Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Proceedings Series 2020/001

Pacific Region

Proceedings of the Pacific regional peer review on Habitat Suitability Modelling Best Practices for Canada’s Pacific Ocean

June 11-12, 2019
Nanaimo, British Columbia

Chairperson: Steven Schut
Editor: Steven Schut and Jill Campbell

Pacific Biological Station
Fisheries and Oceans Canada
3190 Hammond Bay Road
Nanaimo, BC V9T 6N7

March 2020

Canada

Foreword

The purpose of these Proceedings is to document the activities and key discussions of the meeting. The Proceedings may include research recommendations, uncertainties, and the rationale for decisions made during the meeting. Proceedings may also document when data, analyses or interpretations were reviewed and rejected on scientific grounds, including the reason(s) for rejection. As such, interpretations and opinions presented in this report individually may be factually incorrect or misleading, but are included to record as faithfully as possible what was considered at the meeting. No statements are to be taken as reflecting the conclusions of the meeting unless they are clearly identified as such. Moreover, further review may result in a change of conclusions where additional information was identified as relevant to the topics being considered, but not available in the timeframe of the meeting. In the rare case when there are formal dissenting views, these are also archived as Annexes to the Proceedings.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1701-1280

Correct citation for this publication:

DFO. 2020. Proceedings of the Pacific Regional Peer Review on Habitat Suitability Modelling Best Practices for Canada’s Pacific Ocean; June 11-12, 2019. DFO Can. Sci. Advis. Sec. Proceed. Ser. 2020/001.

Aussi disponible en français :

MPO. 2020. Compte rendu de l’examen par les pairs de la région du Pacifique sur la Modélisation de la qualité de l’habitat – Pratiques exemplaires pour l’océan Pacifique du Canada; 11 et 12 juin 2019. Secr. can. de consult. sci. du MPO, Compte rendu 2020/001.

TABLE OF CONTENTS

SUMMARY

These proceedings summarize the relevant discussions and key conclusions that resulted from a Fisheries and Oceans Canada (DFO), Canadian Science Advisory Secretariat (CSAS) Regional Peer Review (RPR) meeting on June 11-12, 2019 at the Pacific Biological Station in Nanaimo, B.C. A working paper establishing the framework for habitat suitability models and species distribution models and an accompanying example application of this framework was presented for peer review.

The major topics discussed were the differences between habitat suitability models and species distribution models and what the models generated within the framework are predicting, the inclusion of the habitat suitability index model in the ensemble model, the inclusion of additional environmental predictors, how to aggregate multiple data points within a single raster cell, and if data obtained with different survey gear can or should be combined.

In-person and web-based participation included Fisheries and Oceans Canada (DFO) Science and Ecosystem Management; and external participants from First Nations organizations, and academia.

The conclusions and advice resulting from this review will be provided in the form of a Science Advisory Report providing advice to DFO Science to inform habitat suitability and species distribution modelling.

The Science Advisory Report and supporting Research Document will be made publicly available on the Canadian Science Advisory Secretariat (CSAS) website.

INTRODUCTION

A regional peer review meeting was held on 11-12 June 2019 at the Pacific Biological Station in Nanaimo to evaluate the best practices framework for developing habitat suitability and species distribution models as established in the working paper.

The terms of reference (ToR) for the science review (Appendix A) were developed in response to a request for science advice from the Science Section of Fisheries and Oceans Canada. Notifications of the science review and conditions for participation were sent to representatives with relevant expertise from Fisheries and Oceans Canada, First Nations, non-governmental organizations, and academia.

The following working paper was prepared and made available to meeting participants prior to the meeting:

Development and Application of a Habitat Suitability Modelling Framework to a Collection of Species on Canada's Pacific Coast by Nephin, J., Gregr, E.J., St. Germain, C., Fields, C., and Finney, J.L. CSAS Working Paper 2018SCI01.

The meeting chair, Steven Schut, welcomed participants, reviewed the role of CSAS in the provision of peer-reviewed advice, and gave a general overview of the CSAS process. The Chair discussed the role of participants, the purpose of the various RPR publications (Science Advisory Report, Proceedings, and Research Document), and the definition and process around achieving consensus decisions and advice. Everyone was invited to participate fully in the discussion and to contribute knowledge to the process, with the goal of delivering scientifically defensible conclusions and advice. It was confirmed with participants that all had received copies of the Terms of Reference, working paper, and draft SARs.

The Chair reviewed the Agenda (Appendix B) and the Terms of Reference for the meeting, highlighting the objectives and identifying the Rapporteur for the review as Jill Campbell. The Chair then reviewed the ground rules and process for exchange, reminding participants that the meeting was a science review and not a consultation. The room was equipped with microphones to allow remote participation by web-based attendees, and in-person attendees were reminded to address comments and questions so they could be heard by those online.

Members were reminded that everyone at the meeting had equal standing as participants and that they were expected to contribute to the review process if they had information or questions relevant to the paper being discussed. In total, 39 people participated in the RPR (Appendix C).

Participants were informed that Chris Rooper, Javier Murillo-Perez and Anders Knudby had been asked before the meeting to provide detailed written reviews for the working paper to assist everyone attending the peer-review meeting (Appendix D). Participants were provided with copies of the written reviews.

The conclusions and advice resulting from this review will be provided in the form of a Science Advisory Report to Science to inform the development of habitat suitability and species distribution models. The Science Advisory Report and supporting Research Document will be made publicly available on the Canadian Science Advisory Secretariat (CSAS) website.

PRESENTATION OF THE WORKING PAPER

All authors were present including Nephin, J., Gregr, E.J., St. Germain, C., Fields, C., and Finney, J.L. An oral presentation was given by Jessica Nephin, Jessica Finney, and Candice St. Germain to summarize the working paper described in the following abstract.

ABSTRACT OF THE RESEARCH DOCUMENT

Models depicting suitable habitat for species are central to making informed decisions about the management and conservation of marine species and places. Habitat suitability models (HSMs) are relevant to several national objectives relating to marine spatial planning, vulnerability assessments, emergency response, and stock assessment. As the lead agency in developing the science underlying these policy goals, Fisheries and Oceans Canada (DFO) will benefit from a consistent application of HSM methods. The effective application of HSMs depends on the relevant ecological and management contexts, the consistent preparation of the available data, and the application of appropriate analytical methods to construct and evaluate the resulting models.

This framework, implemented with purpose-built scripts written in the R statistical programming language, has been prepared as a set of guidelines for the development of consistent, interpretable, and defensible HSMs to support DFO's contribution to Canada's ocean policies. The framework is intended to support analysts interested in developing HSMs, and others interested in best practices related to understanding, integrating, and communicating the results of such models to a broader audience. Specifically for DFO, the challenges related to preparing observational data and marine predictors may be of particular interest.

Where practical, the framework implements accepted best practices to key aspects of data preparation, model fitting, and evaluation. These methods are illustrated using three model building approaches of increasing complexity, which are combined into a final ensemble model. By providing a collection of statistical, graphical, and mapped results, the framework provides a platform for the efficient, standardized development of a suite of HSMs suitable for informing the management of living marine resources.

The application of the framework is demonstrated by using a suite of the best available environmental predictors to predict potential habitat for twelve valued species on Canada's Pacific coast. Species ranged from invertebrates, to habitat-forming plants, to finfish. Hypothesis-based modelling was applied to all species, and emphasized for those found to be data deficient. Future development of HSMs for additional species in Pacific Canada will be greatly facilitated by the set of common predictors, methods, and evaluation tools assembled here.

FRAMEWORK

PRESENTATION OF WRITTEN REVIEWS OF THE FRAMEWORK

Chris Rooper, DFO Science, Ecosystem Sciences

Please refer to Appendix D for full written review. The main comments presented by Chris Rooper in the RPR meeting are listed below:

- Other types of modelling than the three here could be considered, including maximum entropy as it is becoming more commonly used.

