Methodological Applications

The importance of temporal resolution for niche modelling in dynamic marine environments

Fernandez, M[1], Yesson, C[2], Gannier, A[3], Miller, PI[4], Azevedo, JMN[1]

1. cE3c - Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group, and Faculdade de Ciências e Tecnologia, Universidade dos Açores, Rua da Mãe de Deus, 13A, 9501-801. Ponta Delgada, Azores, Portugal.
   Email: marc.f.morron@uac.pt

2. Institute of Zoology, Zoological Society of London, UK.

3. Groupe de Recherche sur les Cétacés, Antibes, France.

4. Remote Sensing Group, Plymouth Marine Laboratory, UK.

Keywords: species distribution modelling; temporal scales; highly mobile species; marine environment; cetaceans; dynamic environment

Word count abstract: 264

Word count main text: 7023

Running head: Temporal resolution niche models marine habitat

ABSTRACT

Aim

Highly dynamic ocean environments can experience dramatic changes over relatively short timeframes, affecting the spatial distribution of resources and therefore the presence or absence of highly mobile species. We use simulation studies to investigate how different temporal resolutions might affect the results of species distribution models for highly mobile species (e.g. cetaceans) in marine environments.

Location

Azores archipelago, Portugal

Methods

We developed 3 virtual species with different habitat preferences influenced by (i) only static (topographic), (ii) only dynamic (oceanographic), and (iii) both dynamic and static variables. Assuming that species would reposition themselves daily according to these preferences (as has been observed for large marine foragers such as cetaceans), we used two different approaches (generalized linear model and generalized boosted model) to test the effect of using daily, weekly and monthly environmental datasets to model distributions.

Results

The results showed that the selection of different temporal scales has a very important effect on model predictions. When dynamic variables are important components of habitat preference, models based on daily or weekly timeframes performed best at reconstructing the known niche.

Main conclusion

It is important that we consider temporal resolution when applying species distribution models. Several factors (e.g. species ecology and oceanographic characteristics of the ecosystem) should be taken into consideration when selecting an adequate temporal scale for niche modelling. For fine scale applications (e.g. dynamic ocean management), highly dynamic ecosystems, and highly mobile species, our results suggest exploring temporal resolution of 7-8 days rather than coarser temporal scales. For some applications annual, seasonal or even monthly averages may produce inferior or inaccurate models.

Author contributions: M.F. conceived the ideas; M.F., P.M. and C.Y. provided and analysed data; all authors contributed to the writing and revision processes.

1. INTRODUCTION

It is important to understand the factors that influence species distributions within the application of environmental niche models (Fryxel et al., 2014). Species distribution models (SDMs) are useful tools and can have many applications including informing management and conservation decisions (Hirzel et al., 2001; Peterson et al., 2011). Their widespread use has led to useful discussions regarding their utility and accuracy (Brotons et al., 2004; Lawson et al., 2014). One important factor in the construction of SDMs is the resolution of environmental variables. Typically, the temporal and spatial resolutions of analyses are determined by the availability of environmental data, rather than by a considered assessment of species' characteristics (Barry & Elith, 2006; Jetz et al., 2012). Some studies have investigated how different spatial resolutions affect modelling results (Guisan et al., 2007; Svensson et al., 2013). However, the temporal resolution of environmental variables has received far less attention (e.g. Araújo et al., 2005) and in some cases those studies have focused on scales of centuries to millennia. Many variables used for SDM show significant variation over a variety of timescales. In the natural world, we see potentially significant temperature variations over years, seasons, months, weeks and even days, and these may be important for determining or limiting species distribution. It is common practice for SDM studies to incorporate seasonal variations (e.g. bioclimatical Bioclim/WorldClim variables), but less common to examine variability over larger (multi-year) or shorter (monthly/weekly) periods.

When working with very dynamic environments (e.g. small-scale oceanic frontal areas) and/or with species with high mobility (e.g. cetaceans or sharks), short term temporal variation could be an important issue, as ephemeral environmental conditions may determine distribution over short time- frames. When working with top predators, the oceanographic dynamic variables will not affect the species distributions directly, but could be used as distal variables (Austin 2002), e.g. a proxy of prey density. For example biophysical coupling at frontal areas can lead to the formation of pelagic foraging hotspots (Scales et al. 2014), creating aggregations zones for zooplankton advected from surrounding water masses driving bottom-up processes across multiple trophic levels up to apex predators (Bakun 2006).

Typically, incorporating temporal dynamics of the environment does not extend beyond the inclusion of seasonal or monthly climatological variables, e.g. data from Bioclim (Busby 1991) for terrestrial studies or from MARSPEC (Sbrocco & Barber 2013) for marine studies. Some SDM studies based in the marine environment have used annual or seasonal averages (Cañadas & Hammond 2008; Praca & Gannier 2007), while others have employed monthly averages (MacLeod et al., 2007, Moura et al., 2012, Panigada et al., 2008), and a notable few have considered weekly means (Becker et al., 2010; Becker et al., 2016; Howell et al., 2008; Mannocci et al., 2014; Roberts et al., 2016). Few studies have investigated the adequacy of temporal resolution of environmental data, such as Forney et al. (2012) and Scales et al. (2017).

Mannocci et al. (2014), grouping species at broad scales, found improved model results when using seasonal oceanographic data, leading the authors to suggest that there was no apparent short-term reaction of top predators towards oceanographic variability. Conversely, Scales et al. (2017) found that models fitted using seasonal or climatological data fields can introduce bias in presence-availability models. Biologically relevant time scales can vary from thousands of years to minutes, depending on oceanographic processes (Mann & Lazier, 2013). Therefore, it is important we gain a better understanding of how different temporal scales might affect SDMs in the marine realm.

The aim of this study is to investigate the effect of using different temporal resolutions in developing SDMs for highly mobile species in dynamic environments.

2. MATERIALS AND METHODS

2.1. Study area

The study area is located in the Azores archipelago, a group of North Atlantic oceanic islands located approximately 1,800 km west of Lisbon. The region is strongly influenced by the Gulf Stream and all the branches of this current. Its large-scale oceanic circulation is dominated by the Azores Current, which generates considerable mesoscale variability (Santos et al. 1995).

2.2. Environmental data

A set of real marine environmental variables was selected to represent the variability and dynamism of an oceanic system. Variables were chosen based on their reported influence on cetacean distributions (see Appendix S1). These were divided into two thematic groups: static (little or no short term variation – i.e. topographic variables) and dynamic (rapidly changing variable, such as temperature) (Table 1).

Four static variables were derived from the digital elevation model (DEM) of the EMODnet Bathymetry portal: depth (the DEM); slope and curvature, calculated using DEM Surface Tools for ArcGIS 10.2; distance to the nearest 200 m bathymetric line, calculated using QGis 2.12. Curvature was used as a proxy of sea bottom roughness, providing an estimate of sea floor relief, which can influence some cetacean species (Lindsay et al., 2016). All static variables were calculated at a spatial resolution of 0.5 x 0.5 km.

Daily dynamic variables were derived from NASA's Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (SST) dataset, which merges many infrared and passive microwave datasets, gathered from satellites, into global daily maps at 1 km resolution. Thermal ocean fronts were detected from each MUR SST daily map (Miller, 2009) and used to generate daily ocean front metrics. Front distance (Fdist) quantifies the distance to the closest major front (Miller et al., 2015).

We calculated weekly and (approximately) monthly layers using the mean values of daily layers. All pairs of variables were tested for pairwise correlation;

the final set of selected variables all showed Pearson correlation under 0.75. All layers were rescaled to a 2x2 km grid using bilinear interpolation (Fig. 1).

