

Supplement of Earth Syst. Sci. Data, 12, 1003–1023, 2020
<https://doi.org/10.5194/essd-12-1003-2020-supplement>
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.



Supplement of

An integrated compilation of data sources for the development of a marine protected area in the Weddell Sea

Katharina Teschke et al.

Correspondence to: Katharina Teschke (katharina.teschke@awi.de)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

Supplementary Material

S1 Data processing and analysis

The workflow of data preparation and cleaning always followed the same pattern. First we downloaded the data from the respective repository or requested the data from contact persons (see details in the main text, Section 2.3.1 to 2.4.5 and Tab. 1, 2). The collected data, which were available in various file formats (e.g., *.csv, *.shp, *.dat, *.hdf), were then checked for quality, i.e. for missing data/information, duplicates and correctness of Lat/Long coordinates. Where data sets contained missing data, "empty" cells were marked with the abbreviation NA (Not Available) and were not used for the subsequent calculations. Duplicates in the data set were deleted. If Lat/Long coordinates were missing these were subsequently added from cruise reports or other sources if possible, otherwise the data entry was also deleted. In addition, Lat/Long coordinates were checked for the correct position. Obviously wrong Lat/Long coordinates (e.g. coordinates on land, in another geographical region, etc.) were corrected along the cruise reports or checked against the principal investigators of the data sets. After the data quality check a shape file was created for each data set, which was then projected onto South Pole Lambert Azimuthal Equal Area (<https://spatialreference.org/ref/esri/102020/>). When working with raster data sets, the default raster cell size was 6.25 km x 6.25 km (raster size of AMSR-E 89 GHz sea ice concentration maps). Finally, the data layers were cut to the size of the study site (see main text, Section 2.1). Our data processing and statistical analyses as well as the map compilation were mainly performed using the R software (Version 3.1.2, R Core Team, 2014), QGIS (Version 2.10 "Pisa", QGIS Development Team, 2015) and the ESRI's ArcGIS desktop software suite (Version 10.2, ESRI 2013).

S1.1 Environmental data (IBCSO data, AMSR-E sea ice maps, FESOM data)

For our pelagic regionalisation analysis (Fig. S2) we focused on the austral summer (December to March). For each raster cell (i) the mean of depth (IBCSO data) and depth range (i.e. difference between maximum and minimum depth), (ii) the relative number of days with ice cover $\leq 70\%$ (AMSR-E sea ice maps) and (iii) the mean of temperature and salinity at the sea surface and the sea bottom (FESOM data) were calculated over the respective time periods of the environmental data sets (detailed description of environmental data sources see main text, Section 2.3, and Tab. 1). The mean of depth and depth range were ln-transformed. The parameters chosen for the pelagic regionalisation analysis are major structuring components of the Weddell Sea ecosystem and are consistent with the variables used by Raymond (2014) in a cluster approach for a circum-antarctic pelagic regionalisation. For clustering, we used *k*-means clustering (Han et al., 2011), the most widely used numerical method for partitioning abiotic and/or biotic data in a predefined number of groups (*k*) (ecological examples from marine realm see e.g., Legendre et al., 2002, Hewitt et al., 2004, Zharikov et al., 2005, Verfaillie et al., 2009). To estimate the optimal number of clusters we used the gap statistic of the R package cluster (Maechler et al., 2014). The first local maximum in the gap statistic was used to define the optimal number of cluster. Due to the large amount of data, the gap statistic could not be applied to the complete data matrix (119,862 samples x 7 variables). Therefore, data subsets were extracted from the complete data matrix using a permutation approach and the gap statistic were applied to each of the data subsets. Finally, the median of the data subsets with respect to the optimal number of clusters was used for *k*-means clustering.

S1.2 Ecological data

S1.2.1 Zooplankton

Antarctic krill (adults)

The habitat suitability model of the Antarctic krill (Fig. S3) was developed with R (R Core Team, 2014) using the biodiversity modelling package *biomod2* (Thuiller et al., 2009 and 2014). *Biomod* is freely available and is probably the best known and most established software in the modelling world of ecologists, geographers and conservationists, combining predictive results from different models (Hao et al., 2018 and references therein).

