



Freer, J. J., Tarling, G., Collins, M., Partridge, J. C., & Genner, M. J. (2020). Estimating circumpolar distributions of lanternfish using 2D and 3D ecological niche models. *Marine Ecology Progress Series*, 647 (2020), 179-193. <https://doi.org/10.3354/meps13384>

Peer reviewed version

Link to published version (if available):
[10.3354/meps13384](https://doi.org/10.3354/meps13384)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Inter-Research Science Publisher at <https://doi.org/10.3354/meps13384> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/>

Estimating circumpolar distributions of lanternfish using 2D and 3D ecological niche models

Running page head: 3D niche models of lanternfish

**Jennifer J. Freer^{1, 2*}, Geraint A. Tarling², Martin A. Collins², Julian C. Partridge³, and
Martin J. Genner¹**

1. University of Bristol, School of Biological Sciences, 24 Tyndall Avenue, Bristol, BS8
1TQ, UK
2. British Antarctic Survey, NERC, High Cross, Madingley Road, Cambridge, CB3 0ET,
UK
3. University of Western Australia, Oceans Institute, 35 Stirling Highway, WA 6009,
Australia

*Corresponding author: *jen.free@bas.ac.uk*, +44 (0) 7533621000

1 **ABSTRACT**

2 Ecological niche models (ENMs) can be a practical approach for investigating distributions and
3 habitat characteristics of pelagic species. In principle, to reflect a species ecological niche well,
4 ENMs should incorporate environmental predictors that consider its full vertical habitat, yet
5 examples of such models are rare. Here we present the first application of ‘3D’ ENMs to ten
6 Southern Ocean lanternfish species. This 3D approach incorporates depth-specific environmental
7 predictor data to identify the distribution of suitable habitat across multiple depth levels. Results
8 were compared to those from the more common ‘2D’ approach, which uses only environmental
9 data from the sea surface. Measures of model discriminatory ability and overfitting indicated that
10 2D models often outperform 3D methods, even when accounting for reduced available sample
11 size in the 3D models. Nevertheless, models for species with a known affinity for deeper habitat
12 benefitted from the 3D approach, and our results suggest that species can track their ecological
13 niche in latitude and depth leading to equatorward or poleward range extensions beyond that
14 expected from incorporating only surface data. However, since 3D models require
15 comprehensive depth-specific data, both data availability and the need for depth-specific model
16 outputs must be considered when choosing the appropriate modelling approach. We advocate
17 increased effort to include depth-resolved environmental parameters within marine ENMs. This
18 will require collection of mesopelagic species occurrence data using appropriate temporal and
19 depth stratified methods, and inclusion of accurate depth information when occurrence records
20 are submitted to global biodiversity databases.

21 **Keywords:** 3D modelling, ecological niche model, species distribution model, biogeography,
22 lanternfish, Myctophidae, mesopelagic, Southern Ocean

1 **1. INTRODUCTION**

2 From the mid-20th century, the development of methods to sample organisms from discrete
3 depths using midwater trawls (Isaacs & Kidd 1951, Clarke 1969, Baker et al. 1973) quickly
4 helped to revolutionise mesopelagic biogeography. This is because studies were at last able to
5 describe both fine-scale latitudinal and vertical patterns in species distributions, and begin to
6 understand the processes maintaining them (Sutton 2013). Increased access to these records via
7 online global biodiversity databases has considerably enhanced our understanding of marine
8 species distributions, and specifically the data have found an important role in underpinning
9 modelling approaches such as Ecological Niche Modelling or Species Distribution Modelling. In
10 general, these methods correlate records of species presence with biologically relevant
11 environmental variables enabling predictions of the probability of occurrence of suitable habitat
12 for a species across a region of interest (Peterson et al. 2011a). Examples of such methods being
13 applied to mesopelagic fauna range from helmet jellyfish (Bentlage et al. 2013), giant squid
14 (Coro et al. 2015), snipe eels (DeVaney 2016), and lanternfishes of the family Myctophidae
15 (Loots et al. 2007, Flynn & Marshall 2013, Duhamel et al. 2014, Freer et al. 2018).

16 Ecological Niche Models (ENMs) of mesopelagic fishes have demonstrated that subsurface data
17 (often conditions at 200m or deeper) may be better predictors of distribution patterns than
18 surface environmental conditions (Duhamel et al. 2000, Loots et al. 2007, Koubbi et al. 2011b).
19 However, in marine ENMs, it is often the case that surface (or near-surface) environmental
20 variables are paired with occurrence records of animals that were caught from a range of depths,
21 despite the surface and deeper waters having very different environmental conditions. In such
22 cases the model may provide a poor representation of the normal environment of a species, and

1 therefore their physical and ecological niche, ultimately leading to inaccurate predictions of
2 suitable habitat (Duffy & Chown 2017).

3 A common method of including environmental information from multiple depths into an ENM is
4 to treat data from each depth layer as a separate environmental predictor, for example, using both
5 sea surface temperature and sea bottom temperature. Although such ENMs do not rely on surface
6 data alone, they remain relatively simplistic (Duffy & Chown 2017) and do not fully account for
7 variability in the association between depth and species occurrence throughout the pelagic
8 environment, which plausibly may be relevant to species with pelagic life history stages.

9 Other methods exist to integrate a third (i.e. vertical/depth) dimension into ENMs such as the
10 method described by Duffy and Chown (2017). This approach uses species occurrence records
11 and their associated *latitude*, *longitude*, and *depth* information, matching these records with
12 environmental variables from a similar depth from which the animals were caught. For example,
13 a record caught at 250 m would be matched to the temperature at 250 m (or the closest available
14 depth level) rather than using temperature at the surface or any other depth level. As depth is
15 considered as discrete (i.e. discontinuous) levels rather than as a continuous variable, a
16 distinction can be made between this modelling approach and truly three-dimensional models in
17 which environmental data and species' depths of occurrence would be known in more detail.
18 Nevertheless, this approach can be used to create separate habitat suitability maps for each of the
19 depth 'slices' used and, as only the most appropriate environmental data are paired with each
20 occurrence record, the accuracy of the ENM can, in principle, be improved.

21 However, applying a vertical dimension in this way is rare (but see Bentlage et al., 2013) and
22 models of deep sea species are often constrained by issues of low sampling effort, which can

1 lead to problems of small sample sizes, spatial sampling bias, and/or a poor representation of
2 their environmental or geographic distribution (Robinson et al. 2011, Bentlage et al. 2013). Sub-
3 selecting only those occurrence records that contain associated depth information may
4 exacerbate these issues, meaning that trade-offs between data quantity and quality likely exist
5 when choosing between 2D and 3D approaches (Duffy & Chown 2017).

6 Lanternfishes (Myctophidae) are the dominant family of offshore fish in the Southern Ocean in
7 terms of biomass, abundance and diversity (Barrera-Oro 2002). As such they are key consumers
8 of zooplankton and krill (Saunders et al. 2015, Saunders et al. 2019), and are major components
9 of the diets of predators such as birds, seals and fish (Guinet et al. 1996, Olsson & North 1997,
10 Casaux et al. 1998, Collins et al. 2007, Cherel et al. 2008, Casaux et al. 2009). Through their diel
11 vertical migration behaviour, they also have a key role in the active transport of carbon to deeper
12 waters (Collins et al. 2012). Whilst their latitudinal biogeography is well understood through
13 large-scale modelling efforts (Loots et al. 2007, Koubbi et al. 2011b, Duhamel et al. 2014, Freer
14 et al. 2019), their vertical distribution patterns remain less clear. The vertical habitat use of
15 Southern Ocean lanternfish have been described from regional sampling efforts (Lancraft et al.
16 1989, Duhamel et al. 2000, Pusch et al. 2004, Collins et al. 2008, Hulley & Duhamel 2011,
17 Koubbi et al. 2011a, Collins et al. 2012), yet vertical migration behaviour and a capability to
18 evade net capture means that sampling can be prone to false negative results (Kaartvedt et al.
19 2012). Thus, 3D modelling approaches have the potential to bring novel insights to vertical
20 distribution patterns and the complex use of the environment by these species.

21 In this study, we aim to investigate the three-dimensional distribution of ten Southern Ocean
22 lanternfish species by building depth-specific ENMs following the method of Duffy and Chown

1 (2017). We also compare the predictive performance of models and the spatial overlap of 3D
2 ENMs to those built using the less complex ‘2D’ approach. Therefore, as well as improving our
3 understanding of these species’ vertical habitat, we aim to evaluate the relative utility of 2D and
4 3D approaches when modelling suitable habitat of mesopelagic species.

5 **2. MATERIALS AND METHODS**

6 **2.1. Species occurrence records**

7 Occurrence records of the family Myctophidae were downloaded from the Global Biodiversity
8 Information Facility (GBIF; <http://www.gbif.org/>). The ten species with the highest number of
9 records in the Southern Ocean were retained for analyses, these being *Electrona antarctica*,
10 *Electrona carlsbergi*, *Gymnoscopelus bolini*, *Gymnoscopelus braueri*, *Gymnoscopelus fraseri*,
11 *Gymnoscopelus nicholsi*, *Gymnoscopelus opisthopterus*, *Krefftichthys anderssoni*,
12 *Protomyctophum bolini*, and *Protomyctophum tenisoni*. All occurrence records were subject to
13 quality assurance and control processes. Unreliable data, which included records with identical
14 latitude and longitude, and records with a latitude and longitude corresponding to a terrestrial
15 location, were removed. Due to high sampling bias towards austral spring and summer seasons,
16 only records from the months October-March were kept for analyses. Furthermore, only
17 occurrence records from 1960 to 2010 were retained, to correspond to a similar baseline period
18 of available environmental data. Along with latitude and longitude, depth-of-catch information
19 was retained for each occurrence record, obtained from the “depth” field of the GBIF database,
20 when this was available. It is possible that some records reflect only the maximum depth that the
21 net reached, and not the actual depth that the fish was caught. Hence, only records falling within
22 the expected vertical bounds for these fishes (i.e. 0-1500 m) were retained.

