labelled training dataset—are fully adaptable and reproducible to mapping different habitats and plant species. By recognising the potential of these cutting-edge technologies as a vital source of valuable information, it is anticipated that they can be operationalised and integrated into existing monitoring initiatives.

The methodologies and results developed in this thesis were communicated with the National Parks and Wildlife Service team, which is responsible for managing and conserving nature within Ireland. The team was receptive to these new techniques in mapping and monitoring protected habitats. This integration of techniques into their work would mean that an effort to build capacity for those working in the conservation field is needed. This capacity building will involve providing targeted training on remote sensing and machine learning applications for conservation purposes, along with more user-friendly open-access software and freely available, easy-to-access remote sensing data. The same applies to other relevant national and local authorities working on biodiversity management, such as the city and county councils.

5.5.1. Recommendations for future work

A number of challenges related to data acquisition and processing have been identified in this research. The recognition of these challenges has indicated significant areas for future work. Detailed technical suggestions for further research were presented in Chapters 2, 3

and 4. In the broader context of habitat mapping with remote sensing and machine learning, examples of areas for future work are outlined below.

a. Improving the generalisation ability of machine learning models

One of the key recommendations of this thesis is to improve the generalisation ability of machine learning-based models developed for habitat mapping and monitoring. This improvement can be achieved by building a model using large amounts of labelled data acquired from various habitat sites (spatial generalisation) at different periods of the year (temporal generalisation). However, labelling data is a time-consuming and tedious task. Thus, it is recommended that future studies explore how the number of labelled samples influences the accuracy of the model. The findings will have an implication for streamlining the workflow process by reducing the number of labelled samples required without compromising the accuracy of the model. Also, it is crucial that a labelled data repository be established. Such a repository would enable the storage, sharing and public access of these labelled data, together with their metadata. The metadata must include essential details, such as the accuracy, source, description, acquisition date and location.

In some cases, remotely sensed images may contain shadows, which could be due to the time of the day when these images were captured. For example, shadows may be more noticeable on images taken in the early morning or late afternoon when the position of the sun is lower. Furthermore, the presence of steep slopes or tree/plant canopies can cast shadows on their surrounding areas. In such cases, including these images in the training data is important to help improve the ability of the model to handle future images containing shadows. This way, the model can learn to differentiate shadowed and non-shadowed areas. With the correct training data, shadows can be classified as a separate class for pixel-based classification (as shown in Chapter 4) or as part of the background for binary semantic segmentation (as shown in Chapter 3).

The DSMs used in this thesis were not normalised, i.e., the actual pixel values representing elevation, were used. It is recommended that future studies investigate the potential benefits of including normalised elevation models in the modelling process. By representing the feature/vegetation heights above the ground, these elevation data could be useful in applying models across different areas at varying elevations.

b. Testing of the presented methodologies on images acquired by occupied airborne platforms

This research focused on developing habitat mapping methodologies using high-resolution images acquired by UAVs and commercial satellite platforms. Another important remote sensing platform is an occupied aircraft. An occupied aircraft (or occupied aerial vehicle, as opposed to unoccupied aerial vehicles) is considered to be an intermediate platform between satellite and UAV platforms in terms of the spatial extent that can be covered, flexibility and endurance in data acquisition (Matese et al., 2015). Like UAVs, occupied aircraft can mount a wide range of sensors (e.g., LiDAR, RGB, multi-spectral and hyperspectral cameras) for acquiring submetre spatial resolution data suited for many remote sensing applications, such as land cover mapping (X. Li & Shao, 2014), wildlife monitoring (Groom et al., 2013) and biomass estimation (Q. Chen et al., 2012). Similarly, occupied aircraft remote sensing surveys can be scheduled flexibly, which means that the key phenological stages of vegetation can also be targeted. Additionally, aircraft platforms can cover significantly larger areas than a UAV can cover (a few km² for UAVs vs up to hundreds of km² for occupied aircraft), as they can be operated at higher flying altitudes and for longer distances (Colefax et al., 2018; Dash et al., 2019). Further work is suggested to explore the applicability of the presented mapping methodologies on similar types of datasets (i.e., multi-spectral image and elevation data) obtained from aircraft surveys.

c. Using data acquired by UAV-mounted hyperspectral camera and LiDAR.

A recent technological innovation is a UAV platform that can carry more sophisticated sensors, including hyperspectral cameras and LiDAR (Almeida et al., 2021; Qin et al., 2022; Sankey et al., 2018). These sensors can acquire data with greater detail (higher spectral resolution) than conventional multi-spectral/RGB cameras, allowing for more detailed habitat classification. A hyperspectral camera is a sensor that can capture images of hundreds of narrow and contiguous spectral bands. The information from these bands can allow for detecting subtle spectral differences between vegetation types (Meerdink et al., 2019). Light Detection and Ranging, or LiDAR, is a type of remote sensing that uses laser pulses to measure ranges (variable distances to the Earth's surface), providing information to estimate the structure and height of vegetation (Guo et al., 2017; Yin &

Wang, 2019). This information can improve habitat monitoring in areas like forestry and woodlands. As technologies continue to progress, the methodologies are also expected to be updated accordingly.

d. Further collaboration among remote sensing analysts, ecologists and conservation managers

Habitat mapping using remote sensing and machine learning technologies requires extensive knowledge from domain experts in ecology. This thesis is a product of many dialogues between ecologists, remote sensing and machine learning experts. It shows the significance of open communication and collaboration between different field experts. Ecologists are knowledgeable about the habitats to be mapped, which is crucial for collecting high-quality field data for training and evaluation of models. Remote sensing experts are skilled in acquiring and processing remotely sensed datasets. Machine learning experts specialise in the automated extraction of information from combined remote sensing and field datasets. These experts—typically with backgrounds in geography, engineering, ecology, and environmental science—are available to support the conservation community. They are trained to interpret and analyse remote sensing data using machine learning techniques to answer specific ecological questions and address conservation challenges, such as monitoring habitat loss and assessing ecosystem health. In other words, interdisciplinary collaboration is critical for producing accurate and reliable models and can be considered a key component in most habitat monitoring efforts.

Nevertheless, further collaboration between experts is needed for more specific applications of these mapping technologies in habitat and biodiversity conservation. These include assessing functional plant traits, producing essential biodiversity variables and modelling the impact of natural and anthropogenic factors on habitats.

5.6. References

Abrams, J. F., Vashishtha, A., Wong, S. T., Nguyen, A., Mohamed, A., Wieser, S., Kuiper, A., Wilting, A., & Mukhopadhyay, A. (2019). Habitat-Net: Segmentation

- of habitat images using deep learning. *Ecological Informatics*, 51, 121–128.
<https://doi.org/10.1016/j.ecoinf.2019.01.009>
- Agrillo, E., Filipponi, F., Salvati, R., Pezzarossa, A., & Casella, L. (2023). Modeling approach for coastal dune habitat detection on coastal ecosystems combining very high-resolution UAV imagery and field survey. *Remote Sensing in Ecology and Conservation*, 9(2), 251–267. <https://doi.org/10.1002/rse2.308>
- Almeida, D. R. A. de, Broadbent, E. N., Ferreira, M. P., Meli, P., Zambrano, A. M. A., Gorgens, E. B., Resende, A. F., de Almeida, C. T., do Amaral, C. H., Corte, A. P. D., Silva, C. A., Romanelli, J. P., Prata, G. A., de Almeida Papa, D., Stark, S. C., Valbuena, R., Nelson, B. W., Guillemot, J., Féret, J. B., ... Brancalion, P. H. S. (2021). Monitoring restored tropical forest diversity and structure through UAV-borne hyperspectral and LiDAR fusion. *Remote Sensing of Environment*, 264, 112582. <https://doi.org/10.1016/j.rse.2021.112582>
- Alshehhi, R., Marpu, P. R., Woon, W. L., & Mura, M. D. (2017). Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 139–149. <https://doi.org/10.1016/j.isprsjprs.2017.05.002>
- Alvarez-Vanhard, E., Houet, T., Mony, C., Lecoq, L., & Corpetti, T. (2020). Can UAVs fill the gap between in situ surveys and satellites for habitat mapping? *Remote Sensing of Environment*, 243, 111780. <https://doi.org/10.1016/j.rse.2020.111780>
- Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, 11(3), 138–146. <https://doi.org/10.1890/120150>
- Barrett, B., Raab, C., Cawkwell, F., & Green, S. (2016). Upland vegetation mapping using Random Forests with optical and radar satellite data. *Remote Sensing in Ecology and Conservation*, 2(4), 212–231. <https://doi.org/10.1002/rse2.32>
- Barron, S.J., Daly, O.H. & Hodd, R.L. (2022). *National Survey of Upland Habitats (Phase 5, 2020), Summary Report*. National Parks and Wildlife Service, Department of Housing, Local Government and Heritage, Ireland.
https://irishriverproject.com/wp-content/uploads/2023/04/NSUH20_Summary-

- Casella, E., Rovere, A., Pedroncini, A., Stark, C. P., Casella, M., Ferrari, M., & Firpo, M. (2016). Drones as tools for monitoring beach topography changes in the Ligurian Sea (NW Mediterranean). *Geo-Marine Letters*, 36, 151–163.
<https://doi.org/10.1007/s00367-016-0435-9>
- Chen, B., Yang, Y., Wen, H., Ruan, H., Zhou, Z., Luo, K., & Zhong, F. (2018). High-resolution monitoring of beach topography and its change using unmanned aerial vehicle imagery. *Ocean and Coastal Management*, 160, 103–116.
<https://doi.org/10.1016/j.ocecoaman.2018.04.007>
- Chen, Q., Vaglio Laurin, G., Battles, J. J., & Saah, D. (2012). Integration of airborne LiDAR and vegetation types derived from aerial photography for mapping aboveground live biomass. *Remote Sensing of Environment*, 121, 108–117.
<https://doi.org/10.1016/j.rse.2012.01.021>
- Colefax, A. P., Butcher, P. A., & Kelaher, B. P. (2018). The potential for unmanned aerial vehicles (UAVs) to conduct marine fauna surveys in place of manned aircraft. *ICES Journal of Marine Science*, 75(1), 1–8.
<https://doi.org/10.1093/icesjms/fsx100>
- Cruz, C., McGuinness, K., Perrin, P. M., O’Connell, J., Martin, J. R., & Connolly, J. (2023). Improving the mapping of coastal invasive species using UAV imagery and deep learning. *International Journal of Remote Sensing*, 44(18), 5713–5735.
<https://doi.org/10.1080/01431161.2023.2251186>
- Cruz, C., O’Connell, J., McGuinness, K., Martin, J. R., Perrin, P. M., & Connolly, J. (2023). Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping. *European Journal of Remote Sensing*, 56(1), 2191870.
<https://doi.org/10.1080/22797254.2023.2191870>
- Cruzan, M. B., Weinstein, B. G., Grasty, M. R., Kohrn, B. F., Hendrickson, E. C., Arredondo, T. M., & Thompson, P. G. (2016). Small unmanned aerial vehicles (micro-UAVs, drones) in plant ecology. *Applications in Plant Sciences*, 4(9), 1600041. <https://doi.org/10.3732/apps.1600041>

- Dabski, M., Zmarz, A., Rodzewicz, M., Korczak-Abshire, M., Karsznia, I., Lach, K., Rachlewicz, G., & Chwedorzewska, K. (2020). Mapping glacier forelands based on UAV BVLOS operation in Antarctica. *Remote Sensing*, 12(4), 630. <https://doi.org/10.3390/rs12040630>
- Dash, J. P., Watt, M. S., Paul, T. S. H., Morgenroth, J., & Pearce, G. D. (2019). Early detection of invasive exotic trees using UAV and manned aircraft multispectral and LiDAR Data. *Remote Sensing*, 11(15), 1812. <https://doi.org/10.3390/rs11151812>
- Duffy, J. P., Cunliffe, A. M., DeBell, L., Sandbrook, C., Wich, S. A., Shutler, J. D., Myers-Smith, I. H., Varela, M. R., & Anderson, K. (2018). Location, location, location: considerations when using lightweight drones in challenging environments. *Remote Sensing in Ecology and Conservation*, 4(1), 7–19. <https://doi.org/10.1002/rse2.58>
- Fraser, R. H., Olthof, I., Lantz, T. C., & Schmitt, C. (2016). UAV photogrammetry for mapping vegetation in the low-Arctic. *Arctic Science*, 2(3), 79–102. <https://doi.org/10.1139/as-2016-0008>
- Ge, S., Zhang, J., Pan, Y., Yang, Z., & Zhu, S. (2021). Transferable deep learning model based on the phenological matching principle for mapping crop extent. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102451. <https://doi.org/10.1016/j.jag.2021.102451>
- Gonçalves, J. A., & Henriques, R. (2015). UAV photogrammetry for topographic monitoring of coastal areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 101–111. <https://doi.org/10.1016/j.isprsjprs.2015.02.009>
- Groom, G., Stjernholm, M., Nielsen, R. D., Fleetwood, A., & Petersen, I. K. (2013). Remote sensing image data and automated analysis to describe marine bird distributions and abundances. *Ecological Informatics*, 14, 2–8. <https://doi.org/10.1016/j.ecoinf.2012.12.001>
- Guo, X., Coops, N. C., Tompalski, P., Nielsen, S. E., Bater, C. W., & John Stadt, J. (2017). Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data. *Ecological Informatics*, 38, 50–61. <https://doi.org/10.1016/j.ecoinf.2017.01.005>

- Hamylton, S. M., Morris, R. H., Carvalho, R. C., Roder, N., Barlow, P., Mills, K., & Wang, L. (2020). Evaluating techniques for mapping island vegetation from unmanned aerial vehicle (UAV) images: Pixel classification, visual interpretation and machine learning approaches. *International Journal of Applied Earth Observation and Geoinformation*, 89, 102085.
<https://doi.org/10.1016/j.jag.2020.102085>
- Hengl, T. (2006). Finding the right pixel size. *Computers and Geosciences*, 32(9), 1283–1298. <https://doi.org/10.1016/j.cageo.2005.11.008>
- Impollonia, G., Croci, M., Martani, E., Ferrarini, A., Kam, J., Trindade, L. M., Clifton-Brown, J., & Amaducci, S. (2022). Moisture content estimation and senescence phenotyping of novel *Miscanthus* hybrids combining UAV-based remote sensing and machine learning. *GCB Bioenergy*, 14(6), 639–656.
<https://doi.org/10.1111/gcbb.12930>
- Ji, S., Wei, S., & Lu, M. (2019). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574–586.
<https://doi.org/10.1109/TGRS.2018.2858817>
- Kattenborn, T., Eichel, J., & Fassnacht, F. E. (2019). Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific Reports*, 9, 17656.
<https://doi.org/10.1038/s41598-019-53797-9>
- Li, X., & Shao, G. (2014). Object-based land-cover mapping with high resolution aerial photography at a county scale in midwestern USA. *Remote Sensing*, 6(11), 11372–11390. <https://doi.org/10.3390/rs61111372>
- Li, Y., Chen, W., Huang, X., Gao, Z., Li, S., He, T., & Zhang, Y. (2023). MFVNet: a deep adaptive fusion network with multiple field-of-views for remote sensing image semantic segmentation. *Science China Information Sciences*, 66(4), 140305.
<https://doi.org/10.1007/s11432-022-3599-y>
- Li, Z., Xin, Q., Sun, Y., & Cao, M. (2021). A deep learning-based framework for automated extraction of building footprint polygons from very high-resolution

- aerial imagery. *Remote Sensing*, 13(18), 3630. <https://doi.org/10.3390/rs13183630>
- Madurapperuma, B., Close, P., Fleming, S., Collin, M., Thuresson, K., Lamping, J., Dellysse, J., & Cortenbach, J. (2018). Habitat mapping of Ma-le'l dunes coupling with UAV and NAIP imagery. *Proceedings*, 2(7), 368. <https://doi.org/10.3390/ecrs-2-05182>
- Manfreda, S., McCabe, M. F., Miller, P. E., Lucas, R., Madrigal, V. P., Mallinis, G., Dor, E. Ben, Helman, D., Estes, L., Ciruolo, G., Müllerová, J., Tauro, F., de Lima, M. I., de Lima, J. L. M. P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., ... Toth, B. (2018). On the use of unmanned aerial systems for environmental monitoring. *Remote Sensing*, 10(4), 641. <https://doi.org/10.3390/rs10040641>
- Mao, P., Ding, J., Jiang, B., Qin, L., & Qiu, G. Y. (2022). How can UAV bridge the gap between ground and satellite observations for quantifying the biomass of desert shrub community? *ISPRS Journal of Photogrammetry and Remote Sensing*, 192, 361–376. <https://doi.org/10.1016/j.isprsjprs.2022.08.021>
- Martin, J. R., O'Neill, F. H., & Daly, O. H. (2018). *The monitoring and assessment of three E.U. Habitats Directive Annex I grassland habitats (Irish Wildlife Manuals, No. 102)*. National Parks and Wildlife Service, Department of Culture, Heritage and the Gaeltacht, Ireland.
- Marzioletti, F., Giulio, S., Malavasi, M., Sperandii, M. G., Acosta, A. T. R., & Carranza, M. L. (2019). Capturing coastal dune natural vegetation types using a phenology-based mapping approach: The potential of Sentinel-2. *Remote Sensing*, 11(12), 1506. <https://doi.org/10.3390/rs11121506>
- Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., & Gioli, B. (2015). Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sensing*, 7(3), 2971–2990. <https://doi.org/10.3390/rs70302971>
- Meerdink, S. K., Roberts, D. A., Roth, K. L., King, J. Y., Gader, P. D., & Koltunov, A. (2019). Classifying California plant species temporally using airborne hyperspectral imagery. *Remote Sensing of Environment*, 232, 111308. <https://doi.org/10.1016/j.rse.2019.111308>

- O'Neill, F. H., Perrin, P. M., Denyer, J., Martin, J. R., Brophy, J. T., & Daly, O. H. (2023). *Scoping study and pilot survey of fens (Irish Wildlife Manuals No. 143)*. National Parks and Wildlife Service, Department of Housing, Local Government and Heritage, Ireland.
<https://www.npws.ie/sites/default/files/publications/pdf/IWM143.pdf>
- Parente, L., Taquary, E., Silva, A. P., Souza, C., & Ferreira, L. (2019). Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. *Remote Sensing*, 11(23), 2881. <https://doi.org/10.3390/rs11232881>
- Pyšek, P., & Richardson, D. M. (2010). Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources*, 35, 25–55. <https://doi.org/10.1146/annurev-environ-033009-095548>
- Qin, H., Zhou, W., Yao, Y., & Wang, W. (2022). Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data. *Remote Sensing of Environment*, 280, 113143. <https://doi.org/10.1016/j.rse.2022.113143>
- Sankey, T. T., McVay, J., Swetnam, T. L., McClaran, M. P., Heilman, P., & Nichols, M. (2018). UAV hyperspectral and LiDAR data and their fusion for arid and semi-arid land vegetation monitoring. *Remote Sensing in Ecology and Conservation*, 4(1), 20–33. <https://doi.org/10.1002/rse2.44>
- Scarelli, F. M., Sistilli, F., Fabbri, S., Cantelli, L., Barboza, E. G., & Gabbianelli, G. (2017). Seasonal dune and beach monitoring using photogrammetry from UAV surveys to apply in the ICZM on the Ravenna coast (Emilia-Romagna, Italy). *Remote Sensing Applications: Society and Environment*, 7, 27–39.
<https://doi.org/10.1016/j.rsase.2017.06.003>
- Sparrow, B. D., Edwards, W., Munroe, S. E. M., Wardle, G. M., Guerin, G. R., Bastin, J.-F., Morris, B., Christensen, R., Phinn, S., & Lowe, A. J. (2020). Effective ecosystem monitoring requires a multi-scaled approach. *Biological Reviews*, 95(6), 1706–1719. <https://doi.org/10.1111/brv.12636>
- Suo, C., McGovern, E., & Gilmer, A. (2019). Coastal dune vegetation mapping using a multispectral sensor mounted on an UAS. *Remote Sensing*, 11(15), 1814.