All models were run with presence-absence data on Antarctic krill (detailed description of data sources see main text, Section 2.4.1, and Tab. 2). The predictor variables used in our final model were defined in a stepwise procedure. First, we fed *biomod2* with more than 20 environmental predictors and the model was run. The relative importance of each variable was evaluated by the following permutation procedure: Once the model is calibrated, a standard prediction is generated. Then, one of the predictor variables is randomised and a new prediction is made. The Pearson's correlation coefficient (r) between that new prediction and the standard prediction is used to measure this variable's relative importance in the model ($= 1 - r$; for more details on the permutation procedure see Thuiller et al., 2012). Variables with low importance were then excluded from the subsequent permutation, and the relative importance in the model of each remaining variable was measured again. Based on this permutation procedure (10 permutations in total) we reduced the number of variables to the most important predictors without negatively influencing the model performance. Thus, for our final predictive model we used the following five environmental variables (ranked by decreasing mean importance value calculated by *biomod2*): (i) dissolved oxygen (WOA13 data), (ii) ice coverage (AMSR-E sea ice maps), (iii) temperature (FESOM data), (iv) bathymetry (IBCSO data) and (v) chlorophyll-a concentration (SeaWiFS data) (detailed description of environmental data sources see main text, Section 2.3, and Tab. 1). All data used in the models came from near the sea surface in austral summer (January to March).

In our modelling approach, we focused on nine commonly used modelling techniques, which include regression, classification and machine learning methods, as described by Elith and Graham (2009): generalised linear model (GLM), generalised boosting model, generalised additive model, classification tree analysis, artificial neural network, surface range envelope, flexible discriminant analysis, multiple adaptive regression splines, random forest). Three evaluation methods, i.e. relative operating characteristic (ROC), true skill statistic (TSS) and accuracy, were used. Each modelling technique was calibrated with 70 % of the data (random sample from the total data set) and the remaining 30 % of the data were used to evaluate their performances (Thuiller, 2003). In total 270 calibrated models (9 different models x 10 replicates of pseudo-absences x 3 evaluation runs) were used for the model synthesis where the different models were combined into a single ensemble model (EM). For the development of our EM, all models were scaled applying a binomial GLM as implemented in *biomod2* to ensure comparable model results. Out of the 270 individual models we selected those models for our EM with a TSS threshold higher than 0.65 (i.e., good prediction accuracy accord to Thuiller et al., 2010). Furthermore, we ground-truthed our EM against krill catch data from CCAMLR (see main text, Section 2.4.1, and Tab. 2) by calculating the percentage of krill catches in the areas with different predicted habitat suitability (high to unsuitable).

Antarctic krill (larvae)

The map of the interpolated abundances of krill larvae in the WSMMPA Planning Area (see Fig. S4) was done with the ArcGIS spatial analyst in the ArcGIS desktop software suite (ESRI Inc., 2011) using the inverse-distance

weighting (IDW) method, one of the most commonly used deterministic models in spatial interpolation (e.g., Lu and Wong, 2008). The interpolation was performed with log-transformed abundance data (detailed description of data sources see main text, Section 2.4.1, Tab. 2). The output cell size (x, y) was set to 1000 m and the distance coefficient power to 2. The interpolated abundances were finally expressed for a radius of 30 km around each data record.

Ice krill

The map of the interpolated abundances of ice krill in the WSMPA Planning Area (Fig. S5) was developed in the same way as the Antarctic krill map (see use of interpolation in previous paragraph). For a detailed description of the data sources see Section 2.4.1 and Table 2 in the main text.

The potential ice krill habitat (Fig. S6) was approximated by water depth from 0 m to 550 m (IBCSO data) and mean sea surface temperature $\leq 0^{\circ}\text{C}$ (FESOM data) (detailed description of environmental data sources see main text, Section 2.3, Tab. 1). The biological characteristics of ice krill were taken from the Biogeographic Atlas of the Southern Ocean (Cuzin-Roudy et al., 2014).