2.2. Environmental predictors

Five environmental predictors were selected on which to build the ENMs. These comprised temperature, salinity, dissolved oxygen, primary productivity and bathymetry, and were chosen based on their physiological importance for marine ectotherms and on previous results demonstrating their importance for determining marine species distributions (Loots et al. 2007, Koubbi et al. 2011b, Flynn & Marshall 2013, Duhamel et al. 2014). Climatological means for temperature, oxygen and salinity predictors were extracted from the World Ocean Atlas 2013 database (Locarnini et al. 2013, Zweng et al. 2013, Garcia et al. 2014a, b) at a resolution of $0.25^\circ \times 0.25^\circ$ (~27.75km at the equator) for the months October-March across the baseline temporal period 1956-2005. These data were extracted from seven vertical depth layers: sea surface, 50 m, 100 m, 200 m, 500 m, 800 m, and 1000 m with the greater resolution at shallow depths representing increased variability in conditions within the epipelagic zone. Primary productivity data represent the primary organic carbon production by all types of phytoplankton ($\text{nmol m}^{-2} \text{s}^{-1}$) in the upper 150 m for the months October-March between 1956-2005. Bathymetric data correspond to maximum water depth and had an original spatial resolution of 30 arc seconds (Becker et al. 2009), being re-sampled to the same resolution as the other variables (i.e. $0.25^\circ \times 0.25^\circ$) using the bi-linear resample tool in ArcGIS v.10.5.1 (ESRI, Redlands, California). All data were delimited to 35-75°S as this region encompasses the known geographic extent of these species (Duhamel et al. 2014) and environmental data south of 75°S are often missing or imprecise.

We acknowledge that some environmental predictors (e.g. oxygen and temperature) are highly correlated, having Pearson's $r > 0.9$ (Table S1) and that including correlated predictors can make

1 it difficult to assess the relative importance of each due to issues of collinearity. Nevertheless,
2 there is evidence to suggest that, when dealing with correlated variables that are each
3 biologically meaningful, including all predictors can produce models with better predictive
4 performance, in addition to a better fit, than a model parameterized on only one of the correlated
5 predictors (Braunisch et al., 2013). MaxEnt (Phillips et al. 2006, Phillips et al. 2017), the
6 maximum entropy modelling approach we use here, is (i) particularly effective in dealing with
7 collinearity through its iterative model fitting approach which can consider variables
8 independently, (ii) can include non-linear interactions between variables, and (iii) has a robust
9 ability to rank variables according to their importance (Braunisch et al. 2013).

10 **2.3. Building 2D and 3D models**

11 The 2D ENMs were built with no reliance on depth-of-catch information from occurrence
12 records. The five environmental predictors (sea surface temperature, sea surface salinity, sea
13 surface dissolved oxygen, primary productivity, and bathymetry) were matched to the longitude
14 and latitude fields of each occurrence record. These models, that include all available species
15 presence data, are hereafter referred to as “2D_{all}” models.

16 As many occurrence records were missing the depth-of-catch information required for the 3D
17 models. To enable direct comparisons between 2D and 3D models based upon the same sample
18 size, a “2D_{sub}” model was fitted for each species, which only used the occurrence records that
19 are available for the 3D models.

20 To build the 3D models, occurrence records with associated depth-of-catch information were
21 used in combination with the five environmental predictors from each of the seven depth layers;

1 0 m, 50 m, 100 m, 200 m, 500 m, 800 m and 1000 m. Data were extracted from each
2 environmental predictor raster and were matched to occurrence records based upon their
3 longitude, latitude and depth fields, following the methodology of Duffy and Chown (2017).
4 Specifically, the depth field from each occurrence record was used to identify the closest depth
5 layer from which to extract the most appropriate environmental data. Primary productivity data
6 was assigned to all occurrences regardless of vertical position. The resulting ENMs were then
7 projected spatially onto each of the aforementioned depth layers. This resulted in habitat
8 suitability maps for each modelled species and each depth layer, each representing a two-
9 dimensional slice within a three-dimensional environment.

10 **2.4.MaxEnt ecological niche models**

11 For each species, occurrence and environmental data were fitted to the presence-only ecological
12 niche modelling algorithm MaxEnt v. 3.4.1. MaxEnt estimates the conditional probability of
13 presence of a species relative to locations where the species has been observed by sampling the
14 environment at a range of locations across the study region (“background sites”) and
15 discriminating these from locations and environments where species are known to be present
16 (“presence sites”). MaxEnt assumes background locations adequately cover areas accessible to
17 the species and that presence localities are unbiased and cover important environmental gradients
18 (Jarnevich et al. 2015). While a lack of absence data prevents probability estimates of a species
19 presence and predictions of species’ realised distributions, presence-only outputs can be
20 interpreted as showing the existing, fundamental niche and the potential distribution of a species
21 (*sensu* Hutchinson, 1957) (Peterson et al. 2011b). MaxEnt was chosen for its repeatedly high
22 performance against other ENM algorithms (Elith et al. 2006, Ortega-Huerta & Peterson 2008,

1 Monk et al. 2010). Moreover, MaxEnt’s capacity to use presence-only data is particularly
2 appropriate for mesopelagic species given the high potential for errors under a presence-absence
3 approach due to the low sampling effort relative to the potential habitat area (or volume)
4 available, as well as the net avoidance behaviour common among lanternfishes (Collins et al.
5 2008, Kaartvedt et al. 2012). Using an ensemble of model algorithms rather than relying on a
6 single approach can be the best methodological practice when the aim is to gain robust
7 predictions of a species distribution (Araujo & New 2007, Araujo et al. 2019). However, as the
8 main aim of this work was to understand the appropriateness and utility of 3D models in
9 comparison to traditional 2D methods, we give emphasis to this methodological comparison
10 rather than comparing between algorithms.

11 All ENMs were fitted using the “SDMtune” R package (Vignali et al. 2019). Occurrence data
12 were partitioned into “calibration” and “evaluation” data using 3 sub-samples, with 30% of data
13 used for model evaluation each time. Only linear, quadratic and hinge feature classes were
14 selected in order to avoid fitting overly complex responses (Elith et al. 2010). 10,000 background
15 data points were selected from within 2 decimal degrees of all mesopelagic fish records within
16 the study region. This ensures both the background and presence sites have the same spatial and
17 environmental bias (Phillips et al. 2009). Similarly, for 3D models the depth distribution of
18 background points reflected the same depth sampling bias as occurrence records. 5000 model
19 iterations were used and all other MaxEnt settings were kept as default.

20 **2.5.Comparing 2D and 3D modelling approaches**

21 The outputs of the *2Dall*, *2Dsub* and 3D modelling approaches were compared in terms of their
22 discriminatory ability (AUC_{TEST} , TSS), calibration ($AICc$) and overfitting (AUC_{DIFF}).

1 Discriminatory ability was determined by the Area Under the receiver operating characteristic
2 Curve (AUC) calculated on the evaluation data (AUC_{TEST}). The AUC score is a widely used
3 rank-based measure of predictive accuracy that can be interpreted in the context of MaxEnt as
4 the probability that a randomly chosen presence location is ranked higher than a randomly
5 chosen background point (Merow et al. 2013). A model with no discriminatory power will have
6 an AUC value equal to 0.5 (no better than random) whilst a model with perfect fit would have an
7 AUC value of 1.0. Discriminatory accuracy of each model was also measured using the True
8 Skill Statistic (TSS) (Allouche et al. 2006). TSS values range from -1 to 1 with 0 reflecting a
9 model that is no different than random and values closer to 1 are better at discerning presence
10 and background points. Overfitting was measured using AUC_{DIFF} (Warren & Seifert 2011, Bohl
11 et al. 2019), the difference between the AUCs calculated with calibration records (AUC_{TRAIN})
12 and evaluation records (AUC_{TEST}). This metric is based on the notion that overfitted models
13 generally perform well on training data but poorly on test data and will therefore yield relatively
14 high AUC_{DIFF} values. Lastly, Akaike's Information Criterion corrected for small sample size
15 (AICc) was used as a metric of calibration performance as it has been demonstrated that
16 information criterion based approaches to model selection may be particularly useful when
17 sample sizes are small (Warren & Seifert 2011, Lawson et al. 2014).

18 We also compared how model outputs differed across environmental space by reporting the
19 permutation importance of each environmental predictor for *2Dall*, *2Dsub* and 3D model
20 outputs. Response curves, which describe the relationship between a modelled species
21 occurrence and an individual environmental predictor, were also compared between all model
22 outputs. Finally, the niche overlap between the *2Dall* or *2Dsub* output and each of the seven
23 depth predictions of the 3D model (e.g. *2Dall* vs 3D 0 m, and *2Dall* vs 3D 50 m etc.) were

1 calculated using the “niche overlap” tool in the software ENMTools v.1.3 (Warren et al. 2010).
2 Both the Spearman rank correlation coefficient and *I* metrics are reported for comparison. The *I*
3 metric reflects the similarity of environmental suitability predicted by the ENMs yet tends to
4 overestimate similarity when many grid cells are of similar values. Rank correlation is more
5 reflective of similarity between ENMs in their underlying response to environmental gradients.

