WORKSHOP ON THE USE OF PREDICTIVE HABITAT MODELS IN ICES ADVICE (WKPHM)

VOLUME 3 | ISSUE 67

This document is the product of an expert group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the view of the Council.

ICES Scientific Reports

WORKSHOP ON THE USE OF PREDICTIVE HABITAT MODELS IN ICES ADVICE (WKPHM)

Editors

Chris Rooper

Authors

Chris Rooper • Laura Robson • Julian Burgos • Daniel van Denderen • Anna Downie • Jodi Pirtle • Patricia Puerta • Peter Auster • Pål Buhl-Mortensen • Jackson Chu • Ana Colaço • Jessica Finney • Ana Garcia-Alegre • Matthew Gianni • José Manuel González • Miriam Tuaty Guerra • Kerry Howell • Rebecca Langton • Anna Metaxas • Telmo Morato • Aurora Nastasi • Eugene Nixon • Javier Murillo Perez • Irene Pérez-Rodríguez • Roland Pesch • Jesus Rivera • David Stirling • Patrick Thompson • Ana de la Torriente • Beatriz Vinha

ICES
CIEM

International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

Contents

i Executive summary

The UNGA Resolution 61/105 and Regulation EU 2016/2336 require Regional Fisheries Management Organisations (RFMOs) and EU Member States to prevent significant adverse impacts on vulnerable marine ecosystems (VMEs) by deep-sea fisheries by identifying areas where VMEs occur or are likely to occur. In recent years, ICES experts have suggested the use of predictive habitat models (PHM) to identify areas where VME are likely to occur in the absence of documented observations. The ICES workshop WKPHM was therefore tasked with developing standards for data and modeling approaches that could be accepted for use in providing ICES advice. In addition, the WKPHM was requested to provide some desired criteria for models, both in terms of application and presentation that would be useful in communicating ICES advice WKPHM reviewed published standards for PHM from the peer-reviewed literature and utilised expert opinion from workshop participants to develop a list of recommended criteria that focused on a few key themes in PHM. These themes included transparency in the decisions made about the data and models used, for PHM and the assumptions that were inherent in the treatment of these data and models. Furthermore, they included the need for reproducibility of results and clear reporting of key components of the modeling, such as independent variables utilized for predictions. Standards were also suggested for communication of model results and uncertainty, so that managers could have a basis for evaluating the information to support decision making. Finally, suggestions on the best methods to evaluate models, including using independent data sets collected for that purpose, were generated.

Future steps suggested by WKPHM included the development of a systematic approach to generating new models, following the guidelines set out in this report, and a process for evaluating existing PHM against these guidelines, so that they could also be considered for supporting ICES advice. These guidelines will not only formalize a process for incorporating new information, but it will allow ICES to move forward with new and improved techniques as better data and models become available

ii Expert group information

Expert group name	Workshop on the Use of Predictive Habitat Models in ICES Advice (WKPHM)
Expert group cycle	Annual
Year cycle started	2021
Reporting year in cycle	1/1
Chair	Chris Rooper, Canada
Meeting venue(s) and dates	1-5 February 2021, by correspondence (30 participants)

1 Background and introduction

1.1 Introduction to current VME process in ICES and aims of the workshop

In 2006 the UN General Assembly resolution 61/105 adopted language to prevent significant adverse impacts on vulnerable marine ecosystems (VMEs) by deep-sea fisheries "consistent with the precautionary approach and ecosystem approaches" through, amongst other things, identifying areas where VMEs are known or likely to occur. The FAO Guidelines (FAO, 2009) establish criteria for identifying areas where VMEs are known or likely to occur and outline the types of information required to conduct assessments of the potential impacts of deep-sea bottom fisheries (FAO Guidelines, para 42 and 47). These include using the 'best available scientific and technical information on the current state of fishery resources and baseline information on the ecosystems, habitats and communities in the fishing area, against which future changes are to be compared, including through identification, description and mapping of VMEs known or likely to occur in the fishing area.'

The provisions of the UNGA resolutions and the FAO Guidelines related to identifying areas where VMEs are known or likely to occur have been incorporated into the bottom fisheries regulations adopted by the North East Atlantic Fisheries Commission (NEAFC) and the Northwest Atlantic Fisheries Organization (NAFO). The EU regulation for the management of deep-sea fisheries in EU waters (EU Regulation 2016/2336) also requires that measures adopted for the management of deep-sea fisheries in the areas covered by the regulation are consistent with the UNGA resolutions and that these measures should be supported by the "best available scientific and technical information…to identify where VMEs are known to occur or are likely to occur." The regulations encourage the use of "different types of marine scientific research, such as, inter alia, seabed mapping, mapping of VMEs based on information from the fishing fleet, on-site camera observations from remote vehicles, benthic ecosystem modelling, comparative benthic studies and predictive modelling to identify areas where vulnerable marine ecosystems are known or are likely to occur and in the adoption of conservation and management measures to prevent significant adverse impacts on such ecosystems, including the closure of areas to bottom fishing …"

ICES has provided scientific advice on where VMEs are known or likely to occur within the Northeast Atlantic following requests from NEAFC and the EU (e.g. ICES, 2019a, 2021). Current ICES advice is based on records of VME indicator taxa (species or taxonomic groups that may indicate the presence of a VME) contained within the ICES VME database, which is subject to quality control standards via the ICES Data Centre and the Working Group on Deep-water Ecology (WGDEC). Data submitted to the ICES VME database either confirms the actual presence of a VME on the sea floor from dedicated deep-sea surveys (e.g. a cold-water coral reef), or provides records that suggest the presence of a VME with varying degrees of uncertainty, e.g., bycatch of a VME indicator taxa, such as gorgonians (sea fans), from a fishing vessel (see section 1.2 for further information on the VME database).

Over the last ten years, ICES experts have advocated for the potential use of Species Distribution Models (SDM) and Habitat Suitability Models (HSM), referred to as 'Predictive Habitat Models' (PHM) herein, as

a tool to identify areas where VMEs are likely to occur. In 2014, WGDEC reviewed existing 'terrain-based models' for predicting VME indicator taxa distribution and concluded that published (peer reviewed) models should be taken into consideration for management decisions on deep-sea ecosystems (ICES, 2014). In 2018, the Review Group on Vulnerable Marine Ecosystems (RGVME) also recommended that statistical modelling techniques that produce predicted probability surfaces for VME indicator taxa should be further investigated, for the provision of new information on where VMEs are likely to occur and their spatial extent, to support ICES advice to the EU and NEAFC.

Following the recommendations of RGVME, the 2019 WGDEC meeting was jointly held with the Working Group on Marine Habitat Mapping (WGMHM), who explored the use of PHM for mapping VME indicator taxa distribution in the North Atlantic. Through this work, WGDEC identified the availability of existing models with differing taxonomic resolutions, spatial extents and spatial resolutions and proposed next steps for model use (ICES, 2019b), whilst WGMHM developed a 'roadmap' setting out the proposed steps to facilitate the adoption of PHM in ICES advice (ICES, 2019c). The 'roadmap' clarified the need to generate a standard set of model outputs, to identify aspects such as which habitats/species to model, the spatial extent of the model, the minimum mapping resolution and how often the model should be revisited (e.g. re-running models when new data becomes available). In addition, WGMHM recommended a trial run for a subset of VME indicator taxa, to optimize the model approaches, with the final methods to be published as part of the ICES 'Transparent Assessment Framework'.

In support of these developments in the application of predictive habitat models for advice, ICES included four types of VME geophysical elements (i.e. seamounts, banks, coral mounds, and mud volcanoes), based in part on geomorphological maps, in their latest advice to identify areas that potentially support VMEs in EU waters (ICES, 2021). Nonetheless, to date, ICES advice has not used PHM to provide an evaluation of whether a VME or VME indicator is `*likely to occur*' in an area.

Together with the 'roadmap' developed by WGMHM, it was also proposed that a set of criteria needed to be derived, against which new and existing models could be reviewed, to determine appropriate standards for their use for scientific advice (ICES, 2020). These criteria could form the backbone of a 'benchmarking' approach to assess PHMs, ensuring that only models which meet specific quality criteria are used within ICES advice. Benchmarking[1] is a system commonly used for fish stock assessments which aims to gather consensus agreement on a method that is to be used in future update assessments. This sets the 'benchmark level' for assessment, equating to the 'best available' method that ICES advice can be based on. If new data or methods become available that might improve the assessment, proposals can be made for a renewed benchmark.

The overall aim for the Workshop on the Use of Predictive Habitat Models in ICES Advice (WKPHM) was to develop benchmark standards for the use of PHM in ICES advice related to the distribution of VMEs. The 'roadmap' developed by WGMHM and the need for benchmarking criteria formed the basis for the Terms of Reference (see Section 1.3).

---

[1] http://www.ices.dk/community/documents/advice/introduction%20to%20benchmarks%20at%20ices.pdf

1.2   Data and models available for VME in ICES regions

PHM have been widely used for terrestrial and marine habitats to predict the occurrence of suitable habitat, the presence of a taxon or habitat, or the abundance of a taxon or taxonomic group. All PHM link environmental predictor variables (independent variables; for VME PHM these are typically variables describing the physical characteristics of the environment such as bathymetry, geomorphology and oceanographic conditions) and response variables (dependent variables which tend to be biological in nature) in a statistical or machine learning framework to predict a continuous or categorical output that can be geo-located or mapped.

The dependent data or response variables (e.g. presence, presence-absence, or abundance) available in the ICES regions can be found in the ICES VME database or from other directed studies or databases (e.g. OBIS; PANGAEA). The ICES VME database stores records of VMEs, VME indicators and the locations of where neither of these have been observed (absence data), as described by a VME database schema[2].

Data from the ICES VME database are collated during an annual VME data call and are accessible via the VME data portal[3]. Much of the data are openly available to download, with a small number of restricted-access records.

Independent data or environmental predictor variables of VME distribution are generally available for the ICES regions on a global or basin scale. For broad-scale models, environmental predictors that are themselves the outputs of models, are typically used. They tend to make use of global bathymetry models (Becker *et al.,* 2009; Weatherall *et al.,* 2015) for deriving depth and topographic variables such as slope, rugosity, and bathymetric position index; and until recently (Tozer *et al.,*2018), this has limited the resolution of model outputs to 30 arc seconds (~ 1's of km). Oceanographic data used in these basin-scale models are generally only available from models at coarser resolution than the bathymetry models (e.g. 10's of km). A number of global and basin scale models for VME distribution have already been published using these methods (e.g. Davies *et al.,* 2008; Tittensor *et al.,* 2009; Davies and Guinotte 2011; Yesson *et al.,* 2012; Howell *et al.,* 2016).

For fine-scale models, there may be finer and more informative predictor variables available that can be used in generating higher resolution models of VME indicator taxa distribution. However the spatial coverage of the fine-scale model predictions is usually limited by the spatial extent and coverage of the predictor variables. These higher resolution regional and national models tend to make use of bathymetry (and possible backscatter) data collected by multibeam sonar to derive terrain variables, but may also utilize a much wider range of potential environmental variables relevant for species distributions, and therefore tend to produce more accurate models (e.g. Ross *et al.,* 2015). Although these models have limited spatial coverage, they are particularly useful for spatially constrained areas where there is a need to understand and define VME distribution more precisely (e.g. individual nation's EEZs or individual seafloor features such as seamounts). The resolution of published fine-scale models is varied but tends to be in the order of a < 200 m grid cell size. As a result of the smaller grid cell size in these models versus basin-scale / global models, the estimates of area of predicted suitable habitat for VME and VME indicator species derived from such models tends to be much smaller, and this may be an important consideration when assessing spatial closures.

---

[2] http://www.ices.dk/data/Documents/VME/VME_Reporting_Format.zip

[3] http://vme.ices.dk/map.aspx

Within ICES management areas there are a number of existing models in the published literature that vary in spatial coverage, species considered, output variables and overall objectives. For example, there are published models of small-scale (both in terms of spatial resolution and spatial extent) regional predictions of the probability of suitable habitat for individual species such as Lophelia, *Desmophyllum pertusum*, (e.g. Dolan *et al.,* 2008; Rengstorf *et al.,* 2013) and basin-wide predictions of the impacts of climate change on the probability of VME presence under future climate scenarios (e.g. Morato *et al.,* 2020). Some of these existing models may be useful in delineating probable VME locations that can be considered for spatial management and implementing closures in the context of providing ICES advice.

There are also opportunities to improve and build on the models available in the ICES region by utilizing the data from the ICES VME database or other marine biodiversity databases, new and higher resolution environmental variables as they become available (e.g. GEBCO or World Ocean Atlas updates) and new, different or improved modelling techniques (such as joint species distribution models; e.g. Thorson et al., 2015). New models developed specifically to address ICES management issues, using data available in the ICES regions can improve the information available for providing ICES advice.

1.3   Need and terms of reference

In order for ICES to utilize PHM to provide an evaluation of whether a VME is `*likely to occur*', an agreed set of modelling standards is required. To develop these standards, the workshop had four terms of reference:

a)   Based on existing approaches, identify the methods for modelling vulnerable marine ecosystems (VMEs) that would be most appropriate for use within ICES advice, detailing 'required' and 'desirable' criteria, with emphasis on the deep-sea environment greater than 200 m (considering bias of preferential sampling), PHM techniques (including spatial display of uncertainty) and required validation steps for the modelled outputs);

b)   Develop clear standards for recording the caveats and assumptions inherent in the modelling method, for future use;

c)   Conduct a trial run for a small number of existing models to ensure that both the approach and outputs are fit-for-purpose;

d)   Review and recommend a set of criteria, similar to the existing ICES benchmarking system for regional fish stock assessments, under which new and existing predictive habitat models can be used for ICES scientific advice related to the distribution of VMEs.

1.4   Structure of the workshop and report

The workshop was conducted virtually over five consecutive days (February 1-5, 2021). The work was organized around plenary sessions and breakout groups. Both plenary and breakout group sessions were held daily except on a single day that was dedicated to conducting trial modelling work using provided data sets. There were three breakout groups covering the main foci of the terms of reference. The three breakout groups covered 1) independent and dependent data used for modelling, 2) modelling methods, uncertainty and model validation, and 3) required and desired criteria for modelling to be used in providing ICES advice. Each breakout group developed sets of assumptions regarding the topic of PHM and minimal and desired characteristics for PHM.

In the plenary sessions these products were then summarized and more fully developed into a set of standards that could be used to evaluate PHM to be used to provide ICES advice, as well as recommendations to ICES on the use of PHM and future directions in VME benchmarking.

The structure of this report follows the general structure of the workshop, with a summary from each of the major breakout group topics and a general summary with some recommendations and future directions. Finally, a set of standards for PHM and a required set of components to be reported for each PHM to be considered for providing ICES advice are presented in Annexes 2 and 3. These contain the minimal (required) criteria that PHM input data, methods and outputs must meet to be considered for use in ICES advice as well as a template for reporting the necessary characteristics of the PHM (an example is provided in Annex 4).

References:

Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B. and O'Hara, R.B. 2019. Standards for distribution models in biodiversity assessments. Science Advances, 5, p.eaat4858.

Becker, J.J., Sandwell, D.T., Smith, W.H.F., Braud, J., Binder, B., Depner, J.L., Fabre, D., Factor, J., Ingalls, S., Kim, S.H. and Ladner, R. 2009. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS. Marine Geodesy, 32(4), pp.355-371.

Davies, A.J., Wisshak, M., Orr, J.C., and Roberts, J.M. 2008. Predicting suitable habitat for the cold-water coral *Lophelia pertusa* (Scleractinia). Deep-Sea Research I55:1048–1062.Davies, A.J. and Guinotte, J.M. 2011. Global habitat suitability for framework-forming cold-water corals. PloS one, 6(4), p.e18483.

Dolan, M. F. J., Grehan, A. J., Guinan, J. C., and Brown, C. 2008. Modelling the local distribution of cold-water corals in relation to bathymetric variables: adding spatial context to deep-sea video data. Deep Sea Res. I 55, 1564–1579. doi: 10.1016/j.dsr.2008.06.010

Howell, K.L., Piechaud, N., Downie, A.L. and Kenny, A. 2016. The distribution of deep-sea sponge aggregations in the North Atlantic and implications for their effective spatial management. Deep Sea Research Part I: Oceanographic Research Papers, 115, pp.309-320.

ICES. 2014. Report of the ICES/NAFO Joint Working Group on Deep-water Ecology (WGDEC), 24–28 Feb-ruary 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:29. 70 pp.

ICES. 2019a. New information regarding vulnerable habitats in the NEAFC Regulatory Area. *In* Report of the ICES Advisory Committee, 2019. ICES Advice 2019, vme.neafc, *https://doi.org/10.17895/ices.advice.5580*

ICES, 2019b. ICES/NAFO Joint Working Group on Deep-water Ecology (WGDEC). ICES Scientific Reports. 1:56. 119 pp. *http://doi.org/10.17895/ices.pub.5567*

ICES. 2019c. Working Group on Marine Habitat Mapping (WGMHM). ICES Scientific Reports. 1:54. 28 pp. *http://doi.org/10.17895/ices.pub.5578*

ICES. 2020. ICES/NAFO Joint Working Group on Deep-water Ecology (WGDEC). ICES Scientific Reports. 2:62. 188 pp. *https://doi.org/10.17895/ices.pub.7503*

ICES. 2021. EU Request to advise on the list of areas where VMEs are known to occur or are likely to occur and on the existing deep-sea fishing areas (ref. (EU)2016/2336.). *In* Report of the ICES Advisory Committee, 2021. ICES Advice 2021, sr.2021.01. *https://doi.org/10.17895/ices.advice.7507*. The electronic data outputs can be accessed at *https://doi.org/10.17895/ices.data.7506*.

Morato, T, González-Irusta, J-M, Dominguez-Carrió, C, *et al.*,2020. Climate-induced changes in the suitable habitat of cold-water corals and commercially important deep-sea fishes in the North Atlantic. Global Change Biology. 26: 2181– 2202. https://doi.org/10.1111/gcb.14996

Piechaud, N., Downie, A., Stewart, H.A. and Howell, K.L. 2015. The impact of modelling method selection on predicted extent and distribution of deep-sea benthic assemblages. Earth and Environmental Science Transactions of The Royal Society of Edinburgh, 105(4), pp.251-261.

Rengstorf, A.M., Yesson, C., Brown, C. and Grehan, A.J. 2013. High-resolution habitat suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea. Journal of Biogeography, 40(9), pp.1702-1714.

Ross, L.K., Ross, R.E., Stewart, H.A., and 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. doi:10.1371/journal. pone.0140061

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast ground- fishes. ICES Journal of Marine Science, 72: 1297–1310.

Tittensor, D.P., Baco, A.R., Brewin, P.E., Clark, M.R., Consalvey, M., Hall-Spencer, J. Rowden, A.A., Schlacher, T., Stocks, K.I., and Rogers, A.D. 2009. Predicting global habitat suitability for stony corals on seamounts. J. Biogeogr. 36:1111-1128.

Tozer, B., Sandwell, D.T., Becker, J.J., Tea, B., Freund, Y., Harper, H., Beale, J., Wessel, P. and Smith, W.H. 2018. Global bathymetry at 15 arc seconds resolution with implications for the isostatic compensation of geological features in the MH370 survey area. In AGU Fall Meeting Abstracts.

Weatherall, P., Marks, K.M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J.E., Rovere, M., Chayes, D., Ferrini, V. and Wigley, R. 2015. A new digital bathymetric model of the world's oceans. Earth and Space Science, 2(8), pp.331-345.

Yesson, C., Taylor, M.L., Tittensor, D.P., Davies, A.J., Guinotte, J., Baco, A., Black, J., Hall-Spencer, J.M. and Rogers, A.D. 2012. Global habitat suitability of cold-water octocorals. Journal of Biogeography, 39(7), pp.1278-1292.

2 Considerations for dependent and independent data

2.1 General introduction to data issues

Implementation of PHM to inform management advice and conservation of VMEs requires careful consideration of the input data. It is important to bear in mind that while data and associated modelling needs should be driven by specific management objectives, the data sets can often accommodate multiple management objectives. Most important to consider is that data availability and level of detail will drive which PHM are possible and associated management objectives will determine which PHM to consider and implement. It is important therefore to consider the relationship between the data and the management objective, for instance the question of how VMEs are defined by the management objective (e.g. a single occurrence of as single indicator species or a dense aggregation of multiple taxa) will determine the type and amount of data required for PHM.

Both dependent and independent data sources are required to implement PHM. Dependent data are observations of taxon distribution with requisite geospatial and temporal information (e.g., coral observations with sample locations across the spatial-temporal domain of the model; Winship *et al.,* 2020). Presence data is sometimes the only available data for VME indicator taxa and can be derived from multiple sources and combined for certain PHM methods. Pseudo-absences or background points selected from within the model domain are used to parameterize some presence-only PHM. Absence data, either true or inferred, can be applied more broadly across PHM methods. Inferred absences can be generated from sampled locations where other species were recorded as present during the same survey. These inferred absences can have additional uncertainty associated with them if there was unknown consistency in the observations or survey methods, but they can be useful in facilitating the application of a presence-absence PHM, which is preferred over a presence-only method (Winship *et al.,* 2020). Abundance data is often rare for VME taxa, but allows for predictions of habitat-related numerical abundance, density, or biomass. Independent data are those that define the seascape setting in which the taxa are distributed and can be based on direct observation (e.g., bathymetry, temperature), derived from observations (e.g., seafloor slope, rugosity), or modelled (e.g., regional biophysical oceanographic modelling systems). Multiple dimensions of data quality exist on a continuum (Figure 2.1.1). In practice, it has been impossible to maximize data quality along all dimensions, as this would take unlimited resources. There are always trade-offs in data quality that must be assessed. For example, one of the most common trade-offs is between spatial extent and spatial resolution, as it is difficult to have a large spatial extent of data collected at a high resolution. The following sections address data caveats and assumptions and data standards for VME PHM. Our advice on VME PHM data is not restricted by current VME data limitations. Data collection and PHM methods may advance in time to get closer to our recommended good practices detailed below (e.g., a perfect dataset for building PHMs with the most accurate predictions will come from a randomly stratified survey inclusive of strata parsed by known environmental gradients and human activities). In the absence of such complementary data sets, or when those available data sets have high uncertainty (e.g., due to under sampling, spatial and temporal resolution), there may be simpler approaches that can be used to infer aggregations of VMEs (e.g., kernel-density analysis; Kenchington *et al.*, 2016). VME indicator taxa and their distributions can also be delimited using limited observations or inference of taxon distributions and other environmental parameters using qualitative methods, such as habitat suitability indices (Nephin *et al.,* 2020). We expect that data collection and modelling methods will continue to advance, offering many solutions to address information needs with respect to conservation and management of VMEs. However, careful preparation

and consideration of the input data, in terms of the quality, type and quantity need to achieve PHM objectives, is the most critical step in improving the performance of all models, regardless of choice of method.
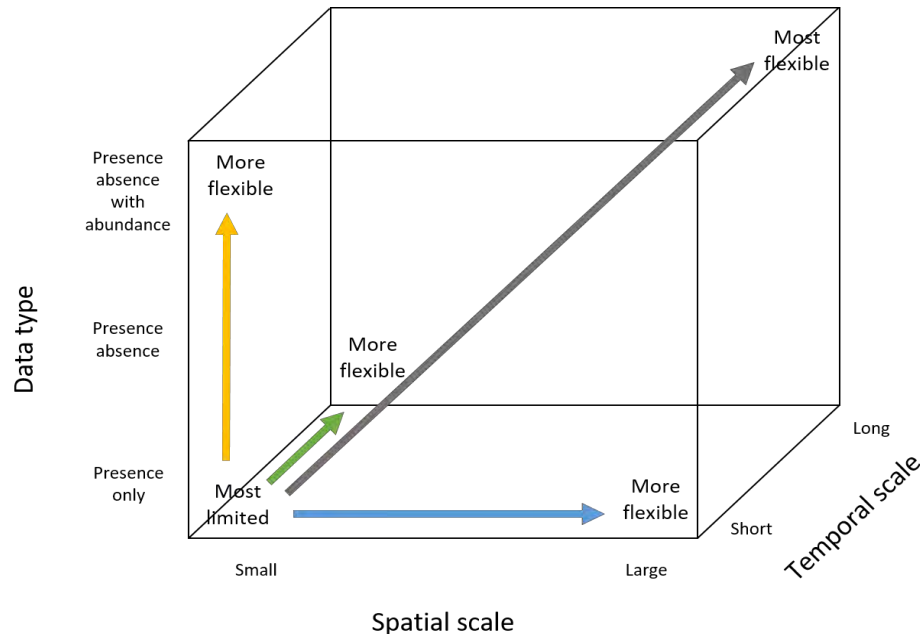

Figure 2.1.1: Dimensions of data quality. Type of data (y-axis); spatial scale (resolution and extent) (x-axis); temporal scale (resolution and extent) (z-axis). Other dimensions, such as taxonomic resolution and other sources of uncertainty, are not displayed and should also be considered. In this context, resolution refers to the relative spacing of data in time (e.g. data from monthly collections) and space (e.g. data from a 100 m by 100 m grid), while extent refers to the data coverage in time (e.g. data collected in 2015-2019) and space (e.g. data collected over the entire North Atlantic).

2.2 Caveats and assumptions to be reported for data used in PHM

PHM models rely on a series of assumptions about the link between the distribution of taxa and the spatial-temporal distribution of environmental conditions (Zurell *et al.*, 2020). It is important to remember that even the highest quality datasets and models will inevitably fail to fully meet these assumptions (Jarnevich *et al.*, 2015). Failure to fully meet these assumptions leads to error and uncertainty in the predictions of the model. Thus, the degree to which models can provide accurate predictions and useful inference depends on how closely these assumptions are met. Therefore, it is critical that these assumptions, and the degree to which they are met by the data, are clearly communicated when interpreting the results of the model.

Critical assumptions include those about the ecology of the species or taxonomic group and the quality of the dependent and independent data (Table 2.2.1). The two key ecological assumptions focus on the link between the species distribution and the environmental gradient. The assumptions about the dependent data focus on whether the observations of the species/taxon are representative of the true distribution of the species in space and time. The assumptions about the independent data focus on whether the environmental variables included in the model are accurate representations of the environmental gradients that structure the distribution of the species in space and time. The assumptions about the model focus on how the dependent and independent data are linked statistically in the model that is used. Finally, it is important to remember that the predictions made by the model are hypotheses about the distribution of the

species/taxon and that these hypotheses require independent validation (Jarnevich *et al.*, 2015). A conceptual illustration is provided of how the assumptions outlined in Table 2.2.1 relate to the ecological processes that determine species distributions, the independent and dependent data, the model, the predictions, and interpretation and inference (Figure 2.2.1).

Uncertainty that results from failure to fully meet the assumptions of the model should be reported clearly as a series of caveats (Jarnevich *et al.*, 2015). These provide guidance for end users of the model outputs for how to interpret the model predictions, given the limitations of the data or model. For example, if part of the study region was less comprehensively sampled, it should be communicated that the model predictions are less certain in that area. Areas that were well sampled and thus model predictions may be more certain should also be clearly reported. In combination, clear assumptions and caveats provide a framework for acknowledging and communicating sources of uncertainty and error so that they can be taken into account by managers and decision makers.