2.3. Virtual species

Three virtual species were created, with different habitat preferences, based on varying responses to static and dynamic variables (Table 2; for full details of construction see Appendix S1). The 'Dynamic' species reacted only to dynamic variables. The 'Static' species was influenced solely by topographic parameters. The 'Pseudoreal' virtual species was influenced by both dynamic and static variables, with dynamic variables having twice the weight of static ones, so that only when dynamic characteristics were suitable (e.g. temperature) would the species prefer a specific static environment (e.g. depth).

Ecological niches were simulated in a multidimensional space following Hirzel et al. (2001). We defined the ecological niche of each virtual species as the weighted sum of its hypothetical response curves to three different sets of environmental variables. The ecological niche suitability can be expressed as $H_i \times W_i$, where H_i represented the virtual species' niche suitability index for an ith space and the W_i the weight of this suitability (Duan et al., 2014). Therefore the final suitability, H, was calculated as:

$$H = \frac{1}{\sum_{i=1}^{n} W_i} \sum_{i=1}^{n} W_i H_i + e \qquad \text{Eq. 1}$$

For each virtual species a suitability index (H) was calculated for all the areas sampled each day, using a threshold approach ($H \geqslant 0.6$) to select the suitable area for presences (for more details see Appendix S2).

The virtual species' responses to environmental variables were either linear or unimodal. The final species distribution was based on a weighted combination of responses to each variable (Figs. S1.1 to S1.3, see Appendix S1 for full details). Suitable areas for each species were projected onto a 2x2 km map of the study area (Figs. S2.1 to S2.3).

2.4. Sampling survey design

Environmental and effort data for the virtual species mirrored the timeframe of a simulated cetacean detection survey for the Azores archipelago. Surveys were restricted to the Central and Eastern island groups, covering 20,415 km^2. We modified Faustino et al (2010) tracks (Fig. 2; for more details of construction see Appendix S3) to last two months (8 weeks) per year, with 4 days of sampling per week over two years (July-August 2013 and 2014). Survey transects were mapped onto the 2 km grid that matched the environmental data.

As a complementary analysis a second survey was used to test for potential effects of survey design on the results. A non-linear survey design was used; see Appendix S3 for more details.

We simulated detections of the target species to infer presence (and absence) data in our models. We randomly selected 300 detection points (150 for training and 150 for testing) from the sampled suitable area over the entire sampling period.

This generated two datasets, each with 150 presences (or detections) with the other grids cells, noted as absences (details in Appendix S2). This mimics a real world encounter rate of c. 1.22 groups/100 km (Silva et al. 2014). We allowed the encounter rate to change through time; therefore it was related to the amount of suitable habitat per kilometre. With this design, for the species influenced by dynamic variables (Dynamic and Pseudoreal), the daily encounter rate will start low and steadily increase over the course of the season (as suitability areas are more available), with some days with rates much higher than 1.22 groups/100km. Consequently we assumed that these species were some sort of large-ranging seasonal migrant, not present at the beginning of the season and then disperses to the area. To simplify the analysis we assumed a perfect presence/absence detection scenario (all the groups encountered during the sampling were detected), although this is unusual for marine species (Katsanevakis et al., 2012).

This random selection was repeated 1,000 times for each species. Data were grouped according to three temporal aggregations. Niche estimates were calculated using two modelling approaches.

2.5. Temporal aggregations

Three temporal aggregations were created: daily, weekly (7-days) and monthly (4 weeks). For daily data, we constructed a data frame containing the environmental data, the sampling effort and the presence or absence of species for all the grid squares sampled each day. Using this approach a given location can be regarded as a presence one sampling day and an absence the next. For the weekly aggregation, data were grouped by 7-day periods, calculating mean values for each environmental variable and aggregating the sightings, recording a single presence on each grid square with a species observation, regardless of the number of times a species was recorded over the 7-day period. The monthly dataset involved the calculation of the average values of the environmental variables corresponding to the four months virtually sampled (with 16 days sampled per month). Presence grids were computed for each period, as in the weekly data.

There was almost no reduction of the number of presences with the coarsening of temporal resolution, due to the virtual sampling design. No reduction was found between the daily and the weekly approach, and a very small amount (less than 5 over 150 sightings) was found, for the monthly approach.

2.6. Modelling approaches

There are many SDM methods with variable accuracy and applicability, and notably performance may depend upon the characteristics of the target species (Quiao et al., 2015). Therefore, two modelling approaches with different theoretical bases were used: generalized linear models and boosted regression trees. The analyses were performed using the 'MASS', dismo, SDMTools, ecodist and gbm (Ridgeway et al., 2015) packages for R 3.2.2 (R Core Team ,2015).

Binomial generalized linear models (GLM) are used widely for predicting species distributions, and perform well when applied to the detection of the most influential environmental variables (Peterson et al., 2011). Models used both linear and quadratic terms for all explanatory variables to allow greater flexibility in fitting. Model selection utilised a stepwise (forward and backward) Akaike information criterion (AIC) procedure, obtaining the best explanatory variables for each case (James et al., 2013).

Boosted regression trees (BRT) or generalized boosted regression models (GBM) are a combination of classical statistics approach (regression trees) and a machine learning (ML) technique (boosting). The inclusion of ML adds considerable advantages compared to conventional methods, including the improvement of model selection (Elith et al., 2008). This approach examines a large number of trees and uses a boosting approach to select a linear combination of many trees (usually from hundreds to thousands). Fitted values in the final model are computed as the sum of all trees weighted by an estimate of the contribution of each tree to the growing model. A relatively slow learning rate (0.001) with a higher tree complexity (5), was selected to aim for more than 1,000 trees in the final model, avoiding a potential overfitting (Elith et al. 2008).

2.7. Model evaluation

SDM performance was evaluated using two metrics: (i) a variable contribution index; (ii) the area under the curve (AUC) of the receiver operator curve (ROC) for the training and test dataset. The analyses were performed using the SDMTools, ecodist, pROC and PMCMR packages for R.

GLM variable contributions were based on a tally of their inclusion in each stepwise selection procedure. Variable contribution for GBMs was estimated using the relative importance selection tool in the gbm R package.

Test AUCs were evaluated using daily environmental data. For each model the training AUC and test AUC were calculated (test dataset = 150 randomly selected presences from the sampled suitable area different from the training data).

To support the results, explanation suitability maps were produced for an extended area for randomly chosen dates for the three virtual species (Dynamic, Static and Pseudoreal) using the GLM approach. All analysis and figures were produced using R.

3. RESULTS

We built 6,000 ecological niche models (three temporal aggregations and two modelling algorithms) for each of the three virtual species, making a total of 18,000 models. For the two species influenced by dynamic variables, there were important differences in the evaluation metrics between the three temporal aggregations. In general, results improved when using the daily or weekly environmental layers. For species influenced solely by static variables, differences in accuracy between temporal aggregations were smaller.

3.1. Variable contributions

There were some differences among variable contributions by modelling method (Fig. 3), detailed below.

3.1.1. GLM

Temporal scale affected the models for the dynamic species; models based on daily and weekly aggregations successfully detected the two most important variables (SST and Fdist), but the monthly-based models were unable to detect the influence of Fdist.

For the Pseudoreal species, influenced by two dynamic (SST, Fdist) and one static variable (slope), models based on daily and weekly environmental data recovered all influencing variables. The greater relative importance of the dynamic variables was reflected by the contribution measures. Models based on monthly data performed poorly in selecting the influence of the Fdist variable for the niche of the Pseudoreal species.