<https://doi.org/10.3390/rs11151814>

- Tuanmu, M. N., Viña, A., Roloff, G. J., Liu, W., Ouyang, Z., Zhang, H., & Liu, J. (2011). Temporal transferability of wildlife habitat models: Implications for habitat monitoring. *Journal of Biogeography*, 38(8), 1510–1523. <https://doi.org/10.1111/j.1365-2699.2011.02479.x>
- van der Sluijs, J., Saitet, E., Bakelaar, C., Wentworth, A., Fraser, R., & Kokelj, S. (2023). Beyond visual-line-of-sight (BVLOS) drone operations for environmental and infrastructure monitoring: a case study in northwestern Canada. *Drone Systems and Applications*, 11, 1–15. <https://doi.org/10.1139/dsa-2023-0012>
- Vanden Borre, J., Haest, B., Lang, S., Spanhove, T., Förster, M., & Sifakis, N. I. (2011). Towards a wider uptake of remote sensing in Natura 2000 monitoring: Streamlining remote sensing products with users' needs and expectations. *2nd International Conference on Space Technology, ICST 2011*. <https://doi.org/10.1109/ICSpT.2011.6064686>
- Vanden Borre, J., Paelinckx, D., Múcher, C., Kooistra, L., Haest, B., De Blust, G., & Schmidt, A. (2011). Integrating remote sensing in Natura 2000 habitat monitoring: Prospects on the way forward. *Journal for Nature Conservation*, 19(2), 116–125. <https://doi.org/10.1016/j.jnc.2010.07.003>
- Wang, L., Wang, J., Liu, Z., Zhu, J., & Qin, F. (2022). Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. *Crop Journal*, 10(5), 1435–1451. <https://doi.org/10.1016/j.cj.2022.01.009>
- Wang, Y., Yang, Z., Gert, K., & Khan, H. A. (2023). The impact of variable illumination on vegetation indices and evaluation of illumination correction methods on chlorophyll content estimation using UAV imagery. *Plant Methods*, 19(1), 51. <https://doi.org/10.1186/s13007-023-01028-8>
- Whitehead, K., & Hugenholtz, C. H. (2014). Remote sensing of the environment with small Unmanned Aircraft Systems (UASs), part 1: a review of progress and challenges. *Journal of Unmanned Vehicle Systems*, 2, 69–85. <https://doi.org/10.1139/juvs-2014-0007>

- Xian, G., Homer, C., & Aldridge, C. (2012). Assessing long-term variations in sagebrush habitat - characterization of spatial extents and distribution patterns using multi-temporal satellite remote-sensing data. *International Journal of Remote Sensing*, 33(7), 2034–2058.
<https://doi.org/10.1080/01431161.2011.605085>
- Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, 2017, 1353691.
<https://doi.org/10.1155/2017/1353691>
- Yang, Q., Shi, L., Han, J., Zha, Y., & Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*, 235, 142–153.
<https://doi.org/10.1016/j.fcr.2019.02.022>
- Yi, S. (2017). FragMAP: a tool for long-term and cooperative monitoring and analysis of small-scale habitat fragmentation using an unmanned aerial vehicle. *International Journal of Remote Sensing*, 38, 2686–2697.
<https://doi.org/10.1080/01431161.2016.1253898>
- Yin, D., & Wang, L. (2019). Individual mangrove tree measurement using UAV-based LiDAR data: Possibilities and challenges. *Remote Sensing of Environment*, 223, 34–49. <https://doi.org/10.1016/j.rse.2018.12.034>
- Zmarz, A., Karlsen, S. R., Kycko, M., Korczak-Abshire, M., Gołębiowska, I., Karsznia, I., & Chwedorzewska, K. (2023). BVLOS UAV missions for vegetation mapping in maritime Antarctic. *Frontiers in Environmental Science*, 11, 1154115.
<https://doi.org/10.3389/fenvs.2023.1154115>
- Zmarz, A., Rodzewicz, M., Dąbski, M., Karsznia, I., Korczak-Abshire, M., & Chwedorzewska, K. J. (2018). Application of UAV BVLOS remote sensing data for multi-faceted analysis of Antarctic ecosystem. *Remote Sensing of Environment*, 217, 375–388. <https://doi.org/10.1016/j.rse.2018.08.031>

6. Appendices

Appendix A: Guidelines and best practices in data collection

This section contains best practices and guidelines used in data collection for this thesis. A selection of photographs taken during fieldwork is also presented in Appendix A4.

Appendix A1: UAV survey protocols and best practices

The following outlines general protocols and best practices for UAV surveys, with the aim of capturing high-quality data for habitat and vegetation mapping.

1. Flight parameters and sensors

- The survey spatial extent, the sensor (i.e., camera) and required image resolution determine the flight altitude above ground level and the flight time. Multiple flights are sometimes needed to cover the survey site due to the UAV battery capacity. Most consumer and professional UAV systems have different flight times, ranging between 20 to 40 minutes.
- Setting a minimum of 75% image overlap for both side and front overlaps is recommended. The side overlap corresponds to the overlap between two images from adjacent flight lines, while the front overlap corresponds to the overlap between two successive images along a flight line.
- For most vegetation-related mapping applications, the use of multispectral sensors (e.g., MicaSense Altum and DJI P4 multispectral) is recommended. These sensors have a near-infrared band, which is useful for identifying plant distribution, abundance or health. Including a red-edge band can also improve vegetation analysis.

2. Weather conditions and schedule of the flight

The weather and time of the day in which images are acquired can affect the quality of the captured images. For example, fog or low light conditions can result in poor-quality images (Figure A.1). Similarly, taking images around sunrise and sunset times can lead to images with shadows. Therefore, checking the weather forecast and assessing prevailing weather conditions an hour before the flight is recommended.



Figure A.1. Heavy fog in Magharees, Kerry. The UAV operator decided to postpone the UAV survey.

- It is important to ensure that the wind speed during a UAV survey does not exceed the wind resistance rating of the UAV being used. For instance, the wind speed should not exceed 10 m/s for DJI Phantom 4 Pro and 12 m/s for DJI Matrice 300/350 RTK. If the wind speed exceeds the maximum wind resistance, the UAV survey needs to be postponed. The UAV operator will decide whether to proceed with the survey or not. High winds can cause movement of ground features, particularly vegetation, which can lead to issues during image matching in post-processing. Additionally, high winds can increase the risk of crashing the UAV and cause the batteries to drain faster than expected if flying against the wind.
- To ensure high-quality images from UAV flights, it is recommended that the flights be conducted within 2 to 3 hours of solar noon (i.e., when the sun reaches its maximum altitude). This will help avoid shadows on the images and provide near-nadir illumination. UAV survey dates for coastal habitats (e.g., dunes and salt marshes) should be chosen so that the low tide is close to solar noon.

3. Geometric correction

Geometric correction is a vital step in UAV image processing. This step aims to correct images for any positional errors and geometric distortions that may have occurred during the flight. Two methods are commonly used for this correction: (1) direct georeferencing using onboard positioning sensors—Global Positioning System and Inertial Measurement Unit—in UAVs, and (2) using Ground Control Points (GCPs) measured before or after the flight. Direct georeferencing involves estimating the positions of the images using the onboard sensors. This method is useful for dangerous or inaccessible areas where GCPs cannot be deployed on the ground. On the other hand, conducting a GCP survey is recommended if the project aims for a time-series analysis of a particular site (e.g., monitoring vegetation growth). The following points outline best practices for using GCPs.

- GCPs, or any large markers on the ground used to spatially reference the images, must be clearly marked on the ground and be visually distinct and identifiable in the image. A checkerboard pattern or alternating black and white squares can be used because of its high contrast patterns (Figure A.2). The recommended minimum size of these checkerboard pattern GCPs is six times the spatial resolution of the image. Natural points like road corners can also be used as GCPs, but avoid using points on the riverbank or a non-paved road, as they can move over time. It is not recommended to use spray-painted symbols as an alternative for any site covered with vegetation.

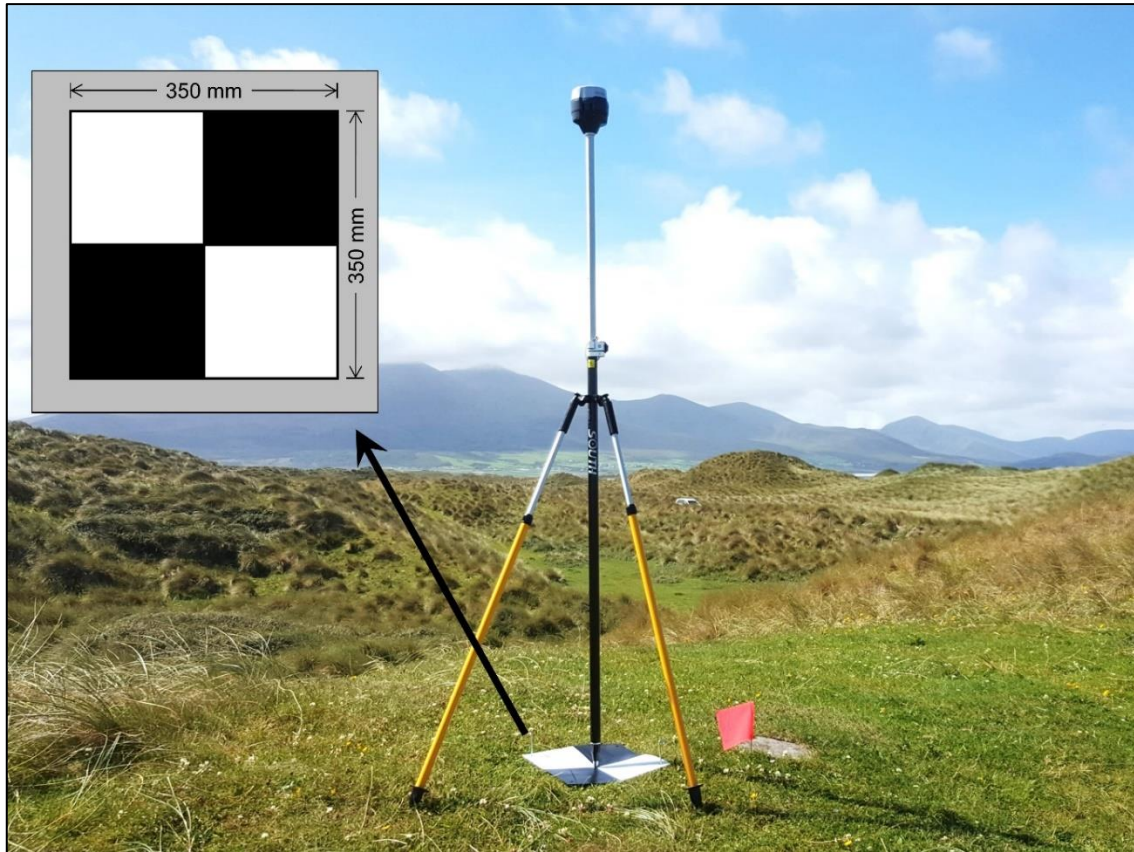


Figure A.2. Measurement of the coordinates of a GCP target using a GNSS device.

- Place a minimum of five GCPs throughout the survey area before the flight starts. These GCPs need to be located on relatively flat terrain and distributed evenly, emphasising the boundaries of the study area rather than the centre. More GCPs can be added for sites with varying topography. To fix the GCPs in place, use steel pegs and place a small flag beside each one to make it easier to locate after the survey. It is not necessary for the GCPs to be in the exact same location for each survey date.
- The coordinates at the centre of each GCP marker are measured and recorded using a GNSS receiver before or after the UAV flight. The observation time for each GCP is set to two minutes. Surveying can be done using either the Real-Time Kinematic (RTK) or Post-Processed Kinematic (PPK) method. In the RTK method, the GNSS receiver can be linked to the closest Ordnance Survey Ireland (OSI) reference station, which can be found here: osi.ie/services/geodetic-services/active-gnss-data/. Corrections are applied to the GNSS observations in real-time. In the PPK method, the GNSS receiver is not connected to any reference

station, and the corrections are applied after the survey. The data to be used for correction can be downloaded from the OSI website. Detailed steps to process PPK-based data are presented in Appendix A3. To avoid confusion across survey dates, follow a logical GCP naming convention. For example, MAH1_1 corresponds to the first three letters of the name of the survey site, followed by the number of the current survey and the GCP number.

- Only collect GCPs once the UAV flight has been completed to ensure that the captured images contain all the GCPs.

4. Radiometric calibration

When analysing data from a single UAV survey, the use of Digital Numbers (DNs) is usually sufficient. DNs are pixel brightness values that range from 0 to 255. However, when analysing data from multiple UAV surveys, radiometric calibration becomes necessary, especially if the data will be used to compare and track changes over time. Radiometric calibration is the process of converting DNs into reflectance values.

- Calibrated reflectance panels should be placed on a flat surface and not affected by shadows or any reflection from the surrounding objects.
- Using the UAV camera, take images of the panels immediately before and after the flight. It is crucial to ensure that the lighting conditions in the field remain consistent during the flight.
- If calibration images are to be taken during the flight, the panels should be large enough to be visible from the air and cover multiple pixels in the image.

Reference:

Assmann, J. J., Kerby, J. T., Cunliffe, A. M., & Myers-Smith, I. H. (2019). Vegetation monitoring using multispectral sensors — Best practices and lessons learned from high latitudes. *Journal of Unmanned Vehicle Systems*, 7(1), 54–75. <https://doi.org/10.1139/juvs-2018-0018>

Appendix A2: Field data collection protocols and best practices

Field surveys are considered the main method for habitat mapping, as they can provide comprehensive information about the habitats. This in-depth ecological knowledge of the study sites is vital to the success of this research. However, there are factors that can affect the accuracy of the field survey methods. These include the positioning accuracy of GPS when recording the habitat location, the experience of the surveyor (i.e., an ecologist), human error in recording habitat information and the complexity of the habitats. For this research, these factors are considered during the planning phase, and efforts are made to minimise the errors related to them.

1. Survey schedule

- Ecological field data will be collected after UAV surveys to avoid disturbance on the site. The aim is to collect field data after one week of the UAV survey.

2. GNSS positioning accuracy

- Ecologists use GNSS receivers (e.g., Emlid Reach RS+) during the survey (Figure A.3). These receivers can provide more accurate positions, ranging from centimetres to two metres, compared to conventional handheld GPS devices with an accuracy range of three to ten metres.



Figure A.3. Recording the location of an upland habitat using a GNSS receiver.

- To ensure consistency among ecologists, a manual on how to use the GNSS receivers was prepared and included in Appendix A3.
 - Post-field survey, the ecologists and the remote sensing analyst must review the collected training samples to ensure conformity with the correct habitats as seen in the UAV image.
3. Ecologists to undertake the surveys.
- Trained and experienced ecologists should collect the data.
 - The same ecologists should survey a particular site if multiple visits are required to avoid inconsistencies by different ecologists.
4. Data transcription
- The aim is to measure the coordinates of a point, which is located at the centre of an area of relatively homogeneous habitat cover, and then record the radius of this area around each point.
 - All collected data should include the Annex I, Fossitt and Irish Vegetation Classification codes.
 - Each habitat recorded in the field is photographed in the north-view direction. The photos serve as a guide in reviewing if the correct habitat names/codes are recorded.
5. Habitat complexity
- Ecologists can use ERICA, an Irish vegetation classification tool. ERICA is designed to help users identify the community based on their vegetation data (e.g., species present, percentage). The link to access the website is biodiversityireland.shinyapps.io/vegetation-classification.
 - Subsequent site visits throughout the growing season can be conducted to allow the ecologists to confirm the habitats identified from the previous survey. Any changes that occurred in the site during these visits must be recorded.
 - Full botanical 2 × 2 vegetation plots will also be collected to define the most common habitats (Figure A.4). All plant species within the plot will be recorded, including their corresponding percentage cover. The coordinates of each corner of the plot will be recorded with a GNSS receiver and will be labelled from 'a' to 'd' in a clockwise direction.



Figure A.4. Recording botanical information in a 2×2 vegetation plot.

Appendix A3: Manual in using a GNSS receiver.

This guide is prepared based on the specifications and operation of the Reach RS+ receivers. The processing of data will be carried out using RTKlib.