Table 2.2.1: Assumptions of dependent and independent data in VME PHM.

|	Assumption	Description
Ecological	The species/taxon is at equilibrium with respect to the environment	Assumes no historical contingency (e.g., historical disturbances, fishing, environmental change, colonization history, interactions between species)
	Niche conservatism	Species/taxon responses to environmental conditions are consistent across the entire study extent/range
Dependent data	Presences or absences are true	Data are free from observational bias, or differences in observations have been accounted for in the model
		When randomly chosen pseudo absences are used they are representative of the distribution and prevalence of true absences
	Observations are independent	Spatial and temporal autocorrelation have been adequately addressed
		Individuals are only reported once
	Data are representative of the study extent	The full environmental gradient present in the study extent has been representatively sampled
	Data units meaningfully represent the species distribution	Sampling was adequate across region or feature of interest (e.g. North Atlantic, seamount, bank), with respect to the life history characteristics of the species in question.
Independent data	Independent data capture key factors structuring species distribution	Predictor variables are assumed to influence species distributions and key variables are not missing
	Predictor variables are independent and uncorrelated	Collinearity among predictors has a negligible impact on the results of the model and the model predictions
	Independent data are accurate representations of the true values	Measurement, interpolation, or modelling error has negligible influence on the modelling results
	Spatial and temporal scales of the data are appropriate for goals of the study	Resolution is appropriate to capture the relevant ecological processes
Model based assumptions	Data meet the assumptions of the model	Data are appropriate for the model
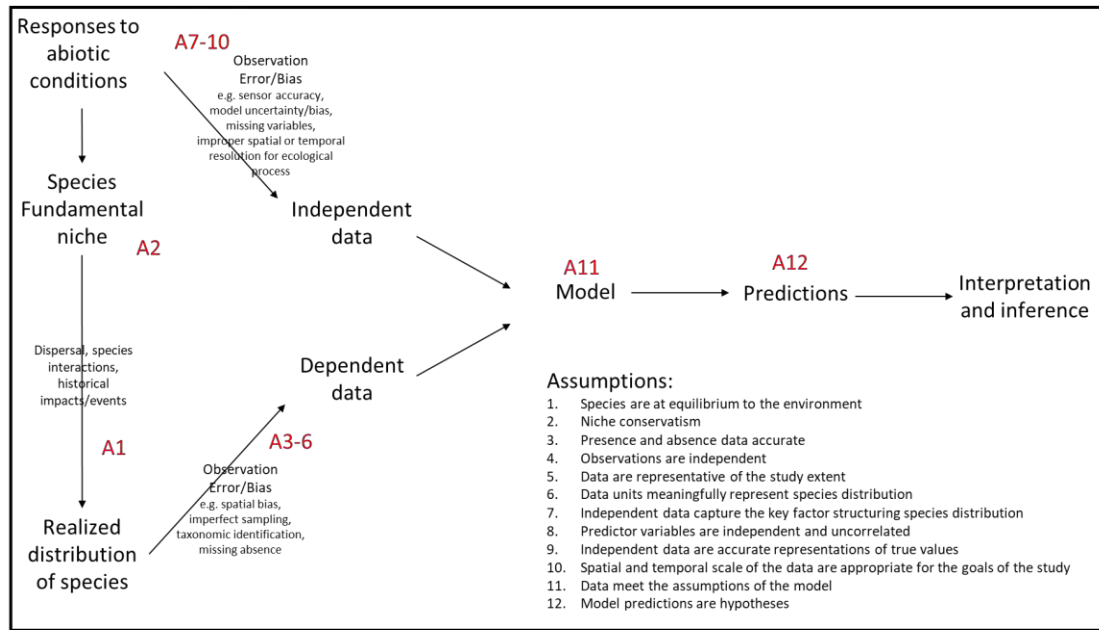
Figure 2.2.1:
Conceptual illustration of how the assumptions provided in Table 2.2.1 relate to the ecological processes that determine species distributions, the independent and dependent data, the model, the predictions, and interpretation and inference. The arrows illustrate the logical flow between these different components with indications of where each of the assumptions is relevant. The smaller text associated with the arrows provides examples of why the assumptions may not be met. The red text indicates the corresponding assumption number listed in the Assumptions list.

2.3 Minimum standards for data used in PHM

Several good practice protocols for reporting data considerations used in PHM have been developed in the wider ecology, evolution, and conservation science communities (Araújo *et al.,* 2019; Winship *et al.,* 2020; Zurell *et al.,* 2020). The general goal with applying such recommended protocols is to maximize transparency and reproducibility of the models while outlining data assumptions and considerations that might influence how predictions could be interpreted should they be applied in management decisions. We suggest reporting on relevant metadata and information that applies to how data were collected, accuracy and precision (resolution) of the data, processing methods applied to the data, and model assumptions or factors that may introduce uncertainty into the model results (Table 2.3.1). Gaps or missing variables with known effects should be acknowledged and assumptions regarding the data used should be clearly outlined.

Dependent data Table 2.3.1 outlines commonly agreed upon standards for describing dependent data used in the development of PHM in relation to VMEs. The type of dependent data should be clearly stated. For example, as absence or abundance data are rarely available for VMEs, clear assumptions and caveats associated with those data should be outlined if they are used in the model (e.g., true or inferred absences used in presence-absence models). Any processing of the dependent data prior to inclusion in the development of PHM must also be documented. For example, for some presence-only models, the probability of presence of suitable habitat may not be easily distinguished from clustering (Boria *et al.,* 2014) or spatial bias in sampling (Fithian *et al,* 2015). In these cases, techniques such as spatial thinning or bias correction of the observations may be used, but should be documented in the meta-data associated with the observations used in developing the PHM.

Dependent data for use in PHM may come from existing biodiversity databases (e.g. OBIS), the ICES VME database or new research cruises. The group agreed on an overarching criterion that data sources should not be limited to the ICES VME database, and should be all the available data. However datasets must be quality controlled, and should, at least, meet the minimum quality control (QC) standards of the VME database found in ICES (2020). These QC standards include basic data checks, such as all the required data fields are populated, the positional data does not occur on land or outside the ICES region, codes for habitat type are consistent with the database, etc. The 'required' criteria for use in PHM are therefore that the best available data are used, but also that data used must have been quality assured (QA) using national or international best practice guidelines. Metadata/data should also be reported on following the reporting standard in Annex 3. If multiple sources of dependent data are combined for modelling, documentation of any methods used to prepare or combine the data must also be provided. There is no specific QA standard set for new data submissions to the VME database, however there is an expectation that all data suppliers follow national or international best practice guidelines in the QA of their data, for example, Batley (1999); Rumohr (2009); Howell *et al.* (2014); Turner *et al.* (2016) (ICES 2020). Examples of these standards include the use of high definition cameras in the collection of image data and the identification of VME indicator taxa by trained observers using a common and regionally specific identification guide (Howell *et al.* (2014). In addition, the VME database uses a standardised data submission template which details mandatory, conditional and optional data fields for all new VME data submissions. These data fields therefore form the metadata requirements of the VME database records and would enable completion of the reporting standard proposed in Annex 3.

Consideration was also made of the taxonomic resolution of data. Confidence in the prediction of different taxonomic or functional groups increases if predictions are based on the highest possible taxonomic resolution (Jansen *et al.,* 2018). According to Winship *et al.* (2020) models developed using data with low taxonomic resolution may mix species with different life-histories and environmental requirements, potentially resulting in inaccurate (broad) predicted distributions and increased model uncertainty. However, models of functional groups, or otherwise reduced taxonomic resolution, may be sufficient to address some management applications, but insufficient for others such as particular species of concern (e.g., endangered species) (Winship *et al.,* 2020). The lists of VME indicator taxa used for different RFMOs generally include the Class as the lowest taxonomic rank (Thompson *et al.,* 2016). The 'required' criterion is therefore that the dependent data uses a taxonomic level from Class to species or a defined assemblage type (e.g. sea-fan type corals), based on quality controlled data and expert opinion, unless otherwise justified.

Independent data. The types of independent data e.g., how those were collected, derived, or modelled, should also be clearly reported. For independent (environmental) data, ICES do not have specific QA standards that are required to be followed when using these types of data within ICES work. However, as a minimum, it is expected that datasets are validated following best practice QC standards, for example those made available via the EMODnet Physics data portal[4].

The 'required' criteria for independent data are that the range of the environmental parameters used (e.g. depth limits) are informed by any known the biological tolerance limits for the specific VME indicator taxa being modelled, or at minimum, they are inferred from those evidenced and documented for 'proxy' taxa, with expert evaluation of their suitability for the taxa being modelled. Such inference from 'proxy' taxa will be unacceptable without this expert evaluation and description. In some cases, 'proxy' taxa may not be known or there may be incomplete overlap between the modelled species and the 'proxy' taxa (for example, where a species might extend to a deeper depth than its 'proxy' taxa). In these cases, the relationship

---

[4] https://emodnet.eu/en/physics

between the range of predictions from a model and the environmental coverage of the data can be used to measure uncertainty in the inferred relationships and predictions (Stephenson *et al.,* 2020). For the desired criterion, the tolerance limits for the VME indicator taxa must be known and documented (Table 2.3.1). In all cases, the environmental conditions in the location (or time) where predictions are made must be within a plausible range for extrapolation of response – predictor relationships. It is important to note that in many cases the experts on VME indicator taxa may be the modellers themselves, part of the modelling team or expertise cited from previously published research from the literature. Thus, the use of the term expert in the context of evaluation of independent data does not always imply an external review of the PHM by an outside expert, but integration of expertise that can speak to the ecological relevance and appropriate range of conditions of environmental variables for the taxa whose distribution is being modelled.

PHM generally require that independent variables continuously cover the entire model domain. Most of the time, these variables are obtained by interpolating point estimates to a grid, generating gridded observations from a oceanographic model, or generating a grid from remotely sensed observations, such as satellite data. Depending on the spatial resolution of the model domain, this can result in either aggregation to a higher spatial resolution or downscaling to a smaller spatial resolution. The method for interpolating a variable to a different resolution and the original spatial scale of the observations should always be reported (Table 2.3.1). Ideally, goodness-of-fit measures (such as correlation measures) between the interpolated data and the original scale should be provided so that the variability introduced by changing data resolution can be assessed. Even when available, this uncertainty is usually not accounted for when they are used in PHM (Hijmans *et al.*, 2005; Stoklosa *et al.*, 2015), but at a minimum this issue should be considered explicitly in developing PHM for ICES management advice (Table 2.3.1).

Collinearity in independent variables can also be a concern in PHM (Araújo *et al.,* 2019, Zurell *et al.,* 2020, Winship *et al.,* 2020). Collinearity can make it impossible to clearly separate the confounded effects of related variables in PHM. If the relationship between the variables is non-stationary, this can result in poor predictive ability in areas where the collinear relationship changes. Ideally, only variables that are not correlated would be used in a PHM (Araújo *et al.,* 2019), however, if collinear variables need to be used in a PHM, the extent of the correlation and its potential effects on the retention on the PHM and the robustness of the results should be demonstrated.

Spatial and temporal scale. Details of the spatial and temporal scales and resolution of model input data (dependent and independent) are considered mandatory in standard protocols for reporting models (Table 2.3.1), to ensure reproducibility and transparency of the methods and to facilitate model reliability and relevance (Araújo *et al.,* 2019, Zurell *et al.,* 2020, Winship *et al.,* 2020). The predictive accuracy of a model can be reduced if inappropriate spatial and temporal resolutions of dependent and independent data are chosen, which in turn, can mask the relationships between species and the environment (Araújo *et al.,* 2019). The spatial and temporal scales for both dependent and independent data is also a required criteria for reporting in Annex 3.

Table 2.3.1: Minimum required and desirable criteria for PHM development to predict the distribution of VME in providing ICES advice.

		UNACCEPTABLE	REQUIRED	DESIRED
DEPENDENT (BIOLOGICAL) DATA	Data quality	Sampling design for data collection not described.	All the available data that meet QC standards are used, with a clear description of sampling design(s) and data collection.	Data are sampled via systematic sampling design (which are the same for biological and environmental data) and standardized methods are used for sample collection and processing. A clear description of a robust sampling design is provided.
		Data have no quality control and/or associated metadata. When multiple data sources are combined, there is no description or consideration of the differences.	Quality control of data undertaken, based on metadata, meeting the minimum standards of ICES VME database (e.g., data QA follows national and/or international best practice guidelines and details of which guidelines followed provided). When multiple data sources are combined, the same quality control criteria are followed.	Same as required criteria
		No metadata provided on data sources or the treatment of data.	Metadata/data are reported, following the reporting standard in Annex 3, including a description of the data sources, sampling effort, resolution and extent, and any pre-processing of the data, such as combining multiple data sources and possible spatial thinning or bias correction of data.	Same as required criteria
		Presence-only data is used when absence data is available or the information exists to infer absence data at sampled locations	Presence-only data available.	Presence and absence data and/or abundance data available.
	Caveats, bias	Caveats, bias and assumptions in dependent data are not reported.	Caveats, bias and assumptions independent data are clearly reported, acknowledging sources of uncertainty and error.	Same as required criteria
	Taxonomy	Response variable type is not clearly explained.	Response variable type is clearly explained with any caveats on interpretation of outputs.	Same as required criteria
		Taxonomic level is above Class or uses vernacular names.	Taxonomic level is from Class, species or a defined assemblage type (e.g. sea-fan type corals), based on quality-controlled data and expert opinion.	Taxonomic level is species, based on expert identification.

		UNACCEPTABLE	REQUIRED	DESIRED
INDEPENDENT (ENVIRONMENTAL) DATA	Data quality	Data have no quality control and/or associated metadata.	Quality control of data undertaken, based on metadata of quality assured (QA) databases or reported survey design and methodology.	Data are sampled via systematic sampling design (same for biological and environmental data) and standardized methods are used for sampling. Clear description of robust sampling design is provided.
		Source(s) of independent data not provided.	Independent data source(s) explained and reported following reporting standards in Annex 3.	Same as required criteria
		Uncertainty is not considered in the environmental data.	Uncertainty of environmental predictors is characterized and accounted for (e.g., use of different analyses and scenario testing).	Uncertainty of environmental predictors is minimized (e.g., use of different analyses and scenario testing).
	Variable choice	Large number of predictor variables included without *a priori* consideration of their ecological relevance evidence of the ecological relevance and/or relevance of the range of predictor variables used not evidenced or reported.	Predictor variables have a demonstrable association with the response variable (observational, statistical or theoretical) at the spatial and temporal resolution of the model.	Predictor variables are proximal variables with a confirmed ecologically relevant association with the response variable at the spatial resolution and extent of the model.
		Predictor variables and their ranges are informed by the biological tolerances inferred from other 'proxy' taxa with no expert evaluation for their use.	Predictor variables and their ranges are informed by any known biological tolerances of the VME indicator taxa being modelled, as documented in peer-reviewed studies; or are inferred from those evidenced and documented for 'proxy' taxa with expert evaluation approval for their use.	Predictor variables and their ranges used are defined by the biological tolerance limits of the VME indicator taxa being modelled, as documented in peer-reviewed studies.
	Data processing	Spatial accuracy and resolution are substantially inconsistent between biological and environmental data, e.g., m to hundreds km.	Spatial accuracy and resolution may be inconsistent between biological and environmental data, but data selected are quality-controlled and based on ecological relevance.	Spatial accuracy and resolution of biological and environmental data are the same, or consistent, and are ecologically relevant.
		Native spatial resolution and/or downscaling / aggregation processing methods to match spatial accuracy and resolution between biological and environmental data are not reported.	Native spatial resolution and/or downscaling / aggregation processing methods to match spatial accuracy and resolution between biological and environmental data are fully reported.	Downscaling / aggregation and method used reported and goodness of fit measures for downscaling/aggregation method reported. The effects of variable downscaling/aggregation on model output are tested.
		Derived variables and methods of calculation not described.	Derived variables described and calculations and references provided.	Same as required criteria
	Collin-	Collinearity has not been investigated or is simply acknowledged or uncorrected.	Collinearity is addressed (e.g. correlation coefficient (r) and variance inflation factor (VIF)) and alleviated, or an approach/model insensitive or able to handle it is used.	Lack of collinearity is demonstrated.

	UNACCEPTABLE	REQUIRED	DESIRED
	Collinearity has not been investigated or acknowledged.	Selection of variables to retain from collinear pairs is transparent and has a logical basis. If collinear variables are retained, robustness of the results to collinearity are demonstrated.	Selection of variables to retain from collinear pairs is transparent and is based on significant contribution to model and ecologically relevant association. No collinear pairs of variables are retained in the model.
SPATIAL AND TEMPORAL	Spatial and temporal extents, resolutions and location of the study used are not justified.	The spatial and temporal extents, resolutions and location of the study are justified as evidenced from peer-reviewed studies, data availability and/or quality-controlled databases.	The full spatial and temporal, extent, resolution and distribution of the VME indicator taxa are known and used, including current and historical distribution of the VME/indicator.
	Model includes outdated data from locations where natural or anthropogenic influences have changed the response – predictor dynamics.	Model includes data that is relevant to current conditions (including anthropogenic influences).	Model is updated regularly with new data.

References:

Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B. and O'Hara, R.B., 2019. Standards for distribution models in biodiversity assessments. Science Advances, 5, p.eaat4858.

Batley. G.E., 1999. Quality Assurance in Environmental Monitoring. Marine Pollution Bulletin. 39: 23-31.

Boria, R.A., Olson, L.E., Goodman, S.M., and Anderson, R.P. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecol. Model. 275, 73–77.

Fithian, W., Elith, J., Hastie, T., and Keith, D. A. 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods Ecol. Evol. 6, 424–438. doi: 10.1111/2041-210X.12242Howell, K.L., Bullimore, R.D. & Foster, N.L. 2014. Quality assurance in the identification of deep-sea taxa from video and image analysis: response to Henry and Roberts, ICES Journal of Marine Science. 71:899–906. https://doi.org/10.1093/icesjms/fsu052.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology: A Journal of the Royal Meteorological Society, 25: 1965–1978.

ICES. 2020. ICES/NAFO Joint Working Group on Deep-water Ecology (WGDEC). ICES Scientific Reports. 2:62. 188 pp. https://doi.org/10.17895/ices.pub.7503

Jansen J, Hill NA, Dunstan PK, Eléaume MP, Johnson CR. 2018. Taxonomic resolution, functional traits, and the influence of species groupings on mapping Antarctic seafloor biodiversity. Frontiers in Ecology and Evolution. 19:6-81.

Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morisette, J.T., Holcombe, T.R., 2015. Caveats for correlative species distribution modeling. Ecological Informatics, 29: 6-15. https://doi.org/10.1016/j.ecoinf.2015.06.007.

Kenchington, E., Beazley, L., Lirette, C., Murillo, F.J., Guijarro, J., Wareham, V., Gilkinson, K., Koen Alonso, M., Benoît, H., Bourdages, H., Sainte-Marie, B., Treble, M., and Siferd, T., 2016. Delineation of Coral and Sponge Significant Benthic Areas in Eastern Canada Using Kernel Density Analyses and Species Distribution Models. DFO Can. Sci. Advis. Sec. Res. Doc. 2016/093. vi + 178 p. Accessed at: https://www.dfo-mpo.gc.ca/csas-sccs/Publications/ResDocs-DocRech/2016/2016_093-eng.html.

Nephin, J., Gregr, E., St.Germain, C., Fields, C., and Finney, J.L. 2020. Development of a species distribution modelling framework and its application to twelve species on Canada's Pacific coast. Research Document 2020/004. Canadian Science Advisory Secretariat.

Rumohr, H. 2009. Soft-bottom macrofauna: collection, treatment, and quality assurance of samples. ICES Techniques in Marine Environmental Sciences No. 43. 20 pp.

Stephenson, F., Goetz, K., Sharp, B. R., Mouton, T. L., Beets, F. L., Roberts, J., MacDiarmid, A. B., Constantine, R., Lundquist, C.J. 2020. Modelling the spatial distribution of cetaceans in New Zealand waters. Diversity and Distributions, 26: 495-516.

Stoklosa, J., Daly, C., Foster, S. D., Ashcroft, M. B., and Warton, D. I. 2015. A climate of uncer-tainty: Accounting for error in climate variables for species distribution models. Methods in Ecol-ogy and Evolution, 6: 412–423.

Thompson A, Sanders J, Tandstad M, Carocci F, Fuller J. 2016. Vulnerable marine ecosystems: processes and practices in the high seas. FAO Fisheries and Aquaculture Technical Paper. 595.

Turner S. , Manderson J., Richardson D., Hoey J., Hare J. 2016. Using habitat association models to predict Alewife and Blueback Herring marine distributions and overlap with Atlantic Herring and Atlantic Mackerel: can incidental catches be avoided. ICES Journal of Marine Science, 10.1093/icesjms/fsv166.

Winship, A.J., Thorson, J.T., Clarke, M.E., Coleman, H.M., Costa, B., Georgian, S.E., Gillett, D., Grüss, A. Henderson, M.J., Hourigan, T.F., Huff, D.D., Kreidler, N., Pirtle, J.L., Olson, J.V., Poti, M., Rooper, C.N., Sigler, M.F., Viehman, S. and 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.

Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A. and Lahoz-Monfort, J.J., Leitao, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., and Merow, C. 2020. A standard protocol for reporting species distribution models. Ecography, 43:1261-1277. doi: 10.1111/ecog.04960.

3 Approaches to modelling

3.1 General introduction to PHM as applied to VME

Predictive habitat modelling methods can generally be sorted into four main types. Kriging and other geostatistical methods utilize the correlation between observations in space to make inference about the distribution of a species (e.g. Kenchington *et al.,* 2017). Standard statistical methods such as generalized linear models (e.g. Krigsman *et al.,* 2012) and generalized additive models (e.g. Rooper *et al.,* 2017) develop linear or non-linear relationships between dependent and independent data. Machine learning methods, such as random forests (e.g. Beazley *et al.,* 2018) and boosted regression trees (e.g. Rowden *et al.,* 2017) utilize computer algorithms to determine the distribution of a species. More complex statistical methods such as point process models (e.g. Renner *et al.,* 2015, Howell *et al.,* 2016), multivariate mixture models (e.g. Dunston *et al.,* 2011, Dunston *et al.,* 2013, Foster *et al.,* 2013), generalized linear and generalized additive mixed models (e.g. Thorson *et al.,* 2016), and joint species distribution models (Ovaskainan *et al.,* 2017) are more recent developments that can incorporate increased complexity or decreased data quality in PHM. Kenchington *et al.,* (2019) recommend that statistical model-based approaches, which describe the error distribution and explain it with respect to described models should be prioritised when possible. Robinson *et al.,* (2017) who reviewed the use of PHM in the marine environment and Norberg *et al.,* (2019) who compared the predictive performance of commonly used PHM methods recommend fitting multiple models and applying a cross-validation procedure to establish which is most suitable for the goal of the study, before selecting a modelling method. Finally, others have suggested fitting multiple models or model types and using ensemble approach in order to take advantage of the relative strengths of individual models and reduce uncertainty in the predicted distributions of VME indicator taxa (Rooper et al., 2016, Rowden *et al.,* 2017). Ensemble approaches and combining models with differing spatial extents and resolutions to provide ICES advice will be a key area for work in the future in cases where there are multiple existing PHM for species (e.g. Lophelia). The challenges of combining PHM across differing scales and extents include defining how their predictions should be averaged (e.g. weighted or unweighted) or not, the ultimate resolution of the final prediction, or whether multiple models should even be used in providing ICES advice. The approach outlined during this workshop was to determine minimum criteria that could be used to judge individual models, but how information from multiple models should be combined was not directly addressed.

The specific PHM methods currently being used in the fields of ecology and biogeography to predict the distributions of species, biological communities, habitats and habitat types encompass a large variety of modelling approaches. The number of methods is too large for a thorough review in this report, but Norberg *et al.,* (2019) provides a comprehensive list of method comparison studies published between 1995 and 2019. Here we will discuss the main issues to consider when evaluating the appropriateness of a method for modelling the distribution of VMEs in the deep sea.

3.2 Characteristics of different modelling methods

Data requirements. The data requirements (including data type, the number of records, the statistical distribution of data and requirement for balanced data from factor groups) vary between modelling methods. As a rule of thumb, increasing model complexity (higher number of terms and increased flexibility of fit) comes with an increased data requirement. Regression-based models and geostatistical models are the most data hungry, whilst machine learning methods can fit models to much smaller datasets. Modelled

data types range from presence only data to abundance or biomass of entire communities, depending on the method and aims of the study. Table 3.2.1 gives details of the data types accepted by the most commonly used modelling methods.

Assumptions. The statistical assumptions of PHM range from strict expectations of data fit to a specified statistical distribution and parametric responses to predictor variables seen in regression-based methods, to the relative absence of assumptions in data distribution for machine learning methods such as random forest (Cutler *et al.,* 2007), boosted regression trees (Elith and Leathwick, 2008), or maximum entropy (Phillips *et al.,* 2006). In all cases the dependent data are assumed to be independent of one another. Geostatistical methods and Spatial Point Process Models (SPPM) also assume the existence of spatial autocorrelation and structure in the dependent (response) variable. Regression-based methods, including methods such as general linear models, generalized additive models (Wood, 2006), mixture models (Dunstan *et al.,* 2011, Foster *et al.,* 2013) and joint species distribution models (Ovaskainen *et al.,* 2017), which include regression fits, also assume normality of the residuals, error independence (no spatial autocorrelation in the residuals) and in the case of continuous response data no overdispersion. Universal kriging with covariate trends, includes the same assumptions, but autocorrelation in residuals is expected and utilized for fitting in kriging. Where statistical assumptions for regression cannot be met, the machine learning methods are a better choice. Whilst free of strict statistical assumptions, there are still assumptions regarding the input data with the machine learning methods. For example, random forest classification is sensitive to the class ratio, which is an important consideration for the very unbalanced datasets often available for deep-sea VME taxa. This imbalance can be offset by balanced subsampling inside the algorithm. Similarly, maximum entropy (MaxEnt) assumes an equal likelihood of sampling over the environmental gradients (i.e., random or systematic sampling), which is rarely ever the case in deep-sea environments. The unequal likelihood of sampling can be accounted for by targeted selection of background points from the environmental gradients.

Spatial structure. Spatial autocorrelation is present to some extent in all species and habitat distribution data. It is particularly relevant in data collected in the deep sea, which are often very clumped and collated from multiple sources. Whilst some methods such as universal kriging, kernel density estimation (KDE) and SPPM are primarily centred around utilizing the spatial structure in the dependent data, other modelling methods can incorporate it to differing extents. The simplest, but most incomplete way is through inclusion of x and y coordinates or their interaction surfaces in the model, which allows the location to influence predictions, but does not include the distance-based spatial correlation structure addressed by geostatistics. Mixed models, such as general linear mixed models (GLMM) and generalized additive mixed models (GAMM) can also incorporate various formulations of distance-based spatial correlation structure in data as random variable (e.g. Thorson *et al.,* 2016). In fact, the boundary between a universal kriging approach, which includes a complex regression model of covariate trends, and a mixed model incorporating a geostatistical random variable becomes unclear.

Ecological relevance. The interpretability of relationships and ecological relevance of a model and its outputs is important for assessment of the feasibility of the modelled response of the species or habitat on the independent (predictor) variables. The mechanics of model fitting make the response curves of the regression-based models easier to interpret than those of many machine learning methods, such as random forest, where the curve is not fitted to the data, but is the result of predictions from the forest's component trees. Models with curves that are fitted directly to the data are more useful for explaining responses, and also more generalised (when not overfitted), potentially making them more transferable to novel data sets or areas. It is also important to assess the response curves of model relationships and their importance in the context of expectations based on prior knowledge. For example, the absence of a strong explanatory relationship between substrate hardness and abundance of a gorgonian coral might indicate potential problems

with the model or variables that needs to be addressed. Predictor importance can be assessed using a number of methods that can be specific to the PHM used. For example, variable importance for random forest models can be determined by misclassification rates when variables are randomly permutated (Cutler *et al.,* 2007), while for statistical models variable importance can be measured by removing terms sequentially from the best fitting model to determine the relative contribution to overall model fit.

Outputs. The type of output varies between modelling methods and modes of prediction. Whilst regression-based models of presence / absence output a probability estimate within the best fitting model when assumptions are being met, other methods approximate probability in other ways. Random forest models, for example, outputs the proportion of trees predicting the presence, whereas the raw output from MaxEnt and SPPM is a relative occurrence rate (or intensity). MaxEnt can also produce logistic, log-log and complementary log-log (clog-log) transformed outputs, which are an unscaled approximation of a probability of presence. However, it is important to note that probability estimates from different methods are not directly comparable, due to the effect of the data distribution and probability estimation methods. Some models offer the option to output a presence/absence grid, but probability outputs that can be dichotomised afterwards using an appropriate calibrated threshold are preferable (e.g. Liu *et al.,* 2013). Other output types include the most probable factor class and continuous outputs at the scale of the response variable (such as abundance or biomass).

Spatial uncertainty. A spatial representation of uncertainty enables advice and management bodies to assign confidence to decisions made in different areas of the predicted distribution. Some modelling methods intrinsically produce error surfaces, such as standard error or coefficient of variation. With other methods bootstrapping can be used to highlight areas where multiple model runs vary in their predictions, indicating locations where the model is not fitting consistently to combinations of environmental variables. Uncertainty in model outputs is addressed more fully in Section 4 of this report.

Transferability. The transferability of models into new places and times hinges on the ability of models to represent a generalised response of the species or habitat to the environmental conditions. Consequently, geostatistical models such as Kriging and KDE that have been fitted using local spatial data are not transferable in space or time. Although the predictor variables used play a large part, where predictors should have direct effects on the response (e.g., temperature) instead of acting as proxies that are valid only under local conditions (e.g., depth), the generality of the model fit has a big influence. Regression-based models, and other models that rely on a generalised fit to data, are best for predicting outside of the training spatiotemporal envelope. Whilst these types of models can still overfit, tailoring the model fit to the specific conditions encountered in the training data, the fit is easier to control through model settings. MaxEnt can also be parameterised to provide a smoother fit, avoiding overfitting. Random forest and boosted regression trees, through their mechanisms of model fitting are the most prone to overfitting to local conditions.