The GLM analysis for the Static species showed smaller differences in variable selection between temporal aggregations. The two most important variables (SST and slope) were selected in more than 80% of iterations for all temporal groupings.

3.1.2. GBM

The GBM models performed well for the dynamic and the Pseudoreal species. For the Dynamic species, the 'daily' model was able to successfully detect all influencing variables, although the contribution of the main variable (SST) was smaller than its theoretical weight. The 'weekly' models detected the two main variables (SST and Fdist). However, for the 'monthly' models, only the influence of SST was detected.

For the Pseudoreal species, the daily approach identified the effects of SST and slope. However a relative upweighting of Fdist was found. The weekly analysis showed an almost perfect correlation between the variable contributions and their theoretical weight. Models using monthly data had Fdist contributions lower than expected, while the curvature contribution was overestimated. In general, models from this scenario showed the poorest accuracy regarding variable selection.

For the Static species, the three temporal aggregations produced similar results, with depth as the main contributing variable, as expected. However, some noise can be observed in the model for the monthly scenario, which exhibited a larger variation of contribution values.

3.2. Train and test AUC results

Ignoring the influence of mobility, we would expect that coarsening temporal resolution would decrease model performance for the Dynamic species, but would have little impact on the Static species, with the Pseudoreal species (influenced by both static and dynamic variables) showing an intermediate position. The AUC train and test results from the GBM and GLM approach confirm this hypothesis (Fig. 4). However when looking at the Static species, the monthly models performed slightly worse for the GLM modelling approach in the AUC test and for the GBM approach in the AUC train.

Results of the AUC test for the non-linear survey design (transects not following a pre-designed line and with unequal effort distribution) showed the same patterns (Fig. S3.2): finer temporal resolutions produced better AUC values for the Dynamic and the Pseudoreal species.

3.3. Suitability map projections

The visualized predictions (Fig. 5), showed the same pattern as previous evaluations. For the Dynamic and Pseudoreal differences are visible as temporal resolution coarsens. The Static species were less influenced by the temporal resolution of environmental variables, with no difference between temporal grain selections.

4. DISCUSSION

Selection of temporal resolution can be important for SDMs. When working in highly dynamic areas like the marine environment, and with species responding to daily environmental changes, the selection of temporal resolution can play an important role for environmental niche modelling procedures. In particular, the use of models based on an environmental dataset with finer temporal resolution can improve predictions of distribution.

The results obtained suggest these findings are not related to survey design, although further analysis with other designs and applying detectability indexes would be useful to discard any potential undetected effects.

4.1. Dynamic cetacean movements

The virtual species used in the present study were designed based on a review of previous distributional cetacean studies. A daily response to rapidly changing oceanographic patterns, as assumed for the present study, has been described or suggested for some cetacean species, such as baleen whales (Doniol-Valcroze et al. 2007, Druon et al. 2012). Similarly, small delphinids seem to be strongly influenced by dynamic oceanographic structures (Balance et al., 2006; Becker et al., 2010). Furthermore, daily environmental variation at small to medium spatial scales (approx. 5 km) can be important for other marine mobile pelagic species such as tuna (Hobday & Hartman, 2006). These responses are probably related with prey movements associated with local/regional oceanographic features. However other cetacean species appear to respond to broad-scale oceanographic patterns (Becker et al., 2010). Non-dynamic factors, such as bathymetric features (e.g. seamounts) can also play an important role for some cetacean species (e.g. bottlenose dolphins, Risso's dolphins, or pilot whales; Azzellino et al., 2008; Cañadas et al., 2002). Therefore, for species that may be more influenced by topographic features (such as deep-diving cetaceans) or broad/medium scale oceanographic features (such as the year-round presence of blue whales in the Costa Rica Dome; Reilly & Thayer, 1990), the dynamism captured by fine (temporal) scale oceanographic patterns may not be relevant. Even so, some of the results presented here suggest that a species responding to static factors could still be influenced by the temporal scale selected. Although it might be expected that dynamic variables would have no impact on models for these species, implicit relationships between static and dynamic variables can result in some explanatory power for dynamic variables.

4.2. Temporal resolution of dynamic variables

Generally, modelling with weekly environmental data produced the best results. Using monthly aggregations produced inconsistent results, with SST patterns more routinely detected than frontal distance. This might be a consequence of two factors: the variable dynamism and the species relation with the predictors. The level of variable dynamism could have a clear effect: SST is typically slower to change, while the location of thermal fronts can move rapidly. Consequently, a finer temporal resolution might be needed to detect the effects of highly dynamic variables (Fdist in this case). Moreover, the species ecology and their relation with the environment could be also essential. Response curves for SST in the present study were always based on a linear function, while those for distance to fronts were created using a unimodal function, leading to a more restricted range of suitable values for distances to frontal areas than for SST. Therefore the species modelled will be more sensitive to changes on thermal front locations than to SST changes. Likewise, the use of a finer temporal grain might be important when species are strongly related to specific ranges of one or more dynamic variables. However, for species with a more generalist relation with dynamic predictors, a coarser resolution could be suitable.

Scales et al. 2017 found that models using broader temporal scales can introduce bias in presence-availability for simulated blue whale movements for the California upwelling system. However, Mannocci et al. (2014) concluded that modelling using a climatological temporal scale (corresponding to seasonal oceanographic conditions averaged over 7 years) performed better than using weekly data. These authors examined a tropical system, which are typically more constant, with stable oceanographic phenomena that can be used by top predators in a predictable fashion. In contrast, the (temperate) Azores region has been described as an area with high mesoscale activity strongly influenced by the Gulf Stream and associated currents (Santos et al., 1995). In order to produce accurate models it is essential to have a good understanding of the oceanographic characteristics of the study area. When producing distribution estimates for areas with higher dynamism (such as temperate oceanic islands or coastal upwelling systems) the use of fine temporal resolution may be important.

We found little evidence that modelling with daily (rather than weekly) environmental data could lead to significant improvement in model performance. Weekly environmental aggregations may prove a fairly consistent representation of average daily conditions, as has been suggested for SST products in relatively dynamic environments, such as the California current (Becker et al., 2010).

It is important to consider the quality of the environmental data being analysed, particularly the characteristics of gap-free remote sensing products. Remote sensing datasets can have cloud-masked missing data which may reduce the predictive ability of the models (Scales et al. 2017). Some products include large areas of interpolation in order to cover cloud gaps. For example the MUR SST dataset used in this study performs spatio-temporal interpolation to fill gaps, but does this at multiple resolutions in order to preserve small-scale features (Vazquez-Cuervo et al., 2013). There is a trade-off to be made in deciding between daily and weekly aggregates. Our study indicates that weekly means may be the best choice at present.

4.3. Relationship between spatial and temporal scales

In the present study we found differences in model predictions between the different temporal grain sizes, although we did not test the combined effects of spatial and temporal scales. The temporality of oceanographic and biological processes can be dependent on spatial scale, consequently temporal variability tends to be higher at finer spatial scales (Haury et al., 1978; Hunt & Schneider 1978). Both Redfern et al. 2006 and Balance et al. 2006 reinforced the importance of using adequate resolutions to the scale of the data collected, matching spatial and temporal grain size to the specific research question. However, Becker et al. (2010) suggested that effects of the use of different spatial grain sizes are relatively small, finding similar functional relationships between SST response variables across different spatial resolution. Additionally, Scales et al. (2017) found that spatial effects at small temporal grain sizes (daily-monthly) are relatively small compared to climatological scales. Other studies of modelling applications suggest similar effects of spatial resolution for different areas and taxonomic groups (Guisan et al., 2007; Redfern et al., 2008, Becker et al., 2010). Therefore, the results obtained in the present study might be useful for different spatial scales when working at relatively small temporal scales.