Objectives:

1. To obtain the coordinates of the GCPs/training and validation data points using Reach RS+ receivers by:
 - a. Establishing a local base station, then connecting a rover receiver to the base receiver (RTK); or,
 - b. Collecting points (without connecting to the base), then processing the data after the field survey (PPK)

The following Global Navigation Satellite System (GNSS) positioning techniques will be used in the survey:

1. Static – observation is done to establish a base point in which the corrections in the horizontal position will be based
2. Real-Time Kinematic (RTK) – rover receiver will communicate directly with the base to calculate corrections in a horizontal position. In this method, the base and the rover should be visible to each other through a radio signal medium. An advantage of RTK is that the position is instantaneously corrected, reducing post-processing work. On the other hand, using RTK constricts the users to always maintain communication between base and rover receivers, thus limiting the area of the fieldwork site.
3. Post Processed Kinematic (PPK) - accurate positioning does not happen in real-time; the corrections are applied after data collection

Materials:

- 2 Reach RS+ units
- 1 survey tripod/camera tripod
- 1 survey rod
- 1 tribrach with optical plummet (optional)
- Tape measure
- Ruggedised phones
- ReachView - a web-based app that allows you to take control of the Reach RS+ units
- RTKlib – an open-source post-processing software
 - Windows: <https://files.emlid.com/RTKLIB/rtklib-qt-win-b33.zip>
 - Linux: <https://files.emlid.com/RTKLIB/rtklib-qt-linux-b33.zip>

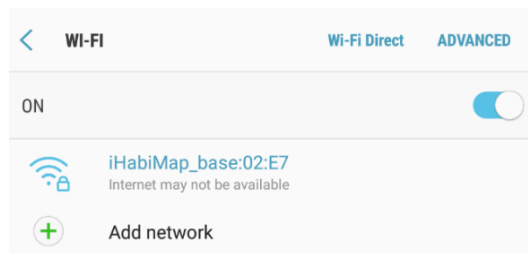
Before fieldwork:

1. Ensure that Reach RS+ units and phones are fully charged.

Base and rover settings

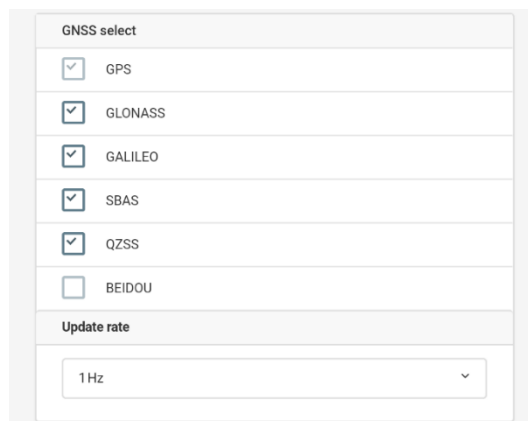
A. Base

1. Connect to the base unit using the ReachView app.

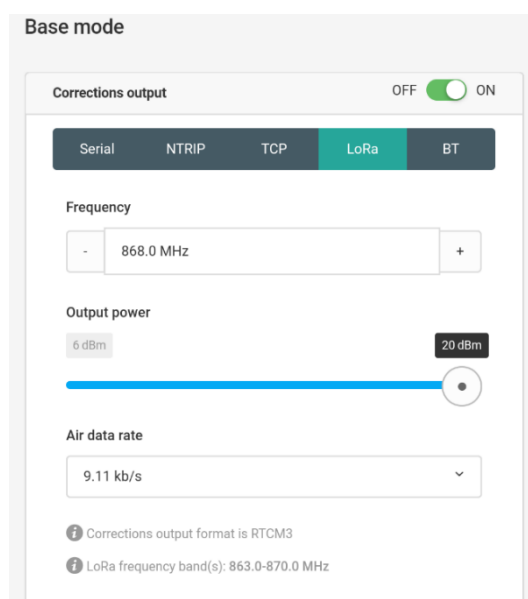


2. Open the **RTK Settings** tab and pick each of the satellite systems. In this case, choose GPS, GLONASS, GALILEO, SBAS and QZSS.

Note: GLONASS covers the world. BEIDOU is more efficient in the Asia Pacific region.

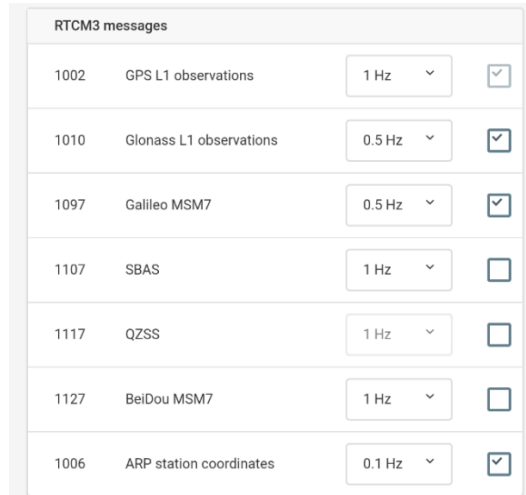


3. Set the **Update rate** at 1 Hz, i.e., data will be updated every second.
4. Go to the **Base mode** tab, and in the **Corrections output** section, select **LoRa**. Set the **Output Power** to 20 dBm and **Air data rate** to 9.11 kb/s



5. In the list of **RTCM3 messages**, select the following parameters:

- GPS L1 observations: 1 Hz
- ARP station coordinates: 0.1 Hz
- GLONASS L1 observation: 1 Hz
- GALILEO or BeiDou: 1 Hz



RTCM3 messages			
1002	GPS L1 observations	1 Hz	☒
1010	Glonass L1 observations	0.5 Hz	☒
1097	Galileo MSM7	0.5 Hz	☒
1107	SBAS	1 Hz	☐
1117	QZSS	1 Hz	☐
1127	BeiDou MSM7	1 Hz	☐
1006	ARP station coordinates	0.1 Hz	☒

6. Click **[Apply]**.

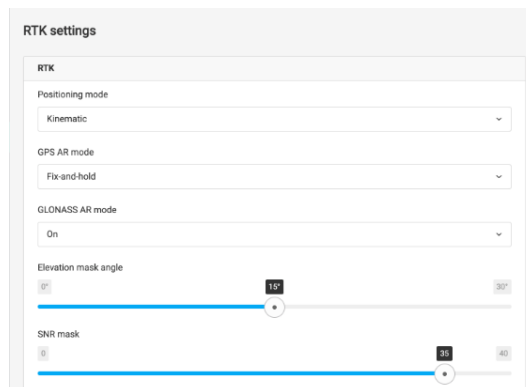
B. Rover

1. Connect to the rover unit.

2. Go to the **RTK settings** tab.

3. Set the following parameters:

- Positioning mode: Kinematic
- GPS Ambiguity resolution mode: Fix-and-hold
- GLONASS Ambiguity resolution mode: ON



RTK settings	
RTK	
Positioning mode	Kinematic
GPS AR mode	Fix-and-hold
GLONASS AR mode	On
Elevation mask angle	15°
SNR mask	30

4. Select the same GNSS systems as the base. Set a 5 Hz update rate and click **[Apply]**.
5. Go to the **Correction input** tab to configure the LoRa radio on the rover unit to receive the corrections.
6. In the **Base correction**, pick **LoRa**.
7. Frequency and air rate settings must match what was configured on the base
8. Click **[Apply]**. The rover is now connected to the base.

Setting the base:

1. Mount the base unit and accurately level the tripod. Consider the following when establishing a base station:
 - a. To have good satellite visibility and to avoid multipath, there should be no obstacles near the antenna that could block the sky view higher than 30 degrees above the horizon. Do not place the set-up near trees, buildings, metal fences, etc.
PDOP: high PDOP - satellites are close together; low PDOP - satellites are widely spaced
- The point should be rugged and difficult to move/disturb
- The point should be easy to re-locate in the future



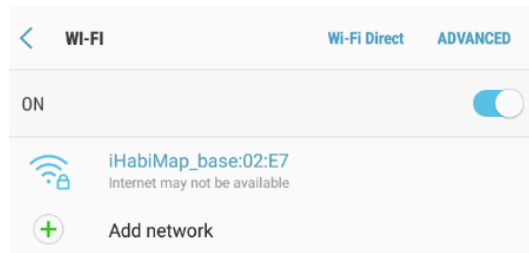
2. Accurately measure the Height of the Instrument (HI) using tape. HI is the distance (in metres) from the survey mark to the base of the receiver. Add 65mm to this measurement to get the total height of the instrument (HI). Take note of this measurement.



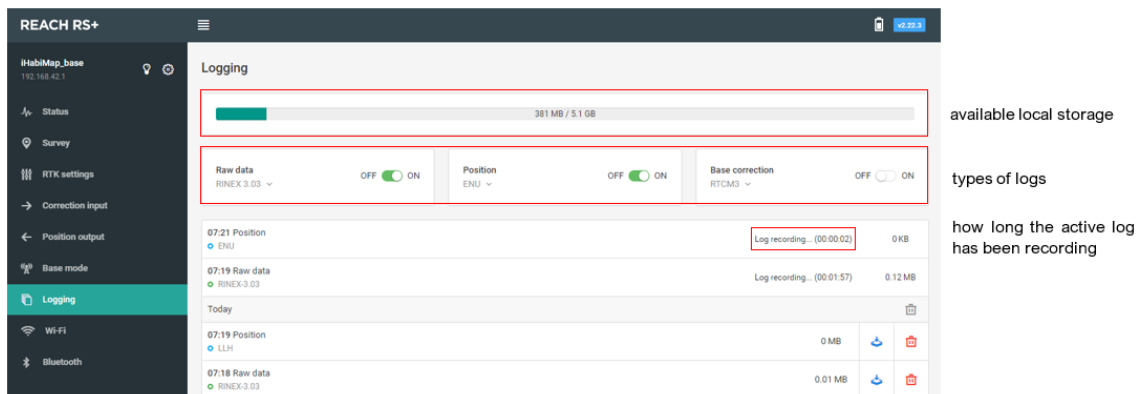
3. Turn on the device by holding the power button for 3 seconds.



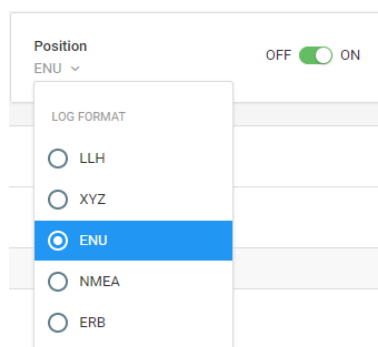
4. Connect to the base unit. The password is “emlidreach”.



5. Once connected to the Reach hotspot, open the ReachView  app.
6. Go to the **Logging** tab in the menu. Enable the logging of **Raw Data**. Select **Rinex 3.03** as the format. Ensure that the screen shows an active log recording.



7. Position can be logged in different formats – LLH, XYZ, ENU, NMEA, ERB. Choose **ENU**.



- LLH is a simple text protocol for Latitude, Longitude, and Height in WGS84
 - XYZ is a simple text protocol for X, Y and Z ECEF coordinates
 - ENU is also a simple text protocol for East, North and UP components of the baseline
 - NMEA 0183 is the most popular standard in the industry. It is usually supported by most software and hardware. NMEA messages supported: GPRMC, GPGGA, GPGSA, GLGSA, GAGSA, GPGSV, GLGSV, GAGSV.
 - ERB format is used for communication with Ardupilot.
8. Log data continuously for 3 – 4 hours (L1 frequency) or 1 – 2 hours (L2 frequency).
 9. Once the survey is completed, download the raw files. Make sure to download the correct files by checking the date and length of the recording.
 10. There are two buttons on the right side of each log; the blue arrow button allows you to save it, and the red garbage can button allows you to delete it.

09 April 2020			
16:24 Raw data		1.86 MB	
UBX			
16:24 Position		0.06 MB	
LLH			

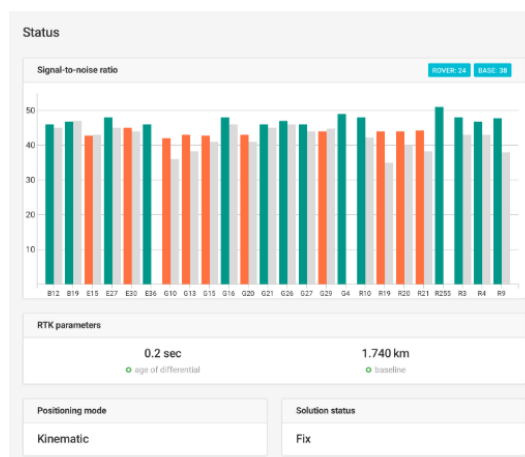
- Go to the **Logging** tab in the menu. Click the  button to download the file. The downloaded file is a zipped folder called <raw_YEARDATE_TIME_Rinex-3_03>.
- Unzip the folder. It should contain three files (.nav, .obs, .sbs). These are the files needed for the post-processing.

Setting the rover

- Mount the other Reach RS+ unit to a survey pole using the brass adaptor. Attach the black LoRa antenna on the bottom of the device.



- Accurately measure the distance from the tip of the survey pole to the base of the RS+ unit and add 65mm. Record this information as the Rover HI (Height of Instrument).
- Turn on the device by holding the power button for 3 seconds.
- Connect to the rover unit. The password is “emlidreach”.
- Once connected to the Reach hotspot, open the ReachView  app.
- Go to the **Status** tab in the menu. If everything is configured correctly, green/orange bars will be accompanied by grey bars. At this point, the rover is connected to the base unit. (Note: No grey bars if there is no base station during the field survey.)



- Go to the **Logging** tab in the menu. Enable the logging of **Raw Data**. Select **Rinex 3.03** as the format. Ensure that the screen shows an active log recording.

The screenshot shows the REACH RS+ app interface. On the left is a sidebar menu with options: Status, Survey, RTK settings, Correction input, Position output, Base mode, **Logging** (highlighted), Wi-Fi, and Bluetooth. The main area is titled 'Logging' and features a storage bar at the top indicating 381 MB / 5.1 GB. Below this are three toggle sections: 'Raw data' (RINEX 3.03) with an 'ON' toggle, 'Position' (ENU) with an 'ON' toggle, and 'Base correction' (RTCM3) with an 'OFF' toggle. A list of logs follows, with the first entry '07:21 Position' (ENU) highlighted and showing 'Log recording... (00:00:02)' and '0 KB'. The second entry '07:19 Raw data' (RINEX 3.03) shows 'Log recording... (00:01:57)' and '0.12 MB'. A 'Today' section lists '07:19 Position' (LLH) at '0 MB' and '07:18 Raw data' (RINEX 3.03) at '0.01 MB'. Red boxes highlight the storage bar, the toggle sections, and the active log recording time.

available local storage

types of logs

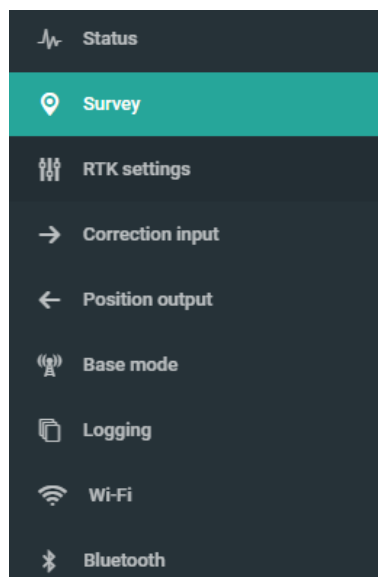
how long the active log has been recording

- Position can be logged in different formats – LLH, XYZ, ENU, NMEA or ERB. Choose **ENU**.

Point collection

A. Creating a new project

- Go to the **Survey** tab in the menu.



- To create a new project, select **New project** or click on the project name to open it. For a new project, indicate the Project name, Author, and Comment. Use the naming convention for the Project name. Click **[Next]**.

Survey - New project
Projects list / New project

Basic information
Step 1/2

Project name: GLE_20200413

Author: Charm

Comment: UAV Survey

10/400

Cancel Next ➔

3. Auto-save rules are set for Single, Float, or Fix statuses, allowing them to automate the point collection process. For this survey, leave these auto-save rules unchecked. Click **[Done]**.

Survey - New project
Projects list / New project

Auto-save rules
Step 2/2

Enable up to three solution statuses acceptable for point collecting. For every status you need to configure auto-save rule.

Optional: Auto-save point if...

- ☐ Solution status FIX
Rule is disabled
- ☐ Solution status FLOAT
Rule is disabled
- ☐ Solution status SINGLE
Rule is disabled

You can skip this step and save points manually

Cancel Back Done

By default, the required observation times are the following:

- Fix 40 sec
- Float 5 min
- Single 10 min

B. Antenna height set-up

1. Click  to open the Pole height menu
2. Enter the pole height value on the next page
3. Click **[Save]**.

C. Point collection

1. To record a point, click  at the bottom of the screen. At the **Record Point** screen, the basic information about the current state of the receiver, including “Solution” status (fixed, float, single, or “-”), PDOP (Position Dilution of Precision), the number of visible satellites, age of differential, Ambiguity Resolution (AR) ratio, and the current baseline length (distance between the rover and base) are included.
 - a. It is recommended to use only FIXED ambiguity solutions

- i. Single: the rover has found a solution relying on its own receiver, and base corrections are not applied. Precision in standalone mode is usually metre-level
 - ii. Float: the base corrections are now taken into consideration
 - iii. Fix: all ambiguities are resolved, and the RTK solution is centimetre-level accurate
- b. Position Dilution of Precision (PDOP) is a unitless measure of the satellite geometry relative to the roving receiver. The lower the PDOP value (i.e., less than 2), the better the geometry. Collecting points when the PDOP value is high is not recommended because bad satellite geometry might affect the accuracy.

Status		
- solution	0 PDOP	R: 0 / B: 0 satellites
0.0 s age	0.0 AR ratio	0.000 m baseline

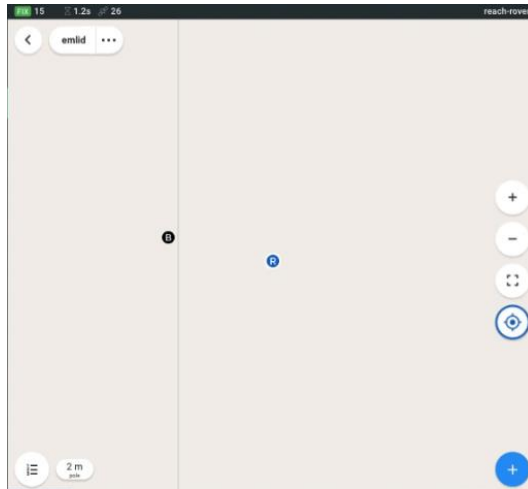
Point name Point 1 Lat: 0.00000000° Lon: 0.00000000° Ht: -2.000 m	Pole height 2 m σ _E = 0.000 m σ _N = 0.000 m σ _U = 0.000 m 
---	---

Waiting for solution status...

