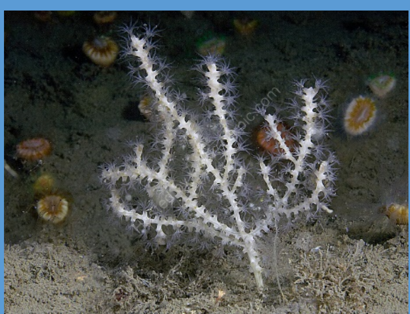



D.T2.1.4 Report on Species Distribution Modelling (SDM) workshop 26-27th March 2019

April 2019

MarPAMM
Marine Protected Area Management and Monitoring

Interreg
Northern Ireland - Ireland - Scotland
European Regional Development Fund



Executive summary

The T2 work package will provide one of four models contributing to the evidence base available for design of Marine Protected Area (MPA) management plans being delivered by the MarPAMM project; specifically the benthic species and habitats model. Species distribution models (SDMs) will be developed across the INTERREG VA region for the seven selected species and habitats in this work package as listed below.

1. Flapper skate (*Dipturus intermedia*)
2. Sea pen species characteristic of burrowed mud and other protected sedimentary habitats (*Virgularia mirabilis, Pennatula phosphorea, Funiculina quadrangularis*)
3. Horse mussel (*Modiolus modiolus*)
4. Sea fans complex (*Swiftia pallida* & *Eunicella verrucosa*). Inclusion of *Eunicella verrucosa* to be confirmed pending data availability sense check.
5. Maerl (*Phymatolithon calcareum & Lithothamnion glaciale*)
6. Ocean quahog (*Arctica islandica*). Inclusion to be confirmed pending data availability sense check.
7. Fan mussel (*Atrina fragilis*)- inclusion to be confirmed pending data availability sense check

To aid with the development of these models, a species distribution modelling workshop was hosted by MarPAMM at the main AFBI site in Belfast between 26-27th March 2019. This workshop brought together participants from academia, policy, NGOs and government bodies to help identify current barriers with the use of SDMs, how SDM outputs might be presented in a meaningful way to end users and identification of best practice methods to overcome more technical issues with producing such models.

The following recommendations were made to help overcome barriers limiting the use of SDMs in marine management and monitoring

- Improve dialogue between modellers and stakeholders who wish to use models, to ensure that they are fit for the intended purpose and are up to date.
- Increase coordination between organisations who undertake surveys to maximise on potential opportunities for field verification
- Set up a central hub of current knowledge and status of benthic species distribution models in the Interreg VA area, to include links to appropriate datasets and technical guidance.

- Develop standard metadata which can be used to assess the quality of the SDM output and communicate uncertainty more effectively.
- Produce a benthic modelling toolkit for stakeholders on how SDMs can be used to improve MPA management in the region

1. Workshop purpose

This workshop was organised to investigate the range of methods currently used to produce SDMs, showcase examples where such approaches have been used (both from on-going research and literature reviews) and highlight current barriers faced with such models in an attempt to create action plans as to how we might overcome these within MarPAMM.

The workshop was carried out over two days (26th & 27th March). Day one gave MarPAMM team members an opportunity to introduce the project to the wider audience (presented by Annika Clements), as well as providing a basic introduction as to what SDMs are, how and why we use them and how the outputs can be interpreted by the end user (presented by Phil Boulcott). Once all participants had been given a basic understanding of these models, nine case studies were presented to demonstrate how such models are currently being used to address key policy issues. The second half of day one aimed to explore these case studies in small groups to consider what aspects of each worked well or if any limitations were revealed. Furthermore, the workshop aimed to discuss the potential management applications of these case studies before summarising group discussions to the wider group. By the end of day one it was hoped that barriers to SDMs would be identified, requirements for evidence underpinning management obtained and how to communicate results with end users agreed.

The purpose of day two was to explore some of the issues highlighted from day one in slightly more detail. Following a recap of everything that was discussed during day one, a more technical exercise was planned for day two where it was hoped that data limitations, modelling approaches, validation and verification methods, interpretation of results and updating SDMs could all be discussed with the input of policy leads as well as marine scientists.

