Maya McCracken

Site description and investigation of the use of SfM photogrammetry for generating ultra-fine scale variables with species distribution modelling of *Lophelia pertusa* at various spatial scales in Fangorn Bank, Northeast Atlantic

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

2023

Maya McCracken

Site description and investigation of the use of SfM photogrammetry for generating ultra-fine scale variables with species distribution modelling of *Lophelia pertusa* at various spatial scales in Fangorn Bank, Northeast Atlantic

Mestrado em Biologia Marinha

Supervisors:

Dr. Chris McGonigle

Co-supervisor

Dr. Aschwin Hillebrand Engelen

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

2023

Declaração de autoria de trabalho/statement of authorship

Site description and investigation of the use of SfM photogrammetry for generating ultra-fine scale variables with species distribution modelling of *Lophelia pertusa* at various spatial scales in Fangorn Bank, Northeast Atlantic

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída

I hereby declare to be the author of this work, which is original and unpublished. Authors and works consulted are properly cited in the text an included in the reference list.

____

(Maya McCracken)

Acknowledgements

Firstly I would like to thank chief scientists onboard RV Celtic Explorer for collection of data that has been used for this study. I would also like to thank both my supervisors: Dr. Chris McGonigle at Ulster University and Dr. Aschwin Hillebrand Engelen at University of the Algarve; without their support and guidance this study would not be possible.

I would also like to give my personal thanks and gratitude to Ulster University and Dr. Chris McGonigle for providing me with the facilities to carry out this thesis, not to mention their welcoming nature and constant support. I have gained so many enriching experiences and have gained a strong skillset, which I know will only make my future brighter. I have been given so many opportunities from my time there and was able to work on my dream dataset, while picking up proficiency in new pieces of software and attending field schools.

A warm and special thank you to my friends, family and partner who have been my rock and guide through all of this. I can't imagine any of this would be possible without the constant phone calls and WhatsApp messages to my fellow marine biology colleagues and endless phone calls to my mum and dad, urging me on and motivating me to never give up. I can't thank my partner Stra enough for their unbending generosity and multiple proofreads of my final draft. Their patience, kindness and love has guided me through.

It is with my utmost gratitude to the University of the Algarve for my experience completing my Marine Biology master's degree. The lecturers during my course have inspired me from their infectious passion and knowledge. Furthermore and most valuably, I have gained a really strong group of connections, colleagues and friends who I know I will have for the rest of my life!

Resumo

Nas últimas duas décadas, houve um aumento significativo no uso de Modelos de Distribuição de Espécies (MDE) para estudar e gerenciar espécies marinhas. Esse aumento se deve principalmente à disponibilidade generalizada de variáveis preditoras globais e à facilidade de implementação das técnicas de MDE. Os pesquisadores estão cada vez mais focados em entender a distribuição de espécies no mar profundo. No entanto, a tradução dessa pesquisa em ferramentas práticas de gestão para lidar com as mudanças climáticas e a crescente exploração dos recursos do mar profundo tem sido mais lenta, mas é imperativa.

O MDE, conforme definido por Zurrel (2020), é um modelo de biodiversidade baseado empiricamente que utiliza métodos estatísticos e de aprendizado de máquina. Esses modelos associam registros de biodiversidade geográfica, incluindo presenças observadas e às vezes ausências ou contagens medidas, às características abióticas e/ou bióticas em locais específicos. Os MDEs são usados de forma intercambiável com termos como Modelos de Nicho Ecológico (MNE), modelos de adequação de habitat e outros, destacando diferentes aspectos das entidades que estão sendo modeladas e do tipo de dados usados.

O crescimento dos MDEs é atribuído a dois fatores-chave: aumento da capacidade computacional e aprimoramento dos métodos de aquisição de dados. Esse crescimento tornou os Modelos de Adequação de Habitat (MAHs) ferramentas essenciais para avaliar implicações ecológicas e ambientais, especialmente diante de pressões sociais e da expansão da economia azul. Além disso, o aquecimento dos oceanos e as mudanças na concentração de aragonita aumentaram a necessidade de tais modelos.

A contínua diminuição da biodiversidade, conhecida como a sexta extinção em massa, destaca a urgência de reverter essa tendência. Informações confiáveis sobre a ecologia e a distribuição das espécies são essenciais para a proteção de habitats, a mitigação das mudanças ambientais e a tomada de decisões informadas. No entanto, o uso de modelos inadequados pode levar a resultados custosos e ineficazes, enfatizando a importância da identificação de baixo desempenho do modelo e alta incerteza. A formalização de métodos de modelagem e critérios de avaliação é crucial para previsões precisas e para evitar decisões equivocadas.

A Teoria do Nicho Ecológico é fundamental para as abordagens de MDE, postulando que as espécies prosperam em condições ambientais específicas. Modelos de adequação de habitat (MAHs) têm sido usados para entender os impactos da conectividade de habitats em larga escala, prevendo a distribuição

de espécies e os fatores subjacentes. Adotar uma abordagem multi-escala pode aumentar a precisão preditiva dos MAHs, mas frequentemente é estatisticamente complexa e computacionalmente intensiva.

A transparência e a reprodutibilidade são essenciais para os modelos usados em avaliações de impacto ecológico, planejamento de conservação, tomada de decisões e análises de biodiversidade. Zurrel et al. (2020) desenvolveram um procedimento de modelagem chamado ODMAP, que aprimora a transparência e a reprodutibilidade nos MDEs.

Embora os MDEs tenham se tornado amplamente utilizados em ecologia, evolução e conservação, ainda há falta de padronização e documentação em relação aos protocolos de modelagem. Essa revisão, incluída neste documento, tem como objetivo avaliar o conhecimento atual em modelagem de distribuição de espécies, seleção de dados e abordagens de modelagem em ecologia, com foco em ambientes marinhos e terrestres.

A seleção de dados e preditores é uma etapa inicial crítica. Os autores devem fornecer nomes taxonômicos, sistemas de referência taxonômica, desenhos de amostragem espacial e temporal, tamanhos de amostra e prevalência de táxons focais. Eles também devem considerar erros e vieses potenciais nos dados.

A seleção de preditores envolve a escolha de variáveis ambientais que influenciam a distribuição das espécies. Isso pode incluir características topográficas, fatores bióticos e variáveis abióticas. Testar a colinearidade é crucial para evitar redundância e superajuste.

A escala é outra consideração crítica. A modelagem de única escala usando um grão e extensão de preditor universais é comum, mas MAHs multi-escala, que integram preditores medidos em sua "escala de efeito", fornecem previsões mais precisas e insights ecológicos mais profundos. A escolha da escala pode influenciar significativamente a precisão e a utilidade do modelo.

Dados de Muito Alta Resolução (MAR), como fotogrametria Structure from Motion (SfM), oferecem dados em escala fina para prever a distribuição de espécies, especialmente em ambientes complexos como recifes de coral. Essas tecnologias aprimoram a precisão e a exatidão dos MDEs, capturando variações microambientais anteriormente inacessíveis.

A seleção de algoritmos é essencial, com várias opções como Modelos Lineares Generalizados (GLM), MaxEnt, Random Forest e Support Vector Machine disponíveis. Modelos de conjunto que combinam vários algoritmos podem fornecer previsões robustas.

Métodos de avaliação desempenham um papel crucial na avaliação do desempenho do modelo. Métricas como Estatísticas de Habilidades Verdadeiras (TSS), coeficiente Kappa, curvas Receiver Operating Characteristic (ROC) e Área Sob a Curva (AUC) devem ser usadas para medir a precisão e a adequação do modelo.

No mar profundo, os MDEs são ferramentas valiosas para identificar potenciais habitats para corais e esponjas de águas profundas (CCEAs) e informar decisões de gerenciamento. Modelos de conjunto e aqueles que usam dados de presença-ausência são preferidos, aproveitando informações entre táxons e facilitando inferências comunitárias.

A Modelagem de Distribuição de Espécies tornou-se uma ferramenta vital em ecologia e conservação. A seleção adequada de dados, a escolha de preditores, a consideração da escala e os métodos de avaliação são cruciais para previsões precisas e confiáveis. Fontes de dados de MAR, como fotogrametria SfM, têm potencial para aprimorar a precisão do modelo. No mar profundo, os MDEs são valiosos para o gerenciamento do habitat de CCEAs. Compreender e implementar esses princípios de modelagem é essencial para a tomada de decisões informadas em um mundo em rápida transformação.

A batimetria de alta resolução (HR) derivada de dados MBES e dados de vídeo em alta definição (HD) foi adquirida na expedição CE21010 'Recursos de Rockall'. Os dados primários desta expedição foram utilizados neste estudo para prever o habitat adequado do Coral de Água Fria (CWC) Lophelia pertusa. Um fluxo de trabalho para a geração de modelos digitais de elevação em escala muito fina usando técnicas de fotogrametria SfM é delineado neste estudo usando o software Agisoft Metashape. Registros de ocorrência foram desenvolvidos a partir de anotações de imagens de vídeo de ROV de observações de estruturas de coral vivo. Um conjunto de camadas ambientais foi gerado a partir de dados de batimetria HR usando a Ferramenta TASSE no ArcGIS Pro e utilizado como camadas de dados ambientais para o procedimento de modelagem. A técnica de Árvores de Regressão Impulsionada (BRT) foi utilizada para criar diversos modelos em resoluções amplas, de escala fina e ultra fina. Restrições de encolhimento e validação cruzada foram usadas para evitar o excesso de ajuste, e restrições monotônicas foram usadas para reduzir a complexidade do modelo. Em todos os modelos, L. pertusa mostrou preferência por alta complexidade de terreno e picos topográficos. As faixas de profundidade para a ocorrência da espécie estavam entre -1.000 m e -800 m. O modelo de escala fina obteve o melhor desempenho, com um valor de AUC de 0,970, e uma pontuação de sensibilidade de 0,980 e especificidade de 0,965. O modelo de escala ampla obteve um valor de AUC de 0,951 e uma pontuação de sensibilidade de 0,875, com uma especificidade de 0,939 e um valor geral de TSS de

0,814. A precisão das previsões para o modelo de escala ampla não foi tão alta quanto a do modelo de escala fina, indicando que os modelos treinados com dados de escala fina produzem previsões mais confiáveis. O modelo de ultra fina escala (UFS) obteve um valor de AUC de 0,834. Embora o modelo UFS tenha tido o pior desempenho geral, a técnica de fotogrametria SfM apresentada aqui mostra promessa para pesquisas futuras na geração de modelos digitais de elevação em ultra fina escala derivados de dados de vídeo para serem usados em modelos de distribuição de espécies para melhorar a precisão das previsões.

Vinte e seis grupos faunísticos distintos foram registrados durante a anotação do ROV, juntamente com quatro tipos de substrato primário. Uma cadeia de características de montes de águas profundas foi revelada por meio da análise GIS de dados MBES, sendo a característica de monte mais alta medida com 2,4 km de elevação. A Fangorn Bank demonstrou ser um local biologicamente complexo sustentado por um alto grau de complexidade de terreno de altas rochas inclinadas e fundos lamacentos planos que levam aos ricos e diversos ecossistemas encontrados aqui. A identificação de 26 grupos faunísticos e suas características neste local de estudo destaca sua notável biodiversidade e a importância de sua preservação.

Palavras-chave: Dados MBES, ROV, Fotogrametria SfM, SDMs, CWCs, BRT

Table of Contents

This page has intentionally been left blank.

List of Tables

List of Figures

List of Abbreviations and Acronyms

AUC	Area under the curve
BRT	Boosted regression tree
CWC	Cold water coral
DEM	Digital Elevation Model
DSCS	Deep-sea coral and sponge grounds
FS	Fine scale
GLM	Generalised linear modelling
HD	High definition
HR	High resolution
HSM	Habitat suitability modelling
MaxEnt	Maximum Entropy
MBES	Multibeam Echosounder
ML	Machine learning
MPA	Marine Protected Area
ROV	Remotely operated vehicle
SDM	Species distribution modelling
SfM	Structure from motion
UFS	Ultra-fine scale
VHR	Very high resolution
VME	Vulnerable marine ecosystem

A review of species distribution (SDM) modelling approaches: From land to the deep sea

General Introduction

Over the past two decades, there has been a significant increase in the use of Species Distribution Modelling (SDM) for studying and managing marine species (Melo-Merino et al., 2020). This rise can be attributed to the widespread availability of global predictor variables and the ease of implementing SDM techniques. Consequently, there has been a surge in research focused on understanding species distribution in the deep sea. However, the translation of these research findings into practical management tools that can aid decision-making in the face of climate change and increasing exploitation of deep-sea resources has been slower but imperative (Oldfield., 2015).

To provide clarity, Zurrel (2020) defines SDM as an empirically based biodiversity model that utilises statistical and machine learning methods. These models associate geographic biodiversity records, including observed presences, and sometimes absences/non-detections or measured counts, with the abiotic and/or biotic characteristics at those specific locations. Similar terms used interchangeably with SDMs or closely related models include ecological niche models (ENM), species range models, environmental or climate envelopes, habitat suitability and habitat distribution models, occupancy models, resource selection functions, abundance and N-mixture models. Often, these names emphasise different aspects of the entities being modelled: the niche, the distribution or the habitat preferences of species, or the data type used (Zurrel et al., 2020; Elith and Leathwick 2009; Peterson and Soberón 2012; Guisan et al., 2017).

Our understanding of the living world and the factors driving species distribution has greatly improved, as has our ability to predict it. This growth can be attributed to two key factors. Firstly, there has been a rapid increase in computing power, enabling the development of higher quality datasets and improved data acquisition methods. This has provided greater freedom to analyse, predict, and model various taxa across different spatial scales and extents (Zurrel et al., 2020). Habitat Suitability Models have become essential tools for evaluating ecological and environmental implications, particularly in the face of societal pressures and the rapid expansion of the blue economy and marine spatial planning. Additionally, the warming oceans and shallowing of aragonite concentration horizons have further stressed the need for such models (Roberts and Cairns, 2014; Bowden et al., 2021).

The ongoing rapid decline in biodiversity, known as the sixth mass extinction (Barnosky et al., 2011), emphasises the urgent need for reversal before it is too late. To protect habitats, mitigate environmental change, and make informed decisions, reliable information on species ecology and distributions is essential (Loiselle et al., 2003). Habitat Suitability Models (HSM) and Species Distribution Models (SDMs) provide landscape-scale predictions of habitat suitability and associated environmental factors, filling gaps in occurrence data (Guisan et al., 2013). However, the use of poor models can lead to costly and ineffective outcomes, highlighting the importance of identifying low model performance and high uncertainty (Loiselle et al., 2003). Formalising modelling methods and evaluation criteria is critical for accurate predictions and avoiding misinformed decisions that could have a high cost to conservation efforts (Banard and Boyce, 2013; Zurrel et al., 2020).

A fundamental assumption related to such modelling approaches relies on the hypothesis that individual species only thrive within a definite range of environmental conditions, and this has stimulated one of the most fertile fields in ecology: The Ecological Niche Theory (Chase and Leibold, 2003). The requirement-based concept of the ecological niche (Grinnell, 1917) defines it as a function that links the fitness of individuals to their environment (Hirzel and Lay, 2008). HSMs have been used to understand the impacts of landscape-scale habitat connectivity (Wright et al., 2021). This approach predicts the distribution of a species and provides information on the underlying drivers using occurrence records and environmental data layers (Guisan et al., 2017). HSMs can improve our understanding of a species' ecology and are valuable tools for informing landscape-scale decisions. It is possible to increase HSM predictive accuracy and derive more realistic conclusions by taking a multi-scale approach. However, this process is often statistically complex and computationally intensive (Bellamy et al., 2020). A major parameter that can influence the predictive accuracy and usefulness of an HSM is the spatial scale at which predictors are considered (Guisan and Thuillier., 2005).

Ensuring transparency and reproducibility is also crucial for models used in ecological impact assessments, conservation planning, decision making, and biodiversity analyses (Fitzpatrick et al., 2021). This involves providing sufficient information about input data, model implementation, evaluation, validation, and output processing, allowing end-users (e.g., conservationists, evaluators) to assess the reliability and relevance of the model without direct communication with the authors (Araújo et al., 2019). To address these needs, Zurrel et al. (2020) developed a modelling procedure called ODMAP, which outlines the important considerations, methodologies, and design aspects that should be followed and published to enhance transparency and reproducibility in the context of SDMs.

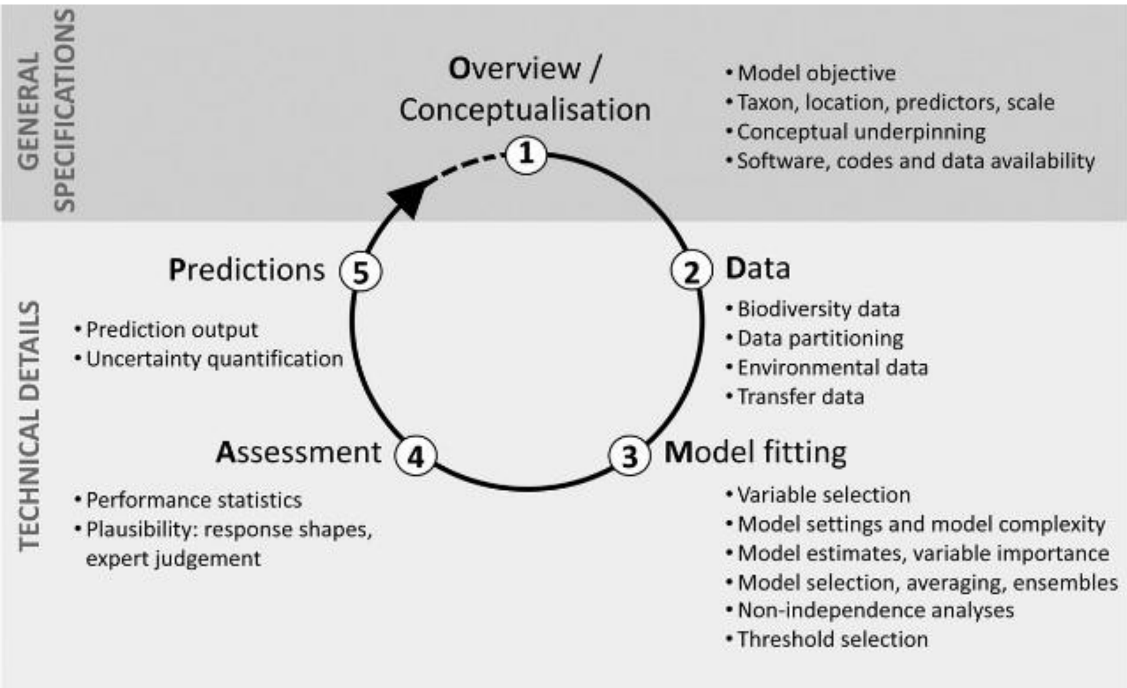

Figure 1. Graphic of ODMAP steps proposed by Zurrell et al. (2020)

SDMs have become widely used in ecology, evolution, and conservation as they provide valuable insights into species distribution. The availability of user-friendly software and digital geoinformation has greatly facilitated the application of SDMs, allowing for their increased use in conservation and management efforts, as well as in assessing the impacts of global change. However, there is still a lack of standardisation and documentation regarding modelling protocols for SDMs (Zurrell et al., 2020). It is therefore the aim of this review to assess current knowledge in Species Distribution Modelling, data inputs and modelling approaches in the field of ecology, with the objective to better understand how these factors may come together to affect accuracy and predictions within the modelling process to infer accurate predictions for making sound management decisions. This review will aim to encompass both realms – marine and terrestrial – to better understand appropriate methods for model species of choice.

Data and Predictor Selection

The initial steps in data inclusion and description involve providing taxon names and information on the taxonomic reference system. Furthermore, it is important to define the focal taxonomic units that are being modelled. Authors should also describe the spatial and temporal sampling design, as well as the sample size and prevalence of the focal taxa in the case of presence-absence data (Zurrell et al., 2020). It is essential to consider potential errors and biases in the data, such as variations in spatial and temporal precision, which can have a significant impact on the accuracy of the models (Meyer et al., 2016; Park and Davis, 2017).

Data Inputs

It is essential to provide detailed information on how absence data were obtained and their accuracy. False absences can result from low detection probability (Guillera-Arroita, 2017). In many SDM studies, presence-only data are used, requiring the use of background data, also known as "pseudo-absences" or "quadrature points," against which the observed presences are compared (Renner et al., 2015). For instance, in cases where presence records exhibit spatial bias, background data can be sampled to reflect the same spatial bias (Phillips et al., 2009; Kramer-Schadt et al., 2013).

Predictor Selection

Early biogeographers determined the same species could occur in different places with different environmental conditions and therefore concluded certain species had a tolerance range in which they are able to persist in space and time (Hirzel and Lay, 2008) thereby occupying different habitats (Guissan et al., 2017). Yet, most species have a limited geographic range (Woodward and Kelly, 2003). However, outside this general tolerance range it is accepted there will be a species range shift or an inability to persist in that space and time any longer (Elith and Leathwick 2009; Peterson and Soberón 2012). From this it is possible to deduce that species can colonise a range of conditions along certain environmental gradients, but in most cases the range only occupies a proportion of all possible habitat available, resulting in species only occupying a limited environmental and geographic range (Guisan et al., 2017). This theory is based on physiological assumptions and responses of the model species (Grinnell, 1917; Hutchinson, 1957; Peterson and Soberón, 2012) based on its ecological niche.