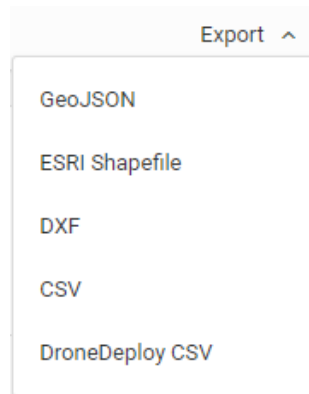
Back
Collect

2. Set the point's name or use the default value. The default point name will be "Point 1", "Point 2", etc..
4. Click [**Collect**] to start the point-collecting process.

Note: If there is a message "Waiting for solution status," the **Collect** button is unavailable, and the point collection won't be able to start. Check the solution status. At least a "Single" solution status is needed to record a point.
5. To stop the point collection, click the **Accept** button.
6. To cancel the point collection, click the **Cancel** button. In this case, the point will not be saved.
7. The Survey tool interface shows the map with the base and rover on it.
 - **B** stands for the base station
 - **R** stands for the rover



8. Repeat the above steps until all the points are recorded.
9. Once the survey is completed, download the files.
 - a. Go to the **Survey** tab and select the project created in the field.
 - b. Click **[Export]** and choose a format. File format options for downloading the project points are GeoJSON, Esri Shapefile, DXF, CSV, or DroneDeploy CSV. Select **CSV**.



File naming convention

1. Project Name
Format: SiteCode_YYYYMMDD

Site Name	Site Code
Liffey Head	LIF
Magharees	MAG
Slieve Mish	SLI

2. Ground Control Points
Format: <SiteCode><NumberOfSurvey>_GCP<GCPNumber>

Example:

Marking GCP1 during the first survey in Magharees: MAG1_GCP1

Field data post-processing

A. OSI Reference Data

1. Go to <https://gnss.osi.ie/download-rinex.php> and download the reference data of the nearest OSI active stations.

**Ordnance
Survey
Ireland**

[Active GNSS Station Data](#)
> Search by coordinates > Search by station > Health by Station > Health by Date

[View GNSS Station Location Map](#)

Find GNSS Data

Select Station:

Select Station 1

Select Station 2

Select Station 3

Select Date:

Select Time:

00:00

01:00

FIND DATA

By clicking Download I agree to the [Terms of Use](#).

To determine the closest stations from the survey site, go to <https://gnss.osi.ie/download-rinex.php?action=searchcoord>. Note that the coordinates are in ITM. Go to <https://gnss.osi.ie/new-converter/> to convert the coordinates.

**Ordnance
Survey
Ireland**

[Active GNSS Station Data](#)
> Search by coordinates > Search by station > Health by Station > Health by Date

[View GNSS Station Location Map](#)

Find GNSS Data by ITM Coordinates

Select Date:

Select Time:

00:00

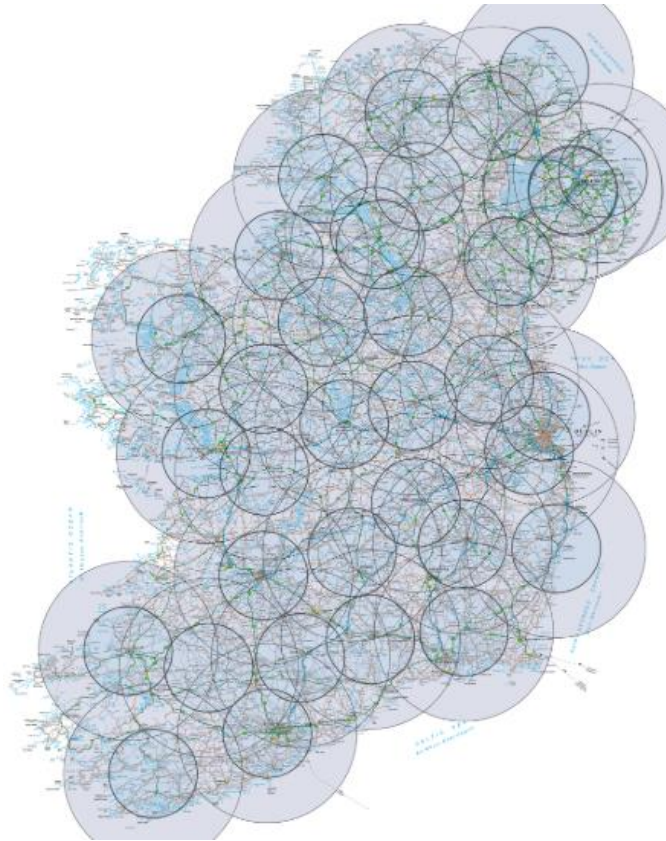
01:00

ITM Easting:

ITM Northing:

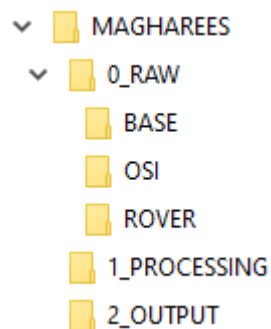
FIND NEAREST STATIONS

By clicking Download I agree to the [Terms of Use](#).



B. Base Station Post Processing (PPK) - optional

1. Ensure that the following files are ready.
 - a. Base station log observation file ending in “.obs”
 - b. Base station log navigation file ending in “.nav”
 - c. OSI reference data file ending in .obs (time range should cover the time covered by the base log)
2. The output folders should be created before the processing. Save these folders in the desired drive. These folders will serve as the directory for the processed files.



3. Run **RTKPost.exe**.

4. Load the files on the main screen using the  icon:
 - a. **RINEX OBS: Rover:** select the base log observation file ending in .obs.
 - b. **RINEX OBS: Base Station:** select the downloaded OSI log observation file ending in .obs
 - c. **RINEX NAV/CLK/, SP3, FCB, IONEX, SBS/EMS, or RTCM:** select the .nav file from the base station log
5. The default output folder and the file name (in .pos) are the same as the file under RINEX OBS: Rover. To change it, browse to the desired output directory by clicking the  icon in the **Solution** line or edit the file name directly in the text box.
6. Adjust the processing parameters. Click **[Options]** and change settings on the “Setting 1”, “Setting 2” and “Positions” tabs.

a. Setting 1

Options ? X

Setting1 Setting2 Output Statistics Positions Files Misc

Positioning Mode Static

Frequencies / Filter Type L1 Forward

Elevation Mask (°) / SNR Mask (dBHz) 15 ...

Rec Dynamics / Earth Tides Correction OFF OFF

Ionosphere Correction Broadcast

Troposphere Correction Saastamoinen

Satellite Ephemeris/Clock Precise

☐ Sat PCV ☐ Rec PCV ☐ PhWU ☐ Rej Ed ☐ RAIM FDE ☐ DBCorr

Excluded Satellites (+PRN: Included) G04, G05, G09, R06, R20, G07, G10, G11, G29, R04, R23, G31

☒ GPS ☒ GLO ☒ Galileo ☒ QZSS ☒ SBAS ☒ BeiDou ☐ IRNSS

Ext Opt... Load... Save... OK Cancel

b. Setting 2

Options ? X

Setting1 Setting2 Output Statistics Positions Files Misc

Integer Ambiguity Res (GPS/GLO/BDS) Fix and Hold ON OFF

Min Ratio to Fix Ambiguity 3

Min Confidence / Max FCB to Fix Amb 0.1 0.25

Min Lock / Elevation (°) to Fix Amb 0 0

Min Fix / Elevation (°) to Hold Amb 100 0

Outage to Reset Amb/Slip Thres (m) 100 0.050

Max Age of Diff (s) / Sync Solution 60.0 OFF

Reject Threshold of GDOP/Innov (m) 30.0 1000.0

Max # of AR Iter/# of Filter Iter 1 1

☐ Baseline Length Constraint (m) 0.000 0.000

AR filtering ON

Min # of sats to Fix Amb/Hold Amb 4 5

Allow base multi epoch/Residual mode ON ON

Ext Opt... Load... Save... OK Cancel

c. Positions

- Check the Antenna Height and enter the HI of the base recorded in the field.

Options

Setting1 Setting2 Output Statistics Positions Files Misc

Rover

Lat/Lon/Height deg/m

0.000000000 0.000000000 0.0000

☒ Antenna Type (*: Auto) Delta-E/N/U (m)

0.0000 0.0000 0

Base Station

RINEX Header Position

90.000000000 0.000000000 -6335367.6285

☐ Antenna Type (*: Auto) Delta-E/N/U (m)

0.0000 0.0000 0.0000

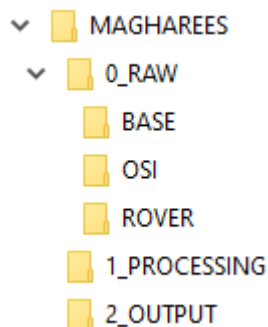
Station Position File

Ext Opt... Load... Save... OK Cancel

7. Click **[OK]**.
8. On the main window, click **[Execute]**. Depending on the options selected and the length of the log files, it may take several minutes for processing to be completed.

C. Post-Processing Rover Logs for PPK Survey

1. Ensure that the following files are ready.
 - d. Base station log observation file ending in “.obs”
 - e. Base station log navigation file ending in “.nav”
 - f. Rover log file ending in .obs
 - g. Precise orbit and clock data ending in .sp3 and .clk
2. The output folders should be created before the processing. Save these folders in the desired drive. These folders will serve as the directory for the processed files.



3. Run **RTKPost.exe**.

4. Load the files on the main screen using the  icon:
 - d. **RINEX OBS: Rover:** select the rover log observation file ending in .obs.
 - e. **RINEX OBS: Base Station:** select the base log observation file ending in .obs
 - f. **RINEX NAV/CLK/, SP3, FCB, IONEX, SBS/EMS, or RTCM:** select the .nav file from the base station log
5. The default output folder and file name (in .pos) are the same as the file under RINEX OBS: Rover. To change it, browse to the desired output directory by clicking the  icon in the **Solution** line or edit the file name directly in the text box.
6. Adjust the processing parameters. Click **[Options]** and change settings on the “Setting 1”, “Setting 2” and “Positions” tabs.

a. Setting 1

Options

Setting 1 | Setting 2 | Output | Statistics | Positions | Files | Misc

Positioning Mode: Kinematic

Frequencies / Filter Type: L1 | Forward

Elevation Mask (°) / SNR Mask (dBHz): 15 | ...

Rec Dynamics / Earth Tides Correction: OFF | OFF

Ionosphere Correction: Broadcast

Troposphere Correction: Saastamoinen

Satellite Ephemeris/Clock: Precise

☐ Sat PCV ☐ Rec PCV ☐ PhWU ☐ Rej Ed ☐ RAIM FDE ☐ DBCorr

Excluded Satellites (+PRN: Included): G04, G05, G09, R06, R20, G07, G10, G11, G29, R04, R23, G31

☒ GPS ☒ GLO ☒ Galileo ☒ QZSS ☒ SBAS ☒ BeiDou ☐ IRNSS

Ext Opt... Load... Save... OK Cancel

b. Setting 2

Options

Setting 1 | Setting 2 | Output | Statistics | Positions | Files | Misc

Integer Ambiguity Res (GPS/GLO/BDS): Fix and Hold | ON | OFF

Min Ratio to Fix Ambiguity: 3

Min Confidence / Max FCB to Fix Amb: 0.1 | 0.25

Min Lock / Elevation (°) to Fix Amb: 0 | 0

Min Fix / Elevation (°) to Hold Amb: 100 | 0

Outage to Reset Amb/Slip Thres (m): 100 | 0.050

Max Age of Diff (s) / Sync Solution: 60.0 | OFF

Reject Threshold of GDOP/Innov (m): 30.0 | 1000.0

Max # of AR Iter/# of Filter Iter: 1 | 1

☐ Baseline Length Constraint (m): 0.000 | 0.000

AR filtering: ON

Min # of sats to Fix Amb/Hold Amb: 4 | 5

Allow base multi epoch/Residual mode: ON | ON

Ext Opt... Load... Save... OK Cancel

c. Positions

- Check the **Antenna Height** and enter the HI of the base recorded in the field.
- Under the **Base Station**, select **Lat/Long/Height deg/m**. Enter the base station coordinates in the fields below. The assumption here is that the base station has known coordinates.

7. Click **[OK]**.
8. On the main window, click **[Execute]**. Depending on the options selected and the length of the log files, it may take several minutes for processing to be completed.

PROCESSING OF SURVEY POINTS

1. Open the CSV file.

	A	B	C	D	E	F	G	H	I	J	K
1	index	name	longitude	latitude	elevation	collection start	collection end	solution status	lateral rms	sample count	antenna height
2	1	Point 1	-6.262092969	53.37075535	66.49958763	2020-04-02 16:14:15 UTC	2020-04-02 16:14:34 UTC	FIX	0.0801	97	2.065
3	2	Point 2	-6.261981668	53.37061858	65.84237857	2020-04-02 16:15:10 UTC	2020-04-02 16:15:15 UTC	FIX	0.0077	28	2.065
4	3	Point 3	-6.262091542	53.3703599	71.04945455	2020-04-02 16:16:18 UTC	2020-04-02 16:16:23 UTC	FLOAT	0.0801	22	2.065

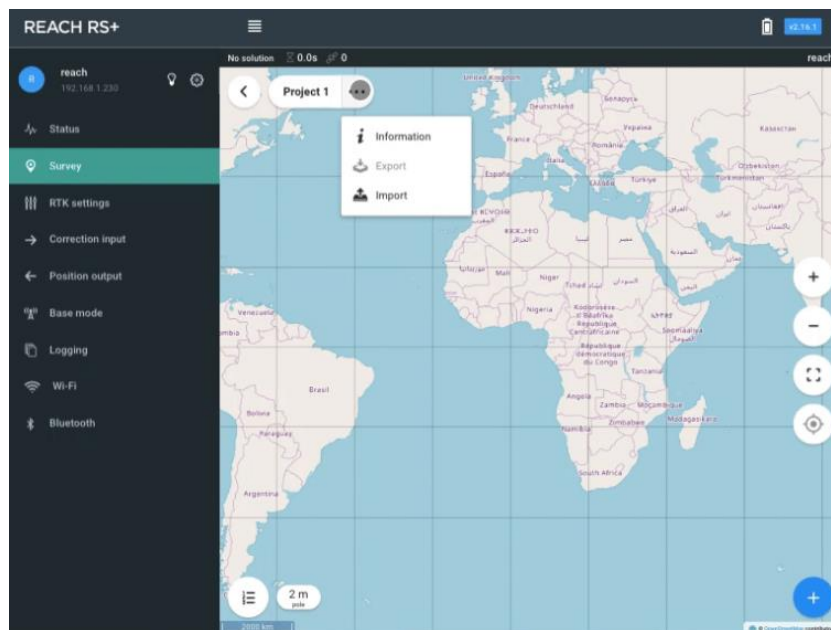
2. Add three new columns in the spreadsheet called “PPKLat”, “PPKLon”, and “PPKHeight”. Copy the latitude, longitude, and height values (columns 3, 4, and 5) from the generated .pos file during the correct time range for the first survey point into the new “PPKLat” “PPK Lon” and “PPKHeight” columns for that point in the .csv file.
3. Repeat step 2 for each record in the CSV file.

Point stakeout

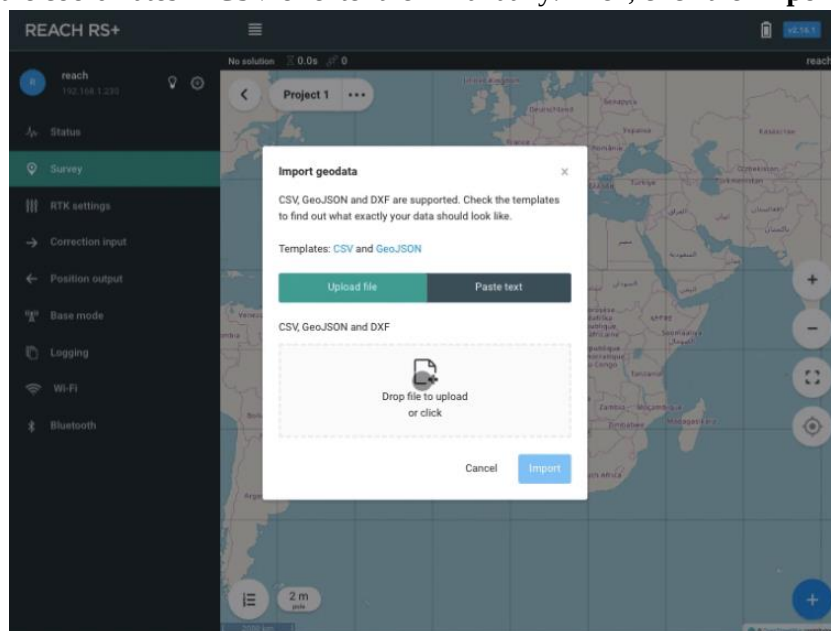
Point stakeout is a feature that allows finding points with known coordinates.

1. Open the **Survey** tab in ReachView.
2. Create a New Project.

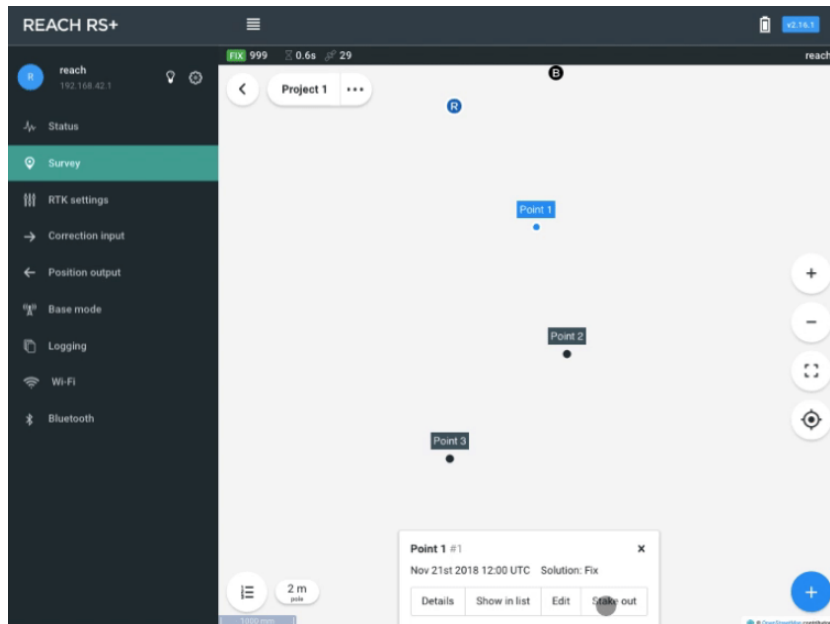
3. Click the **...** button on the right side of the project name and choose **Import** in the drop-down list.