- The framework models should be tested with either independently collected data or data collected using survey equipment that is different from that used in model development.
- A clear definition of HSMs and species distribution models (SDMs) should be included as they are unclear in the working paper and unclear more broadly in the literature. As well, a further discussion of realized and fundamental niches as they apply to HSM and SDM would be beneficial.

The authors did not require further clarification of this presentation.

Javier Murillo-Perez, DFO Science, Maritimes

Please refer to Appendix D for full written review. The main comments presented by Javier Murillo-Perez in the RPR meeting are listed below:

- Other types of modelling than the three here could be considered, including generalized additive models (GAM) or random forest models. Using GAMs could provide other techniques for defining spatial predictors and spatial autocorrelation (SAC).
- The framework models should be tested using data collected using survey equipment that is different from that used in model development.
- Similar models developed in the Maritimes used overlay layers to demonstrate predictor uncertainty and similar methods could be employed here.

The authors did not require further clarification of this presentation.

Anders Knudby, University of Ottawa

Please refer to Appendix D for full written review. The main comments presented by Anders Knudby in the RPR meeting are listed below:

- A clear definition of HSMs and species distribution models (SDMs) should be included as they are unclear in the working paper and unclear more broadly in the literature. As well, a further discussion of realized and fundamental niches as they apply to HSM and SDM would be beneficial. The distinction between realized and fundamental niche may have implications to end model objectives.
- Question raised about how models compare with or without the use of SAC. Concern raised about Figure 3.13 (*Map of spatial auto-covariate predictor that represents the distance weighted sum of nearby observations*) and how the cells with no calculated residual were dealt with during the spatial autocorrelation step.
- The relative performance of each model should be displayed for each species as the habitat suitability index (HSI) is different from the generalized linear model (GLM) and boosted regression tree (BRT). A step could be added to the framework to decide if HSI should be added to the ensemble.

The authors clarified that the auto-covariate (ACV) was not interpolated and was only used for model fit and evaluating model test data. The role of the HSI in the ensemble is defined by its AUC value.

GENERAL DISCUSSION OF THE FRAMEWORK

Terminology

The reviewers, and subsequently participants, requested habitat suitability models (HSM) be more clearly defined in the paper. It was noted that the habitat suitability index (HSI) models, built on expert advice, more realistically map fundamental niches and should therefore be described as HSMs, whereas the GLM and BRT models (correlative models), built on presence/absence observations, more realistically map realized niches/distributions and should therefore be described as species distribution models (SDMs). However, a participant noted that depending on the interview questions asked of the experts, the HSI models may be providing either fundamental or realized niche information. However, it was decided that the HSI models are predicting different things than the correlative models as expert advice doesn't include probability of occurrence factors such as effort or survey gear. Additionally, a participant brought up the point that the probability of having species presence observations outside of expert advice is likely as experts are more likely to err on the side of caution (realized niches). It was also noted that the HSI models generally under predicted habitat due to the limiting parameter methods. It was recommended that an ensemble model of all HSI models be developed from individual expert interviews. A participant noted that a spectrum exists from habitat suitability to species distribution and the models generated here fall somewhere along that spectrum.

Outcome: Clear definitions of HSM, SDM, and HSI be added to the paper glossary and an expanded discussion of fundamental and realized niches will be included in the body of the paper. The name of the paper will be changed to reflect that SDMs (as opposed to HSMs) are created in this framework.

Ensemble Model Composition

Since it was determined that the HSI models are modelling species habitat suitability whereas the correlative models are modelling species distribution, there was concern among participants about including the HSI model in the ensemble model. The authors demonstrated that when the HSI model was not included in the ensemble model, there was very little difference in output. The authors also developed a figure to show where the two types of models differ spatially, and it was noted by participants that the HSI model appears to under predict suitable habitat. A participant suggested that the areas of increased difference in habitat suitability between the model types would be good places to sample in subsequent years to test the models.

Outcome: Recommend end users examine the implications of including HSI in their ensemble models. HSI can be more useful with data poor species. Recommend end users develop maps of each model type individually and map of ensemble model including correlative models only. Recommend end users develop maps pairing each set of models to determine where modelling prediction differences exist. Add sample locations on the maps to determine spatial correlation or sampling bias. Recommend end users have clear modelling objectives to help them make these decisions.

Model Selection

Participants noted that there are other model types that could be used in this framework that were not mentioned (GAMs, random forest models, Maxent). The authors noted that other models were not excluded purposely but rather the models used were selected to cover a range of modelling complexities. It was decided that it is not part of the framework to recommend the end user employ specific models.

Outcome: A series of considerations was developed to aid end users in model selection. The authors will mention there is a body of work on model types.

Use of Application Models

It was noted that the species models were developed for oil spill response planning and for framework testing. However, this was not a CSAS meeting to evaluate the acceptance of these twelve species models. It was decided that this meeting process is not sufficient to accept the species models as they are and this paper does not make any recommendations to carry these models forward without further development. However, a participant pointed out that these models represent the best available data, it is likely that they will be used until more comprehensive models can be developed.

Outcome: It was decided that text be added to the paper indicating these models be used with caution and that similar text also be added to all species model figure captions. The application models were determined to be case studies rather than accepted species distribution models.

Model Uncertainty

There was discussion among participants about the difference between cross validation uncertainty and model uncertainty and how to best communicate those differences. A participant noted that cross validation uncertainty has more of an effect on model uncertainty than parameter uncertainty. Additionally, there are differences between model uncertainty and fold uncertainty that should be clarified. Participants determined the current best practice is to look at the residuals against the predicted values and model parameters. However, it was noted that as the data is binary (presence/absence) interpreting the residuals can be tricky, and further analysis may be required (binning, using randomized quantile residuals, or finding an optimal spatial block size). By looking at the map of residuals, any spatial patterns may become evident, in which case an auto-covariate term may need to be used.

Outcome: It was decided that guidance will be provided to the end user regarding the areas of uncertainty and how to best identify them with specific wording developed to reflect framework uncertainty and data uncertainty.

Miscellaneous Topics

A participant expressed concern TEK was not included in the HSI models. It was determined that TEK should be included in the HSI models and that it could act as a model test of the existing HSI models. TEK could also be employed to review species habitat maps.

A reviewer mentioned that independent data should be used to test the models. The group determined that this is a current best practice. It was determined that testing against another survey gear type would not be sufficient as it may highlight bias as not whole species distribution or habitat type may be covered.

A reviewer suggested an ideal model be defined, but it was decided that this cannot be created as model objectives will vary by end user and a gold standard model is not achievable which could deter the use of this framework.

The framework flowchart was slated to be updated to capture model testing with independent data and to make it more general as opposed to specific to the methods used in this application.

A participant noted that there are multiple ways to include spatial structure such as Gaussian random fields. It was decided that text will be added to the working paper to this effect.

A reviewer suggested that an overlay of extrapolation on the model output maps be added. It was decided that this was outside the scope of this framework.

A participant requested clarification be added around how the spatial autocorrelation (SAC) term was used. There was uncertainty if this term was included in the final predictions. The authors will add text in the paper to clarify that SAC was used only to determine if there was an impact of SAC.

Clarification was requested around how the BRT tuning was conducted. The authors will state that the parameter data points are removed randomly and the tuning and testing datasets are spatial.

Participants suggested the authors include text to the end user recommending they consider the spatial coverage of their data and the differing conditions along the coast when denoting model domain.

APPLICATION

PRESENTATION OF WRITTEN REVIEWS OF THE APPLICATION

Chris Rooper, DFO Science, Ecosystem Sciences

Please refer to Appendix D for full written review. The main comments presented by Chris Rooper in the RPR meeting are listed below:

- Many species were not sampled over their entire habitat range. The data sources used are not optimal for the purpose of this model development. There needs to be justification around the choice of data sets included in the models as there appears to be inconsistency.
- The cut-off for model inclusion in the ensemble model (AUC <0.5) could be higher.
- Conclusion section should include more broad conclusions.
- The HSI and correlative models are predicting different things and need to be handled differently.
- Add more figures to show all model maps together, and show all data relationships in one figure.
- It appears the average AUC of the five cross validation models comes out higher than the individual models and this doesn’t make sense.