Topographic variation is generally accepted to underpin many processes and patterns in hydrology, climatology, geography and ecology. Different terrain features are therefore key to understanding species distribution and prevalence, as well as life on earth in general. Terrain variables therefore have the potential to be powerful predictors during the selection process of model building for species distribution (Ammatuli, 2022). The biotic environment is also an important part of predictor selection. A study by Lecours et al. (2016) developed a selection of six abiotic surrogates of terrain. Marker variables included easternness, local mean, northernness, relative deviation, slope and standard deviation (StDev) which is a measure of rugosity or roughness, or proxy for complexity (Lecours et al., 2017). These marker variables provided were concluded to be the best trade-off between predictive ability and overfitting of the data (Lecours et al., 2018). Moreover, terrain characteristics may represent vital refugia for species under climate change, a function that likely varies with spatial grain (Ammulti et al. 2020).

Additionally, the most important variables in predicting habitat suitability can be influenced by the scale of the study. For example, Valle et al. (2013) found that wave exposure and current velocity were the key variables predicting *Zostera marina* distribution when using broadly distributed data, whereas slope and depth were important at predicting species distribution of conservative presence areas. This has implications for the creation of HSMs for informing restoration and may require the need for different models and variables dependent on the scale, local topography and location in question (Bertelli et al., 2022). In a study carried out by Bertelli et al. (2022) the most used predictor variables were bathymetry (74%), salinity (57%), light (51%) and temperature (51%) for HSMs of seagrasses.

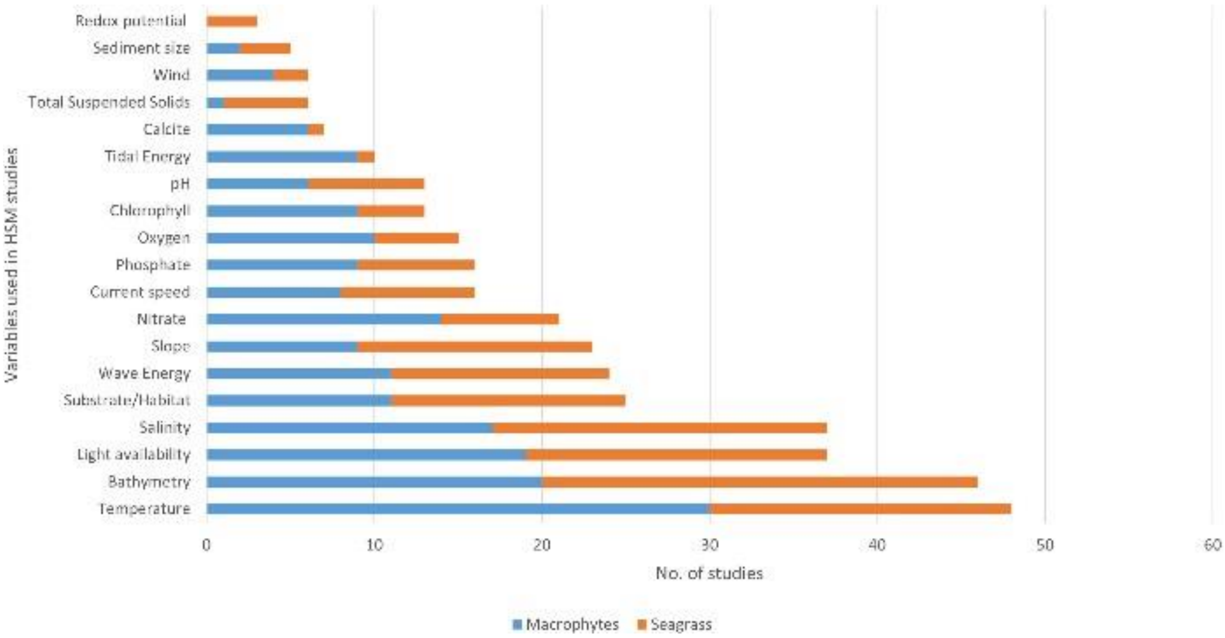

Figure 2. Most common variables selected in Macrophyte HSM studies (Bertelli et al. 2022)

Testing for collinearity is widely recognised as a crucial step after selecting predictors, as it helps mitigate the influence of confounding variables and eliminate redundancy in the model, aligning with the goal of limiting the number of variables to prevent overfitting and creating simpler, more generalisable models (Elith et al., 2010). Guisan et al. (2017) also highly recommends reducing variables with Variance Inflator Factor (VIF) analyses or simple pairwise correlations to test for collinearity (Dormann et al., 2013) to ultimately select a series of non-correlated ecologically relevant variables.

A Question of Scale

Single-scale modelling using a universal predictor grain and extent is widely adopted (Vincente at al. 2014), even though multi-scale HSMs that integrate predictors measured at their 'scale(s) of effect'

(Holland et al. 2004) provide more accurate predictions, give deeper ecological inference and avoid spurious conclusions caused by scale mismatches (Poizat and Pont, 1996; Vincente et al., 2014; Bellamy at al., 2013).

Combining predictors measured over a range of scales increases the likelihood of multicollinearity (Lipsey et al. 2017) which breaks statistical assumptions and confounds model inference. HSMs that integrate drivers at their scale of effect across multiple levels, and that encompass the full range of conditions experienced across the population range, provide a more complete characterisation of a species' niche and prevent truncation of modelled species-environment relationships, i.e., avoid cutting extent short: no top ends or bottom ends (Holland et al. 2004, Vincente et al. 2014), therefore encompassing boundaries or limits of the environmental conditions experienced across the species' distribution range. This allows for a more comprehensive understanding of the species' habitat requirements and potential distribution patterns.

Effect of Ecological Grain and Spatial Extent on Model Accuracy

A study by Gogol-Prokurat (2011), examined the effects of grain size, spatial extent, and fine-grain environmental predictors on local-scale model accuracy using occurrence data of four rare plant species in California. They found a trade-off between model extent and the use of fine-grain environmental variables; goodness-of-fit was improved at larger extents, and fine-grain environmental variables improved local-scale accuracy, but fine-grain variables were not available at large extents. No single model met all habitat prioritisation criteria, and the best models were overlaid to identify consensus areas of high suitability (Gogol-Prokraut et al., 2011). Although the four species modelled here co-occur and are treated together for conservation planning, model accuracy and predicted suitable areas varied among species. Similarly, Barton et al. (2018) examined running their model with geographic data from a mountainous terrain in South Africa using climate data at three spatial resolutions. They found that the modelling framework revealed how spatial resolution and topography influence predictions of species distribution models, including the potential impacts of climate change. These systematic biases must be accounted for when interpreting the outputs of future modelling studies, in addition to designing the model architecture (Barton et al., 2018).

In a study by Ammatuli et al. (2018) reviewing topographic variables for ecological modelling, slope, aspect, easternness and northernness, roughness, rugosity, and terrain ruggedness were used. Different terrain features are key to understanding variation of life on Earth and have potential to be powerful predictors in the selection process of model building and for species distribution (Stein et al., 2015; Ammatuli et al., 2018). In their study a moving window analysis to aggregate variables to coarser spatial grains was used for data to be gridded at 1, 5, 10, 50 and 100km. They concluded that environmental and biological processes depend strongly on the spatial grain of the input variables. For instance, to successfully predict species occurrences using species distribution models, the spatial grain of the environmental variables must match and agree with the sample size and spatial data accuracy (Ammatuli et al. 2018).

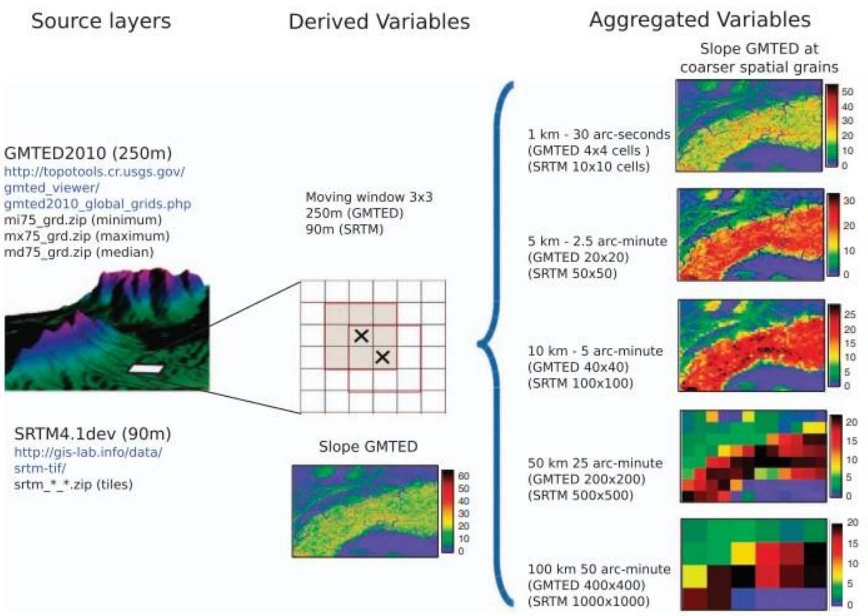

Figure 3. Process of aggregating source layers which provide basis for derived variables to be gridded to coarser spatial grains and used as data layers in SDMs (Ammatuli et al., 2018)

A study by Ross et al. (2015) aimed to assess the impact of different data resolutions on predictive models using real-world data. High-resolution multibeam bathymetric data and terrain variables were used to construct distribution maps for three deep-sea habitats. The study found that the predicted spatial extent of two habitats decreased with higher model resolution, while the suitable area for another habitat increased. Notably, high-resolution models performed better than low-resolution models, and it was recommended to use high-resolution data for modelling instead. The study also emphasised that predictive models should be used to identify likely areas of suitable habitat rather than producing absolute values of habitat extent (Ross et al., 2015).

However, disagreements arise regarding the influence of scale on projected species distributions, particularly in rugged terrains where coarse-scale climate grids fail to capture topographically controlled climate variation relevant to microhabitats or refugia (Franklin et al., 2012). Few formal analyses have explored how modelled distributions differ with scale, but a study examining 52 plant species in the California Floristic Province by Franklin et al. (2012) revealed that as climate data resolution became coarser, SDMs predicted larger habitat areas with reduced spatial congruence between fine- and coarse-scale predictions. These differences in projected habitat change locations can have significant implications for conservation decision-making based on predictive maps, beyond the mere net habitat area.

Very High Resolution

Recent advancements in remote sensing technologies and computing capabilities have paved the way for the creation of Very High Resolution (VHR) variables and Digital Elevation Models (DEMs) at sub-meter scales (Ehlers et al., 2006). While primarily used in engineering studies for precise geomorphological analyses like flood calculations and wind turbine modelling (Kalbermatten et al., 2012; Vaze et al., 2010), these VHR DEMs could also enhance the current terrain attributes and abiotic surrogates of predictors used in SDMs (Pradervand et al., 2013; Lecours, 2016). The incorporation of VHR predictors can potentially improve the precision and accuracy of SDMs by capturing micro-environmental variations previously inaccessible to modellers or limited to small geographic extents (Gottfried et al., 1999; Lassueur et al., 2006; Le Roux et al., 2013a; Wang et al., 2012).

Potential Application of Structure from Motion (SfM) Photogrammetry to Produce VHR Datasets

SfM, as an emerging and cost-effective photogrammetric method, enables high-resolution 3D topographic reconstruction of reefs and similarly complex environments (Burns et al., 2015). The structural complexity of certain ecosystems such as coral reefs is crucial for reef ecosystems, and a 3-dimensional (3D) approach, such as SfM, is better suited to quantify important reef characteristics compared to traditional 2-dimensional metrics (Burns et al., 2015; Quintata and Wilson, 2023). SfM provides a more effective means of quantifying topography, rugosity, and other structural characteristics that play a vital role in the ecological dynamics of coral reef communities.

Quintana and Wilson (2023) hypothesised that the three-dimensional (3D) habitat structure influences species distributions and abundances through biotic interactions (Ehbrecht et al., 2021; Gámez & Harris, 2022). Components of 3D habitat structure, ranging from kilometre to decimetre scales, including geomorphological features and vertical space, as well as the roles of ecosystem engineers, are crucial (Quintana and Wilson, 2023). To reflect this fine-scale resolution, they employed SfM photogrammetry in their SDM of gorgonians and scleractinian corals (Quintana and Wilson, 2023). Additionally, Storlazzi et al. (2016) concluded that SfM techniques provide spatial resolution that is orders of magnitude greater than large-scale LIDAR and sonar mapping for coral reef ecosystems. This finer-scale data offers more applicable measurements of bathymetry and rugosity, particularly valuable for ecological studies of coral reefs and a cost-effective alternative to traditional geophysical methods (Storlazzi et al., 2016).

Modelling Approaches and Algorithm Selection

Habitat Suitability Modelling (HSM) is an approach to create inferences of suitable habitat where the model species is likely to persist based on ecological niche theory (Guisan and Zimmerman, 2000). A general methodology for HSMs includes a set of techniques, procedures and definitions built on an understanding of a core foundation built on ecological and biogeographical concepts (or other biotic components that may be affecting species distribution) about species distribution and the physical (abiotic) environment (Elith and Franklin, 2013). Biotic factors that may impact species distribution, other than biogeographic features, could include predation, disease, competition, scarcity of resources and inter/intra specific variation (Kearney and Porter, 2009). Several approaches can be considered, including but not limited to hierarchical models, correlative modelling and ensemble models (Guisan et al. 2002; Norberg et al. 2019).

Regression based approaches are the most used approaches used in ecology, in particular for HSM (Guisan et al., 2002). They usually rely on robust statistical theories (e.g. sum of squares, maximum likelihood). Regression relates a response variable (e.g. presence-absence, abundance or biomass) to a set of pre-selected environmental predictors (e.g. climate, land use, resources available etc.). In addition to model structure and the selection of predictors, the statistical inference framework within which the model is fit to data can have a major impact on predictive performance (Norberg et al., 2019). As technological advances in machine learning are reshaping the way scientists develop models and analyse data, researchers are increasingly aware that one specific algorithm won’t offer a tailor-made solution to every problem (Chatfield, 1993).

Algorithm Selection

GLM (Generalised Linear Model) is a flexible statistical modelling technique that extends linear regression to handle different types of response variables. MaxEnt (Maximum Entropy) estimates the probability distribution of species occurrence based on environmental variables.

Random Forest is known for its robustness and high predictive accuracy, while GAM (Generalised Additive Model) accommodates non-linear relationships. GAMs are non-parametric regression techniques that offer flexibility in modelling relationships between variables without assuming a specific form for the regression function. By incorporating smooth functions as regressors, GAMs provide greater flexibility compared to linear or other parametric models (Lauria et al., 2017). BRT (Boosted Regression Trees) combines regression trees and boosting algorithms to capture complex relationships. SVM (Support Vector Machine) is versatile, handling both linear and non-linear relationships (Guisan et al., 2017). Each method has its strengths and suitability for different data types and distributions, but proper data assessment and model evaluation are crucial for reliable results.

Table 1. Model approaches for marine macrophytes and recommended model approaches from Bertelli et al. (2020)

Algorithm	Model selection	Model validation	Uncertainty	Comment
MaxEnt	Stepwise selection (e.g. 'MaxEntVariableSelection')	Threshold probability	N/A	Presence-only data
Ensemble (RF, GBM, MARS, SVM, GAM)	Cross-validation	Independent data set (Continuous Boyce Index)	Predictive power	Presence-only data used e.g. Folmer et al., 2016
Ensemble (GLM, GBM, GAM, FDA, MARS, RF)	Cross-validation	Threshold probability	Predictive power	Presence-only data but added pseudo-absences. e.g. Chefaoui et al., 2016
Ensemble (BRT, MaxEnt, ANN, RF, SVM, GAM, GLM, MARS)	Multimodel Inference	Threshold probability	N/A	Valle et al., 2013

RF, Random Forests; GLM, Generalized Additive Model; GBM, Generalized Boosting Model; SVM, Support Vector Machines; BRT, Boosted Regression Trees; ANN, Artificial NeuralNetwork; MARS, Multivariate Adaptive Regression Splines; N/A, Not Applicable.

Recent studies (Jones et al., 2012; Jones and Cheung, 2015; Anderson et al., 2016; Ovaskainen et al., 2016) have shown that the combination of multiple Species Distribution Models (SDMs) offers a systematic approach to include important model characteristics, such as uncertainty and agreement levels, into the framework (Robinson et al., 2017). A paper by Bertelli et al. (2022) found ensemble models, using models added together and averaged, more reliable for predictions. Therefore, a useful approach could be to build a few models by selecting algorithms based on their assumptions and data inputs available, gain predictive outputs and build an ensemble model to then compare predictive robustness and performance of pre-selected modelling approaches.

Evaluation Methods

Qazi et al. (2022) conducted a systematic review on SDM methodology, outputs and shortfalls and concluded that with 81% of the studies, degree of uncertainty and error estimates were not disclosed. When model performance estimates are disclosed, the model results are more transparent and therefore more reliable, allowing for more assurance in the predictions they make (Qazi et al., 2022). Therefore, it is important to define evaluation methods prior to designing model architecture and publishing model results with full transparency.

Once an ecological niche model (ENM) is developed, it is important to assess how well it predicts species distribution. There are various methods to assess model performance, validation and accuracy. Model accuracy can be assessed using True Skills Statistics (TSS) and Kappa coefficient (Allouche et al., 2006), while model performance can be assessed using AUC or ROC metrics. These model performance metrics function as follows:

- Kappa coefficient measures agreement between model predictions and observed species distribution, considering the possibility of chance alone. Kappa coefficient ranges from -1 to 1, where a value of 1 indicates perfect agreement between model's predictions and observed distribution (Cohen, 1960; Fernandes, 2018; Xia, 2020). A value of 0 indicates agreement that is no better than chance, whereas a value of -1 suggests a disagreement worse than chance (Cohen et al., 1960; Guisan et al., 2017; Zurrel et al., 2020).
- TSS is a statistical metric that considers a model's ability to predict the proportion of correctly predicted presences of a species (sensitivity) and proportion of absence (specificity) (Allouch et al., 2006).
- Receiver Operating Characteristic (ROC) can offer a visual representation of how model performance varies as the threshold for classifying a presence or absence occurrence is adjusted (Zurrel et al., 2020).
- Finally, Area Under the Curve (AUC) values are used to summarise the overall performance of the model across all possible thresholds. The AUC values range from 0 to 1, where a value of 1 indicates a prefect model, that reliably discriminates between absence and presence. A value of 0.5 or less indicates a model that performs no better than random, as it is unable to distinguish between presence and absence (Swetts, 1988; Guisan et al., 2017; Zurrel et al., 2020).

It is important to note that Guisan et al. (2017) concluded that TSS was found to be independent of prevalence, where Kappa shows a unimodal response to prevalence, therefore suggesting Kappa values may be nothing more than a statistical artefact, in addition to the TSS metric showing a decreasing linear response, which could be interpreted as reflecting true ecological phenomena and therefore not a statistical artefact. A similar pattern was also found for ROC values. This further highlights the need for multiple methods when assessing model performance and accuracy.

Generally, performance should be assessed with respect to the aim of the application and to the response variable. For presence–absence data, we may distinguish threshold-independent measures such as the AUC and ROC (Swets, 1988; Guisan et al., 2017). Response shapes are one of the most important outputs of SDMs as they summarise the estimated species–environment relationship and can thus be directly subjected to plausibility checks against available biological knowledge (Zurrell et al., 2020).

Habitat Suitability Modelling in the Deep Sea

SDM provides a cost-effective means of identifying potential deep-sea coral and sponge (DSCS) habitat over large areas to inform management decisions and data collection (Winship et al., 2020). Using three case studies encompassing DSCS habitat in the Pacific, Winship et al. (2020) concluded that, when possible, models fitted to presence-absence and density data are preferred over models fitted only to presence data, which are difficult to validate and can confound estimated probability of occurrence or density with sampling effort (Winship et al., 2020). Ensembles of models can provide robust predictions, while multi-species models leverage information across taxa, and facilitate community inference (Winship et al., 2020, Bertelli et al., 2022).

In SDM literature, there are several examples of using multiple models in an ensemble (Robert et al., 2016, Rooper et al., 2017). In some cases, models have been constructed at different spatial scales; for example a PICES working group has been conducting broad-scale modelling of corals and sponges for the entire North Pacific Ocean (Rooper et al., 2020). This is an effort that is fairly data-poor for some regions due to limited data acquisition and uses only presence data. However, within the North Pacific Ocean there are several Exclusive Economic Zones (EEZs) and seamounts for which better models have been produced. For example, Miyamoto et al. (2017) constructed models for the Emperor Seamount chain. It would be beneficial to assemble these smaller scale and larger scale models to come up with better overall predictions (Davies and Roberts, 2018).

Conclusion

There is no one-box-fits-all modelling approach or algorithm selection when building an SDM in the field of ecology. However, is it of pivotal importance to carefully consider the data available, model species physiology and drivers of its distribution, in addition to scale of study and extent, before designing model architecture to reliably answer the research question or meet project goals adequately.

Critical steps of the modelling process are:

- In particular, to specify the model objectives in addition to defining the focal organism(s) with a degree of taxonomic certainty.
- Consider the type of biodiversity data, type of environmental predictor variables, spatiotemporal scale of the analyses and the underlying hypotheses about the biodiversity–environment relationship.
- Critically evaluate model assumptions, chosen SDM algorithms, desired model complexity, and, lastly, the software used.
- Furthermore, using robust and multiple evaluation techniques. Ensemble models and those using presence-absence data sets, where available, are currently accepted to be sound modelling approaches to better inform management and conservation decisions.