S1.2.2 Zoobenthos

Sponges

The map of the occurrence of sponges in the WSMPA Planning Area (Fig. S7) was finally also generated using the IDW method (see use of interpolation in Section S1.2.1 "Antarctic krill (larvae)"). The previous data processing focused on the consolidation of two different data sets (one quantitative, one semi-quantitative; for detailed description of data sources see main text, Section 2.4.2, and Tab. 2). We transformed the quantitative data into the same four-category system as the semi-quantitative data (i.e. absent, rare, common, very common) by creating a Monte Carlo sample using Sobol low-discrepancy sequences to develop a Weibull distribution ($n = 10,000,000$). Within the Weibull distribution, the following classes were identified (i) class 0 (absent) = 0, (ii) class 1 (rare) = 0 to mean - standard deviation (SD), (iii) class 2 (common) = mean - SD to mean and (iv) class 3 (very common) = mean to mean + SD. The quantitative data were classified according to these classes and merged with the semi-quantitative data. The interpolated data were finally expressed for a 10 nm radius around each data record according to CCAMLR Conservation Measure 22-09 (2012).

Echinoderms

The potential habitat for echinoderms in the WSMPA Planning Area (Fig. S8) was developed with JMP (S.A.S. Institute Inc.) using Ward's (1963) minimum variance method, which has been widely used for calculating distances between clusters since its first description (examples from marine realm see e.g., Verfaillie et al., 2009, Weise et al., 2010, Neukermans et al., 2016). A cluster analysis with a species x station matrix was performed for Asterozoa, Ophiurozoa and Holothurozoa respectively (detailed description of data sources see main text, Section 2.4.2, and Tab. 2). All species occurred only in two stations or less were excluded from the clustering. The results of the cluster analyses were then linked to various environmental data sets. Water temperature best reflected the occurrence of a particular echinoderm community, and therefore their habitat was approximated by bottom water temperature $\leq -1^{\circ}$ (FESOM data).

S1.2.3 Fish

Antarctic silverfish (larvae and adults)

The map of the interpolated abundances of Antarctic silverfish in the WSMPA Planning Area (Fig. S9) was developed using the IDW method (see use of interpolation in Section S1.2.1 "Antarctic krill (larvae)"). For a detailed description of the data sources see Section 2.4.1 and Table 2 in the main text. The interpolated data were finally expressed for a 10 nm radius around each data record according to CCAMLR Conservation Measure 22-09 (2012).

Demersal fish

The habitat suitability model of demersal fish (Fig. S10) was developed with R (R Core Team, 2014) using the biodiversity modelling package *biomod2* (Thuiller et al., 2009 and 2014).

All models were run with presence-absence data on demersal fish (detailed description of data sources see main text, Section 2.4.3, and Tab. 2). The predictor variables that were used for our final predictive model were defined in the same iterative process as described under "Antarctic krill (adults)". The environmental variables finally used for our modelling approach were (with decreasing variable importance): (i) distance to coast, (ii) bathymetry, (iii) calcium carbonate, (iv) broad benthic positioning index, (v) silica, (vi) dissolved oxygen, (vii) biogenic silica, (viii) total organic carbon, (ix) nitrate, (x) salinity, (xi) temperature, (xii) current velocity, (xiii) slope and (xiv) phosphate (detailed description of environmental data sources see Section 2.3 and Tab. 1 in main text). All data used in the models came from the sea bottom in austral summer (January to March). Distance to coast, i.e. the Euclidean distance to the nearest land from each raster cell centroid (cell size: 8.02 km x 8.02 km), was calculated with the GRASS GIS package *v.distance* (Soimasuo et al., 1994) in QGIS 2.10 "Pisa". The coastline derived from IBCSO Version 1.0 DBM (Arndt et al., 2013). Slope and broad scale benthic positioning index (BPI) was also derived from IBCSO and were calculated with the Benthic Terrain Modeler Version 3.0 extension for the ArcGIS desktop software suite (ESRI Inc., 2011). For the calculation of the broad scale BPI, the inner radius was set to 5 km and the outer radius to 125 km according to Jerosch et al. (2016).