The authors demonstrated that the ensemble model does not have a higher AUC than the individual models. As well, if one or two of the cross validation models does poorly, the HSI model could compensate when included in the ensemble model.

Javier Murillo-Perez, DFO Science, Maritimes

Please refer to Appendix D for full written review. The main comments presented by Javier Murillo-Perez in the RPR meeting are listed below:

- Other predictor variables could be used. These could also be added to account for inconsistencies due to survey gear type.
- Creating a secondary substrate layer could help with the loss of data when converting line data to point data. Or convert substrate into proportional data points.

- It appears the environmental predictors are for the spring/summer only.
- Including a full analysis of the four species not fully modeled would be helpful to discuss how the quality of the data influences model output.
- Recommend including the presence/absence data points on the maps.

The authors noted that the spring and summer environmental predictor data is more highly correlated with species distributions. Clarification will be added in the paper.

Anders Knudby, University of Ottawa

Please refer to Appendix D for full written review. The main comments presented by Anders Knudby in the RPR meeting are listed below:

- Clarity around what the ecological model is, and how it is applied is needed.
- Further discussion needed around how the GLM interaction terms were derived.
- Concern surrounding how to include the HSI model with the correlative models in the ensemble.

The authors discussed how the ecological model was included in determining if predictors have quadratic or interaction effect in the GLM, and in the predictor selection process, as some predictors were not used if a relationship was not noted in the ecological model.

GENERAL DISCUSSION OF THE APPLICATION

Data Aggregation

The data was prepared by aggregating presence/absence data over time within a single raster cell. There was concern among participants that this method was not a good way to thin the data and that observations were being lost. It was suggested that changing the model input from a binary presence/absence value to a probability of occurrence at that raster cell would be more accurate. It was noted that aggregating data will affect the weighting of cells with many observations compared with cells with few observations. This would be especially prevalent during line surveys over multiple cells, as they would tend towards over predicting presence. Another alternative considered was random sampling of data points, but this method was also deemed to result in a loss of observations, and could be influenced by environmental predictor layers that vary by year and are highly correlated to species presence. Another participant suggested averaging data across the trawls and creating proportional data instead of binary data, but then it was unclear how this could be incorporated into the model framework and was rejected. Ultimately the method used will depend on the objectives of the end user.

Outcome: Authors will clarify in the text how their data was aggregated, the implications of this, and that alternative methods are available in the literature.

Data Set Inclusion in Models

Concern was raised over the justification of survey data inclusion in the Rockfish and Dungeness crab models. Rockfish trawl data was not used as the authors did not want to combine different survey gear type data together (as was the case with the other species models), but for the Dungeness crab model, data from trap, trawl, and dive surveys was used in model building. Participants noted that gear type alters detectability, predictability of catch, and habitat type. The authors prepared a presentation on the Dungeness species model and requested advice from participants on how to best handle combining different data sets. There was concern among participants that the dive data was predicting habitat over a deeper area than it should potentially due to georeferencing errors in the data. It was noted that each data set produced different species distribution maps, but by combining the data sets into one model, the map was able to predict over a wider spatial scale. There was concern that the training models had a high AUC (>0.7), and high variation in ROC curves which shows that the statistics used can be misleading. A participant suggested developing a scaling or calibration factor between gear types to smooth out the catchability differences. There was discussion over which data set to use, whether it be your "best" data set, or the data set that most directly targets the modeled species (in the case of Dungeness crab this would be the trap data) as it is over known suitable habitat. It was suggested that running each data set through the model as well as the combined data set and comparing the maps, as the authors did in their presentation, would be a current best practice which would allow end users to see how their data sets compare to each other. Guidance will be provided in the paper as to how to choose the most appropriate data set to use. The end user needs to keep their modelling objectives in mind when doing this as different types of data will predict different things (i.e., trawl data is over soft substrate habitat only and indicate where crab are likely to be caught, but does not include all suitable habitat). As well, end users need to be aware of data gaps and false absences (i.e., estuaries or deltas are rarely sampled, yet this is where female and juvenile crabs are more likely to be found and these locations are not identified in the models). A participant proposed that each data set modeled could be combined into an ensemble model by adding a weighing factor to the AUC values, but it was noted that the dive data had a high AUC value but was not predicting habitat well. A participant questioned if the cross validation models with low AUC scores should be included in the models. The authors indicated that all five cross validation models were included and that it results in a large standard deviation. A participant raised the point that these data sets are spatially biased, yet extrapolate over a wide area, and questioned the justification for doing so. The authors chose to do this because they didn't feel there was much of a difference of how the data was collected over the north and south shelf bioregions. The authors also noted that this was addressed in their choice of nearshore and shelf spatial scales.

Outcome: The authors will provide further clarification in their paper to justify their decision in data selection differences between Rockfish and Dungeness crab. The authors will recommend end users create maps using each of their data sets to explore how their data may be biased. As well, to address concerns around the implications of extrapolating over a larger spatial scale than the original data set, it was decided that guidance in the paper was required to alert end users of this concern and to encourage them to consider the scale of their data sets, the environmental predictors they choose, and to have clear model objectives. Guidance will also be added that end users should consider using the HSI model instead of the correlative models if the survey gear data sets are not providing accurate species distribution maps.

GLM Interaction Terms

The point was raised by a participant that interactions other than those identified in the GLM could occur. Only known interactions were included in the model, not all possible interactions. It was not clear to some participants if interaction terms were omitted due to computational reasons or ecological reasons. The authors stated that interaction terms were included using expert advice, and since some environmental predictors are highly correlated, including those interaction terms could result in over correlation.

Outcome: It was decided that it should be noted in the text that ecological reasons limited the number of interaction terms used and that interactions may change due to subsequent research or climate change.

Additional Environmental Predictors

There was discussion among participants about creating additional environmental predictors, perhaps based on individual species ecology, over smaller spatial scales, or based on conditions associated with mortality events. It was noted that the intent of this paper was not to develop the best models for each species but to develop a framework.

Outcome: Text will be added to the paper to alter end users to the selection of species specific environmental predictors and that many more environmental predictors exist than those used in this application.

AUC Cut-off for Inclusion in Ensemble Model

There was discussion among participants about increasing the AUC cut-off value for the inclusion of cross validations models in the ensemble model. In the working paper, 0.5 was selected, as anything above this value is better than random, and any other cut-off value may be interpreted as being arbitrary. As well, too high of an AUC cut-off value could result in no models being created.

Outcome: The value will be left at 0.5 and the authors will provide the end user guidance on how to examine model agreement, the standard deviation between the cross validation models, and data quality, which will allow end users to make their own decisions about AUC cut-off values when creating their models.

Parameter Relationship and Model Output Figures

Additional figures were requested including a figure showing all parameter relationships, and a six panel figure with each model output map. These figures will help guide end users in fully analysing their model outputs. To increase the resolution of the nearshore models, additional zoomed in model overlay will be added over the terrestrial area on the larger shelf bioregion maps.

Outcome: To reduce the amount of work for the authors, it was agreed that a full suite of figures for one high-quality data and one low-quality data species example should be produced.

Miscellaneous Topics

Participants noted that species were not sampled over their entire habitat range and that this could skew the models. Text will be added to alert end users to consider the spatial extent of their data when developing their own models.

It will be clarified in the text that the Spearman correlation is in addition to the variance inflation factor.

Some participants are involved in archetype modelling and wondered how this could be applied to the framework. It was decided that archetype modelling is outside the scope of the paper but will be mentioned as a future possibility.

CONCLUSIONS

The working paper was accepted with revisions. It was agreed that the Terms of Reference were met. The following conclusions also appear in the Science Advisory Report.

- The Research Document presents an appropriate review of background materials along with challenges and best practices for HSM development.

- The application of the framework, implemented through R scripts, is a scientifically defensible, transparent, and reproducible method for producing SDMs. It can be used to produce additional species models, or adapted for other modelling extents.
- A multi-model approach facilitates uncertainty estimates by allowing a comparison between different model types.
- The framework has been applied to twelve species using variable quality and quantity of presence-absence data as case studies. The models generally fit the data well as indicated by the performance metrics. The application demonstrates the importance of clarifying key components including data exploration, model interpretation, etc.
- The appropriateness of the model predictions from the application of the framework will need to be assessed by the user.