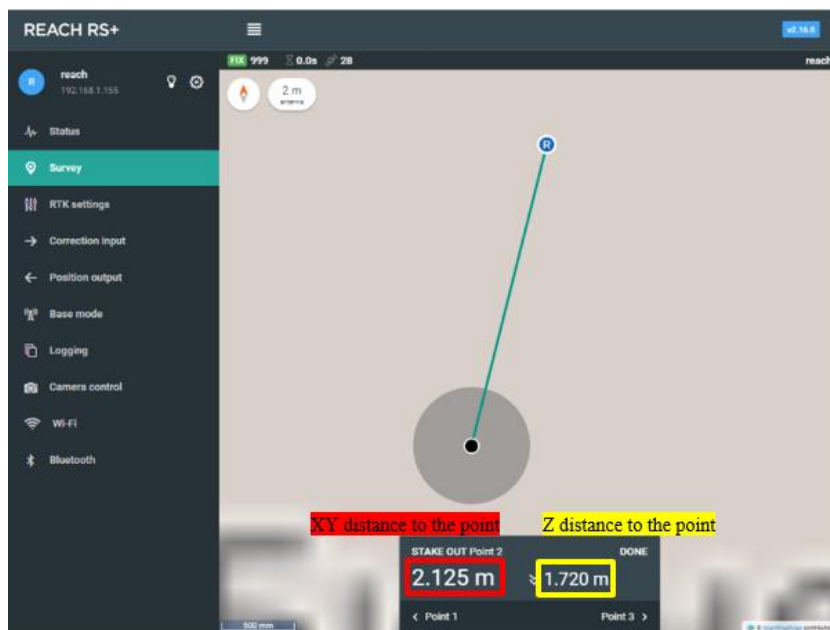
4. Upload the coordinates in CSV or enter them manually. Then, click the **Import** button.



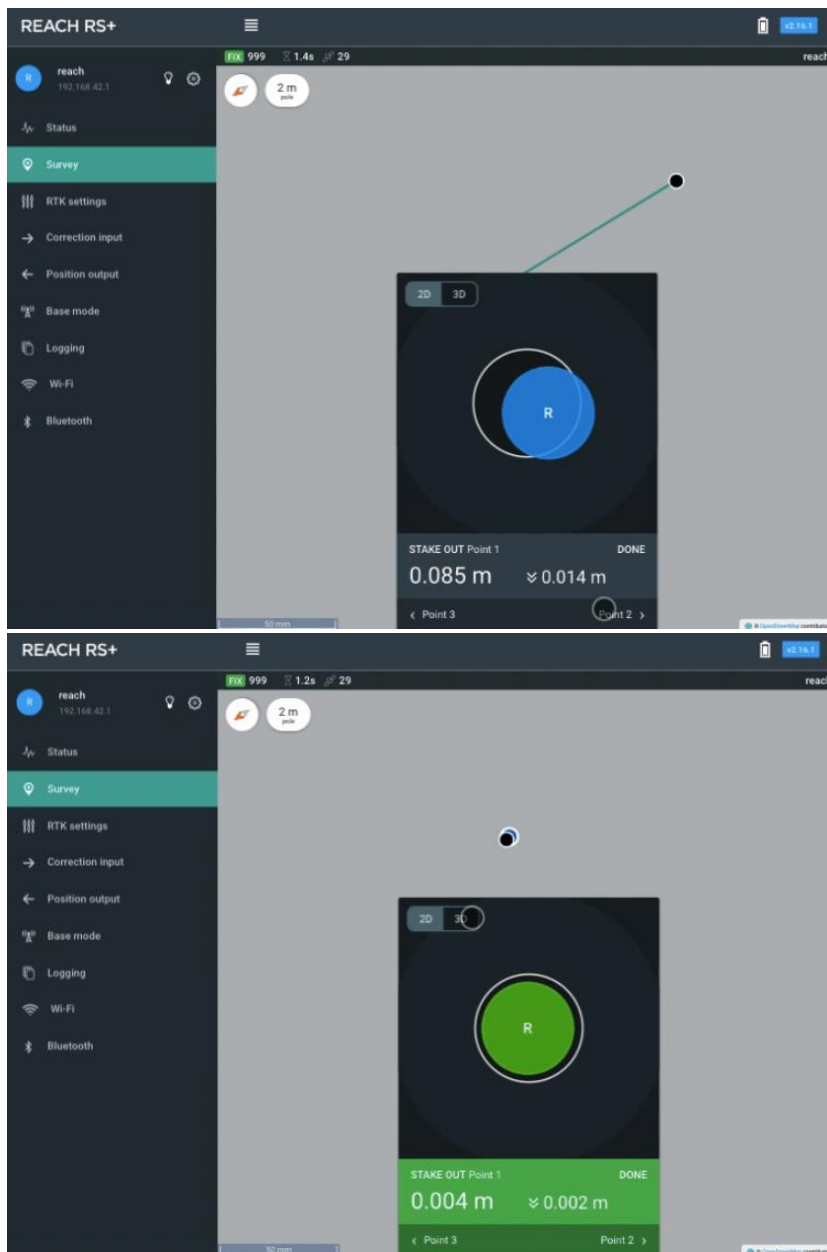
5. Tap the point to stake out. It will be highlighted in blue.
6. Click the **Stakeout** button to start.



7. Try to move in the direction of the point. Note that while walking, the Rover symbol changes to the arrow.



8. The goal is to put the blue rover circle into the white ring. Continue moving the rover to the point until the rover symbol becomes green.



9. Click **[Done]** to finish.

REFERENCES:

<https://docs.emlid.com/reach/>

<https://docs.emlid.com/reachrs/common/reachview/survey/#point-stakeout>

<https://documents.cast.uark.edu/emlidreach/Guide3.html>

<https://documents.cast.uark.edu/emlidreach/Guide5.html>

Appendix A4: A selection of photographs taken during fieldwork.



Figure A.5. DJI Phantom Pro 4 RTK drone and accessories (batteries, remote controller and reference panels).



Figure A.6. DJI Matrice 300 RTK flying over a saltmarsh site in Bull Island, Dublin.



Figure A.7. Flooded dune slack (Annex I habitat code 2190) in the Magharees Co. Kerry.



Figure A.8. Dry heath (Annex I habitat code 4030) in Slieve Mish, Co. Kerry.



Figure A.9. Fixed coastal dunes (Annex I habitat code 2130) in the Magharees Co. Kerry. A priority habitat under the EU Habitats Directive.



Figure A.10. An example of a front dune showing stages of succession (i.e., strandline to mobile to fixed dunes)



Figure A.11. C. Cruz measuring the coordinates of a training point using a GNSS receiver.



Figure A.12. Dense *Spartina anglica* (common cordgrass) plants in a saltmarsh site in Bull Island Dublin.



Figure A.13. A GNSS receiver recording the coordinates of a training point in Slieve Mish, Co. Kerry.

Appendix B: Supplementary materials for Chapter 2

This section contains all the published supplementary materials for Chapter 2. This can also be accessed through the link: <https://doi.org/10.1080/22797254.2023.2191870>.

Appendix B1: Annex I dune habitat descriptions and plot data¹

In addition to the training points described in the paper, 11 quantitative vegetation plots were recorded within the Magharees survey area to allow the Annex I habitats at the site to be described and for comparisons to be made with the descriptions published within NPWS (2019). The data recorded within these 2 × 2 m vegetation plots are presented in Tables B.1 and B.2, wherein characteristic vascular plant species are indicated for each of the Annex I habitats which had plots recorded within the study area. Data from the 11 plots were classified to the Irish Vegetation Classification (IVC) using the ERICA application (Version 5.1, 5th March 2020). Descriptions of the Annex I habitats present follow. The photographs of the plots shown below (Figures B.1-B.6) do not always match the date the vegetation data were recorded as the plots were often photographed over multiple visits.

Annual vegetation of drift lines (habitat code 1210)

This habitat was recorded covering small areas along the front of the Magharees dune system. Three 1 × 1 m training points were recorded and photographed but no 2 × 2 m vegetation plots were recorded. All three points that represented this habitat were characterised by a low cover, usually less than 5%, of *Honckenya peploides* growing

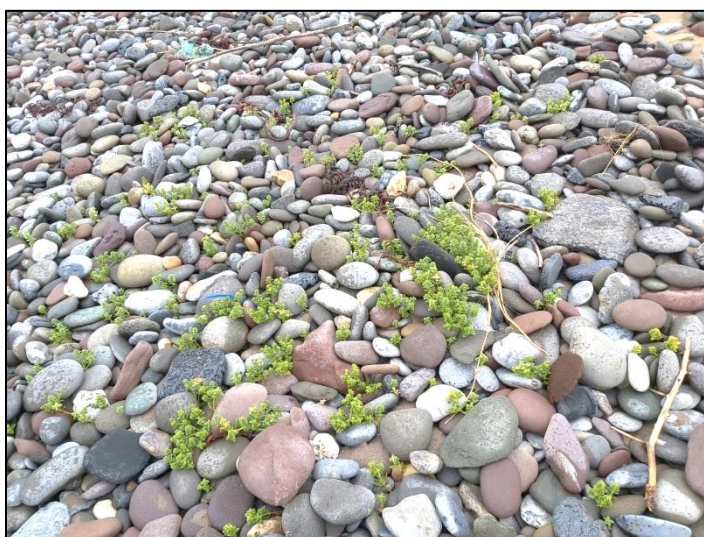


Figure B.1. Training point recorded within the 1210 habitat (Magharees, June 2020)

¹ Appendix B1 was prepared by the co-authors James Martin and Philip Perrin (Cruz et al., 2023).

amongst strandline debris such as driftwood or seaweed. The substrate was usually shingle or a combination of sand and shingle. This vegetation was classified within the IVC as the ST1C *Honckenya peploides* – *Eryngium maritimum* strandline community. As described in NPWS (2019) this habitat exists in a state of instability and may be absent in some years due to natural causes, such as storms, or anthropogenic causes such as leisure activities.

Embryonic shifting dunes (habitat code 2110)

This habitat was recorded covering relatively extensive areas along the front of the Magharees dune system. One plot (M25) was recorded within this Annex I habitat and was classified within the IVC as the DU1A *Elytrigia juncea* duneland community. There was a high cover of bare sand within the plot (70%). *Elytrigia juncea* was the only



Figure B.2. Plot M25 recorded within the 2110 habitat (Magharees, June 2021)

characteristic species for the 2110 habitat at the site and the species was recorded with a relatively high cover within the plot (25%). The only other species recorded within the plot was *Ammophila arenaria*.

Shifting dunes along the shoreline with *Ammophila arenaria* (white dunes) (habitat code 2120)

This habitat was recorded covering extensive areas towards the front of the Magharees dune system and sometimes in disturbed areas farther back in the dunes. One plot (M18)

was recorded within this Annex I habitat. It was classified to the IVC community DU1B *Ammophila arenaria* duneland and was characterised by a high cover of bare sand, a relatively high cover of *Ammophila arenaria* and *Eryngium maritimum*, and a low cover for all other species present within the plot.



Figure B.3. Plot M18 recorded within the 2120 habitat (Magharees, July 2020)

Fixed coastal dunes with herbaceous vegetation (grey dunes) (habitat code 2130)

This was one of the most extensive Annex I habitats recorded within the study area and 4 of the 11 vegetation plots recorded within the Magharees were recorded within the 2130 habitat. Within the site, three distinct IVC communities that correspond to the 2130 habitat were recorded. All were found farther back in the dune system. One of the 2130 plots (M15)



Figure B.4. Plot M11 recorded within the 2130 habitat (Magharees, June 2020)

was recorded within the IVC community DU2A *Ammophila arenaria* – *Festuca rubra* duneland. This represented a tall, rank version of the 2130 habitat with a high cover of both *Ammophila arenaria* and *Festuca rubra* in the plot. No other plant species recorded within the plot was abundant. The cover of bare sand was low. Species richness was relatively high at 22 and notable plant species that were recorded within the plot included *Galium verum*, *Hypochaeris radicata*, *Plantago lanceolata*, *Ranunculus bulbosus*, *Veronica chamaedrys* and *Viola riviniana*.

Two of the 2130 plots (M12 and M13) were recorded within the IVC community DU2B *Festuca rubra* – *Syntrichia ruraliformis* duneland, a less common community within the Magharees study area, that is distinguished by both high bryophyte diversity and cover. Due to the relatively high number of moss species recorded within these plots they were some of the most species-rich recorded within the dune system, with 32 and 30 species respectively. The moss species *Syntrichia ruraliformis* and *Homalothecium lutescens* were recorded within both plots. Notable vascular plant species recorded within the two plots included *Anthyllis vulneraria* and *Thymus polytrichus*.

One of the 2130 plots (M11) was recorded within the IVC community DU2C *Festuca rubra* - *Plantago lanceolata* – *Carex arenaria* duneland. There was no bare sand within the plot which had a relatively short sward and lacked *Ammophila arenaria*. Species richness was 22 and notable plant species that were recorded included *Bromus hordeaceus*, *Cerastium fontanum*, *Festuca rubra*, *Galium verum*, *Plantago lanceolata* and *Senecio jacobaea*. This community tended to occur in flatter and more stable areas of the Magharees dune system that were subject to higher levels of cattle grazing and human impacts such as trampling.

Dunes with *Salix repens* ssp. *argentea* (*Salicion arenariae*) (habitat code 2170)

Two plots (M14 and M20) were recorded within the 2170 habitat within the study area, and both were characterised by a cover of *Salix repens* of 20% or over. No bare sand was recorded in either plot. Both plots were recorded within the IVC community DU3A *Salix repens* – *Lotus corniculatus* duneland. Other notable species that were recorded within both plots included *Centaurea nigra* and *Trifolium pratense*.



Figure B.5. Plot M14 recorded within the 2170 habitat (Magharees, June 2020)

Humid dune slacks (habitat code 2190)

The extensive areas of 2190 habitat within the Magharees are an important conservation feature of the dune system. All areas of 2190 habitat at the site are subject to some flooding due to a combination of heavy rainfall and saturated ground water; the waterlogged condition of the soil is an important determinant of the plant community recorded



Figure B.6. Plot M16 recorded within the 2190 habitat (Magharees, June 2020)

(NPWS, 2019). Two of the 2190 plots (M17 and M23) were recorded within the IVC community DU3B *Agrostis stolonifera* – *Calliergonella cuspidata* – *Carex arenaria* duneland, the most distinguishing character of which was a high cover of *Agrostis stolonifera* in both plots.

The plant communities within some dune slacks correspond more closely to fens than dunes and one of the 2190 plots (M16) corresponded to FE3D *Carex nigra* – *Calliergonella cuspidata* fen. This plot had high covers of *Carex nigra*, *Calliergonella cuspidata* and *Hydrocotyle vulgaris*, and the cover of *Agrostis stolonifera* was also relatively high. The species richness of 16 was relatively low for a dune slack plot.

Table B.1. Vegetation plot data for Annex I habitats 2110, 2120 and 2130. Percentage cover of individual species is shown.

Plot	M25	M18	M15	M12	M13	M11
Annex I code	2110	2120	2130	2130	2130	2130
IVC community code	DU1A	DU1B	DU2A	DU2B	DU2B	DU2C
% affinity to community	100	100	100	72	99	41
Date	29/06/21	06/06/20	04/06/20	04/06/20	04/06/20	04/06/20
% Bare sand	70	65	1	20	25	0
% Bare rock	0	0	0	0	0	0
% Surface water	0	0	0	0	0	0
Field height 1 (cm)	48	53	64	15	48	19
Field height 2 (cm)	45	48	64	16	54	18
Field height 3 (cm)	36	55	72	16	55	16
Field height 4 (cm)	66	59	78	20	58	15
Characteristic species 2110						
<i>Elytrigia juncea</i>	25.0	3.0	0.0	0.0	0.0	0.0
Characteristic species 2120						
<i>Ammophila arenaria</i>	7.0	30.0	50.0	0.5	10.0	0.0
<i>Eryngium maritimum</i>	0.0	10.0	0.0	0.5	0.0	1.0
<i>Euphorbia paralias</i>	0.0	0.5	0.0	0.0	0.0	0.0
Characteristic species 2130						
<i>Anthyllis vulneraria</i>	0.0	0.5	5.0	15.0	15.0	0.0
<i>Asperula cynanchica</i>	0.0	0.0	0.0	1.0	3.0	0.0
<i>Bromus hordeaceus</i>	0.0	0.0	0.5	0.0	0.0	10.0
<i>Carex arenaria</i>	0.0	0.0	0.5	0.5	5.0	1.0
<i>Cerastium fontanum</i>	0.0	0.0	0.5	0.5	0.0	20.0
<i>Festuca rubra</i>	0.0	3.0	40.0	10.0	7.0	1.0
<i>Galium verum</i>	0.0	0.0	3.0	0.5	0.0	30.0
<i>Hypochaeris radicata</i>	0.0	0.0	3.0	0.5	1.0	0.5
<i>Leontodon saxatilis</i>	0.0	0.0	1.0	0.5	0.5	0.5
<i>Lotus corniculatus</i>	0.0	0.0	0.0	0.5	0.0	0.0
<i>Pilosella officinarum</i>	0.0	0.0	1.0	0.0	0.5	0.0

Table B.1 (continued). Vegetation plot data for Annex I habitats 2110, 2120 and 2130

Plot	M25	M18	M15	M12	M13	M11
<i>Plantago lanceolata</i>	0.0	0.0	1.0	0.5	1.0	20.0
<i>Polygala vulgaris</i>	0.0	0.0	0.5	0.5	0.0	0.0
<i>Ranunculus bulbosus</i>	0.0	0.0	0.5	0.0	0.0	0.0
<i>Senecio jacobaea</i>	0.0	0.0	0.5	0.0	1.0	7.0
<i>Thymus polytrichus</i>	0.0	0.0	0.0	25.0	7.0	0.0
<i>Trifolium campestre</i>	0.0	0.0	1.0	3.0	1.0	0.0
<i>Trifolium dubium</i>	0.0	0.0	0.0	0.5	0.0	0.5
<i>Trifolium repens</i>	0.0	0.0	0.0	0.0	0.0	0.5
<i>Veronica chamaedrys</i>	0.0	0.0	0.5	0.0	0.0	0.0
<i>Viola riviniana</i>	0.0	0.0	1.0	0.0	0.0	0.0
<i>Viola tricolor</i> subsp. <i>curtisii</i>	0.0	0.0	0.0	0.0	0.5	0.0
General plant species						
<i>Agrostis capillaris</i>	0.0	0.0	0.0	0.0	0.0	3.0
<i>Aira praecox</i>	0.0	0.0	0.0	0.0	0.5	0.0
<i>Anagallis arvensis</i>	0.0	0.0	0.0	0.0	0.5	0.0
<i>Arenaria serpyllifolia</i>	0.0	0.0	0.0	0.0	0.5	0.5
<i>Bellis perennis</i>	0.0	0.0	0.5	0.5	0.0	0.0
<i>Carex flacca</i>	0.0	0.0	0.0	0.0	0.0	0.5
<i>Catapodium maritimum</i>	0.0	0.0	0.0	0.0	0.5	0.0
<i>Centaureum erythraea</i>	0.0	0.0	0.0	1.0	0.0	0.5
<i>Cerastium diffusum</i>	0.0	0.0	0.0	3.0	1.0	0.0
<i>Cuscuta epithymum</i>	0.0	0.0	0.0	1.0	0.5	0.0
<i>Geranium molle</i>	0.0	0.0	0.0	0.0	0.0	0.5
<i>Holcus lanatus</i>	0.0	0.0	0.5	0.0	0.5	0.0
<i>Phleum arenarium</i>	0.0	0.0	0.0	0.5	0.5	0.0
<i>Poa annua</i>	0.0	0.0	0.0	0.0	0.0	1.0
<i>Prunella vulgaris</i>	0.0	0.0	0.0	0.5	0.0	0.0
<i>Sedum acre</i>	0.0	0.0	0.0	1.0	0.5	0.0
<i>Taraxacum officinale</i> agg.	0.0	0.0	0.5	0.5	0.0	0.5
<i>Trifolium pratense</i>	0.0	0.0	0.0	0.0	0.0	3.0
Bryophytes						
<i>Bryum</i> sp.	0.0	0.0	0.0	0.5	0.5	0.0
<i>Ceratodon purpureus</i>	0.0	0.0	0.0	0.0	0.5	0.0
<i>Homalothecium lutescens</i>	0.0	0.0	3.0	25.0	3.0	3.0
<i>Hypnum lacunosum</i>	0.0	0.0	0.0	1.0	1.0	0.0
<i>Scleropodium purum</i>	0.0	0.0	0.0	0.5	0.5	0.0
<i>Syntrichia ruraliformis</i>	0.0	0.0	0.0	3.0	55.0	0.0
<i>Trichostomum crispulum</i>	0.0	0.0	0.0	1.0	0.5	0.0
Species richness	2	6	22	32	30	22

Table B.2. Vegetation plot data for Annex I habitats 2170 and 2190. Percentage cover of individual species is shown.