In the modelling approach, we focused on the same modelling techniques as described under "Antarctic krill (adults)". ROC and TSS were used as evaluation methods. Each modelling technique was calibrated with 70 % of the data (random sample from the total data set) and the remaining 30 % of the data were used to evaluate their performances (Thuiller, 2003). A total of 135 calibrated models (9 different models x 3 replicates of pseudo-absences x 5 evaluation runs) were used for the EM synthesis where all models were scaled applying a binomial GLM as implemented in *biomod2* to ensure comparable model results. Out of the 135 individual models we selected those models for our EM with a TSS threshold higher than 0.9 (i.e., high or excellent prediction accuracy accord to Thuiller et al., 2010).

Antarctic toothfish (adults)

The probability model of Antarctic toothfish occurrence in the WSMPA Planning Area (Fig. S11) was developed as a function of depth as recommended by the CCAMLR Working Group on Ecosystem Monitoring and Management (WG-EMM). Following analytical steps were performed:

- (i) We calculated the standard descriptive parameters of catch per unit effort (CPUE) data on the Antarctic toothfish (CCAMLR fisheries data) per depth interval i ($mCPUE_i$) with a depth interval width of 100 m

(depth interval mean depth: $0 \text{ m} \leq D_i \leq 2600 \text{ m}$, $D_{i+1} - D_i = 50 \text{ m}$). Depth intervals with less than five CPUE data points were not considered.

- (ii) A Monte Carlo sample was built for each depth interval i ($n = 10,000$) by randomly drawn samples from a log-normal distribution with the same mean and standard deviation as the CPUE data in each depth interval.
- (iii) Outliers were defined as data points below $Q1 - 3.0 \times \text{IQR}$ or above $Q3 + 3.0 \times \text{IQR}$ per depth interval i where $Q1$ and $Q3$ are the 25% and 75% quartiles, respectively, and IQR is the interquartile range, i.e. the difference between $Q1$ and $Q3$. Thus, only extreme data points, that are "far out" (Tukey 1977), were excluded from the subsequent model fit.
- (iv) We fitted a 4 parameter Weibull model to the simulated median $mCPUE_i$ per depth interval i ,

$$mCPUE_i = \text{if}(D_i \leq x_0 - b \times ((c - 1)/c)^{(1/c)}, 0, a \times ((c - 1)/c)^{((1 - c)/c)} \times (\text{abs}((D_i - x_0)/b + ((c - 1)/c)^{(1/c)})^c + (c - 1)/c)$$
The model selection based on R (R Core Team 2014) using the package `fitdistrplus` (Delignette-Muller et al. 2014).
- (v) The median water depth (IBCSO data) was calculated for each raster cell (cell size: 6.25 km x 6.25 km), was assigned to the respective depth interval and the corresponding $mCPUE_i$ value - calculated by the Weibull model for this depth interval - was mapped (detailed description of IBCSO data see main text, Section 2.3, and Tab. 1). Finally, the potential habitat of the Antarctic toothfish was bounded from 550 to 2 000 m according to CCAMLR Conservations Measures and fishing practice as recommended by WG-EMM (WG-EMM-16 report, para. 3.6).

S1.2.4 Flying and non-flying seabirds

Breeding and non-breeding Adélie penguins

The probability of occurrence of breeding and non-breeding Adélie penguins during foraging (Fig. S13, S14) was developed with R (R Core Team, 2014) using the R package `crawl` (Johnson, 2015). The continuous-time correlated random walk model developed by Johnson et al. (2008) has become established in Antarctic science in recent years in order to estimate more accurately the locations of tracked seabirds and pinnipeds along their trajectory (see e.g., Warwick-Evans et al., 2018 and 2019; Baylis et al., 2019).