Finally, the overall aim of this workshop was to create a set of action plans for MarPAMM to help improve the uptake of SDMs by environmental managers. More specifically, where SDMs are appropriate in guiding management and how these can be an effective tool for marine management.

2. Summary of participants

This workshop hosted a range of participants from across the UK and Ireland. In attendance were staff from the Agri-Food and Biosciences Institute (x5), Marine Science Scotland (x3), Scottish Natural Heritage (x1), Department of Agriculture, Environment and Rural Affairs (x1), National Parks and Wildlife Service (x1), Scottish Association for Marine Science (x3), Heriot-Watt University (x2), University of Edinburgh (x1), National University of Ireland Galway (x2), Joint Nature Conservation Committee (x1) and Ulster University (x3), see table 1 for full details.

3. Case studies

Nine case studies were presented by participants from various organisations. During the presentations, four rapporteurs took notes specifically focussing on the modelling approach used, data limitations, how/if the model was used for marine management and the take home messages. Where possible the presentations and rapporteur notes are available in the appendix.

3.1 Summary of case studies

Case study 1: Fan mussels (David Stirling & Lisa Kamphausen)

Fan mussels *(Atrina fragilis*) are a rare benthic species and as a result there are few current occurrence records. The case study illustrated how historic data could be found and used to increase the available occurrence dataset and allow a species distribution model to be performed. A Poisson point process approach was used. The model predictions were verified using field surveys. The output of this model has been used by SNH as part of the Priority Marine Features review, which proposed new fisheries management areas to protect fan mussel. Although the model output alone could not be used to identify locations for fisheries management (there needs to be confirmation of presence from survey data) the model output helped guide the boundary for the management area.

Case study 2: Mingulay reef (Laurence De Clippele)

This case study introduced the ATLAS project, specifically focusing on *Lophelia pertusa* on the Mingulay reef. The use of multibeam echosounder was shown to act as powerful tool for locating and mapping deep-water coral reefs, however more environmental data would have increased the resolution in this example. A retracing approach was used to create separate polygons based on changes in backscatter as well as the BGS tool in Arc GIS to create detailed habitat maps leading to the designation of an SAC in August 2011. Further to this, a random forest (RF) method was used to predict the presence of *L. pertusa* on 'minimounds' within Mingulay reef.

Case study 3: North Sea (Paul Mayo)

Taken from the literature (Reiss *et al.,* 2011), this review paper compared nine modelling approaches commonly used in species distribution modelling. Twenty benthic marine species were modelled using ten environmental covariates. Drawing comparisons between species was difficult as species with a narrower environmental range could be modelled more accurately than those with a wider range. This paper also failed to include biological traits such as species interaction which may further influence the accuracy of the models. Of all methods considered, maximum entropy (MAXENT) emerged as the model with the highest accuracy and predictive ability, whereas bioclimatic envelope (BIOCLIM) proved to be the least. Some methods such as random forest (RF), flexible discriminant analysis (FDA) and multivariate adaptive regression splines (MARS) often under-predicted, suggesting these methods may not be as useful for species protection application such as marine protected area designation. On the contrary, genetic algorithm for rule-set prediction (GARP) over-predicted suggesting this approach could be more useful for precautionary management strategies such as limiting the spread of invasive species.

Ref: Reiss, H., Cunze, S., König, K., Neumann, H. and Kröncke, I. (2011). Species distribution modelling of marine benthos: a North Sea case study. *Marine Ecology Progress Series*, **442**: 71-86.

Case study 4: Flame shells (Euan Mackenzie)

Here, the outline of a recently started PhD project was presented. The large extent of flame shell range was highlighted however, it was noted that there was positional uncertainty in some records and some of the main issues with mapping their distribution was differentiating between the different types of aggregation as well as assigning environmental variables. This project hopes to improve the resolution of flame shell mapping using a range of methods including MAXENT, FVCOM and Ecospat (multi-species habitat comparability). Application of this study will be most useful for subsea cable routing.

Case study 5: Maerl (Cornelia Simon Nutbrown)

A second PhD study was presented, one which focused on the conservation of maerl beds in Scotland, combining a genetic and predictive modelling approach. One of the main issues encountered was that of sampling bias, where many data points are clumped together and may have arisen as a result of sampling effort. Second to this was the issue of obtaining data. A MAXENT approach was used to help predict suitable habitats for maerl along the western coast of Scotland using a range of environmental variables. One recommendation put forward here was to ensure that the variables put into the model make ecological sense.

