Explicitly integrating a third dimension in marine species distribution modelling

G. A. Duffy*, S. L. Chown

School of Biological Sciences, Monash University, Victoria 3800, Australia

ABSTRACT: Marine species distribution modelling (SDM) studies frequently pair surface environmental variables with deep-water occurrence records despite the disparity of environmental conditions at the surface and at depth. Straightforward methods, which require only simple modification of existing SDM techniques, have been recently proposed as a means to overcome this problem. Nevertheless, subsequent marine SDM studies have not adopted these modified techniques, continuing to draw conclusions from analyses that associate 2-dimensional (2D; i.e. surface) environmental data with occurrence points distributed in a 3-dimensional (3D) space (i.e. across a range of depths). These include studies involving multi-species macroecological assessments, where SDMs are rarely tailored to suit each species modelled. Here, we use traditional (2D) and modified (2.5D) SDM methods to emphasise the importance of incorporating depth in marine ecological modelling. We demonstrate that ignoring depth may lead to mis-representative and potentially misleading results, and show that relatively straightforward 2.5D techniques provide an enhanced representation of the environmental niche and, therefore, produce more representative range estimates than their 2D counterparts. Explicit 3D approaches need to be developed to better represent marine diversity and distributions and to test macroecological generalisations.

KEY WORDS: 3D modelling · Macroeconomy · Niche model · Pelagic fish · Species distribution · Species–environment relationships

INTRODUCTION

In ecological modelling, environments are frequently treated as flat, 2-dimensional (2D) spaces, despite having both 2D and 3-dimensional (3D) spatial components. The influence of a temporal dimension is sometimes considered, for example in studies of environmental change (e.g. Elith et al. 2011, Bellard et al. 2013), but the third spatial dimension (z) is often overlooked. Spatial simplification can be rationalised when applied to relatively flat benthic or terrestrial ecosystems, but such simplification is inappropriate when applied to systems where the environmental and ecological importance of a third spatial dimension is more explicitly recognised (Fig. 1). Examples of spatially 3D systems include pelagic habitats, where the extra dimension, introduced by the depth of the water column, adds to environmental variation and has considerable ecological significance (Pawar et al. 2012, Bentlage et al. 2013). Envisaging a scenario where the environmental variation of a 3D space can be confidently realised...
using data that represent only 2 dimensions is, in consequence, difficult. Such simplification contravenes an important principle of sound species distribution modelling (SDM) practice. That is, environmental data should be causally linked, or at the very least thought for sound biological reasons to be causally linked, to the distribution of the species of interest, and should represent key environmental factors of its habitat (Guisan & Thuiller 2005, Guillera-Arroita et al. 2015, Reiss et al. 2015). With the increasing application of SDM methods to marine ecosystems (Robinson et al. 2011, Monk 2014), the potential bias that is introduced when the third dimension is overlooked warrants further consideration.

Despite past and recent emphasis of the 3D nature of most marine environments (Carr et al. 2003, Bentlage et al. 2013), the coupling of occurrence records from a range of depths with surface environmental variables is not uncommon in marine SDM studies (Fig. 2; e.g. García-Roselló et al. 2015). Some studies use ocean depth, often in addition to surface variables, in an attempt to account for the third dimension. Using ocean depth as a predictor variable may, however, prove problematic because depth is a distal predictor, which correlates inconsistently with causative predictors (e.g. temperature, salinity) across large geographic ranges (Elith 2015, see also McArthur et al. 2010). In contrast, other variables derived from bathymetry, such as seabed rugosity and slope (see Lecours et al. 2016 for a comprehensive review of marine geomorphometry), can improve models of habitat suitability for sessile benthic organisms as submarine topography is expected to