Here, we used the random walk model to generate predictions of the location of each tracked Adélie individual on an hourly time scale (detailed description of penguin tracking data see main text, Section 2.4.4, and Tab. 2). Raw ARGOS data were first processed by assigning error values to the different ARGOS location quality codes, i.e. location code 3 (= highest accuracy of ARGOS position estimate) was set off against the lowest error value, the highest error was assigned to location code B (= lowest accuracy of ARGOS position estimate). Subsequently, simulated track-lines between the temporally sequenced ARGOS positions or each tracked individual, were generated by the continuous-time correlated random walk model, binned onto a 6.25 km x 6.25 km spatial grid and pooled per raster cell so that the final data layers (one for breeding, one for non-breeding Adélies) identifies the areas that were used most often by tracked Adélies. Buffer areas (i.e. a 50 km buffer and a 50-100 km ring buffer) around each colony - shown on the final map of breeding Adélie penguins - were adopted in accordance with the recommendations of the 2nd international workshop on the identification of CCAMLR MPAs in Planning Domain 1 (WG-EMM-15/42 and references therein).

Breeding Emperor penguins

The probability model of Emperor penguins occurrence during foraging in breeding season (Fig. S15) was developed as a function of distance from colony and colony size (Fretwell et al., 2012 and 2014) as well as sea ice concentration (AMSR-E sea ice maps; detailed description of data source see main text, Section 2.3, and Tab. 1).

Analysis 1: Probability model of penguin occurrence as a function of distance from colony and of colony size

To calculate the distances from colony for foraging, we used a raster grid with a spatial resolution of 6.25 km x 6.25 km (as for sea ice concentration). We calculated the Euclidian distance for each raster pixel centroid j to each emperor penguin breeding colony i . Thus, the probability of occurrence $PI_{i,j}$ of one penguin from colony i in centroid j was calculated by the following approximation:

$$P1_{i,j} = \left(\frac{1}{\sqrt{\pi}}\right) * e^{\left(\frac{-\left(3 * \frac{d_{i,j}}{d_{max}}\right)^2}{2}\right)} \quad (1)$$

where d_{max} is the maximum foraging distance to breeding colony (here $d_{max} = 190$ km; derived from Zimmer et al. (2008) and reference therein by mean maximum foraging distance to the colony of male penguins in winter of 106 km (standard deviation (SD) = 28 km) plus three SD, i.e. 106 km + 3*28km = 190 km), and $d_{i,j}$ is the Euclidean distance (in km) between colony i and centroid j , which was calculated by:

$$d_{i,j} = \left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\right) - d_{ice_edge_i} \quad (2)$$

where $d_{ice_edge_i}$ is the distance of colony to the shelf ice edge. Distances $d_{i,j} \leq 0$ were set to 1. Subsequently, different boundaries of ice shelf edge were adjusted by a 10 km puffer, which was subtracted from the distances $d_{i,j}$, too, and a reclassification was performed again ($d_{i,j} \leq 0$ were set to 1).

Then, the probability of penguin occurrence $PI_{i,j}$ from colony i in centroid j was normalised between 0 and 1 (i.e. $0 \leq PI_{i,j} \leq 1$). Finally, all $PI_{i,j}$ were added for each centroid j and normalised again to a range between 0 and 1:

$$P1_j = \frac{\sum_{i=1}^n P1_{i,j}}{\max(\sum_{i=1}^n P1_{i,j})} \quad (3)$$

where n is the number of emperor penguin breeding colonies.

To account for breeding colony size (number of animals), each probability of penguin occurrence $PI_{i,j}$ was weighted with the best population estimate (BE) for this emperor penguin colony according to Fretwell et al. (2012).

$$P1'_{i,j} = P1_{i,j} * BE_i \quad (4)$$

Subsequently, all $P1'_{i,j}$ were added for each centroid j and normalised to a range between 0 and 1 (i.e. $0 \leq P1'_{i,j} \leq 1$):

$$P1'_j = \frac{\sum_{i=1}^n P1'_{i,j}}{\max(\sum_{i=1}^n P1'_{i,j})} \quad (5)$$

where n is the number of emperor penguin breeding colonies.