REVISIONS FOR THE WORKING PAPER

Framework

- Mention other models (e.g., Maxent)
- Mention that collecting new independent data to test the model is the gold standard, and add to recommendations
- Clear definition of HSM, SDM, and realized vs fundamental niche to paper and glossary
- Clarify that the predictions for HSI vs GLM and BRT are different and if HSI is included in the ensemble it is unclear what is being predicted
- Discuss GLM vs GAMs and justify why they weren't included
- Mention that including more methods in an ensemble could be a valid approach
- Mention other ways that spatial structure could be included, (i.e., Gaussian random fields)
- Include discussion of how TEK/LEK could be included in the framework analysis
- Clarify SAC was used only to see if there was an impact of SAC and that it wasn't included in the final predictions
- State that the parameters data points are removed randomly in BRT tuning, and that the tuning and testing datasets are spatial
- Clarify that application uncertainty is mostly cross validation uncertainty, not really parameter uncertainty
- Clarify the differences between model uncertainty and fold uncertainty
- Look for an alternative approach to analyse residuals via bin or randomized quantile residuals, and provide a more complete explanation of the methods surrounding residual creation and interpretation
- Provide guidance on when it is appropriate to include HSI in the ensemble models
- State that the species models are not final models and may require further development for use in management applications
- Include a paragraph about framework "red flags" that end users should consider before/during model development

- Generalize framework flowchart
- Add reference for AUC value cut-off for cross validation models to be included in the ensemble model
- Clarify why and how point observations were aggregated, and the implications of those decisions. Explain other options are available to aggregate data
- Provide guidance on model domain scope with respect to the spatial coverage of your data and differing conditions along the coast. A discussion on stationarity will be included

Application

- Note that not all species were sampled across the entire range of their habitat
- Provide justification for the inconsistencies between what was recommended for data selection for Dungeness and Quillback Rockfish
- Make recommendations for future work (e.g., validation, simulation testing)
- Include six panel figure of all models, add data points to maps, add figure of ensemble SD, include an inset of nearshore maps
- Include a figure showing all parameter relationships
- Include text under Next Steps discussing the inclusion of other predictor variables (e.g., mixed layer, aragonite, fishing effort, other species (e.g., coral and sponges), peak temperature, salinity dips, lethal events)
- State that the Spearman correlation is in addition to VIF
- Clarify why other seasons weren't included in the environmental predictors
- Mention archetype modelling as a future possibility
- State if ecological or computational reasons were the reason for not including all predictor interaction terms in the HSI models
- Provide comments on what would make the best model
- Recommend building multiple HSI models based on various species experts
- Add text to recommend that when datasets are different, each should be modelled individually and the end user should decide if it's appropriate to combine all the data sets or just use one. End users should make this decision based on their final objectives, data and model limitations and/or bias. The Dungeness crab species model will be used as an example.

RECOMMENDATIONS

- Effective application of the framework to approaching management questions requires matching management objectives to model inputs and outputs.
- The appropriateness of the case study predictions from the application of the framework will need to be assessed by the user.
- When applying the framework:
    - Clearly identify model objectives.

    - Examine the spatial and temporal coherence in resolution between the predictor and species data.
    - Diagnose possible errors by looking for unexpected relationships in marginal effects and relative influence of predictors.
    - Consider how the quality of species occurrence data influences model performance.
    - Consider reducing the sample size if a portion of the occurrence data lowers the overall data quality.
    - Examine validation plots and residuals (including spatial patterns in the residuals) in addition to individual statistics to assess model performance.
    - Use AUC and TSS statistics over Kappa and Accuracy to evaluate model performance.
    - Employ spatial block CV to build models and evaluate model fit and transferability.
    - Consider modelling with a spatial predictor to aid in model interpretation.
    - Provide a spatially explicit measure of uncertainty by examining variation across multiple models.
- Use independently collected data for model evaluation where practical.

ACKNOWLEDGMENTS

We appreciate the time contributed to the RPR process by all participants. In particular, we thank the reviewers, Chris Rooper, Javier Murillo-Perez and Anders Knudby, for their time and expertise.

APPENDIX A: TERMS OF REFERENCE

Habitat Suitability Modelling Best Practices for Canada's Pacific Ocean

Regional Peer Review Process – Pacific Region

June 11-12, 2019
Nanaimo, British Columbia

Chairperson: Steven Schut

Context

Habitat Suitability Models (HSMs) are a tool that can be used to predict the distribution of a species' habitat by relating observations of species occurrence to environmental data. This class of models is diverse, and HSMs serve a range of applications, such as predicting habitat for vulnerable species (Rengstorf et al. 2013, Anderson et al. 2016, Rowden et al. 2017), determining important areas in the life history of fish (Le Pape et al. 2014, Rooper et al. 2019), identifying candidate areas for protection (Embling et al. 2009), and forecasting changes in aquatic species habitat due to climate change (Cheung et al. 2010). HSMs also have the potential to maximize the utility of future research efforts and funding by informing survey planning, and directing research towards existing knowledge gaps.

HSMs can therefore help meet conservation and management needs by filling some of these information gaps and increasing our understanding of species distributions for many marine spatial planning initiatives, such as marine protected area network design (Abecasis et al. 2014), oil spill response, and identification of ecologically and biologically significant areas (Beazley et al. 2016). However, HSMs can also be misapplied if species or environmental data are not appropriately screened and prepared, and proper model validation is not carried out (Roberts et al. 2017, Hawkins et al. 200row3, Elith and Leathwick 2009).

The purpose of this peer review is to produce a comprehensive best practices framework to guide future development and application of HSMs, based on 12 species found in the Pacific Region. Although the framework will be developed using Pacific species, there is the potential to apply this framework to other species/regions. Discussions and guidance on appropriate data usage, pre-processing and data preparation, modelling approaches and development, model validation, interpretation of results, and how to convey underlying uncertainty will be provided.

DFO Pacific Science Branch proposes to develop a standardized approach to building HSMs based on best practices to ensure consistent quality and rigor in their use and application. The assessment and advice arising from this Canadian Science Advisory Secretariat (CSAS) Regional Peer Review (RPR) will be used to develop HSMs, and integrate them into science and policy decisions related to the management and conservation of marine species.

Objectives

The following working paper will be reviewed and provide the basis for discussion and advice on the specific objectives outlined below.

Nephin, J., E.J. Gregr, C. St. Germain, C. Fields, J.L. Finney. Habitat Suitability Modelling Best Practices for Canada's Pacific Ocean. CSAP Working Paper 2018SCI01

The specific objectives of this review are to:

Provide background methods, challenges and current best practices for HSM development.

Assess the challenges associated with finding appropriate species and environmental data, and provide guidance on best practices for preparing data for HSMs.

Develop a framework that applies best practices to HSM development; including how to validate models and present model assumptions and uncertainty.

Demonstrate an application of the HSM framework using a case study in Pacific Regionz.

Examine findings relevant to building HSMs and provide recommendations for future applications.

Expected Publications

- Science Advisory Report
- Research Document
- Proceedings

Expected Participation

- Fisheries and Oceans Canada (Ecosystems and Oceans Science, Fisheries Management, Ecosystems Management)
- Academia
- Indigenous Groups
- Non Government Organizations

References

Abecasis, D., Afonso, P. and Erzini, K., 2014. Combining multispecies home range and distribution models aids assessment of MPA effectiveness. Marine Ecology Progress Series, 513, pp.155-169.

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., & 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. https://doi.org/10.1016/j.dsr.2016.07.006

Beazley, L., Kenchington, E., Murillo, F.J., Lirette, C., Guijarro, J., McMillan, A. and Knudby, A., 2016. Species distribution modelling of corals and sponges in the maritimes region for use in the identification of Significant Benthic Areas. Canadian Technical Report of Fisheries and Aquatic Sciences 3171.