Fig. 1. Graphical representation of a theoretical environmental data layer. (a) 2D environment layers can be used to represent relatively flat benthic (i.e. seabed) or terrestrial (i.e. land-surface) environments. (b) While topographic variation (e.g. bathymetry) often influences these 2D layers, in most instances the utilised space of the environment can still be adequately represented by a flat 2D skin overlaid onto this topography (i.e. a = b). (c) However, when the utilised space is explicitly 3D (e.g. aquatic or arboreal systems), a single 2D layer cannot accurately represent the environment in its entirety (i.e. a ≠ b and b ≠ c)

Fig. 2. Environmental variables used for single genus (n = 17) or multiple genera (n = 38) species distribution modelling studies published in the last decade (2005−2015) for marine organisms living at depths greater than 200 m. Studies are classified as using only surface variables (e.g. sea surface temperature), using variables derived from bathymetry (e.g. ocean depth, rugosity), using simplified at-depth variables (e.g. only bottom temperature rather than a whole depth profile of water temperature), or using depth-specific variables (e.g. temperature at the depth of recorded occurrence; as in Bentlage et al. 2013). See Supplement 1 for details of literature search and Table S1 in Supplement 2 (www.int-res.com/articles/suppl/m564p001_supp/) for full list of studies
influence the distributions of groups closely associated with the benthos (Vierod et al. 2014). An increasing number of marine modelling studies use simplified at-depth variables (e.g. surface temperature and bottom temperature) to represent marine environments (Fig. 2). This represents a clear improvement over the use of surface variables alone and may be sufficient for modelling of benthic taxa or taxa with narrow depth distributions. Simplified at-depth data may, however, fail to capture the intricacies of environmental variation throughout the water column (e.g. Vecchione et al. 2015) and are, therefore, poorly suited for cross-taxa studies. Indeed, of the 55 studies reviewed in Fig. 2, only Bentlage et al. (2013) built SDMs using predictor data that were representative of the specific depth at which an occurrence point was recorded. No cross-taxon studies that use depth-specific data were identified.

Of particular concern are marine SDM studies that model multiple pelagic taxa using only surface variables (Fig. 2; e.g. García-Roselló et al. 2015) and do not consider environmental variation associated with depth. In essence, using only 1 set of 2D environmental layers (e.g. sea surface temperature, sea surface salinity) to represent a 3D marine habitat has the potential to misrepresent the niches (sensu Kearney & Porter 2009) of species that inhabit deeper water because environmental variables vary substantially with depth. Moreover, models trained on inaccurate environmental predictor data may provide a flawed representation of the environmental drivers of species occurrence, resulting in erroneous predictions of species distributions. Minor modifications to SDM methods can, however, be applied to address these errors and will encourage discussion of how 3D environments should be modelled and compared. We therefore emphasise the importance of explicitly incorporating the third dimension where it is likely to be biologically relevant (Dambach & Rödder 2011, Bentlage et al. 2013) and suggest that all 3 spatial dimensions of the study environment should be considered during the construction of fit-for-purpose marine SDMs (Guillera-Arroita et al. 2015).

**SDMs IN 3D MARINE ENVIRONMENTS**

Environmental data from sea-surface datasets arguably provide a functional approximation of environmental conditions associated with shallow occurrence points (Tyberghein et al. 2012). Deep-water occurrence records, however, exhibit a much greater disparity between surface and at-depth environmental conditions (Fig. 3). Using environmental data that

---

Fig. 3. Mean depth and water temperature (±1 SD) of occurrence points of common fish species (>25 valid occurrence records in the Global Biodiversity Information Facility, GBIF), sorted by mean recorded depth, from the Northeast Atlantic (45–55° N, 0–30° W). Depth-corrected mean temperature estimates for each species (black points; linearly interpolated from the World Ocean Atlas 2013 dataset (Locarnini et al. 2013) using geographic location and depth of occurrence of each GBIF occurrence point) and uncorrected estimates from the sea surface layer (grey points) shown. See Supplement 1 for full methods and Table S2 in Supplement 2 for list of fish species.
represent only the topmost layer of the marine environment (e.g. Bio-ORACLE; Tyberghein et al. 2012) to model marine organisms would, therefore, mean that, for many meso-, bathy-, and abyssopelagic species, the assumption that environmental data are causally linked to species distributions (Guisan & Thuiller 2005, Guillera-Arroita et al. 2015, Reiss et al. 2015) is not met. In consequence, generating SDMs this way and assessing their validity, such as for estimating species ranges and testing macroecological patterns, is, by default, problematic.