Table 3.2.1. Characteristics of different model types. P/A = presence / absence of a species, classified assemblage group or habitat type. A Factor variable consists of multiple discrete classes, X and Y are longitude and latitude. A continuous dependent variable can be count, abundance, biomass etc. The relative usefulness ranking for PHM methods is scaled from 1 (useful) to 4 (most useful).

Model type	Data requirements	Assumptions	Treatment of spatial structure in data	Ecological relevance	Type of output	Spatial uncertainty	Transferability	Relative usefulness for VME PHM
Universal kriging (AKA Regression Kriging and Kriging with external drift) (Bivand *et al.,* 2008),	P/A or Continuous dependent variable Independent variables when co-variate trends included Even spatial spread of observations	Spatial autocorrelation Normality (in residuals if co-variate trend model fitted)	Variogram model fitted to represent spatial dependence among (residuals at) points	Variogram depicting spatial relation Response curves for co-variate trend functions	P/A Abundance	Kriging variances / standard errors	Not transferable in space or time	1
Kernel Density Estimation (KDE) (Bivand *et al.,* 2008),	Continuous variable Even spatial spread of observations	Spatial autocorrelation	Weighted density evaluated within defined spatial neighbourhood	Kernal density estimate	Weighted density raster	Not estimated	Not transferable in space or time	2
Generalized linear models and general additive models (GLM/GAM) (McCullagh and Nelder, 1989, Wood 2006)	P/A or continuous dependent variable Independent variables	Normality in residuals Appropriate link function for data distribution Error independence No overdispersion in abundance data	X and Y and/or their interaction as independent variables	Smooth response curves fitted to data	Probability of P/A Continuous on scale of dependent variable	Standard error	Easy to generalise Good for transfer in time or space	4
Generalized linear mixed models and general additive mixed models (GLMM/GAMM) (Wood 2006, Zuur *et al.,* 2009)	P/A or continuous dependent variable Independent variables	Normality in residuals Appropriate link function for data distribution Error independence No overdispersion in abundance data	X and Y and/or their interaction as independent variables Various ways to include spatial random effects	Smooth response curves fitted to data	Probability of P/A Continuous on scale of dependent variable	Standard error	Easy to generalise Not transferable in time or space	3

Model type	Data requirements	Assumptions	Treatment of spatial structure in data	Ecological relevance	Type of output	Spatial uncertainty	Transferability	Relative usefulness for VME PHM
Boosted regression trees (Elith *et al.,* 2008)	P/A or continuous dependent variable Independent variables	None	X and Y as predictor variables	Response curves produced by model prediction – not always interpretable	Probability of P/A or Factor class Continuous on scale of dependent variable	Bootstrap estimates of prediction variability	Transferability questionable	4
Random forest (Cutler *et al.,* 2007)	P/A, Factor or continuous dependent variable Environmental variables	None	X and Y as predictor variables	Response curves produced by model prediction – not always interpretable	Probability of P/A or Factor class – proportion of trees predicting presence Continuous on scale of dependent variable	Bootstrap estimates of prediction variability Proportion of trees (factor classes) Standard error (continuous variables)	Transferability questionable	4
Maximum entropy (Phillips *et al.,* 2006)	Presence only (possibly with user defined background points) Environmental variables	Equal likelihood of sampling over background (random or constant sampling) Constant detectability	No explicit spatial structure	Representative response curves depending on the complexity allowed in the model responses	Raw output is a relative occurrence rate Logistic, log-log or clog-log output approximates presence probability	Bootstrap estimates of prediction variability	Easy to generalise Good for transfer in time or space	3
Multivariate Mixture Models (e.g. species archetype models, regions of common profiles) (Dunston *et al.,* 2011)	P/A or continuous dependent variable Independent variables Usually a community matrix	Parametric species response to their environment	No explicit spatial structure	Plots to choose the number of species archetypes/RCP Archetype/RCP response to the covariate	Predicted probability of each species archetype or RCP Archetype/RCP membership probabilities	Standard error Confidence intervals	Can be transferable in space and time	3

Model type	Data requirements	Assumptions	Treatment of spatial structure in data	Ecological relevance	Type of output	Spatial uncertainty	Transferability	Relative usefulness for VME PHM
Spatial point process models (for presence only data – specifically) (Bivand *et al.,* 2008),	Presence only Independent variables	Different classes of PPM have different assumptions Points are independent The intensity of points varies spatially with the environment	Yes. The object of primary interest in a PPM is the spatial location of the presence points	Influence, leverage and partial residual plots	Intensity of observations Raw output is a relative occurrence rate Logistic, log-log or clog-log ouput approximates presence probability	Depends on software and class of PPM model used	Can be transferable in space and time	3
Joint Species Distribution Models (Ovaskainen *et al.,* 2017)	P/A or continuous dependent variable Usually community matrix Independent variables Can include species traits and phylogenetic data Spatial-temporal data can be included	Parametric species response to their environment	Yes. Spatially structured random effect which can capture species associations irrespective of independent data	Variance partitioning plot Smooth response to covariates Species traits environmental responses Species residual associations	Probability of P/A Species richness Community-weighted mean traits Regions of common profile	Standard error, credible intervals	Transferable in space or time, but not if using random spatial effects	3

3.3 Minimum standards for description and outputs from models used in PHM

This section lays out the required standards and additional desirable best practices for reporting PHM outputs, including associated caveats and assumptions. All PHM are approximations of the real world and therefore require careful interpretation. The proper interpretation of models and assessment of the suitability of model outputs for management advice depends on the user being aware of the particulars of the model development, its assumptions and their implications. Numerous authors have described good practice for PHM over the years. Robinson *et al.* (2017), Araújo *et al.* (2019a,b), Feng *et al.* (2019), Sofaer *et al.* (2019) and Zurell *et al.* (2020) have all suggested comprehensive assessments and checklists for evaluating such predictive models. Here we summarise the recommendations made by these authors adapting them to the specific requirements of VMEs and the deep sea, augmented by the specific suggestions relating to modelling species and habitats in the deep sea by Kenchington *et al.* (2019) and Winship *et al.* (2020).

Table 3.3.1. sets out the minimum requirements for models to be considered suitable for use in ICES advice and Annex 3 lists the information that needs to be reported with any PHM when providing ICES advice. The model requirements pertain to the technicalities of the modelling procedure from aims and objectives to model outputs, whilst input data and model validation requirements are addressed in Sections 2 and 4, respectively.

A clear description of the model aims and objectives, along with the type of variable being predicted is essential for deciding whether or not a model's outputs are suited for a specific purpose. Models built to (1) explain ecological linkages between the environment and species, (2) provide the most accurate prediction of current species distributions, and (3) project distributions outside the spatiotemporal setting of the training data all have a different emphasis on the requirements for model fitting and interpretation. Most PHM to be used in providing ICES advice are likely to be used for predictions of VME distributions. A transparent description of the modelling steps, including input data, software and tools used, provides information that both makes the study replicable and allows for an informed assessment of its applicability for any specific use. In their quantitative review of recent literature Feng *et al.*, (2019) learned that methods sections often lack the details needed for reproducibility. Zurell *et al.*, (2020) found similar deficiencies and have produced an online tool which generates a checklist as an appendix for any modelling study (Overview/Conceptualisation, Data, Model fitting, Assessment and Prediction, ODMAP). The tool produces a model metadata record consistent with most of the reporting requirements set out in Annex 3.

For PHM to be used in ICES advice, the choice of the modelling method should be explained, and be suitable for the modelling objective and input data. In an ideal situation, multiple alternatives would be evaluated before choosing the most appropriate method or ensemble of methods (Sofaer *et al.*, 2019). Statistical assumptions for the selected method need to be recognized and considered. As a minimum, any violations of assumptions, and their consequences for interpretation of the results should be discussed (Araújo *et al.*, 2019b). Ideally, assumptions should be formally evaluated and lack of, or robustness to, violation demonstrated (Araújo *et al.*, 2019; Sofaer *et al.*, 2019). The modelling method should be well described and referenced with transparent reporting of model-specific settings. The most comprehensive way to communicate the applied model settings is to provide both the data and well annotated code scripts used for the analysis (Sofaer *et al.*, 2019). As a minimum, settings which influence the model fit, and hence predictions should be reported and discussed. The rationale used for decisions made on model settings and the associated model complexity (number of parameters, flexibility of the modelling approach, and the number of operations involved) should be made clear. Over-complex models

risk overfitting the data and reducing its predictive power. Araújo *et al.* (2019a) define their high standard (gold-level) for models as including a full exploration of the consequences of all choices in model building through result comparison via cross-validation or the use of independent data. However, these choices are most commonly decided using broadly agreed rules of thumb.

Model results need to be reported in enough detail to convey the goodness and ecological relevance of the model fit. Model method-specific term estimates or coefficients and goodness-of-fit statistics should be reported, where relevant (see Section 4). Variable importance, the relative contribution of each predictor variable to model fit, indicates which of the independent data layers are the most influential in determining the prediction output. Both the relative contributions of predictors, and the shape of response curves (for methods where possible) should be checked for plausibility based on expert judgement and compared to existing knowledge. Spatial autocorrelation in model residuals indicates that the model has been unable to adequately represent the distribution of the response variable, either due to missing environmental variables or the predominance of biological interactions. It is not always possible to remove all spatial autocorrelation, but it should be investigated, reported and the implications of any remaining autocorrelation on interpretation of results discussed. Map outputs should also be checked for plausibility by comparison with known distributions and even expert opinions where experts can be identified.

Understanding of the model output hinges on the prediction unit being clearly defined and explained. Predictions may be dichotomous presence/absence of a species or habitat type, a combination of factor classes (such as multiple VME types), a probability of presence (0-1) of a species or habitat type, or a continuous variable such as density or biomass. Especially when the output is continuous or a probability, the meaning of the values and how this affects interpretation must be clearly explained. Probability outputs are driven by the prevalence of the input data (zero inflation in data leads to very low probabilities even at locations where taxa are observed) and are often not intuitively interpretable. The raw model predictions are sometimes transformed via different methods under different assumptions before mapping and such post-processing steps need to be clearly outlined (Feng *et al.*, 2019). Threshold selection for dichotomising a continuous likelihood of presence output to presence / absence must be clearly described and appropriate for intended use. Ideally thresholds should be selected via model assessment with exploration of sensitivity (Sofaer *et al.*, 2019). Liu *et al.* (2013) outlines the most commonly used thresholding methods and discusses their implications to the emphasis placed on false positives and negatives.

Finally, we must consider model transferability in geographical space and time. Transferring a model across space and/or time leads to potential over-extrapolation, where the projected environments are outside of the environmental envelope of the training data. In the deep sea, where field observations are often sparse and clumped, with an uneven coverage of environmental gradients, some level of extrapolation may also be required in local models. In both cases, the environmental conditions in the location (or time) where predictions are made must be within a plausible range for extrapolation of response – predictor relationships. Model extrapolation is statistically challenging and different extrapolation strategies can lead to very different model predictions. Therefore, the choice of extrapolation method should be clearly described (Feng *et al.*, 2019). Furthermore, a spatial representation of novel environments would help to communicate the uncertainties associated with predictions outside of training data range (e.g. Stephenson *et al.*, 2020). For model results to remain current over time, environmental conditions and potential human impacts must remain relatively stable. When using older model outputs, consideration must be made for the vintage of both input data and model. In an ideal situation a model would be updated via targeted field sampling, incorporated into subsequent model iterations (Sofaer *et al.*, 2019).

The main recommendations and criteria to be applied in use of PHM for ICES advice, require the entire modelling process to be transparent, and include sufficiently detailed explanations and justifications using plain language for all choices, along with any necessary material (e.g. selection criteria, code) to make it reproducible.

The selection of the most suitable modelling technique for PHM will depend on the characteristics of the independent (e.g. collinearity) and dependent data (e.g. sampling bias) and the management needs or questions. The selection needs to be fully, accurately and clearly explained, including known advantages and disadvantages of the modelling technique. If more than one modelling technique is suitable, either multiple models should be explored or the prioritisation of one particular technique explained and justified. If an ensemble modeling approach is used, each model within the ensemble should meet the minimum criteria and the method of ensembling be fully explained and justified. Ideally model assumptions (e.g. stationarity, homocedasticity, normality, constant error; Dormann *et al*, 2012) should not be violated; or the robustness of the model predictions despite any violation should be addressed. The same criteria apply to the presence of collinearity among the predictor variables included in the models and the potential bias in any of the data (dependent or independent). Realiability and performace of the model need to be reported based on multiple metrics and criteria, which are described in detail in the following section.

Table 3.3.1. Description of components of PHM to consider, evaluate and present when describing a model in support of scientific or management advice. Standards are also provided based on previous literature.

	UNACCEPTABLE	REQUIRED	DESIRED
Objective	No objectives stated.	Model objective (to explain, predict or project) is stated.	Model objective (to explain, predict or project) is stated and hypotheses for model linkages are clearly stated.
Modelling method	*A priori* model selection is made, without taking into account the available data characteristics.	Transparent and clear communication of pros and cons of each model type is provided to inform selection of best possible model.	Same as required criteria.
	No rationale given for choice of modelling method.	Selection of the modelling method is appropriate for the study objective and available data types.	Selection of the modelling method is aligned with objectives and appropriate for the input data type. Multiple models have been evaluated at the model development stage.
	Method is not described.	Method is thoroughly described and/or referenced.	Same as required criteria.
	Assumptions not described.	Assumptions are recognized and their effects on model results are considered and explained.	Assumptions formally evaluated in statistical analysis.
	Violation of assumptions is not reported.	Robustness of model outcomes to violation of assumptions is demonstrated.	Absence of violation of assumptions is demonstrated.
Model settings	Description of settings used in model is incomplete.	Model specific settings are reported.	Model specific settings, including default ones, should be reported. Model sensitivity to the choice of settings assessed and reported.

	UNACCEPTABLE	REQUIRED	DESIRED
Model terms/coefficients	No information or explanation provided on model terms.	Method of extracting relevant method-specific term estimates or coefficients and how they were evaluated is reported.	Same as required criteria
	Model complexity has not been considered or justified.	Model complexity has been decided/optimised using justified methods or broadly agreed rules of thumb.	Model complexity has been optimised through comparison of multiple models and cross-validation.
	Model outputs have not been evaluated, or model output is not considered plausible.	Model outputs have been evaluated and match common understanding of the response taxon's ecology or habitat requirements and the expected distribution.	Model outputs have been evaluated and compared with independent data or established references.
	The relative contribution of predictor variables has not been considered.	Variable importance and how it was determined is reported.	Same as required criteria
Model fit	Goodness-of-fit not considered.	Goodness of fit statistics, and where appropriate residuals, have been checked and their implications to model interpretation are reported.	Goodness of fit statistics, and where appropriate residuals, have been checked and their implications to model interpretation are reported. data and code are also provided.
	Model performance is not reported.	Multiple measures of model performance are reported.	Same as required criteria

References

Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., et al. 2019. Supplementary Materials for Standards for distribution models in biodiversity assessments. Sci. Adv. 5.

Beazley L, Wang Z, Kenchington E, Yashayaev I, Rapp HT, et al. 2018. Predicted distribution of the glass sponge *Vazella pourtalesi* on the Scotian Shelf and its persistence in the face of climatic variability. PLOS ONE 13(10): e0205505. *https://doi.org/10.1371/journal.pone.0205505*.

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. 2008. Applied Spatial Data Analysis with R. Springer Science+Business Media, LLC, New York, New York. 378 p.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. 2007. Random forests for classification in ecology. Ecology, 88(11):2783–2792. https://doi.org/10.1890/07-0539.1.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop.). 36. doi:10.1111/j.1600-0587.2012.07348.x.

Dunstan, P.K., Foster, S.D., and Darnell, R. 2011. Model based grouping of species across envi-ronmental gradients. Ecological Modelling, 222: 955–963.

Dunstan, P.K., Foster, S.D., Hui, F.K., and Warton, D.I. 2013. Finite mixture of regression mod-eling for high-dimensional count and biomass data in ecology. Journal of agricultural, biolog-ical, and environmental statistics, 18: 357–375.

Elith, J., Leathwick, J. R., & Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology, 77(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Feng, X., Park, D. S., Walker, C., Peterson, A. T., Merow, C., and Papeş, M. 2019. A checklist for maximizing reproducibility of ecological niche models. Nat. Ecol. Evol. 3. doi:10.1038/s41559-019-0972-5.

Foster, S., Givens, G., Dornan, G., Dunstan, P., and Darnell, R. 2013. Modelling biological re-gions from multi-species and environmental data. Environmetrics, 24: 489–499.

Howell, K. L., Piechaud, N., Downie, A. L., & Kenny, A. 2016. The distribution of deep-sea sponge aggre-gations in the North Atlantic and implications for their effective spatial management. Deep-Sea Re-search Part I: Oceanographic Research Papers, 115, 309–320. https://doi.org/10.1016/j.dsr.2016.07.005.

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., et al. 2019. Use of Species Distribution Modeling in the Deep Sea.

Kenchington, E., Murillo, F.J., Lirette, C., Sacau, M., Koen-Alonso, M., Kenny, A., Ollerhead, N., Wareham, V. and Beazley, L. 2014. Kernel density surface modelling as a means to identify significant concentra-tions of vulnerable marine ecosystem indicators. PLoS ONE 9(10): e109365. doi:10.1371/jour-nal.pone.0109365.

Krigsman, L. M., Yoklavich, M. M., Dick, E. J., & Cochrane, G. R. 2012. Models and maps: predicting the distribution of corals and other benthic macro-invertebrates in shelf habitats. Ecosphere, 3(1), art3. https://doi.org/10.1890/es11-00295.1.

Liu, C., White, M., and Newell, G. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. J. Biogeogr. 40, 778–789. doi:10.1111/jbi.12058.

McCullagh, P., and Nelder, J., 1989. Generalized Linear Models, 2nd ed. Chapman & Hall, New York.

Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community lev-els. Ecol. Monogr. 89. doi:10.1002/ecm.1370.

Ovaskainen, O., Tikhonov, A.G., Norberg, F.G., Blanchet, L., Duan, D.B., Dunson, T.R., and Abrego, N. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters, 2: 561–576.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026.

Renner I.W., Elith J., Baddeley A., Fithian W., Hastie T., Phillips S.J., Popovic G., & Warton D.I. 2015 Point process models for presence-only analysis. Methods in Ecology and Evolution, 6, 366–379.

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E., and Lundquist, C. J. 2017. A systematic review of marine-based Species Distribution Models (SDMs) with recommendations for best practice. Front. Mar. Sci. 4. doi:10.3389/fmars.2017.00421.

Rooper, C. N., Sigler, M. F., Goddard, P., Malecha, P., Towler, R., Williams, K., and Wilborn, R. *et al.* 2016. Validation and improvement of species distribution models for structure-forming in-vertebrates in the eastern Bering Sea with an independent survey. Marine Ecology Progress Series, 551: 117–130.

Rowden, A. A., Anderson, O. F., Georgian, S. E., Bowden, D. A., Clark, M. R., Pallentin, A., & Miller, A. 2017. High-resolution habitat suitability models for the conservation and man-agement of vulnerable marine ecosystems on the Louisville Seamount Chain, South Pacific Ocean. Frontiers in Marine Science, 4(OCT). https://doi.org/10.3389/fmars.2017.00335

Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., et al. (2019). De-velopment and Delivery of Species Distribution Models to Inform Decision-Making. Bioscience 69. doi:10.1093/biosci/biz045.

Stephenson, F., Goetz, K., Sharp, B. R., Mouton, T. L., Beets, F. L., Roberts, J., MacDiarmid, A. B., et al. 2020. Modelling the spatial distribution of cetaceans in New Zealand waters. Diversity and Distributions, 26: 495-516.

Thorson, J.T., Pinsky, M.L., Ward, E.J., 2016. Model-based inference for estimating shifts in species distri-bution, area occupied, and center of gravity. Methods Ecol. Evol. 7(8), 990-1008. doi:10.1111/2041-210X.12567.

Winship, A. J., Thorson, J. T., Clarke, M. E., Coleman, H. M., Costa, B., Georgian, S. E., *et al.* 2020. Good Practices for Species Distribution Modeling of Deep-Sea Corals and Sponges for Resource Manage-ment: Data Collection, Analysis, Validation, and Communication. Front. Mar. Sci. 7. doi:10.3389/fmars.2020.00303.

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC. 398 p.

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., *et al.* 2020. A standard protocol for reporting species distribution models. Ecography (Cop.). 43. doi:10.1111/ecog.04960.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev A. A., and Smith, G. M. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer Science+Business Media, LLC. New York, New York. 579 p.

4 Model validation and uncertainty

4.1 Approaches used for model validation

Model validation with an external independent dataset is considered the most robust type of evaluation (Araújo *et al.*, 2005; Winship *et al.*, 2020). Ideally, the new data should be collected after fitting the model, using the model output to inform the sampling design of the validation survey (Newbold *et al.*, 2010; Williams and Brown, 2019). Simulations can contribute to the selection of the best sampling design (Hirzel and Guisan, 2002). With consideration of the assumptions of data from differing sources, sampling methods, spatial scales, etc., it is also possible to use other independent datasets (e.g. using fisheries bycatch data to validate a model based on ROV observations). There are few examples of validations of models of VMEs using fully independent data (Anderson *et al.*, 2016a; Rooper *et al.*, 2016, Rooper *et al.*, 2018). In most cases this is not feasible, especially when VMEs are located in the deep sea (> 50 m depth), a rather remote and challenging environment and therefore expensive in terms of time and budget to survey (Serrano *et al.*, 2017).

More often, model validation is carried out through internal evaluation by resampling. In this approach, known as cross-validation, observations are divided into two groups, for fitting (training set), and evaluation (test set). Partitions can be done one single time (e.g. k-fold, leave-one-out cross-validation), or multiple times (e.g. bootstrap). The model is fitted using the training set, and predictions are carried out for the test set. The predictive performance of the model is evaluated by comparing the predicted values to the test set itself using different statistical metrics. Cross-validation produces two useful outputs: a) cross-validated or bootstrapped predictions for all observations, and b) estimates of variation in model parameters that can be used to evaluate model stability (Arlot *et al.*, 2010; Guisan *et al.*, 2017). Broadly, there are three main approaches for cross-validation in PHM:

K-fold cross-validation. A common approach to internal model validation is k-fold cross-validation. In this method the data is divided into k groups or folds (usually k = 5-10) of equal size. The model is fitted k times, using in turn k-1 folds as training data and the remaining fold as test data. K-fold validation has been used in many PHM (e.g. Edwards Jr *et al.*, 2006; Dormann *et al.*, 2008). Usually the folds are selected at random, although as with all cross-validation methods, it is possible that partitions are selected in geographic space (Osborne and Suárez-Seoane, 2002). K-fold has the limitation that the number of replicate models is low, and that for datasets with low prevalence the number of observations in a particular model could be too low (Guisan *et al.*, 2017). Leave-one-out cross-validation where k is equal to the number of observations (Hastie *et al.*, 2009) is a similar method in which each individual observation is used in turn as a test set. Leave-one-out cross-validation is useful when the number of observations is very low (Guisan *et al.*, 2017).

Repeated split sample cross-validation. Also known as a validation set approach, repeated split-sample cross-validation is a method in which the observation dataset is split randomly in two partitions, typically 70% for model fitting, and 30% for model testing. This splitting is done multiple times, usually many more than in the case of k-fold cross-validation. The split cross-validation approach can be more informative that the k-fold validation because it generates multiple estimates of model parameters and evaluation metrics from multiple data splits (Guisan *et al.*, 2017). This can reduce the variance of the estimates due to the higher sample size and be used to more fully evaluate model stability.

Bootstrapping. Bootstrapping is a method in which the original dataset is sampled multiple times with replacement, with each sample having the same size as the original data. The bootstrap builds an empirical distribution function of the model parameters which provides an estimate of the variance and the bias in model parameters (Guisan *et al.*, 2017). However, it is difficult to accurately estimate prediction error using the bootstrapping method, as bootstrapped training and testing data sets can share observations resulting in positively biased performance for the model predictions. To alleviate this problem a .632+ bootstrap (Efron and Tibshirani, 1997) can be used. The .632+ bootstrap mimics cross-validation by tracking predictions for observations only in the subset of bootstrap samples that do not contain that observation. The method gets is name because on average 63.2% of the original data is selected in each subset, so the remaining 36.8% can be used for testing.. The bootstrap (and in particular the .632+ bootstrap) can be used as an alternative to k-fold cross-validation for producing less variable estimates of model parameters with potentially smaller and less biased prediction errors (Efron and Tibshirani, 1997).

In spite of cross-validation being a popular and widely accepted technique, there are ongoing discussions about the best method for assessing the performance of PHM. In recent years the best way to allocate data to training and testing datasets and the potential consequences of each election in the validation results has been discussed in the literature (Valavi *et al.*, 2018) and important concerns about the way how PHM models are evaluated have emerged (Fourcade *et al.*, 2018). One of the most important problems highlighted by Fourcade *et al.* (2018) is the lack of independence between training and test datasets, especially when these points are generated via conventional random cross-validation. A general strategy applied in PHM literature to avoid this problem (which especially affects the capacity of the approach to evaluate model transferability) is to use some type of spatial approach when the data are split between training and testing datasets (Muscarella *et al.*, 2014; Fourcade *et al.*, 2018, Valavi et al., 2018). Recently at least two R packages ENMEval (Muscarella *et al.*, 2014) and block cross-validation (Valavi *et al.*, 2018) have been released that provide a wide set of different techniques to spatially (and not only randomly) divide the data. Some of these strategies include spatial blocks, checkboard strategies, environmental profiling, or buffering and we strongly advise exploring at least one of them for a more appropriate cross-validation evaluation of the models (e.g. Anderson *et al.,* 2016).

4.2 Metrics used to evaluate models and model validation

Evaluation metrics. A number of different evaluation metrics can be used to assess the performance of species distribution models. Standard goodness-of-fit methods common to all models, such as deviance explained or correlation metrics are always reported. However, PHM that predict the presence or absence of a taxa also use a set of specialized metrics to evaluate their performance. In general, these can be threshold independent metrics or threshold dependent metrics which are derived from a confusion matrix (e.g. Table 4.2.1).

Table 4.2.1. Confusion matrix showing the two type of possible errors in a presence/absence model; False positives (b) and false negatives (c).

	Observed presence	Observed absence
Predicted presence	a	b
Predicted absence	c	d

The confusion matrix is a very simple table, obtained after converting continuous predictions of PHM into binary predictions of presence and absence using a threshold probability and is the

base for most of the metrics used in PHM literature (Fielding and Bell, 1997). Some of the most popular metrics in this field such as sensitivity, specificity, true skill statistic (TSS) or Kappa are just combinations of the 4 potential outputs of the confusion matrix. (see formulae in Table 4.2.2, Fielding and Bell, 1997, Allouche *et al.*, 2006). The main limitation of evaluation metrics that use the confusion matrix is that they are calculated using a threshold (hence threshold dependent), and as such are sensitive to the choice of the threshold or method used to choose the threshold. To overcome this limitation, most studies use the Area Under the Curve Receiver Operating Characteristic (ROC) plots which is considered standard practice in the PHM literature (Lobo *et al.*, 2008; Fourcade *et al.*, 2018). The main advantage of this metric (usually called just AUC) is that it is threshold independent, since it is obtained by using a plot which is generated by using all possible thresholds to compute sensitivity (y axis) and 1-specificity (x axis). Using AUC, the performance of a model can also be classified into easily understood categories that describe its usefulness in predicting distributions (AUC ≤ 0.5 is no better than random performance, ≥ 0.7 is adequate, ≥ 0.8 is excellent; Hosmer et al., 2013). Unfortunately, AUC also has important limitations that are not always sufficiently incorporated in the model evaluation, but which limit its capacity to accurately evaluate the model (Lobo *et al.*, 2008; Jiménez-Valverde, 2012; Fourcade *et al.*, 2018). These limitations are especially important when using pseudo-absence or background points because the weight of commission errors is much lower than that of omission errors or in the case of rare species where low prevalence is difficult to disentangle from model performance (Jiménez-Valverde, 2012). In spite of these well-known limitations, AUC is still the most common metric used in the PHM literature, probably because of the lack of an adequate alternative. The other metrics also have important limitations besides being threshold dependant. For instance, kappa is prevalence dependant which can seriously affect its performance under certain circumstances (Allouche *et al.*, 2006). Fourcade *et al.* (2018) highlight the need to select biologically relevant independent data, as models built without ecologically relevant environmental data (e.g., using representations of paintings) can produce models that perform well according to evaluation metrics. Because of these limitations, we advise to always use more than one metric to correctly evaluate the performance of PHM, while always providing values of sensitivity and specificity to offer information about the discrimination capacity of the model. Finally, in addition to test metrics results, other evaluation process such as the plausibility of the response curves and their consistency with the ecological knowledge of the species are equally important (see Section 3) and must be included in the evaluation process, especially in "presence-only" models (Warren *et al.*, 2020).