Cheung, W.W., Lam, V.W., Sarmiento, J.L., Kearney, K., Watson, R.E.G., Zeller, D. and Pauly, D., 2010. Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. Global Change Biology, 16(1), pp.24-35.

Elith, J., & Leathwick, J., 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Embling, C. B., Gillibrand, P. A., Gordon, J., Shrimpton, J., Stevick, P. T., & Hammond, P. S., 2009. Using habitat models to identify suitable sites for marine protected areas for harbour porpoises (*Phocoena phocoena*). Biological Conservation, 143(2), 267–279. https://doi.org/10.1016/j.biocon.2009.09.005

Hawkins, D. M., Basak, S. C., & Mills, D., 2003. Assessing model fit by cross-validation. Journal of Chemical Information and Computer Sciences, 43(2), 579–586. https://doi.org/10.1021/ci025626i

Le Pape, O., Delavenne, J., & Vaz, S., 2014. Quantitative mapping of fish habitat: A useful tool to design spatialised management measures and marine protected area with fishery objectives. Ocean and Coastal Management, 87, 8–19. https://doi.org/10.1016/j.ocecoaman.2013.10.018

Rengstorf, A. M., Yesson, C., Brown, C., & 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), 1702–1714. https://doi.org/10.1111/jbi.12123

Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40:913-929.

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 Management of Vulnerable Marine Ecosystems on the Louisville Seamount Chain, South Pacific Ocean. Frontiers in Marine Science, 4(October). https://doi.org/10.3389/fmars.2017.00335

Rooper, C.N., Hoff, G.R., Stevenson, D.E., Orr, J.W. and Spies, I.B., 2019. Skate egg nursery habitat in the eastern Bering Sea: a predictive model. Marine Ecology Progress Series, 609, pp.163-178.

APPENDIX B: AGENDA

Habitat Suitability Modelling Best Practices for Canada's Pacific Ocean

June 11-12, 2019
Nanaimo, BC

Chair: Steve Schut

DAY 1 – Tuesday, June 11

Time	Subject	Presenter
0900	Introductions Review Agenda & Housekeeping CSAS Overview and Procedures	Chair
0915	Review Terms of Reference	Chair
0930	Presentation of Working Paper – HSM Framework	Authors
1015	***Break***	
1045	Overview Written Reviews – HSM Framework	Chair + Reviewers & Authors
1130	Identification of Key Issues for Group Discussion - HSM framework	Group
12:00	***Lunch Break***	
1300	Presentation of Working Paper – Application of Framework	Authors
1330	Overview Written Reviews – Application of Framework	Chair + Reviewers & Authors
1430	***Break***	
14:45	Identification of Key Issues for Group Discussion - Application of framework	RPR Participants
1515	Discussion & Resolution of Results & Conclusions	RPR Participants
1700	**Adjourn for the Day**	

Day 2 - Wednesday, June 12

Time	Subject	Presenter
0900	Introductions Review Agenda & Housekeeping Review Status of Day 1 (*As Necessary*)	Chair
0915	Carry forward outstanding issues from Day 1	RPR Participants
1030	***Break***	
1045	*Science Advisory Report (SAR)* Develop consensus on the following for inclusion: • Summary bullets • Sources of Uncertainty • Results & Conclusions • Figures/Tables • Additional advice to Management (as warranted)	RPR Participants
1200	***Lunch Break***	
1300	*Science Advisory Report (SAR) cont'd*	RPR Participants
1445	***Break***	
1500	Next Steps – Chair to review • SAR review/approval process and timelines • Research Document & Proceedings timelines • Other follow-up or commitments (*as necessary)*	Chair
1545	Other Business arising from the review	Chair & Participants
1600	***Adjourn meeting***	

APPENDIX C: PARTICIPANT LIST

Last Name	First Name	Affiliation
Anderson	Sean	DFO Science, Stock Assessment
Beazley	Lindsay	DFO Science, Maritimes
Benoy	Nicholas	DFO Ecosystem Management, Oceans
Campbell	Jill	DFO Science
Candy	John	DFO Science, Centre for Science Advice
Chiang	Eric	DFO Oceans
Christensen	Lisa	DFO Science, Centre for Science Advice
Curtis	Janelle	DFO Science, Ecosystems Sciences
Davies	Sarah	DFO Science, Ecosystems Sciences
Dudas	Sarah	DFO Science, Ecosystems Sciences
English	Philina	DFO Science, Stock Assessment
Ferguson	Kiyomi	DFO Science
Fields	Cole	DFO Science, Ecosystems Sciences
Finney	Jessica	DFO Science, Ecosystems Sciences
Gale	Katie	DFO Science, Ecosystems Sciences
Gomez	Catalina	DFO Science, Newfoundland
Goulet	Pierre	DFO Science, Newfoundland Region
Gregr	Ed	University of British Columbia
Gullage	Lauren	DFO Science, Newfoundland Region
Herborg	Matthias	DFO Science, Ecosystems Sciences
Hubley	Brad	DFO Science, Maritimes Region
Kenchington	Ellen	DFO Science, Maritimes
Knudby	Anders	University of Ottawa
Lessard	Joanne	DFO Science, Ecosystems Sciences
Leus	Dan	DFO Science, Ecosystems Sciences
Murillo-Perez	Javier	DFO Science, Maritimes
Nephin	Jessica	DFO Science, Ecosystems Sciences
Novaczek	Emilie	DFO Science, Newfoundland Region
O	Miriam	DFO Science, Ecosystems Sciences
Pretty	Christina	DFO Science, Newfoundland
Robinson	Cliff	DFO Science, Ecosystems Sciences
Rooper	Chris	DFO Science, Ecosystems Sciences
Rubidge	Emily	DFO Science, Ecosystems Sciences
Sameoto	Jessica	DFO Science, Maritimes
Schut	Steve	DFO Science, Fishery & Assessment Data
St. Germain	Candice	DFO Science, Oceans Science Division
Warren	Margaret	DFO Science, Newfoundland Region
Wells	Nadine	DFO Science, Newfoundland
Yakgujaanas	Jaasaljuus	Council of Haida Nations

APPENDIX D: WORKING PAPER REVIEWS

REVIEWER: CHRIS ROOPER, DFO SCIENCE, ECOSYSTEMS SCIENCES

Thank you for the opportunity to participate in this review. I really admire and enjoyed the work on the CSAS paper, the authors should be commended for their incredible efforts on putting together this framework.

Below are comments that address the specific questions of this review, followed by some specific comments that should be addressed/justified. I have also included more specific comments and some minor grammatical suggestions in the text using track changes.

1. *Provide background methods, challenges and current best practices for HSM development.*

   The field of habitat/species distribution modeling is large and there have been a fair number of reviews and detailed discussions of methodologies, challenges and best practices. Many of these, such as which test statistics to use when evaluating models have been going on for many years with no resolution. The authors did a nice job of providing background on these topics and referred to the relevant papers for those that are interested in the details of these types of discussions. In general, the authors did a nice job of identifying points where they made choices and justifying those choices by bringing forward relevant literature or examples. The treatment of modeling types other than those considered in this framework was relatively light, but understandable given the space needed to do this topic justice and the fact that it was a little outside the focus of this paper.

2. *Assess the challenges associated with finding appropriate species and environmental data,and provide guidance on best practices for preparing data for HSMs.*

   The authors provide a good overview of the data challenges and examples using the species examined in this framework. There are a few issues that I have highlighted for further explanation in the text comments. Overall, I think they lay out a pathway for data compilation that can be followed, especially relevant to the coast of BC.

3. *Develop a framework that applies best practices to HSM development; including how to validate models and present model assumptions and uncertainty.*

   I really liked the flow charts that provided the pathway to HSM development. Those were very useful and provided an easy to follow framework. It was easy to see places where new data/methods could be inserted into the flow chart. I also really like the figures used for showing their analyses of model evaluation using threshold dependent and independent data. Those were very useful and clear. The CV method of model validation was presented very well including the explanation of the block-CV method. Where choices were made (such as using SD v. coefficient of variation to show uncertainty) these were well explained and well thought out decisions. There are certainly other ways to address some of the validation exercises, but highlighting the need for independent data to validate models is consistent with best practices.