In addition to the accessibility of VHR DEMs, it has become an important challenge to test whether an increase in the resolution of a DEM and the associated variables (e.g., topographic and climatic variables calculated at a VHR) allows improvement of the predictive power of SDMs compared to models computed at a standard resolution. With a rise in accurate remote sensing and cost-effective sampling of some of the deepest areas on Earth, SDMs of previously inaccessible species has never been more exciting. However, it is still of great importance to apply firm rigour in following good modelling practices in deep-sea models as to those in shallow and terrestrial environments.

The rugosity or complexity of the seafloor has been shown to be an important ecological parameter for fish, algae, and corals. SfM photogrammetry therefore provides an interesting opportunity to examine in the role of in producing VHR predictor data inputs for SDMs and their potential for increasing predictive performance of deep sea SDMs, which may in the future be utilised for making sound and accurate management decisions in relation to marine spatial planning, fisheries and beyond.

References

Amatulli, G., Domisch, S., Tuanmu, M. N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific data, 5(1), 1-15.

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Peterson, A. T., & Soberón, J. (2016). Are species occurrence data in global online repositories fit for modeling species distributions? The case of the Global Biodiversity Information Facility (GBIF). Final Report of the Task Group on GBIF Data Fitness for Use in Distribution Modelling. Global Biodiversity Information Facility (GBIF).

Appah, J. K. M., Lim, A., Harris, K., O'Riordan, R., O'Reilly, L., & Wheeler, A. J. (2020). Are non-reef habitats as important to benthic diversity and composition as coral reef and rubble habitats in submarine canyons? Analysis of controls on benthic megafauna distribution in the Porcupine Bank Canyon, NE Atlantic. Frontiers in Marine Science, 7, 571820.

Barnard, S., & Boyes, S. J. (2013). JNCC Report No: 490.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O., Swartz, B., Quental, T. B., ... & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived?. Nature, 471(7336), 51-57.

Barton, M. G., Clusella-Trullas, S., & Terblanche, J. S. (2019). Spatial scale, topography and thermoregulatory behaviour interact when modelling species' thermal niches. Ecography, 42(2), 376-389.

Bowden, D. A., Anderson, O. F., Rowden, A. A., Stephenson, F., & Clark, M. R. (2021). Assessing habitat suitability models for the deep sea: is our ability to predict the distributions of seafloor fauna improving?. Frontiers in Marine Science, 8, 632389.

Baussant T, Arnberg M, Lyng E, Ramanand S, Bamber S, Berry M, et al. (2022) Identification of tolerance levels on the cold-water coral Desmophyllum pertusum (Lophelia pertusa) from realistic exposure conditions to suspended bentonite, barite and drill cutting particles. PLoS ONE 17(2)

Bellamy, C., Boughey, K., Hawkins, C., Reveley, S., Spake, R., Williams, C., & Altringham, J. (2020). A sequential multi-level framework to improve habitat suitability modelling. Landscape ecology, 35(4), 1001-1020.

Bertelli, C. M., Stokes, H. J., Bull, J. C., & Unsworth, R. K. (2022). The use of habitat suitability modelling for seagrass: A review. Frontiers in Marine Science, 9, 997831.

Bowden, D. A., Anderson, O. F., Rowden, A. A., Stephenson, F., & Clark, M. R. (2021). Assessing habitat suitability models for the deep sea: Is our ability to predict the distributions of seafloor fauna improving? Frontiers in Marine Science, 8, 632389.

Burns JHR, Delparte D, Gates RD, Takabayashi M (2015) Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. Peerj 3:e1077.

Chase, J. M., & Leibold, M. A. (2003). Comparing Classical and Contemporary Niche Theory. In Ecological niches: linking classical and contemporary approaches (pp. 51-59). Chicago: University of Chicago Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1), 37-46.

Davies, A. J., & Roberts, E. M. 3.6. GlobENV: Towards a High-resolution Climatology for the Seafloor. Use of Species Distribution Modeling in the Deep Sea, 34.

Dormann et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. – Ecography 36: 27–46.

Ehlers, M., Gaehler, M., & Janowsky, R. (2006). Automated techniques for environmental monitoring and change analyses for ultra high resolution remote sensing data. Photogrammetric Engineering & Remote Sensing, 72(7), 835-844.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A Working Guide to Boosted Regression Trees. Journal of Animal Ecology, 77(4), 802-813.

Elith Jane, and Franklin Janet (2013) Species Distribution Modeling. In: Levin S.A. (ed.) Encyclopedia of Biodiversity, second edition, Volume 6, pp. 692-705. Waltham, MA: Academic Press

Fitzpatrick, M. C., Lachmuth, S., & Haydt, N. T. (2021). The ODMAP protocol: a new tool for standardized reporting that could revolutionize species distribution modeling. Ecography, 44(7), 1067-1070.

Franklin, J., Davis, F. W., Ikegami, M., Syphard, A. D., Flint, L. E., Flint, A. L., & Hannah, L. (2013). Modeling plant species distributions under future climates: How fine scale do climate projections need to be? Global Change Biology, 19(2), 473-483.

Friedman A, Pizarro O, Williams SB, Johnson-Roberson M (2013) Correction: Multi-Scale Measures of Rugosity, Slope and Aspect from Benthic Stereo Image Reconstructions. PLOS ONE 8(12): 10.1371

Gogol-Prokurat, M. (2011). Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. Ecological Applications, 21(1), 33-47.

Grinnell, J. (1917). Field tests of theories concerning distributional control. The American Naturalist, 51(602), 115-128.

Guillera-Arroita.2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. – Ecography 40: 281–295

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). Habitat suitability and distribution models. With applications in R. Ecology, biodiversity and conservation.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., ... & Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. Ecology letters, 16(12), 1424-1435.

Hirzel, A. H., & Le Lay, G. (2008). Habitat suitability modelling and niche theory. Journal of applied ecology, 45(5), 1372-1381.

Jones, C. C. (2012). Challenges in predicting the future distributions of invasive plant species. Forest Ecology and Management, 284, 69-77.

Jones, M. C., & Cheung, W. W. (2015). Multi-model ensemble projections of climate change effects on global marine biodiversity. ICES Journal of Marine Science, 72(3), 741-752.

Ketchington et al. (2019). Use of Species Distribution Modeling in the Deep Sea. Canadian Technical Report of Fisheries and Aquatic Sciences, 3296, ix + 76 p.

Kramer-Schadts. et al. 2013. The importance of correcting for sampling bias in maxent species distribution models. – Divers. Distrib. 19: 1366–1379.

Lauria, V., Garofalo, G., Fiorentino, F., Massi, D., Milisenda, G., Piraino, S., ... & Gristina, M. (2017). Species distribution models of two critically endangered deep-sea octocorals reveal fishing impacts on vulnerable marine ecosystems in central Mediterranean Sea. Scientific reports, 7(1), 8049.

Lecours, V., Brown, C. J., Devillers, R., Lucieer, V. L., & Edinger, E. N. (2016). Comparing selections of environmental variables for ecological studies: A focus on terrain attributes. PLoS ONE, 11(12), e0167128.

Lecours, Vincent, et al. "Towards a framework for terrain attribute selection in environmental studies." Environmental Modelling & Software, 89, 19-30.

Melo-Merino, S. M., Reyes-Bonilla, H., & Lira-Noriega, A. (2020). Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. Ecological Modelling, 415, 108837.

Meyer. et al. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. – Ecol. Lett. 19: 992–1006.

Miyamoto, M., Kiyota, M., Hayashibara, T. and Nonaka, M. 2017. Megafaunal composition of cold-water corals and other deep-sea benthos in the southern Emperor Seamounts area, North Pacific Ocean. Galaxea, Journal of Coral Reef Studies 19: 19–30.

Norberg, A. et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. – Ecol. Monogr. 89: e01370.

Oldfield, F. (2005). Environmental change: key issues and alternative perspectives. Cambridge University Press.

Ovaskainen, O., Roy, D. B., Fox, R., & Anderson, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. Methods in Ecology and Evolution, 7(4), 428-436.

Park and Davis. C.2017. Implications and alternatives of assigning climate data to geographical centroids. – J. Biogeogr. 44: 2188–2198.

Peterson, A. T., & Soberón, J. (2012). Species distribution modeling and ecological niche modeling: getting the concepts right. Natureza & Conservação, 10(2), 102-107.

Phillips et al. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – Ecol. Appl. 19: 181–197.

Pradervand, J. N., Dubuis, A., Pellissier, L., Guisan, A., & Randin, C. (2014). Very high resolution environmental predictors in species distribution models: Moving beyond topography? Progress in Physical Geography, 38(1), 79-96.

Qazi, A. W., Saqib, Z., & Zaman-ul-Haq, M. (2022). Trends in species distribution modelling in context of rare and endemic plants: a systematic review. Ecological Processes, 11(1), 1-11.

Martínez-Quintana, Á., Lasker, H. R., & Wilson, A. M. (2023). Three-dimensional species distribution modelling reveals the realized spatial niche for coral recruitment on contemporary Caribbean reefs. Ecology Letters, 26(9), 1497-1509.

Renner. W. et al. (2015). Point process models for presence-only analysis. – Methods Ecol. Evol. 6: 366–379.

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E., & Lundquist, C. J. (2017). A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. Frontiers in Marine Science, 4, 421.

Rooper, C. N., Georgian, S. E., Guinotte, J. M., & Watling, L. (2020). 10. Studies of the distribution and diversity of biogenic habitat forming taxa in the USA. Report of Working Group 32 on Biodiversity of Biogenic Habitats, 88.

Ross, L. K., Ross, R. E., Stewart, H. A., & Howell, K. L. (2015). The influence of data resolution on predicted distribution and estimates of extent of current protection of three 'listed' deep-sea habitats. PloS ONE, 10(10), e0140061.

Stein, A., & Kreft, H. (2015). Terminology and quantification of environmental heterogeneity in species-richness research. Biological Reviews, 90, 815–836.

Storlazzi, C. D., Dartnell, P., Hatcher, G. A., & Gibbs, A. E. (2016). End of the chain? Rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (SfM) technology. Coral Reefs, 35(3), 889-894.

Winship, A. J., Thorson, J. T., Clarke, M. E., Coleman, H. M., Costa, B., Georgian, S. E., ... & Whitmire, C. E. (2020). Good practices for species distribution modeling of deep-sea corals and sponges for resource management: data collection, analysis, validation, and communication. Frontiers in Marine Science, 7, 303.

Wright, P. G. R., Bellamy, C., Hamilton, P. B., Schofield, H., Finch, D., & Mathews, F. (2021). Characterising the relationship between suitable habitat and gene flow for Myotis bechsteinii and Eptesicus serotinus in Britain. Landscape Ecology, 36(12), 3419-3428.

Wu, B. M., Subbarao, K. V., Ferrandino, F. J., & Hao, J. J. (2006). Spatial analysis based on variance of moving window averages. Journal of phytopathology, 154(6), 349-360.

Xia, Y. (2020). Correlation and association analyses in microbiome study integrating multiomics in health and disease. Progress in Molecular Biology and Translational Science, 171, 309-491.

Ward, G., Hastie, T., Barry, S., Elith, J., & Leathwick, J. R. (2009). Presence-only data and the EM algorithm. Biometrics, 65(2), 554-563.

Wright, P. G. R., Bellamy, C., Hamilton, P. B., Schofield, H., Finch, D., & Mathews, F. (2021). Characterising the relationship between suitable habitat and gene flow for Myotis bechsteinii and Eptesicus serotinus in Britain. Landscape Ecology, 36(12), 3419-3428.

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., ... & Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, *43*(9), 1261-1277.

Site description and investigation of the use of SfM photogrammetry for generating ultra-fine scale variables with species distribution modelling of *Lophelia pertusa* at various spatial scales in Fangorn Bank, Northeast Atlantic

Mestrado em Biologia Marinha

Supervisors:

Dr. Chris McGonigle

Co-supervisor

Dr. Aschwin Hillebrand Engelen

Abstract

High Resolution (HR) bathymetry derived from MBES data and High Definition (HD) video data was acquired on the CE21010 ‘Resources of Rockall’ expedition. Primary data from this expedition was used in this study for predicting suitable habitat of Cold-Water Coral (CWC) *Lophelia pertusa*. A workflow for rendering very fine scale DEMs using SfM photogrammetry techniques is outlined in this study using Agisoft Metashape. Records of occurrence were developed using ROV video footage annotations of live coral framework observations. A suite of environmental layers were generated from HR bathymetry data using TASSE Toolbox in ArcGIS Pro and used as environmental data layers for the modelling procedure. The Boosted Regression Tree (BRT) technique was used for the creation of various models at broad scale, fine scale and ultra fine scale resolutions. Shrinkage and cross-validation were used as constraints to prevent overfitting and monotonic constraints were used for reducing model complexity. In all models *L. pertusa* showed a preference for high terrain complexity and topographic highs. Depth ranges for species occurrence were between -1,000 m and -800m. The fine scale model performed best with an AUC value of 0.970, and a sensitivity scoring of 0.980 and specificity of 0.965. Broad scale model obtained an AUC value of 0.951 and a sensitivity score of 0.875, with a specificity of 0.939 and an overall TSS value of 0.814. Accuracy of predictions for broad scale model was not as high as fine scale model, indicating that models trained with fine scale data produce more reliable predictions. Ultra fine scale (UFS) model obtained an AUC value of 0.834. Although UFS model performed the worst overall, the SfM photogrammetry technique presented here shows promise for future research generating UFS DEMs derived from video data to be used in species distribution models to improve accuracy of predictions.

Twenty-six distinct faunal groups were recorded during ROV annotation, along with four primary substrate types. A chain of deep-sea mound features were revealed through GIS analysis of MBES data and highest mound feature measured 2.4km in elevation. Fangorn Bank has been shown to be a biologically complex site underpinned by a high degree of terrain complexity of sloping rock highs and flat muddy bottoms which lead to the rich and diverse ecosystems found here. The identification of 26 faunal groups and their characteristics in this study site underscores its remarkable biodiversity and the significance of its preservation.

Key words: MBES data, ROV, SfM Photogrammetry, SDMs, CWCs, BRT

1. Introduction

Cold-Water Corals (CWC), which not only form habitats but also serve as indicators of vulnerable marine ecosystems (VMEs), face potential risks posed by human activities such as fisheries and oil/mineral exploration (Burns et al., 2016). To ensure the effective protection of VMEs, it is crucial to have a deep understanding of their habitat needs and where they are found (Sampaio et al., 2012; Armstrong et al., 2014). However, extensive sampling in the deep sea presents challenges due to accessibility and costs. *Lophelia pertusa* is a reef-forming scleractinian deep water coral that has received much attention within research and management due to its accessibility, wide extent globally, and its status as a species for deep-sea conservation (Davies and Guinotte, 2011). Their complex three-dimensional structures supply microhabitats for both sessile and mobile epifauna and are considered as 'biodiversity hotspots' (De Oliveira et al., 2021), as well as providing breeding grounds, shelter and protected feeding ground for numerous fish, such as redfish (*Sebastes* spp.), ling (*Molva molva*), and ray species. Their skeleton and tissue may also serve as a host to various endoparasites and cryptofauna, e.g., sponges, fungi and nematodes (Buhl-Mortensen et al., 2017). The scleractinian *Lophelia pertusa* (Linnaeus, 1758), now formerly synonymised to *Desmophyllum pertusum* (Linnaeus, 1758; Baussant et al., 2022) is widely distributed across a wide depth range globally. In this study however, the model species will be referred to as *Lophelia pertusa* for reasons of consistency with the data that was used to conduct this work.

In the deep sea, where data is sparse due to surveys being constrained by costs and technological capabilities, predictive mapping offers a powerful tool for such studies (Robert et al., 2016; Anderson et al., 2016). Predictive mapping relies on models that describe relationships between species and their environment, allowing the forecasting of where species are likely to be found, even in areas where direct sampling has not taken place, based on probability (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005). Therefore, Species Distribution Modelling (SDM) is frequently employed to estimate the general distribution patterns and ecological preferences of species, even when data is limited (Sundahl et al., 2020). As presence of CWC reefs and deep-sea sponge aggregations are generally indicators of VMEs, having information on their presence or absence can aid planning of marine protected areas (MPAs), thereby having implications in terms of marine spatial planning, or when advocating for construction of windfarms and offshore structures (Gilbert et al., 2015).

To make accurate predictions, it is essential to incorporate ecologically relevant environmental data collected at resolutions that capture the scale at which these variables influence species spatial patterns (Lecours et al., 2015). These techniques are rooted in the principles of niche theory, which asserts that species distributions are shaped by the environmental dimensions of their ecological niche (Guisan and Zimmermann, 2000). In the realm of deep-sea research, environmental factors derived from acoustics, such as depth and slope, are commonly used as indirect indicators for more direct resource-related variables (Pearman et al., 2020). Moreover, bathymetric maps enable the calculation of various environmental descriptors, including rugosity (a measure of topographic complexity), aspect (indicating the steepest slope's orientation), bathymetric position index (BPI, assessing a pixel's height relative to its surroundings), or Relative Distance from Mean Value (RDMV) when using the TASSE Toolbox (Lecours et al., 2016), and curvature (Jordan, 2007).

Ecologists have identified multiple measures of complexity as primary drivers of biodiversity in deep-sea ecosystems, with rugosity serving as a 3D complexity proxy (Burns et al., 2015). High-resolution data has proven essential for investigating ecosystem dynamics and capturing the nuanced characteristics of coral reefs (Ferrari et al., 2016). Thus, the incorporation of high-resolution mapping of these ecosystems plays a pivotal role in studying these intricate interactions (Jackson et al., 2014; Lim et al., 2020)

The utilisation of underwater imagery and Remotely Operated Vehicles (ROVs) has been on the rise for gathering biological, ecological, and geomorphological data (Ware and Downie, 2020). This technology not only facilitates increased data collection but also provides access to deeper ocean regions that were previously challenging to explore. It has opened opportunities for groundbreaking discoveries in ocean areas that were previously inaccessible to surveys.

In the past, reliance on images alone had limitations, potentially missing fine-scale details at survey sites. However, the advent of 3D imagery from video footage has significantly improved the precision of annotations and the overall data quality. This advancement offers essential insights into the physical and biological attributes of the seafloor, including their spatial and temporal distribution and abundance (Goes et al., 2019).

Recent advancements in remote sensing technologies and computational capabilities have ushered in a new era of data collection, enabling the creation of Very High Resolution (VHR) variables and digital elevation models (DEMs) at scales finer than a meter (Ehlers et al, 2006), using novel approaches such as Structure from Motion (SfM) photogrammetric techniques (Mcgready et al., 2023; De Oliveira et al,, 2023). While these technologies have primarily been applied in engineering disciplines, particularly for precise geomorphological assessments like flood calculations (Vaze et al., 2010), their potential extends into the realm of Species Distribution Models (SDMs) with significant promise (Pradervand et al., 2014).

By integrating VHR predictors into SDMs, we unlock the potential to greatly improve the precision and accuracy of these models. This is achieved by capturing subtle variations in the environment that were previously challenging for modellers to discern or confined to limited geographic areas, such as proxies of terrain complexity, i.e., rugosity (Le Roux et al., 2013a; Wang et al., 2012). In essence, the utilisation of VHR data expands our ability to model and understand species distributions in ways that were once beyond our reach.

The aim of this research was threefold. The first objective was to investigate whether scales of predictors and their resolution, when included in model training, have the ability to significantly improve prediction accuracy of species distribution when using *Lophelia pertusa* as the model species. The secondary objective was to investigate whether high definition (HD) video data can be used to derive very fine scale (VFS) very high resolution (VHR) variables for model training and projection to a study site within a small spatial extent through the utilisation of SfM photogrammetry software. The third objective was to generate annotations from video data in order to begin to describe species composition in a previously undescribed study site in the Northeast Atlantic, with the aim of shedding light on using SfM photogrammetry software as a tool for rendering data layers with more robust and accurate information needed to make sound predictions in the realm of SDM in the deep sea and the implications this could have for marine spatial planning.

2. Methods

2.1 Environmental Setting

Fangorn Bank is located within the marine region of Rockall in the NE Atlantic, situated approximately at latitude 55.5° and longitude -20.166°. This deep-sea marine feature is positioned about 125 km southwest of the Rockall Plateau. Fangorn Bank lies southwest of Rockall Bank and to the east of Rockall Trough, as documented in GEBCO's 2003 dataset (GEBCO, 2003).

The Rockall Plateau, found south of Iceland and west of the British Isles, comprises a thick layer of basalt and is part of the North Atlantic Igneous Province (NAIP), one of Earth's Large Igneous Provinces (LIPs), as noted by Wilkinson et al. in 2016 (Wilkinson et al, 2016; Hansen and Ganerød, 2022). Historically, it was interpreted as a microcontinental fragment (Roberts, 1975), formed during the sea floor spreading of the North Atlantic Ocean, however this theory has been disproved. Separating it from the British Isles is the deep Rockall Trough, which reaches depths of 3,000 meters (Roberts, 1975).