Analysis 2: Probability model of penguin occurrence as a function of sea ice concentration

The probability model of penguin occurrence as a function of sea ice concentration was calculated in the following steps: (1) A sigmoid transfer function was applied (eq. 6) to achieve an even distribution of the mean sea ice concentration data; (2) the ice index data (IC_j) were normalised to a range between 0 and 1 (eq. 7); and (3) the probability of penguin occurrence was calculated using the transformed data and a hyperbolic \tanh -function (eq. 8). The mean sea ice concentration was calculated for the breeding period of emperor penguins (June to January) from 2002 to 2011.

$$IC_j = \frac{1}{1+e^{(-\ln(x+10^{-5})*gain)}} \quad (6)$$

with x = mean sea ice concentration/100 and gain set to 6.23.

Subsequently, the ice index data (IC_j) were normalised to a range between 0 and 1:

$$IC_j = std_{IC_j} = \frac{IC_j - \min(IC_{j_1}, IC_{j_2}, \dots, IC_{j_n})}{\max(IC_{j_1}, IC_{j_2}, \dots, IC_{j_n}) - \min(IC_{j_1}, IC_{j_2}, \dots, IC_{j_n})} \quad (7)$$

For the probability model of penguin occurrence we have assumed that the penguin preference does not relate linearly to sea ice conditions, but with a sigmoid pattern, i.e. areas with medium sea ice concentration are already suitable foraging grounds. This sigmoid pattern was modelled by the following \tanh -function:

$$P2_j = \frac{\tanh(\pi*(IC_j*2-1))+1}{2} \quad (8)$$

Analysis 3: Combining the distance/colony size model with the sea ice concentration model

An overall probability of penguin occurrence P_j , i.e. a combination of the distance/colony size model and the sea ice concentration model, was calculated by the following equation:

$$P_j = \frac{(P1_j*P2_j) - \min(P1_j*P2_{j_1}, P1_j*P2_{j_2}, \dots, P1_j*P2_{j_n})}{\max(P1_j*P2_{j_1}, P1_j*P2_{j_2}, \dots, P1_j*P2_{j_n}) - \min(P1_j*P2_{j_1}, P1_j*P2_{j_2}, \dots, P1_j*P2_{j_n})} \quad (9)$$

Antarctic petrel

The potential foraging habitats of the Antarctic petrel (Fig. S16) was developed as a function of (i) sea ice concentration (AMSR-E sea ice maps), (ii) bathymetry (IBCSO data) and (iii) sea water temperature (FESOM data; detailed description of data sources see main text, Section 2.3, and Tab. 1).

As preferred ice regime of the Antarctic petrel we focused on the marginal ice zone, i.e. 15 % - 80 % ice coverage, according to van Franeker (1996) and Ainley et al. (1984, 1994). Data on sea ice concentration were reclassified as first step, i.e. a value of 1 was assigned to each cell with ice cover 15 % - 80 %, whereas cells with ice cover less than 15 % and more than 80 % were set to 0. Then, for each grid cell, the relative number of days (in %) for which a given grid cell had an ice cover between 15 % and 80 % was calculated for the breeding period (January to March) from 2002 to 2011. Subsequently, eight classes regarding the frequency of occurrence of the marginal ice zone were defined and scaled between 0 and 1.

We used abundance data from Ainley and Jacobs (1981) and calculated mean Antarctic petrel densities for three depth classes, i.e. (1) deep ocean: > 2600 m, (2) continental slope and shelf break: 2600 to < 600 m and (3) continental shelf: the remainder of the continental shelf. Then, the mean densities were scaled between 0 and 1. Finally, bathymetric data (IBCSO) were used to identify the three different depth zones in the Weddell Sea Planning Area.

According to Ainley et al. (1984) Antarctic petrels seem to prefer water temperatures colder than 0.5°C. Thus, sea surface temperature (SST) data (FESOM) were reclassified for each raster cell, i.e. value 3 = SST ≤ 0.5°C in all three months (January to March), 2 = SST ≤ 0.5°C in only two months, 1 = SST ≤ 0.5°C in only one month and 0 = SST > 0.5°C in all three months. Subsequently, the values were scaled between 0 and 1.