Case study 6: AUVs as tools in Marine Spatial Planning (Karen Boswarva)

This case study focussed on AUV surveys in Chilean Patagonia. The aims of the surveys were to generate habitat maps and seabed imagery of fjordic systems and verify sites of the cold-water coral *Desmophyllum dianthus*. The AUV was deployed 19 times and led to 30,810 usable images of the seabed. Issues that were encountered during the deployments related to the enclosed nature of the fjord environment, seabed topography and inaccurate nautical charts. Predictive habitat maps of the surveyed areas were created using a modified version of EUNIS/Marine Habitat Classification for Britain and Ireland. The maps will be used to inform MPA designation and management.

Case study 7: Benthic community modelling in Gulf of Maine (Jay Calvert)

Rather than modelling a single species, this case study used a Joint Species Distribution Model to model the benthic community at Cashes Ledge in Gulf of Maine. Presence and absence data were collected using still images of the seabed and were modelled using HMSC. The results show that the level of association detected between species in the community depends on the spatial scale being analysed. The study also found that varying detection rates between different species can influence habitat maps. Results can therefore be impacted by false absences, particularly if there is only data from a single snap-shot in time, mis-identification of taxa and limited field of view in images. Other data limitations identified were scale mismatches between species' data and covariate data, and auto-correlation violating independence assumptions of the model.

Case study 8: Benthos distribution modelling for marine ecosystem management (Beckie Langton)

The main findings of a review paper taken from the literature (Reiss *et al.* 2015) were presented. The authors of the paper noted that species distribution models of benthic species have been produced at a range of geographic scales and that these could have applications for marine ecosystem management. Three broad approaches to SDMs were identified; statistical approaches, mechanistic models and Bayesian belief networks. Statistical approaches are by far the most common method used, and the authors did not find any examples of mechanistic models or Bayesian belief networks being applied in marine management. Reasons for this could be that statistical models are easier to understand and interpret and require less detailed biological information than mechanistic models. Four broad applications for SDMs were discussed in the paper; marine spatial planning, marine monitoring, non-indigenous species and future scenario prediction. Recommendations for improving the use of SDMs in marine ecosystem management were to include cause-effect relationships as well as correlative relationships, ensure the physical variables being used in the model are relevant to benthic species (e.g. relate to conditions at the seabed) and include biological interactions.

Case study 9: Mapping species that move (Patricia Breen)

Unlike the other case studies, this one related to mobile marine species rather than benthic species. Initially, much of the occurrence data came from a number of different sources and was presence-only. Liklihood of presence of different species each quarter were mapped using Maxent with the environmental variables; depth, temperature, salinity and distance to coast. The distribution of cetaceans is more likely to be dependent on the location of prey, but this is not known, so environmental variables are used as a proxy. The advantages of using Maxent for the study were that it can combine data from different sources and model rare species. However, Maxent predicts likelihood of presence and not a measure of abundance or density, therefore there is not usually enough evidence to make management decisions. An aerial survey was carried out and the abundance of different species estimated using distance sampling techniques. SDMs were generated using Generalised Additive Models (GAMs) to map abundances. The GAMS produced estimates of abundance, however a good number of reliable sightings and obtaining this amount of data is expensive and difficult. Model outputs are likely to be used in environmental assessments.

3.2 Key issues from case studies

The following technical issues were identified from the introductory presentations and case studies.

- How to model reef or aggregation-forming species?
- How to deal with historical data?
- Positional accuracy and scale between occurrence data and environmental layers
- How is ‘over-fitting’ or ‘under-fitting’ determined?
- Knowing when to give up
- Validation best practice
- Sense-checking: Biological knowledge and causative effects
- When to use an ensemble approach

<u>3.3 Recommendations</u>

The issues listed above are relevant for both modellers developing SDMs and stakeholders who apply the models in management. However, many of the issues listed have been discussed in the literature and information is already available. There was a recommendation that, during the MarPAMM project, briefing notes on these issues will be produced which will sign-post to key information and resources.