The western margin of the Rockall Trough, situated at the foot of the Rockall Plateau Continental Slope in water depths of 2,000 to 2,400 meters, experiences strong to moderate currents, as reported by Vermeulen (1996). Lonsdale and Hollister (1979) detail the primary sedimentary features and current dynamics of the Rockall Trough. Both its eastern and western margins exhibit current ripples, with the eastern margin characterised by a fast, narrow, erosive flow of Labrador Sea Water and Northeast Atlantic Deep Water (N. J. Vermeulen, 1996; Londsdale and Hollister, 1979).

Notably, East Rockall Bank holds the status of a designated Marine Protected Area (MPA). However, it is worth mentioning that as of 2021, no specific management plan has been established for Fangorn Bank according to OSPAR (OSPAR, 2021).

2.2 Data Acquisition

High Resolution (HR) bathymetry derived from MBES data and High Definition (HD) video data (Figure 4) was acquired on the CE21010 'Resources of Rockall' expedition between the 22nd of July and 13th August 2021.

The original survey objectives were to collect HD video footage and MBES data of specific mounds in Fangorn Bank which likely encompass vulnerable marine ecosystems (VMEs). Primary data used in this study was limited to Fangorn Bank only, a previously undescribed study site in the Northeast Atlantic, and the utilisation of Multibeam Echosounder (MBES) data and HD video footage was relevant to this study.

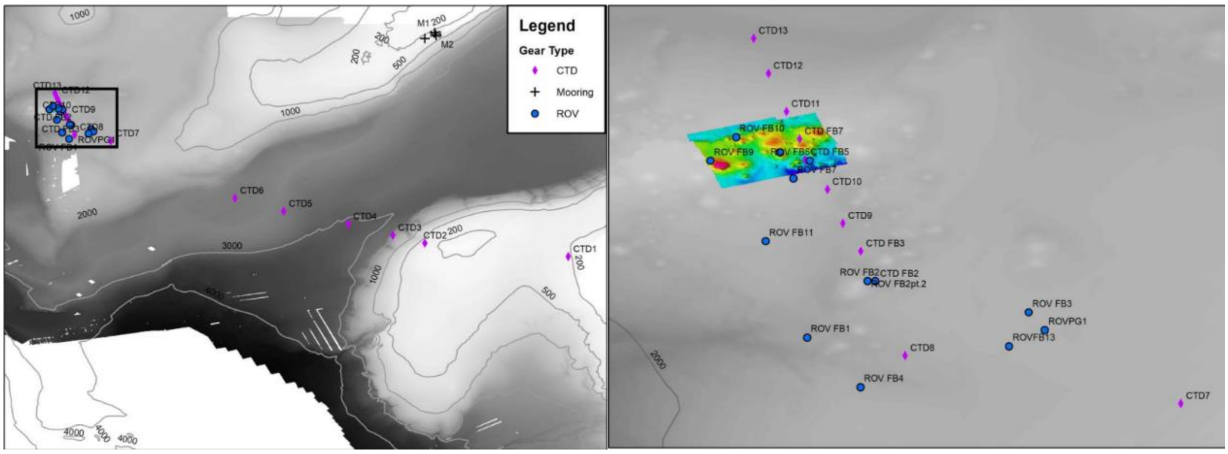

Figure 4. Sampling station maps obtained from CE2010 cruise report. Left: overview of full area. Right: inset of detailed area surveyed at Fangorn Bank including the newly acquired area of multibeam used for this study and location of ROV stations.

2.2.1 MBES

Data was obtained using a Kongsberg EM302 (30 kHz) system (1 [Tx] x 2 [Rx] degree) with the beam angles constrained to 35° for both the port and starboard sectors. Although this led to a reduction in the total achievable coverage, it was determined to be the best balance with the along-track data density as the dual ping system was in operation allowing for 864 effective soundings per ping at the survey speeds of between 2 and 3 ms^{-1}. Runtime parameters pulse form was set to CW, and all other parameters were constant throughout the survey. Sound velocity observations were taken daily Ami SVpLUS, and the Valeport SV probe at the transducer head when the deviance from the sound speed at head was different from the profile loaded. The data was collected to ensure 100% overlap between adjacent lines and crosslines were taken for error evaluation and verification of tidal corrections. The data were processed in QPS Qimera using CUBE algorithm (Calder, 2007), supplemented with manual cleaning to remove erroneous data points. ASCII gridded surfaces were exported and prepared for geomorphometric analysis in ArcGIS and R.

2.2.2 Video Data

The video data was collected using a *Mini-Zeus* downward facing HD camera positioned at 30º angle to horizontal from the seafloor mounted on the *Holland I* deep-water ROV. ROV footage was obtained and in total there were thirteen ROV transects, equating to 26 hours 47 minutes. In this study eleven have been analysed and annotated; details of the dives used for this study can be found in Table 2.

Table 2. Details of ROV stations in relation to video transects used for annotations

ROV Station	Date	Latitude (ºN)	Longitude (ºE)	Max Depth (m)
FB2	06/08/21	55.250	-19.823	1,503
FB2pt.2	08/08/21	55.250	-19.839	1,362
FB3	06/08/21	55.186	-19.510	1,524
FB4	06/08/21	55.034	-19.853	1,591
FB7	07/08/21	55.459	-19.989	1,354
FB5	07/08/21	55.495	-19.957	1,276
FB12	07/08/21	55.512	-20.017	924
FB9	07/08/21	55.495	-20.159	956
FB10	08/08/21	55.543	-20.106	1,144
FB11	08/08/21	55.331	-20.046	1,053
FB13	09/08/21	55.117	-19.550	1,492

2.2.3 GEBCO Legacy Data

Bathymetry data was obtained from GEBCO legacy data (GEBCO, 2022) encompassing the upper and lower bounds of 55.5425 ºN to 55.4235 ºN and -20.2313 ºE to -20.0701 ºE. Geographic bounds were chosen to replicate the geographic bounds of the HR bathymetry data obtained from primary cruise data. The GEBCO_2022 Grid is a comprehensive worldwide topographical model encompassing both ocean and land areas. It supplies elevation information in meters, organised on a grid with a 15 arc-second spacing, resulting in a grid size of 43,200 rows by 86,400 columns, equating to 3,732,480,000 data points (GEBCO, 2022). It's important to note that the data values pertain to the elevations in meters at the central points of the grid cells.

2.3 Data Processing

2.3.1 MBES

Multibeam Echosounder (MBES) data was processed with QPS Qimera using CUBE algorithm (Calder, 2007) by first cleaning and correcting raw sonar data for factors like water column interference and motion artifacts to account for yaw, pitch and roll of the ship. Processed data was used to create HR bathymetric maps of the study site in ArcGIS Pro.

2.3.2 SfM Photogrammetry Workflow

3D models of the study site were rendered by using a free 30-day trial of Agisoft Metashape software v2.0. Model site was selected by referring to the biological and geomorphological variation in substrata at a finer scale by observing ROV footage and referring to the detailed annotations yielded from the ROV footage.

ffmpeg software was used to extract frames from the raw Apple ProRes ROV footage at 1fps to obtain a 70-80% overlap. Once each frame was extracted, *ExifTool* software was used to give each frame a georeferenced latitude, longitude and elevation position, to produce more accurate models, due to EXIF data not being extracted with each frame by *ffmpeg*.

These georeferenced images were then imported into Agisoft Metashape software, checked for blurring or inconsistency and, if needed, removed. The software analyses the images, identifies scale invariant features ("key points") and extracts them from the overlapping images.

Photos were then aligned in steps going from medium accuracy to high and then super-high accuracy. Key point limit and tie point limits were kept at their default ranges within the software; 40,000 key point limit and 4,000 tie point limit. These key points were then matched and aligned to develop a photomosaic of the substrate.

The SfM software performs bundle adjustments to determine camera positions and construct a 3D point cloud which is then processed with soft-copy triangulation to reconstruct the scene geometry and create a solid 3D mesh that can be shaded or textured using the high-resolution photographs. The 3D point cloud was rendered in similar steps-ups of accuracy to the photo alignment procedure to increase accuracy and limit processing time. Finally, a mesh was generated, a texture rendered, and a digital elevation model (DEM) produced.

The workflow followed was in general accordance with the Agisoft Metashape Pro v2.0 handbook (Agisoft LLC, 2023).

2.3.3 Calculating Rugosity and Derivatives

A depth layer was derived from MBES data in ArcGIS Pro and gridded at a resolution of 10m. Seven terrain variables were chosen to be the primary used environmental data variables for this study and prepared in ArcGIS Pro. The geomorphometric variables rugosity/ruggedness/roughness (StDev), slope, relative distance from mean value (RDMV) and statistical aspect (northernness and easternness) (Lecours et al., 2016) were derived from the depth layer using TASSE Toolbox plug-in for ArcGIS Pro.

Standard Deviation (StDev) in the case of this toolbox is a measure of rugosity or roughness, a proxy for biological complexity and was computed with the TASSE Toolbox which utilises the Focal Statistics tool in ArcGIS Pro. Raster datasets were exported as ASCII files and used in the R programming language for the modelling procedure.

Table 3. Table of derivatives: rugosity (StDev), mean, slope, RDMV, easternness and northernness calculated utilising TASSE Toolbox utilising HR bathymetry data obtained from MBES providing information on resolution, type of data, source and units of measurement

Variable	Unit	Descriptor	Type of Data	Data Source
Mean	m	Statistical mean of depth information. Smooths out the DTM and potentially removes local errors.	Derived from bathymetry	TASSE
StDev	n/a	Standard deviation is a measure of rugosity/ruggedness/roughness. 0 indicates no rugosity, the higher the number the higher the rugosity.	Derived from bathymetry	TASSE
Slope	Degrees	Slope, in degrees, is measured using Horn's (1981) algorithm.	Derived from bathymetry	TASSE
RDMV	n/a	Measure of relative position that identifies peaks (positive values) and pits (negative values). It is unit-less and computed using a combination of focal statistics: RDMV = -(local mean – initial value of elevation or depth)/(local range).	Derived from bathymetry	TASSE
Easternness	Radians	Derived from aspect and converted into radians before calculating its sine (for easternness).	Derived from aspect (in degrees) that is computed using Horn's (1981) algorithm.	TASSE
Northernness	Radians	Derived from aspect and converted into radians before calculating its cosine (for northernness).	Derived from aspect (in degrees) that is computed using Horn's (1981) algorithm.	TASSE

Terrain complexity is strongly correlated to biodiversity in marine environments (McCormick, 1994; Sleeman et al., 2005). It is a ratio between the actual length (or area) along the undulating terrain and the straight-line distance or planar projected area (Friedman et al., 2013). Units are those of initial surface and lightest colour correlating to highest measure of seabed rugosity or complexity, darker colours representing the lowest measures of rugosity.

Slope identifies steepness at each cell of a raster surface and is measured in degrees. The lower the degree of slope value, the flatter the terrain; the higher the degree of slope value, the steeper the terrain.

Aspect identifies the maximum rate of change in value from each cell to its neighbours in the downslope direction (Addedrian et al, 2004); the values of the output raster are the compass direction of the aspect. Easternness and northernness are derivatives of aspect from west to east and measured in radians.

2.3.4 Calculating Rugosity and Derivatives for UFS data

DEMs were exported from Agisoft Metashape v2.0 at 1cm resolution. While the software was able to compute sub-centimeter values, file size, compute power and bandwidth limitations restricted the usable resolution to 1cm.

The DEM datasets from the 3D rendered model were exported from Agisoft Metashape as ASCII datasets, and then imported into ArcGIS Pro. A boundary polygon was drawn for clipping the geometry to the extent of the transect using the Clip Raster Tool. Finally, the derivatives were computed using TASSE Toolbox within ArcGIS Pro.

An attempt was made to clip both broad scale and fine scale datasets to same extent as UFS dataset. However, broad scale data was too coarse to be clipped to such a small extent and when fine scale data was clipped to this extent, TASSE Toolbox was not able to capture all information for each variable. It was only possible to calculate mean, StDev and RDMV (Figure 18) derived from HR bathymetry data and can be found in the supplementary material provided with this text.

2.3.5 Video Annotation and Classification

Annotation of ROV transects was carried out manually on a CSV file in Microsoft Excel with co-ordinating video timestamps and georeferenced ROV positions extracted from the USBL positioning system onboard the ship for underwater acoustic positioning of the ROV. All eleven video transects (Table 2) were annotated with a total of 26 hours 47 minutes.

Particular attention was taken to characterise seafloor and geomorphology of sample sites through adhering to a set of specific classes that are repeatable and follow a certain set of semantics in line with most recent updates of classification schemes by JNCC (2022).

Annotations were done following a hierarchal format, with particular attention paid to the primary substrate, secondary (or veneer) and fauna present. Annotations were done in line with the most recent updates of the marine habitat classification scheme by JNCC (2022) for the Atlantic mid bathyal region, which pertains to regions in the 600-1,300m depth range (JNCC, 2022), and recorded using a carefully developed set of classes (Table 4).

Table 4. Classing key for ROV annotations giving the class label codes used for manual annotations and a brief description

ID	Class code	Description	Notes
1	LCF	Live coral Framework	Coral presenting any signs of identifiable living parts (polyps or mucus-covered frameworks evident) although major proportions of the coral framework may be dead. Coral polyps, skeletal casing are usually bright, white or orange in colour (Appah et al., 2020)
2	SD	Sediment and dropstones	Sediment (sand or mud) and dropstones
3	DCF	Dead coral framework	Coral framework which has no identifiable living parts. Identified by darker gray or brown skeleton (Appah et al., 2020)
4	CR	Coral rubble	Coral rubble is recognisable by biogenic material where >50% of detached dead coral fragments, shell fragments and sediment is observed (Lim et al., 2017)
5	SD	Sediment and dropstones	Fine Sand/Mud
6	TM	Trawl mark	Tool marks
7	SP	Sponge aggregations	Massive, Cup-like, simple (CATAMI, 2014)
8	HC	Hexacoralia	Non framework building coral assemblage
9	OC	Octocoralia	Non framework building coral assemblage
10	B	Bioturbation	Visible evidence of crawling traces, burrows etc.
11	MS	Mixed sediment	Sand/mud/shells/pebbles/ripples/waves
12	BR	Bedrock	Mainly hard consolidated substrata

2.3.6 Coral Presence Data

Coral presence data was obtained from all primary ROV data from all dives during the CE2021 expedition on the Celtic Explorer. Presence records for the model species *Lophelia pertusa* were generated by extracting all georeferenced recorded presences from ROV transects. The presence of species was given an integer value of 1. Presences were recorded in a CSV file in Microsoft Excel, to be used as an attribute table in ArcGIS Pro to give a visualisation of presence of *L. pertusa* within the study site. Since this data is expressed in global coordinates, it was transformed into the UTM coordinate system, WGS 84 zone 27N, which is the appropriate zone for this study site. This transformation was carried out in R during the data preparation steps of the modelling procedure.

2.4 Analysis Framework

2.4.1 GIS – Geospatial Analysis

ArcGIS Pro was utilised to analyse processed MBES data for production of DEMs and yield a bathymetric map for the study site. The Define Projection tool was used to assign correct coordinate system used for the map to avoid geographic misalignment; WGS_1984_UTM_Zone_27N was used for all data sets. Processed MBES data was added to ArcGIS Pro and using the ‘symbology’ ribbon the ‘precipitation’ colour ramp was chosen and inverted to give greater definition to the depth profile. The spatial analysis tool was utilised for artificial illumination of the offshore landscape and to render the derived hillshade (HS) layer. An azimuth value of 315º and an altitude of 45º was set with a stretch type of ‘standard deviation’. The bathymetry layer was then overlain with the HS layer with a transparency of 44.9% to produce an illuminated DEM and further aid interpretation of the offshore landscape and provide depth information for the study site.

3D exploration of several mound features was carried out using the 3D Analysis Tool in ArcGIS Pro; elevation profiles were created and elevation of mound features and distance between mounds were measured in meters.

2.4.2 Scale of Analysis

Three scales of analysis were chosen for this study; broad scale, fine scale (FS) and ultra fine scale (UFS). Broad scale data used in this study pertained to the legacy GEBCO data gridded at 15 arc second resolution. Fine scale data was derived from MBES data and processed in GIS to obtain bathymetry data gridded at 10m resolution. Ultra fine scale data in this study was obtained from the SfM photogrammetry outputs obtained in Agisoft Metashape and gridded at a 1cm resolution. The seven terrain variables chosen to be used as environmental layers for this study were prepared in ArcGIS Pro for all three scales of data.

2.5 Modelling in R

Helper functions written by Assis (2023) were used for the purposes of this research at commit ID dfef8a6 (Assis, 2023). The script has been modified for this study and the full script can be found in the supplementary material. The R libraries which were used are marmap, plyr, devtools, Rcpp, ENMeval, plotROC, dismo, sp, rgdal, rgeos, raster, ggplot2, rnaturalearth, rnaturalearthdata, viridis, leaflet, leaflet.extras, robis, sdmpredictors, SDMtune, RColorBrewer and readr (RStudio Team, 2020). Data preparation steps for all environmental layers was also taken, like that of presence data, whereby all ASCII files were transformed into the WGS 84 UTM coordinate system, zone 27N, in R.

Records of occurrence of the model species were used along with raster data sets developed in ArcGIS Pro for modelling in R. Raster data sets obtained from HR bathymetry and geomorphometric derivatives obtained from TASSE Toolbox v 1.1 developed by Lecours et al. (2016) were used as environmental layers for model training and can be found in the supplementary material. Geomorphometric variables used were mean (depth), easternness, northernness, standard deviation (StDev), relative distance from mean value (RDMV), aspect and slope. Model training for this study was limited to ROV data analysed from stations FB13, 12, 11, 9, 7 and 5, due to modelling not being carried out in this area.

2.5.1 Model Selection

Boosted Regression Trees (BRT) are one of several techniques that aim to improve the performance of a single model by fitting many models and combining them for prediction. The BRT technique was used to build a series of models of decision trees iteratively with the environmental layers in the training dataset from data points that were misclassified by the previous tree, until the data points converge, or number of trees is reached (e.g. 1,000) (Elith et al., 2008), with the intention to increase predictive performance due to this ensemble model method of combining multiple decision trees. In comparison, other widely used techniques in ecology such as Maximum Entropy (MaxEnt) or Generalised Linear Modelling (GLM) do not use this ensemble model technique. The boosting approach used in BRT places its origins within Machine Learning (ML) (Schapire, 2003), but further developments in the statistical community reinterpret it as an advanced form of regression (Friedman et al, 2002). BRT shows promise in this method due to its ability to handle complex datasets with non-linear relationships between environmental variables and occurrence of model species (Yu et al., 2020, Hallman and Robinson, 2020).

Similarly to GLM, BRT models can be fitted to a variety of response types (Gaussian, Poisson, Bernoulli, binomial, etc.) by specifying the error distribution and the link. In this case Bernoulli distribution has been selected to enable a probability outcome i.e. presence probability (Gormley et al., 2011). However, BRT models can be prone to overfitting, and this is best understood in the context of other model-fitting practices. For all prediction problems, overfitting models to training data reduces their generality, so regularisation methods are used to constrain the procedure so that it balances model fit and predictive performance (Hastie et al. 2001). In this case shrinkage and cross-validation have been used as constraints to prevent overfitting of the model, known as model tuning.

2.5.2 Monotonic Constraints

Monotonic constraints are useful for reducing model complexity by 'hinting' to relationships between certain variables in the model (Stegmann et al, 2022). Monotonicity can be neutral (0), increasing (+1) or decreasing (-1). In the case for L. pertusa, research indicates that the likelihood of presence (and therefore +1) can be selected for mean depth, rugosity (StDev) and slope. However, due to the unclear relationships between other variables and less degrees of certainty, a parametric study to iteratively test the remaining four monotonic constraints was executed with the aim of improving model predictions (see supplementary material for the used script). This was carried out with one of the datasets and the conclusions taken forward.

Table 5. Parametric study of BRT monotonicity constraints: RDMV, northernness, easternness derived from aspect

Slope	Mean	StDev	RDMV	Aspect	N'ness	E'ness	TSS	Sens.	Spec.	AUC
1	1	1	-1	-1	-1	-1	0.945	1.000	0.945	0.991
1	1	1	-1	-1	-1	0	0.990	1.000	0.990	0.989
1	1	1	-1	-1	-1	1	0.945	1.000	0.945	0.989
1	1	1	-1	0	-1	-1	0.945	1.000	0.945	0.989
1	1	1	-1	0	-1	0	0.990	1.000	0.990	0.989

2.5.3 Hyperparameter Tuning

Tuning of hyperparameters was also adjusted within the model with the aim of producing the highest possible quality outputs. The hyperparameters which could be adjusted were distribution, number of trees (the number of gradient boosting iterations), interaction depth (the maximum number of nodes per tree), shrinkage (learning rate) and bag fraction (subsampling fraction) (Elith et al, 2008). To generate the best performing model, the following parameters were tested:

- Interaction depth: 1, 2, 3 and 4.
- Shrinkage: 1E-2, 1E-3 and 1E-4.
- Bag fraction: 0.5, 0.6 and 0.7.
- Number of trees: 1000 and 2000.

The hyperparameter permutation with the highest value of AUCtest was taken forward for training the model. The bag fraction and number of trees were left at their default values of 0.5 and 1000 respectively after a parametric study. A similar approach was carried out by Lecour et al. (2016) where models with higher AUCtest and lower standard deviations were found to be more robust and present higher predictive capacity, whereas models with higher AUCtrain fitted better the training data while models with low AUCdiff are more generalisable (Lecours et al., 2016). Models with a high AUCtrain and low AUCdiff are therefore the most generalisable and brought forward to train the model in this study.