Using temperature as an example, the species mean (i.e. the mean of temperature values associated with all occurrence records of a species) of fishes can change substantially when corrected for depth of occurrence points (Fig. 3). Depth-corrected mean temperature estimates for each species show little agreement with uncorrected estimates from the sea surface layer. Changes in the environmental predictors associated with occurrence points and of the overall species mean will considerably alter the probability density of occurrence points within a multivariate environmental space, from which a realised niche (the hypervolume occupied by species within this space; Kearney & Porter 2009) is derived. Realised niches of deep-water species generated using surface environmental predictors will be inaccurate and will substantially change when depth-corrected environmental data are used. This will subsequently alter the results generated by distribution-modelling methods that use a species’ realised niche to estimate potential distributions (e.g. MaxEnt; Phillips et al. 2006, Elith et al. 2011). Use of misrepresentative environmental data, such as that resulting from positional error (Osborne & Leitão 2009, Mitchell et al. 2017) or temporal mismatches between species and environmental sampling dates (Roubicek et al. 2010), is known to reduce SDM performance. Osborne & Leitão (2009) hypothesised that, with positional errors, spatial autocorrelation of environmental data may mitigate substantial mismatch errors between occurrence and environment. This protection is unlikely, however, to be afforded to depth-driven mismatches, which often span greater distances than those of relatively localised positional errors (i.e. 100s of metres along the z axis for depth mismatches versus 10s of metres along the x and y axes for positional errors). Although surface productivity influences deep-sea organic carbon flux (Lutz et al. 2002, Ruhl & Smith, 2004), other surface variables (e.g. sea surface temperature) are often poor indicators of conditions at depth.

**QUANTITY–QUALITY TRADE-OFFS**

The modelling of 3D aquatic systems using 2D surface variables as environmental predictors is a conspicuous example of an increasingly recognised problem in ecological modelling where third-party environmental datasets might not provide an accurate representation of the environment from which samples are taken (Storlie et al. 2014, Woods et al. 2015). Ideally, environmental data associated with a species occurrence should be collected at the time and place of sampling to avoid spatial (Osborne & Leitão 2009, Mitchell et al. 2017) or temporal (Roubicek et al. 2010) mismatches between a confirmed presence and associated environmental data. Doing so is not, however, always practical. When using occurrence data from repositories such as the Global Biodiversity Information Facility (GBIF; www.gbif.org), environmental data are frequently unobtainable and must be added retroactively. If environmental data are collected post hoc from a third-party dataset, the limitations of this data should be considered and any datasets used should be as suitable as possible for the intended purpose.

Datasets that represent 3D environmental variation are a relatively recent addition to ecological modelling toolsets, but are now available for a range of environments. NOAA’s 2013 World Ocean Atlas (WOA13; Locarnini et al. 2013, Zweng et al. 2013 Garcia et al. 2014a,b) is a prime example of such a dataset with global coverage of marine environments. The WOA13 comprises multiple 2D layers (1° × 1° resolution) representing environmental variables (temperature, salinity, oxygen, and nutrients) from a range of depths (0–5500 m), which constitutes an approximate representation of 3D space (i.e. 2.5D). Using these layers in combination with the ‘depth’ field of occurrence data provides a means to ensure that environmental predictors used for subsequent modelling are representative of the habitat where species occurrence is recorded.