6 To quantify how model outputs differed geographically, range overlap scores were also
7 calculated between the *2Dall* or *2Dsub* output and each of the seven depth predictions of the 3D
8 model using ENMTools. This requires a threshold value over which a species is classified as
9 present rather than using a continuous score between 0 and 1. To transform outputs into binary
10 maps we used the “Equal test sensitivity plus specificity” threshold criteria. Selecting threshold
11 criteria is somewhat arbitrary and each approach can generate different results. Our selection was
12 based on the results of Liu et al. (2016) who found that sensitivity-specificity approaches
13 performed best compared to others.

14 **3. RESULTS**

15 **3.1. Comparing model predictive performance**

16 A total of 2918 unique occurrence records were used in analyses. For five out of the ten species
17 (*Electrona carlsbergi*, *Gymnoscopelus bolini*, *Gymnoscopelus fraseri*, *Gymnoscopelus*
18 *opisthopterus*, *Protomyctophum tenisoni*), the total number of occurrence records available for
19 the 3D approach fell below 90 records per species due to absence of depth information for the
20 occurrence records. Depending on species, between 21.4 and 74.5% of records did not have
21 associated depth information (Table 1).

1 Highest scoring metrics of model performance varied greatly between species and model
2 approaches (Table 1). In comparing overall performance between 2Dall and 3D models, 2Dall
3 models had, on average, higher AUC_{TEST} and lower AUC_{DIFF} scores, whilst 3D models had
4 higher TSS and lower AICc scores (Table 1). Three of the four metrics were stronger in the
5 2Dall approach for *G. fraseri*, *G. nicholsi* and *Kreftichthys anderssoni*, whilst the opposite was
6 found for *E. antarctica*, *G. braueri* and *G. opisthopterus*. In comparing performance between
7 2Dsub and 3D models, AUC_{TEST}, AUC_{DIFF} and TSS scores were all stronger, on average, for
8 2Dsub approach (Table 1). Whilst the 2Dsub approach yielded higher performing metrics for
9 most species, indicators of discrimination accuracy (AUC_{TEST} and TSS) remained higher under
10 the 3D approach for *E. antarctica* and *G. opisthopterus*.

11 **3.2.Comparing predictions in environmental space**

12 Temperature was the variable with highest permutation importance for *E. antarctica*, *G. bolini*,
13 *G. braueri*, *G. fraseri*, *G. opisthopterus*, and *K. anderssoni* regardless of modelling approach.
14 2Dall and 3D approaches both found primary productivity to be the most important variable for
15 *E. carlsbergi* whilst bathymetry was highest for *G. nicholsi* under both 2Dsub and 3D models
16 (Table S2).

17 Comparing the niche overlap between 2Dall and 3D model predictions, we find that overlap
18 values are high across all species, though relative rank values were lower and more variable
19 (mean \pm 1SD across all comparisons = 0.8 ± 0.04) than the *I* similarity metric (mean \pm 1SD
20 across all comparisons = 0.93 ± 0.01 ; Table 2). Per species, niche overlap values remain stable
21 across the 3D model predictions at different depth bands and often peak at mid depths (100-500

1 m; Table 2). A similar pattern is found when comparing the 2D_{sub} and 3D model predictions
2 (Table S3).

3 Response curves describing the relationship between a modelled species' occurrence and an
4 individual environmental predictor were compared across the modelling approaches (Figure S11-
5 41). The 3D models demonstrate a slightly lower optimal temperature and narrower thermal
6 tolerance ranges. Most species were predicted to tolerate a broader range of salinity and
7 dissolved oxygen under the 3D models.

8 **3.3. Comparing predictions in geographic space**

9 Examples of suitable habitat predicted by the different model approaches are shown for *G.*
10 *opisthopterus* (Figure 1) and *G. fraseri* (Figure 2) while all other outputs are given in Figures S1-
11 10. Comparisons of geographic range overlap between 2D_{dall} and 3D model outputs varied
12 considerably though were found to have higher overlap in the upper 200 m depth bands than at
13 deeper depths (Table 2). Spatial overlap was consistently high at all depth bands for *E.*
14 *antarctica*, *G. opisthopterus* and *G. braueri* (Table 2). A similar pattern is found when
15 comparing the 2D_{sub} with 3D model predictions, and overlap values are slightly elevated in
16 these comparisons (Table S3).

17 For some species, the 3D outputs reveal vertical distribution features that are not resolved by the
18 2D_{dall} or 2D_{sub} approaches. For example, *E. antarctica* (Figure S1) is estimated to have greater
19 suitable habitat in the upper 200 m, whilst *G. opisthopterus* is estimated to have greater suitable
20 habitat below 500 m (Figure 1). The 3D outputs also suggest that the habitat suitability of some
21 species, notably *E. carlsbergi*, *G. bolini*, *G. fraseri* and *K. anderssoni*, extends equatorward at

1 deeper depths (Figure 2 and Figures S2, 3, 8, 9). *P. tenisoni* had the highest loss of occurrence
2 records and the resulting habitat suitability maps under the 3D approach have low estimates of
3 presence throughout the study region compared to the circumpolar prediction under the 2D
4 approach (Figure S10).

5 **4. DISCUSSION**

6 In this study we present results of Southern Ocean lanternfish distributions obtained from both
7 simplistic (2D_{all}, 2D_{sub}) and depth-integrated (3D) ecological niche models. In comparing
8 metrics of model performance, and how outputs differ in geographic and environmental niche
9 space, we have gained valuable knowledge of the benefits and trade-offs presented by each
10 approach and how to interpret the ecological information they provide.

11 **4.1. Distribution patterns**

12 Using a variety of model performance metrics, we have shown that, even after accounting for
13 sample size differences, 2D ENMs can perform better than 3D ENMs for the same species. For
14 certain species, this suggests that surface data alone can largely determine the abiotic drivers of
15 their fundamental geographic distributions. This is also reflected in the high niche overlap
16 between modelling approaches, and the similar, broad scale patterns in habitat suitability they
17 generate. Given that many myctophids and other mesopelagic species spend time feeding in
18 surface zones, and that their prey depend upon surface productivity, high performance of 2D
19 ENMs in comparison to 3D methods could be anticipated and is reassuring for previous ENM
20 predictions that have not used fully depth resolved methods for these species (Loots et al. 2007,
21 Duhamel et al. 2014, Freer et al. 2018), and other mesopelagic fishes (DeVaney 2016).

1 Nevertheless, we also find examples of species whose model performance benefitted from
2 including depth resolved environmental parameters and, for all ten species investigated, the 3D
3 model predictions give novel insights in to their vertical habitat suitability which until now were
4 only described from trawl data covering areas between 100-1000 kms (Lancraft et al. 1989,
5 Duhamel et al. 2000, Pusch et al. 2004, Collins et al. 2008, Hulley & Duhamel 2011, Koubbi et
6 al. 2011a, Collins et al. 2012).

7 The apparent preferences for shallow or deep habitat in our study species largely follow
8 expectations from observed data. For example, the prediction of increased suitability of *E.*
9 *antarctica* habitat in the upper 200 m matches previous observational records (Hulley 1981,
10 Lancraft et al. 1989, Hulley 1990). Suitable habitat for *G. bolini* and *G. nicholsi*, both
11 benthopelagic as adults, is increased along shelf regions at depths below 200 m and between
12 100-500 m, respectively. This reflects patterns from catch data at similar depths (Duhamel et al.
13 2014). *G. braueri* and *K. anderssoni* display suitable habitat throughout the water column, which
14 may reflect the extensive vertical migrations known in these species (Collins et al. 2012,
15 Duhamel et al. 2014). As adults, *G. opisthopterus* is known to inhabit deeper (> 600 m) waters
16 associated with eastern slope regions of Kerguelen (Hulley & Duhamel 2011), the continental
17 slope of Antarctica (Koubbi et al. 2011a), and southern Scotia Sea (Collins et al. 2012) which
18 reflects our findings of suitable habitat throughout these shelf and slope regions at depths below
19 500 m. The suitable habitat predicted in shallow water regions may be explained by the
20 upwelling of deeper water masses towards the surface in these areas (Hulley & Duhamel 2011).

21 Overall, there was a tendency for 3D model predictions at shallow depths to identify suitable
22 habitat further south than the *2Dall* or *2Dsub* model predictions, particularly noticeable at the

1 polar front boundary within the Scotia Sea. This is likely due to the 3D models occupying niche
2 space with lower optimal temperature and higher optimal salinity than the equivalent 2D models,
3 as identified by model response curves. Despite this, the 2D and 3D approaches were found to
4 have high overlap in environmental niche space across all depth predictions but their overlap in
5 geographic space generally declined at deeper depths. This highlights the potential for 3D
6 models to reveal how the latitudinal habitat of a species may change with depth. For example,
7 comparisons amongst 3D model predictions at different depths show that suitable habitat for *E.*
8 *carlsbergi* extends further south at shallow depths and is extended equatorward at deeper layers.
9 Similar patterns are found for *G. bolini*, *K. anderssoni* and *G. fraseri*, which all have extended
10 equatorward distributions that reach into the region of the Subtropical Convergence at depths
11 greater than 500 m. This suggests that these species may follow Antarctic Intermediate Water as
12 it moves from shallow depths in the Antarctic Convergence to deep waters around the
13 Subtropical Convergence (Pardo et al. 2012). Such a hypothesis was first put forward by Hulley
14 (1981) following the disparate vertical ranges of these species from different latitudes.