<u>4. Barriers to using Species Distribution Models in management</u>

In group and plenary discussions, the following barriers limiting the use of SDMs in marine management were identified.

<u>4.1 Stakeholder engagement</u>

The purpose of the Species Distribution Models that will be developed during the MarPAMM project is to inform marine management and monitoring. There are a number of stakeholder groups including government policy departments, industry and other marine users who could benefit from the models. It is recognised that to date SDMs have not been widely picked up by stakeholders. This may be due to: (1) lack of resources such as finance, (2) perceived lack of applicability, (3) lack of awareness of the existence and function of the models and (4) perceived technical complexity. To address these issues, projects intending to develop SDMs should develop a meaningful dialog with stakeholders early in the process. For example, within the MarPAMM project engagement was carried out between key policy leads and stakeholder groups when prioritising which species to include.

At present, it is recognised that generally the dialog existing between modellers and stakeholders is informal, passing through several routes and agencies. SDM work within the Interreg VA Region does not have a visible home or framework. It is the aim of the MarPAMM project to improve this situation by: (1) developing and sharing SDMs for seven species with relevant stakeholders, (2) interface closely with the development of regional management plans also being developed within the MarPAMM project and (3) providing case studies that emphasise the different uses and applications of SDMs to marine management.

A further function of this coordination within the MarPAMM project could be to act as a central hub for the current state of knowledge on SDMs. This information would include what SDM information exists and for what species and where, as well as environmental layers, data portals and relevant technical guidance. This hub could also act as a conduit through which policy stakeholders could make requests for SDMS.

4.2 Identification of evidence needs

Stakeholders will have in mind a clear use for SDMs and they may need to meet certain standards in order for them to be used as evidence in management. This needs to be articulated to the modelling process at the outset. For example, users may require abundance rather than presence/ absence information spread over a defined geographic scale. Therefore dialog should include issues such as geographic area, species of interest, resolution and scale and the need to include the effects of pressures. The need for causation mechanisms, whether current distribution scenarios or general habitat suitability is required, also need to be specified. These need to be communicated to modellers early as they can affect the modelling process.

It should be recognised that dialog between partners has to be two-way. In the first instance, stakeholders may have objectives that SDMs can serve e.g. management objectives can range from species protection through to restoration. SDMs can be seen as a low cost alternative to monitoring. Whilst SDMs do deliver value for money, they should not be seen as a replacement for further monitoring of species distribution in all cases. For example, where data layers are exceptionally poor, SDMs are likely to be imprecise and should not be considered. In this case, physical monitoring should be the preferred option.

SDMs can meet marine spatial planning objectives. The ability to map the distribution of the species assists in spatial management measures, environment assessments and EIAs, and defining ecological niche can inform restoration objectives. The use of future scenarios (such as climate models) can elucidate future species distribution and the spread of invasive non-native species. SDMs can be used in the testing of future technologies for environmental monitoring e.g. underwater robotics and eDNA. The overlaying of SDM models can also inform connectivity (EU Marine Strategy Framework Directives, OSPAR and Natura network requirements) and issues associated with MPA network coherence.

4.3 Clear specification of use

As with all modelling techniques, SDMs come with caveats as to how they should be used. The issue of uncertainty surrounding the model outputs (which is typically a map) needs to be addressed. Presently, uncertainty can be expressed spatially in supplementary maps but these are often technical and can be confusing for end users. Communicating model uncertainty is discussed in more detail in section 4.4. However, a simple approach to the limitations of SDMs would be to highlight what these should and shouldn't be used for. For example, SDMs can inform where monitoring should be targeted but key decisions relating to designation should not be based solely on SDMs without verification monitoring.

One of the objectives of the MarPAMM project is to provide a SDM toolkit to aid policy makers/ end users with key decisions relating to monitoring and management. Recognising that policy sectors are advised by their own scientific advisors, this toolkit should also provide information allows the clear explanation of the underlying techniques and their limitations.

4.4 Communicating "how good" the model is

The output of a Species Distribution Model is typically a map indicating predicted distribution, e.g. probability of presence or habitat suitability. There will be uncertainty in this prediction. The predictive ability of the output of an SDM will vary between different SDMs, but possibly also spatially within a single distribution map. This uncertainty needs to be communicated to the end user. In addition, the data and methods used to produce the SDM may have implications on how the outputs should be applied for marine management.