Table 4.2.2. Some of the most popular evaluation metrics obtained from the confusion matrix. Adapted from: Fielding and Bell (1997) and Allouche *et al.* (2006). The values of a,b,c and d are extracted from the confusion matrix. N is the number of data (a+b+c+d)

MEASURE	CALCULATION
Prevalence	(a+c)/N
Overall diagnostic power	(b+d)/N
Correct classification rate	(a+d)/N
Sensitivity	a/(a+c)
Specificity	d/(b+d)
False positive rate	b/(b+d
False negative rate	c/(a+c)

MEASURE	CALCULATION
Kappa	[(a + d) - (((a + c)(a + b) + (b + d) (c + d))/N)]/[N - (((a + c)(a + b) + (b + d)(c + d))/N)]
True Skill Statistic (TSS)	Sensitivity+Specificity-1

Any model used to support ICES advice should be reliable, and robust in demonstrating its utility and capabilities to predict the occurrence of the defined VME/indicator in unsampled and/or new locations and time periods. Table 4.2.3 outlines the minimum and desired criteria for model validation of PHM to be used to provide ICES advice. Models should ideally be validated against representative and statistically independent data of the defined VME/indicator. Such statistical independence implies that, at least, the validation data were not used to build the model. Ideally, datasets from different spatial, temporal and/or survey collections are desired for validation. A highly reliable model would be validated using data specifically collected in areas identified by the model as having high and low likelihoods of VME/indicator to be present (e.g. Anderson *et al.,* 2016, Rooper *et al.,* 2018). For this purpose, we suggest collaborations with projects that have dedicated ship time for exploration in the deep sea (e.g. EU Horizon projects) to incorporate such sampling and data collection in their plans for new surveys and expeditions.

Table 4.2.3. Criteria for model validation requirements for PHM to be used when supporting ICES advice.

UNACCEPTABLE	REQUIRED	DESIRED
No validation method is provided.	Internal cross-validation (subset of data, not used to build the model).	Validation with truly independent (new collected) data.
No consideration has been given to the assumptions in data structure when splitting data into training and test sets.	Rationale and method of splitting data into training and test datasets is described, and any caveats explained.	Rationale and method of splitting data into training and test datasets is described, and any caveats explained. Splitting rationale includes a spatial and temporal (if applicable) component.

4.3 Description and characteristics of model uncertainty

In ecological models, the term uncertainty refers to the unexplained statistical variation in the model output (Barry and Elith, 2006; Gould *et al.,* 2014). The uncertainty in the output of PHM is not necessarily a problem for their use in management, as long as its effects on model predictions are not ignored. When model uncertainty is assessed, managers can make decisions that are robust to model uncertainty or take into account model uncertainty (Gould *et al.,* 2014). For example, Rowden *et al.* (2019) proposed VME closures giving higher conservation priority to areas with low prediction uncertainty. On the other hand, when model uncertainty is not acknowledged, the model output can potentially be interpreted with unjustifiably high confidence which can be misleading (Beale and Lennon, 2012; Wenger *et al.,* 2013), resulting in management decisions that may have counterproductive effects (Robinson *et al.,* 2017). Estimating prediction uncertainty is essential to effectively inform management decisions (Jones-Farrand *et al.,* 2011, Jones and Cheung, 2015). Unfortunately, many published PHM do not explicitly address uncertainty (Peterson *et al.,* 2018; Araújo *et al.,* 2019). In a recent review, Robinson *et al.* (2017) reported that 94% of published PHM in marine environments failed to report the amount of uncertainty derived from data deficiency and inappropriate model parameterization. The lack of effective

communication of model uncertainty is considered as one of the main barriers against the more frequent use of PHM output in environmental decision making (Rapacciuolo, 2019). Therefore, it is necessary that PHMs used for management purposes should always explain how uncertainties in the model predictions were addressed. In general, the three approaches for addressing uncertainty in PHM include a) reducing model uncertainty by increasing the ecological knowledge and improving the modelling process, b) assessing model uncertainty quantitatively or probabilistically, and c) applying management measures that are robust to model uncertainty (Gould *et al.*, 2014). Models where uncertainty was not addressed should be interpreted with caution (Sofaer *et al.*, 2019).

Major sources and types of model uncertainty

Uncertainty in PHM arise from different sources (Barry and Elith, 2006; Dormann *et al.*, 2008; Link *et al.*, 2012). One of the sources of uncertainty is natural variability, an inherent property of any natural system. PHM are attempts to summarise complex distributional patterns using the relationships between the distribution and a reduced set of environmental covariates (independent variables). Therefore, PHM inevitably will include some degree of uncertainty, because not all factors that determine the distribution of a species or habitat can be included in a model. Uncertainty arises also from the incorrect or incomplete specification of the ecological model, which refers to the relationship between the distribution patterns and independent variables. Some examples are the incomplete knowledge of the species' habitat requirements (Pulliam, 2000; Kearney, 2006), spatial or temporal mismatch between the input data and the ecology of the species (Dormann *et al.*, 2007), and spatial or temporal variations in the ecology of the modelled species of habitats.

Data related uncertainty. Uncertainty is also introduced by the limitations of the dependent and independent data used to fit the model (Barry and Elith, 2006; Dormann *et al.*, 2008). Some of these include:

- Positional errors associated to presence records (Moudr and Šímová, 2012), including variability arising from sampling methods operating at different spatial resolutions (e.g. bottom trawl *vs*. ROV sampling).

- Small sample sizes, in particular when using presence-only data, and low prevalence (number of locations where the species of interest is present).

- The use of presence-only data, which can lead to inaccurate identification of unsuitable sites (Barry and Elith, 2006) (Phillips *et al.*, 2009).

- Sampling bias in observation records (Fourcade *et al.*, 2014; Merow *et al.*, 2016). Most modelling approaches assume that the distribution of observations are representative of the spatial and environmental distributions of the species being modelled (Araújo *et al.*, 2019). In most cases though, the distribution of observations is also strongly influenced by sampling effort and by imperfect detectability.

- Uncertainty in the gridded data used for deriving environmental layers (Elith *et al.*, 2002; McInerny and Purves, 2011). PHM require predictor environmental variables with full-coverage, which are obtained by interpolating discrete observations (e.g World Ocean Atlas, Locarnini *et al.*, 2013), from physical models (e.g. Dolan *et al.*, 2021), or derived from remote sensing observations (Lutz *et al.*, 2002). In all cases full-coverage environmental variables are predicted with uncertainty, although available data products not always

include estimates of this uncertainty. Even when available, this uncertainty is usually not accounted for when they are used in PHM (Hijmans *et al.*, 2005; Stoklosa *et al.*, 2015).

- Collinearity in the set of predictor variables (Beale *et al.*, 2010).

Topics related to uncertainty in the dependent and independent data are further explored in more detail in Section 2. At a minimum, PHM used for management purposes should characterise main sources of uncertainty and explore their potential effect in the interpretation of model results, considering that not all sources of uncertainty have the same effect in management decisions (Rapacciuolo, 2019).

Spatial autocorrelation in residuals. Spatial autocorrelation occurs when the value of variables sampled at nearby locations are not independent from each other (Dormann *et al.*, 2007). In PHM, spatial autocorrelation in the residuals can occur when biological processes (e.g. predation, larval dispersal) are distance-related and not included in the model, when non-linear relationships between environmental variables and the species are modelled erroneously, predictors are poorly scaled to the observation, and when important covariates are missing from the model (Dormann *et al.*, 2007; Elith and Leathwick, 2009). In these situations, the models do not fully explain the species' spatial distribution, resulting in model residuals showing spatial autocorrelation, which can affect statistical inference because it violates the assumption of independence among the data points (Legendre, 1993; Roberts *et al.*, 2017). PHM that do not account for spatial dependence in occurrence data, can potentially lead to misidentification of important predictors and overly optimistic error rates (Dormann *et al.*, 2007; Record *et al.*, 2013).

Some modelling frameworks used for PHM allow for autocorrelated residuals. For example, Generalized Linear Mixed Models (GLMMs) and Generalized Additive Mixed Models (GAMMs) can include spatially correlated errors with different structures (Wood, 2006). Recent attempts to incorporate spatial structure in machine learning models include the random forest regression kriging (Hengl *et al.*, 2015), and the random forest Spatial Interpolation (Sekulić *et al.*, 2020) approaches.

In some cases, the autocorrelation in residuals can be minimized or removed by incorporating "space" as a predictor to model spatial patterns not explained by the environmental predictors (Dormann *et al.*, 2007; Merow *et al.*, 2014). For example, with Generalized Additive Models (GAMs) space can be modelled with a bivariate spline as a function of coordinates (Wood and Augustin, 2002). More generally, the residuals of a preliminary model can be used to obtain a residual auto-covariate (RAC) variable that can be used as a predictor in subsequent models (Hughes *et al.*, 2011; Bardos *et al.*, 2015; Georgian *et al.*, 2019). Approaches like these should remove the problem of independence and aide model interpretation, although in practice specification errors can lead to poor model performance (Roberts *et al.*, 2017). In addition, these methods can cause structural overfitting, when the model explains the spatial structure of the observations with non-causal predictors which themselves have spatial autocorrelation (Roberts *et al.*, 2017).

It is necessary to use robust non-parametric methods for model validation, selection and assessment when using observation data with spatial autocorrelation. As a minimum, we recommend that model residuals are displayed as maps (e.g. Nephin *et al.*, 2020). A map of residuals can be used to identify areas where the model is over or under predicting (Rooper *et al.*, 2018; Nephin *et al.*, 2020). Spatial autocorrelation should also be considered when selecting cross-validation strategies. In particular, spatial blocking strategies where geostatisical approaches or relationships are utilized in choosing training and testing data sets (Muscarella *et al.*, 2014; Roberts *et al.*, 2017; Valavi *et al.*, 2018) are recommended even when no spatial patterns are apparent in the

model residuals. Spatial blocks can account for the spatial dependency of the data and provide more realistic model performance metrics (see Section 4.1).

Extrapolation of model predictions The extent of the area where PHM models are trained have a strong influence on the model result. Ideally, these areas should include current and historical distributions, and be based on the organisms movement patterns and ability and relevant environmental variables (Araújo *et al.*, 2019). Some degree of model transfer is necessary when characterising the full geographic distribution of VMEs in the deep sea, because observations are not available in all areas. Furthermore, for some applications it is necessary to extrapolate PHM to new geographical areas or temporal ranges (i.e. model transfer) than where the data used to fit the model originated. An example could be anticipating the potential effect of climate change on the species' distribution (Owens *et al.*, 2013).

In general, a well-behaved PHM fitted using spatial block cross-validation methods can be used for interpolation, that is for predicting distributions in the region of the training data (Nephin *et al.*, 2020). On the other hand, areas where the model is extrapolating need to be evaluated with care, as the performance of the model in those areas is unknown (Elith and Graham, 2009; Fitzpatrick and Hargrove, 2009).

We recommend that model extrapolation is carefully considered and limited in predictor and geographic space (Austin, 2007; Merow *et al.*, 2014). In all cases, areas where the model is extrapolating outside the geographic or environmental space of the occurrence data should be highlighted. Model extrapolation can occur in geographical and environmental space. In geographical space, convex hulls or spatial buffers can be used to delineate the geographical area around the observations. In environmental space, it is possible to identify areas where prediction occurs outside the range of influential predictor variables (e.g. variables accounting for 95% of the cumulative relative influence of the model; Nephin *et al.*, 2020). Tools like the Multivariate Environmental Similarity Surface (MESS) (Elith *et al.*, 2010), or the Mobility-Oriented Parity (Owens *et al.*, 2013) can also be used to highlight areas of environmental extrapolation (e.g. Stephenson *et al*, 2020). It is also necessary to verify that the model predictions are realistic and that they comply with ecological theory and prior knowledge of the species (Guisan and Thuiller, 2005; Austin, 2007).

Uncertainty mapping. Assessing variability and presenting visualizations of uncertainty is not a common practice (Peterson *et al.*, 2018). Nevertheless, given that one of the main uses of PHM in the management of deep-sea habitats is to aid the delineation of spatial closures and other forms of spatial management, we consider that model output should include maps showing the distribution of uncertainty in the model predictions (Anderson *et al.*, 2016b). Communicating the model uncertainty in a spatially explicit way can increase awareness of the potential impacts of uncertainty and reduce the risks that model outputs are misinterpreted (Elith *et al.*, 2002; Wiens *et al.*, 2009; Rocchini *et al.*, 2011).

Models fitted using statistical approaches like GLMs and GAMs, generate spatial predictions with standard errors, based on the posterior distribution of the model coefficients (Wood and Augustin, 2002). These can be used to construct confidence intervals that can be used to map the uncertainty of the model output.

Uncertainty maps can also be obtained from the cross-validation runs produced during the model validation (Nephin *et al.*, 2020). However, a more robust approach is the use of non-parametric bootstrapping. Bootstrapping provides a set of replicate model predictions from which different statistics can be calculated (mean, SD). Bootstrap is a more robust alternative, because it provides a potentially higher number of alternative maps and therefore a better representation of the spatial uncertainty of the models. Two recent examples of PHM using non-parametric bootstrapping methods to estimate spatial uncertainty for VME indicator models are Anderson

et al. (2016b) and Georgian *et al.* (2019). In this approach a high (>100) number of models is fitted after sampling the presence/absence or presence/background points with replacement. The model uncertainty is estimated by the coefficient of variation calculated in each cell of the prediction raster. The coefficient of variation can be used to compare relative uncertainty in the predictions from different models (Rooper *et al.*, 2017; Georgian *et al.*, 2019), while the standard deviation has been proposed to highlight areas with high suitability and high uncertainty (Nephin *et al.*, 2020).

Table 4.4.1 outlines the required and desired reporting criteria for PHM to be used in providing ICES advice. All sources of uncertainty need to be identified, characterized and quantified at every step of the modelling process, including uncertainty derived from the input data (dependent or independent) or the selection of model algorithm/s and the parameterization (limitations) of such model/s. These uncertainties must be spatially quantified and mapped, acknowledging whether the data are not equally distributed or accurate across the spatial domain (Araújo *et al.*, 2019). Explanatory text should be included to clearly define the metrics or units used to quantify the uncertainty and to help interpret the possible effects on the predictive maps of the VME indicator taxa, including reliability of the predictions and ecological meaning or plausibility. When multiple models are used, their uncertainties should be reported using a directly comparable metric when possible. Ideally, propagation error techniques should be used to produce final predictive maps of the defined VME/indicator that already account for and reflect all the different sources of uncertainty.

Table 4.4.1. Reporting criteria for model uncertainty for PHM to be used for providing ICES advice.

UNACCEPTABLE	REQUIRED	DESIRED
Uncertainty is not reported.	Major sources of uncertainties in the model and data are characterized, quantified, accounted for, and discussed. If multiple models are used or ensembled, uncertainties are assessed and compared among models.	Uncertainty, including uncertainty in the environmental predictors, is fully propagated through the modelling process and quantified, mapped, interpreted and discussed in the results. If multiple models are used or ensembled, all uncertainties are assessed and compared among models.
No consideration for spatial autocorrelation in residuals.	Spatial autocorrelation in residuals has been assessed and, if needed, statistical inference adjusted accordingly.	No spatial autocorrelation present in residuals.
Bias is not considered, only acknowledged or remain uncorrected.	Correction for major bias is made, testing performance by cross-validation.	Absence of bias is demonstrated or bias is fully corrected.

References

Allouche, O., Tsoar, A., and Kadmon, R. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (tss). Journal of applied ecology, 43: 1223–Wiley Online Library.

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Clark, M. R., Mormede, S., Davies, A. J., and Bowden, D. A. 2016a. Field validation of habitat suitability models for vulnerable marine ecosys-tems in the south pacific ocean: Implications for the use of broad-scale models in fisheries man-agement. Ocean & Coastal Management, 120: 110–126. Elsevier.

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., and Clark, M. R. 2016b. Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. Deep Sea Research Part I: Oceanographic Research Papers, 115:265–292.

Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., and Garcia, R. A. *et al.* 2019. Standards for distribution models in biodiversity assessments. Science advances, 5: eaat4858.

Araújo, M. B., Pearson, R. G., Thuiller, W., and Erhard, M. 2005. Validation of species–climate impact models under climate change. Global change biology, 11: 1504–1513.

Arlot, S., Celisse, A., and others. 2010. A survey of cross-validation procedures for model selec- tion. Statistics surveys, 4: 40–79.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling, 200: 1–19.

Bardos, D. C., Guillera-Arroita, G., and Wintle, B. A. 2015. Valid auto-models for spatially auto- correlated occupancy and abundance data. Methods in Ecology and Evolution, 6: 1137–1149.

Barry, S., and Elith, J. 2006. Error and uncertainty in habitat models. Journal of Applied Ecology, 43: 413–423.

Beale, C. M., and Lennon, J. J. 2012. Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society B: Biological Sciences, 367: 247–258.

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., and Elston, D. A. 2010. Regression analysis of spatial data. Ecology letters, 13: 246–264.

Dolan, M. F. J., Ross, R. E., Albretsen, J., Skarðhamar, J., Gonzalez-Mirels, G., Bellec, V. K., and Buhl-Mortensen, P. *et al.* 2021. Using spatial validity and uncertainty metrics to determine the relative suitability of alternative suites of oceanographic data for seabed biotope prediction. A case study from the Barents Sea, Norway. Geosciences, 11.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., and Davies, R. G. *et al.* 2007. Methods to account for spatial autocorrelation in the analysis of species distribu- tional data: A review. Ecography, 30: 609–628.

Dormann, C. F., Purschke, O., Márquez, J. R. G., Lautenbach, S., and Schröder, B. 2008. Compo-nents of uncertainty in species distribution analysis: A case study of the great grey shrike. Ecology, 89: 3371–3386.

Edwards Jr, T. C., Cutler, D. R., Zimmermann, N. E., Geiser, L., and Moisen, G. G. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. ecological modelling, 199: 132–141.

Efron, B., & Tibshirani, R. 1997. Improvements on Cross-Validation: The .632 Bootstrap Method. Journal of the American Statistical Association, 92(438), 548-560. doi:10.2307/2965703

Elith, J., and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography, 32: 66–77.

Elith, J., and Leathwick, J. R. 2009. Species distribution models: Ecological explanation and pre-diction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40: 677–697.

Elith, J., Burgman, M. A., and Regan, H. M. 2002. Mapping epistemic uncertainties and vagueconcepts in predictions of species distribution. Ecological modelling, 157: 313–329.

Elith, J., Kearney, M., and Phillips, S. 2010. The art of modelling range-shifting species. Methods in Ecology and Evolution, 1: 330–342.

Fielding, A. H., and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conser-vation presence/absence models. Environmental conservation, 24: 38–49.

Fitzpatrick, M. C., and Hargrove, W. W. 2009. The projection of species distribution models and the problem of non-analog climate. Biodiversity and Conservation, 18: 2255.

Fourcade, Y., Engler, J. O., Rödder, D., and Secondi, J. 2014. Mapping species distributions with maxent using a geographically biased sample of presence data: A performance assessment of methods for cor-recting sampling bias. PloS one, 9: e97122.

Fourcade, Y., Besnard, A. G., and Secondi, J. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography, 27: 245–256.

Georgian, S. E., Anderson, O. F., and Rowden, A. A. 2019. Ensemble habitat suitability modelling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. Fisheries Research, 211: 256–274.

Gould, S. F., Beeton, N. J., Harris, R. M., Hutchinson, M. F., Lechner, A. M., Porfirio, L. L., and Mackey, B. G. 2014. A tool for simulating and communicating uncertainty when modelling species distributions under future climates. Ecology and evolution, 4: 4798–4811.

Guisan, A., and Thuiller, W. 2005. Predicting species distribution: Offering more than simple habitat models. Ecology letters, 8: 993–1009.

Guisan, A., Thuiller, W., and Zimmermann, N. E. 2017. Habitat suitability and distribution models: With applications in r. Cambridge University Press.

Hastie, T., Tibshirani, R. and Friedman, J. 2009. The Elements of Statistical Learning Data Mining, Inference, and Prediction. 2nd Ed. Springer. 764 p.

Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., and Sila, A. *et al.* 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PloS one, 10: e0125814.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology: A Journal of the Royal Meteorological Society, 25: 1965–1978.

Hirzel, A., and Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling, 157: 331–341.

Hosmer, D. W. Jr., Lemeshow, S., and Sturdivant, R. X. 2013. Assessing the Fit of the Model. Applied Logistic Regression, 3rd Edn. John Wiley & Sons. doi: 10.1002/9781118548387.ch5

Hughes, J., Haran, M., and Caragea, P. C. 2011. Autologistic models for binary data on a lattice. Environmetrics, 22: 857–871.

Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography, 21: 498–507.

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

Jones-Farrand, D. T., Fearer, T. M., Thogmartin, W. E., III, F. R. T., Nelson, M. D., and Tirpak, J. M. 2011. Comparison of statistical and theoretical habitat models for conservation planning: The benefit of ensemble prediction. Ecological Applications, 21: 2269–2282.

Kearney, M. 2006. Habitat, environment and niche: What are we modelling? Oikos, 115: 186–191.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology, 74: 1659–1673.

Link, J. S., Ihde, T., Harvey, C., Gaichas, S. K., Field, J., Brodziak, J., and Townsend, H. *et al.* 2012. Dealing with uncertainty in ecosystem models: The paradox of use for living marine resource management. Progress in Oceanography, 102: 102–114.

Lobo, J. M., Jiménez-Valverde, A., and Real, R. 2008. AUC: A misleading measure of the perfor-mance of predictive distribution models. Global ecology and Biogeography, 17: 145–151.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H.E., K., B. O., and Zweng, M. M. *et al.* 2013. Temperature. In NOAA atlas nesdis: 73, p. 40. Ed. by S. Levitus and A. Mishonov.

Lutz, M., Dunbar, R., and Caldeira, K. 2002. Regional variability in the vertical flux of particulate organic carbon in the ocean interior. Global Biogeochemical Cycles, 16.

McInerny, G. J., and Purves, D. W. 2011. Fine-scale environmental variation in species distribu-tion model-ling: Regression dilution, latent variables and neighbourly advice. Methods in Ecol-ogy and Evolution, 2: 248–257

Merow, C., Smith, M. J., Edwards Jr, T. C., Guisan, A., McMahon, S. M., Normand, S., and Thuiller, W. *et al.* 2014. What do we gain from simplicity versus complexity in species distribu- tion models? Ecography, 37: 1267–1281

Moudr, V., and Šímová, P. 2012. Influence of positional accuracy, sample size and scale on modelling spe-cies distributions: A review. International Journal of Geographical Information Science, 26: 2083–2095.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., and Ander- son, R. P. 2014. ENMeval: An r package for conducting spatially independent evaluations and estimating optimal model complexity for maxent ecological niche models. Methods in Ecology and Evolution, 5: 1198–1205.

Nephin, J., Gregr, E. J., St. Germain, C., Fields, C., and Finney, J. L. 2020. Development of a species distri-bution modelling framework and is application to twelve species on Canada's Pacific coast. Research Document, 2020/004. Canadian Science Advisory Secretariat.

Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W. M., Zalat, S., and El Din, S. B. *et al.* 2010. Testing the accuracy of species distribution models using species records from a new field survey. Oikos, 119: 1326–1334.

Osborne, P. E., and Suárez-Seoane, S. 2002. Should data be partitioned spatially before building large-scale distribution models? Ecological Modelling, 157: 249–259.

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., and Ingenloff, K. *et al.* 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibra-tion areas. Ecological Modelling, 263: 10–18.

Peterson, A. T., Cobos, M. E., and Jiménez-García, D. 2018. Major challenges for correlational ecological niche model projections to future climate conditions. Annals of the New York Acad- emy of Sciences, 1429: 66–77.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., and Ferrier, S. 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-ab-sence data. Ecological Applications, 19: 181–197.

Pulliam, H. R. 2000. On the relationship between niche and distribution. Ecology letters, 3: 349–361.

Rapacciuolo, G. 2019. Strengthening the contribution of macroecological models to conservation practice. Global Ecology and Biogeography, 28: 54–60.

Record, S., Fitzpatrick, M. C., Finley, A. O., Veloz, S., and Ellison, A. M. 2013. Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change. Global Ecology and Biogeography, 22: 760–771.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., and Hauenstein, S. *et al.* 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phyloge- netic structure. Ecography, 40: 913–929.

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E., and 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.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jimenez-Valverde, A., Ricotta, C., and Bacaro, G. *et al.* 2011. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Pro-gress in Physical Geography, 35: 211–226.

Rooper, C. N., Sigler, M. F., Goddard, P., Malecha, P., Towler, R., Williams, K., and Wilborn, R. *et al.* 2016. Validation and improvement of species distribution models for structure-forming in-vertebrates in the eastern Bering Sea with an independent survey. Marine Ecology Progress Series, 551: 117–130.

Rooper, C. N., Zimmermann, M., and Prescott, M. M. 2017. Comparison of modeling methods to predict the spatial distribution of deep-sea coral and sponge in the Gulf of Alaska. Deep Sea Re-search Part I: Oceanographic Research Papers, 126: 148–161.

Rooper, C. N., Wilborn, R., Goddard, P., Williams, K., Towler, R., and Hoff, G. R. 2018. Validation of deep-sea coral and sponge distribution models in the Aleutian Islands, Alaska. ICES Journal of Marine Science, 75: 199–209.

Rowden, A., Stephenson, F., Clark, M., Anderson, O., Guinotte, J., Baird, S., and Roux, M.-J. *et al.* 2019. Examining the utility of a decision-support tool to develop spatial management options for the protection of vulnerable marine ecosystems on the high seas around New Zealand. Ocean & Coastal Management, 170: 1–16.

Sekulić, A., Kilibarda, M., Heuvelink, G., Nikolić, M., and Bajat, B. 2020. Random forest spatial interpolation. Remote Sensing, 12: 1687.

Serrano, A., González-Irusta, J., Punzón, A., García-Alegre, A., Lourido, A., Ríos, P., and Blanco, M. *et al.* 2017. Deep-sea benthic habitats modeling and mapping in a NE Atlantic seamount (Galicia bank). Deep Sea Research Part I: Oceanographic Research Papers.

Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., and Edwards, T. C. *et al.* 2019. Development and delivery of species distribution models to inform decision-making. BioScience.

Stoklosa, J., Daly, C., Foster, S. D., Ashcroft, M. B., and Warton, D. I. 2015. A climate of uncer-tainty: Accounting for error in climate variables for species distribution models. Methods in Ecol-ogy and Evolution, 6: 412–423.

Stephenson, F., Goetz, K., Sharp, B. R., Mouton, T. L., Beets, F. L., Roberts, J., MacDiarmid, A. B., et al. 2020. Modelling the spatial distribution of cetaceans in New Zealand waters. Diversity and Distributions, 26: 495-516.

Valavi, R., J., E., Lahoz-Monfort, J., and Guillera-Arroita, G. 2018. BlockCV: An r package for generating spatially or environmentally separated folds for k- fold cross- validation of species distribution models. Methods in Ecology and Evolution, 10: 225–232.

Warren, D. L., Matzke, N. J., and Iglesias, T. L. 2020. Evaluating presence-only species distribu-tion models with discrimination accuracy is uninformative for many applications. Journal of Bi-ogeography, 47: 167–180.

Wenger, S. J., Som, N. A., Dauwalter, D. C., Isaak, D. J., Neville, H. M., Luce, C. H., and Dunham, J. B. *et al.* 2013. Probabilistic accounting of uncertainty in forecasts of species distributions under climate change. Global Change Biology, 19: 3343–3354.

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., and Snyder, M. A. 2009. Niches, models, and climate change: Assessing the assumptions and uncertainties. PNAS, 106: 2.

Williams, B. K., and Brown, E. D. 2019. Sampling and analysis frameworks for inference in ecol-ogy. Methods in Ecology and Evolution, 10: 1832–1842.

Winship, A., Thorson, J., Clarke, M. E., Coleman, H. M., Costa, B., Georgian, S., and Gillett, D. *et al.* 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.

Wood, S. N., and Augustin, N. H. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecological modelling, 157: 157– 177.