Nonetheless, the detectability of the influence of particular environmental factors can be dependent on the spatial resolution. Guinet et al. (2001) found that different spatial scales resulted in different variable influences on fur seal niche models. The relative importance of oceanic features will change with geographical scale, from oceanic gyres down to random turbulence (Parsons et al., 2013). For example, a weekly dataset might fail to detect an ocean gyre. Thus, coarsening temporal resolution might have a similar effect as coarsening the spatial resolution. The use of a fine temporal grain may negatively impact the detection of some large-scale oceanographic features that can influence cetacean distributions, such as island-generated eddies or domes (e.g. Ballance et al., 2006), adding a temporal dimension to the Redfern et al. (2008) hypothesis of the relationship between signal-to-noise ratio and spatial scale.

It is possible that for studies focusing on mesoscale/global distributions, a coarser temporal and spatial resolution may be more appropriate (e.g. Kaschner et al., 2006; Mannocci et al., 2014), while studies focused on species distribution modelling on regional and local scales may be improved by examining finer temporal resolutions (e.g. Becker et al., 2016). Yet, this might limit model applicability. Models built using seasonal environmental data won't be able to predict distributions at finer grain sizes, and models using a weekly resolution might fail when projected into a global scale (Redfern et al. 2006). In contrast, Scales et al. 2017 found that even when working with large spatial scales (111 km) the use of seasonal and climatological fields increased the model error substantially but admit that this observation may not be valid in all biogeographic provinces. Our simulation results support the suggestion that care is needed when matching different scales (Scales et al., 2017). In fact we found that, in some cases, even if working at small spatial scales the use of a monthly resolution can produce unrealistic predictions.

Redfern et al. (2006) suggest the simultaneous modelling of cetacean distributions at different scales as a way to overcome this problem. Further research is needed to understand better the relationship between temporal and spatial scales.

4.4. Further considerations

It could be argued that temporal dynamics are not an issue for mobile species with high residency, or for those species that are mainly dependent on more stable environmental conditions. In these cases, animals would tend not to move from their preference areas, within reasonable environmental boundaries. Migratory species traveling long distances can have strong site fidelity between migrations (Rasmussen et al. 2007). However it is important to keep in mind that species can interact with the environment at multiple scales (e.g. hourly feeding, daily foraging, seasonal migration). This behaviour-dependent habitat utilisation may be detectable at different scales, for instance in baleen whale migrations (Corkeron & Connor, 1999; Rasmussen et al., 2007).

Given that a species' niche is not usually well understood prior to modelling, it would be a good practice to include dynamic, static, and climatological variables in the model fitting process to test for influences at multiple spatial and temporal scales. However, high quality environmental data for many oceanographic variables rarely exists at daily temporal resolutions in most parts of the ocean and fine-scale prey distribution is non-existent on most temporal scales. As these data become available it would be worth testing their influence. Meanwhile the inclusion of variability measures (e.g. minimum daily temperature in a given month) when using coarser grains can provide a way of adding some finer temporal resolution data, improving model predictions.

Approaches using a finer grain (both on biological, spatial and temporal scales) may be more suitable for effective conservation measures (Stelzenmüller et al., 2013). In fact, recent studies (e.g. Maxwell et al., 2016) highlighted the importance of management that changes in space and time in response to changes in the ocean and its users. Dynamic management techniques are appealing for areas with substantial temporal and spatial variability (e.g. seasonal tourism, Becker et al., 2016). However other approaches might be useful: for species with low or moderate sighting rates the use of models using broader temporal resolutions might be justified (Roberts et al., 2016). For example if the goal of the study is to produce distribution maps of beaked whales, which are rarely sighted, the use of a fine temporal scale might be an unrealistic choice. There are several factors to take into consideration before choosing a specific temporal resolution, such as the ecology of the target species, the dynamism of the environment, the species detectability, the spatial scale to be used, the main objectives of the analysis or even the data availability.

4.5. Final remarks

The combination of mobility and habitat dynamism is a key issue when selecting the best temporal resolution to model a species' ecological niche. In this study we used theoretical species responding to daily changes of environmental variables to test these effects, and we found important differences between temporal resolutions. Even if theoretical species can differ from real world examples, it is important to emphasize the potential impact of these dynamic factors. Assuming that low-frequency environmental data will sufficiently reproduce high-frequency variation in species distributions might lead to inaccurate distribution models.

It is essential to have some knowledge of the species ecology and variable dynamism to select the best predictors and resolutions. For fine scale applications (e.g. dynamic ocean management), when using variables with high temporal dynamism (e.g. distance to frontal areas), and highly mobile species (or for species strongly related to dynamic environmental predictors), our results suggest exploring weekly temporal resolution. Coarser resolutions might be useful when working with variables with low dynamism or for species less dependent on dynamic variables (e.g. some deep diving cetaceans). However one must take into consideration that averaging environmental variables over larger time periods may mask the underlying dynamic patterns and produce a less realistic niche model, which may be misleading and even detrimental for conservation purposes.

ACKNOWLEDGEMENTS

This study was only possible through the collaboration of the Groupe de Recherché sur les Cétacés (GREC), who provided the infrastructure and knowledge needed. We thank Cláudia Faustino for the shapefiles to simulate the sampling transects used in this study. Marc Fernandez is supported by grant M3.1.2/F/028/2011 from the Fundo Regional para a Ciência e Tecnologia (Azores Government). This research was partially supported by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Programme and national funds through FCT - Foundation for Science and Technology, under the project "PEst-C/MAR/LA0015/2013", by the Strategic Funding UID/Multi/04423/2013 through national funds provided by FCT -

Foundation for Science and Technology and European Regional Development Fund (ERDF), in the framework of the program PT2020 and by cE3c funding (Ref: UID/BIA/00329/2013). It was also partly supported by CIRN (Centro de Investigação de Recursos Naturais, University of the Azores), and CIIMAR (Interdisciplinary Centre of Marine and Environmental Research, Porto, Portugal).

REFERENCES

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504-1513.

Azzellino, A., Gaspari, S., Airoldi, S., Nani, B. (2008) Habitat use and preferences of cetaceans along the continental slope and the adjacent pelagic waters in the western Ligurian Sea. *Deep-Sea Research I*, **55**, 296-323.

Austin, M. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological modelling*, **157**, 101-118.

Bakun, A. (2006) Fronts and eddies as key structures in the habitat of marine fish larvae: opportunity, adaptive response and competitive advantage. *Scientia Marina*, **70**, 18.

Ballance, L.T., Pitman, R.L., Fiedler, P.C. (2006) Oceanographic influences on seabirds and cetaceans of the eastern tropical Pacific: A review. *Progress in Oceanography*, **69**, 360-390.

Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413-423.

Becker, E., Forney, K., Ferguson, M., Foley, D., Smith, R., Barlow, J. & Redfern, J. (2010) Comparing California Current cetacean–habitat models developed using in situ and remotely sensed sea surface temperature data. *Marine Ecology Progress Series*, **413**, 163-183.