4. One comment is that although the block CV method helps with reducing independence, it does not eliminate it. As an example, the northern abalone model was based largely on index sites that were chosen initially as areas of suitable habitat. Thus, there is some dependence/bias built into the survey design. The best way to validate models is to actually independently test their predictions by collecting additional data. This is rarely done, but is the “gold” standard. In the case of the abalone in particular they had data from a number of survey types, it would have been relatively easy to use data from one survey to

parameterize the model and then validate with another survey. The independence would have been further reduced in this process.

5. *Demonstrate an application of the HSM framework using a case study in Pacific Regions.*

   The application of the framework to Pacific Region species was very well done. I have some comments on additional presentation and other parts inside the word document.

6. *Examine findings relevant to building HSMs and provide recommendations for future applications.*

   The conclusions section needs a bit of strengthening/expanding (see comment below), but was generally good. The recommendations followed very well from the data and analyses presented. There were a couple of inconsistencies between the recommendations and the choices that were made in the analyses (thinking in particular of the trade-offs of Dungeness crab data, where a few different sources of somewhat poor comparability were used and the rockfish were two reasonably good sources of data were available and only one was used). A tightening of this standard with some explanation in the Dungeness section would be helpful.

There was no explicit recommendation on what types of models to include in the ensembles. The authors used any model that had an AUC over 0.5 (correct?). Is that their recommendation here in the framework?

The one thing that was missing was a recommendation on doing independent validation of the models by collecting new data. I would like to see this included, even though it was not necessarily applied here.

Major comments

Section 1

In section 1.1. An overview of habitat suitability modeling is provided that is confusing when read in the context of this framework. HSM is defined as determining the fundamental niche of species in the absence of barriers and biotic interactions, as opposed to species distribution models that incorporate these processes. It appears this discussion applies mostly to the HSI methods, but not the GLM and BRT methods. However, the data and relationships used to generate all the models is based on either relationships gleaned mostly from field observations (i.e. expert opinion) or from the data themselves. Both of these types of models are at least partially models of the realized niche. It would probably help to define exactly what a HSM is in the context of this framework and then explain how it is defining a fundamental niche. Alternatively (probably my preference) would be to discuss niche theory, define what HSM means for this framework and delete the discussion of how HSM as you've defined it fits into niche theory. The important part that you are trying to communicate is in the next two paragraphs.

If you choose to combine gear/survey types and either the spatial extent or the catchability differ between gear types, this has to be accounted for (or tested for) in the model. A spatial bias can occur in the former case and a non-random bias will occur in the latter case. This has to be accounted for in the modeling or by scaling prevalence predictions. Or, as you have done with rockfish, throw one data set out.

HSM and species distribution are used interchangeably in the first paragraph of the introduction which might be part of the issue. A good clear definition of what you are talking about when you refer to HSM is needed. Based on some of the framework discussion in the methods on scaling temporal predictors, etc. I might be tempted to state explicitly and up front that your interest in this framework is to make static maps of habitat.

Related to this issue, for the GLM/BRT models the manuscript uses HSM in a very specific way, probability that an organism has been found at a location at least once. Since your presences can be 0's in combination with 1's and your absences are never having been found here ever. This is a bit different than the typical presence/absence modeling or even HSI as you are using it. This should be specified somewhere and also the implications of this definition discussed. The observational data from the HSI is also different (there is a lot of mixing of density into the definitions of these functions). Meaning the models are predicting different things, this leads to the question of whether they should be ensemble at all?

Section 1. I would consider doing a bit of reorganizing and rethinking this introduction. The main recommendations that I think you make here are 1) consider the management context of the model, what questions needs to be answered, 2) consider the objective of the models, what product do you need to answer the questions, 3) consider the species characteristics and its ecological relationships, 4) consider the kinds of data you need to address these relationships, and 4) consider the type of model appropriate for the data quality/quantity in the context of the product to be produced. These concepts are a bit mixed in the current structure and although some important points are made (like the ecology and management are linked), it would be better in my opinion to have a pathway that leads the reader to the next sections where you discuss in more detail so that the framework is all organized with the same flow.

Section 3.1

Why did you choose to use any presence as an indication of presence, instead of using the entire data set? From a statistical standpoint I don't immediately see the reasoning other than maybe your point densities were too high (spatial bias) for some surveys? If so, there are other thinning procedures that can be used. Also, how many times did this happen (what was the n for the original and n for the reduced data sets)?

Section 4.4.

This conclusions section focusses much of its attention on the HSI methods and why they should be retained. There should be other important conclusions that are added here.

Section 5.

Validation by independent data is not addressed.

Did you test ensemble model performance without the HSI models included? One of the conclusions was that the increase in AUC from the ensemble model was in part a result of including the HSI model that inflected some biological realism on the statistical models. This may be true, but it wasn't actually shown in the comparisons. To support this point, it would be useful to calculate the ensemble with just the GLM and BRT and compare that to the 3-model version.

Related to this, in the absence of good data, the HSI is very useful, but when there is reasonable data, it would seem better practice to let the data guide the relationships and then look at them in the context of the HSI/literature/expert opinion rather than recommending including a model with poor performance to the ensemble. This may be an unpopular opinion, but if you're going to make the recommendation to always use HSI in this framework, you should demonstrate that the performance of the ensembles is just as good or better when you include this model as when it is eliminated. Using the HSI in data poor situations (like the mussels) seems more appropriate than the data rich species (like the rockfishes).

Appendix B

In appendix B, it would be nice to see 6 panel figures of the maps with each model as a panel, the data on a map as a panel, the ensemble and the ensemble SD maps. As it stands, the only model output we see is from the Ensemble and HSI, and we never see the actual data. Best practices probably demand showing these things. The maps with limiting factors for the HSI can probably be removed.

Related to that with the space saved, I would like to see the model shapes from the various models plotted next to each other. One of the points of the paper is that robust models should exhibit the same shapes over model types, but it would be nice to see this in the appendix so the reader can judge the model fits on their merits and ecological realism.

REVIEWER: JAVIER MURILLO-PEREZ, DFO SCIENCE, MARITIMES

One working paper has been prepared for review under the CSAS meeting Habitat Suitability Modelling Best Practices for Canada’s Pacific Ocean. This document presents a framework for the development and application of Habitat Suitability Models on Canada’s Pacific Coast. Data preparation, model development, model evaluation, and the interpretation of results and uncertainties are discussed in order to standardize and facilitate the model building process. This framework has been implemented in the R statistical programming language.

The authors are to be congratulated on the quality and quantity of the analyses that are presented in the document. The approach includes most of the recent recommendations on habitat suitability modelling and each step of the modelling process, from data preparation to the interpretation of results and uncertainties are discussed. The data and methods are well explained and support the main conclusions.

Most comments and questions that I provide here are minor and reflect potential directions or alternative models that could be used in future revisions of the methodology.

Major points for consideration:

1. MODEL TYPE SELECTION.

   The authors have selected three model types with different degree of complexity that are used in the final ensemble prediction: expert-derived habitat suitability models (HSIs), generalized linear model (GLM), and boosted regression trees (BRT). The use of HSI is discussed and justified in the conclusions and BRT are commonly used in recent applications. However, although it is said that the framework is suitable also for generalized additive models (GAM), there is not discussion about the preference of GLMs over GAMs. The use of GAMs, which can include non-linear relationship between the response and predictors, could deal with some of the problems found in the GLM, like the Temperature relationship on the Dungeness Crab.

2. PREDICTOR DATA (page 6).

   Maximum mixed layer depth is another predictor estimated by ocean circulation models which can have a significant influence on primary production in the surface waters (Polovina et al. 1995, Carstensen et al. 2002). Additionally, fishing effort can also be an important predictor to understand some of the present distribution on impacted ecosystems (e.g. Foster et al. 2015, Tien et al. 2017). And in some cases, biotic variables can also be informative of the distribution of other organisms when there is a strong relationship between them. These predictors could also be discussed.