15 3D model outputs for some species also suggest suitable habitat in areas previously unrecorded
16 or thought to be unsuitable based upon the outputs of the 2D models. For example, *P. bolini* is
17 known to have an oceanic distribution that is bounded to the south by the South Antarctic
18 Circumpolar Current Front (SACCF) (Saunders et al. 2019), and is generally absent inshore of
19 shelf breaks (Duhamel et al. 2014). Results from the 3D model would suggest that regions of the
20 Western Antarctic Peninsula and continental slope approximately 90-140°E are also potential
21 suitable habitats, not dissimilar from the results of Duhamel et al. (2014) that also predict
22 suitable habitat south of the SACCF.

1 **4.2. Challenges of modelling in three dimensions**

2 Integrating a vertical dimension into an ENM, as the 3D models do here, brings with it additional
3 challenges. Diel vertical migrations, by their nature, will result in differences in recorded depth
4 between daytime and night-time sampling efforts (Robison 2003). Extreme seasonal changes in
5 light conditions, as is found in polar regions, may also alter the pattern, timing, or extent of
6 vertical migrations (Cisewski et al. 2010, Dypvik et al. 2012). Diel, seasonal and ontogenetic
7 variability in depth distributions are confounded by net avoidance issues as species are more
8 likely to evade nets deployed above 400 m in daylight resulting in higher catch rates at night
9 (Collins et al. 2012). Overall, light conditions at the time of sampling likely influence the depth
10 at which a fish is caught, which can in turn can affect the vertical component of a model that has
11 incorporated this information. Authors should acknowledge that the vertical distribution patterns
12 described by a 3D niche model are specific to a particular set of diel and seasonal light
13 conditions.

14 To control for these issues of vertical sampling bias, data from a wide range of diel and temporal
15 sampling times are required. One solution is to have sampling protocols that are standardised
16 across day/night cycles and multiple depth bands. The need for systematic sampling across
17 environmental and spatial gradients echoes previous calls for well-designed surveys to obtain the
18 high quality data needed for better performing ENMs (Fei & Yu 2016, Leroy et al. 2018, Araujo
19 et al. 2019). For mesopelagic data, factoring in an additional temporal dimension is necessary to
20 generate a more informative sample of presences, as well as non-detections, which would refine
21 our ability to predict and interpret patterns of vertical habitat use in light of species' diel
22 behaviour.

1 **4.3. Data quantity versus quality**

2 Since many occurrence records do not have associated depth data, they must be excluded from
3 3D ENMs (Duffy & Chown 2017), or else alternative solutions sought, for example it may be
4 possible to assign assumed depths to occurrence records through a hierarchical modelling
5 approach with certain expectations about the distribution-at-depth of the missing data. In the
6 examples considered in our study, the sample size for half the species of 3D ENMs dropped to
7 fewer than 90 occurrence records, and to only 35 records in the case of *P. tenisoni*. This dropout
8 highlights the need to consider the availability of depth-specific presence data when choosing the
9 most appropriate modelling approach for marine species.

10 ENMs built using a small sample size tend to be less powerful because their parameter estimates
11 have higher uncertainty, outliers have a higher weight, and there may not be enough records to
12 represent the complexity of a species' ecological niche comprehensively (Wisz et al. 2008).
13 Previous studies have explored the performance of multiple ENM algorithms with different
14 sample sizes. These studies found that, whilst MaxEnt can be less sensitive to changes in sample
15 size than other algorithms (Hernandez et al. 2006, Wisz et al. 2008), (i) depending on the type of
16 data available, a minimum of 50-100 occurrences are needed to characterise a species niche
17 (Meynard et al. 2019), though some studies have produced useful models with as few as 10
18 records (Stockwell & Peterson 2002, Soultan & Safi 2017); (ii) model performance tends to
19 increase with additional presences but can plateau after reaching a certain threshold where
20 additional records add little to model accuracy (Stockwell & Peterson 2002); and (iii)
21 characteristics of a species e.g. niche breadth, specialisation (Mateo et al. 2010, Soultan & Safi
22 2017), the size of the study area (van Proosdij et al. 2016) can influence minimum prior

1 information needed about a species. Taken together, these studies are a strong indicator that low
2 sample size can affect model performance and may help to explain the subtle differences we
3 observed between the performance of *2Dall*, *2Dsub* and 3D models presented here.

4 That our results have signs of better fit with larger sample size, as measured by the AUC_{DIFF}
5 metric, is consistent with this literature. However, the AUC_{TEST} (Lobo et al. 2008) and recently
6 the TSS (Leroy et al. 2018) metrics have been shown to be prevalence-dependent; i.e. they are
7 dependent on the proportion of the data representing species presence (Phillips et al. 2006, Raes
8 & ter Steege 2007). When there is a strong imbalance between presence and absence, the model
9 is very likely to have a higher probability of occurrence at a random presence point than at a
10 random absence point, resulting in an AUC value that is falsely inflated by statistical artefacts
11 caused by the lower sample size (McPherson et al. 2004, van Proosdij et al. 2016). Here, several
12 models show slightly higher AUC_{TEST} and TSS values under the *2Dsub* and 3D approaches,
13 which had lower sample size, and thus lower prevalence, than comparative *2Dall* models. Thus,
14 it is unclear whether these indicators of performance are a statistical artefact, or a true reflection
15 of a better model performance. Nevertheless, these metrics are able to give an indication of
16 model discriminatory ability and, used in tandem with other methods, such as niche and range
17 overlap tools, the ecological realism of model outputs can be compared (Fourcade et al. 2014).

18 Data quality can be determined by several characteristics of the presence data, including the
19 reliability of species identifications, spatial accuracy, and the degree to which they represent the
20 true distribution of the species (Kadmon et al. 2003). By utilising depth-of-catch information, the
21 3D method employed here matches presence data to abiotic values, which more accurately
22 reflect the conditions at the depth at which they were caught. In doing so, 3D methods

1 theoretically provide an improvement in the spatial accuracy of the presence data and also in the
2 completeness of the representation of the environmental niche provided by the model training
3 data, which has been shown to be important for obtaining reliable models with good calibration
4 (Jimenez-Valverde et al. 2009, Meynard et al. 2019). Our results demonstrate that this may be
5 particularly relevant for species that are known to prefer deep ocean environments. *G.*
6 *opisthopterus* can be characterised as being a deeper living (400-1000 m), bathypelagic species
7 (Hulley & Duhamel 2011, Saunders et al. 2019). Performance metrics agree that the 3D model
8 for this species is better performing than 2D models, supporting the idea that, in some cases, the
9 quality of ecological information provided by the presence (and absence/pseudo-absence) data
10 can be more important than their quantity (Mateo et al. 2010).

11 At depths below the surface, environmental climatologies, particularly for biochemical variables
12 such as dissolved organic carbon or chlorophyll, which are often derived from satellite products,
13 are not as readily available as in surface layers. Global ocean models can provide estimates of
14 carbon and nutrient parameters subject to their own model-based uncertainties (Allen et al. 2007,
15 Aumont et al. 2015), yet overcoming the limitations of creating depth-resolved predictor
16 information will also be important to address the uptake, quality, and performance of future 3D
17 ENM techniques.

18 In reality, both the quantity and quality of data will influence model predictions, alongside
19 sampling biases (Fei & Yu 2016) and the distribution characteristics of the modelled species
20 (Kadmon et al. 2003). The choice of the most appropriate modelling approach will depend upon
21 all of these factors and the intended use of the model. While a large, representative sample size
22 should be favoured in studies interested in defining environmental conditions of a species or

1 complex conservation applications, a smaller sample size may be tolerated when the aims of an
2 ENM are to identify potential sampling locations or to explore macroecological patterns of
3 poorly known regions or taxa (Wisz et al. 2008, Soutan & Safi 2017).

4 Here we have shown that 3D approaches can give results that are insightful and comparable to
5 the more simplistic 2D models for some species. However, the utility of these 3D approaches are
6 limited by the paucity of depth-of-catch data associated with online-sourced occurrence records,
7 at least for the species considered here. We emphasise the need for mesopelagic species data to
8 be collected using appropriate temporal and depth stratified methods, and for providers to upload
9 full and accurate records of occurrence, including depth-of-catch, when submitting data to global
10 biodiversity databases such as GBIF. This should also include the type of net used, deployment
11 method (e.g. oblique or stratified depth), and non-detections of species, as this information gives
12 researchers a fuller understanding of the data and its limitations. Future studies aiming
13 specifically to integrate a third dimension into ENMs should consider the data quality, quantity,
14 the ecological characteristics of the species, and the objectives of the study to clarify the
15 suitability of such methods.

16 **4.3. Concluding remarks**

17 We have found that 3D ENMs developed for mesopelagic species have a higher and more
18 consistent overlap to 2D ENMs in environmental space than geographic space. However, this
19 rarely resulted in differences to inferred broad-scale patterns of habitat suitability. 3D models
20 may outperform models based on two dimensional approaches depending on the metrics used
21 and species of interest, yet trade-offs are required between the quality and quantity of occurrence
22 records which will determine the subsequent choice of ENM (2D or 3D). In particular, the

1 paucity and added uncertainty of depth-of-catch data may limit the widespread use of 3D if this
2 is not corrected through an increased effort of appropriate collection and documentation. 3D
3 models could be a valuable addition to the researcher's toolbox when investigating deep pelagic
4 biogeography and we advocate their continued use and development, where data are appropriate.