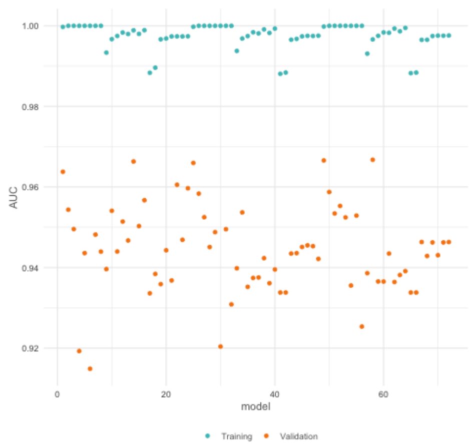

Figure 5. Example grid search of hyperperameter permutations of models, blue depicting AUCtrain and red depicting AUCtest

2.5.4 Variables of Highest Permutation of Importance

Variables of highest permutation of importance for each model (broad scale, fine scale and ultra fine scale) were tested with the 'varImp' function in R and tested for collinearity. Values >0.85 are expected to affect the model negatively and are therefore not defined as variables of highest permutation of importance and can be removed from the model (Gevrey et al., 2003). Examples of these are given in Figure 6 below.

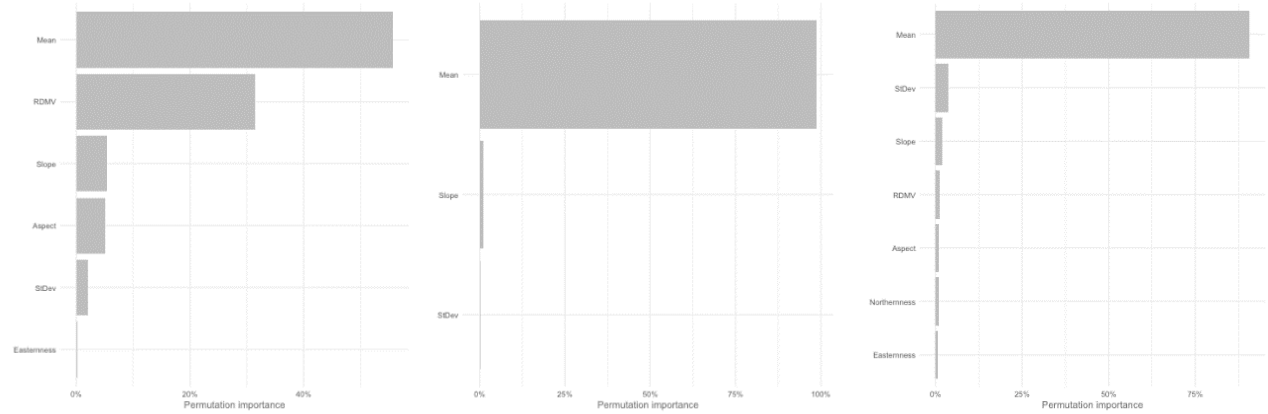

Figure 6. Variables of highest permutation of importance for broad scale, fine scale and ultra fine scale models

The variables mean, RMDV, slope, aspect, StDev and easternness in models trained with broad scale data were deemed to have the highest degree of permutation of importance and only northernness was removed when training the model. However, for the fine scale model with data obtained from the survey data, only mean, slope and StDev were used to train the model, all other variables were removed, greatly reducing complexity of the model. No variables were removed for training of ultra fine scale model.

2.5.5 Model Validation

For validation, True Skills Statistics (TSS) to assess model accuracy was selected for its ability to predict the proportion of correctly predicted presences of a species (sensitivity) and proportion of absence (specificity) (Allouch et al. 2006).

Finally, binomial models were tested for sensitivity by using the area under the curve method (AUC). AUC values are used to summarise the overall performance of the model across all possible thresholds. The AUC values range from 0 to 1, where a value of 1 indicates a perfect model, one which reliably discriminates between absence and presence. Conversely, a value of 0.5 or less indicates a model that performs no better than random, as it is unable to distinguish between presence and absence (Swets, 1988; Guisan et al. 2017; Zurrel et al. 2020).

Data was portioned using the *k-folds* function in R into training and testing data. All binomial models for broad scale (GEBCO), fine scale (FS) data (derived from MBES data) and ultra fine scale (UFS) data (derived from Agisoft Metashape) were then used to train each model. Broad scale data was used to train a model and then the predictions were projected onto the broad scale GEBCO bathymetry map. FS and UFS data were then both individually used to train two more models and predictions projected onto broad scale GEBCO data to investigate transferability of model to a widely available data source and testing data used to validate each model utilising AUC metrics.

Best performing models were then used to predict occurrence of species within the extent of the study site, along with superimposed actual occurrences using the *ggplot* and *viridis* packages.

3. Results

3.1 SfM Photogrammetry

The outputs yielded UFS DEMs and derived aspect and slope from rendering of 3D model in Agisoft Metashape and captured a finer resolution of terrain attributes. Depth DEM was further processed using TASSE Toolbox to obtain a suite of geomorphometric derivatives akin to broad scale and fine scale for use as environmental layers in the SDM.

Figure 7 below shows the results of the DEM as generated in Agisoft Metashape. The left side of the figure shows the DEM of 3D rendered model in Agisoft Metashape software, with deeper areas of DEM in blue colour to the highest depth elevation in red. The middle of the figure shows the aspect of 3D rendered model in Agisoft Metashape software ranging from 0 to 90 degrees. 0 degrees (dark blue) indicates true north, and 90 degrees (red) indicates an eastwardly aspect. The right of the figure shows the slope of 3D rendered model in Agisoft Metashape software with scale bar in degrees showing direction slope is facing in degrees: 0 degrees (blue) indicating degree of slope faces directly north, 180 degrees (green) indicating degree of slope is facing south and 270-360 degrees (yellow to orange) indicating degree of slope is facing west.

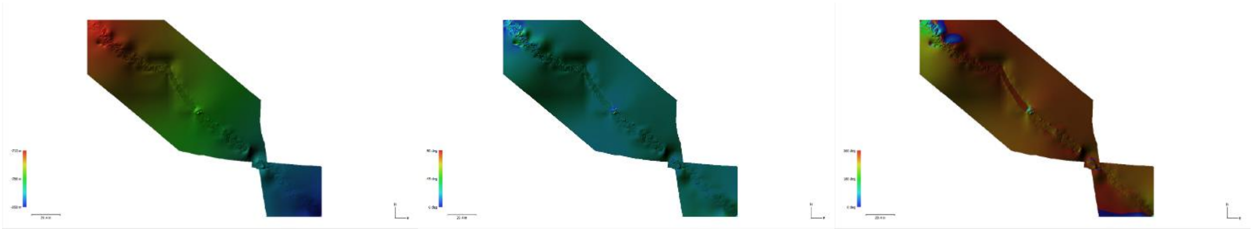

Figure 7. DEM of 3D rendered model in Agisoft Metashape software (left to right: depth, slope and aspect)

Deepest area of transect was -858m and shallowest was -715m, showing the fine detail of depth variation captured. Aspect of transect was facing a northerly direction while slope mainly faced westward. Terrain variations showed many undulations in seabed and were captured to a centimeter resolution with boulder like structures, coral rubble and framework becoming much more defined than in that of HR bathymetry.

3.2 MBES and Derivatives

Figure 8 below shows a map of HR Bathymetry data derived from MBES data on research cruise CE 21010 and position of study site in Fangorn Bank. The position of the ROV transect used for rendering DEMs in Agisoft Metashape is shown in the bottom of the figure as is the scale bar showing the maximum and minimum depths in the area of the study site. The HR bathymetry data obtained fell into the geographic marine region of Fangorn Bank 55° 30' 0" N (55.500°) 20° 10' 0" W (-20.166°) with topographic undulations of deep-sea mounds and flatter plains. The maximum depth of the study site was -1,408.59m below sea level and minimum depth was -655.35m.

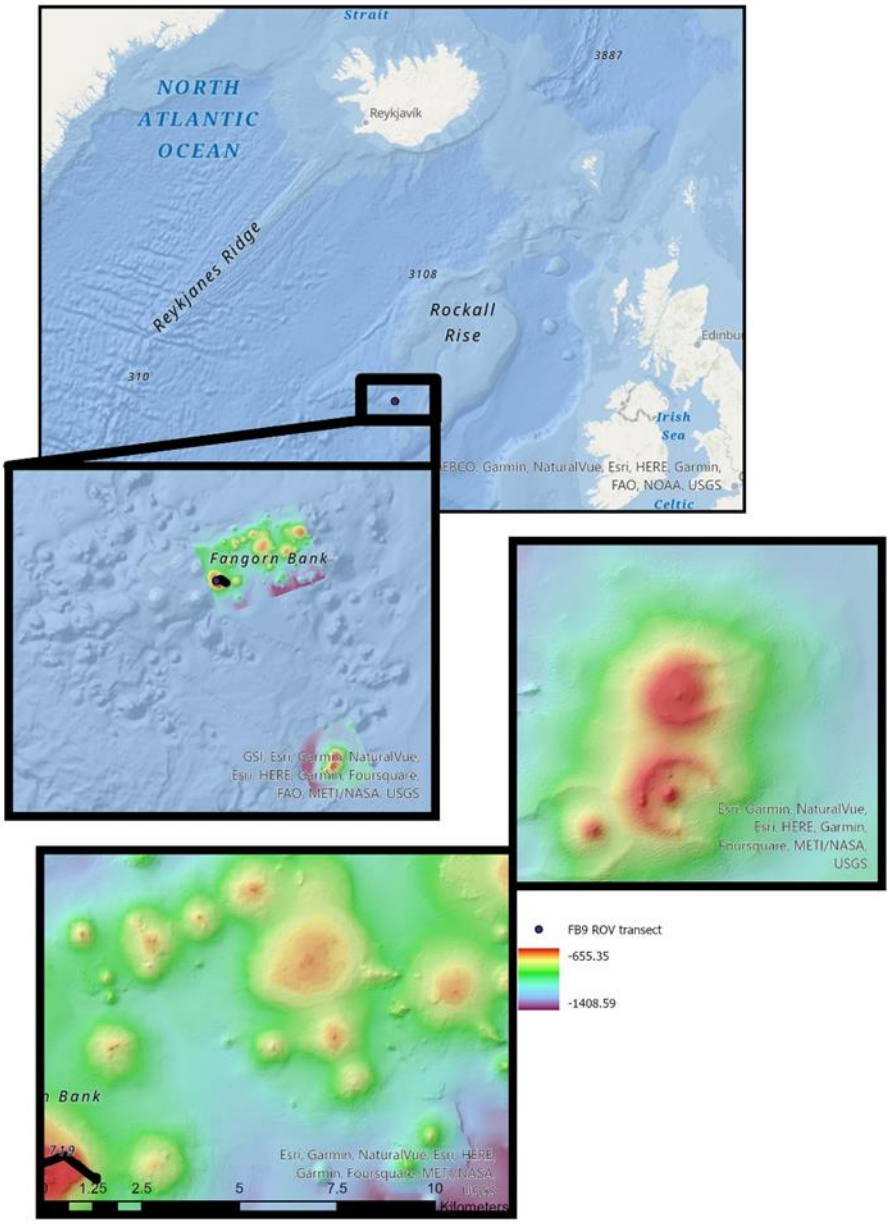

Figure 8. Map of HR bathymetry data derived from MBES data on research cruise CE 21010 and position of study site in Fangorn Bank; see text for further information

The geomorphometric derivatives obtained from TASSE Toolbox using ArcGIS Pro used as predictors in modelling in R for fine scale data are shown below in Figure 9. These are aspect, easterness and northernness as derivatives of aspect, RDMV, StDev (as a proxy of rugosity) and slope.

Figure 9 below shows that higher slope values appear to correspond with measures of StDev at the study site, suggesting that the greater the slope, the greater the ecological complexity. Rugosity increases on flanks of elevated geomorphological features and carbonate mounds, while many summits of mounds and elevated features appear to have particularly high rugosity (StDev), representing highest areas of biological complexity in this study site.

Ledges to the NW flanks also show high StDev, however this could be a consequence of processing of the MBES due to the slight repetition of the mound feature, in essence doubling the apparent rugosity. However, the summit of this feature still has a high level of rugosity in its singular form. The highest value of StDev is around 30, representing higher levels of rugosity and lowest is 0, representing no rugosity. Units are those of initial surface; the lightest colour indicates the highest measure of seabed rugosity or complexity, with darker colours representing the lowest measures of rugosity.

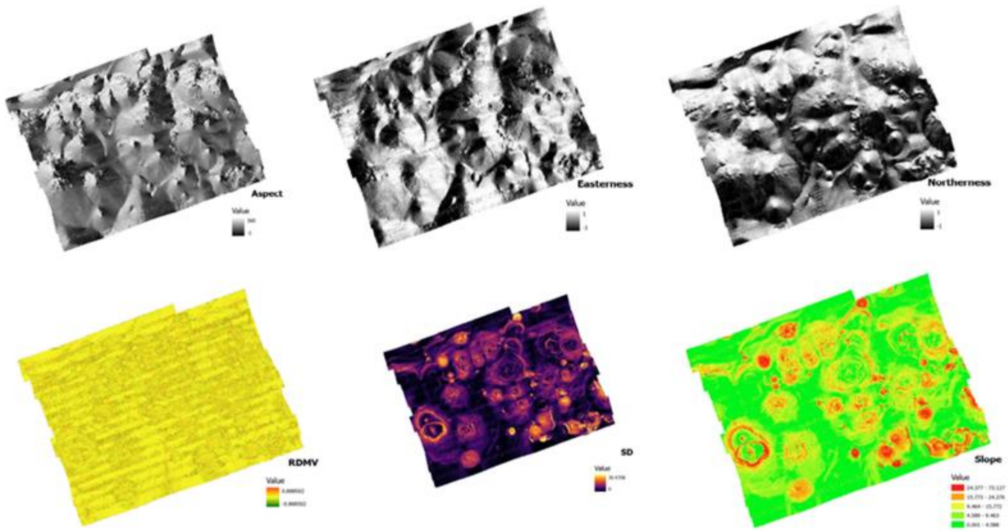

Figure 9. Geomorphometric derivatives obtained from TASSE Toolbox using ArcGIS Pro; see text for further information

Mounds at Fangorn Bank were proven to be sizeable features undulating in height and distance between each other. The highest mound measured was a staggering 2.4km in elevation from the seabed, while another was nearly 2km high. Distance of small sub sections of underwater mounds were also measured, ranging from 8.52km and 6.7km distance.

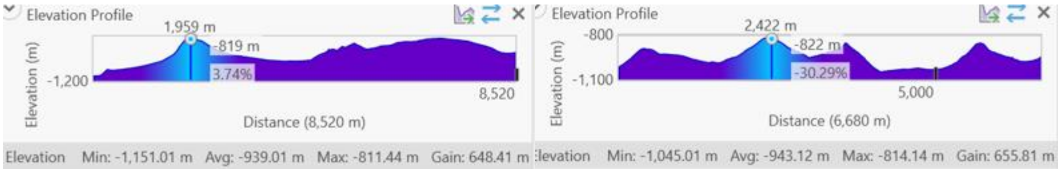

Figure 10. Elevation profiles measured from the seabed of small sub sections of deep-sea mounds in Fangorn Bank yielded from a 3D analysis in ArcGIS Pro and measured in meters

3.3 ROV Annotations

Primary substrates included sediment and drop stones (SD), mixed sediment (MS) and bedrock (BR). Secondary substrates generally included live coral framework (LCF), dead coral framework (DCF) and coral rubble (CR), referring to that of the model species *L. pertusa*. For this study, particular attention was paid to annotations that were that of model species presence so at the third level LCF, DCF were noted along with the presence of any other species in the *Hexacoralia* (HC) or *Octocoralia* (OC) classes, along with presence of large cup-like sponges (SP). Notes on fauna were recorded to the lowest possible morphological taxonomic level. Notes were also recorded on anthropogenic evidence e.g., trawl marks (TM) and those of bioturbation (B). Examples of footage annotated can be found in Figure 11.

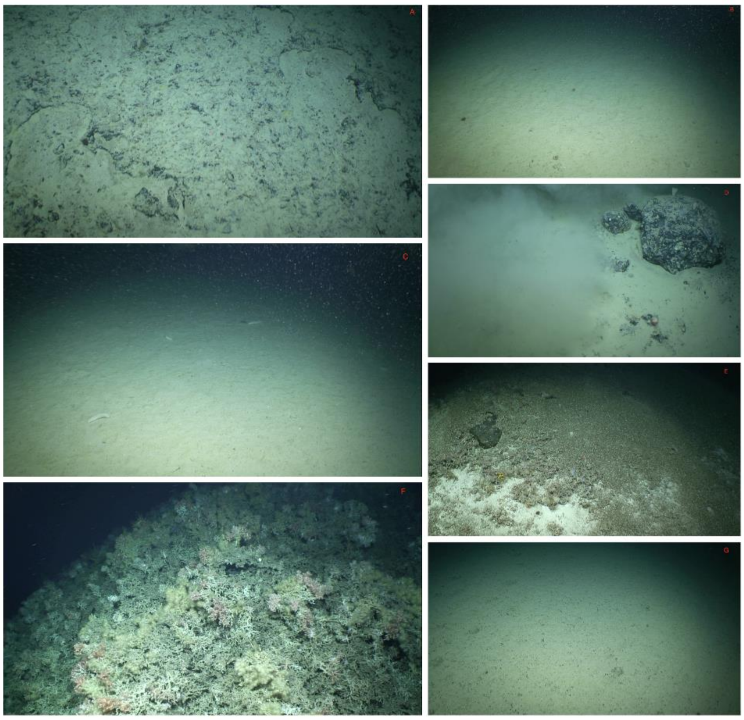

Figure 11. ROV images showing organisms and substrata encountered and annotated using classification scheme in Table 4: A) classed as bedrock (BR), B)Mixed sediment (MS) sand/mud/shells with signs of an anemone field, C) Mixed sediment (MS), D) Sediment and drop stones (SD) with simple cup like sponge, E) Area of dead coral framework (DCF) and (CR) on sediment and drop stones (SD), F) area of Live coral framework and L. pertusa reef, G) Mixed sediment (MS)

3.3.1 Substrate

Sediment and Dropstones (SD) was the most common substrate type encountered at Fangorn and accounted for 56.43% of the substrate in the ROV footage annotated. Meanwhile Bedrock (BR) accounted for 7.11%, Coral rubble (CR) accounted for 35.77% and Mixed sediment (MS) accounted for 0.68% of the substrate annotated in the ROV footage.

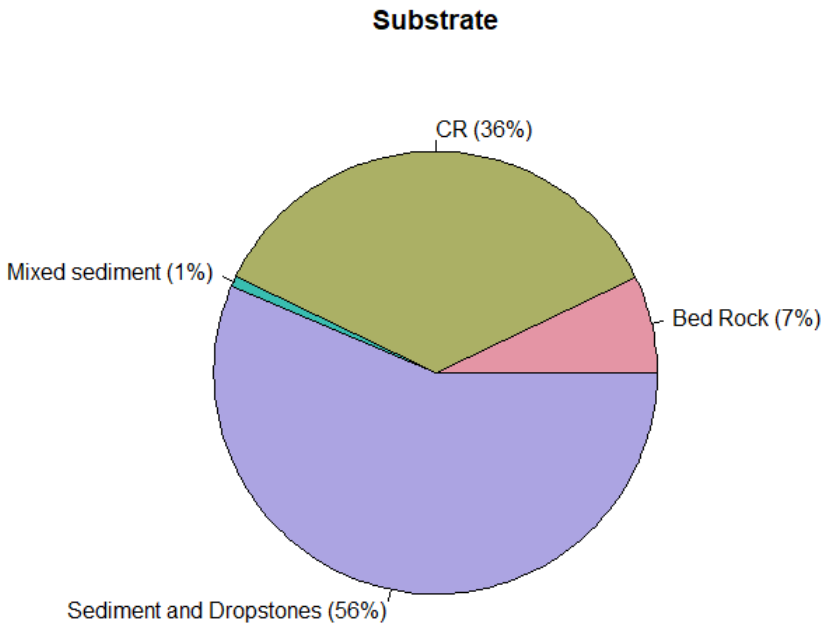

Figure 12. Pie chart showing proportion of primary substrate types annotated in ROV footage consisting of sediment and dropstones, coral rubble (CR), bed rock and mixed sediment

3.3.2 Fauna

Analysis in R of the ROV footage at Fangorn Bank revealed a diverse array of marine life, comprising 26 distinct faunal groups. These groups were identified during annotations to the lowest possible morphotaxa level, for providing insights into the underwater ecosystem. Porifera contributed to the largest proportion of faunal groups annotated (33.72%) at Fangorn Bank; live *Lophelia pertusa* coral framework (LCF) contributed to 23.03%, *Octocoralia* (OC) contributed to 24.26% and *Hexacoralia* (HC) contributed to 13.13% of the total data annotated within faunal groups at Fangorn Bank. All other groups were detected significantly less ranging from *Neprops*, *Pennatulacea*, *Anthomastus* and *Leiopathes*. It is worth noting OC and HC groups were of corals where it was not possible to identify to any lower morphotaxa level, thus incorporating more of the data than individual groups.