Plot	M14	M20	M17	M23	M16
Annex I code	2170	2170	2190	2190	2190
IVC community code	DU3A	DU3A	DU3B	DU3B	FE3D
% affinity to community	84	90	56	91	81
Date	04/06/20	30/07/20	04/06/20	23/09/20	04/06/20
% Bare sand	0.0	0.0	5.0	0.0	0.0
% Bare rock	0.0	0.0	0.0	0.0	0.0
% Surface water	0.0	0.0	0.0	0.0	0.0
Field height 1 (cm)	32.0	41.0	17.0	19.0	23.0
Field height 2 (cm)	53.0	26.0	25.0	19.0	26.0
Field height 3 (cm)	62.0	42.0	21.0	17.0	28.0
Field height 4 (cm)	63.0	32.0	17.0	12.0	29.0
Characteristic species 2170					
<i>Centaurea nigra</i>	10.0	10.0	0.0	0.0	0.0
<i>Salix repens</i>	60.0	20.0	0.0	3.0	15.0
<i>Trifolium pratense</i>	10.0	5.0	0.0	0.0	0.5
Characteristic species 2190					
<i>Agrostis stolonifera</i>	0.0	30.0	45.0	40.0	20.0
<i>Cardamine pratensis</i>	0.0	0.0	0.5	0.0	0.0
<i>Carex nigra</i>	0.0	0.0	1.0	30.0	35.0
<i>Carex panicea</i>	0.0	0.0	20.0	20.0	1.0
<i>Eleocharis palustris</i>	0.0	0.0	0.0	0.0	1.0
<i>Hydrocotyle vulgaris</i>	0.0	0.0	0.5	10.0	40.0
<i>Juncus acutiflorus</i>	0.0	0.0	0.5	0.5	0.5
<i>Juncus bufonius</i>	0.0	0.0	0.5	0.0	0.0
<i>Potentilla anserina</i>	0.0	7.0	1.0	10.0	0.5
<i>Ranunculus flammula</i>	0.0	0.0	0.0	0.0	0.5
<i>Ranunculus repens</i>	0.0	3.0	1.0	15.0	0.5
<i>Bellis perennis</i>	0.0	0.0	1.0	0.0	0.0
<i>Carex arenaria</i>	0.5	0.0	0.5	0.0	0.0
<i>Carex flacca</i>	3.0	30.0	0.0	0.5	0.0
<i>Dactylis glomerata</i>	15.0	0.0	0.0	0.0	0.0
<i>Dactylorhiza</i> sp.	0.0	0.0	0.0	0.0	0.5
<i>Euphrasia officinalis</i> agg.	0.0	0.5	0.5	1.0	0.0
<i>Festuca arundinacea</i>	0.0	0.0	0.0	0.5	0.0
<i>Festuca rubra</i>	5.0	3.0	0.0	0.5	0.0
<i>Galium verum</i>	0.5	0.5	0.0	0.0	0.0
<i>Holcus lanatus</i>	0.5	0.5	0.0	0.0	0.0
<i>Leontodon autumnalis</i>	0.0	0.0	0.5	1.0	0.0
<i>Leontodon saxatilis</i>	0.0	1.0	5.0	0.5	0.0
<i>Linum catharticum</i>	0.0	0.5	0.0	0.5	0.0

Table B.2 (continued). Vegetation plot data for Annex I habitats 2170 and 2190

Plot	M14	M20	M17	M23	M16
<i>Lolium perenne</i>	0.0	0.5	0.0	0.0	0.0
<i>Lotus corniculatus</i>	5.0	25.0	5.0	7.0	0.5
<i>Plantago coronopus</i>	0.0	0.0	0.5	0.0	0.0
<i>Plantago lanceolata</i>	3.0	3.0	3.0	0.0	0.0
<i>Poa annua</i>	0.0	0.0	0.5	0.0	0.0
<i>Poa trivialis</i>	5.0	0.0	0.0	0.0	0.0
<i>Potentilla anglica</i>	0.0	0.0	1.0	0.0	1.0
<i>Prunella vulgaris</i>	0.5	3.0	0.5	0.5	0.0
<i>Ranunculus bulbosus</i>	1.0	0.0	0.0	0.0	0.0
<i>Rosa spinosissima</i>	5.0	0.0	0.0	0.0	0.0
<i>Senecio jacobaea</i>	0.5	0.5	0.5	0.0	0.0
<i>Taraxacum officinale</i> agg.	0.5	0.5	3.0	0.0	0.5
<i>Trifolium repens</i>	1.0	0.5	0.5	5.0	0.0
<i>Veronica chamaedrys</i>	0.5	0.0	0.0	0.0	0.0
<i>Viola riviniana</i>	1.0	0.0	0.0	0.0	0.0
Bryophytes					
<i>Brachythecium mildeanum</i>	0.0	3.0	0.0	0.0	0.0
<i>Brachythecium rivulare</i>	0.0	0.0	5.0	0.0	0.0
<i>Bryum argenteum</i>	0.0	0.0	0.5	0.0	0.0
<i>Bryum</i> sp.	0.0	0.0	0.5	0.0	0.0
<i>Calliergonella cuspidata</i>	0.0	1.0	25.0	55.0	50.0
<i>Fissidens adianthoides</i>	0.0	0.0	0.0	0.5	0.0
<i>Plagiomnium</i> sp.	0.0	0.0	0.5	0.0	0.0
<i>Rhytidiadelphus squarrosus</i>	0.0	1.0	0.5	0.0	0.0
<i>Riccia cavernosa</i>	0.0	0.0	0.5	0.0	0.0
<i>Scleropodium purum</i>	3.0	0.0	0.0	0.0	0.0
Species richness	21	23	30	20	16

Appendix B2: Other supplementary materials

Table B.3. Parameter settings for orthomosaic and DSM generation in Pix4DMapper (ver 4.3.33)

Parameter	Value
Keypoints Image Scale	Full
Matching Image Pairs	Aerial Grid or Corridor
Targeted Number of Keypoints	10000
Calibration Method	Alternative
Point Cloud Densification	1/2 - Optimal - 3
DSM Resolution	1 x GSD
DSM Filters	Use Surface Smoothing - Medium
DSM GeoTIFF	Method - Inverse Distance Weighting

Table B.4. Principal Components (PC) eigenvalues and cumulative eigenvalues obtained from the sets of UAV images acquired in May, July and October.

May

PC bands	Eigenvalues (%)	Cumulative Eigenvalues (%)
PC1	76.363	76.363
PC2	22.395	98.758
PC3	0.675	99.433
PC4	0.444	99.877
PC5	0.123	100.000

July

PC bands	Eigenvalues (%)	Cumulative Eigenvalues (%)
PC1	77.219	77.219
PC2	21.587	98.806
PC3	0.714	99.520
PC4	0.332	99.852
PC5	0.148	100.000

October

PC bands	Eigenvalues (%)	Cumulative Eigenvalues (%)
PC1	89.145	89.145
PC2	9.870	99.015
PC3	0.427	99.442
PC4	0.345	99.787
PC5	0.213	100.000

Table B.5. Variable importance ranked by Mean Decrease in Impurity (MDI). The top five variables from each model are indicated in bold letters. A full list of variable acronyms is provided in Table 2.2.

M_RF _{ST}		J_RF _{ST}		O_RF _{ST}	
Variables	MDI	Variables	MDI	Variables	MDI
Elevation	0.189755	Elevation	0.188144	Elevation	0.241111
NDWI	0.082205	NDWI	0.096405	NDWI	0.118330
Slope	0.074859	GNDVI	0.092190	GNDVI	0.107617
B	0.071734	EVI2	0.073986	Slope	0.078839
GNDVI	0.071026	SR	0.066134	NDRE	0.061086
SR	0.070957	Slope	0.063543	NDVI	0.057231
NDVI	0.066479	B	0.053326	SR	0.054473
EVI	0.062626	PCA1	0.049897	RE	0.039488
EVI2	0.062280	NDVI	0.048751	EVI	0.037066
R	0.055091	NDRE	0.045927	G	0.036853
PCA1	0.045118	MSAVI2	0.041938	B	0.031741
MSAVI2	0.038290	R	0.039870	RE	0.029746
G	0.036206	RE	0.039124	MSAVI2	0.028926
NDRE	0.029462	G	0.037555	NIR	0.028217
NIR	0.022317	EVI	0.034251	PCA1	0.025003
RE	0.021595	NIR	0.028959	EVI2	0.024275

M, J, O – field campaign month

RF_{ST} – Random Forest model using both spectral and topographic variables

Appendix C: Supplementary materials for Chapter 4

This section contains all the supplementary materials for Chapter 4.

Table C.1. Principal Components (PC) eigenvalues and cumulative eigenvalues obtained from the sets of WM and SM images.

Study site	PC bands	Eigenvalues (%)	Cumulative Eigenvalues (%)
SM	PC1	79.70	79.70
	PC2	20.02	99.72
	PC3	0.23	99.95
	PC4	0.05	100.00
WM	PC1	90.40	90.40
	PC2	9.20	99.60
	PC3	0.31	99.91
	PC4	0.09	100.00

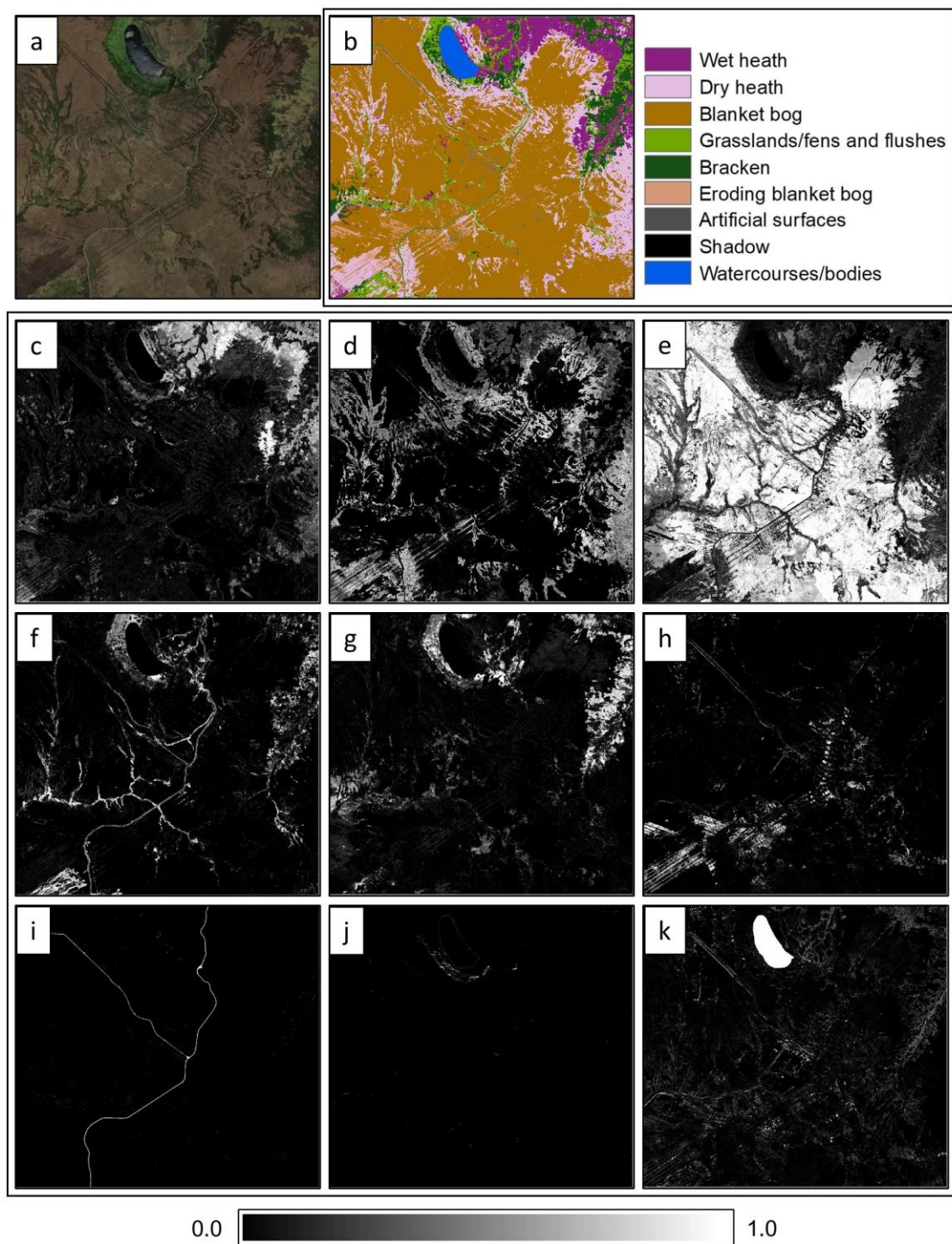


Figure C.1. Basemap of the WM_2m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

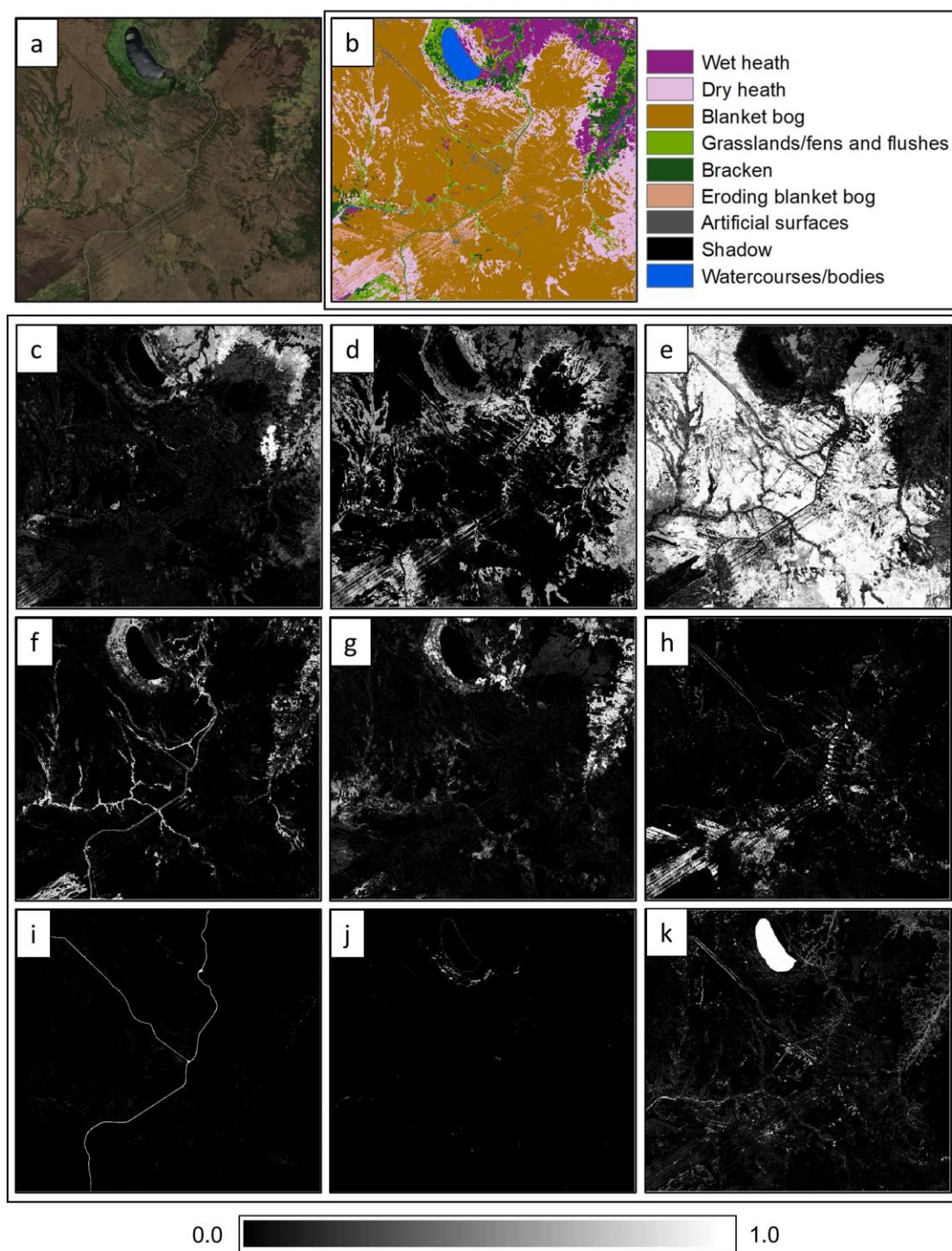


Figure C.2. Basemap of the WM_4m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