5 *Data archive.* All occurrence, background, and associated environmental data utilised in this
6 study can be accessed at the following DOI: [insert data DOI]

7 *Acknowledgements.* This work was supported by a Natural Environment Research Council
8 studentship to JJF [NE/L002434/1]. We acknowledge G. Duffy and S. Chown for providing their
9 reproducible code which helped to match occurrence data to the correct vertical environmental
10 layer. We extend our appreciation to all those who have contributed to and maintained the data
11 used within GBIF, and to the anonymous reviewers for their insightful comments which greatly
12 improved this manuscript.

13 **LITERATURE CITED**

14 Allen JI, Somerfield PJ, Gilbert FJ (2007) Quantifying uncertainty in high-resolution coupled
15 hydrodynamic-ecosystem models. *Journal of Marine Systems* 64:3-14

16 Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models:
17 prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223-1232

18 Araujo MB, Anderson RP, Barbosa AM, Beale CM, Dormann CF, Early R, Garcia RA, Guisan
19 A, Maiorano L, Naimi B, O'Hara RB, Zimmermann NE, Rahbek C (2019) Standards for
20 distribution models in biodiversity assessments. *Science Advances* 5

- 1 Araujo MB, New M (2007) Ensemble forecasting of species distributions. Trends in ecology &
2 evolution 22:42-47
- 3 Aumont O, Ethe C, Tagliabue A, Bopp L, Gehlen M (2015) PISCES-v2: an ocean
4 biogeochemical model for carbon and ecosystem studies. Geoscientific Model
5 Development 8:2465-2513
- 6 Baker ADC, Clarke MR, Harris MJ (1973) The NIO combination net (RMT 1+8) and further
7 developments of rectangular midwater trawls. J Mar Biol Assoc UK 53:167-184
- 8 Barrera-Oro E (2002) The role of fish in the Antarctic marine food web: differences between
9 inshore and offshore waters in the southern Scotia Arc and west Antarctic Peninsula.
10 Antarct Sci 14:293-309
- 11 Becker JJ, Sandwell DT, Smith WHF, Braud J, Binder B, Depner J, Fabre D, Factor J, Ingalls S,
12 Kim SH, Ladner R, Marks K, Nelson S, Pharaoh A, Trimmer R, Von Rosenberg J,
13 Wallace G, Weatherall P (2009) Global bathymetry and elevation data at 30 arc seconds
14 resolution: SRTM30_PLUS. Marine Geodesy 32:355-371
- 15 Bentlage B, Peterson AT, Barve N, Cartwright P (2013) Plumbing the depths: extending
16 ecological niche modelling and species distribution modelling in three dimensions.
17 Global Ecol Biogeogr 22:952-961
- 18 Bohl CL, Kass JM, Anderson RP (2019) A new null model approach to quantify performance
19 and significance for ecological niche models of species distributions. J Biogeogr 0

- 1 Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, Bollmann K (2013) Selecting from
2 correlated climate variables: a major source of uncertainty for predicting species
3 distributions under climate change. *Ecography* 36:971-983
- 4 Casaux R, Baroni A, Ramon A, Carlini A, Bertolin M, DiPrinzio CY (2009) Diet of the Leopard
5 Seal *Hydrurga leptonyx* at the Danco Coast, Antarctic Peninsula. *Polar Biol* 32:307-310
- 6 Casaux R, Soave G, Coria N (1998) Lanternfish (Myctophidae) in the diet of the Cape petrel
7 *Daption capense* at the South Shetland and South Orkney Islands, Antarctica. *Polar Biol*
8 20:364-366
- 9 Cherel Y, Ducatez S, Fontaine C, Richard P, Guinet C (2008) Stable isotopes reveal the trophic
10 position and mesopelagic fish diet of female southern elephant seals breeding on the
11 Kerguelen Islands. *Mar Ecol Prog Ser* 370:239-247
- 12 Cisewski B, Strass VH, Rhein M, Kragefsky S (2010) Seasonal variation of diel vertical
13 migration of zooplankton from ADCP backscatter time series data in the Lazarev Sea,
14 Antarctica. *Deep Sea Res Part I Oceanogr Res Pap* 57:78-94
- 15 Clarke MR (1969) A new midwater trawl for sampling discrete depth horizons. *J Mar Biol Assoc*
16 UK 49:945-960
- 17 Collins MA, Ross KA, Belchier M, Reid K (2007) Distribution and diet of juvenile Patagonian
18 toothfish on the South Georgia and Shag Rocks shelves (Southern Ocean). *Mar Biol*
19 152:135-147

1 Collins MA, Stowasser G, Fielding S, Shreeve R, Xavier JC, Venables HJ, Enderlein P, Cherel
2 Y, Van de Putte A (2012) Latitudinal and bathymetric patterns in the distribution and
3 abundance of mesopelagic fish in the Scotia Sea. *Deep Sea Res Part II Top Stud*
4 *Oceanogr* 59:189-198

5 Collins MA, Xavier JC, Johnston NM, North AW, Enderlein P, Tarling GA, Waluda CM,
6 Hawker EJ, Cunningham NJ (2008) Patterns in the distribution of myctophid fish in the
7 northern Scotia Sea ecosystem. *Polar Biol* 31:837-851

8 Coro G, Magliozzi C, Ellenbroek A, Pagano P (2015) Improving data quality to build a robust
9 distribution model for *Architeuthis dux*. *Ecol Model* 305:29-39

10 DeVaney SC (2016) Species distribution modeling of deep pelagic eels. *Integr Comp Biol*
11 56:524-530

12 Duffy GA, Chown SL (2017) Explicitly integrating a third dimension in marine species
13 distribution modelling. *Mar Ecol Prog Ser* 564:1-8

14 Duhamel G, Hulley PA, Causse R, Koubbi P, Vacchi M, Pruvost P, Vigetta S, Irisson J,
15 Mormede S, Belchier M, Dettai A, Detrich HW, Gutt J, Jones CD, Kock KH, Abellan L,
16 Van de Putte AP (2014) Chapter 7: biogeographic patterns of fish. In: De Broyer C,
17 Koubbi P, Griffiths HJ, Raymond B, d'Udekem dA, C. , Van de Putte AP, Danis B,
18 David B, Grant S, Gutt J, Held C, Hosie G, Huettmann F, Post A, Ropert-Coudert Y (eds)
19 *Biogeographic Atlas of the Southern Ocean*. Scientific Committee on Antarctic Research,
20 Cambridge, United Kingdom

- 1 Duhamel G, Koubbi P, Ravier C (2000) Day and night mesopelagic fish assemblages off the
2 Kerguelen Islands (Southern Ocean). *Polar Biol* 23:106-112
- 3 Dypvik E, Rostad A, Kaartvedt S (2012) Seasonal variations in vertical migration of glacier
4 lanternfish, *Benthosema glaciale*. *Mar Biol* 159:1673-1683
- 5 Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F,
6 Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C,
7 Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K,
8 Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE
9 (2006) Novel methods improve prediction of species' distributions from occurrence data.
10 *Ecography* 29:129-151
- 11 Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods Ecol*
12 *Evol* 1:330-342
- 13 Fei SL, Yu F (2016) Quality of presence data determines species distribution model
14 performance: a novel index to evaluate data quality. *Landscape Ecol* 31:31-42
- 15 Flynn AJ, Marshall NJ (2013) Lanternfish (Myctophidae) zoogeography off eastern Australia: a
16 comparison with physicochemical biogeography. *Plos One* 8:e80950
- 17 Fourcade Y, Engler JO, Roedder D, Secondi J (2014) Mapping species distributions with Maxent
18 using a geographically biased sample of presence data: a performance assessment of
19 methods for correcting sampling bias. *Plos One* 9:e97122

- 1 Freer JJ, Tarling GA, Collins M, Partridge JC, Genner MJ (2019) Predicting future distributions
2 of lanternfish, a significant ecological resource within the Southern Ocean. *Divers Distrib*
- 3 Freer JJ, Tarling GA, Collins MA, Partridge JC, Genner MJ (2018) Predicting ecological
4 responses in a changing ocean: The effects of future climate uncertainty. *Mar Biol* 165:7
- 5 Garcia H, Locarnini RA, Boyer TP, Antonov JI, Baranova OK, Zweng MM, Reagan JR, Johnson
6 DR (2014a) World Ocean Atlas 2013, volume 3: dissolved oxygen, apparent oxygen
7 utilization, and oxygen saturation. In: Levitus S, Mishonov A (eds) NOAA Atlas
8 NESDIS 75
- 9 Garcia H, Locarnini RA, Boyer TP, Antonov JI, Baranova OK, Zweng MM, Reagan JR, Johnson
10 DR (2014b) World Ocean Atlas 2013, volume 4: dissolved inorganic nutrients
11 (phosphate, nitrate, silicate). In: Levitus S, Mishonov A (eds) NOAA Atlas NESDIS 76
- 12 Guinet C, Cherel Y, Ridoux V, Jouventin P (1996) Consumption of marine resources by seabirds
13 and seals in Crozet and Kerguelen waters: changes in relation to consumer biomass 1962-
14 65. *Antarct Sci* 8:23-30
- 15 Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species
16 characteristics on performance of different species distribution modeling methods.
17 *Ecography* 29:773-785
- 18 Hulley PA (1981) Results of the research cruises of FRV "Walther Herwig" to South-America.
19 LVIII. Family Myctophidae (Osteichthyes, Myctophiformes). *Archiv Für*
20 *Fischereiwissenschaft* 31:1-300