5 Model outputs

The use of PHM results within ICES advice, will be reliant on managers reviewing and assimilating the model outputs, and interpreting them correctly to provide management advice. The group therefore also considered the required criteria for model outputs. The suggested outputs are shown in Table 5.1. It is important to note that many of the researchers and managers that will be utilizing information provided by the PHM may not be as familiar with the intricacies and details of each modelling method. Therefore, it is important to produce a summary set of descriptors (1-2 paragraphs) that describes at a high level the model goals, method and key results and uncertainty in lay terms. Although not prescribed in the table below, this simplified description is a required component in Annex 3.

Easily understandably maps produced at the appropriate spatial scale, including relevant legends, scale, etc., are required from PHM to support ICES advice. Predictive maps of distribution of VME/indicator species should be accompanied by a map of the spatial uncertainty in model predictions. In addition, associated metadata and spatial layers should be provided, together with a clear explanatory text in a summary report. This summary can facilitate interpretation of model outcomes and avoid any misinterpretation of results by managers and decision makers. This text should include a concise summary of the dependent and independent data, the model approach, interpretation of the uncertainty metrics and overall confidence in model predictions. When multiple models are used and multiple outputs provided, the desired criteria are to include comparable metrics of performance and spatial uncertainty among the different types of model. Resulting maps from the PHM need to be evaluated for ecological plausibility before being used for ICES advice, i.e., a 'sense check' of the outputs. This evaluation should be based on expert opinion and ideally complemented with comparisons to other independent data or peer-reviewed references.

Table 5.1 – Criteria for model outputs required to be produced for PHM when providing ICES advice.

UNACCEPTABLE	REQUIRED	DESIRED
Analysis is not reproducible.	Full method description provided with software, algorithms and settings applied are fully reported.	All software and algorithms, including versions, and settings are reported and the data and scripts used in the analyses are publicly available.
Outputs are overly complex and uninterpretable.	Outputs are simple to interpret, regardless of the underlying model complexity or the use of multiple models.	Same as required criteria.
Explanatory text is not provided.	Explanatory text is provided for the model outcomes and maps in a summary report, using 'plain' language when possible without losing accuracy. This includes a concise summary of the response and predictor variables used and interpretation of the uncertainty metrics.	"Plain" language explanatory text is provided for the model outcomes, following a standardized template (such as Zurell *et al.*, 2020). This includes a concise summary of the response and predictor variables used and interpretation of the uncertainty metrics.

Ecological plausibility is not evaluated.	Ecological plausibility is evaluated based on expert opinion.	Ecological plausibility is evaluated based on expert opinion and compared with independent data or references.
Maps with unclear or inappropriate explanatory elements (e.g., no legend, mismatch of colour scale, inappropriate coordinates used).	Maps include appropriate legends, north arrow, scale, coordinate system and detail of dependent and independent data units used.	Same as required criteria.
Spatial uncertainty is not quantified or not provided as a map output.	Quantification and mapping of the spatial uncertainty and overall confidence in the model is provided using standardized metrics. When multiple models are used, comparable metrics among different types of model are provided if possible.	Spatial uncertainty and overall confidence are quantified and mapped, with multiple models rather than a single one, using standardized and comparable metrics among different types of model.
No metadata provided.	Associated metadata and spatial layers (e.g., GIS layers and observation positions) are provided with model outputs.	Same as required criteria.
Spatial scale does not match the region to be assessed.	Maps are produced at an appropriate spatial scale, relevant to the region being assessed.	Same as required criteria.
Model output lacks clear description of how to interpret the given range of values and/or, post-processing of output values has been done without being clearly stated.	Prediction unit is clearly explained with guidance on how to interpret the range of values and any steps taken to post-process output values, and their implications, are clearly stated	Prediction unit fits the intended purpose ideally.
A single default threshold (such as 0.5) has been used without consideration of response prevalence of intended use.	Threshold selection is based on an optimisation procedure (e.g., sensitivity equals specificity) and its implications are clearly described but not necessarily linked to intended use.	Threshold has been selected based on intended use and determined through model assessment.
The model extrapolates responses far outside the environmental domain of training data without explanation of extrapolation method not described or a large part of predicted area is outside of the training domain.	Model and predictions include a low degree of extrapolation outside the environmental domain of training data with a well described method and supporting evidence.	Model does not extrapolate outside of the training environmental domain.

6 Synthesis of Workshop Results

6.1 Summary of the workshop

During the five-day workshop, there were extensive discussions of existing approaches to PHM and its use in predicting the distribution of VMEs. Much of this discussion centred around the best utilization of observational data, the standards, pros and cons for considering presence, absence and abundance data. Independent data standards were also developed and in general, there was consistent agreement on the types of independent variables that were the most useful; those derived from bathymetry, environmental variables such as temperature, ocean chemistry, productivity and currents, and measures or proxies of substrate composition. The workshop participants also highlighted the need to include historical fishing activity when conducting PHM, either through direct use as independent data or during model evaluation. Vulnerable marine ecosystem indicator taxa are generally long-lived with low population productivity rates. VME indicator taxa also tend to have patchy distributions, with a strong association to their preferred substrate type and areas where there is delivery of passively drifting food items. The relative stability of environmental conditions in the deep sea allows these VMEs to persist, but also makes them vulnerable to changing conditions due to climate effects and human activities (such as fishing or seabed mining). These attributes of VME indicator taxa make them good candidates for PHM, however, the paucity of observational data on their distribution and the absence of directly measured environmental predictions over management scales makes producing a PHM for VME indicator taxa difficult.

There are a large number of existing modelling approaches that have been applied widely in the past (such as maximum entropy models) and some highly anticipated new methods that are just coming into use (such as joint species distribution modelling). There was general agreement that no single modelling approach is the best for all situations, but instead the modelling approach should be tailored to the objective and the available data. It was also agreed that multiple modelling approaches are useful to attempt, since much of the difficult work in VME modelling is in acquiring and developing the underlying independent and dependent data.

Two of the most important considerations decided by the group were the need for transparency in the choices and assumptions around the modelling (e.g. model settings and parameterization,, collinearity in independent variables, etc.) and presentation of uncertainty in a way that demonstrates the model confidence to managers. These considerations are reflected in the recommendations of the workshop, as well as the tables of required and desirable criteria found in the individual sections of the report.

A clear reporting standard was developed (Annex 3) that includes specific items that must be communicated with each model being used to support ICES scientific advice. This standard includes caveats and assumptions inherent in the modelling method. The caveats and assumptions were also discussed in detail during the workshop and it was agreed that for most data that meets the data quality standards, an appropriate modelling method can be found that can be applied to the data.

There was also discussion of broader use of PHM for questions beyond simply identifying the distribution of VMEs. It was noted that VME PHMs can be extended where appropriate and when possible to meet additional management information needs. For example in stock assessment where model effects and ecological mechanisms identified by PHM can inform the ecology of managed stocks and assist in developing next-generation stock assessment models. Models of

VME distribution can also inform important fish habitat, habitat conservation goals (e.g., to assess fishing impacts to habitat), predictions of future states under climate change, and ecosystem management (e.g., Ecosystem Based Management and Ecosystem Based Fisheries Management)— *model once and use many times.*

On day 3 of the workshop a number of trial runs of modelling methods and approaches were conducted. These trial runs used data made available through a GitHub repository (*www.github.com/ices-eg/WKPHM*), which also contains some minimal model code. The trial runs included a maximum entropy approach (J. Burgos), a point-process approach (D. Stirling), a generalized linear model (C. Rooper), a multivariate clustering approach (A. Downie) and a joint species distribution model (P. Thompson). Unfortunately, there was not enough time during the workshop to fully evaluate or complete the model runs and evaluation, but it was noted that in general the criteria set out in the WKPHM were applicable to these approaches. The generalized linear model is used to provide an example of the reporting criteria in Annex 4. It was also noted that the criteria set out in the WKPHM could likely be applied to some existing and published models developed in the ICES region as well.

Based on the discussions, including tables of criteria and results generated at the workshop, a set of required criteria for inclusion in model reporting was recommended that is loosely based on the information requirements of existing ICES benchmarking for regional fish stock assessments. This required reporting criteria are presented in Annex 2 which incorporates the information from tables and text presented in the previous sections of this report.

6.2 Recommendations

There are specific recommendations with regard to the data used in modelling, the types of models to be incorporated, the process of presenting uncertainty in model results and the recommended methods of model validation. Each of the components of model development and assessment has a minimum 'required' standard and, in most cases, a recommended ' desired' standard for models that can exceed the minimum. These specific recommendations are brought together in Table A.2.1 (Appendix 2). A few of the items stand out as important recommendations to highlight here, or are overarching recommendations that apply to all aspects of data and PHM development.

- **Transparency in data and methods**

The primary recommendation from this workshop is that models to be used in ICES advice should be transparent with respect to the data and methods they use. The standard reporting form set out in Annex 3 attempts to provide guidance on the various model components that should be included with reporting any model. However, it is critically important that documentation of the choices made during the data preparation and model development are clearly presented. Transparent and visually consistent (e.g. similar spatial scale, units etc.) model outputs, with user-friendly overviews of input data and modelling methodologies, will contribute to ensuring model credibility and to increasing the confidence in the use of models for management actions and decision-making (Araújo *et al.*, 2019, Sofaer *et al.*, 2019).

- **Clearly state the objective of the PHM to be developed**

A clearly stated objective of the PHM is a key aspect in providing transparency in model selection and development that will assist managers in evaluating whether the model is suitable for providing ICES advice for spatial management of VMEs. For example, a model with the objective of predicting the future distribution of VME under climate change scenarios will have a fundamentally different design parameters than a model that attempts to predict the presence of a VME indicator species in a particular ICES management block. Managers can assist in this

process by providing scientists with clear guidelines on the questions to be addressed, so that scientists can provide model outputs that are appropriate for decision making.

- **Include all available data that meets criteria and standards for inclusion**

The data utilized in models developed for ICES advice should include not only data from the ICES VME database, but other data that can meet the criteria and standards for inclusion (see Section 2.4 for further detail on recommended data standards). There are many data points that are held nationally or in other international databases on the presence and abundance of VME taxa that will meet the criteria set out in Annex 2. In order to provide the best science-based advice to managers, it is important that all available data that is appropriate for use is included.

- **Collect independent data to validate model predictions**

One of the strongest recommendations from this workshop was to encourage the collection of independent data designed to validate model predictions by ICES and its members. This type of model validation is rarely done, but has been crucial to understanding and evaluating models that have been used for management advice. Scientists and managers will value and believe models that have been independently validated with greater confidence than those subject to cross-validation techniques alone (Anderson *et al.,* 2016b, Winship *et al.,* 2019).

- **Include existing and new models in developing ICES management advice**

Due to the varying scales at which previous modelling has been carried out and the potential for producing new models at different scales with different data sources that have corresponding variability in predictive power (e.g. small scale high resolution models may be more precise than global scale low resolution models), there should be enough flexibility to include nested models in ICES management advice to accommodate a range of intended uses.

- **Facilitate communication between science and management**

Clear, comprehensive, and constant communication between science and management will facilitate an informed choice of the most appropriate model to use to address the assessment objectives (Araújo *et al.,* 2019, Kenchington *et al.,* 2019). Summary set of descriptors (1-2 paragraphs) that describes at a high level the model goals, method and key results and uncertainty in lay terms are useful in communicating the basics facets of highly technical PHM to a wider audience with a diverse background in modelling.

- **Develop a systematic approach to PHM in ICES**

Finally, it was recommended that ICES develop a systematic approach to building and routinely updating models within ICES WGs, parallel to the types of benchmarking process that are currently used for stock assessment models. This systematic approach will not only formalize a process for incorporating new information into model development, but it will allow ICES to move forward with new and improved techniques as better data and models become available.

6.3   Next Steps

The results of this workshop suggest that the next steps in the process of incorporating PHM into ICES advice would be two-fold. Firstly, a review of the existing VME models could be undertaken to assess whether they could meet the standards for use in ICES advice. This step would involve a literature review to identify existing models, followed by a ranking against the criteria developed in Annex 2 to judge whether the models could be appropriately used.

Secondly, new models to predict the distribution of VME using PHM should be developed where gaps in existing PHM are found. These new models should follow the guidelines provided in

this report and provide the level of transparency and reporting that will meet the standards in Annex 2. These models should be focused on providing the probability of the occurrence of VME taxa at a scale that will be useful for ICES advice.

Workshop participants also highlighted that if new models were to be developed during ICES expert group meetings, using response (dependent) variable data from the ICES VME database, then prior intersessional work should be done to clean these data for use. Decisions should also be made in advance on the most appropriate environmental (independent) variable datasets for use. This type of exercise is common for stock assessment benchmarking where a data evaluation workshop is held to examine and peer review data for use in the analyses.

A final step that will eventually need to be addressed is incorporating the existing and new models into a comprehensive map for VME indicator taxa. Since many of the existing models have been produced at a smaller scale (both in terms of their spatial extent and/or their grid size), some care will have to be taken if combining or even considering the results from different models. This exercise should also pay special attention to the uncertainty of model inputs and outputs including predictions and spatial residuals when considering the impacts of combining results across multiple model scales.

References

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., and Clark, M. R. 2016. Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. Deep Sea Research Part I: Oceanographic Research Papers, 115: 265–292.

Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R. et al. 2019. Standards for distribution models in biodiversity assessments. Sci. Adv, 5, p. eaat4858.

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., Davis, A., Dunstan, P., Du Preez, C., Finney, J., González-Irusta, J.M., Howell, K., Knudby, A., Lacharité, M., Lee, J., Murillo, F.J., Beazley, L., Roberts, J.M., Roberts, M., Rooper, C., Rowden, A., Rubidge, E., Stanley, R., Stirling, D., Tanaka, K.R., Vanhatalo, J., Weigel, B., Woolley, S. and Yesson, C. 2019. Use of species distribution modeling in the deep sea. Can. Tech. Rep. Fish. Aquat. Sci. 3296: ix + 76 p.

Sofaer H., Jarnevich C., Pearse I.S., Smyth R.L., Auer S., Cook G.L., Edwards Jr. T.C., Guala G.F., Howard T.G., Morisette J.T., Hamilton H. 2019. Development and delivery of species distribution models to inform decision-making. BioScience 20: 1–14.

Winship, A.J., Thorson, J.T., Clarke, M.E., Coleman, H.M., Costa, B., Georgian, S.E., Gillett, D., Grüss, A., Henderson, M.J., Hourigan, T.F., Huff, D.D., Kreidler, N., Pirtle, J.L., Olson, J.V., Poti, M., Rooper, C.N., Sigler, M.F., Viehman, S. and 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.

Annex 1: List of participants

Name	Institute	Country (of institute)	Email
Ana Colaço	OKEANOS_IMAR_University of Azores	Portugal	maria.aa.colaco@uac.pt
Ana de la Torriente	IEO	Spain	ana.torriente@ieo.es
Ana Garcia-Alegre	IEO	Spain	ana.garciaalegre@ieo.es
Anna Downie	Cefas	UK	anna.downie@cefas.co.uk
Anna Metaxas	Dalhousie University	Canada	metaxas@dal.ca
Aurora Nastasi	FAO, GFCM	Italy	aurora.nastasi@fao.org
Beatriz Vinha	University of Salento	Italy	beatriz.vinha@studenti.unisalento.it
Chris Rooper (*chair*)	Fisheries and Oceans Canada	Canada	Chris.Rooper@dfo-mpo.gc.ca
Daniel van Denderen	ICES	Denmark	daniel.vandenderen@ices.dk
David Stirling	Marine Scotland Science	UK	David.Stirling@gov.scot
Eugene Nixon	ICES	Denmark	*eugene.nixon@ices.dk*
Irene Pérez-Rodríguez	IEO	Spain	irene.perez@ieo.es
Jackson Chu	University of Victoria	Canada	jwfchu@gmail.com
Javier Murillo Perez	Fisheries and Oceans Canada	Canada	javier.murillo-perez@dfo-mpo.gc.ca
Jessica Finney	Fisheries and Oceans Canada (DFO)	Canada	Jessica.Finney@dfo-mpo.gc.ca
Jesus Rivera	IEO	Spain	jesus.rivera@ieo.es
Jodi Pirtle	NOAA	USA	jodi.pirtle@noaa.gov
José Manuel González	IEO	Spain	jmanuel.gonzalez@ieo.es
Julian Burgos	Marine and Freshwater Research Institute	Iceland	julian.burgos@hafogvatn.is
Kerry Howell	Plymouth University	UK	kerry.howell@plymouth.ac.uk
Laura Robson	Joint Nature Conservation Committee	UK	Laura.Robson@jncc.gov.uk
Matthew Gianni	Deep Sea Conservation Coalition	Netherlands	matthewgianni@gmail.com
Miriam Tuaty Guerra	retired from IPMA	Portugal	mguerra@ipma.pt
Pål Buhl-Mortensen	Institute of Marine Research	Norway	paal.mortensen@HI.no

Name	Institute	Country (of institute)	Email
Patricia Puerta	IEO	Spain	patricia.puerta@ieo.es
Patrick Thompson	Fisheries and Oceans Canada	Canada	patrick.thompson@dfo-mpo.gc.ca
Peter Auster	University of Connecticut and Mystic Aquarium	USA	peter.auster@uconn.edu
Rebecca Langton	Marine Scotland Science	UK	Rebecca.Langton@gov.scot
Roland Pesch	Institute for Applied Photogrammetry and Geoinformatics, Jade University of Applied Sciences	Germany	roland.pesch@jade-hs.de
Telmo Morato	University of the Azores	Portugal	t.morato@gmail.com

Annex 2: Required and Desired Criteria

Table A.2.1. Summary of required and desired criteria for use in evaluating PHM for use in ICES advice. This table summarizes the criteria developed in the individual report sections and should be applied to new PHM. Existing PHM should also be reviewed for appropriate use in the context of these criteria.

		UNACCEPTABLE	REQUIRED	DESIRED
DEPENDENT (BIOLOGICAL) DATA	Data quality	Sampling design for data collection not described.	All the available data that meet QC standards are used, with a clear description of sampling design(s) and data collection.	Data are sampled via systematic sampling design (which are the same for biological and environmental data) and standardized methods are used for sample collection and processing. A clear description of a robust sampling design is provided.
		Data have no quality control and/or associated metadata. When multiple data sources are combined, there is no description or consideration of the differences.	Quality control of data undertaken, based on metadata, meeting the minimum standards of ICES VME database (e.g., data QA follows national and/or international best practice guidelines and details of which guidelines followed provided). When multiple data sources are combined, the same quality control criteria are followed.	Same as required criteria
		No metadata provided on data sources or the treatment of data.	Metadata/data are reported, following the reporting standard in Annex 3, including a description of the data sources, sampling effort, resolution and extent, and any pre-processing of the data, such as combining multiple data sources and possible spatial thinning or bias correction of data.	Same as required criteria
		Presence-only data is used when absence data is available or the information exists to infer absence data at sampled locations	Presence-only data available.	Presence and absence data and/or abundance data available.
	Caveats, bias and assumptions	Caveats, bias and assumptions in dependent data are not reported.	Caveats, bias and assumptions independent data are clearly reported, acknowledging sources of uncertainty and error.	Same as required criteria
	Taxonomy	Response variable type is not clearly explained.	Response variable type is clearly explained with any caveats on interpretation of outputs.	Same as required criteria
		Taxonomic level is above Class or uses vernacular names.	Taxonomic level is from Class, species or a defined assemblage type (e.g. sea-fan type corals), based on quality-controlled data and expert opinion.	Taxonomic level is species, based on expert identification.

		UNACCEPTABLE	REQUIRED	DESIRED
INDEPENDENT (ENVIRONMENTAL) DATA	**Data quality**	Data have no quality control and/or associated metadata.	Quality control of data undertaken, based on metadata of quality assured (QA) databases or reported survey design and methodology.	Data are sampled via systematic sampling design (same for biological and environmental data) and standardized methods are used for sampling. Clear description of robust sampling design is provided.
		Source(s) of independent data not provided.	Independent data source(s) explained and reported following reporting standards in Annex 3.	Same as required criteria
		Uncertainty is not considered in the environmental data.	Uncertainty of environmental predictors is characterized and accounted for (e.g., use of different analyses and scenario testing).	Uncertainty of environmental predictors is minimized (e.g., use of different analyses and scenario testing).
	Variable choice	Large number of predictor variables included without *a priori* consideration or evidence of their ecological relevance and/or relevance of the range of predictor variables used not evidenced or reported.	Predictor variables have a demonstrable association with the response variable (observational, statistical or theoretical) at the spatial and temporal resolution of the model.	Predictor variables are proximal variables with a confirmed ecologically relevant association with the response variable at the spatial resolution and extent of the model.
		Predictor variables and their ranges are informed by the biological tolerances inferred from other 'proxy' taxa with no expert evaluation for their use.	Predictor variables and their ranges are informed by any known biological tolerances of the VME indicator taxa being modelled, as documented in peer-reviewed studies; or are inferred from those evidenced and documented for 'proxy' taxa with expert evaluation approval for their use.	Predictor variables and their ranges used are defined by the biological tolerance limits of the VME indicator taxa being modelled, as documented in peer-reviewed studies.
	Data processing	Spatial accuracy and resolution are substantially inconsistent between biological and environmental data, e.g., m to hundreds km.	Spatial accuracy and resolution may be inconsistent between biological and environmental data, but data selected are quality-controlled and based on ecological relevance.	Spatial accuracy and resolution of biological and environmental data are the same, or consistent, and are ecologically relevant.
		Native spatial resolution and/or downscaling / aggregation processing methods to match spatial accuracy and resolution between biological and environmental data are not reported.	Native spatial resolution and/or downscaling / aggregation processing methods to match spatial accuracy and resolution between biological and environmental data are fully reported.	Downscaling / aggregation and method used reported and goodness of fit measures for downscaling/aggregation method reported. The effects of variable downscaling/aggregation on model output are tested.
		Derived variables and methods of calculation not described.	Derived variables described and calculations and references provided.	Same as required criteria
	Collinearity	Collinearity has not been investigated or is simply acknowledged or uncorrected.	Collinearity is addressed (e.g. correlation coefficient (r) and variance inflation factor (VIF)) and alleviated, or an approach/model insensitive or able to handle it is used.	Lack of collinearity is demonstrated.
		Collinearity has not been investigated or acknowledged.	Selection of variables to retain from collinear pairs is transparent and has a logical basis. If collinear variables are retained, robustness of the results to collinearity are demonstrated.	Selection of variables to retain from collinear pairs is transparent and is based on significant contribution to model and ecologically relevant association. No collinear pairs of variables are retained in the model.

SPATIAL AND TEMPORAL SCALES	UNACCEPTABLE	REQUIRED	DESIRED
	Spatial and temporal extents, resolutions and location of the study used are not justified.	The spatial and temporal extents, resolutions and location of the study are justified as evidenced from peer-reviewed studies, data availability and/or quality-controlled databases.	The full spatial and temporal, extent, resolution and distribution of the VME indicator taxa are known and used, including current and historical distribution of the VME/indicator.
	Model includes outdated data from locations where natural or anthropogenic influences have changed the response – predictor dynamics.	Model includes data that is relevant to current conditions (including anthropogenic influences).	Model is updated regularly with new data.

Modelling		UNACCEPTABLE	REQUIRED	DESIRED
	Objective	No objectives stated.	Model objective (to explain, predict or project) is stated.	Model objective (to explain, predict or project) is stated and hypotheses for model linkages are clearly stated.
	Modelling method	A priori model selection is made, without taking into account the available data characteristics.	Transparent and clear communication of pros and cons of each model type is provided to inform selection of best possible model.	Same as required criteria.
		No rationale given for choice of modelling method.	Selection of the modelling method is appropriate for the study objective and available data types.	Selection of the modelling method is aligned with objectives and appropriate for the input data. Multiple models evaluated during model development.
		Method is not described.	Method is thoroughly described and/or referenced.	Same as required criteria.
		Assumptions not described.	Assumptions are recognized and their effects on model results are considered and explained.	Assumptions formally evaluated in statistical analysis.
		Violation of assumptions is not reported.	Robustness of model outcomes to violation of assumptions is demonstrated.	Absence of violation of assumptions is demonstrated.
	Model settings	Description of settings used in model is incomplete.	Model specific settings are reported.	Model specific settings, including default ones, should be reported. Model sensitivity to the choice of settings assessed and reported.

	UNACCEPTABLE	REQUIRED	DESIRED
Model terms/coefficients	No information or explanation provided on model terms.	Method of extracting relevant method-specific term estimates or coefficients and how they were evaluated is reported.	Same as required criteria
	Model complexity has not been considered or justified.	Model complexity has been decided/optimised using justified methods or agreed rules of thumb.	Model complexity has been optimised through comparison of multiple models and cross-validation.
	Model outputs have not been evaluated, or model output is not considered plausible.	Model outputs have been evaluated and match understanding of the response taxon's ecology or habitat requirements and the expected distribution.	Model outputs have been evaluated and compared with independent data or established references.
	The relative contribution of predictor variables has not been considered.	Variable importance and how it was determined is reported.	Same as required criteria
Model fit	Goodness-of-fit not considered.	Goodness of fit statistics, and where appropriate residuals, have been checked and their implications to model interpretation are reported.	Goodness of fit statistics and residuals, have been checked and their implications to model interpretation are reported. Data and code are provided.
	Model performance is not reported.	Multiple measures of model performance reported.	Same as required criteria

		UNACCEPTABLE	REQUIRED	DESIRED
Validation		No validation method is provided.	Internal cross-validation (subset of data, not used to build the model).	Validation with truly independent (new collected) data.
		No consideration has been given to the assumptions in data structure when splitting data into training and test sets.	Rationale and method of splitting data into training and test datasets is described, and any caveats explained.	Rationale and method of splitting data into training and test datasets is described, and any caveats explained. Splitting rationale includes a spatial and temporal (if applicable) component.
Uncertainty		Uncertainty is not reported.	Major sources of uncertainties in the model and data are characterized, quantified, accounted for, and discussed. If multiple models are used or ensembled, uncertainties are assessed and compared among models.	Uncertainty, including uncertainty in the environmental predictors, is fully propagated through the modelling process and quantified, mapped, interpreted and discussed in the results. If multiple models are used or ensembled, all uncertainties are assessed and compared among models.
		No consideration for spatial autocorrelation in residuals.	Spatial autocorrelation in residuals has been assessed and, if needed, statistical inference adjusted accordingly.	No spatial autocorrelation present in residuals.
Model Outputs		Bias is not considered, only acknowledged or remain uncorrected.	Correction for major bias is made, testing performance by cross-validation.	Absence of bias is demonstrated or bias is fully corrected.
		Analysis is not reproducible.	Full method description provided with software, algorithms and settings applied are fully reported.	All software and algorithms, including versions, and settings are reported and the data and scripts used in the analyses are publicly available.
		Outputs are overly complex and uninterpretable.	Outputs are simple to interpret, regardless of the underlying model complexity or the use of multiple models.	Same as required criteria.

		UNACCEPTABLE	REQUIRED	DESIRED
		Explanatory text is not provided.	Explanatory text is provided for the model outcomes and maps in a summary report, using 'plain' language when possible without losing accuracy. This includes a concise summary of the response and predictor variables used and interpretation of the uncertainty metrics.	"Plain" language explanatory text is provided for the model outcomes, following a standardized template (such as Zurell *et al.,* 2020). This includes a concise summary of the response and predictor variables used and interpretation of the uncertainty metrics.
		Ecological plausibility is not evaluated.	Ecological plausibility is evaluated based on expert opinion.	Ecological plausibility evaluated based on expert opinion and compared to independent data or references.
		Maps with unclear or inappropriate explanatory elements (e.g., no legend, mismatch of colour scale, inappropriate coordinates used).	Maps include appropriate legends, north arrow, scale, coordinate system and detail of dependent and independent data units used.	Same as required criteria.
		Spatial uncertainty is not quantified or not provided as a map output.	Quantification and mapping of the spatial uncertainty and overall confidence in the model is provided using standardized metrics. When multiple models are used, comparable metrics among different types of model are provided if possible.	Spatial uncertainty and overall confidence are quantified and mapped, with multiple models rather than a single one, using standardized and comparable metrics among different types of model.
		No metadata provided.	Associated metadata and spatial layers (e.g., GIS layers and observation positions) are provided with model outputs.	Same as required criteria.
		Spatial scale does not match the region to be assessed.	Maps are produced at an appropriate spatial scale, relevant to the region being assessed.	Same as required criteria.
		Model output lacks clear description of how to interpret the given range of values and/or, post-processing of output values has been done without being clearly stated.	Prediction unit is clearly explained with guidance on how to interpret the range of values and any steps taken to post-process output values, and their implications, are clearly stated	Prediction unit fits the intended purpose ideally.
		A single default threshold (such as 0.5) has been used without consideration of response prevalence of intended use.	Threshold selection is based on an optimisation procedure (e.g., sensitivity equals specificity) and its implications are clearly described but not necessarily linked to intended use.	Threshold has been selected based on intended use and determined through model assessment.