Becker, E., Forney, K., Fiedler, P., Barlow, J., Chivers, S., Edwards, C., Moore, A. & Redfern, J. (2016) Moving Towards Dynamic Ocean Management: How Well Do Modeled Ocean Products Predict Species Distributions? *Remote Sensing*, **8**, 149.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437-448.

Busby, J.R. (1991) BIOCLIM-a bioclimate analysis and prediction system. *Plant Protection Quarterly (Australia)*.

Cañadas, A. & Hammond, P.S. (2008) Abundance and habitat preferences of the short-beaked common dolphin Delphinus delphis in the southwestern Mediterranean: implications for conservation. *Endangered Species Reserach*, **4**, 309-331.

Cañadas, A., Sagarminaga, R., García-Tiscar, S. (2002) Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain. *Deep-Sea Research I*, **49**, 2053-2073.

Corkeron, P.J. & Connor, R.C. (1999) Why do baleen whales migrate? *Marine Mammal Science*, **15**, 1228-1245.

Doniol-Valcroze, T., Berteaux, D., Larouche, P., Sears, R. (2007) Influence of thermal fronts on habitat selection by four rorqual whale species in the Gulf of St. Lawrence. *Marine Ecology Progress Series*, **335**, 207 - 216.

Doniol-Valcroze, T., Lesage, V., Giard, J. & Michaud, R. (2012) Challenges in marine mammal habitat modelling: evidence of multiple foraging habitats from the identification of feeding events in blue whales. *Endangered Species Research*, **17**, 255-268.

Druon, J.N., Panigada, S., David, L., Gannier, A., Mayol, P., Arcangeli, A., Cañadas, A., Laran, S., Di Méglio, N. & Gauffier, P. (2012) Potential feeding habitat of fin whales in the western Mediterranean Sea: an environmental niche model. *Marine Ecology Progress Series*, **464**, 289-306.

Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Wu, G.-L. & Wang, Z.-G. (2014) SDMvspecies: a software for creating virtual species for species distribution modelling. *Ecography*, **38**, 108 - 110.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802-813.

Faustino, C.E.S., Silva, M.A., Marques, T.A. & Thomas, L. (2010) Designing a shipboard line transect survey to estimate cetacean abundance off the Azores archipelago *Arquipélago - Life and Marine Sciences*, **27**, 49-58.

Forney, K.A., Ferguson, M.C., Becker, E.A., Fiedler, P.C., Redfern, J.V., Barlow, J., Vilchis, I.L. & Ballance, L.T. (2012) Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. *Endangered Species Research*, **16**, 113-133.

Fryxell, J, Sinclair A. & A. Cughley, G. (2014) Wildlife ecology, conservation, and management, 3rd edn. Wiley-Blackwell

Guinet, C., Dubroca, L., Lea, M.-A., Goldsworthy, S.D., Cherel, Y., Duhamel, G., Bonadonna, F. & Donnay, J.-P. (2001) Spatial distribution of foraging in female Antarctic fur seals Arctocephalus gazella in relation to oceanographic variables: a scale-dependent approach using geographic information systems. *Marine Ecology Progress Series*, **219**, 251-264.

Guisan, A., Graham, C.H., Elith, J., Huettmann, F. & the, N.S.D.M.G. (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332-340.

Haury, L., McGowan, J. & Wiebe, P. (1978) Patterns and processes in the time-space scales of plankton distributions. *Spatial pattern in plankton communities*, pp. 277-327. Springer.

Hirzel, A.H., Helfer, V. & Metral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**, 111-121.

Hobday, A.J. & Hartmann, K. (2006) Near real-time spatial management based on habitat predictions for a longline bycatch species. *Fisheries Management and Ecology*, **13**, 365-380.

Howell, E.A., Kobayashi, D.R., Parker, D.M., Balazs, G.H. & Polovina, A. (2008) TurtleWatch: a tool to aid in the bycatch reduction of loggerhead turtles Caretta caretta in the Hawaii-based pelagic longline fishery. *Endangered Species Research*, **5**, 267-278.

Hunt Jr, G. & Schneider, D. (1987) Scale-dependent processes in the physical and biological environment of marine birds. *Seabirds: feeding ecology and role in marine ecosystems*, 7-41.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An introduction to statistical learning*. Springer.

Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in ecology & evolution (Personal edition)*, **27**, 151-159.

Kaschner, K., Watson, R., Trites, A.W. & Pauly, D. (2006) Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, **316**, 285-310.

Katsanevakis, S., Weber, A., Pipitone, C. et al. (2012) Monitoring marine populations and communities: methods dealing with imperfect detectability. *Aquatic Biology*, **16**, 31-52.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014) Prevalence, thresholds and the performance of presence–absence models. *Methods in Ecology and Evolution*, **5**, 54-64.

Lindsay, R.E., Constantine, R., Robbins, J., Mattila, D.K., Tagarino, A. & Dennis, T.E. (2016) Characterising essential breeding habitat for whales informs the development of large-scale Marine Protected Areas in the South Pacific. *Marince Ecology Progress Series* ,**54** , 263-275.

MacLeod, C.D., Weir, C.R., Pierpoint, C. & Harland, E.J. (2007) The habitat preferences of marine mammals west of Scotland (UK). *Journal of the Marine Biological Association of the United Kingdom*, **87**, 157 - 164.

Mann, K. & Lazier, J. (2013) *Dynamics of marine ecosystems: biological-physical interactions in the oceans*. John Wiley & Sons.

Mannocci, L., Laran, S., Monestiez, P., Dorémus, G., Van Canneyt, O., Watremez, P. & Ridoux, V. (2014) Predicting top predator habitats in the Southwest Indian Ocean. *Ecography*, **37**, 261-278.

Maxwell, S.M., Hazen, E.L., Lewison, R.L., Dunn, D.C., Bailey, H., Bograd, S.J., Briscoe, D.K., Fossette, S., Hobday, A.J., Bennett, M., Benson, S., Caldwell, M.R., Costa, D.P., Dewar, H., Eguchi, T., Hazen, L., Kohin, S., Sippel, T. & Crowder, L.B. (2015) Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. *Marine Policy*, **58**, 42-50.

Miller, P. (2009) Composite front maps for improved visibility of dynamic sea-surface features on cloudy SeaWiFS and AVHRR data. *Journal of Marine Systems*, **78**, 327-336.

Miller, P.I., Scales, K.L., Ingram, S.N., Southall, E.J. & Sims, D.W. (2015) Basking sharks and oceanographic fronts: quantifying associations in the north-east Atlantic. *Functional Ecology*, **29**, 1099-1109.

Moura, A.E., Sillero, N. & Rodrigues, A. (2012) Common dolphin (Delphinus delphis) habitat preferences using data from two platforms of opportunity. *Acta Oecologica-International Journal of Ecology*, **38**, 24-32.

Panigada, S., Zanardelli, M., MacKenzie, M., Donovan, C., Melin, F. & Hammond, P.S. (2008) Modelling habitat preferences for fin whales and striped dolphins in the Pelagos Sanctuary (Western Mediterranean Sea) with physiographic and remote sensing variables. *Remote Sensing of Environment*, **112**, 3400-3412.

Parsons, T.R., Takahashi, M. & Hargrave, B. (2013) *Biological oceanographic processes*. Elsevier.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton University Press.

Praca, E. & Gannier, A. (2007) Ecological niche of three teuthophageous odontocetes in the northwestern Mediterranean Sea. *Ocean Science Discuss.*, **4**, 785-815.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 3.2.2.