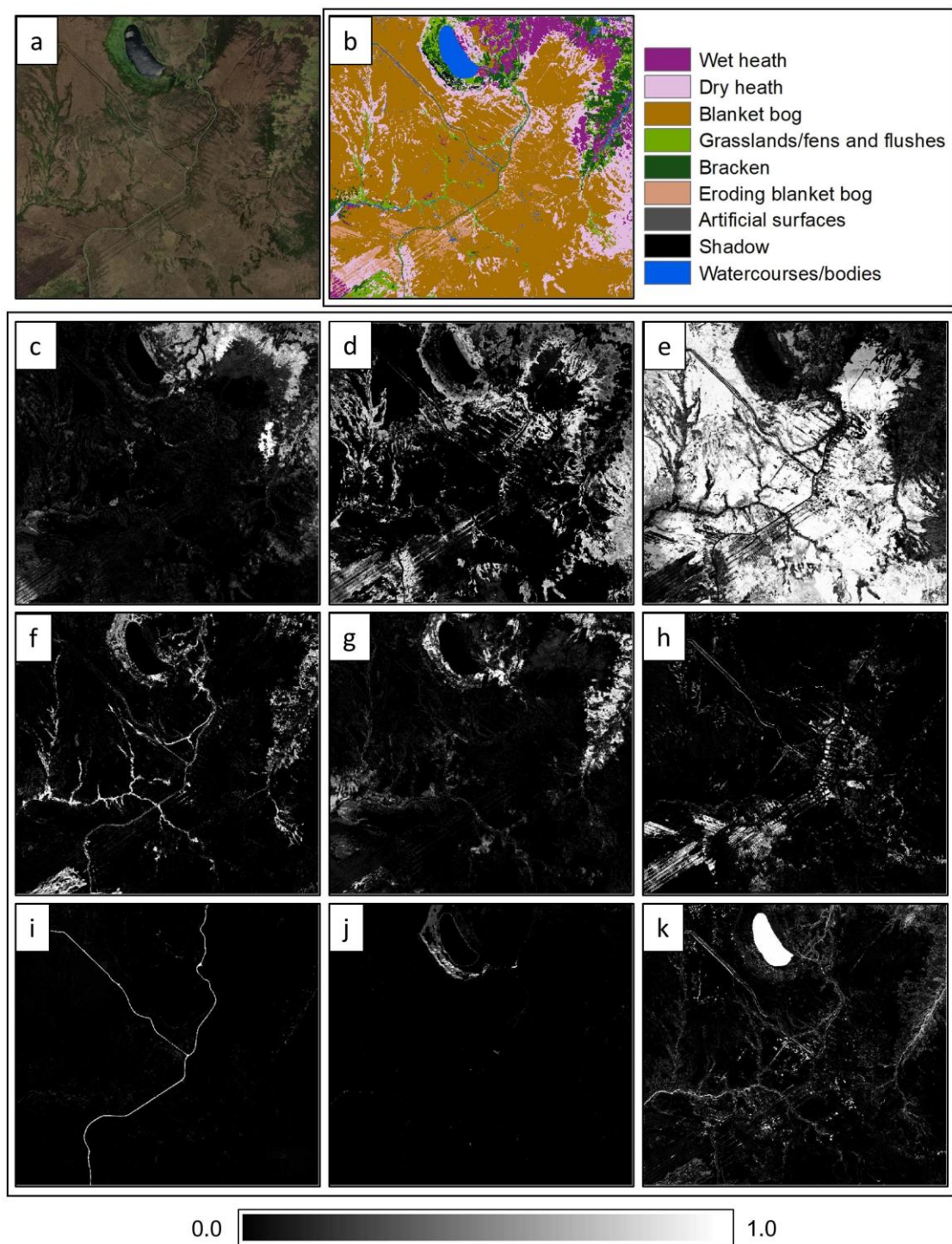


Figure C.3. Basemap of the WM_6m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

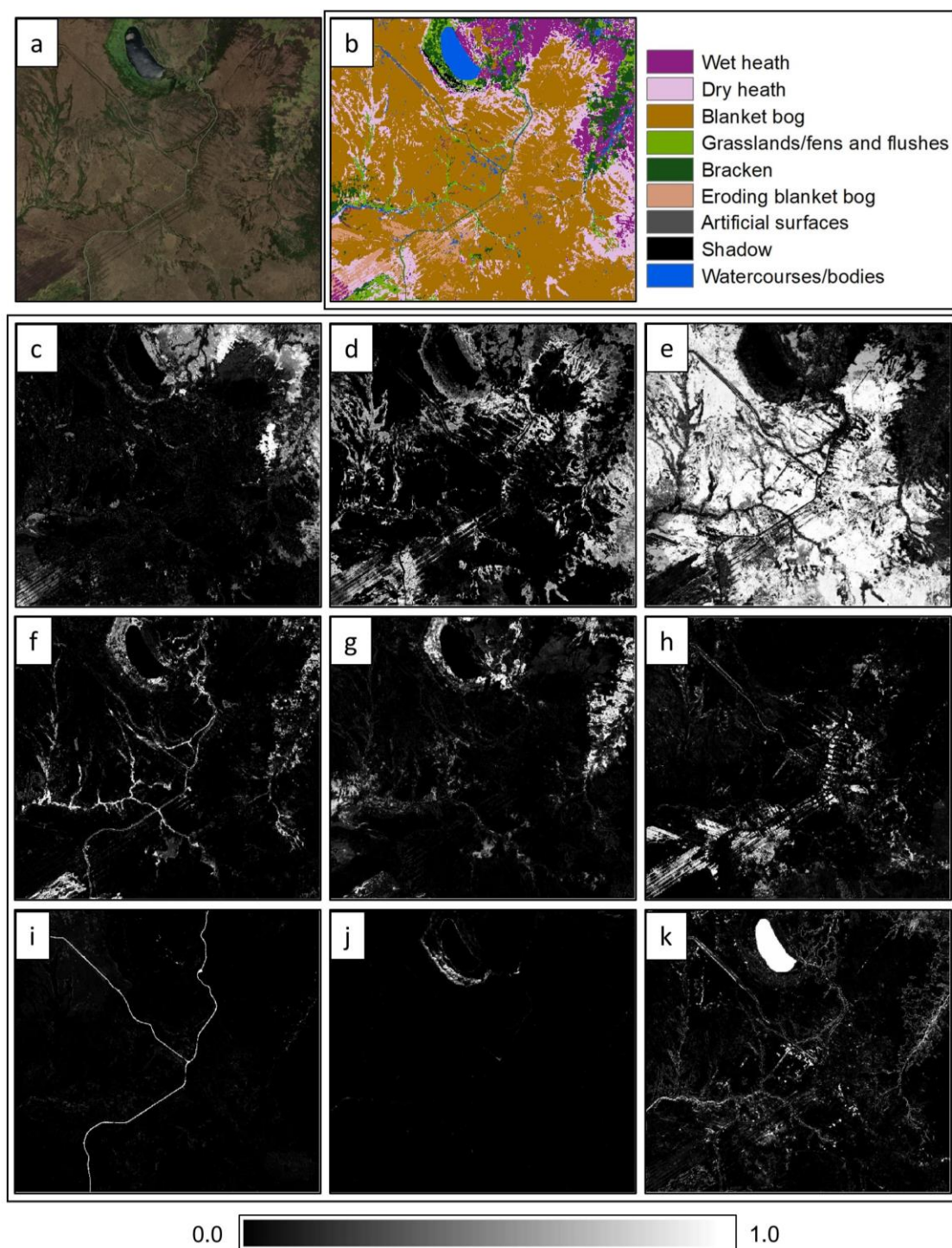


Figure C.4. Basemap of the WM_8m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

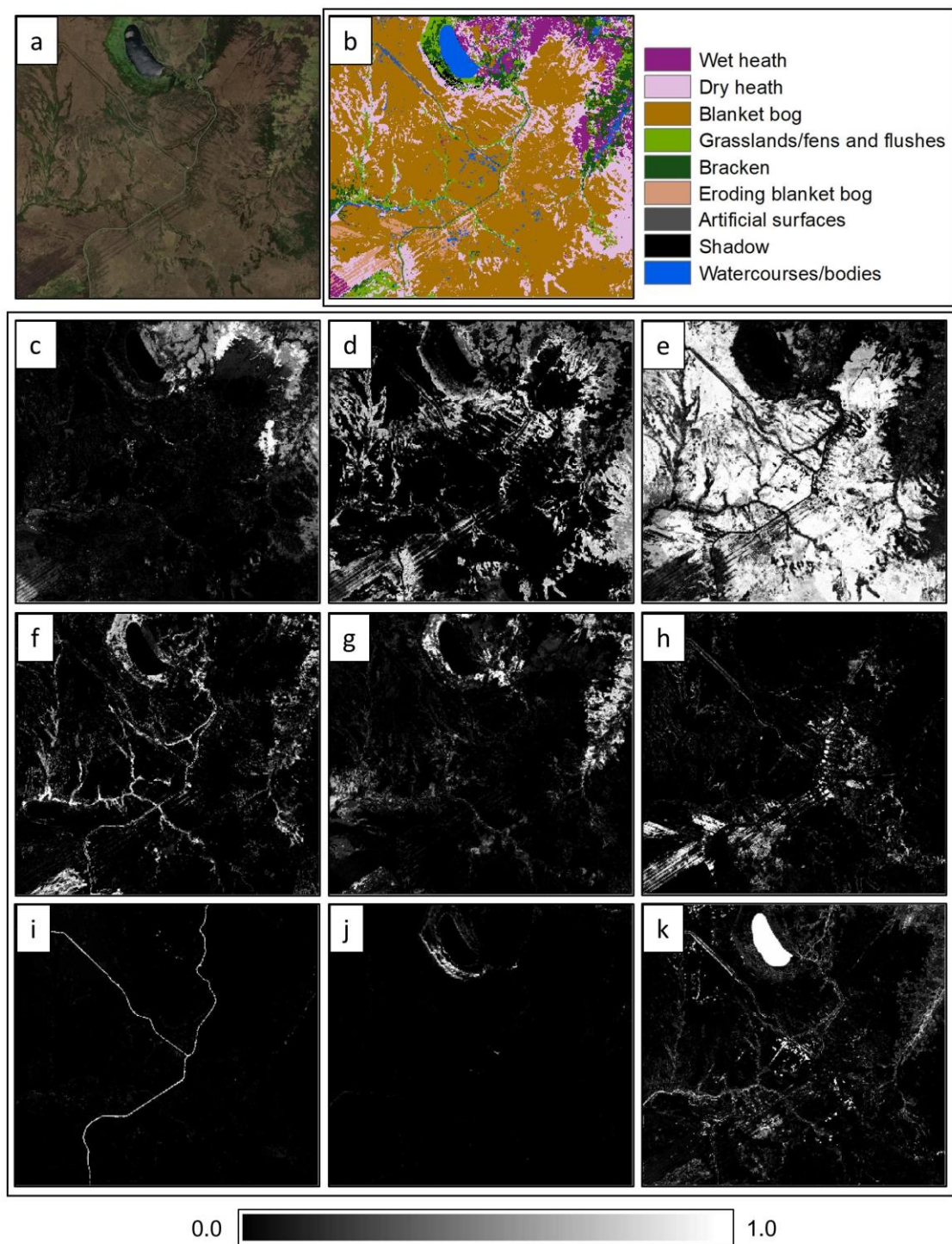


Figure C.5. Basemap of the WM_10m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

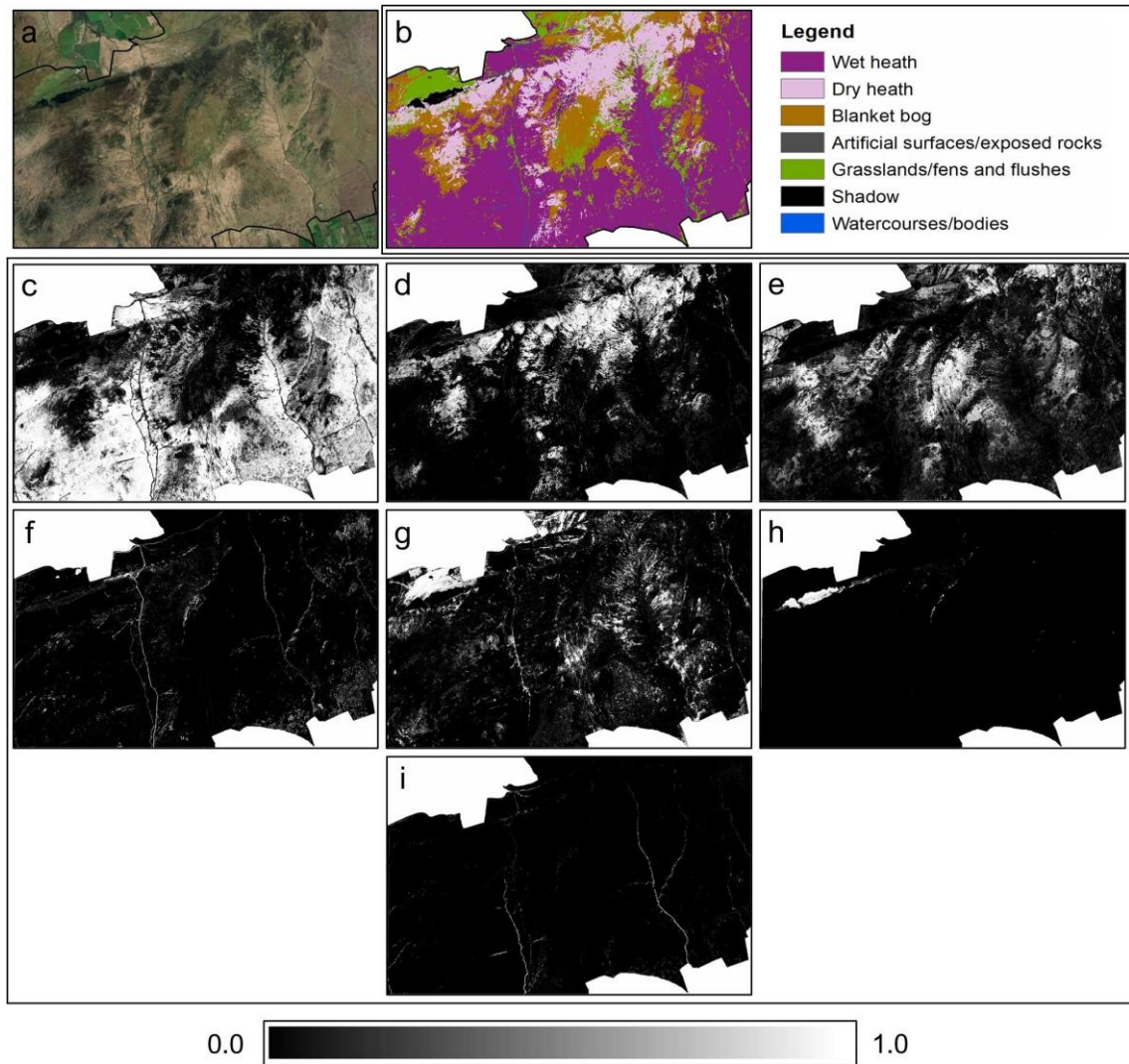


Figure C.6. Basemap of the SM_2m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

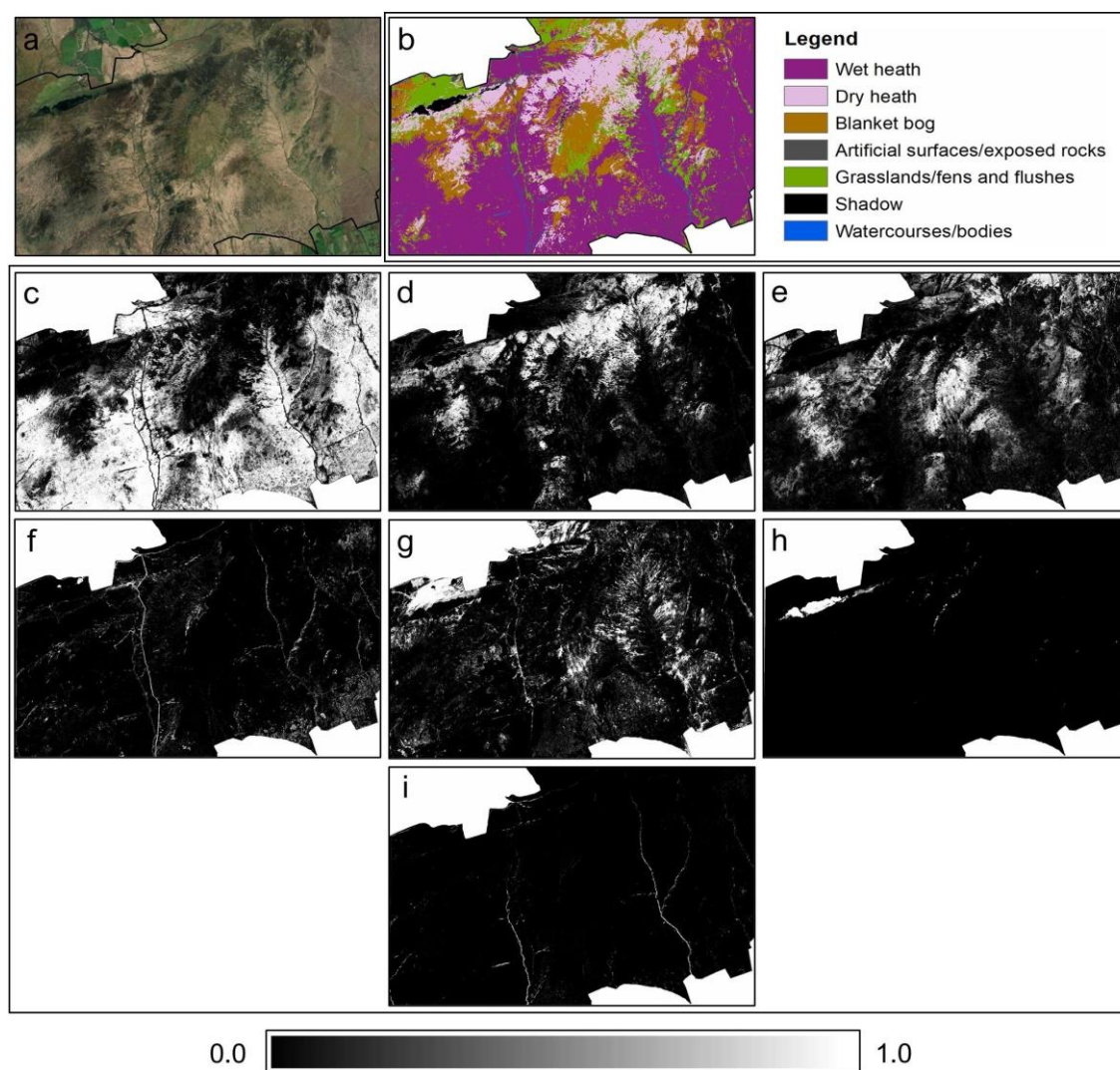


Figure C.7. Basemap of the SM_4m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