Depending on the modelling method and intended use for the SDM, there are statistics which give an indication of the model's accuracy e.g. area under the receiver operating curve (AUC) or amount of variation explained by the model e.g. R^2 and p-values. Maps of uncertainty can also be generated, for example mapping model residuals or though boot strapping techniques. However, these can be technical and their importance is not always communicated to end users. This may result in the uncertainty not being considered when SDM outputs are used in marine management. Detailed descriptions of data and modelling methods used, decisions made during the modelling process and model testing are often described in the academic literature which is inaccessible to policy (or other stakeholders) wishing to use the model. Therefore information needed by stakeholders to determine “how good” the model is, is not always available in a usable format and can easily be separated from the main output map during communication.

The necessary information could be included in the metadata for the output maps of the species distribution models, in addition to other metadata which would already be required for others to use the maps such as coordinate reference system, copyright and licences. The information could include both an assessment of quality of input data and modelling procedure used, as well as predictive accuracy of the map. If a standard structure is adopted for the metadata then this will allow users to become familiar with it, and know where to go for specific information (Table 2). The metadata categories could eventually develop into a quantitative scoring system, similar to MESH scores used for habitat maps from survey (https://webarchive.nationalarchives.gov.uk/20101014083419/http://www.searchmesh.net/Default.aspx?page=1635), which could be used to assess the quality of an SDM. Standardising metadata terminology would mean that only a single glossary of key-words would be needed to cover all the SDMs and would allow SDMs to become searchable, if added to a central repository. There are a number of marine data repositories that cover the Interreg VA area (e.g. NMPI, Marine Atlas, Emodnet) however none currently hold modelled species distribution maps. There could be potential to add SDMs to one of the existing viewers to enhance communication of model availability and quality with stakeholders.

Table 2. An example of metadata fields which could improve the communication of the quality and limitations of SDMs to stakeholders.

Field	Description	Example
Model type	Type of modelling	MAXENT
Model run date	Date/Time	23/02/2019 09:40
Modelling platform/packages	Package used for model development	R mgcv (ver 3.0)
Data used	Source datasets and when accessed	Bathymetry: Emodnet v2.0 accessed on 31/02/2019
Health criteria	1. Input data: Choices of input data have biological basis?	Occurrence and environmental, pressure
	a. Co-linearity checks	Yes
	b. Sense-checking of co-varying variables	Tidal height and current co-vary but both kept in model because both are thought to affect species distribution
	2. Model selection and procedures	AIC
	3. Biological validity of driving variables: critique	Hard substrate removed because no biological link with species biology
Recommendations	Recommendations for application	Recommended for use for survey planning
Limitations	Limitations (plain English) – communication of uncertainty: recommendation of accompanying map output	Because of data uncertainty model is not recommended for MPA designation without further verification monitoring
Uncertainty	Uncertainty map or other summary	Overall model uncertainty described but model technique does not allow uncertainty mapping

4.5 **Need for model verification**

SDMs need to be verified to test how accurate the predictions are and provide confidence in their applicability to marine management decisions. Verification can be done using occurrence data that was available at the start model development process, however it is preferable to collect new data with which compare against model predictions. Other datasets could be searched that may provide true absence data, for example records of bedrock biotopes would indicate definite absences for maerl. Although, for many species it is not possible to identify true absences from existing databases because whether a species was looked for but not found, is not usually recorded.

Collecting new marine benthic data from surveys is expensive, however within the Interreg VA region there are a number of organisations that do undertake surveys. If there is a desire to validate a specific SDM in a certain location, this could be added as a contingency objective of a suitable survey that was already planned. This would require coordination between partners, sharing cruise schedules to help identify potential opportunities, ensuring that wish lists are available in time for cruise planning and prioritising requests. A data wish list could be hosted and updated on a platform available to all stakeholders.

4.6 **Review of models**

The output of an SDM will depend on the species occurrence and environmental data used to produce it. This means that there is a possibility that they can become out of date as input datasets are improved. There may be reluctance by stakeholders to use SDMs in marine management if they are perceived to be out of date. Version control and planned future updates could be included in the metadata. Model code could also be published which would allow users to re-run the model themselves if they feel it is necessary, although this may lead to issues with intellectual property rights. Facilitating a dialog between those who produced the model and those wishing to use it would allow discussions about whether an update is required for a particular use and getting the model updated if deemed beneficial.