		UNACCEPTABLE	REQUIRED	DESIRED
		The model extrapolates responses far outside the environmental domain of training data without explanation of ex-trapolation method described or a large part of predicted area is outside of the training data.	Model and predictions include a low degree of extrapolation outside the environmental do-main of training data with a well described method and supporting evidence.	Model does not extrapolate outside of the train-ing environmental domain.

Annex 3:   Data Reporting Template

VME Modelling template
Authors
Date model developed

1.   VME taxonomic group(s) modelled 2. 3.   Regional Extent 4. 5.   Provide a short summary set of descriptors (1-2 paragraphs) that describes at a high level the model goals, method and key results and uncertainty in lay terms. 6.

A. Study resolution

A.1.   Location of the study area (or management region)
  a.   Spatial extent of the modelled area
  b.   Spatial resolution of the model and independent variables
  c.   Spatial precision (of observations and independent variables)
  d.   Depth resolution/range/extent (of the observations and independent variables)

A.2.   Temporal extent of the data
  a.   Dates of data extent
  b.   Precision of date/time
  c.   Data/time resolution
  d.   Impacts over time to consider in the data set (e.g. historical fishing effort)

B. Dependent data

B.1.   Data type (presence, absence, abundance)
B.2.   Data source (e.g. type of survey(s) combined)
B.3.   Measure of sampling effort (if known)
B.4.   Catchability or detectability (known or assumed)
B.5.   Taxonomic level
B.6.   Functional attributes (its ecology)
B.7.   Taxonomic confidence of species/assemblages
B.8.   Rationale for taxonomic/assemblage level modelled
B.9.   Source of absence data
B.10.   Other potential errors or biases in the data
B.11.   Data filtering steps
B.12.   Taxonomic aggregation steps
B.13.   Method for combining dependent data sources (if done outside the modelling)

C. Independent data

C.1.   Independent data (environmental variables used)
C.2.   Independent data source (source of raw or derived data)

C.3. Native spatial and temporal resolution of the independent data

C.4. Data processing and scaling (method for downscaling or aggregation)

    a. Goodness of fit for downscaled aggregated data

    b. Measurement errors and bias

C.5. Derivation methods and calculations for derived variables

C.6. Rationale for inclusion of independent variables clearly stated and ecologically relevant

D. Modelling approach

D.1. Model steps are clearly described with enough detail to be independently reproduced

    a. Code for model provided

    b. Packages used are referenced

    c. Data is made available as supplementary material

D.2. Biases (spatial, temporal and other) acknowledged and described

D.3. Methods and approaches to collinearity in independent variables are given

    a. Collinearity in independent variables tested

    b. Criteria for variable/dimension reduction provided

D.4. Choice of modelling method is explained and justified

    a. Modelling assumptions are clearly stated

    b. Potential violations of model assumptions are explored

D.5. Model application is clearly detailed

    a. Model settings are comprehensively reported

    b. Model complexity is assessed

D.6. Model response curves are generated (where appropriate) and compared to expectations

    a. Modelling method-specific term estimates or coefficients are reported (where relevant)

    b. Independent variable importance is reported

E. Model uncertainty

E.1. Model specific goodness of fit statistics have been checked and reported

    a. Multiple measures of goodness of fit have been examined

E.2. Spatial autocorrelation in the residuals has been assessed and reported

E.3. Residuals have been tested against assumed distribution (where appropriate)

F. Model validation

F.1. Training and testing data splitting method clearly described

    a. Potential spatial biases were accounted for in splitting the data

    b. A standard method used for cross-validation

F.2. Truly independent data used for model validation if available

G. Model outputs

G.1. Maps of model predictions, model residuals and prediction error have been produced

G.2. Areas of model extrapolation are clearly defined

G.3. The prediction unit is clearly defined (and explained if necessary)

G.4. Thresholding methods (for dichotomising probability into presence or absence) are clearly described and appropriate

    a. The sensitivity of model outcomes to threshold value chosen has been explored

Annex 4: WKPHM advice template

WKPHM Advice Template

WKPHM

May 12, 2021

VME Taxonomic Group(s) modelled: The Order *Antipatharia* (Black Corals) including its Families (Table 1)

Regional Extent: North Atlantic Ocean (ICES management subareas 6, 7, 8, 9, 10, 12)

Summary

The objective of this piece of code was to develop a relatively simple model for a species of coral that could be used to demonstrate the pieces of the proposed ICES PHM advice template. The species chosen was *Antipatharia*. It was chosen simply because it had a fairly large number of observations in the ICES VME database (n = 421). This is not meant to be a realistic model of the distribution of *Antipatharia*, but is instead used here to generate the components of an PHM (data, model, residuals) that can be used to evaluate its predictions and utility. The modelling method used was a general linear model with a binomial distribution. Maps of model predictions are provided in Figure 11. Maps of residuals in Figure 9. Maps of prediction error in Figure 12. The model predicted that the highest probability of presence for *Antipatharia* was in a band from 50-60 degrees North latitude and along areas of moderate slope.

A. Study resolution

A.1. Location of the study area (or management region)

This modelling was carried out for the North Atlantic Ocean.

a. Spatial extent of the modelled area

The specific management regions considered for this modelling exercise were ICES subareas 6, 7, 8, 9, 10, and 12 and comprised the spatial extent of the model (Figure 1).

b. Spatial resolution of the model and independent variables

The spatial resolution of the model and independent variables were 30 arc-second grid (~ 1 km^2).

c. Spatial precision (of observations and independent variables)

The spatial precision of the original observations from the ICES database were unknown, but likely varied from m to km. It was assumed that all observations were precisely located, but the spatial resolution (30 arc-seconds) likely reflected a larger scale than the spatial uncertainty of the individual points. The independent variables were compilations from global scale databases and had spatial resolutions of 15 arc-seconds or 0.5 degrees latitude and longitude.

d. Depth resolution/range/extent (of the observations and independent variables)

The depth range of the observations of *Antipatharia* (from the depth data in the VME database) was from 32.5 to 2123 m (mean = 513.7 m, SE = 481.23). The depth range of the modeled area (ICES subareas 6, 7, 8, 9, 10, and 12) was from 0 to 6481 m.

A.2. Temporal extent of the data

a. Dates of data extent

The dates observations were collected ranged from 1980-01-05 to 2018-09-28.

b. Precision of date/time

The precision of the date and the time of the data was assumed to be the closest day.

c. Data/time resolution

The resolution of the date and time was assumed to be the day.

d. Impacts over time to consider in the data set (e.g. historical fishing effort)

Fishing occurred over the entire time frame from which these data points were collected. We did not attempt to account for historical fishing effort over this time. There may have been climate impacts occurring over the time frame of the data observations as well, however, these were not accounted for in the analyses.

B. Dependent data

The dependent data are shown in Figure 1.

Table 1. Number of records for each taxonomic grouping in the order *Antipatharia* from the ICES VME database.

Species	Count
Stichopathes gravieri	385
Antipatharia	148
Stauropathes arctica	25
Parantipathes	14
Parantipathes hirondelle	6
Bathypathes	3
Tylopathes	2

Species	Count
Leiopathes	2
Stauropathes punctata	1

B.1. Data type (presence, absence, abundance)

The data used for modeling *Antipatharia* distribution were observed presences of the Order (n = 421) that occurred in ICES subareas 6,7,8,9,10 and 12.

B.2. Data source (e.g. type of survey(s) combined)

The data were compiled from the ICES VME database (http://vme.ices.dk/map.aspx). As such, they came from a variety of sources including surveys, fisheries observations, etc. No attempt was made to account for the disperate sampling designs of the observations.

B.3. Measure of sampling effort (if known)

Sampling effort was unknown.

B.4. Catchability or detectability (known or assumed)

Neither catchability or detectability were known for this data We assumed the catchability and detectibility were the same for the different types of observations used in the modelling.

B.5. Taxonomic level

The taxonomic level modelled here was the Order *Antipatharia* (see Table 1 for individual components).

B.6. Functional attributes (its ecology)

Antipatharia are a diverse, long-lived and fragile species. They occur in deep-water and are habitat forming structures important to many fishes, invertebrates other taxonomic groups.

B.7. Taxonomic confidence of species/assemblages

The taxonomic confidence of the assemblage was assumed to be good. Identification of *Antipatharians* to the Order level should be attainable for non-experts.

B.8. Rationale for taxonomic/assemblage level modelled

Antipatharia are a group that shares common habitat requirements and depth distribution. They are closely related and the order is globally distributed at deep depths. This Order has been previously modelled using Maximum entropy methods on a global extent (Yesson et al., 2017).
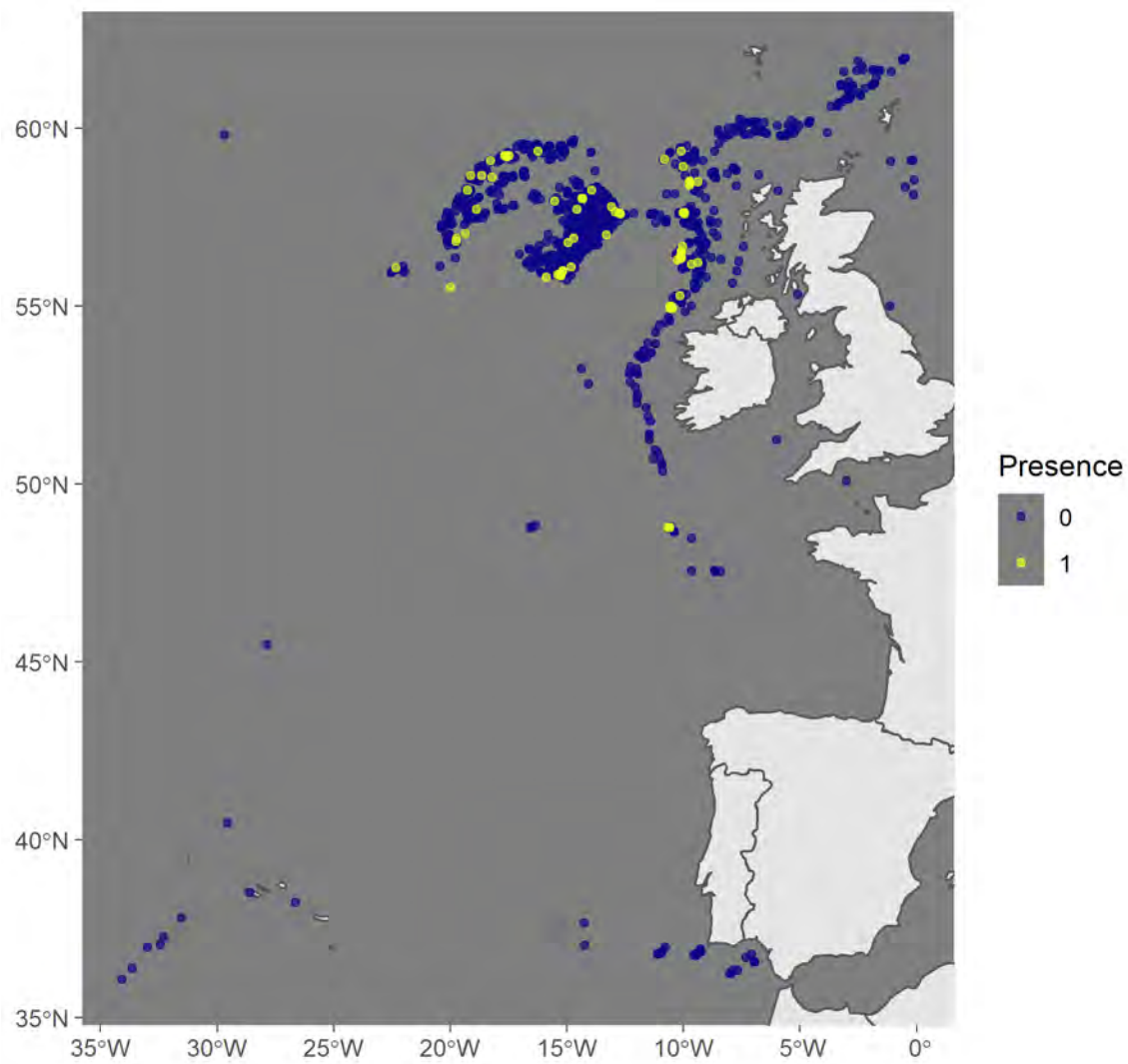
Figure 1: Figure 1. Locations of presence and absence observations for *Antipatharia* from the ICES database

B.9. Source of absence data

Absences were inferred from sample locations where observations for other species were made, but *Antipatharia* were not recorded. In total there were 2745 absences in the ICES VME database.

B.10. Other potential errors or biases in the data

There are many potential sources of error in the data, including errors in positioning of the records, errors in species identification (including both false positives and false negatives). Figure 1 shows that there are some spatial biases in the data set as well.

B.11. Data filtering steps

The code to filter the data completed the following

1. Imported the dependent data from the open ICES data base
2. Subsetted the observations that occurred after 1980
3. Calculated a mean depth for each observation where an upper and lower depth were recorded
4. Compiled a data frame of the unique observation locations (by their recorded middle latitude and longitude)
5. Subsetted all the records where members of *Antipatharia* was observed and assigned them a presence and summed any counts that were recorded at the location (or provides count = 1 where no count was provided, but presence was noted)
6. Merges those presence records with the entire database and assigns presence and counts = 0 to the records with no presence observation
7. There were 165 duplicate record for this the taxonomic grups where the paired entries included different numbers of counts (e.g. a count of *Antipatharia* = 1 and a count of *Stichopathes gravieri* = 3). The counts were summed to 4 for the larger taxonomic group and the duplicate lines removed.

B.12. Taxonomic aggregation steps

The records for the Order *Antipatharia* were aggregated by unique locations (Table 1).

B.13. Method for combining dependent data sources (if done outside the modelling)

No other dependent data sources were used in this modelling.

C. Independent data

C.1. Independent data (environmental variables used)

Four independent variables were used in building a model of *Antipatharia* distribution; bathymetry, topographic position index, seafloor slope and Oxygen concentration (Figure 2).
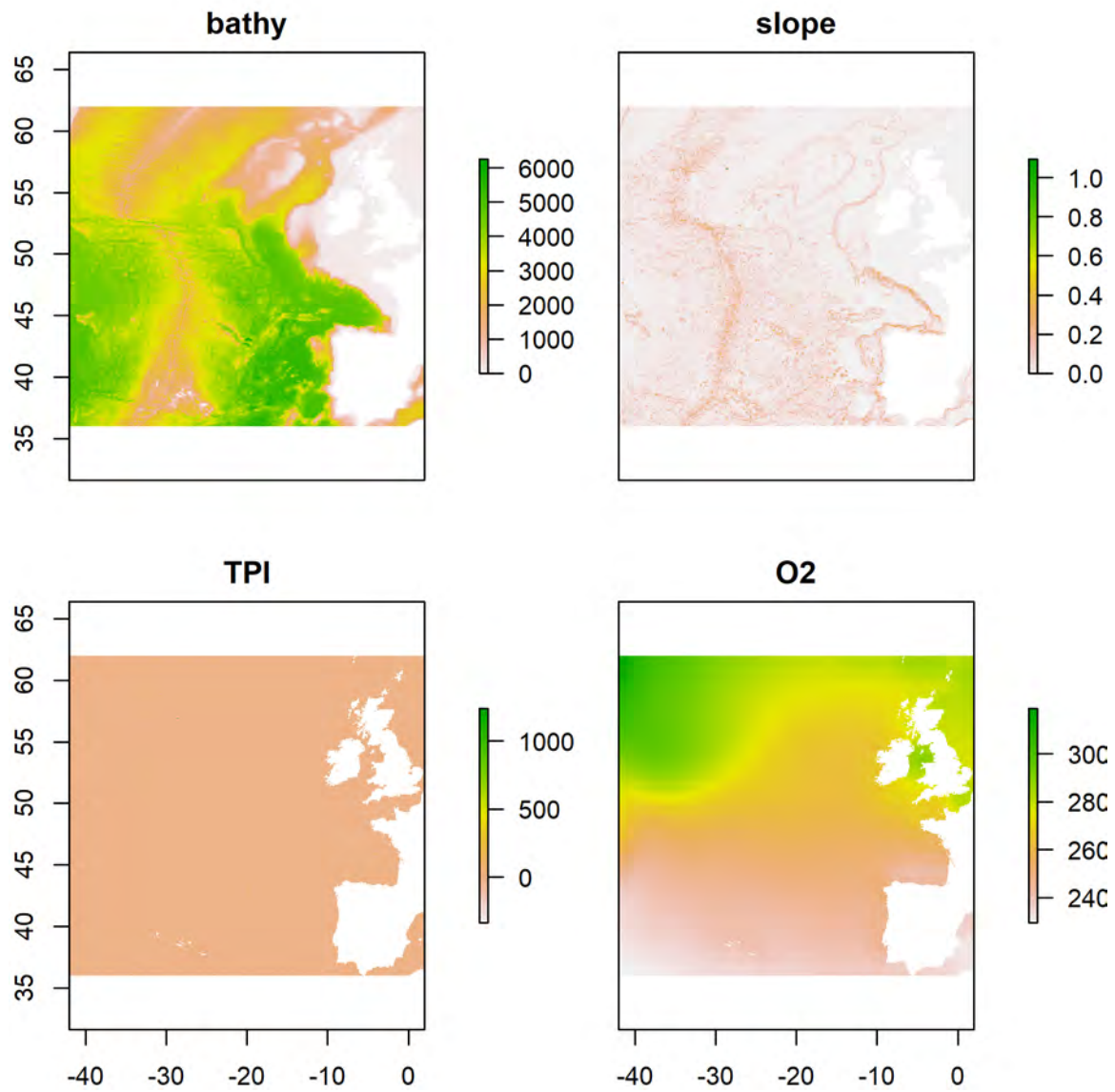
Figure 2: Map of bathymetry, slope, TPI and Oxygen used as explanatory variables in this analysis of ICES VME data

C.2. Independent data source (source of raw or derived data)

The bathymetry used here was downloaded from the GEBCO website (GEBCO_2020 grid; www.gebco.net/data_and_products/
It consists of gridded bathymetry from a wide vareiety of sources on a 15 arc-second grid for the globe. The details of the data sources can be found on the GEBCO website.

From the bathymetry two derived variables (slope and topographic position index) were calculated using the raster package (Hijmans 2019). These variables were calculated on bathymetry aggregated (see below) to a 30 arc-second grid to save processing time.

Oxygen data were downloaded from the World Ocean Atlas 2018 database (https://www.nodc.noaa.gov/OC5/woa18/). This data is a compilation of oxygen measurements averaged over time for as long as there are measurments at standardized depth intervals and on a standard 0.5 degree longitude and latitude grid. These data were clipped to the area of interest and interpolated to the 15 arc-second grid used by the bathymetry. The four explanatory variables are shown in Figure 2.

C.3. Native spatial and temporal resolution of the independent data

The native spatial resolution of the GEBCO bathymetry was 15 arc-second grid. The native spatial resolution for the Oxygen data was 0.5 degrees longitude and latitude. It should be noted that both these data sources are conglomerations of data collected over varying spatial and temporal scales (e.g. the temporal scale is since ~1900's in the case of some bathymetry measurements). For complete documentation of the spatial and temporal scale of the raw data the GEBCO and NODC respective websites should be consulted (www.gebco.net/data_and_products/gridded_bathymetry_data and https://www.nodc.noaa.gov/OC5/woa18/).

Based on the distribution of presences and absences in Figure 1 (and the desire for the code to run fairly fast), a subset of the ICES subareas (SubAreas 6,7,8,9,10 and 12) were chosen for the modeling. All independent data layers were trimmed to include only observations and explanatory variables from this region.

C.4. Data processing and scaling (method for downscaling or aggregation)

Both the bathymetry and oxygen layers were aggregated to a 30 arc-second grid in order to save processing time. For bathymetry, the mean of the 15 arc-second grid cells were aggregated to the 30 arc-second grid. For the Oxygen data, they were first projected and interpolated to the 15 arc-second grid of the GEBCO bathymetry, then these data were aggregated to the 30 arc-second grid (again using the mean value of the smaller grid cells).

a. Goodness of fit for downscaled aggregated data

The aggregated data at the dependent data sites for both Oxygen and bathymetry represented the higher resolution very well (r > 0.9, Figure 3).

b. Measurement errors and bias

Measurement errors in the data or bias in the data were not accounted for beyond the processing conducted on the raw measurements by GEBCO or NODC.

C.5. Derivation methods and calculations for derived variables

From the bathymetry two derived variables (slope and topographic position index) were calculated using the raster package (Hijmans 2019). These variables were calculated on bathymetry aggregated (see below) to a 30 arc-second grid to save processing time.
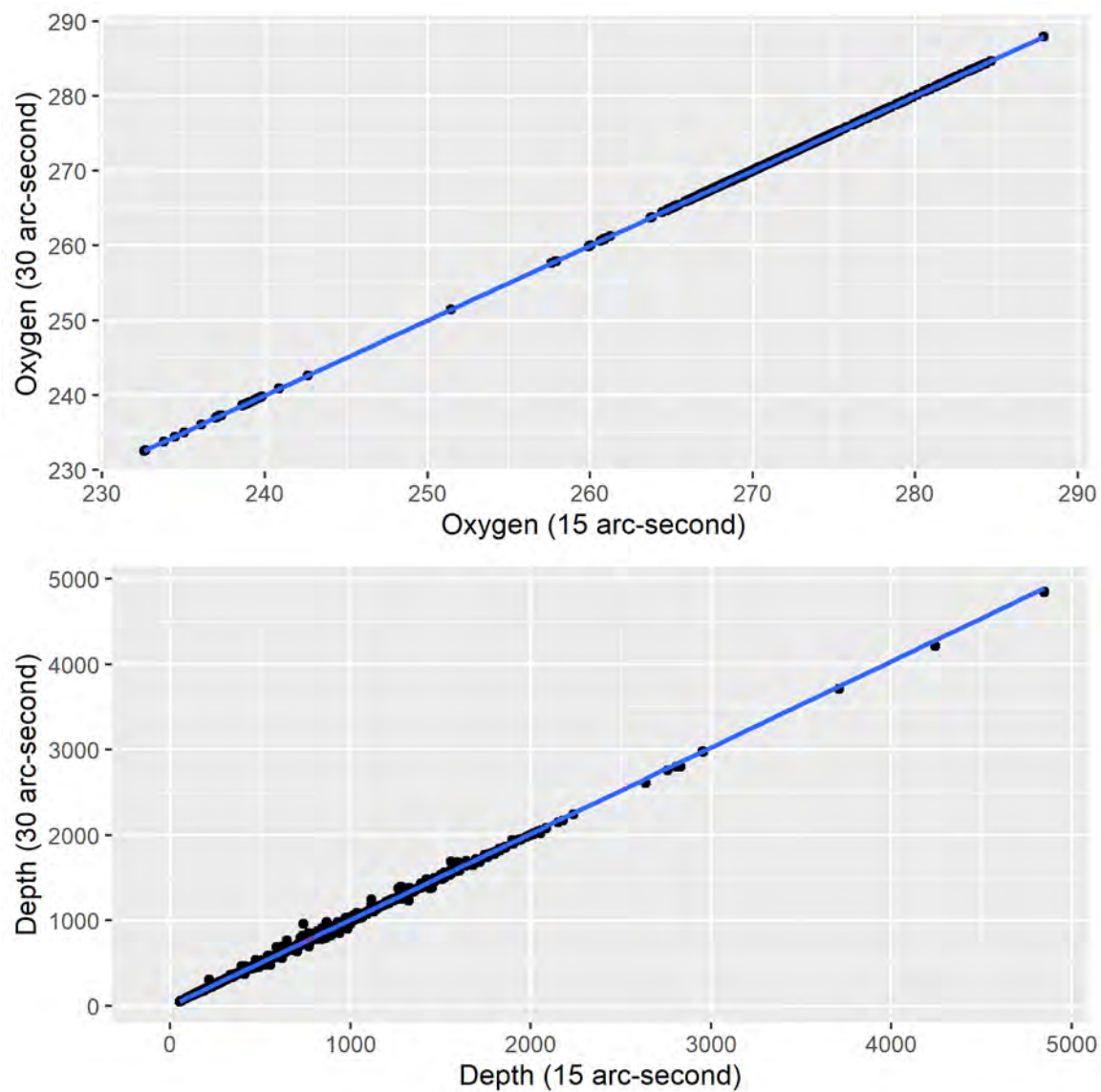
Figure 3: Higher resolution (15 arc-second) and aggregated data (30 arc-second) for Oxygen concentration and bottom depth at the locations of the dependent data.

C.6. Rationale for inclusion of independent variables clearly stated and ecologically relevant

These four variables (depth, slope, topographic position index and oxygen) have been found in previous studies to influence the distribution of *Antipatharia* (Huff et al., 2013, Yesson et al., 2017, Etnoyer et al., 2018).

D. Modelling approach

D.1. Model steps are clearly described with enough detail to be independently reproduced

a. Code for model provided

The code and data used for this model are publically available at https://github.com/ices-eg/WKPHM. The specific code to generate the models and this document are contained in the file "Annex_4_code.Rmd".

b. Packages used are referenced

The packages used to develop this model are referenced in the above .Rmd file. The key packages used were "sf", "rnaturalearth","ggplot2","rgdal","rgeos","gstat","raster","dismo", and "PresenceAbsence" and are all available for download from CRAN. Additional functions are provided in two source scripts at https://github.com/ices-eg/WKPHM. The R version used here was R version 3.6.0 (2019-04-26) – "Planting of a Tree" (R Core Development Team 2019).

c. Data is made available as supplementary material

The independent variables are publically available from a google drive using the code provided in "Annex_4_code.Rmd" or by contacting Chris.Rooper@dfo-mpo.gc.ca. The dependent variables are available from the ICES VME database (see above).

D.2. Biases (spatial, temporal and other) acknowledged and described

There were no inherent biases in the modeling method (although there were biases in the dependent and independant data described above).

D.3. Methods and approaches to collinearity in independent variables are given

a. Collinearity in independent variables tested

The four explanatory variables were examined for collinearity using a pearson correlations (Figure 4). Variance inflation inflation factors (Zuur et al. 2002) were also examined. In both cases the values were low, suggesting that the variables were fairly independent of each other.

Table 2. Variance inflation factors for independent variables using in modeling.

Variable	VIF
Bathymetry	1.399
Slope	1.548

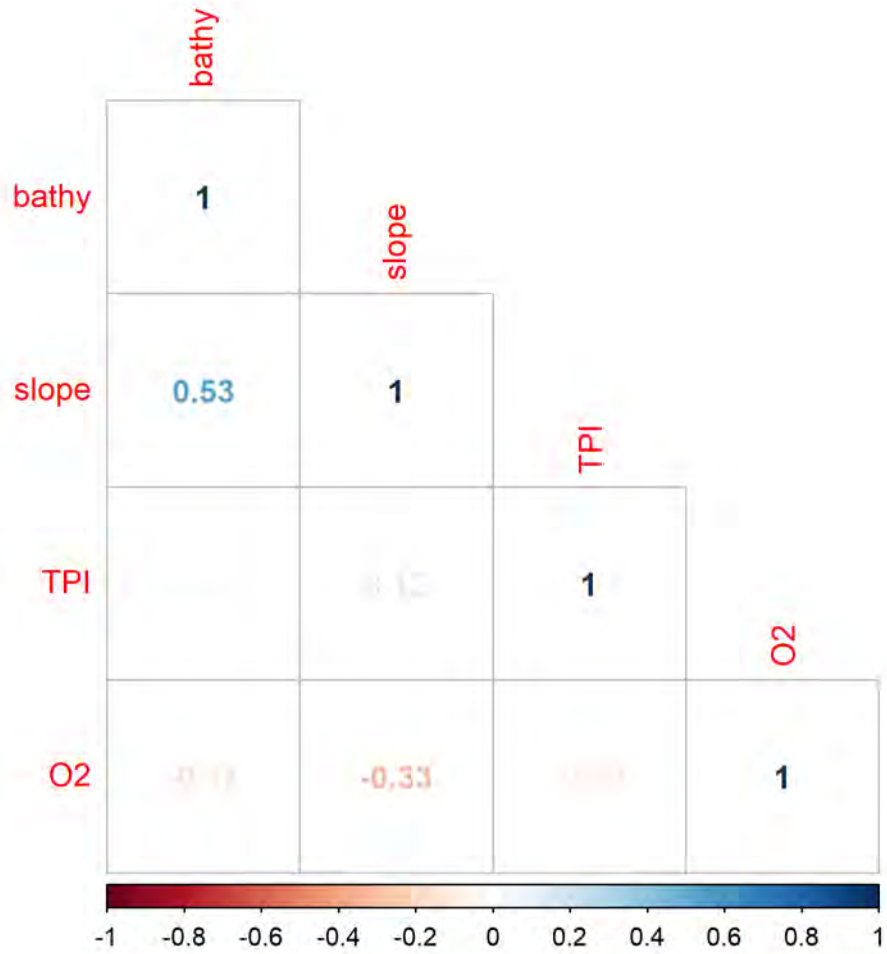| Variable	VIF
TPI	1.018
Oxygen	1.124

Figure 4: Correlation among independent variables used in modeling.

b. Criteria for variable/dimension reduction provided

None of the variance inflation factors exceeded 3, indicating that dimension reduction was not warranted.