Table 6. Table showing each faunal group and their overall percentage contribution to total fauna annotated at Fangorn Bank; Hexacoralia (HC), Octocoralia (OC), LCF pertaining to that of L. pertusa

Group	**Percentage (%)**
Anthomastus	0.02
Bathypathes	0.05
Callegorgia	0.01
Cephalopoda	0.02
Clavularia	0.08
Echinodermata	0.06
Actinoscyphia aurelia	0.12
Hexacorallia (HC)	13.11
Isididae	3.06
Keratoisis	0.01
Lophelia pertusa LCF	23.00
Lepidisis	0.01
Madrepora	0.01
Neprhops	0.05
Octocorallia (OC)	24.23
Ophuiroid	0.13
Paragorgia	0.51
Paramuricea	0.07
Pennatulacea	0.38
Pheronema	0.09
Pisces	0.02
Polychaete	0.03
Porifera	33.63
Gorgonaceae	0.04
Solenosmilia	1.17
Stichopathes	0.07

3.3.3 Coral Presence Data

2,434 records of occurrence were obtained and with only live framework being recorded as presences and not of dead structures or rubble, similar to the approach by Howell et al. (2011). When presence data was loaded into ArcGIS Pro as an attribute table and converted from a table to points on bathymetry layer, a majority of occurrences were recorded at topographic highs), at summits of mound features or on sloping highs towards the summit of mounds; see Figure 13.

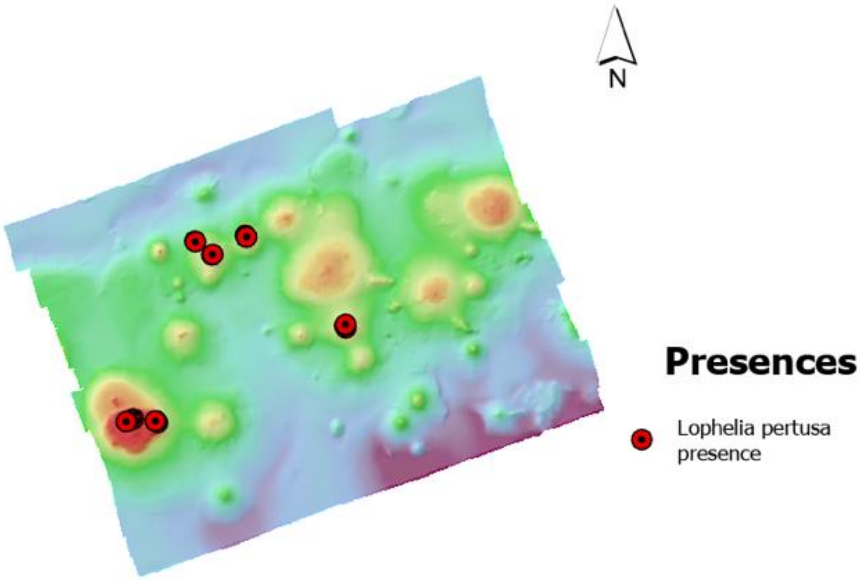

Figure 13. Presences of live coral framework obtained of Lophelia pertusa from ROV annotations as table to point data imported onto HR bathymetry data in ArcGIS Pro

3.4 Modelling in R

3.4.1 Response Curves

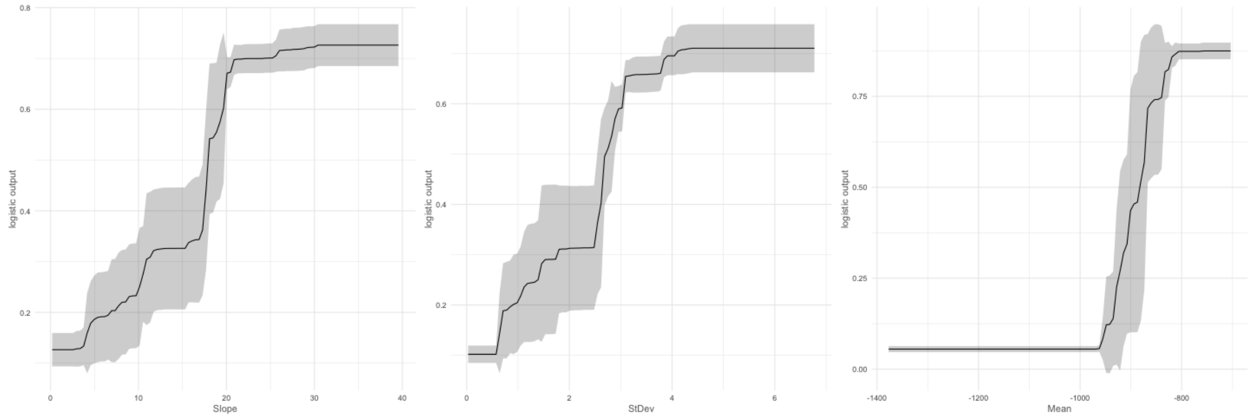

Figure 14. Model response curves for variables of highest permutation of importance for fine scale model; slope in degrees, StDev (rugosity) and mean (depth) in meters

As degree of slope increases, probability of occurrence of *L. pertusa* increases, similarly as rugosity increases, so does probability of occurrence. When a 30-degree slope is reached probability levels off and a threshold is reached. A threshold of probability of occurrence of *L. pertusa* is reached at +4 for rugosity, however as rugosity increased so did probability. With mean depth, probability of occurrence increases after -1,000m and increases in probability until around -800m where a threshold is reached for the probability of occurrence of *L. pertusa*. Similarly, for depths in the range of -1,400m to -1,000m there is little to no probability of occurrence of *L. pertusa*.

3.5 Model Performance

Table 7. Model performance values for each model showing True Skills Statistics (TSS) and Area Under the Curve (AUC)

Model	Threshold	TSS	Sensitivity	Specificity	AUC
Broad scale	0.0100	0.8140	0.8750	0.9390	0.9511
Fine scale	0.3400	0.9452	0.9802	0.9650	0.9697
Ultra Fine Scale	0.6300	0.8990	0.9840	0.9150	0.8338

Overall, the fine scale model performed best with an AUC value of 0.9697, and with a sensitivity scoring of 0.9802 and specificity of 0.9650, showing the model performed much better than random. Broad scale model performed well with an AUC value of 0.9511 and a sensitivity score of 0.8750, with a specificity of 0.9390 and an overall TSS value of 0.8140, indicating the model performs better than random. However, accuracy of predictions is not as high as the fine scale model, indicating that models trained with fine scale data produce more reliable predictions. Ultra fine scale model preformed the worst with an AUC value of 0.8338.

3.6 Reclassified Maps

Reclassified maps are reproduced in the sections below showing mean probability of occurrence of model species, ranging from absolute 0 probability (no presence) to 1 probability of occurrence (highest probability of occurrence) along with superimposed actual occurrences (red dots).

3.6.1 Reclassified Maps with Broad Scale Training Data

The broadscale projection showed a higher probability of *L. pertusa* occurrence in some areas where there are known occurrences, however not with a high degree of certainty due to probability being less than 0.1 (Figure 15). Other areas which coincide with topographic highs of the mound features (Figure 8) also have higher probability scores, but still only between 0.1 and 0.2, indicating that the model is only predicting a 10% chance of *L. pertusa* presence.

By projecting the broadscale model onto the fine scale data, higher levels of probability of up to approximately 0.3 were obtained in areas where the model species is known to occur. This appears to indicate that higher resolution data results in higher degrees of confidence in the predictions.

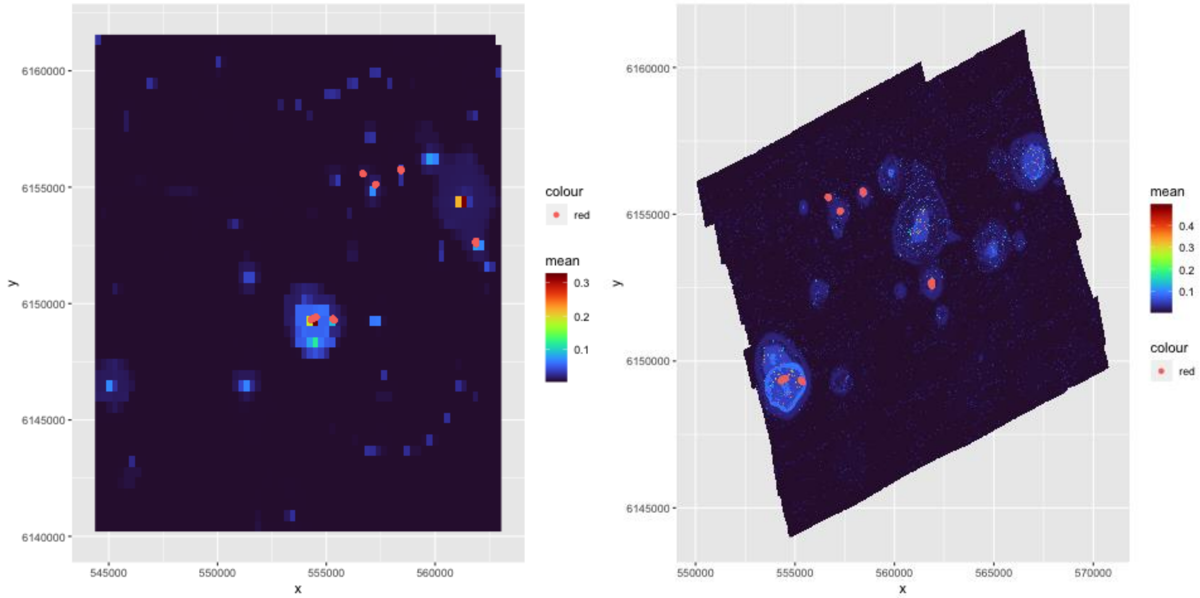

Figure 15. Broad scale trained models: model trained with broad scale data projected onto broad scale data (left), model trained with broad scale data projected onto fine scale data (right) with scalebar showing mean probability of L. pertusa occurrence and red dots actual occurrence verified by ROV annotations

3.6.2 Reclassified Maps with Fine Scale Training Data

The model trained with fine scale data had much higher probability values than that of broad scale trained models. The model trained with fine scale data projected onto broad scale data could predict one area where there would be an 80% chance of occurrence of *L. pertusa*, however the model was not able to adequately predict all areas where actual occurrence data was recorded.

The model trained with fine scale data projected onto fine scale data managed to predict three areas correctly where actual occurrences were recorded. This model also predicted additional areas which had not been surveyed to have an 80% chance presence of model species present. All areas that show a high probability of occurrence for *L. pertusa* are areas that are shown to have topographic highs within the study site, illustrated by the MBES data (Figure 8).

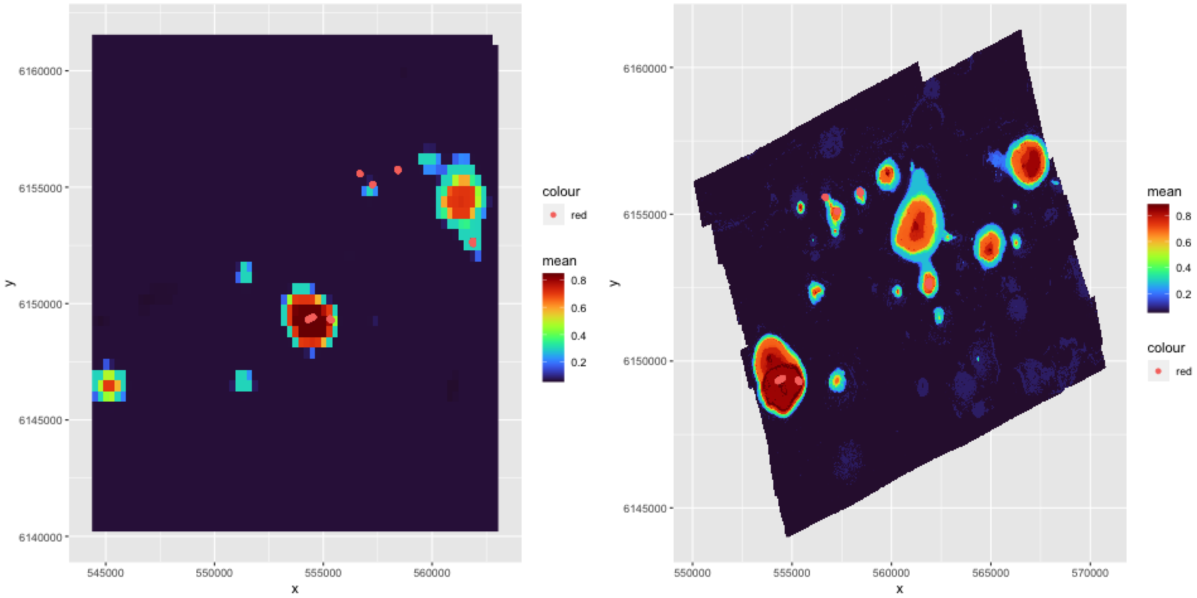

Figure 16. Fine scale trained models: model trained with fine scale data projected onto broad scale data, model trained with fine scale data projected onto fine scale data

3.6.3 Reclassified Maps with Ultra Fine Scale Training Data

There is a high degree of overfitting when the small transect of ultra fine scale data is used to train the model and projected onto fine scale data and broad scale data. The model predicted with a high degree of accuracy that there is >80% probability *L. pertusa* will occur in the area that we know that presences exist, however, in most other areas there is a 40% chance or less. When ultrafine scale data was projected onto broad scale data the model could not predict anything with a probability score higher than 0.28. In areas where we know *L. pertusa* is present the model predicted in the range of 0.26 to 0.28 for probability of occurrence of model species and the degree of error is high.

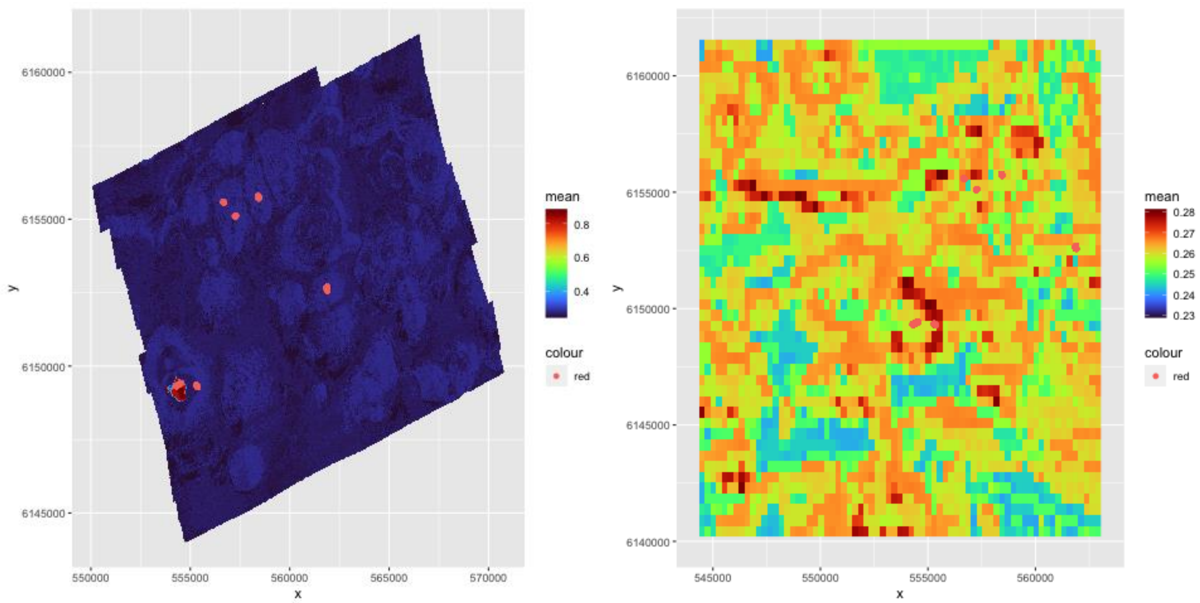

Figure 17. Ultra fine scale trained models: model trained with ultra fine scale data projected onto fine scale data, model trained with ultra fine scale data projected onto broad scale data

4. Discussion

Fangorn Bank is an impressive geomorphological marine region hosting a large group of elevated sea mounds, with the highest mound reaching a significant 2.4km in height. These mounds have a variation of high terrain complexity and encompass a wide range of substrate types. A total of 26 faunal groups were detected for annotation from the footage at Fangorn Bank during analysis of the video data illustrating this as an important and rich study which is both biologically diverse and encompassing a range of ecosystems. Four main primary substrate types (bedrock, coral rubble, mixed sediment and sediment with drop stones) were also identified through analysis of ROV footage. *Lophelia pertusa* presences were found mainly on elevated summits of mound features or on sloping flanks leading to the summit of mounds.

Broad scale model performed poorly in terms of yielding any predictions with probability higher than 0.2 for presence of *L. pertusa* (Figure 15) but overall the model statistically performed well with an AUC value of 0.9510 (Table 7). Fine scale trained model performed best with an AUC value of 0.9697 and yielded areas with 0.8 mean probability of occurrence of *L. pertusa* (Figure 16) for areas where actual occurrences were recorded and additional areas that have not been surveyed yet. The fine scale trained model also showed promise for projecting to areas where only broad scale data is available (Figure 16). This could be useful in situations where only broad scale data is available outside the study site and extrapolations of where *L. pertusa* could be present is needed in the general area for marine spatial planning or MPA design.

It was not possible to project the ultra fine scale (UFS) model to other data with any high degree of predictive accuracy (Figure 17). However, predictive probabilities are high in the extent used to train the model, where the model gave 80% chance or higher of *L. pertusa* presence (Figure 17). UFS model performed worst overall yielding an AUC value of 0.8338 (Table 7), however this is likely to be caused by proportionally less data being used to train the model than that of broad scale or fine scale models due to data and compute power limitations. UFS model still showed promise for use in SDMs as it was still able to predict with a high degree of accuracy where actual coral occurrences were within the extent used to train the model (Figure 17). It is the author's opinion that this shows promise for the use of VHR DEMs generated from the SfM photogrammetry techniques outlined in this paper, albeit at a smaller extent. This approach could therefore also be useful for characterising microhabitats and species distribution modelling within such microhabitats, as shown by the research of Martinez Quintana et al. (2023).

4.1 ROV Annotations

Porifera along with *Octocoralia* and *Hexacoralia* as well as LCF groups were the most abundant faunal groups at the study site indicating it is an important area for CWCs and *Porifera* and could potentially be termed as a deep sea coral and sponge habitat (DSCS). The combination of co-existing cold-water corals and sponges are known to form complex reef structures (Miller et al., 2012) which in turn supports a rich community of suspension feeding fauna and thus play an important role as a refuge, feeding ground and nursery for various fish. This is reflected in the 26 faunal groups recorded during annotations. Furthermore their coexistence can begin to paint a picture of the functioning of this diverse ecosystem found at Fangorn Bank.

CWCs and sponges are also known to play an important role in organic carbon cycling in the ocean. A study by Cathelot et al. (2015) found that the high metabolic activity of CWC and sponge grounds has important implications for carbon cycling and benthic-pelagic coupling at a regional scale of the Træna MPA, and also at the wider scale of the Norwegian Continental Shelf. They found that CWCR and sponge grounds account for 78% of the total benthic respiration and consume 63% of the total OC export flux in the Træna MPA (Slagstad et al., 1999). CWC reefs and sponge gardens, such as those found at Fangorn, could potentially play a vital role in carbon cycling in this area of the Northeast Atlantic, and more research is needed to quantify their role in carbon cycling in this deep sea marine region and what implications this could have for surrounding areas.

Isididae, otherwise known as Bamboo coral, is in the subclass *Octocoralia* and can form extensive aggregations, known as coral gardens (Chimienti et al., 2019b) dwelling on soft bottoms at depths ranging between 115 and 1,650m (Lauria et al., 2017). It plays an important ecological role, similar to *L. pertusa*, increasing the three-dimensional habitat complexity of the otherwise flat bathyal bottoms with its candelabrum-like shape (Carbonara et al., 2020). *Isididae* and other CWCs constitute a relevant habitat for several fishes and crustaceans to feed and shelter, such as *Chimaeras*, *Ophuiroids*, *Echinodermata* and *Brachyura*. Their canopy influences the availability of resources and thus large, suspension feeding colonies of deep-water corals such as those found at Fangorn Bank can play a major role in pelagic-benthic coupling, which comprises a missing link in biological pump (Li and Wang, 2019). *Isididae* in particular, forms an essential habitat for *Nephrops* and red shrimp (Carbonara et al., 2020), therefore indicating that this species may not only play a vital role in creating diverse habitats in this marine region but also act as a medium for replenishing fish stocks, while creating a nursery and feeding ground for commercial fish species.

Other faunal groups identified in the ROV annotations include *Keratosis* which is a CWC and is associated with steep topography (Mortensen and Buhl, 2005). *Pennetuculacea* was also identified and is a species which is associated with sandy or muddy sediment which unlike many other sessile benthic fauna, do not need hard substrata to colonise and are generally found in areas of elevated current (Williams, 2011) due to acquisition of food supply. The presence of this group during the annotations was found in mixed sediment (MS) with a relatively sandy muddy bottom; their presence can also shed light on the hydrodynamic conditions in the area.