- 1 Hulley PA (1990) Family Myctophidae. In: Gon O, Heemstra PC (eds) Fishes of the Southern
2 Ocean. Smith Institute of Ichthyology, Grahamstown, JLB, South Africa
- 3 Hulley PA, Duhamel G (2011) Aspects of lanternfish distribution in the Kerguelen Plateau
4 region. In: Welsford D, Duhamel G (eds) The Kerguelen Plateau: marine ecosystem and
5 fisheries. Société Française d'Ichtyologie, Paris, France
- 6 Hutchinson GE (1957) Population studies - animal ecology and demography - concluding
7 remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415-427
- 8 Isaacs JE, Kidd LW (1951) Isaacs-Kidd midwater trawl. Final Report 1, SIO Ref 53-3. Scripps
9 Institute of Oceanography, University of California
- 10 Jarnevich CS, Stohlgren TJ, Kumar S, Morisette JT, Holcombe TR (2015) Caveats for
11 correlative species distribution modeling. Ecol Inform 29:6-15
- 12 Jimenez-Valverde A, Lobo JM, Hortal J (2009) The effect of prevalence and its interaction with
13 sample size on the reliability of species distribution models. Community Ecology 10:196-
14 205
- 15 Kaartvedt S, Staby A, Aksnes DL (2012) Efficient trawl avoidance by mesopelagic fishes causes
16 large underestimation of their biomass. Mar Ecol Prog Ser 456:1-6
- 17 Kadmon R, Farber O, Danin A (2003) A systematic analysis of factors affecting the performance
18 of climatic envelope models. Ecol Appl 13:853-867

- 1 Koubbi P, Hulley PA, Pruvost P, Henri P, Labat JP, Wadley V, Hirano D, Moteki M (2011a)
2 Size distribution of meso- and bathypelagic fish in the Dumont d'Urville Sea (East
3 Antarctica) during the CEAMARC surveys. *Polar Sci* 5:195-210
- 4 Koubbi P, Moteki M, Duhamel G, Goarant A, Hulley P-A, O'Driscoll R, Ishimaru T, Pruvost P,
5 Tavernier E, Hosie G (2011b) Ecoregionalization of myctophid fish in the Indian sector
6 of the Southern Ocean: results from generalized dissimilarity models. *Deep Sea Res Part*
7 *II Top Stud Oceanogr* 58:170-180
- 8 Lancraft TM, Torres JJ, Hopkins TL (1989) Micronekton and macrozooplankton in the open
9 waters near Antarctic ice edge zones (AMERIEZ 1983 and 1986). *Polar Biol* 9:225-233
- 10 Lawson CR, Hodgson JA, Wilson RJ, Richards SA (2014) Prevalence, thresholds and the
11 performance of presence-absence models. *Methods Ecol Evol* 5:54-64
- 12 Leroy B, Delsol R, Hugueny B, Meynard CN, Barhoumi C, Barbet-Massin M, Bellard C (2018)
13 Without quality presence-absence data, discrimination metrics such as TSS can be
14 misleading measures of model performance. *J Biogeogr* 45:1994-2002
- 15 Liu CR, Newell G, White M (2016) On the selection of thresholds for predicting species
16 occurrence with presence-only data. *Ecol Evol* 6:337-348
- 17 Lobo JM, Jimenez-Valverde A, Real R (2008) AUC: a misleading measure of the performance
18 of predictive distribution models. *Global Ecol Biogeogr* 17:145-151
- 19 Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia H, Baranova OK, Zweng MM,
20 Paver CR, Reagan JR, Johnson DR, Hamilton M, Seidov D (2013) *World Ocean Atlas*

- 1 2013, Volume 1: Temperature. In: Levitus S, Mishonov A (eds) NOAA Atlas NESDIS
2 73
- 3 Loots C, Koubbi P, Duhamel G (2007) Habitat modelling of *Electrona antarctica* (Myctophidae,
4 Pisces) in Kerguelen by generalized additive models and geographic information
5 systems. Polar Biol 30:951-959
- 6 Mateo RG, Felicísimo AM, Muñoz J (2010) Effects of the number of presences on reliability and
7 stability of MARS species distribution models: the importance of regional niche variation
8 and ecological heterogeneity. Journal of Vegetation Science 21:908-922
- 9 McPherson JM, Jetz W, Rogers DJ (2004) The effects of species' range sizes on the accuracy of
10 distribution models: ecological phenomenon or statistical artefact? J Appl Ecol 41:811-
11 823
- 12 Merow C, Smith MJ, Silander JA, Jr. (2013) A practical guide to MaxEnt for modeling species'
13 distributions: what it does, and why inputs and settings matter. Ecography 36:1058-1069
- 14 Meynard CN, Leroy B, Kaplan DM (2019) Testing methods in species distribution modelling
15 using virtual species: what have we learnt and what are we missing? Ecography 42:2021-
16 2036
- 17 Monk J, Ierodiaconou D, Versace VL, Bellgrove A, Harvey E, Rattray A, Laurenson L, Quinn
18 GP (2010) Habitat suitability for marine fishes using presence-only modelling and
19 multibeam sonar. Mar Ecol Prog Ser 420:157-174

- 1 Olsson O, North AW (1997) Diet of the king penguin *Aptenodytes patagonicus* during three
2 summers at South Georgia. *Ibis* 139:504-512
- 3 Orsi AH, Harris U (2001, updated 2015) Locations of the various fronts in the Southern Ocean.
4 Australian Antarctic Data Centre - CAASM Metadata
- 5 Ortega-Huerta MA, Peterson AT (2008) Modeling ecological niches and predicting geographic
6 distributions: a test of six presence-only methods. *Revista Mexicana De Biodiversidad*
7 79:205-216
- 8 Pardo PC, Perez FF, Velo A, Gilcoto M (2012) Water masses distribution in the Southern Ocean:
9 improvement of an extended OMP (eOMP) analysis. *Prog Oceanogr* 103:92-105
- 10 Peterson A, Soberon J, Pearson RG, Anderson R, Martinez-Meyer E, Nakamura M, Araujo MB
11 (2011a) Concepts of niches. In: Peterson A, Soberon J, Pearson RG, Anderson R,
12 Martinez-Meyer E, Nakamura M, Araujo MB (eds) *Ecological Niches and Geographic*
13 *Distributions (MPB-49)*. Princeton University Press, Princeton, NJ, USA and Oxford,
14 United Kingdom
- 15 Peterson A, Soberon J, Pearson RG, Anderson R, Martinez-Meyer E, Nakamura M, Araujo MB
16 (2011b) Niches and geographic distributions. In: Peterson A, Soberon J, Pearson RG,
17 Anderson R, Martinez-Meyer E, Nakamura M, Araujo MB (eds) *Ecological Niches and*
18 *Geographic Distributions (MPB-49)*. Princeton University Press, Princeton, NJ, USA and
19 Oxford, United Kingdom
- 20 Phillips SJ, Anderson RP, Dudik M, Schapire RE, Blair ME (2017) Opening the black box: an
21 open-source release of Maxent. *Ecography* 40:887-893

- 1 Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species
2 geographic distributions. *Ecol Model* 190:231-259
- 3 Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample
4 selection bias and presence-only distribution models: implications for background and
5 pseudo-absence data. *Ecol Appl* 19:181-197
- 6 Pusch C, Hulley PA, Kock KH (2004) Community structure and feeding ecology of mesopelagic
7 fishes in the slope waters of King George Island (South Shetland Islands, Antarctica).
8 *Deep Sea Res Part I Oceanogr Res Pap* 51:1685-1708
- 9 Raes N, ter Steege H (2007) A null-model for significance testing of presence-only species
10 distribution models. *Ecography* 30:727-736
- 11 Robinson LM, Elith J, Hobday AJ, Pearson RG, Kendall BE, Possingham HP, Richardson AJ
12 (2011) Pushing the limits in marine species distribution modelling: lessons from the land
13 present challenges and opportunities. *Global Ecol Biogeogr* 20:789-802
- 14 Robison BH (2003) What drives the diel vertical migrations of Antarctic midwater fish? *J Mar*
15 *Biol Assoc UK* 83:639-642
- 16 Saunders RA, Collins MA, Ward P, Stowasser G, Hill SL, Shreeve R, Tarling GA (2015)
17 Predatory impact of the myctophid fish community on zooplankton in the Scotia Sea
18 (Southern Ocean). *Mar Ecol Prog Ser* 541:45-64

- 1 Saunders RA, Hill SL, Tarling GA, Murphy EJ (2019) Myctophid Fish (Family Myctophidae)
2 Are Central Consumers in the Food Web of the Scotia Sea (Southern Ocean). *Frontiers in*
3 *Marine Science* 6
- 4 Soutan A, Safi K (2017) The interplay of various sources of noise on reliability of species
5 distribution models hinges on ecological specialisation. *Plos One* 12
- 6 Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution
7 models. *Ecol Model* 148:1-13
- 8 Sutton TT (2013) Vertical ecology of the pelagic ocean: classical patterns and new perspectives.
9 *J Fish Biol* 83:1508-1527
- 10 van Proosdij ASJ, Sosef MSM, Wieringa JJ, Raes N (2016) Minimum required number of
11 specimen records to develop accurate species distribution models. *Ecography* 39:542-552
- 12 Vignali S, Barras A, Braunisch V (2019) SDMtune: Species Distribution Model Selection. R
13 package version 1.0.0.
- 14 Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of
15 environmental niche models. *Ecography* 33:607-611
- 16 Warren DL, Seifert SN (2011) Ecological niche modeling in Maxent: the importance of model
17 complexity and the performance of model selection criteria. *Ecol Appl* 21:335-342
- 18 Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, Distribut NPS (2008) Effects
19 of sample size on the performance of species distribution models. *Divers Distrib* 14:763-
20 773