D.4. Choice of modelling method is explained and justified

The modelling method chosen was a general linear model (GLM). This model was primarily chosen for its simplicity of assumptions (stated below), its usefulness in fitting binomial (presence-absence) data, and the

many previous applications of this method to predicting species distributions.

a. Modelling assumptions are clearly stated

The basic GLM assumptions are; 1) Independence among data points, 2) The distribution of the residuals is binomially distributed, 3) homogenous variance across the fitted values, and 4) a linear relationship (in this case second order polynomial) between response and predictor.

b. Potential violations of model assumptions are explored

Diagnostic plots of Pearson residuals are shown in Figure 5. The residuals did not indicate any serious violations of GLM assumptions.
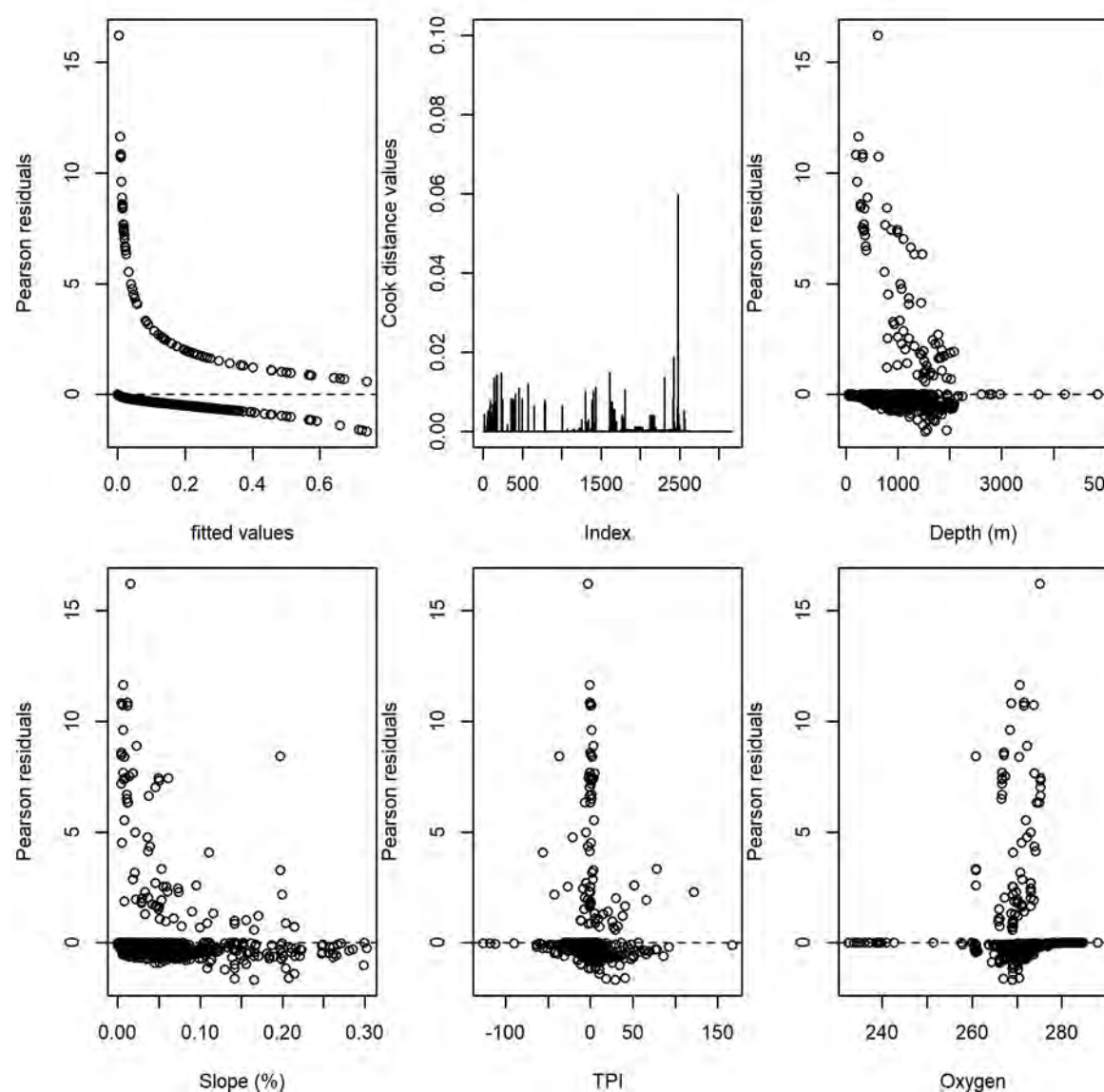
Figure 5: Diagnostic plots for GLM model assumptions.

D.5. Model application is clearly detailed

To build the model of *Antipatharia* a generalized linear model was constructed that contained four explanatory variables (depth, slope, topographic position index and oxygen). Up to second order polynomials were included and the dependent data was presence or absence of *Antipatharia*. The full model was

$$y = \alpha + \beta_1 depth + \beta_2 slope + \beta_3 TPI + \beta_4 O_2 + \beta_5 depth^2 + \beta_6 slope^2 + \beta_7 TPI^2 + \beta_8 O_2^2 + \sigma$$

A binomial error distribution (σ) was used for the model fitting. A full model was fit initially containing all the variables and polynomials. This model was reduced sequentially by removing the least significant term and comparing the AIC for the resulting reduced model. This was repeated until there was no reduction in AIC when removing a variable and all variables remaining in the model were significant.

a. Model settings are comprehensively reported

The default GLM settings in R were used (see Annex_4_code.Rmd). The only setting that was modified was the specification of the binomial error distribution.

b. Model complexity is assessed

The results of the sequential variable reduction resulted in the retention of all four terms; Depth, Slope, TPI and Oxygen (and the polynomials for these variables). The deviance explained by the model (D^2) was 0.438.

The model complexity was assessed against simpler models with less terms during the sequential variable reduction step and the most complex model (containing all terms) was found to be the most appropriate (Table 3).

Table 3. Summary of GLM model predicting presence or absence of *Antipatharia*.

	Df	Deviance	Resid. Df	Resid. Dev
NULL	NA	NA	3165	2482.168
bathy	1	626.057	3164	1856.112
slope	1	158.123	3163	1697.989
TPI	1	5.132	3162	1692.856
O2	1	36.189	3161	1656.667
I(bathy^2)	1	70.881	3160	1585.786
I(slope^2)	1	14.735	3159	1571.051
I(TPI^2)	1	59.421	3158	1511.630
I(O2^2)	1	115.978	3157	1395.652

D.6. Model response curves are generated (where appropriate) and compared to expectations

Model response curves are shown in Figure 6. Probability of presence of *Antipatharia* was highest around 2000 m depth, at a relatively narrow range of Oxygen levels from 260 to 275 ml/L. at moderate slopes > 0.1 and at areas that were slightly elevated from its surrounding bathymetry (on hills). The partial response curves showed a dome shaped response for all four of the significant variables. None of the results were abnormal or unexpected.
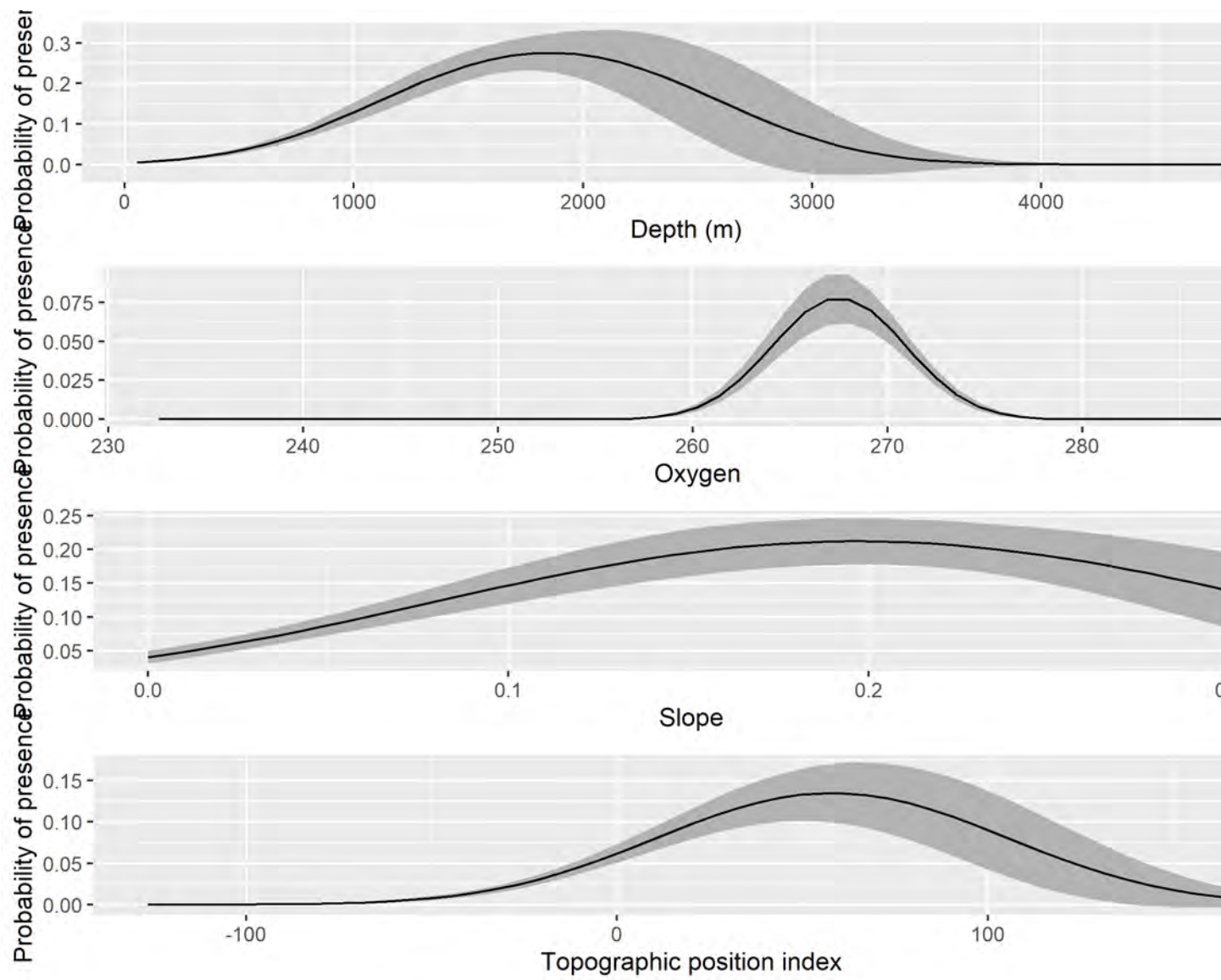
Figure 6: Correlation among independent variables used in modeling.

a. Modelling method-specific term estimates or coefficients are reported (where relevant)

The model specific term estimates are provided in Table 4.

Table 4. Model coefficients, significance and standard error estimates for GLM predicting *Antipatharia* probability of presence.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3264.214	414.094	-7.883	0.000
bathy	0.005	0.001	4.791	0.000
slope	18.800	4.498	4.180	0.000
TPI	0.030	0.004	7.944	0.000
O2	24.364	3.105	7.848	0.000
I(bathy^2)	0.000	0.000	-2.984	0.003
I(slope^2)	-47.744	17.172	-2.780	0.005
I(TPI^2)	0.000	0.000	-4.498	0.000
I(O2^2)	-0.046	0.006	-7.827	0.000

b. Independent variable importance is reported

The relative importance of variables in the model was measured by sequentially removing the individual variables, fitting a new model and calculating the deviance explained. The deviance explained was then scaled to the full model to determine the relative drop in model goodness-of-fit with removal of each variable. The results showed that slope was the most important variable determining the probability of *Antipatharia* presence, and Oxygen was the least important (Figure 7).

E. Model uncertainty

E.1. Model specific goodness of fit statistics have been checked and reported

The *Antipatharia* model AUC was 0.925, an excellent model according to the standards of Hosmer et al., (2013).

Using a threshold of 0.17 resulted in prediction of 358 of the 421 observed presences correctly, while predicting about 86% of the absences correctly (sensitivity = 0.85 and specificity = 0.863)

Table 5. Confusion matrix of predicted and observed presence and absence of *Antipatharia* using a probability threshold 0.17.

	Observed	
Predicted	Absence	Presence
Presence	358	377
Absence	63	2368

a. Multiple measures of goodness of fit have been examined

Commonly used goodness-of-fit measures for binomial models are provided in Table 5 for the GLM predicting *Antipatharia* probability of presence. These include the True Skill Statistic (Allouche et al., 2006), the root-mean-squared-error and the Spearman's rank correlation. Other threshold dependent metrics can be calculated from the confusion matrix (Table 4).
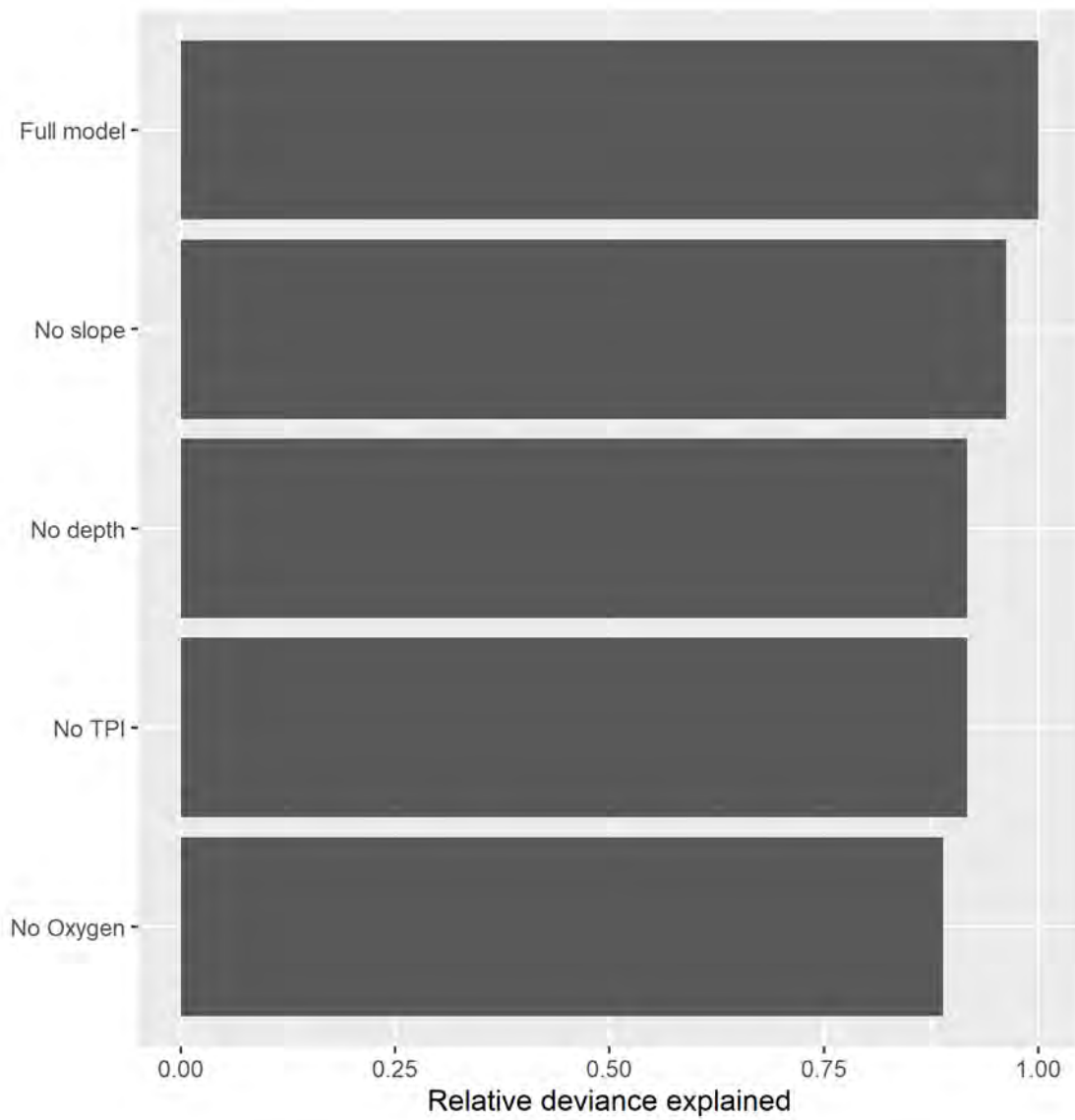
Figure 7: Relative importance of variables included in the Antipatharia presence or absence GLM measured by their contribution to deviance explained when sequentially removed from the model.

Model diagnostics indicated some minor issues with the prediction of presence or absence (Figure 8).The predicted occurence did not always include the 1:1 line (indicating that at some levels of probability the observed occurrences were lower than expected). For example, the model predicted lower than expected probability of presence at ~0.3-0.4 and higher than expected probability of occurrence at ~0.6-0.8.
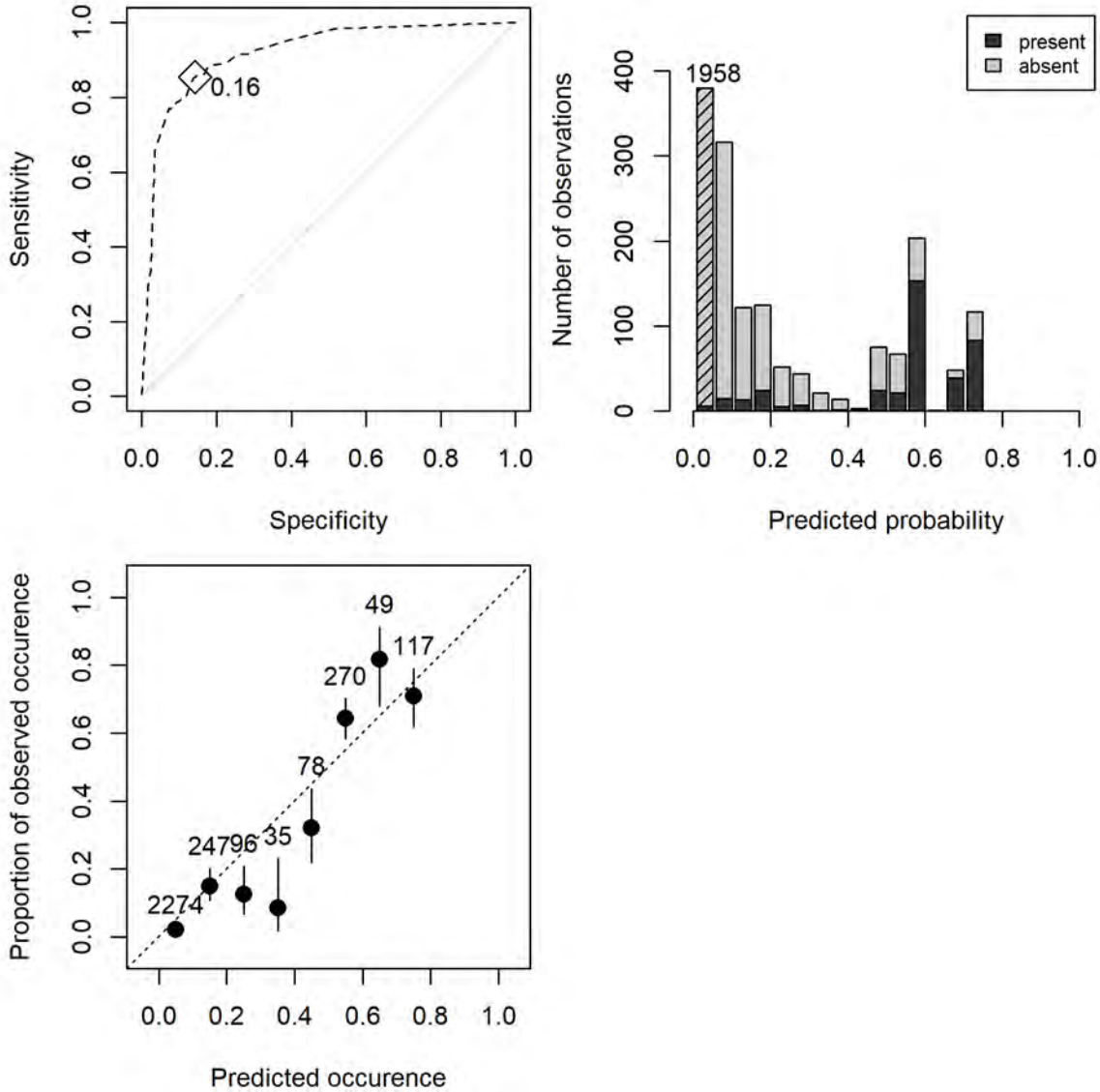
Figure 8: Model diagnostic plots for Antipatharia presence or absence GLM.

E.2. Spatial autocorrelation in the residuals has been assessed and reported

There was significant spatial autocorrelation in the model residuals measured by Moran's I ($I = 3.7747583 \times 10^{-15}$). This was not unexpected given the clumping of observations in the study area.

E.3. Residuals have been tested against assumed distribution (where appropriate)

Not applicable for the binomial distribution. Figure 3 shows model residuals (on the logit scale are shown for each data point used to model *Antipatharia* and diagnostics.

Spatial patterns in residuals

Model residuals are shown in Figure 9. This confirms the results of the Moran's I, in that there is some clustering of larger residuals in the areas sampled at higher densities.

F. Model validation

F.1. Training and testing data splitting method clearly described

An internal model validation method was chosen, since independent data were not available to use as a validation data set. K-fold cross-validation was used here. Five (k) folds were chosen at random. To account for the spatial bias in the data, the spatial block cross-validation package (Valvani et al., 2019) was used to choose the folds.

a. Potential spatial biases were accounted for in splitting the data

The spatial blocking method (Valvani et al., 2019) was used to split the data.

b. A standard method used for cross-validation

k-fold cross-validation is a standard method. The data was divided into 5 equal portions and a model then fit to 80% of the data and tested against the remaining 20% of the data. This was repeated for each subdivision of the data. The same maps and diagnostics were produced for each model fit on the k-folds.

The data folds appeared to show the same patterns as the full model. Relatively high probabilities of presence were apparent in a band from 40-50 North Latitude.

The model performance was similar for all the training data sets (the full model and the individual folds). However the performance of the model on the testing folds was less impressive. For example, the True Skill Statistic for model folds 2 and 5 was very poor, indicating some potential issues with model performance.

F.2. Truly independent data used for model validation if available

No truly independent data was available for model validation.

Table 6. Model goodness of fit measures for the full model and the individual model validation folds

Fold	AIC	threshold	AUC_training	AUC_testing	TSS_training	TSS_testing
Full model	1413.652	0.17	0.925	NA	0.713	NA
GLMFold_1	1202.636	0.09	0.906	0.906	0.697	0.659
GLMFold_2	1208.315	0.15	0.908	0.905	0.679	0.713
GLMFold_3	1202.317	0.19	0.905	0.914	0.694	0.688
GLMFold_4	1173.404	0.17	0.912	0.889	0.701	0.660
GLMFold_5	1222.253	0.11	0.905	0.916	0.690	0.682

Fold	Cor_training	Cor_testing	RMSE_training	RMSE_testing
Full model	0.500	NA	0.256	NA
GLMFold_1	0.481	0.462	0.263	0.269
GLMFold_2	0.480	0.476	0.263	0.262
GLMFold_3	0.475	0.493	0.264	0.264
GLMFold_4	0.482	0.468	0.261	0.273
GLMFold_5	0.477	0.486	0.267	0.253

G. Model outputs

G.1. Maps of model predictions, model residuals and prediction error have been produced

Maps of model predictions are provided in Figure 11. Maps of residuals in Figure 9. Maps of prediction error in Figure 12. The model predicted that the highest probability of presence for *Antipatharia* was in a band from 50-60 degrees North latitude and along areas of moderate slope.

G.2. Areas of model extrapolation are clearly defined

The model was not extrapolated outside ICES subareas 6, 7, 8, 9, 10, and 12, although within this region, there were some areas with little or no sampling.

G.3. The prediction unit is clearly defined (and explained if necessary)

The prediction unit is the probability of presence or absence of *Antipatharia*.

G.4. Thresholding methods (for dichotomising probability into presence or absence) are clearly described and appropriate

No thresholding was done (beyond the thresholding for calculating goodness-of-fit measures).

a. The sensitivity of model outcomes to threshold value chosen has been explored

Sensitivity to threshold values was not explored, but in a formal analysis of the model could be completed using the provided model outputs.