3. SPATIAL PREDICTORS (page 8).

   During last years, it has been a great advance in spatial modelling using Integrated Nested Laplace Approximations (INLA) approach (Rue et al. 2009). The use of an INLA approach allows to include the spatial structure in the parameter estimates and predictions (e.g. Boudreau et al. 2017, Penino et al. 2019). For a review see Bakka et al. (2018).

4. DATA PREPARATION (page 12).

   For observations represented as lines, it is said that mean values are calculated from the raster cells that intersect the line segments. Whereas, for categorical predictor variables, the most dominant category occurring in the intersected raster cells is used. However, in some cases it may be more informative to calculate the percentage of overlap between the categories and the lines. For example, if one transect covers mostly sandy sediment but passes over a rocky patch suitable for a particular organism, using the dominant information would not allow to catch that feature, but using a percentage of overlap (e.g. 80% sand, 20% rock) would add some information that could be useful for the model.

   The use of Spearman's Rank correlation coefficient (or Pearson) in addition to the Variance inflation factor (VIF) could add some information to the correlation between predictors. Additionally, a second threshold based on the Spearman coefficient (>0.8 or 0.9) could be added to the VIF threshold to help the decision in the removal of predictors. This could also add some information on the degree of correlation between Temperature and Salinity (page 35).

5. CROSS-VALIDATION (page 12).

   Model validation usually utilizes the same data type than those used to build the model. However, when new data obtained from a different gear are used to validate the models the results can be quite different (e.g. Rooper et al. 2019). This adds uncertainty on the predictions that could be discussed along the report and perhaps a recommendation to utilize different data type (when they are available) to validate the models could be made.

6. MODEL EVALUATION (page 15).

   More values of AUC and TSS could be added to show the accuracy of the model prediction between 0.5 and 1. According to Pearce and Ferrier (2000) and Jones et al. (2010) values of AUC > 0.9 are considered good, between 0.7 and 0.9 moderate, and < 0.7 poor. Whereas, values of TSS > 0.6 are considered good, between 0.2 and 0.6 fair to moderate, and <0.2 poor (Landis and Koch 1977, Jones et al. 2010).

7. PREDICTIONS AND UNCERTAINTY (Figure 3.7, page 18).

   Other option to delineate areas of extrapolation that has been used in the Maritimes Region is the addition of a layer indicating those areas but with some transparency to still see the predicted values underneath. This allows to see the extrapolated areas in the prediction surfaces but adding a high degree of uncertainty (see figure 1 for an example). In this case, the extrapolated areas are not very informative because a random forest model was used. But in the case of a GLM, GAM or ensemble model the extrapolated areas could provide some information and add some insights for future research.

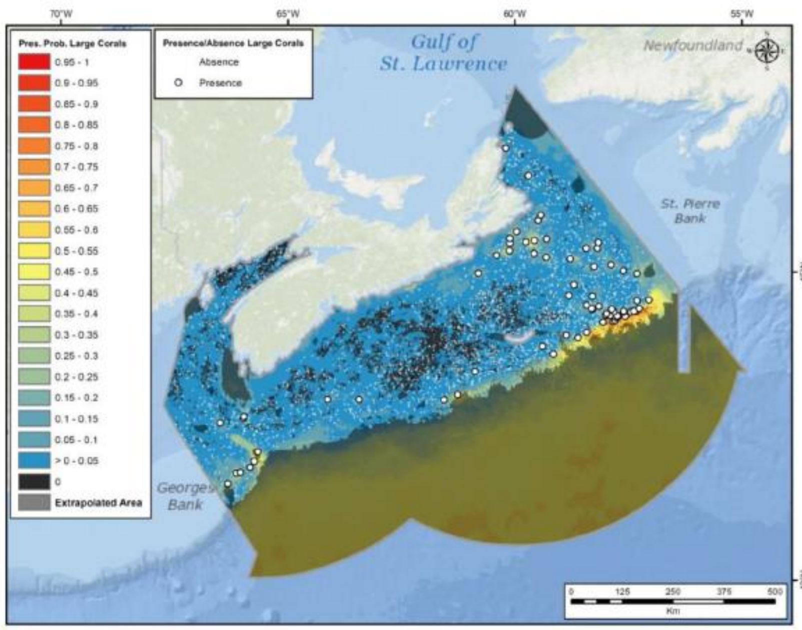

Figure 1. Areas of extrapolation of the random forest model on unbalanced presence and absence large gorgonian catch data collected from the Maritimes Region between 2002 and 2015. Also shown are the large gorgonian presence and absence observations and predictions of presence probability (Beazley et al. 2017).

8. ENVIRONMENTAL PREDICTORS. Oceanographic layers (page 29).

   Oceanographic conditions can undergo strong seasonal variability in this region, mostly for coastal waters. It seems that only the data from spring/summer season was used to calculate the Temperature, Salinity and current speeds. Is there any reason why the fall/winter data were not included? Some clarification or more explanation would help.

9. SPECIES OBSERVATIONS (page 33).

   Four of the twelve species selected were not modelled because they did not have adequate species occurrence data due to low sample sizes, low precision in the spatial location of observations or low taxonomic resolution. However, in some cases it would be interesting to see the results of these models and discuss if the expected lower performance is caused by any of these reasons. For example, in Eastern Canada models for sponges in general performed worse than other benthic groups likely to the low taxonomic resolution considered. But it could be the case that some taxa at high taxonomic resolution (Order Pennatulacea) may have similar environmental affinities. This could add some information to the next step mentioned in the conclusions (page 49) to assess the sample size at which the performance of these models begins to degrade. Additionally, it could provide an example of how to quality of species occurrence data influences model performance statists (4th recommendation).

10. MODEL EVALUATION (page 15) AND COMPARISON OF PERFORMANCE STATISTICS (page 48)

    This framework estimates an optimal threshold which will depend on the purpose for modelling. What is the threshold optimization that has been used along the report? A threshold which maximizes TSS? Prevalence? Some clarification here would be helpful.

Future considerations:

Abundance data

This framework is focused on presence-absence data and estimates of abundance are not considered. However, if the models want to be used for stock assessment, or harvest plans, a framework considering abundance data would be beneficial as it is said in the conclusions (page 49). Although a similar framework could be applied for abundance data, some extra information should be added, mostly related to the model evaluation statistics, in order to use this framework for abundance data in the future.

Community-level models

Twelve species have been selected from a larger group. However, in the future it may be interesting to model all these species trying to include interactions between them. Recent advances in community modelling techniques allow to model community data to look for groups of species sharing common responses to the environment, such as Species Archetype Modelling (Dunstan et al. 2011, Murillo et al. 2018) or the use of joint species distribution models (e.g. Warton et al., 2015; Ovaskainen et al., 2017) that can consider interactions between species and allow the incorporation of species traits into a modelling framework useful for identifying response-traits that can provide functional, mechanistic and predictive perspectives on processes shaping the assembly and dynamics of ecological communities. These models are probably out of the scope of this framework but something to consider in the future when more species are included.

General comments:

Inclusion of input data points (presence and absence) in prediction surfaces could add some information about the data distribution. It may be more complicate to visualize for the shelf study area but something to consider.

Once the results are final and the framework is approved it would be useful to make the R scripts available.

Additional editorial comments were provided to the authors separately.

References

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler. A., Bolin, D., Illian, J., Krainski, E., Simpson, D., Lindgran, F. 2018. Spatial modelling with R-INLA: A review. WIREs Computational Statistics 10, https://doi.org/10.1002/wics.1443.

Beazley, L., Kenchington, E., Murillo, F.J., Lirette, C., Guijarro, J., McMillan, A., and Knudby, A. 2016. Species Distribution Modelling of Corals and Sponges in the Maritimes Region for Use in the Identification of Significant Benthic Areas. Can. Tech. Rep. Fish. Aquat. Sci. 3172: vi + 189p.