Rasmussen, K., Palacios, D.M., Calambokidis, J., Saborío, M.T., Dalla Rosa, L., Secchi, E.R., Steiger, G.H., Allen, J.M. & Stone, G.S. (2007) Southern Hemisphere humpback whales wintering off Central America: insights from water temperature into the longest mammalian migration. *Biology Letters*, **3**, 302-305.

Redfern, J.V., Barlow, J., Ballance, L.T., Gerrodette, T. & Becker, E.A. (2008) Absence of scale dependence in dolphin–habitat models for the eastern tropical Pacific Ocean. *Marine Ecology Progress Series*, **363**, 1-14.

Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Kaschner, K., Baumgartner, M.F., Forney, K.A., Ballance, L.T., Fauchald, P., Halpin, P., Hamazaki, T., Pershing, A.J., Qian, S.S., Read, A., Reilly, S.B., Torres, L. & Werner, F. (2006) Techniques for cetacean habitat modeling. *Marine Ecology Progress Series*, **310**, 271-295.

Reilly, S.B. & Thayer, V.G. (1990) Blue whale (Balaenoptera musculus) distribution in the eastern tropical Pacific. *Marine Mammal Science*, **6**, 265-277.

Ridgeway G. et al. (2015). Gbm: Generalized Boosted Regression Models. R package version 2.1.1.

Roberts, J.J., Best, B.D., Mannocci, L., Fujioka, E., Halpin, P.N., Palka, D.L., Garrison, L.P., Mullin, K.D., Cole, T.V.N., Khan, C.B., McLellan, W.A., Pabst, D.A. & Lockhart, G.G. (2016) Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Scientific Reports*, **6**, 22615.

Santos, R.S., Hawkins, S., Monteiro, L.R., Alves, M., Isidro, E.J. (1995) Marine research, resources and conservation in the Azores. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **5**, 311-354.

Sbrocco, E.J., & Barber, P.H. (2013) MARSPEC: ocean climate layers for marine spatial ecology. *Ecology*, **94.4** , 979-979.

Scales, K.L., Miller, P.I., Hawkes, L.A., Ingram, S.N., Sims, D.W. & Votier, S.C. (2014) REVIEW: On the Front Line: frontal zones as priority at-sea conservation areas for mobile marine vertebrates. *Journal of Applied Ecology*, **51**, 1575-1583.

Scales, K.L., Hazen, E.L., Jacox, M.G., Edwards, C.A., Boustany, A.M., Oliver, M.J. & Bograd, S.J. (2017) Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. *Ecography*, **40,** 210-220.

Stelzenmüller, V., Breen, P., Stamford, T., et al. (2013) Monitoring and evaluation of spatially managed areas: A generic framework for implementation of ecosystem based marine management and its application. *Marine Policy*, **37**, 149-164.

Svensson, J.R., Jonsson, L. & Lindegarth, M. (2013) Excessive spatial resolution decreases performance of quantitative models, contrary to expectations from error analyses. *Marine Ecology Progress Series*, **485**, 57-73.

Vazquez-Cuervo, J., Dewitte, B., Chin, T.M., Armstrong, E.M., Purca, S. & Alburqueque, E. (2013) An analysis of SST gradients off the Peruvian Coast: The impact of going to higher resolution. *Remote Sensing of Environment*, **131**, 76-84.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Virtual species design

Appendix S2 Virtual species suitability projections and presence/absence generation

Appendix S3 Survey details and results

Bioskecth: Marc Fernandez is a PhD candidate of the 3CBIO program at the University of Azores and a current member of the Azorean Biodiversity Group of the Centre for Ecology, Evolution and Environmental Changes. His main research interests are marine spatial ecology, particularly niche modelling of cetaceans. Chris Yesson is a research fellow for the Institute of Zoology (ZSL), focussing his research on the impact of trawling on benthic habitats in Greenland, abundance of large brown seaweeds and phyloclimatic modelling. Alexandre Gannier is the president of the GREC. His work has been related with the study of the general ecology of cetaceans, focusing on populations' ecology, acoustic ecology and acoustic impact.

Editor: Richard Pearson

Table 1. Environmental variables used to construct the virtual species suitability indexes.

Variables	Definition	Source
Oceanographic (dynamic)		
SST	Sea surface temperature (°C)	NASA's Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (SST) - http://mur.jpl.nasa.gov/. Downloaded with spatial resolution of 1kmx1km on a daily basis.
Fdist	Distance from major thermal front (km)	Processed from NASA's Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (SST) following Miller (2009) methodology.
Physiographic (static)		
depth	Depth (m)	Bathymetric metadata and Digital Terrain Model data products derived from the EMODnet Bathymetry portal - http://www.emodnet-bathymetry.eu. Downloaded with a spatial resolution of: 0.125x0.125 minutes.
slope	Slope (degrees)	Processed from EMODnet Bathymetry using DEM Surface Tools for ArcGIS 10.2
d200	Distance from 200 m bathymetric line (km)	Processed from EMODnet Bathymetry using QGIS 2.1.2
curv	Bottom general curvature	Processed from EMODnet Bathymetry using DEM Surface Tools for ArcGIS 10.2

Table 2. Formulas used to build the suitability values for each virtual species according to the environmental variables.

Species	Suitability index
Dynamic	$H_D = \frac{1}{(2 + 1.5)}(2SST + 1.5Fdist)$
Static	$H_S = \frac{1}{(2 + 1.5 + 1)}(2Depth + 1.5D200 + Slope)$
Pseudoreal	$H_{PR} = \frac{1}{(2 + 1.5 + 1)}(2SST + 1.5Slope + Fdist)$

Figure 1: Example of environmental variables for the 20th of August 2013. Variables are categorized as dynamic/oceanographic (SST (a) and Fdist(b)) or static/geographic (Depth(c), Slope(d), Dcoast(e) and Curvature(f)).

Figure 2: Study area map (Eastern Group and Central Group, Azores Archipelago) with the virtual transects (with the nine substratum divisions) used for the niche modelling calculations.

Figure 3: Results of variable selection for the three temporal aggregations (daily, weekly and monthly – in rows), two models algorithms (GLM and GBM), and 3 virtual species (Dynamic, Static and Pseudoreal – in columns). Results of the GBM models are expressed as mean variable contribution over the 1000 iterations according to variable relative importance. Results of the GLM are expressed as the number of times a specific variable was selected for the model after the AIC stepwise selection procedure.

Figure 4: Results for the training and testing AUC using sampling data for the GBM and GLM model algorithms (rows), and the three temporal grain selections (daily, weekly and monthly) and three virtual species (dynamic, static and pseudoreal), (columns). AUC ranges from 0 to 1.

Figure 5: Suitability maps for randomly chosen dates. Projections were made for the three virtual species (Dynamic, Static and Pseudoreal) using the GLM approach. Columns represent the different temporal resolutions and the theoretical suitability (noted as Theoretical in the figure) for each species. The worm-like

pattern observed in the Dynamic species it is related to the preference for a given distance to the thermal front.