Many marine occurrence records held in repositories, such as GBIF, lack associated depth data and must, therefore, be excluded from 2.5D or 3D analyses. Although this reduces the number of usable occurrence records and could potentially increase geographic bias, in explicitly 3D environments records without a z coordinate (e.g. depth or altitude) are comparable to data lacking geographic coordinates in a 2D environment. Thus, we suggest that occurrence records lacking z coordinates should be removed during data cleaning (e.g. occurrence records missing latitude data would not be used to create
SDMs. Ultimately, the decision comes down to whether modelling using fewer data points of enhanced quality is preferable to modelling with more records that are of poor quality (see Lecours et al. 2015 for a recent review of quantity-quality trade-offs in marine habitat mapping).

**ADDING AN EXTRA DIMENSION**

Ideally, SDM methods specifically intended for use in a 3D environment should be developed for use in 3D systems. Minor modification of existing SDM techniques (e.g. MaxEnt; Phillips et al. 2006) can, however, provide an adequate alternative until specific 3D SDMs are widespread (Bentlage et al. 2013). We consider such methods to be 2.5D SDMs, as they rely on multiple 2D layers to approximate 3D systems. For example, a third, temporal dimension is sometimes added to SDM studies through the projection of a SDM trained on current environmental variables onto climate layers representing future climate scenarios (e.g. Elith et al. 2011, Bellard et al. 2013). Although the high likelihood of imperfect detection of marine species distributed across a 3D space and the impacts of other sampling biases should also be considered (Monk 2014), modification of this temporal-layers approach can be used to incorporate depth (or an equivalent third spatial dimension) into SDMs (see Supplement 1 at www.int-res.com/articles/suppl/m564p001_supp/ for details of methods used to create the examples in Figs. 4 & S2, and https://doi.org/10.4225/03/58507ddae3022 for modelling code and an example using simulated species data).

Occurrence data with a depth or other z-dimension field can be used in combination with multiple environmental layers representing each depth or z-dimension layer (e.g. WOA13, discussed above) to create a matrix of occurrence points and their associated environmental data. All environmental data coupled with an occurrence point should be extracted from the most appropriate environmental layer available (e.g. the depth layer closest to occurrence depth) or calculated through interpolation across layers. Such use of 2.5D data for SDM training ensures that models are built using environmental data that better represent each occurrence point (as recommended by Guisan & Thuiller 2005, Reiss et al. 2015), therefore producing a more reliable representation of the species’ realised niche. SDM models for each species can then be projected sequentially onto each depth or z-dimension layer, akin to how models are projected onto time layers in temporal change studies, to estimate environment suitability. The product of this method is multiple habitat-suitability layers at the same intervals of the original environmental data (as in Figs. 4 & S2). Each of these represents a 2D slice of habitat suitability within the 3D environment. When used in combination, these multiple layers can represent habitat suitability and range estimates in 3 dimensions more accurately than if the environment were represented as an individual 2D layer.

Although we acknowledge that indiscriminate cross-taxa modelling with generic SDMs is not recommended practice for a variety of reasons (Guiller-Arroita et al. 2015) and that the likelihood of imperfect detection is particularly high for marine species (Monk 2014), we applied modified SDM methods to illustrate how outcomes may change when using 2.5D approaches (Fig. 4). The influence of the North Atlantic sub-Polar Front is particularly apparent in all 2D SDM estimates with areas of high probability of occurrence tracking the front toward the mid-Atlantic. However, the direct environmental effects of this front do not penetrate into deep water (Vecchione et al. 2015), where the conditions are much more homogenous either side of the frontal zone. A frontal signal in distributions of deep-dwelling fishes is, therefore, contrary to expectations. Incorporating depth to create 2.5D SDMs reduces this frontal signal and generates predicted occurrence estimates that are closer to expectations for deep-dwelling species. Furthermore, a single 2D map fails to capture the intricacies of species that are distributed across 3D space. Improvements in SDM predictive ability are also observed when using a simulated species, which is free from the sampling bias associated with GBIF data (Fig. S2; see Supplement 1 for details of methods used).