Boudreau, S. A., Shackell, N. L., Carson, S., and den Heyer, C. E. 2017. Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the northwest Atlantic ocean. Ecology and evolution, 7(22):9739{9749.

Carstensen, J., Conley, D.J., Lophaven, S., Danielsson, A., Rahm, L., and Toompuu, A. 2002. Statistical analysis and modelling of phytoplankton dynamics: exploitation of data in the Nordic and Baltic Monitoring Programs, Nordic Council of Ministers.

Dunstan PK, Foster SD, Darnell R (2011) Model based grouping of species across environmental gradients. Ecol Modell 222:955–963

Foster, S.D., Dunstan, P.K., Althaus, F., Williams, A. 2015. The cumulative effect of trawl fishing on a multispecies fish assemblage in south-eastern Australia. J. Appl. Ecol., 52:129-139.

Jones, C.C., Acker, S.A., Halpern, C.B. 2010. Combining local- and large-scale models to predict the distributions of invasive plant species. Ecol. Appl., 20: 311-326.

Landis, J.R., and Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics 33: 159- 174.

Murillo, F.J., Kenchington, E., Tompkins, G., Beazley, L., Baker, E., Knudby, A., Walkusz, W. 2018. Sponge assemblages and predicted archetypes in the eastern Canadian Arctic. Mar. Ecol. Prog. Ser., 597: 115-135.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., Abrego, N. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol. Lett., 20:561–576.

Pearce, J., and Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model., 133: 225-245.

Pennino, M.G., Guijarro-Garcia, E., Vilela, R. del Rio, J.L., Bellido, J.M. 2019. Modeling the distribution of thorny skate (*Amblyraja radiata*) in the Southern Grand Banks (Newfoundland, Canada). Can. J. Fish. Aquat. Sci., https://doi.org/10.1139/cjfas-2018-0302

Polovina, J.J., Mitchum, G.T., and Evans, G.T. 1995. Decadal and basin-scale variation in mixed layer depth and the impact on biological production in the Central and North Pacific, 1960-88. Deep-Sea Res. 42: 1701-1716.

Rooper, C., Wilborn, R., Goddard, P. 2019. Determining the best methods for model validation. In: Use of species distribution modeling in the deep sea (Kenchington et al., eds.). Can. Tech. Rep. Fish. Aquat. Sci. 3296. pp.24-26.

Rue, H., Martino, S., Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. B, 71: 319-392.

Tien, N.S.H., Craeymeersch, J., van Damme, C., Couperus, A.S., Adema, J., Tulp, I. 2017. Burrow distribution of three sandeel species relates to beam trawl fishing, sediment composition and water velocity, in Dutch coastal waters. J. Sea Res., 127: 194-202.

Warton, D.I., Blanchet, F.G., O-Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., Hui, F.K.C. 2015. So many variables: joint modeling in community ecology. Trends Ecol. Evol., 30:766–779.

REVIEWER: ANDERS KNUDBY, UNIVERSITY OF OTTAWA

For this review I let myself be guided by the suggested questions to consider, and have provided responses related to each one below in bold.

1. Is the purpose of the working paper clearly stated?

   Yes, the subject is clearly outlined, and the purpose of the paper clearly stated (and meaningful).

2. Are the data and methods adequate to support the conclusions?

   Yes, but I do have a few suggestions for things that could/should be tested. These are all outlined in my comments below.

3. Are the data and methods explained in sufficient detail to properly evaluate the conclusions?

Yes, the paper is very detailed and clear.

4. If the document presents advice to decision-makers, are the recommendations provided in a useable form, and does the advice reflect the uncertainty in the data, analysis or process?

   I don't think the paper provides advice to decision-makers as such, more advice to scientists that may in turn provide advice to decision-makers.

5. Can you suggest additional areas of research that are needed to improve our assessment abilities?

I have some suggestions for consideration – none of which are crucial, but some of which might improve the paper.

- From the figure on page 13, it seems that the blocking method does not maximally reduce interspersion of training and test CV folds. The principal weakness is the random assignment of blocks to folds, which seems like only a small step away from random assignment of points to folds. I would suggest at least testing a clustering method instead.
- On page 46, and 51, you mention that this approach 'removes the spatial dependence between the training and testing data sets', but that's not correct. It reduces it somewhat, but you still have blocks of test data surrounded on all sides by blocks of training data.
- For a given prevalence, Accuracy and Kappa will always be linearly related, so it is not necessary to use them both. I think the reason you get some variation between them is because your prevalence changes between different CV folds (page 15-16).
- As Jessica already knows, I am working on some code to assess whether a given uncertainty metrics (e.g. like the one outlined on pages 17-18) is actually related to the accuracy of predictions. The lack of evidence for this, despite what intuition suggests, might be worth noting.
- Page 21-22: When you mentioned variable importance and methods to calculate it, you should at least mention:
  - The permutation-based approach.
  - That the drop-one method will produce very low importance for predictors strongly correlated with other predictors (even though this will have been taken care of, to some extent, but the variable removal earlier in the process).
- Would it not be a natural extension of the SAC approach to run a kriging interpolation on the auto-covariate predictor, to enable its extension to the entire model extent? Alternatively do the same but for the initial model residuals (a method commonly called regression kriging).
- I think a weakness of the paper is that you i) describe in the framework how e.g. the ecological model forms an important foundation for the data model, but then proceed to ii) construct a single data model for species that presumably have widely varying ecology. This is not unusual at all because most HSMs are created without careful consideration of the characteristics of the species, predictors, and the scales of the imagined ecological relationships, but it seems contrary to your stated objective of following best practices. Examples:
  - You mention that the substrate was mapped in a neighbourhood rather than per cell because it is assumed better for mobile species. But you then also use that substrate layer for non-mobile species.

- You do not vary spatial resolution for any of the predictors to reflect the scale at which each species may “interact with” a given predictor.
- You do not tailor any predictor to be specifically relevant to the time period in question (e.g. the oceanographic layers are quantified for the spring/summer season identically for all species).

I actually think what you are doing is correct, because, as you mention on page 49, ‘The use of a single, suite of predictor data (i.e., a consistent data model) was central to the ability of the framework to handle multiple habitat models and species in an efficient way.’ It’s just that you had previously mentioned how these species-specific characteristics should be taken into account when developed the data and models. But then you also write (page 50): ‘Building an HSI model, regardless of available observation data, is helpful to the modelling process on two fronts. First, it contributes to the ecological model, clarifying the current understand the ecology of the species.’ But then you don’t actually go on to really use this ‘ecological model’ for anything. In fact it is not entirely clear to me at this point what this ‘ecological model’ even is.

- For chlorophyll, you should keep in mind that the chlorophyll-estimation algorithms themselves perform very poorly near land, so it would probably be better to simply exclude chlorophyll data for pixels that contain any amount of land.
- Was the frequency distribution of wind directions and wind speeds included in the fetch calculation? That seems to be a relevant thing to include, but it was not mentioned in the paper.
- I’m curious about the substrate predictor being itself predicted from other predictors. It seems unnecessary, as at least in the highly fitting BRT models this kind of approach should be unnecessary (although it may lead to more interpretable models).
- You should include mention of who developed the HSI models, and what knowledge they had of the occurrence data at the time they did so. It is virtually impossible to develop HSI models without being biased by the data, if one is aware of what the data distribution looks like.
- Page 42: You have negative depths in figure 4.4
- Page 43: I’m not sure if I have misunderstood something, but did you include all occurrence data in your development of the auto-covariate predictor?
    - If yes, then it is somewhat unfair to compare models with and without this term, because the ‘test’ data have then been used to develop one of the predictors for the models including the auto-covariate term.
    - If no, then how did the term contribute to making predictions for the ‘test’ data?
- You mention that the HSI provides a moderating effect because it is free of data-driven bias. I understand the logic here, but did you test whether an ensemble of just your GLM and BRT models performed better or worse than your HSI+GLM+BRT ensemble, when evaluated against the test data? I am a firm believer that the proof is in the pudding, so unless the HSI contributes to improving some test statistic, then I don’t see its value.