References

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Etnoyer, P. J., Wagner, D., Fowle, H. A., Poti, M., Kinlan, B., Georgian, S. E., & Cordes, E. E. (2018). Models of habitat suitability, size, and age-class structure for the deep-sea black coral Leiopathes glaberrima in the Gulf of Mexico. Deep-Sea Research Part II: Topical Studies in Oceanography, 150, 218–228. https://doi.org/10.1016/j.dsr2.2017.10.008
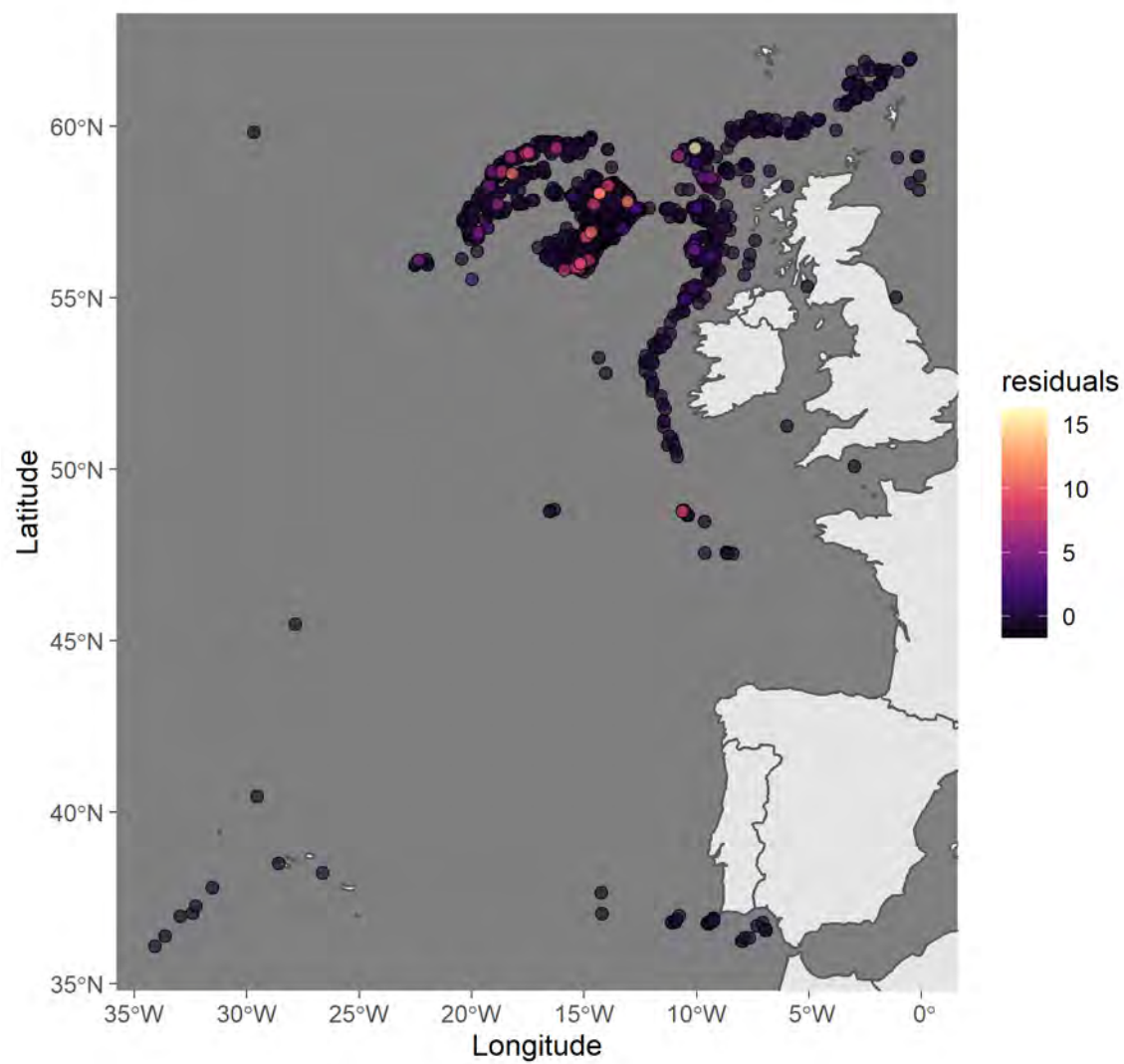
Figure 9: Spatial patterns in model Pearson residuals for GLM predicting probability of *Antipatharia*.
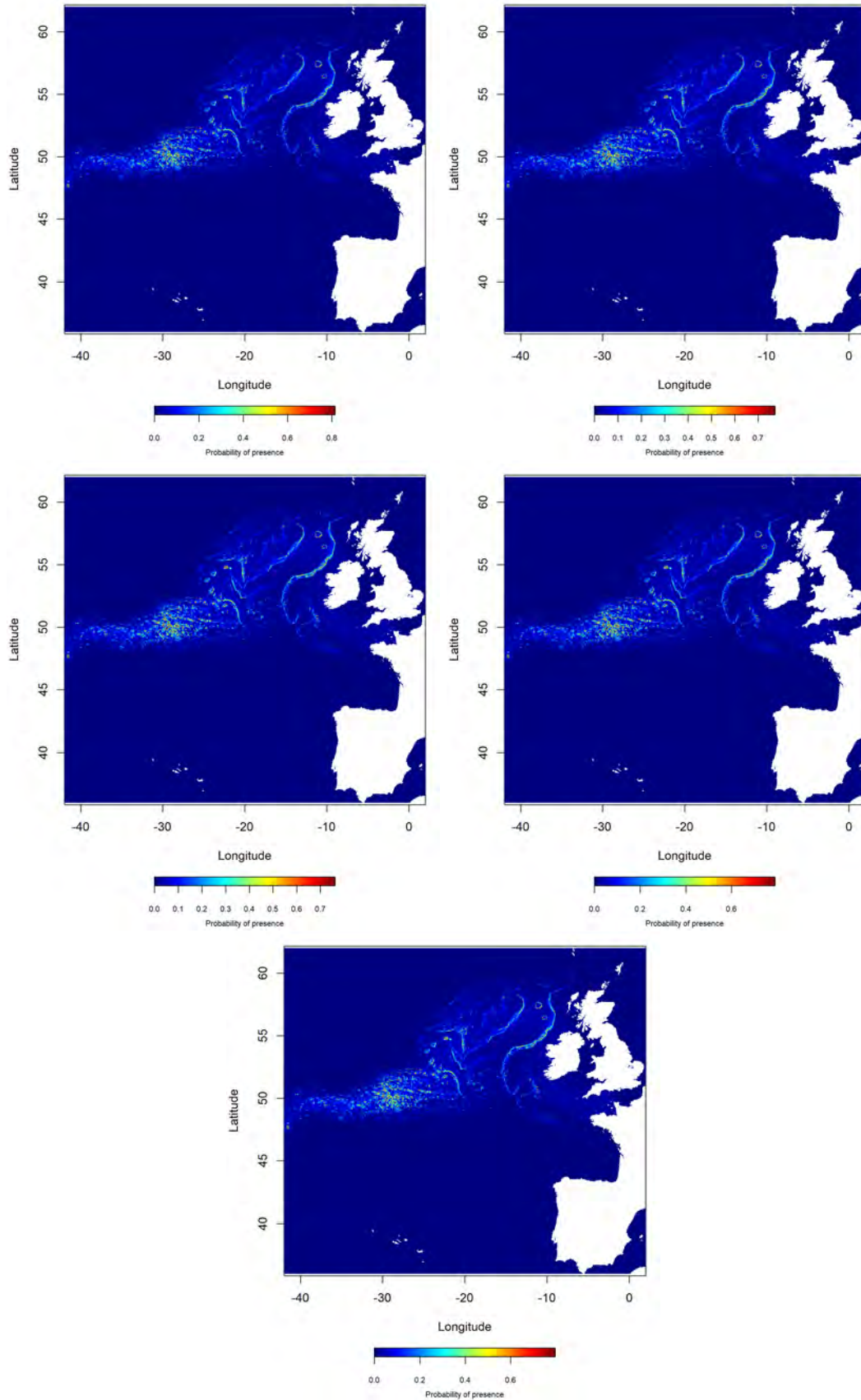
Figure 10: Figure 5. Maps of model predictions for 5 randomly selected folds of the data.
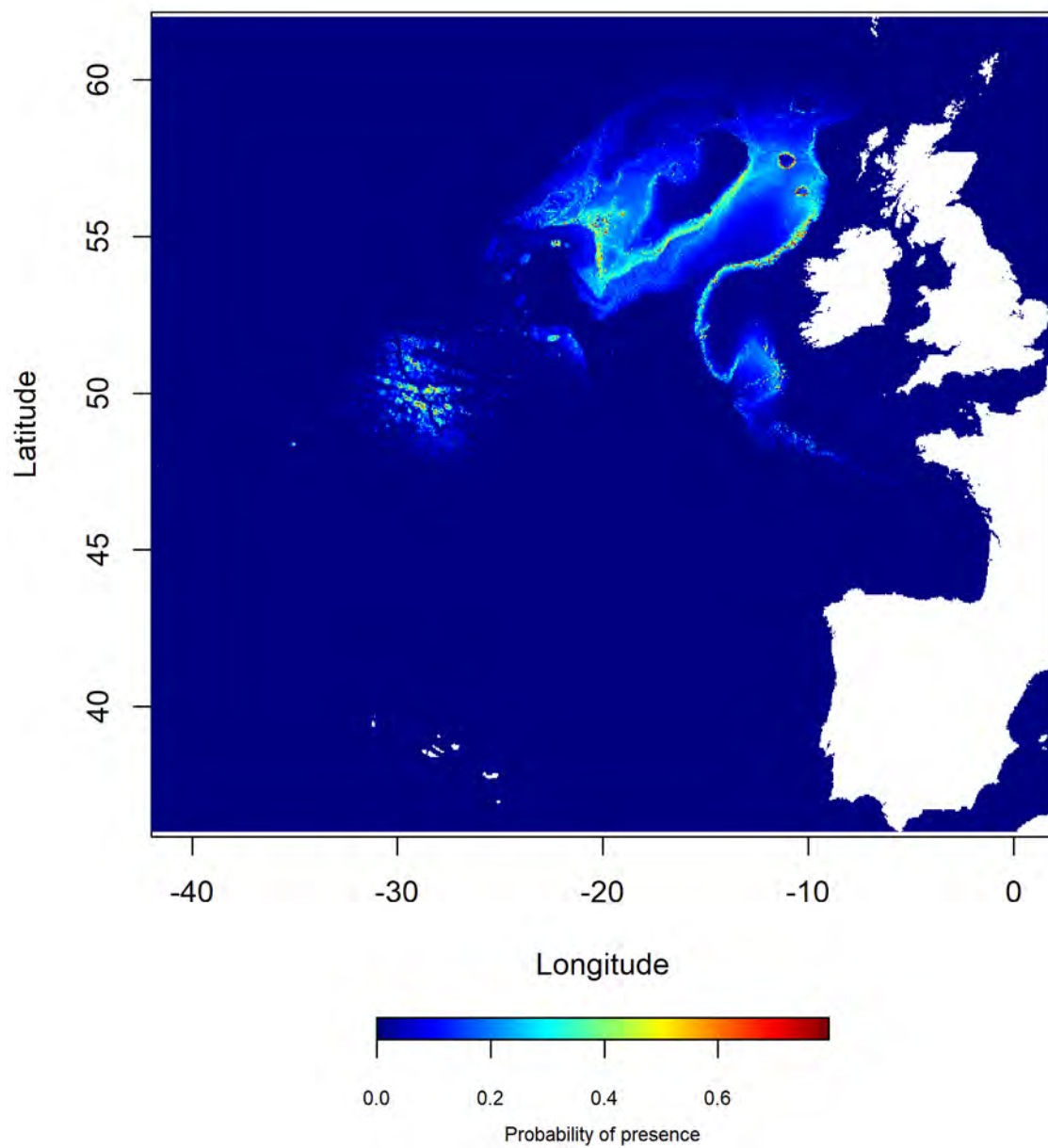
Figure 11: Predicted probability of presence for Antipatharia in SubAreas 6,7,8,9,10 and 12.
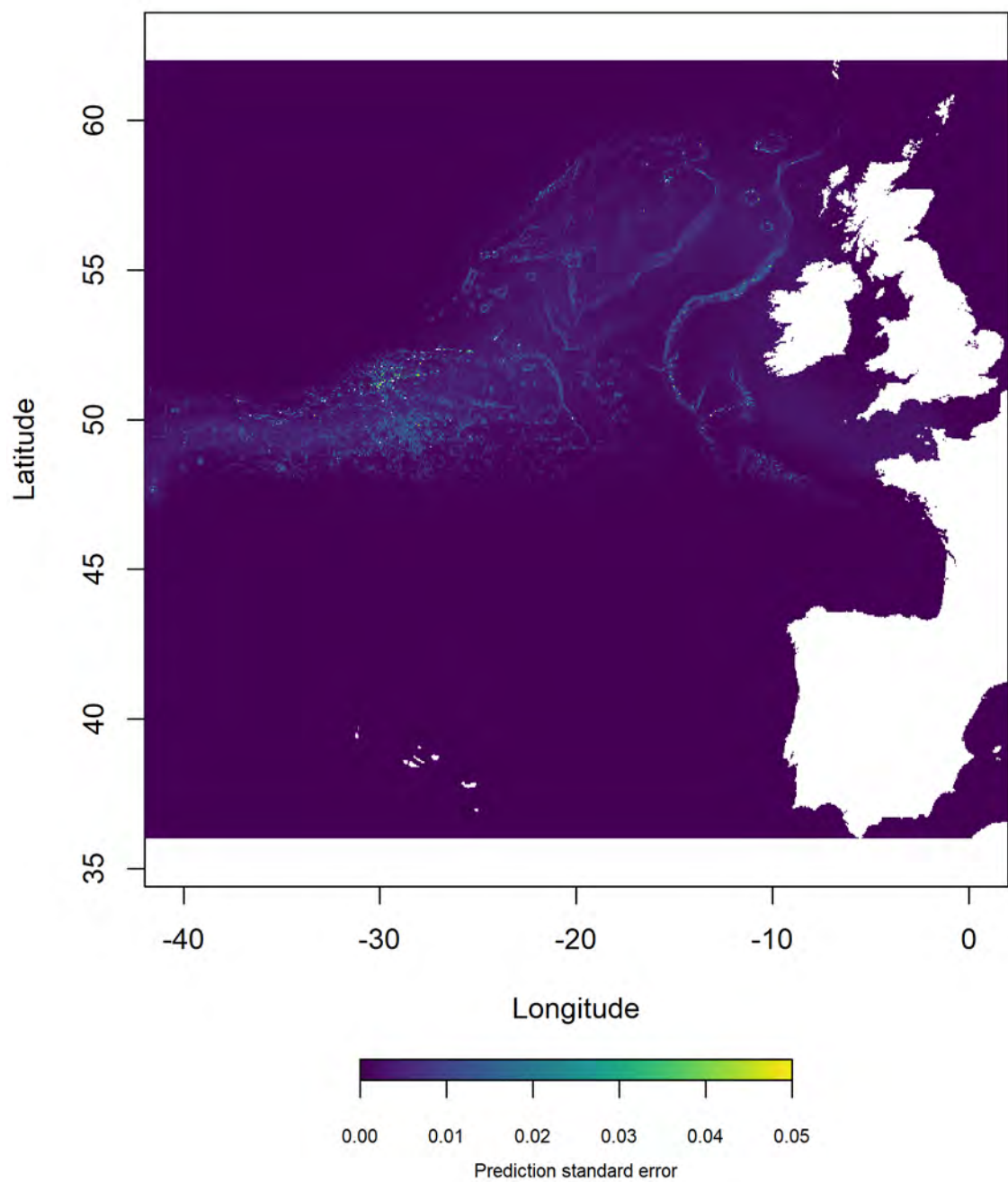
Figure 12: Spatial patterns in model prediction error for GLM predicting probability of Antipatharia in SubAreas 6,7,8,9,10 and 12.

Hijmans R. J. (2019). raster: Geographic Data Analysis and Modeling. R package version 3.0-7. https://CRAN.R-project.org/package=raster

Hosmer, D. W. Jr., Lemeshow, S., and Sturdivant, R. X. (2013). Assessing the Fit of the Model. Applied Logistic Regression, 3rd Edn. John Wiley & Sons. doi: 10.1002/9781118548387.ch5

Huff, D. D., Yoklavich, M. M., Love, M. S., Watters, D. L., Chai, F., & Lindley, S. T. (2013). Environmental factors that influence the distribution, size, and biotic relationships of the Christmas tree coral Antipathes dendrochristos in the Southern California Bight. Marine Ecology Progress Series, 494, 159–177. https://doi.org/10.3354/meps10591

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol Evol. 2019; 10:225–232. https://doi.org/10.1111/2041-210X.13107

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. https://doi.org/10.1016/j.dsr2.2015.12.004

Annex 5: Review of the report from the Workshop on the Use of Predictive Habitat Models (WKPHM) in ICES Advice.

Review of report from the Workshop on the Use of Predictive Habitat Models (WKPHM) in ICES Advice.

Ashley Rowden[1], Andrew Kenny[2], Matthew Poti[3]

[1] National Institute for Water & Atmospheric Research, Wellington, New Zealand

[2] Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, United Kingdom

[3] National Oceanic and Atmospheric Administration, Washington DC, United States of America

Terms of reference for the workshop

(a) Based on existing approaches, identify the methods for modelling vulnerable marine ecosystems (VMEs) that would be most appropriate for use within ICES advice, detailing 'required' and 'desirable' criteria, with emphasis on the deep-sea environment greater than 200m (considering bias of preferential sampling), predictive habitat model (PHM) techniques (including spatial display of uncertainty) and required validation steps for the modelled outputs);

(b) Develop clear standards for recording the caveats and assumptions inherent in the modelling method, for future use;

(c) Conduct a trial run for a small number of existing models to ensure that both the approach and outputs are fit-for-purpose;

(d) Review and recommend a set of criteria, similar to the existing ICES benchmarking system for regional fish stock assessments1, under which new and existing predictive habitat models can be used for ICES scientific advice related to the distribution of VMEs.

The review task

Extracted from email from Daniel Van Denderen, ICES (18/3/21)

"Please comment on the completeness of the work and not on style or general editing. Your review should focus on whether the working group missed important points relevant to the Terms of References [see above], and if you disagree with the conclusions made. If you find that an aspect of the issue was overlooked entirely then please try to draft text to address the point in question, including references as required."

"Your review will be added as an annex to the WKPHM report before the publication of this report. As such, the review text should be a combined product, which, as much as possible, is based on consensus among you."

Review

Overall, the WKPHM report represents a considerable amount of effort and thought by the workshop participants, and is a very useful synthesis of PHM with respect to the variations in approaches, the inherent issues in these approaches, the need for transparency in reporting, as well as standards. With the increasing use of PHMs to inform the management of impacts to VMEs within Regional Fisheries Management Organisations areas, the report is very timely and welcome.

In terms of the completeness of the report, we consider Term of References (a) and (b) have been adequately addressed, but that (c) has not been addressed. We are uncertain about (d), because of our unfamiliarity with the existing ICES bench marking system for regional fish stock assessments. However, ToR (d) simply requests a set of criteria be reviewed and recommended, which the report does, although we note below some further discussion and consideration of how the criteria will be used and applied in a potential benchmark process would be useful.

While the report provides a thorough evaluation of the existing PHM approaches, provides tables of 'required' and 'desirable' criteria, and reporting standards for models of VMEs (i.e., TOR's (a) and (b)), we believe that there are elements of the text/tables that can be expanded upon/clarified to enhance the usefulness of the report (including to widen the readership of the report). In summary these are:

- Consider providing more information/descriptions about the PHM methods themselves – including some discussion of the advantages/disadvantages of each modelling method.
- More attention should be given to the quality, types and quantity of data required to achieve the PHM objectives – i.e., carefully preparing the input data can go a long way to improving the performance of all models irrespective of model choice.

- Include more text about the limitations of presence-only models, including further comment on the implications of using inferred absences (pseudo absences).

- Include some more text on the important issue about not accounting for uncertainty in the independent environmental data.

- The section on model validation approaches (4.1) should be revised to provide clearer descriptions of the methods for each approach. In particular, there needs to be a description of how the .632+bootstrap method works. It uses bootstrapping, but it is not simply bootstrapping. Also, in describing the advantages/disadvantages of each approach, the focus should be on the differences in bias and variance in the estimated performance measures.

- Useful to include some text on the common use of AUC value thresholds to judge whether a model is good/useful or not.

- For some of the required and desired criteria, there were statements that a specific issue needed to be addressed without enough information in the body of the report describing how one could/should adequately address each issue. For example, sampling bias in the dependent data (which is often an issue).

- Include text on use of multiple model and ensemble model approaches (these have already been used for some RFMO areas).

- Include some text in the main body of report on a point that is made in the Conclusion. That is, on the potential nesting of models made at different spatial scales.

- Consider priority ranking of the criteria in Tables 2.3.1, 3.3.1, and 5.1

- Attempt to minimise further the use of jargon and increase the use of plain language for some of the technical details, especially in Section 3.

- Some of the text in the tables need some further elaboration to improve the clarity of meaning.

With respect to the non-inclusion of a trial run for a small number of existing models to ensure that both the approach and outputs are fit-for-purpose (i.e., TOR (c)), we acknowledge that the report contains reference to attempts to undertake an evaluation of different models but that unfortunately time did not allow for its completion. We also acknowledge that the report notes that the examination of existing models is suggested as part of future ICES activities

(this is also noted in the email from Daniel Van Denderen; *"…planning several steps… to identify existing PHMs that meet the standards proposed"*). Nonetheless, we believe that, as originally envisioned by the TORs, the report should contain an assessment of whether any of the existing models for the area of ICES concern in the North Atlantic (i.e., the North Atlantic Fisheries Organisation and North East Atlantic Fisheries Commission Convention areas) currently match the standards developed by the workshop. Including such an evaluation in the report, we believe, would enhance its usefulness considerably. Therefore, we encourage the workshop participants to complete TOR (c) before the report is published in final form. This report has the potential to have high impact, especially in advancing the work around spatial fisheries management and MPAs more generally. Therefore, addressing the points above will ensure the report meets all of its intended objectives, whilst offering the greatest possible benefit to the reader.

With respect to meeting the requirements of ToR (d), the list of essential criteria to be considered when evaluating PHMs as part of a suggested benchmark process is given in Annex 3. The list is comprehensive, although we note some potential ambiguity in the use of terminology that should ideally be resolved. In addition, it would be helpful if the report could include a worked example using the Annex 3 template, so as to guide the user in the types of specific information and detail required when undertaking a benchmark process (possibly to be done as part of completing ToR (c) above). In this respect, it would be helpful if there was further consideration and discussion in the main body of the report of the ICES benchmark process itself and how it could be applied to PHMs using the Annex 3 template. For example, under what circumstances and how often should the PHMs be reviewed or up-dated and is there a requirement for multiple PHM approaches to be considered (where possible or available) to ensure the result is based on the 'best available' method in support of ICES advice.

We attach to this review our detailed comments within the report (see attached). We believe that these comments will assist the authors of the report in addressing some of our concerns noted above. Despite the request to not provide comments on "*style or general editing*", we felt compelled to add some of those also. There are typos, cases of inconsistent use of terms etc., that when addressed will improve the readability of the report.

Annex 6: Responses from WKPHM to the Reviewer reports

Review of report from the Workshop on the Use of Predictive Habitat Models (WKPHM) in ICES Advice.

Reviewer comments below are in blue text, replies from workshop participants are in black text.

Review

In terms of the completeness of the report, we consider Term of References (a) and (b) have been adequately addressed, but that (c) has not been addressed. We are uncertain about (d), because of our unfamiliarity with the existing ICES bench marking system for regional fish stock assessments. However, ToR (d) simply requests a set of criteria be reviewed and recommended, which the report does, although we note below some further discussion and consideration of how the criteria will be used and applied in a potential benchmark process would be useful.

We thank the reviewers for their kind words regarding the report and the completion of ToRs (a, b, and d). We agree that ToR (c) was not adequately addressed during the workshop and initial report (see explanation of revisions below). For ToR (d) we have provided some further comment on revisions included in the report below.

We believe that there are elements of the text/tables that can be expanded upon/clarified to enhance the usefulness of the report (including to widen the readership of the report). these are:

• Consider providing more information/descriptions about the PHM methods themselves – including some discussion of the advantages/disadvantages of each modelling method.

> Although we did not provide much more additional information/descriptions to the PHM methods in the revised report (beyond some modifications to Table 3.2.1), we did provide a relative usefulness ranking for the model types with regard to PHM in Table 3.2.1, as well as the primary reference describing the method. There were ~10 major model types listed in the table (with many subtypes). Describing each type of model was seen as slightly outside the purposes of this report (and better done by examining the references for the models provided). Each of these types of models has a number of data for which it may be better suited, but in honesty there was no consensus within the workshop on the best methods (and likewise this is still an active discussion occurring in the literature). It is likely that the data sets used, familiarity with the method and required outputs drives the decisions on which model researchers choose to use.

• More attention should be given to the quality, types and quantity of data required to achieve the PHM objectives – i.e., carefully preparing the input data can go a long way to improving the performance of all models irrespective of model choice.

> This is a good point by the reviewer and in the revised report, we have reproduced it in Section 2.1. In addition, we have added additional text throughout Section 2 to address this comment (and the individual comments within). We have attempted to focus more on the transparency of presenting the data quality, standards for judging its quality and measures of uncertainty (in particular for independent data).

• Include more text about the limitations of presence-only models, including further comment on the implications of using inferred absences (pseudo absences).

> Additional text and explanation regarding the use of pseudo and inferred absences has been added to Section 2.1 and in Table 2.3.1.

• Include some more text on the important issue about not accounting for uncertainty in the independent environmental data.

> A paragraph describing the issue of accounting for uncertainty and testing for goodness-of-fit in the environmental data and its interpolation methodology has been added to Section 2.3 (lines 438-448) and the accompanying table.

• The section on model validation approaches (4.1) should be revised to provide clearer descriptions of the methods for each approach. In particular, there needs to be a description of how the .632+bootstrap method works. It uses bootstrapping, but it is not simply bootstrapping. Also, in describing the advantages/disadvantages of each approach, the focus should be on the differences in bias and variance in the estimated performance measures.

> Additional details on the .632+bootstrap method and how it differs from both k-fold cross-validation and the traditional bootstrap has been added to Section 4.1. In addition, some additional text has been added that describes the different bias and variance trade-offs for the two validation methods (bootstrapping and k-fold cross-validation.

• Useful to include some text on the common use of AUC value thresholds to judge whether a model is good/useful or not.

> The traditional cutoffs for using AUC (random, adequate, excellent) have been added to Section 4.2.

• For some of the required and desired criteria, there were statements that a specific issue needed to be addressed without enough information in the body of the report describing how one could/should adequately address each issue. For example, sampling bias in the dependent data (which is often an issue).

> In particular, this issue was highlighted for Section 2 (the data section) where a number of items in the criteria table did not have associated text explanations. Here and throughout the report, we have added additional paragraphs and statements to clarify the criteria that are provided in the tables.

• Include text on use of multiple model and ensemble model approaches (these have already been used for some RFMO areas).

> Multiple model and ensemble approaches have now been added in their appropriate locations (eight in total) where needed. For example, on lines 738-739 in Section 3.3 on minimum standards for outputs.

• Include some text in the main body of report on a point that is made in the Conclusion. That is, on the potential nesting of models made at different spatial scales.

> Additional text was added in Section 3.1 to introduce this question. The nesting and ensembling of models will be an interesting and exciting issue to address, but was not directly talked about by this workshop, as we focused our efforts on defining criteria for individual models. Workshop participants envisioned the combining of results as a second step to be taken for models that met the minimum criteria.

• Consider priority ranking of the criteria in Tables 2.3.1, 3.3.1, and 5.1

> No priority ranking of the criteria was formally discussed during the workshop. Although some criteria in these tables are likely more important than others for some applications, we

attempted here to reach consensus on the minimum standards across a broad group of criteria that could be widely applied. For example, a description of the sources of dependent data may be more important than fully testing for collinearity in the independent variables when using a random forest model to conduct PHM. However, accounting for collinearity in independent variables is probably much more important when using a general linear model to conduct PHM. It was not feasible in the context of this report to rank the criteria across the multiple PHM scenarios in which they might be applied.

• Attempt to minimise further the use of jargon and increase the use of plain language for some of the technical details, especially in Section 3.

Throughout the report, with particular attention to Section 3 and other areas that were indicated by reviewers, we have revised the text to provide more explanation of terms and attempted to include more "plain" language explanations and additional examples where possible.

• Some of the text in the tables need some further elaboration to improve the clarity of meaning.

The revised tables (including in the body of the report and Annex 2) have been elaborated on as suggested to improve the clarity of the meaning.

With respect to the non-inclusion of a trial run for a small number of existing models to ensure that both the approach and outputs are fit-for-purpose (i.e., TOR (c)), we acknowledge that the report contains reference to attempts to undertake an evaluation of different models but that unfortunately time did not allow for its completion. We also acknowledge that the report notes that the examination of existing models is suggested as part of future ICES activities (this is also noted in the email from Daniel Van Denderen; *"…planning several steps… to identify existing PHMs that meet the standards proposed"*). Nonetheless, we believe that, as originally envisioned by the TORs, the report should contain an assessment of whether any of the existing models for the area of ICES concern in the North Atlantic (i.e., the North Atlantic Fisheries Organisation and North East Atlantic Fisheries Commission Convention areas) currently match the standards developed by the workshop. Including such an evaluation in the report, we believe, would enhance its usefulness considerably. Therefore, we encourage the workshop participants to complete TOR (c) before the report is published in final form. This report has the potential to have high impact, especially in advancing the work around spatial fisheries management and MPAs more generally. Therefore, addressing the points above will ensure the report meets all of its intended objectives, whilst offering the greatest possible benefit to the reader.

We agree with the reviewers that the failure to meet ToR(c) was unfortunate and that completion of this ToR would be very useful to the target audience for this report. During the next WGMHM meeting ToR(c) will be addressed and 2-3 existing models will be added as Annex's to the WGMHM report. We did add a trial run for a model developed during the workshop as Annex 4 to this report.

With respect to meeting the requirements of ToR (d), the list of essential criteria to be considered when evaluating PHMs as part of a suggested benchmark process is given in Annex 3. The list is comprehensive, although we note some potential ambiguity in the use of terminology that should ideally be resolved. In addition, it would be helpful if the report could include a worked example using the Annex 3 template, so as to guide the user in the types of specific information and detail required when undertaking a benchmark process (possibly to be done as part of completing ToR (c) above). In this respect, it would be helpful if there was further consideration and discussion in the main body of the report of the ICES benchmark process itself and how it could be applied to PHMs using the Annex 3 template. For example, under what circumstances and how often should the PHMs be reviewed or up-

Annex 3 has been revised as suggested to eliminate ambiguity. Annex 4 provides a worked example of the Annex 3 template using one of the trial models completed during the workshop as the example. Although the WKPHM did address the criteria for developing species distribution models for VME, we were not tasked with developing the benchmarking process. It is anticipated that this will happen in subsequent workshops or through the efforts of existing ICES working groups, such as WGMHM and WGDEC.

We appreciate the efforts of the reviewers to provide extensive detailed comments. They were very helpful in indicating weak areas of the report that could be strengthened with additional text and explanation. The detailed comments in particular helped identify specific confusing issues and inconsistent use of terms. All of these comments added considerable value to the report and we thank the reviewers whole-heartedly for their contributions.