1 Zweng MM, Reagan JR, Antonov JI, Locarnini RA, Mishonov AV, Boyer TP, Garcia H,
 2 Baranova OK, Johnson DR, Seidov D, Biddle MM (2013) World Ocean Atlas 2013,
 3 volume 2: salinity. In: Levitus S, Mishonov A (eds) NOAA Atlas NESDIS 74

4 **Table 1** Comparison of 2Dall, 2Dsub and 3D model performance for each species indicated by
 5 metrics of Area Under the receiver operating Characteristic curves (AUC_{TEST} and AUC_{DIFF}),
 6 True Skill Statistic (TSS) and Akaike’s Information Criterion corrected for small sample size
 7 (AICc). Better performance is indicated by high AUC_{TEST} and TSS scores and low AUC_{DIFF} and
 8 AIC_c scores. The number of occurrence records per species and modelling approach is also
 9 given.

| Species | Sample Size | | 2Dall | | | | 2Dsub | | | | 3D | | | |
|-------------------------|-------------|-----|---------------------|---------------------|-------|-------|---------------------|---------------------|-------|-------|---------------------|---------------------|-------|-------|
| | 2D | 3D | AUC _{test} | AUC _{diff} | TSS | AICc | AUC _{test} | AUC _{diff} | TSS | AICc | AUC _{test} | AUC _{diff} | TSS | AICc |
| <i>E. antarctica</i> | 876 | 688 | 0.833 | 0.007 | 0.514 | 13155 | 0.860 | 0.007 | 0.575 | 10049 | 0.870 | 0.007 | 0.588 | 10239 |
| <i>E. carlsbergi</i> | 141 | 87 | 0.759 | 0.034 | 0.409 | 2392 | 0.754 | 0.058 | 0.459 | 1613 | 0.751 | 0.050 | 0.435 | 2293 |
| <i>G. bolini</i> | 106 | 75 | 0.824 | 0.032 | 0.532 | 1669 | 0.830 | 0.044 | 0.566 | 1427 | 0.782 | 0.053 | 0.482 | 1234 |
| <i>G. braueri</i> | 356 | 242 | 0.745 | 0.033 | 0.387 | 5709 | 0.760 | 0.032 | 0.428 | 3842 | 0.750 | 0.031 | 0.396 | 3933 |
| <i>G. fraseri</i> | 124 | 81 | 0.857 | 0.018 | 0.584 | 2059 | 0.833 | 0.039 | 0.628 | NA | 0.776 | 0.081 | 0.531 | 1791 |
| <i>G. nicholsi</i> | 228 | 151 | 0.850 | 0.020 | 0.548 | 3617 | 0.830 | 0.029 | 0.525 | 2667 | 0.812 | 0.033 | 0.491 | 2456 |
| <i>G. opisthopterus</i> | 152 | 65 | 0.810 | 0.031 | 0.493 | 3358 | 0.864 | 0.017 | 0.596 | 2815 | 0.869 | 0.021 | 0.616 | NA |
| <i>K. anderssoni</i> | 436 | 256 | 0.779 | 0.015 | 0.437 | 6878 | 0.780 | 0.019 | 0.471 | 4068 | 0.731 | 0.030 | 0.370 | 4208 |
| <i>P. bolini</i> | 362 | 135 | 0.802 | 0.015 | 0.461 | 5597 | 0.822 | 0.020 | 0.505 | 2142 | 0.804 | 0.026 | 0.492 | 2339 |
| <i>P. tenisoni</i> | 137 | 35 | 0.877 | 0.015 | 0.624 | 2113 | 0.939 | 0.015 | 0.744 | NA | 0.895 | 0.034 | 0.704 | NA |
| Mean | 292 | 182 | 0.814 | 0.022 | 0.499 | 4655 | 0.827 | 0.028 | 0.550 | 3578 | 0.804 | 0.037 | 0.511 | 3562 |
| ±1 SD | 237 | 193 | 0.043 | 0.010 | 0.076 | 3483 | 0.055 | 0.016 | 0.093 | 2781 | 0.057 | 0.020 | 0.103 | 2879 |