**THINKING IN 3 DIMENSIONS**

Beyond SDM approaches, 3D methods can also be incorporated into other means of estimating and representing the ranges of species that inhabit 3D spaces. Simple shape-drawing methods can be extended to incorporate 3D alternatives (i.e. polyhedrons rather than polygons), and 3D analogues of the 2D α-shape method are commonly used in medical imaging to select regions of interest (e.g. Al-Tamimi et al. 2015) and can be readily applied in geographic rather than tomographic space. Although representing the range of one species using either of these methods, or using 2.5D SDM methods (as in Fig. 4), is
relatively straightforward and intuitive, generating richness estimates for 3D ecosystems is more complicated.

Clearly, careful consideration should be given to how the outputs of 2.5D or 3D models are interpreted. Combining multiple 2D species distribution maps to produce a single composite map of species richness simply requires the total number of predicted species per cell to be calculated (i.e. the sum of all species presence-absence maps). Comparable composite maps of species richness can be created for each depth layer of a 2.5D SDM to represent estimated richness geographically at multiple depths, but the way in which multiple depth layers should be combined, or whether they should be combined at all, is open to debate (see especially the discussion of Guillera-Arroita et al. 2015). When creating 2D species richness maps, equal area geographic projections are commonly used (e.g. Tittensor et al. 2010, Bellard et al. 2013) to avoid introducing area effects that arise from comparing 2 map cells of unequal size. It could, therefore, be argued that cells of varying depth, and hence of unequal volume, cannot be directly compared (e.g. a cell with a surface area of 1 km² would have a volume of 0.2 km³ assuming a depth of 200 m or 4 km³ assuming a depth of 4000 m).

Ideally, the effects of topographic and environmental heterogeneity should also be taken into account, but this is theoretically and practically complex in marine environments, particularly as topographic complexity of the benthos varies substantially and is still poorly understood at fine scales in the deep-sea environments that comprise the majority of the marine realm (Lecours et al. 2015). This is highlighted by

Fig. 4. Estimated probability of occurrence for 4 representative fish species, from a range of depths (mean depth of occurrence ± 1 SD shown) in the Northeast Atlantic, modelled using 2D and 2.5D ensemble species distribution modelling methods (Supplement 1). The distribution of occurrence points for each species is shown in 2 dimensions in Fig. S1 in Supplement 1 and rendered in 3 dimensions in the associated HTML file (Supplement 3 at www.int-res.com/articles/suppl/m564p001_supp/)
abyssal hill habitats, where small-scale topographic features provide a hidden source of habitat heterogeneity that has only recently been identified but influences species richness of deep-sea megafauna (Durden et al. 2015).

Extending analyses of marine biodiversity to better incorporate depth could potentially alter current perceptions of global patterns of marine diversity, which thus far have been analysed as flattened 2D environments (as in Tittensor et al. 2010), and improve understanding of marine macroecological patterns. Vertical macroecological patterns, comparable to latitudinal patterns observed in terrestrial systems, are well documented in marine environments (e.g. depth-diversity gradients, Howell et al. 2002, Rex & Etter, 2010), and so the interaction between latitudinal and vertical factors may mean that standard (i.e. 2D) macroecological patterns fail to capture the intricacies of the marine environment. To study latitudinal patterns using existing methods, the 3D environment could be represented by multiple 2D layers (2.5D) with analyses of macroecological patterns applied to each of these layers sequentially. Moving forward, however, the field would perhaps best be served by implementing Bentlage et al.’s (2013) recommendation to focus on the development of ecological models that are explicitly 3D. Such models can better represent organisms distributed within a 3D space, enabling improved conceptualisation and testing of macroecological patterns that incorporate variability across all 3 (x, y, z) geographic dimensions.

Acknowledgements. We thank Bernard Coetzee and 3 anonymous reviewers for their valuable feedback on previous versions of this manuscript. This work was funded by Australian Research Council grant DP140102815.

LITERATURE CITED


Monk J (2014) How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? Fish Fish 15:352–358