Lophelia pertusa live framework accounted for one of the highest proportions of overall groups annotated. Their reefs occur on hard substrata, which there is plenty of at this study site (Figure 12). This may be *Lophelia* rubble from an old colony or on glacial deposits (O'Connell and Collins, 2012). For this reason, *Lophelia pertusa* reefs can be associated with iceberg plough-mark zones (Hall Spencer and Stehfast, 2005) which may explain the formation of this richly diverse area. Similarly to much of the other fauna found here they are typically found in areas of elevated currents. They are associated with other hard corals such as *Madrepora oculata* and *Solenosmilia variabilis* (Rogers, 1999) which have been evidenced to occur at this study site, further deepening the rich complexity of the ecosystems found here. *Leiopathes* in the Order Antipatharia (Antipathes – Bushy or fan shaped) is associated with rocky substrata, characterised by a branched canopy and associated with *Madrepora oculata* and *Lophelia pertusa* while also providing essential habitat and increasing benthic biodiversity of an otherwise barren bathyl region. Occurrences are strongly influenced by hydrodynamic conditions and substrata. Their colonies have a fragmented spatial distribution (Lauria et al., 2021) indicating that Fangorn Bank could be an essential habitat for this species.

The black coral *Bathypathes* in the family *Schizopathidae* form Monopodial colonies, is often associated with hard substrata and are key components of hard substrata ecosystems in the deep sea and have been known to be associated with deep-sea mineral resources such as cobalt-rich crusts and polymetallic nodules (Molodstova et al., 2022). They are slow growing and have slow rates of repopulation (Yesson et al., 2017) while being indicators of VMEs thus providing evidence of the fragility and importance of this relatively small (in the grand scheme of things) area in the Northeast Atlantic.

The presence of these faunal groups illustrates the importance of the variation of substrate types and varied topography available at Fangorn to support and influence the diverse habitats constituting such a variety of benthic and pelagic fauna. Fangorn Bank has also been shown to house a range of habitat types including dense CWC reef structures, sponge grounds, hard substrata CWC gardens and muddy bottom sea pen fields.

4.2 Terrain Complexity Influencing Faunal Distribution

Terrain complexity serves as a representative indicator of seafloor heterogeneity and exhibits a positive correlation with biodiversity (Robert, 2014). In deep-sea environments, elevated terrain complexity contributes to the creation of variations in substratum composition (Stewart et al., 2014), and exposure to ocean currents (Jarnegren et al., 2014).

As a filter-feeding organism, *Lophelia pertusa* exhibits a preference for areas with heightened terrain complexity (Davies et al., 2009, Howell et al., 2011). This is solidified by findings by Davies et al., (2008) where they state 'Most observations of cold-water corals have been in areas with accelerated currents, such as on sloping topography and topographic highs' (Davies et al., 2008). These organisms tend to colonise topographic high points to take advantage of localised current patterns, thereby increasing their chances of encountering food (Lo Iacono et al., 2018). This preference for topographic highs is consistently observed in both broad scale and fine scale models, where the highest probability of the presence of our model species was identified on these elevated mound features, indicating that the models successfully in part were able to capture the ecological niche of *L. pertusa*. Slope, depth (mean) and rugosity (StDev) were selected for all models as variables with highest permutation of importance to predict presence of *L. pertusa* (Figure 14) further illustrating that these variables are key environmental conditions for defining habitat of *L. pertusa*.

In our study, increased probability of CWC occurrence was associated with areas of high terrain complexity (slope and rugosity) over the various spatial scales. It was no surprise that bathymetric mean was the variable with highest permutation of importance for all models, due to model species being a deep water benthic macrofauna. This observation was further corroborated by response curves of fine scale data where *L. pertusa* habitat is defined by depth ranges between -800m and -1,000m showing highest probability of occurrence (Figure 14). Similarly, as slope increased to 40 degrees probability of occurrence also increased. These response curves suggested that a threshold of persistence for *L. pertusa* is reached outside these value ranges and therefore can begin to explain the ecological niche of *L. pertusa* in our study site. Rugosity was seen to play an important role in defining habitat of *L. pertusa* as with 0 rugosity there was little to zero probability of occurrence (Figure 14). However the actual presence of *L. pertusa* creates rugosity and therefore biological complexity so it is a difficult link to unravel. In reference to the hypothesis that Davies et al. (2008) previously mentioned the fine scale model was able to adequately capture the ecological niche of this species and that terrain complexity has a positive influence of the faunal distribution of *L. pertusa*.

4.3 Scales of Data Used

Data for the fine scale model preformed best suggesting it performed better than ultra fine scale data. The transect used for training the ultra-fine scale model due to data processing steps was a small proportion of overall data which may have greatly impeded model performance. However SfM photogrammetry showed that it is possible to obtain VHR data from video data which can then be used to compute UFS topographical variables to be used in modelling. However, a greater dataset spanning a somewhat larger spatial extent, with more variation, should be used to investigate whether it prevents overfitting of the model and poor predictive performance. Due to the trade-off between cost and time it would be recommended to generate a smaller portion from numerous transects to cover more of the study site, yet limit processing time by rendering a selection of smaller models encompassing all substrate types in the annotations. However, this study does present the plausibility of the SfM technique to render ultra fine scale variables to be used in modelling, albeit on a smaller spatial scale within the actual transect used for model training. This shows promise for utilising this approach in generating SDMs in microhabitats and for investigating macrofaunal community relationships, albeit if they can be captured at the resolution within the video data being used for SfM technique.

This paper illustrates that through reverse engineering it is possible to produce ultra fine scale data with video data alone. This paper contains a possible workflow for SfM photogrammetry that could be reproduced with data from this study site for future studies. Even though the UFS data generated in this study was a relatively small proportion of what could have been processed there is enough evidence gathered from this study to show promise in this method for improving predictive accuracy within the realm of species distribution modelling and to warrant further investigation into this method.

Additionally, it is apparent that the fine scale model performed well in terms of AUC values, the model still did not adequately predict areas of actual occurrence in all cases and even less so on the broad scale data. However, the fine scale model predicted the probability of actual occurrences and absences with a higher degree of mere chance suggesting that training a model with fine scale data could infer use in wider areas of this geographic area outside the study site, even if only broad scale data was available. Paradoxically, the model cannot adequately predict all instances of occurrence and therefore absences, thus suggesting that using models trained on such data may have potential risks and implications if used in marine spatial planning or policy.

Future research could aim to model species distribution at nested scales of data. VHR rugosity at centimeter scale data obtained through the SfM workflow outlined in this study could be nested into other FS variables gridded at a 0.1km resolution for topographic variables such as aspect and slope, along with depth information. This could supply the possibility of making more accurate predictions by incorporating ecologically relevant environmental data at resolutions that capture the scale at which these variables influence species spatial patterns, as previously highlighted by Lecours et al. (2015).

4.4 Modelling Limitations

As CWC are suspension feeders, currents play an important role in their distribution due to increasing accessibility to food supply. No local scale current data was available for this study and topography was inferred as a proxy for current speed. A future study could consider carrying out the same modelling procedure, but with modelled localised current information as another environmental layer to investigate if this would improve model accuracy.

5. Conclusion

Modelling has become an indispensable tool in the fields of ecology and conservation, enabling informed decision-making in our rapidly changing world. This study highlights the importance of proper data selection, predictor choice, scale considerations, and evaluation methods for achieving accurate and reliable predictions. The fine scale data models showed an improvement in prediction quality over broad scale data but still had limitations in predicting occurrences correctly, emphasising the challenges and constraints of using species distribution modelling for policy and marine spatial planning. Despite the study's limitations in ultra-fine scale data, it was demonstrated that VHR data sources utilising ROV footage can be derived from SfM photogrammetry and holds promise for enhancing model precision, particularly in microhabitats and in deep-sea environments where data acquisition can be challenging. A detailed workflow for the derivation of UFS data predictors through Agisoft Metashape has been developed and detailed within this document. Further exploration of this method is recommended to enhance species distribution model predictions, especially when paired with other, species-specific predictors.

Fangorn Bank has been identified to encompass a high diversity of marine life while being host to a group of deep-sea mounds with mounds reaching heights of 2.4km in elevation from the seabed. This study shows evidence for the presence of sponge grounds, CWC reefs, muddy bottom sea pen fields and hard substrata CWC gardens at Fangorn Bank. This study provides evidence for the importance of this study site as a potential area for protection. Further site descriptions and research are warranted to better understand this previously undescribed study site's biodiversity and ecological dynamics. This study underscores the complex topographic variations and substrate diversity within Fangorn Bank, influencing and supporting a mix of ecosystems crucial for deep-sea functioning and resource production.

In a site-specific context, Fangorn Bank has been shown to be a biologically complex site underpinned by a high degree of terrain complexity of sloping rock highs and flat muddy bottoms which lead to the rich and diverse ecosystems found here. The identification of 26 faunal groups in this study site can begin to highlight its remarkable biodiversity and the significance of its preservation. In the broader context, Fangorn Bank may hold implications for carbon cycling within this marine region in the Northeast Atlantic and surrounding areas. CWC reefs and sponge grounds, like those found at Fangorn Bank, could play a critical role in regional carbon cycling in the deep-sea while playing an important role in benthic pelagic coupling. These habitats can account for a substantial portion of benthic respiration and organic carbon export flux, further highlighting the ecological importance of this marine region.

6. References

Agisoft LLC. (2023). Agisoft Metashape User Manual: Professional Edition, Version 2.0. Retrieved from [https://www.agisoft.com/pdf/metashape-pro_2_0_en.pdf]

Armstrong, C. W., Foley, N. S., Kahui, V., & Grehan, A. (2014). Cold water coral reef management from an ecosystem service perspective. Marine Policy, 50, 126-134.

Assis, J. (2023). courseMarineEcologicalModelling: Course:: Marine Ecological Modelling and Global Climate Change. GitHub. https://github.com/jorgeassis/courseMarineEcologicalModelling

Baussant, T., Arnberg, M., Lyng, E., Ramanand, S., Bamber, S., Berry, M., ... & Van Breugel, P. (2022). Identification of tolerance levels on the cold-water coral Desmophyllum pertusum (Lophelia pertusa) from realistic exposure conditions to suspended bentonite, barite and drill cutting particles. Plos one, 17(2), e0263061.

Burns JHR, Delparte D, Gates RD, Takabayashi M (2015) Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. Peerj 3:e1077.

Calder, B. R., & Wells, D. E. (2007). CUBE user's manual. URL: https://scholars.unh.edu/cgi/viewcontent.cgi?article=2217&context=ccom [Date Accessed 29/09/2023].

Carbonara, P., Zupa, W., Follesa, M. C., Cau, A., Capezzuto, F., Chimienti, G., ... & Maiorano, P. (2020). Exploring a deep-sea vulnerable marine ecosystem: Isidella elongata (Esper, 1788) species assemblages in the Western and Central Mediterranean. Deep Sea Research Part I: Oceanographic Research Papers, 166, 103406.

Chimienti, G., Mastrototaro, F., & D'Onghia, G. (2019). Mesophotic and deep-sea vulnerable coral habitats of the Mediterranean sea: overview and conservation perspectives. *Advances in the Studies of the Benthic Zone*, *20*.

Davies, A. J., & Guinotte, J. M. (2011). Global habitat suitability for framework-forming cold-water corals. PloS one, 6(4), e18483.

De Oliveira, L. M. C., Lim, A., Conti, L. A., & Wheeler, A. J. (2021). 3D classification of cold-water coral reefs: a comparison of classification techniques for 3D reconstructions of cold-water coral reefs and seabed. Frontiers in Marine Science, 8, 640713.

E. L. Jackson, A. J. Davies, K. L. Howell, P. J. Kershaw, J. M. Hall-Spencer (2014). Future-proofing marine protected area networks for cold water coral reefs. ICES Journal of Marine Science, Volume 71, Issue 9, November/December 2014, Pages 2621–2629.

Ferrari, R., McKinnon, D., He, H., Smith, R. N., Corke, P., González-Rivero, M., ... & Upcroft, B. (2016). Quantifying multiscale habitat structural complexity: a cost-effective framework for underwater 3D modelling. Remote Sensing, 8(2), 113.

Gilbert, A. J., Alexander, K., Sardá, R., Brazinskaite, R., Fischer, C., Gee, Varjopuro, R. (2015). Marine spatial planning and Good Environmental Status: a perspective on spatial and temporal dimensions. Ecology and Society, 20(1).

Goes, E. R., Brown, C. J., & Araújo, T. C. (2019). Geomorphological classification of the benthic structures on a tropical continental shelf. Frontiers in Marine Science, 6, 47.

Gormley, A. M., Forsyth, D. M., Griffioen, P., Lindeman, M., Ramsey, D. S., Scroggie, M. P., & Woodford, L. (2011). Using presence-only and presence–absence data to estimate the current and potential distributions of established invasive species. Journal of Applied Ecology, 48(1), 25-34.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. Ecological modelling, 135(2-3), 147-186.

Guisan, A. and Thuiller, W. (2005) Predicting Species Distribution: Offering More than Simple Habitat Models. Ecology Letters, 8, 993-1009.

Hansen, J., & Ganerød, M. (2022). On the timing and nature of magmatism in the North Atlantic Igneous Province: New implications from basaltic rocks of the Faroe Islands.

Hall-Spencer, J., & Stehfest, K. (2009). Background Document for Lophelia pertusa reefs. Background document for Lophelia pertusa reefs.

Howell, K. L., Holt, R., Endrino, I. P., & Stewart, H. (2011). When the species is also a habitat: comparing the predictively modelled distributions of Lophelia pertusa and the reef habitat it forms. *Biological Conservation*, *144*(11), 2656-2665.

Howell K. L. , Davies J. S. , Jacobs C., Narayanaswamy B. E. , Broad scale survey of the habitats of Rockall Bank and mapping of Annex I “Reef” habitat, 2009 JNCC Report 422.

Jordan, G. (2007). Digital terrain analysis in a GIS environment. Concepts and development. In Digital terrain modelling: development and applications in a policy support environment (pp. 1-43). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lauria, V., Garofalo, G., Fiorentino, F., Massi, D., Milisenda, G., Piraino, S., ... & Gristina, M. (2017). Species distribution models of two critically endangered deep-sea octocorals reveal fishing impacts on vulnerable marine ecosystems in central Mediterranean Sea. Scientific reports, 7(1), 8049.

Lecours, V., Dolan, M. F., Micallef, A., & Lucieer, V. L. (2016). A review of marine geomorphometry, the quantitative study of the seafloor. Hydrology and Earth System Sciences, 20(8), 3207-3244.

Lecours, V. (2015) Terrain Attribute Selection for Spatial Ecology (TASSE) v. 1.1.

Li, J., & Wang, P. (2019). Discovery of deep-water bamboo coral forest in the South China Sea. Scientific Reports, 9(1), 15453.

Lim, A., Wheeler, A. J., Price, D. M., O’Reilly, L., Harris, K., & Conti, L. (2020). Influence of benthic currents on cold-water coral habitats: a combined benthic monitoring and 3D photogrammetric investigation. Scientific Reports, 10(1), 1-15.

Lim, A., Wheeler, A. J., & Conti, L. (2020). Cold-water coral habitat mapping: trends and developments in acquisition and processing methods. Geosciences, 11(1), 9.

Lonsdale, P., & Hollister, C. D. (1979). Near-bottom traverse of Rockall Trough-Hydrographic and geologic inferences. Oceanologica Acta, 2(1), 91-105.

Martínez-Quintana, Á., Lasker, H. R., & Wilson, A. M. (2023). Three-dimensional species distribution modelling reveals the realized spatial niche for coral recruitment on contemporary Caribbean reefs. *Ecology Letters*, *26*(9), 1497-1509.

McCormick MI (1994) Comparison of field methods for measuring surface topography and their associations with a tropical reef fish assemblage. Marine Ecology Progress Series 112: 87–96.

McGeady, R., Runya, R. M., Dooley, J. S., Howe, J. A., Fox, C. J., Wheeler, A. J., ... & McGonigle, C. (2023). A review of new and existing non-extractive techniques for monitoring marine protected areas. Frontiers in Marine Science, 10, 1126301.

Miller, R. J., Hocevar, J., Stone, R. P., and Fedorov, D. V. (2012). Structure-forming corals and sponges and their use as fish habitat in bering sea submarine canyons. PLoS ONE 7:e33885. doi: 10.1371/journal.pone.0033885

Muriel Gevrey, Ioannis Dimopoulos, Sovan Lek (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological Modelling, Volume 160, Issue 3, 2003, Pages 249-264.

O'Connell, K. E., & Collins, P. M. (2012, September). Safeguarding Coldwater Coral Reefs: A Method For Mapping And Characterisation. In SUT Offshore Site Investigation and Geotechnics (pp. SUT-OSIG). SUT.

Pradervand, J. N., Dubuis, A., Pellissier, L., Guisan, A., & Randin, C. (2014). Very high resolution environmental predictors in species distribution models: moving beyond topography?. Progress in Physical Geography, 38(1), 79-96.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/
Roberts D. G. (1975). Marine geology of the Rockall Plateau and Trough. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences. **278**: 447–509.

Roberts, J. M., and Cairns, S. D. (2014). Cold-water corals in a changing ocean. Curr. Opin. Environ. Sustain. 7, 118–126. doi: 10.1016/j.cosust.2014.01.004

Robert, K., Jones, D. O., Roberts, J. M., & Huvenne, V. A. (2016). Improving predictive mapping of deep-water habitats: considering multiple model outputs and ensemble techniques. Deep Sea Research Part I: Oceanographic Research Papers, 113, 80-89.

Rogers, A. D. (1999). The Biology of Lophelia pertusa (Linnaeus 1758) and Other Deep-Water Reef-Forming Corals and Impacts from Human Activities. International review of hydrobiology, 84(4), 315-406.

Robert, K. (2014). Evaluation of local-and medium-scale habitat heterogeneity as proxy for biodiversity in deep-sea habitats. Doctoral dissertation, University of Southampton.

Sampaio, I., Braga-Henriques, A., Pham, C., Ocaña, O., De Matos, V., Morato, T., & Porteiro, F. M. (2012). Cold-water corals landed by bottom longline fisheries in the Azores (north-eastern Atlantic). Journal of the Marine Biological Association of the United Kingdom, 92(7), 1547-1555.

Schapire, R. E. (2003). The boosting approach to machine learning: An overview. Nonlinear estimation and classification, 149-171.

Slagstad, D., Tande, K. S., and Wassmann, P. (1999). Modelled carbon fluxes as validated by field data on the north Norwegian shelf during the productive period in 1994. Sarsia 84, 303–317.

Sleeman J, Boggs G, Radford B, Kendrick Ga (2005) Using Agent Based Models to Aid Reef Restoration: Enhancing Coral Cover and Topographic Complexity through the Spatial Arrangement of Coral Transplants. Restoration Ecology 13: 685–694.

Stegmann, P. G., Johnson, B., Moradi, I., Karpowicz, B., & McCarty, W. (2022). A deep learning approach to fast radiative transfer. Journal of Quantitative Spectroscopy and Radiative Transfer, 280, 108088.

Storlazzi, C. D., Dartnell, P., Hatcher, G. A., & Gibbs, A. E. (2016). End of the chain? Rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (SfM) technology. Coral Reefs, 35(3), 889-894.

Sundahl, H., Buhl-Mortensen, P., & Buhl-Mortensen, L. (2020). Distribution and suitable habitat of the cold-water corals Lophelia pertusa, Paragorgia arborea, and Primnoa resedaeformis on the Norwegian continental shelf. Frontiers in Marine Science, 213.

Vaze, J., Teng, J., & Spencer, G. (2010). Impact of DEM accuracy and resolution on topographic indices. Environmental Modelling & Software, 25(10), 1086-1098.

Ware, S., & Downie, A. L. (2020). Challenges of habitat mapping to inform marine protected area (MPA) designation and monitoring: An operational perspective. Marine Policy, 111, 103717.

Wilkinson, C. M., Ganerød, M., Hendriks, B. W., & Eide, E. A. (2017). Compilation and appraisal of geochronological data from the North Atlantic Igneous Province (NAIP). Geological Society, London, Special Publications, 447(1), 69-103.

Williams, G. C. (2011). The global diversity of sea pens (Cnidaria: Octocorallia: Pennatulacea). PLoS one, 6(7), e22747.

Yesson, C., Bedford, F., Rogers, A. D., & Taylor, M. L. (2017). The global distribution of deep-water Antipatharia habitat. Deep Sea Research Part II: Topical Studies in Oceanography, 145, 79-86.

Yu, H., Yang, W., Liu, C., Tang, Y., Song, X., & Fang, G. (2020). Relationships between community structure and environmental factors in Xixiakou artificial reef area. Journal of Ocean University of China, 19, 883-894.

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., ... & Merow, C. (2020). A standard protocol for reporting species distribution models. Ecography, 43(9), 1261-1277.