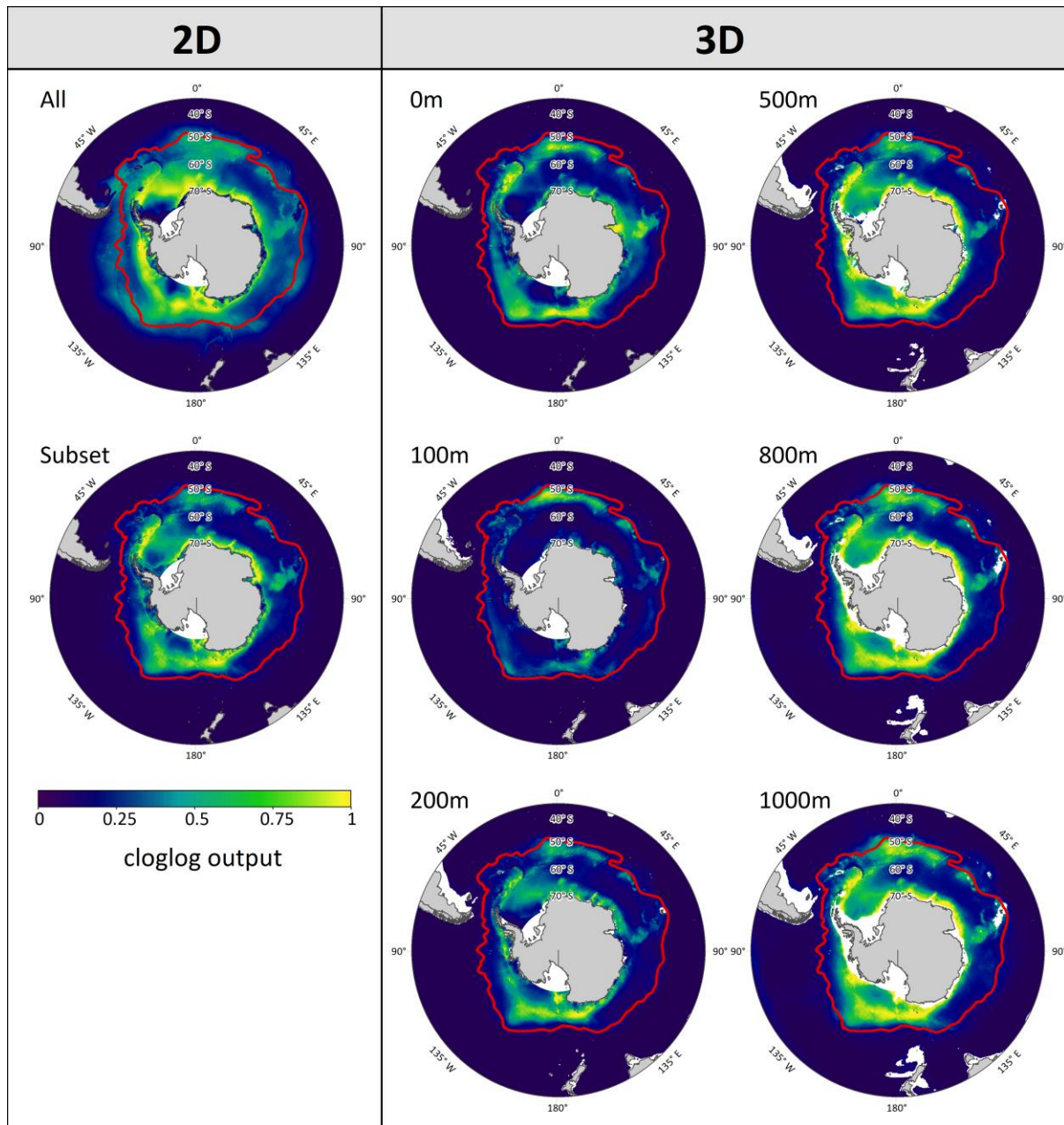
10

11

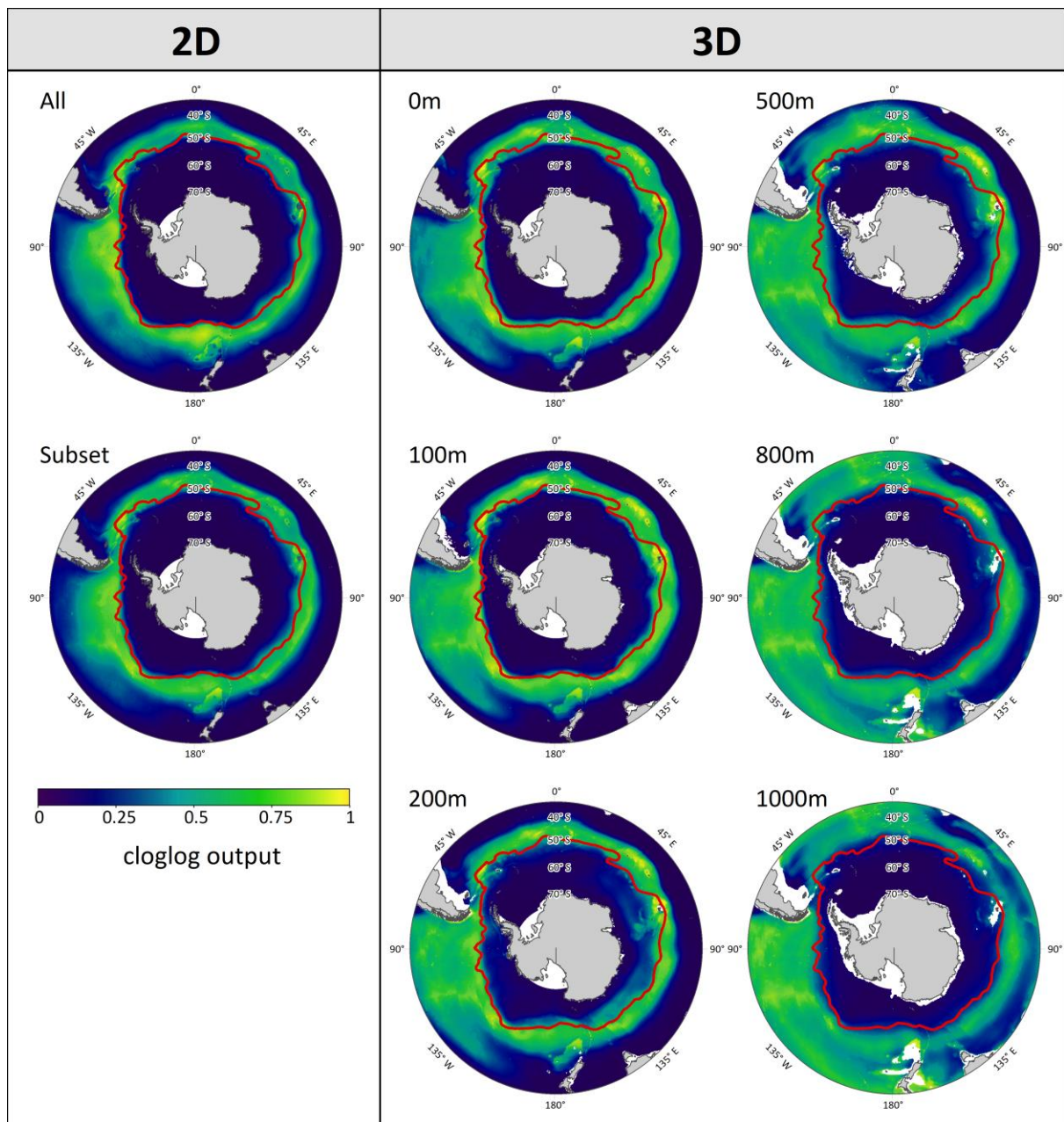
- 1 **Table 2** Comparison of 2Dall 3D model predictions in relation to their spatial range overlap and niche overlap as measured by relative
- 2 rank and *I* metrics. “n records” indicates the number of occurrences available for 3D models. Mean and standard deviation (± 1 SD)
- 3 values for each depth band are also given.

| 3D model depth band (m) | metric | <i>E. antarctica</i> | <i>E. carlsbergi</i> | <i>G. bolini</i> | <i>G. braueri</i> | <i>G. fraseri</i> | <i>G. nicholsi</i> | <i>G. opisthopterus</i> | <i>K. anderssoni</i> | <i>P. bolini</i> | <i>P. tenisoni</i> | Mean | ± 1 SD |
|-------------------------|--------------------------|----------------------|----------------------|------------------|-------------------|-------------------|--------------------|-------------------------|----------------------|------------------|--------------------|------|------------|
| 0 | n records | 106 | 3 | 0 | 24 | 3 | 4 | 7 | 23 | 11 | 2 | 18 | 32 |
| | Range overlap (%) | 0.90 | 0.60 | 0.99 | 0.73 | 0.94 | 0.99 | 0.89 | 0.67 | 0.66 | 0.77 | 0.81 | 0.15 |
| | Niche overlap (rel.rank) | 0.93 | 0.75 | 0.85 | 0.86 | 0.85 | 0.84 | 0.87 | 0.79 | 0.81 | 0.69 | 0.82 | 0.07 |
| | Niche overlap (I) | 0.95 | 0.95 | 0.95 | 0.95 | 0.97 | 0.93 | 0.90 | 0.94 | 0.89 | 0.74 | 0.92 | 0.07 |
| 50 | n records | 73 | 4 | 6 | 14 | 10 | 13 | 0 | 34 | 11 | 1 | 17 | 22 |
| | Range overlap (%) | 0.91 | 0.59 | 0.99 | 0.69 | 0.95 | 0.96 | 0.90 | 0.73 | 0.73 | 0.80 | 0.82 | 0.14 |
| | Niche overlap (rel.rank) | 0.93 | 0.75 | 0.87 | 0.83 | 0.85 | 0.83 | 0.85 | 0.80 | 0.83 | 0.71 | 0.83 | 0.06 |
| | Niche overlap (I) | 0.96 | 0.95 | 0.96 | 0.94 | 0.98 | 0.94 | 0.89 | 0.95 | 0.91 | 0.75 | 0.92 | 0.07 |
| 100 | n records | 69 | 6 | 4 | 16 | 7 | 8 | 2 | 15 | 6 | 3 | 14 | 20 |
| | Range overlap (%) | 0.83 | 0.59 | 0.91 | 0.75 | 0.95 | 0.79 | 0.95 | 0.80 | 0.80 | 0.88 | 0.83 | 0.11 |
| | Niche overlap (rel.rank) | 0.91 | 0.76 | 0.86 | 0.82 | 0.86 | 0.83 | 0.84 | 0.83 | 0.85 | 0.72 | 0.83 | 0.06 |
| | Niche overlap (I) | 0.96 | 0.96 | 0.97 | 0.95 | 0.98 | 0.94 | 0.90 | 0.97 | 0.94 | 0.77 | 0.93 | 0.06 |
| 200 | n records | 192 | 24 | 8 | 39 | 10 | 55 | 1 | 38 | 20 | 3 | 39 | 57 |
| | Range overlap (%) | 0.88 | 0.69 | 0.66 | 0.83 | 0.93 | 0.65 | 0.97 | 0.77 | 0.81 | 0.73 | 0.79 | 0.11 |
| | Niche overlap (rel.rank) | 0.89 | 0.78 | 0.81 | 0.85 | 0.83 | 0.81 | 0.88 | 0.87 | 0.85 | 0.70 | 0.83 | 0.05 |
| | Niche overlap (I) | 0.95 | 0.96 | 0.97 | 0.98 | 0.95 | 0.96 | 0.94 | 0.98 | 0.96 | 0.74 | 0.94 | 0.07 |
| 500 | n records | 113 | 35 | 37 | 60 | 31 | 52 | 12 | 78 | 61 | 25 | 50 | 29 |
| | Range overlap (%) | 0.98 | 0.49 | 0.75 | 0.89 | 0.93 | 0.71 | 0.93 | 0.68 | 0.79 | 0.83 | 0.80 | 0.15 |
| | Niche overlap (rel.rank) | 0.86 | 0.78 | 0.81 | 0.83 | 0.83 | 0.75 | 0.86 | 0.81 | 0.82 | 0.78 | 0.81 | 0.03 |
| | Niche overlap (I) | 0.94 | 0.97 | 0.97 | 0.98 | 0.94 | 0.94 | 0.94 | 0.97 | 0.96 | 0.81 | 0.94 | 0.05 |
| 800 | n records | 96 | 14 | 18 | 66 | 16 | 17 | 34 | 49 | 24 | 1 | 34 | 29 |
| | Range overlap (%) | 0.99 | 0.37 | 0.70 | 0.86 | 0.72 | 0.48 | 0.93 | 0.55 | 0.61 | 0.78 | 0.70 | 0.20 |
| | Niche overlap (rel.rank) | 0.83 | 0.72 | 0.77 | 0.82 | 0.79 | 0.64 | 0.85 | 0.73 | 0.75 | 0.78 | 0.77 | 0.06 |
| | Niche overlap (I) | 0.96 | 0.96 | 0.96 | 0.97 | 0.92 | 0.90 | 0.95 | 0.95 | 0.93 | 0.79 | 0.93 | 0.05 |
| 1000 | n records | 39 | 1 | 2 | 23 | 4 | 2 | 9 | 19 | 2 | 0 | 10 | 13 |
| | Range overlap (%) | 0.98 | 0.34 | 0.59 | 0.81 | 0.47 | 0.38 | 0.92 | 0.31 | 0.49 | 0.69 | 0.60 | 0.24 |
| | Niche overlap (rel.rank) | 0.80 | 0.69 | 0.70 | 0.77 | 0.74 | 0.57 | 0.85 | 0.62 | 0.67 | 0.76 | 0.72 | 0.08 |
| | Niche overlap (I) | 0.92 | 0.95 | 0.94 | 0.94 | 0.89 | 0.88 | 0.96 | 0.92 | 0.90 | 0.75 | 0.91 | 0.06 |

4



14 **Figure 1** Comparison of estimated conditional probability of presence (cloglog output) for
 15 *Gymnoscopelus opisthopterus* using 2Dall (using all available occurrence records and sea
 16 surface environmental data), 2Dsub (as 2Dall, but built using only the same occurrence
 17 records as in the 3D model) and 3D (matches occurrence records with environmental
 18 variables from a similar depth from which the animals were caught) approaches. The red line
 19 denotes the position of the Polar Front from Orsi and Harris (2001, updated 2015).



14 **Figure 2** Comparison of estimated conditional probability of presence for *Gymnoscopelus*
 15 *fraseri* (cloglog output) using 2Dall (using all available occurrence records and sea surface
 16 environmental data), 2Dsub (as 2Dall, but built using only the same occurrence records as in
 17 the 3D model) and 3D (matches occurrence records with environmental variables from a
 18 similar depth from which the animals were caught) approaches. The red line denotes the
 19 position of the Polar Front from Orsi and Harris (2001, updated 2015).