4.7 **Use in scenario modelling**

SDMs can be used to predict the distribution of species under different scenarios such as climate change or the spread of invasive non-native species, and hindcasting. This requires environmental data layers for the new (or previous) environments e.g. ocean acidification, wave base under climate change etc. These datasets are beginning to become available. As mentioned in section 4.1, one aim of MarPAMM could be to act as a central hub for the current state of SDM knowledge. Links to available environmental data suitable for scenario modelling for benthic SDMs could be included here.

4.8 **Summary of Recommendations**

The discussions around barriers led to the following recommendations which the MarPAMM benthic species distribution modelling work package will aim to implement and test:

- Improve dialogue between modellers and stakeholders who wish to use models, to ensure that they are fit for the intended purpose and are up to date.

- Increase coordination between organisations who undertake surveys to maximise on potential opportunities for field verification
- Set up a central hub of current knowledge and status of benthic species distribution models in the Interreg VA area, to include links to appropriate datasets and technical guidance.
- Develop standard metadata which can be used to assess the quality of the SDM output and communicate uncertainty more effectively.
- Produce a benthic modelling toolkit for stakeholders on how SDMs can be used to improve MPA management in the region

5 Review of MarPAMM SDM objectives

The workshop ended with a review of the MarPAMM benthic SDM objectives to consider which of the issues and barriers identified during the workshop are of particular relevance to the MarPAMM SDMs, and how these might be overcome.

Environmental and species occurrence data are stored in various formats and locations in the different countries within the Interreg VA region. For example, particle size analysis (PSA) data for seabed sediments are available for UK waters from the British Geological Survey (BGS) data portal, however this does not include any data from the Republic of Ireland. There are also additional records held by AFBI and MSS which may not be in the BGS dataset. These datasets need identifying and integrating into a single dataset to allow SDMs to be created for the whole Interreg VA region. Communication between the individuals undertaking the SDMs is therefore important to coordinate finding available data and, where possible, ensure consistency between data layers used. One suggestion was to have a forum or message board just for those individuals from each project partner who are developing the models, so that they can share information. This forum could also be used to discuss any problems faced during the SDM development and to suggest solutions, thereby increasing knowledge exchange between the project partners. By using a written forum, there is a record of the discussions which could be useful when developing the benthic modelling toolkit towards the end of the project.

Some of the technical issues identified during the case studies (section 3.2) are likely to be met during the development of the MarPAMM SDMs. In particular, as MarPAMM is aiming to model the distribution of horse mussel beds and maerl, how to model reef or aggregation forming species is already a known issue. Some solutions to this were suggested at the workshop, however a way forward can only be decided once the occurrence data have been collated. These technical issues could be discussed on the modellers' forum, and any sources of information and solutions could help populate the briefing notes suggested in section 3.3.

There are specific joint surveys planned in the MarPAMM project for model validation. However, project partners will undertake surveys as part of other work which may provide additional opportunities for validation. Communication between partners, at the work package teleconferences, about specific data gaps and survey plans should ensure that any additional field validation opportunities are identified. It would be useful if partners are made aware of any data request while the cruise programmes are being planned.

The benthic SDMs produced by work package T2 will feed into the regional MPA management plans that are being developed by MarPAMM work package T5. The SDM outputs will provide feature presence and extent maps which could guide management measures. Conversations are needed between the two work packages to ensure that the outputs produced by work package T2 are usable by work package T5. Information which would be useful for work package T2 includes; 1) the preferred metric to be mapped by the SDM e.g. probability of presence or presence/absence – the latter would require decisions about what threshold to use to classify as present, 2) how will the SDMs be used in the management plans, e.g. combined with sensitivity information and overlaid with pressures and activity information and 3) what format to provide the outputs, e.g. static map or GIS layer.

There are also other Interreg VA projects, specifically COMPASS, which could provide environmental data which may benefit the benthic SDMs being produced by MarPAMM. Therefore, communication is also needed between the projects to ensure that SDMs are based on the best available environmental data for the region.