Figure 1:

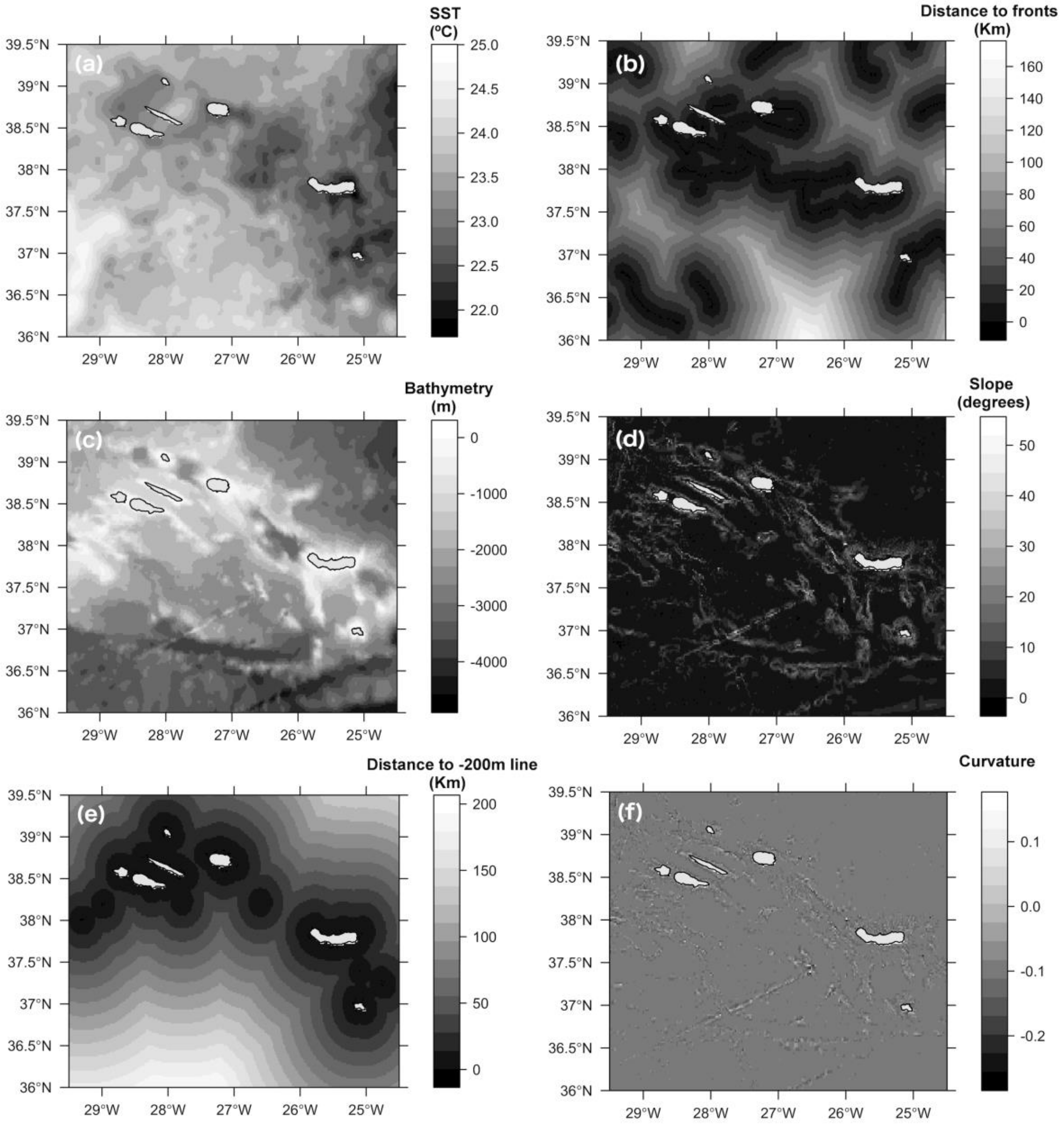

(a)
SST
(°C)
25.0
24.5
24.0
23.5
23.0
22.5
22.0
(b)
Distance to fronts
(Km)
160
140
120
100
80
60
40
20
0
(c)
Bathymetry
(m)
0
-1000
-2000
-3000
-4000
(d)
Slope
(degrees)
50
40
30
20
10
0
(e)
Distance to -200m line
(Km)
200
150
100
50
0
(f)
Curvature
0.1
0.0
-0.1
-0.2
39.5°N
39°N
38.5°N
38°N
37.5°N
37°N
36.5°N
36°N
29°W
28°W
27°W
26°W
25°W

Figure 2:

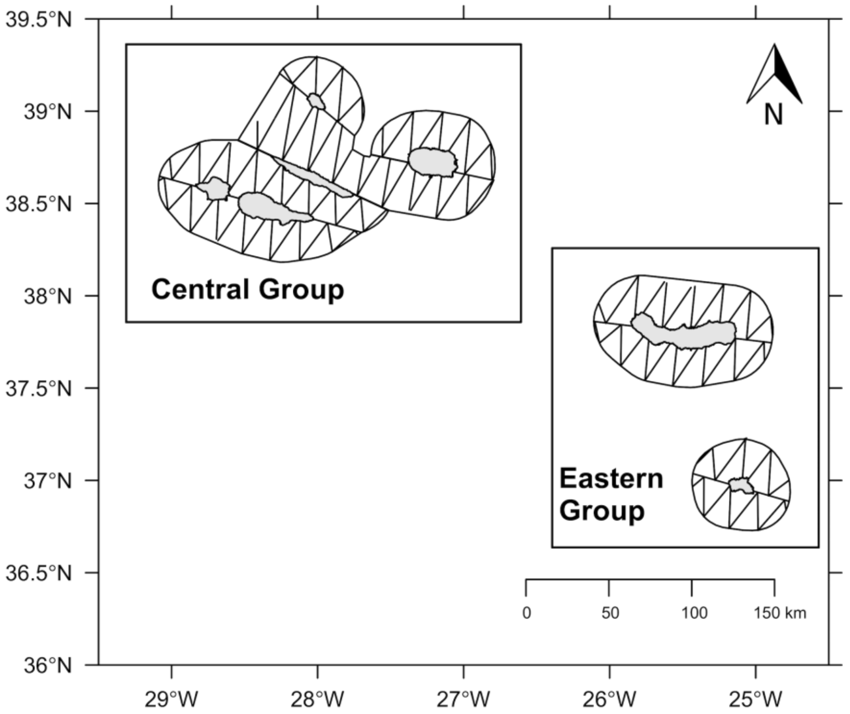

39.5°N
39°N
38.5°N
38°N
37.5°N
37°N
36.5°N
36°N
29°W
28°W
27°W
26°W
25°W
N
Central Group
Eastern Group
0
50
100
150 km

Figure 3:

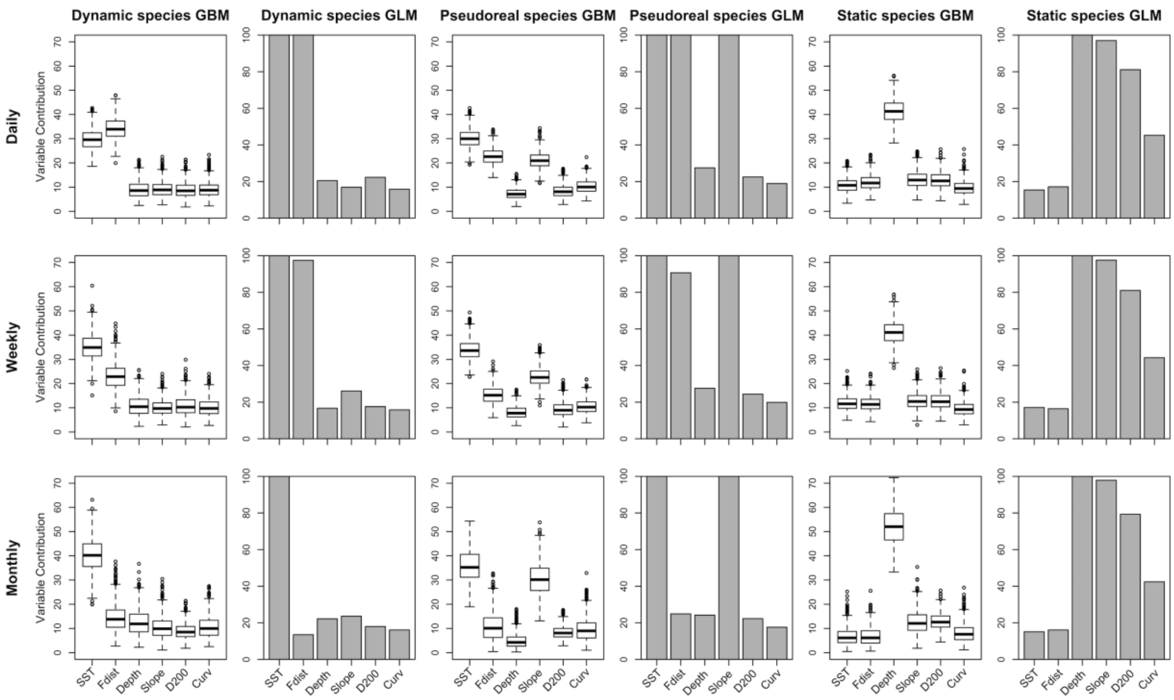

Dynamic species GBM
Dynamic species GLM
Pseudoreal species GBM
Pseudoreal species GLM
Static species GBM
Static species GLM
Daily
Weekly
Monthly
Variable Contribution
0
10
20
30
40
50
60
70
80
100
SST
Fdist
Depth
Slope
D200
Curv

Figure 4:

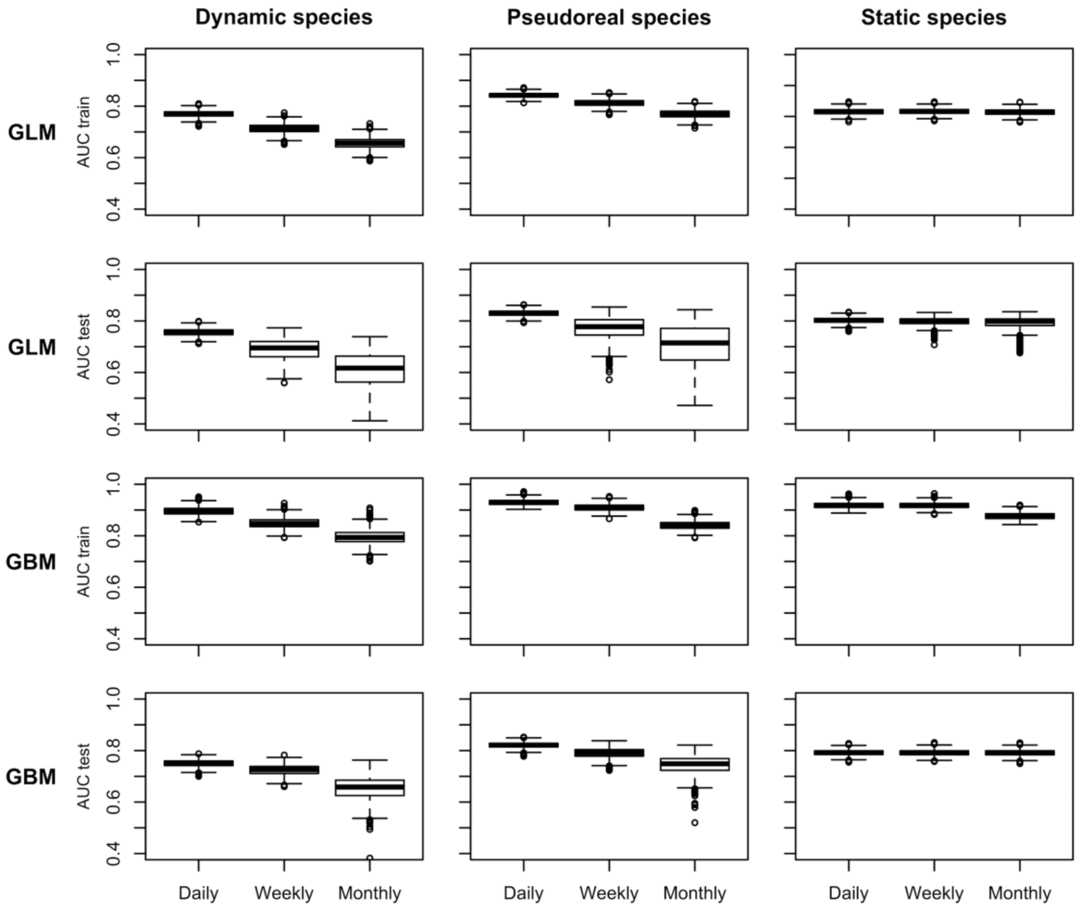
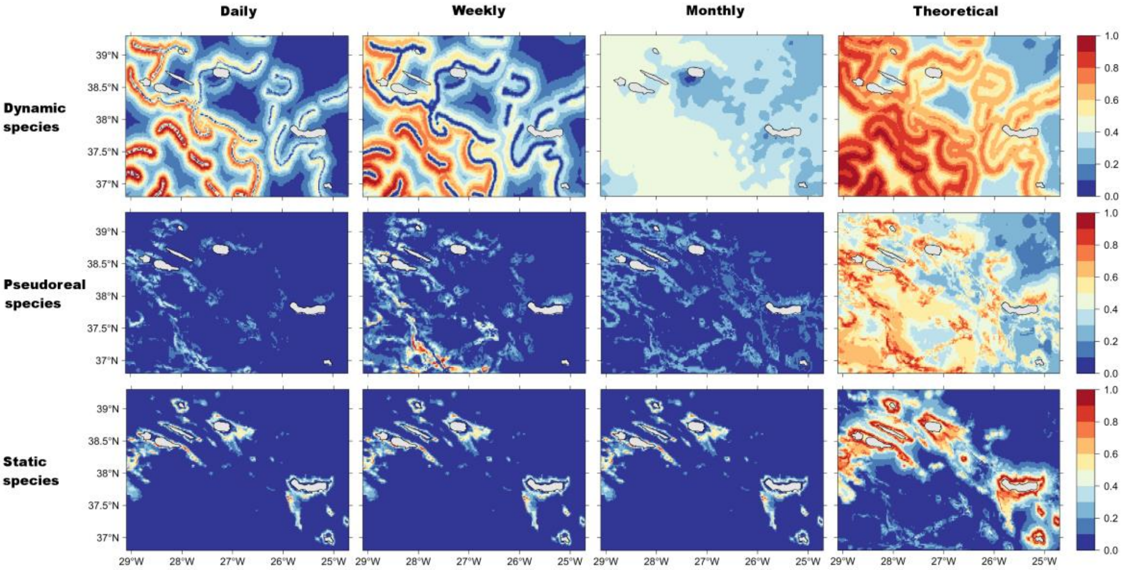

Dynamic species
Pseudoreal species
Static species
GLM
AUC train
GLM
AUC test
GBM
AUC train
GBM
AUC test
0.4
0.6
0.8
1.0
Daily
Weekly
Monthly

Figure 5:


Daily
Weekly
Monthly
Theoretical
Dynamic species
Pseudoreal species
Static species
39°N
38.5°N
38°N
37.5°N
37°N
29°W
28°W
27°W
26°W
25°W
1.0
0.8
0.6
0.4
0.2
0.0