7. Supplementary Material

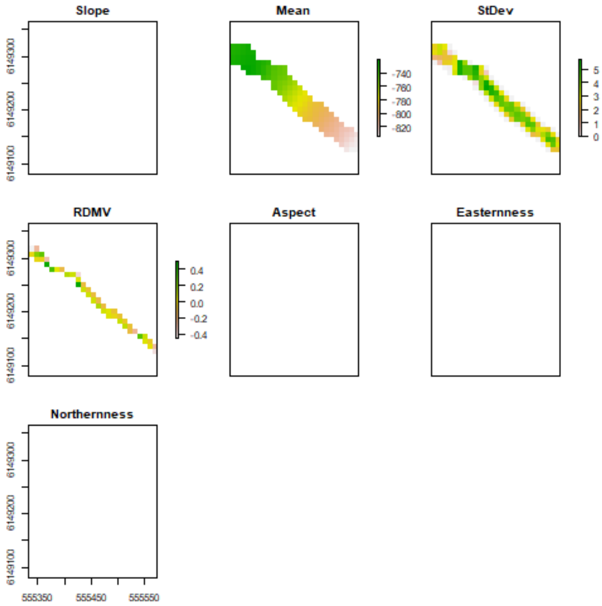

Figure 18. Fine scale environmental variables derived from bathymetry data using TASSE Toolbox when clipped to raster extent of ultra fine scale transect and processed in R Studio showing not all variables were computed

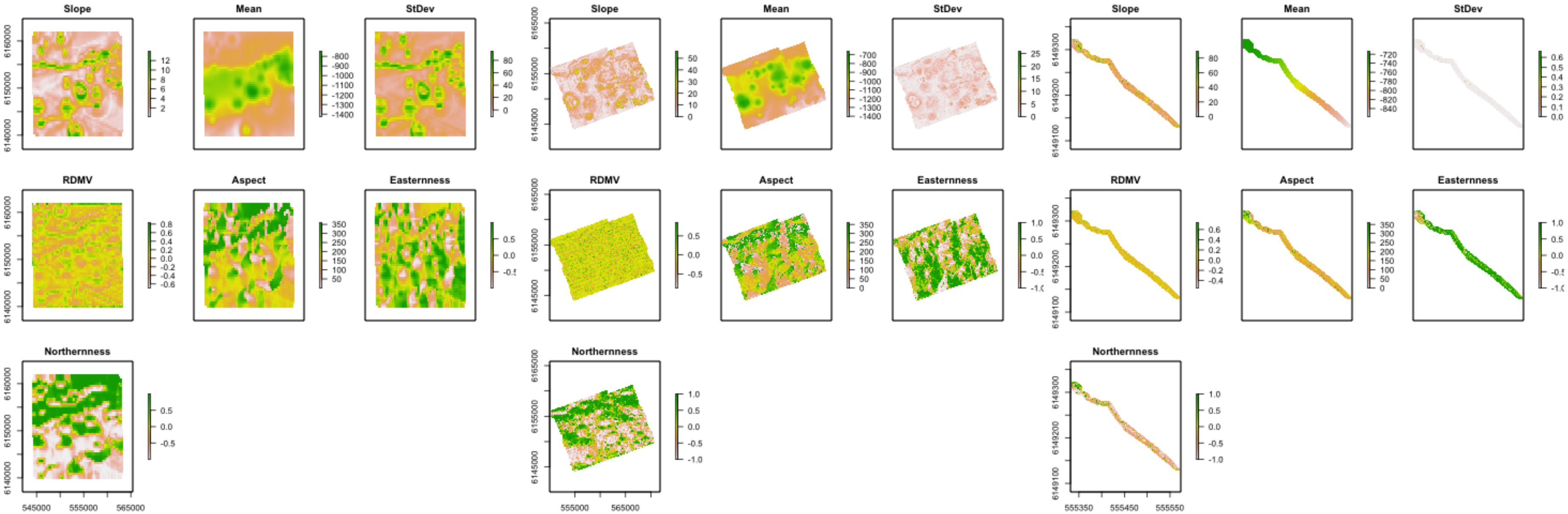

Figure 19. Environmental layers used in model training for broad scale. Top: Broad scale. Middle: fine scale. Bottom: ultra fine scale data; Slope, Mean (Depth), Standard deviation (StDev) a proxy for rugosity, Relative Distance from Mean (RDMV), Aspect, Easteness and Northernness

Appendix A – R Script

```
#PART A: PREPARATION
print("PART A: PREPARATION")
#_________________________________________________________________
#1. Set working directory
setwd("~/Desktop/R")

#2. Load supplementary script source
source("~/Desktop/R/Functions/sourceFunctions.R")

#3. Load additional libraries (the rest are in sourceFunctions.R)
library(readr)

#4. Assign projection systems to proj4 strings
utmProjection <- "+proj=utm +datum=WGS84 +zone=27"
gpsProjection <- "+proj=longlat +datum=WGS84"

#5. (Get bathymetry data for plotting - this is for future use)
print("Downloading bathymetry data...")
lim = data.frame(ylim=c(55, 56), xlim=c(-20.5, -19.5))
bathy = getNOAA.bathy(
  lon1 = lim$xlim[1],
  lon2 = lim$xlim[2],
  lat1 = lim$ylim[1],
  lat2 = lim$ylim[2], resolution = 0.01)
bathyRasterised = marmap::as.raster(bathy)
bathyRasterised <- projectRaster(bathyRasterised, crs=utmProjection)
bathyGplot <- gplot_data(bathyRasterised)

#PART B: PROCESSING FINE SCALE DATA
print("PART B: PROCESSING FINE SCALE DATA")
#_________________________________________________________________
#1. Read presences data from CSV file
presences <- read.csv("~/Desktop/R/Data/Database/LP_records1.csv", sep = ",")
print("Reading presences completed")

#2. Our presences data is expressed in global coordinates (deg, min, sec). Transform them to UTM
coordinates.
#2a. Assign individual lat/long variables for manipulation
gpsLongitudes = presences$Lon
gpsLatitudes = presences$Lat

#2b. Assign the points, and construct vectors of UTM coordinates from GPS data
gpsPoints <- cbind(gpsLongitudes, gpsLatitudes)
gpsVect <- terra::vect(gpsPoints, crs=gpsProjection)
utmVect <- terra::project(gpsVect, utmProjection)
utm <- geom(utmVect)[, c("x", "y")]

#2c. Finally, assign the UTM coordinates to the presences variable
```

```
presences['Lat'] <- utm[,1]
presences['Lon'] <- utm[,2]
print("Transforming presences to UTM coordinates completed")

#3. Generate raster objects
#3a. Read rasters
slope <- raster("~/Desktop/R/Data/Ascii/slo_ds3.asc")
mean <- raster("~/Desktop/R/Data/Ascii/mean_ds3.asc")
StDev <- raster("~/Desktop/R/Data/Ascii/sd_ds3.asc")
rdmv <- raster("~/Desktop/R/Data/Ascii/rdmv_ds3.asc")
aspect <- raster("~/Desktop/R/Data/Ascii/asp_ds3.asc")
east <- raster("~/Desktop/R/Data/Ascii/east_ds3.asc")
north <- raster("~/Desktop/R/Data/Ascii/nort_ds3.asc")

#3b. Stack layers to generate a combined object
environmentalConditions <- stack(slope, mean, StDev, rdmv, aspect, east, north)

#3c. Assign more readable/simpler names
names(environmentalConditions) <- c("Slope", "Mean", "StDev", "RDMV", "Aspect", "Easternness",
"Northernness")

#3d. Plot predictors for checking
png(file="~/Desktop/R/Plots/1. Fine Scale Environmental Conditions.png", width=1500)
plot(environmentalConditions)
dev.off();
print("Fine scale raster reading complete")

#4. Generate pseudo-absences
print("Generating pseudo-absences")
pseudoAbs <- pseudoAbsences(environmentalConditions,presences,n=200)
print("Generating pseudo-absences completed")

#5. Fit a model
print("Generating model data")
modelData <- prepareModelData(presences,pseudoAbs,environmentalConditions)
print("Generating model data complete")

#6. Generate cross validation folds
folds <- getBlocks(modelData)

#7. Define monotonicity constraints (-1 for negative, +1 for positive, 0 for non-monotonicity)
monotonicity = data.frame(Slope=+1,Mean=+1,StDev=+1,RDMV=0,Aspect=0,Easternness=0,Northernness=0)

#8. Fit a BRT model with cross-validation
print("Fitting a BRT model")
model <- train("BRT", modelData, folds = folds)
print("Fitting BRT model complete")

#9. Given a set of possible hyperparameter values for BRT, test all possible combinations of
hyperparameter values
print("Testing hyperparameters")
```

```
h <- list(interaction.depth = c(1,2,3,4) , shrinkage = c(0.1,0.01,0.001) )
exp1 <- gridSearch(model, hypers = h, metric = "auc")
png(file="~/Desktop/R/Plots/2. Grid Search.png", width=1500)
plot(exp1)
dev.off()
exp1@results
print("Testing hyperparameters complete")

#10. Fit a BRT model to the dataset with the best hyperparameter values
print("Fitting model with best hyperparameters")
model <- train("BRT", modelData, folds = folds , interaction.depth=2, shrinkage=0.010 )
print("Fitting model with best hyperparameters complete")

#11. Determine relative variable contribution
viModel <- varImp(model, permut = 5)
viModel

#12. Reduce model complexity by dropping one variable at a time
print("Reducing the model")
reducedModel <- reduceVar(model, th = 0.1, metric = "auc", permut = 5)
print("Model reduction complete")

#13. Determine the final relative variable contribution
print("Computing relative variable contribution")
viModel <- varImp(reducedModel, permut = 5)
viModel
png(file="~/Desktop/R/Plots/3. Relative Variable Contribution.png", width=1500)
plotVarImp(viModel)
dev.off()

#PART C: CHECK FINE SCALE MODEL QUALITY
print("PART C: CHECK FINE SCALE MODEL QUALITY")
#__________________________________________________________________
#1. Inspect response curves
png(file="~/Desktop/R/Plots/4. Slope Response Curve.png", width=1500)
plotResponse(reducedModel, var = "Slope", type = "logistic", only_presence = FALSE, marginal = FALSE,
rug = FALSE, color="Black")
dev.off();
png(file="~/Desktop/R/Plots/5. Mean Response Curve.png", width=1500)
plotResponse(reducedModel, var = "Mean", type = "logistic", only_presence = FALSE, marginal = FALSE,
rug = FALSE, color="Black")
dev.off();

#2. Determine the threshold maximizing the sum of sensitivity and specificity
threshold <- thresholdMaxTSS(reducedModel)

#3. Quantitatively determine model performance
getAccuracy(reducedModel,threshold = threshold)
getAUC(reducedModel, test = TRUE)

#4. Create a raster stack using the BRT predictions
```

```
mapPresent <- predict(reducedModel, environmentalConditions, type=c("logistic"))
mapPresentDataFrame <- as.data.frame(mapPresent, xy=T)

#5. Extract minima and maxima for extents
xMin = min(mapPresentDataFrame['x'])[1]
xMax = max(mapPresentDataFrame['x'])[1]
yMin = min(mapPresentDataFrame['y'])[1]
yMax = max(mapPresentDataFrame['y'])[1]

#6. Plot the combined map
png(file="~/Desktop/R/Plots/6. Fine Scale Model Response.png", width=1500, height=1500)
ggplot()+
  xlim(xMin, xMax)+
  ylim(yMin, yMax)+
  geom_raster(data=mapPresentDataFrame, aes(x=x,y=y,fill=mean))+
  geom_point(data=presences, aes(x=Lat, y = Lon, color="red"))
dev.off()

#PART D: TEST THE MODEL ON GEBCO DATA TO CHECK SUITABILITY
print("PART D: TEST THE MODEL ON GEBCO DATA TO CHECK SUITABILITY")
#_________________________________________________________________
#1. Generate GEBCO raster objects
#1a. Read rasters
gebcoSlope <- raster("~/Desktop/R/Data/Ascii/slop_gebco_2.asc")
gebcoMean <- raster("~/Desktop/R/Data/Ascii/mean_gebco_2.asc")
gebcoStDev <- raster("~/Desktop/R/Data/Ascii/stde_gebco_2.asc")
gebcoRdmv <- raster("~/Desktop/R/Data/Ascii/rdmv_gebco_2.asc")
gebcoAspect <- raster("~/Desktop/R/Data/Ascii/asp_gebco_2.asc")
gebcoEast <- raster("~/Desktop/R/Data/Ascii/east_gebco_2.asc")
gebcoNorth <- raster("~/Desktop/R/Data/Ascii/nort_gebco_2.asc")

#1b. Stack layers to generate a combined object
gebcoEnvironmentalConditions <- stack(gebcoSlope, gebcoMean, gebcoStDev, gebcoRdmv, gebcoAspect,
gebcoEast, gebcoNorth)
#1c. GEBCO rasters are in GPS coordinates. Reproject to UTM coordinates
crs(gebcoEnvironmentalConditions) <- gpsProjection
gebcoEnvironmentalConditions <- projectRaster(gebcoEnvironmentalConditions, crs=utmProjection)
#1d. Assign more readable/simpler names
names(gebcoEnvironmentalConditions) <- c("Slope", "Mean", "StDev", "RDMV", "Aspect", "Easterness",
"Northernness")

#1e. Plot predictors for checking
png(file="~/Desktop/R/Plots/7. Broad Scale Environmental Conditions.png", width=1500)
plot(gebcoEnvironmentalConditions)
dev.off()

#2. Create a raster stack using the fine scale BRT predictions on the broad scale data
mapPresentGebco <- predict(model, gebcoEnvironmentalConditions, type=c("logistic"))
mapPresentGebcoDataFrame <- as.data.frame(mapPresentGebco, xy=T)

#3. Extract minima and maxima for extents
```

```
#xMin = min(mapPresentGebcoDataFrame['x'])[1]
#xMax = max(mapPresentGebcoDataFrame['x'])[1]
#yMin = min(mapPresentGebcoDataFrame['y'])[1]
#yMax = max(mapPresentGebcoDataFrame['y'])[1]

#4. Plot the combined map
png(file="~/Desktop/R/Plots/8. Broad Scale Model Response.png", width=1500, height=1500)
ggplot()+
  xlim(xMin, xMax)+
  ylim(yMin, yMax)+
  geom_raster(data=mapPresentGebcoDataFrame, aes(x=x,y=y,fill=mean))+
  geom_point(data=presences, aes(x=Lat, y = Lon, color="red"))
dev.off()
```

Appendix B – Monotonicity Constraint Iteration Script

```
SCRIPT NAME: mainWithMonotonicityIteration.R

#PART A: PREPARATION
print("PART A: PREPARATION")
#__________________________________________________________________
#1. Set working directory
setwd("~/Desktop/R")

#2. Load supplementary script source
source("~/Desktop/R/Functions/sourceFunctions.R")

#3. Load additional libraries (the rest are in sourceFunctions.R)
library(readr)

#4. Assign projection systems to proj4 strings
utmProjection <- "+proj=utm +datum=WGS84 +zone=27"
gpsProjection <- "+proj=longlat +datum=WGS84"

#5. (Get bathymetry data for plotting - this is for future use)
print("Downloading bathymetry data...")
lim = data.frame(ylim=c(55, 56), xlim=c(-20.5, -19.5))
bathy = getNOAA.bathy(
  lon1 = lim$xlim[1],
  lon2 = lim$xlim[2],
  lat1 = lim$ylim[1],
  lat2 = lim$ylim[2], resolution = 0.01)
bathyRasterised = marmap::as.raster(bathy)
bathyRasterised <- projectRaster(bathyRasterised, crs=utmProjection)

#PART B: PROCESSING FINE SCALE DATA
print("PART B: PROCESSING FINE SCALE DATA")
#__________________________________________________________________
#1. Read presences data from CSV file
presences <- read.csv("~/Desktop/R/Data/Database/LP_records1.csv", sep = ",")
print("Reading presences completed")

#2. Our presences data is expressed in global coordinates (deg, min, sec). Transform them to UTM coordinates.
#2a. Assign individual lat/long variables for manipulation
gpsLongitudes = presences$Lon
gpsLatitudes = presences$Lat

#2b. Assign the points, and construct vectors of UTM coordinates from GPS data
gpsPoints <- cbind(gpsLongitudes, gpsLatitudes)
gpsVect <- terra::vect(gpsPoints, crs=gpsProjection)
utmVect <- terra::project(gpsVect, utmProjection)
utm <- geom(utmVect)[, c("x", "y")]
```

```
#2c. Finally, assign the UTM coordinates to the presences variable
presences['Lat'] <- utm[,1]
presences['Lon'] <- utm[,2]
print("Transforming presences to UTM coordinates completed")

#3. Generate raster objects
#3a. Read rasters
slope <- raster("~/Desktop/R/Data/Ascii/slo_ds3.asc")
mean <- raster("~/Desktop/R/Data/Ascii/mean_ds3.asc")
StDev <- raster("~/Desktop/R/Data/Ascii/sd_ds3.asc")
rdmv <- raster("~/Desktop/R/Data/Ascii/rdmv_ds3.asc")
aspect <- raster("~/Desktop/R/Data/Ascii/asp_ds3.asc")
east <- raster("~/Desktop/R/Data/Ascii/east_ds3.asc")
north <- raster("~/Desktop/R/Data/Ascii/nort_ds3.asc")

#3b. Stack layers to generate a combined object
environmentalConditions <- stack(slope, mean, StDev, rdmv, aspect, east, north)

#3c. Assign more readable/simpler names
names(environmentalConditions)  <-  c("Slope",  "Mean",  "StDev",  "RDMV",  "Aspect",  "Easterness",
"Northernness")

#3d. Plot predictors for checking
png(file="~/Desktop/R/Plots/1. Fine Scale Environmental Conditions.png", width=1500)
plot(environmentalConditions)
dev.off();
print("Fine scale raster reading complete")

#4. Generate pseudo-absences
print("Generating pseudo-absences")
pseudoAbs <- pseudoAbsences(environmentalConditions,presences,n=200)
print("Generating pseudo-absences completed")

#5. Fit a model
print("Generating model data")
modelData <- prepareModelData(presences,pseudoAbs,environmentalConditions)
print("Generating model data complete")

#6. Generate cross validation folds
folds <- getBlocks(modelData)

monotonicityVariables = c(-1, 0, 1)
best = ""
auc = 0
resultsList <- data.frame (
  Title = c("Title"),
  AccuracyThreshold= c(0),
  AccuracyTss= c(0),
  AccuracySensitivity= c(0),
  AccuracySpecificity= c(0),
  AUC = c(0)
```

```
)

slopeVar = 1
meanVar = 1
StDevVar = 1
iteration = 0
for (rdmvVar in monotonicityVariables)
{
  for (aspectVar in monotonicityVariables)
  {
    for (easternnessVar in monotonicityVariables)
    {
      for (northernnessVar in monotonicityVariables)
      {
        iteration = iteration + 1
        print(sprintf("Iteration %d of %d", iteration, 3*3*3*3))
        title <- sprintf("Slope = %d Mean = %d StDev = %d RDMV = %d Aspect = %d Easternness = %d
Northernness = %d", slopeVar, meanVar, StDevVar, rdmvVar, aspectVar, easternnessVar, northernnessVar)

        #7. Define monotonicity constraints (-1 for negative, +1 for positive, 0 for non-monotonicity)
        monotonicity                                                                              =
data.frame(Slope=slopeVar,Mean=meanVar,StDev=StDevVar,RDMV=rdmvVar,Aspect=aspectVar,Easternness=eas
ternnessVar,Northernness=northernnessVar)

        #8. Fit a BRT model with cross-validation
        print("Fitting a BRT model")
        model <- train("BRT", modelData, folds = folds)
        print("Fitting BRT model complete")

        #9. Given a set of possible hyperparameter values for BRT, test all possible combinations of
hyperparameter values
        print("Testing hyperparameters")
        h <- list(interaction.depth = c(1,2,3,4) , shrinkage = c(0.1,0.01,0.001) )
        exp1 <- gridSearch(model, hypers = h, metric = "auc")
        png(file=sprintf("~/Desktop/R/Plots/2. Grid Search - %s.png", title), width=1500)
        plot(exp1)
        dev.off()
        bestHyper = exp1@results[which.min(exp1@results$diff_AUC),]
        print(bestHyper)
        print("Testing hyperparameters complete")

        #10. Fit a BRT model to the dataset with the best hyperparameter values
        print("Fitting model with best hyperparameters")
        model        <-        train("BRT",        modelData,        folds        =        folds,
interaction.depth=bestHyper$interaction.depth[1], shrinkage=bestHyper$shrinkage[1])
        print("Fitting model with best hyperparameters complete")

        #11. Determine relative variable contribution
        viModel <- varImp(model, permut = 5)
        viModel
```

```
        #12. Reduce model complexity by dropping one variable at a time
        print("Reducing the model")
        reducedModel <- reduceVar(model, th = 0.1, metric = "auc", permut = 5)
        print("Model reduction complete")

        #13. Determine the final relative variable contribution
        print("Computing relative variable contribution")
        viModel <- varImp(reducedModel, permut = 5)
        viModel
        png(file=sprintf("~/Desktop/R/Plots/3. Relative Variable Contribution - %s.png", title),
width=1500)
        plotVarImp(viModel)
        dev.off()

        #PART C: CHECK FINE SCALE MODEL QUALITY
        print("PART C: CHECK FINE SCALE MODEL QUALITY")
        #__________________________________________________________________

        #2. Determine the threshold maximizing the sum of sensitivity and specificity
        threshold <- thresholdMaxTSS(reducedModel)

        #3. Quantitatively determine model performance
        accuracy = getAccuracy(reducedModel,threshold = threshold)
        auc2 = getAUC(reducedModel, test = TRUE)
        resultsList[nrow(resultsList) + 1,] = c(title, accuracy$threshold, accuracy$tss,
accuracy$sensitivity, accuracy$specificity, auc2)
        if (auc2 > auc){
          auc = auc2
          best = title
        }
      }
    }
  }
}

install.packages("writexl")
library("writexl")
write_xlsx(resultsList,"~/Desktop/R/MonotonicityIteration.xlsx")
```