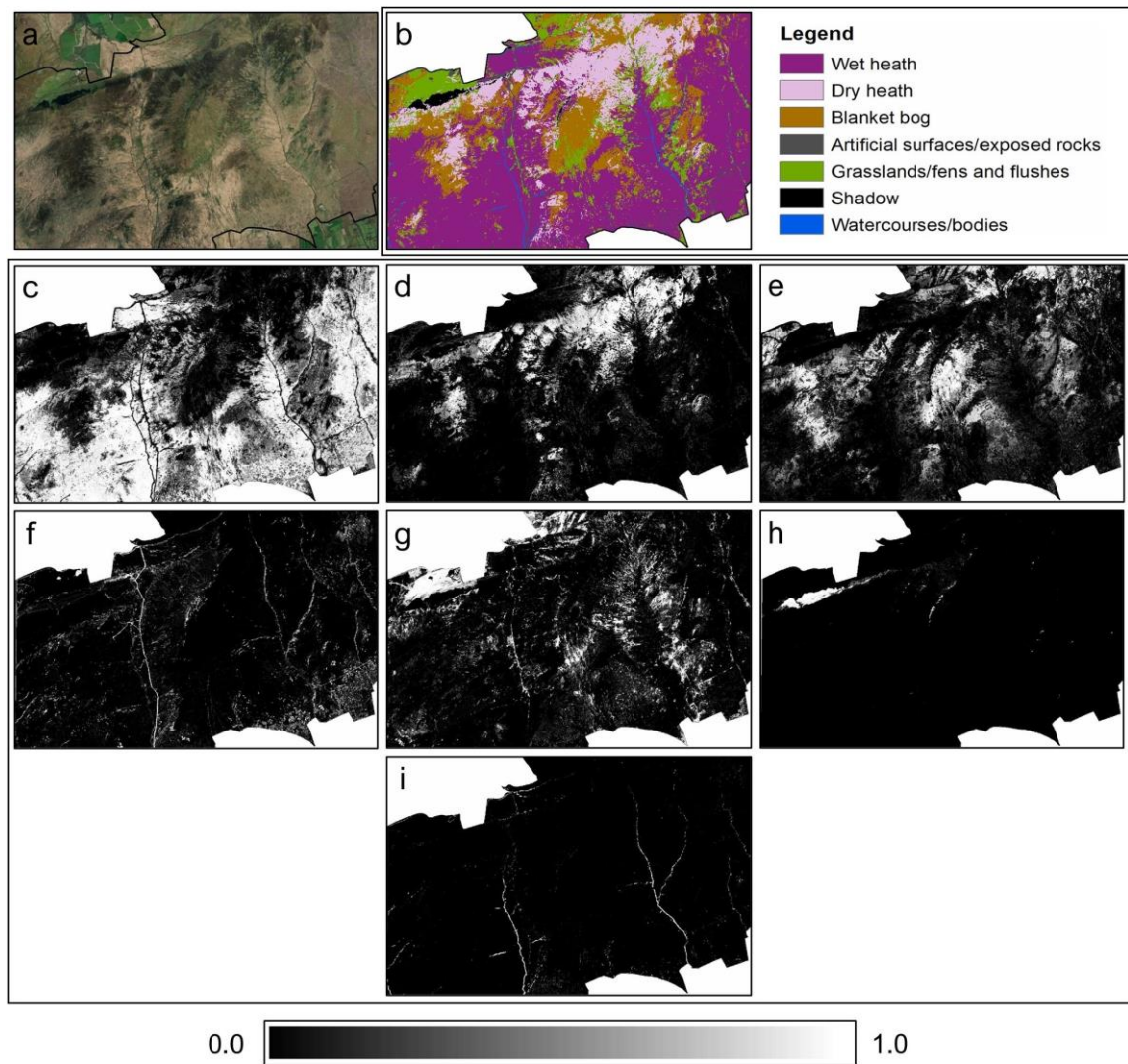


Figure C.8. Basemap of the SM_6m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

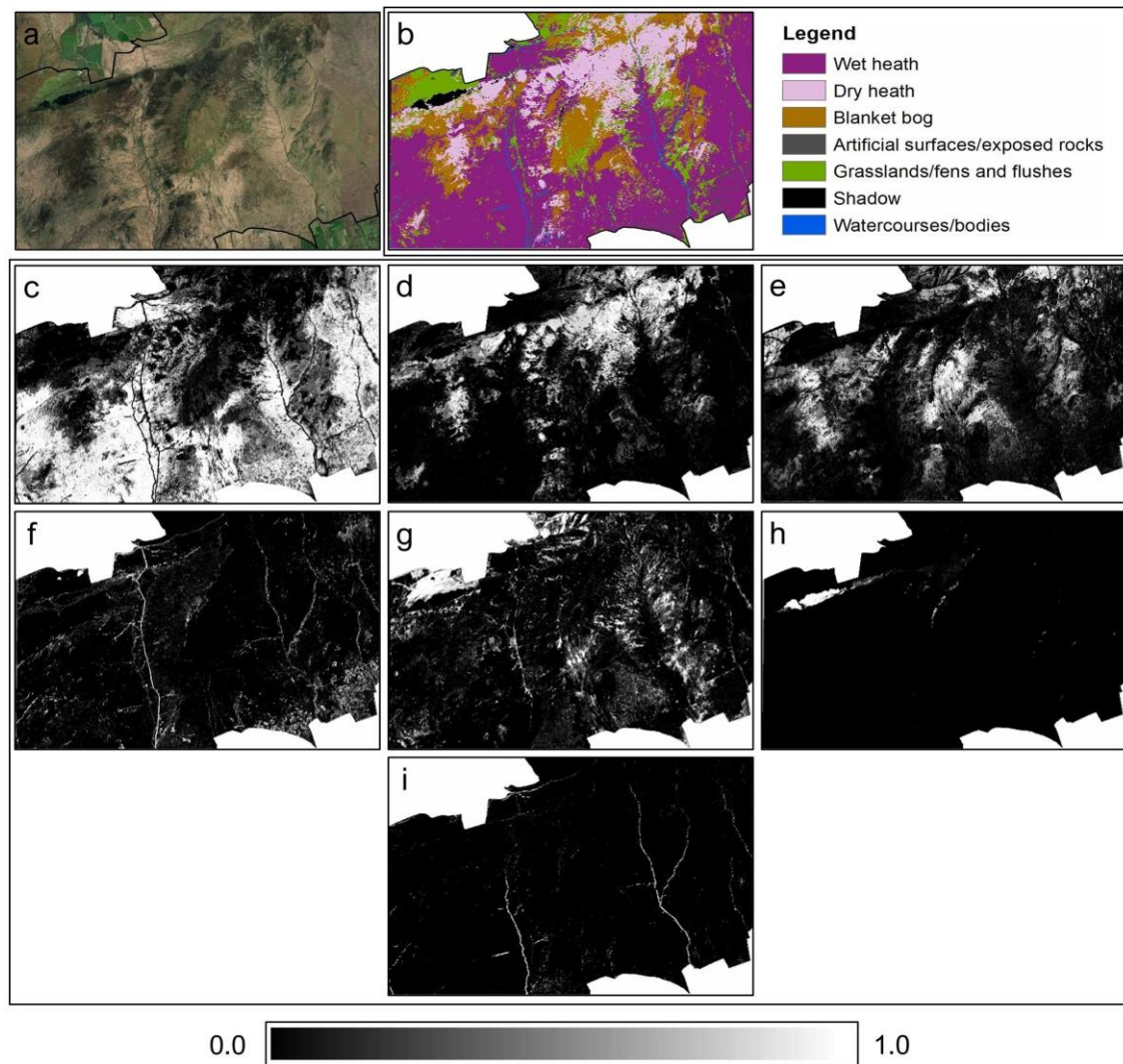


Figure C.9. Basemap of the SM_8m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

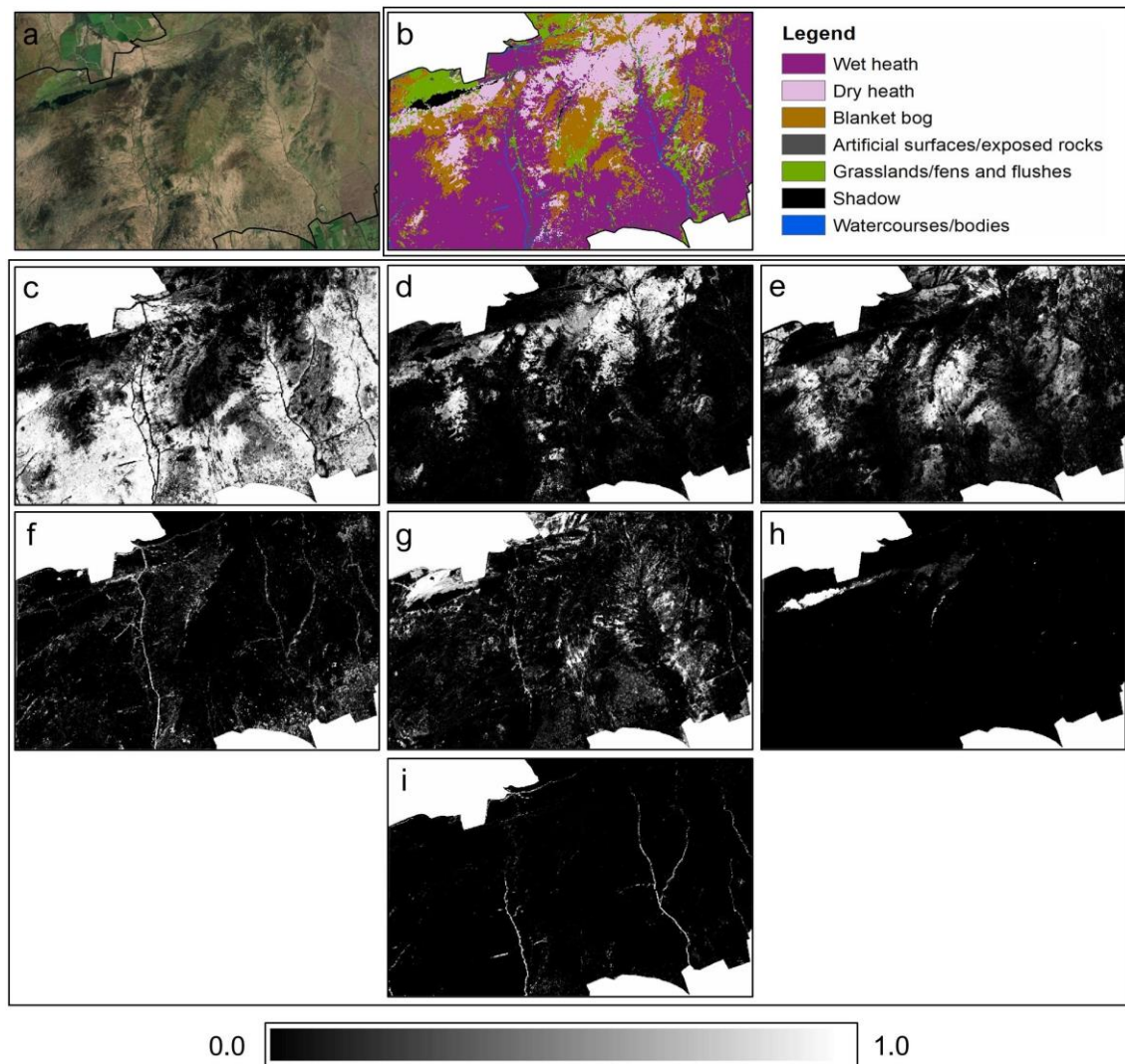


Figure C.10. Basemap of the SM_10m site (a) and the corresponding crisp (b) and fuzzy (c-k) classification results. The colour scale below represents the probability of each habitat being present in every pixel.

Appendix D: List of publications, conference presentations, training and other outreach activities during PhD

This section lists peer-reviewed and submitted articles, conference presentations as well as other outreach activities during the PhD programme.

Peer-reviewed articles

Cruz, C., McGuinness, K., Perrin, P. M., O’Connell, J., Martin, J. R., & Connolly, J. (2023). Improving the mapping of coastal invasive species using UAV imagery and deep learning. *International Journal of Remote Sensing*, 44(18), 5713–5735. <https://doi.org/10.1080/01431161.2023.2251186>

Cruz, C., O’Connell, J., McGuinness, K., Martin, J. R., Perrin, P. M. & Connolly, J. (2023). Assessing the effectiveness of UAV data for accurate coastal dune habitat mapping, *European Journal of Remote Sensing*, 56(1), 2191870. <https://doi.org/10.1080/22797254.2023.2191870>

Under review

Cruz, C., Perrin, P. M., Martin, J. R., O’Connell, J., McGuinness, K., & Connolly, J. Mapping of temperate upland habitats using high-resolution satellite imagery and machine learning. (Under review)

Conference presentations

Cruz, C., O’Connell, J., Martin, J. R., Perrin, P. M., McGuinness, K., and Connolly, J. (2024). Drones and machine learning for habitat monitoring, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-17039, <https://doi.org/10.5194/egusphere-egu24-17039>.

Cruz, C., McGuinness, K., Perrin, P. M., O’Connell, J., Martin, J. R. & Connolly, J. (2023). Improving deep learning-based mapping of coastal invasive

species from UAV imagery. *Remote Sensing and Photogrammetry Society (RSPSoc2023)*. Nottinghamshire, UK. 5 – 6 September 2023.

Cruz, C., O'Connell, J., McGuinness, K., Martin, J., Perrin, P. & Connolly, J. (2023). Mapping Irish habitat types with drones and machine learning. *33rd Irish Environmental Researchers Colloquium*. Atlantic Technological University-Donegal, Ireland. 3 – 4 April 2018.

Cruz, C., O'Connell, J., McGuinness, K., Martin, J., Perrin, P. & Connolly, J. (2022). Drones and Machine Learning for accurate and fine-grained mapping of Irish habitat types. *14th Irish Earth Observation Symposium*. Technological University Dublin Bolton Street Campus, Dublin, Ireland. 3 – 4 November 2022. (Won the best poster)

Cruz, C., O'Connell, J., McGuinness, K., Martin, J., Perrin, P. & Connolly, J. (2022). Can UAVs be used as an effective tool to capture the spatial distributions of habitat types? *XXIV International Society for Photogrammetry and Remote Sensing Congress*. Nice, France. 6 – 11 June 2022.

Cruz, C., Connolly, J., Martin, J., McGuinness, K., O'Connell, J. & Perrin, P. (2021). UAV-based approach for fine-scale mapping of habitats. *Conference of Irish Geographers* (online). 18 – 21 May 2021.

Cruz, C., McGuinness, K., O'Connell, J., Martin, J., Perrin, P. & Connolly, J. (2021). Habitat mapping in coastal dunes using Random Forest classification of UAV images. *EGU General Assembly 2021* (online). 19 – 30 Apr 2021. EGU21-6301, <https://doi.org/10.5194/egusphere-egu21-6301>.

Cruz, C., O'Connell, J., McGuinness, K., Martin, J., Perrin, P. & Connolly, J. (2019) iHabiMap: habitat mapping, monitoring and assessment using high-resolution imagery. *13th Irish Earth Observation Symposium*. National University of Ireland Galway, Ireland. 5 – 6 November 2019. <https://doras.dcu.ie/24004/>

Training

UAS Remote Pilot Competency Certificate. European Union Aviation Safety Agency. Issued on 3 January 2023.

Photogrammetric Image Processing with MicMac & Point Cloud Processing with OPALS (International Society for Photogrammetry and Remote Sensing Summer School). University Campus St. Jean d'Angely, Nice, France. 13th – 17th June 2022.

Open Data Science Europe Workshop. Wageningen International Congress Centre, Wageningen, the Netherlands. 6th – 10th September 2021

Other outreach activities

Insight Sustainability: iHabimap stewarding coastal dunes with drone tech. Insight Spotlight on Research (2023). Link to the article: <https://www.insight-centre.org/insight-sustainability-ihabimap-stewarding-coastal-dunes-with-drone-tech/>

Multi-temporal classification of coastal dune habitats using UAV images. SENSECO COST action CA17134 - Young SENSECO webinars (2021). Link to the video: <https://bit.ly/3RF77yY>

iHabiMap: Habitat Mapping, Assessment, and Monitoring using High-Resolution Imagery. Article published in the International Society for Photogrammetry and Remote Sensing Student Consortium Newsletter (2021), Volume 14 Issue 4, pages 12-14. Link to the article: https://sc.isprs.org/media/bqkempkz/isprs-sc_vol14-issue-4.pdf

Data in a time of Covid podcast. The Insight SFI Research Centre for Data Analytics (2021). Link: <https://www.insight-centre.org/data-in-a-time-of-covid-listen-to-ep-08-of-the-insight-podcast-now/>

How data science and drones combine to assess Ireland's habitats. Article published in Silicon Republic (2020).

Link to the article: <https://www.siliconrepublic.com/machines/charmaine-cruz-insight-dcu-drones-geospatial-data>