Finally, we approximated the potential foraging habitat of Antarctic petrel by stacking the three environmental proxies and corresponding data layers, respectively, and assigning different weighting factors to the proxies. The highest weighting factor was assigned to sea ice concentration (weighting factor: 1) as we assume sea ice as the major structuring component of the Antarctic petrel foraging habitat. Bathymetry and sea water temperature, in contrast, got lower weighting factors of 0.75 and 0.25, respectively.

Subsequently, we combined our model approach with the model results from Descamps et al. (2016) as recommended by the CCAMLR Scientific Committee (SC-CAMLR-XXXV report, paras. 5.14 - 5.28). Descamps et al. (2016) kindly provided us with the shape files showing the modelled kernel utilization summer and winter distribution of Antarctic petrel breeding at Svarthamaren.

We combined the kernel utilization distribution (hereafter kernel UD) model from Descamps et al. (2016) with our model by the following procedure:

- (i) We calculated a weighting factor wf_i for each level of kernel UD (i.e. for 30, 60 and 95 % kernel UDs) by the following equation:

$$wf_i = \frac{\max(k_{UD})}{k_{UDi}} \quad (1)$$

where $\max(k_{UD})$ is 30 derived from the 30 % kernel UD, i.e. core area - high intensity of use, and k_{UDi} is the respective kernel UD.

- (ii) We computed the probability of Antarctic petrel occurrence P_i for each grid cell (i) by:

$$P_i = \frac{(x_{iAWI_model} + (100 * w_{fi_Descamp\ et\ al.\ summer}) + (100 * w_{fi_Descamp\ et\ al.\ winter}))}{100 * \max(x_{iAWI_model}, 100 * w_{fi_Descamp\ et\ al.\ summer}, 100 * w_{fi_Descamp\ et\ al.\ winter})} \quad (2)$$

where x_{iAWI_model} is our model value (i.e. 5, 20, 35, 50 or 100).

S1.2.5 Pinnipeds

The probability of pinniped occurrence based on tracking data (Fig. S17) was developed with R (R Core Team, 2014) using the R package *crawl* (Johnson, 2015; see examples of Antarctic studies using *crawl* in Section S1.2.4 "Breeding and non-breeding Adélie penguins").

Here, we used the random walk model from Johnson et al. (2008) to generate 100 simulated track-lines between the temporally successive ARGOS positions for each tracking data set on pinnipeds (detailed description of pinniped tracking data see main text, Section 2.4.5, and Tab. 2). Only random track-lines were generated where the maximum speed of a pinniped between successive positions was ≤ 2.5 m s⁻¹. The simulated track-lines were binned onto our standard spatial grid (cell size: 6.25 km x 6.25 km) and pooled per raster cell so that the final data layer identifies the areas that were used most often by tracked pinnipeds.

The map on seal densities in the WSMPA Planning Area (Fig. S18) was developed combining modelled and interpolated densities of seals. Predictive density values on crabeater seals were derived from Flores et al. (2008) and Forcada et al. (2012) and were pooled in case of areas where both studies presented model results.

Interpolated densities of seals were derived from APIS point data (unspecified taxa) and observed crabeater seal densities (see Bester et al., 1995, 2002; see detailed data description data in main text, Section 2.4.5, and Tab. 2). From APIS point data, seal densities (i.e. individuals/km²) were calculated using the count method for line transect data (e.g., Bester and Odendaal, 2000, Hedley and Buckland, 2004). We used non-standardised data for the density calculations as the APIS data set is based on video material, and thus at least observer related factors potentially influencing the probability of animal detection are not relevant to consider. The seal densities from Bester et al. (1995) were averaged over the different sampling dates for each transect, and the densities per sampling zones (inner, middle, outer zone; see Bester et al., 2002) were converted from square nautical mile to square kilometer. Finally, all transects were subdivided in sections of circa 5.5 km according to Bester et al. (2002) using QGIS 2.0 "Dufour" with the QChainage plugin and the density values of the respective transect was assigned to each section for the interpolation approach. We applied the IDW method (see also Section S1.2.1 "Antarctic krill (larvae)") with the output cell size (x, y) of 2000 m and the distance coefficient power of 2. The search radius setting, i.e. the number of points, was set to 10.

S2 Figures

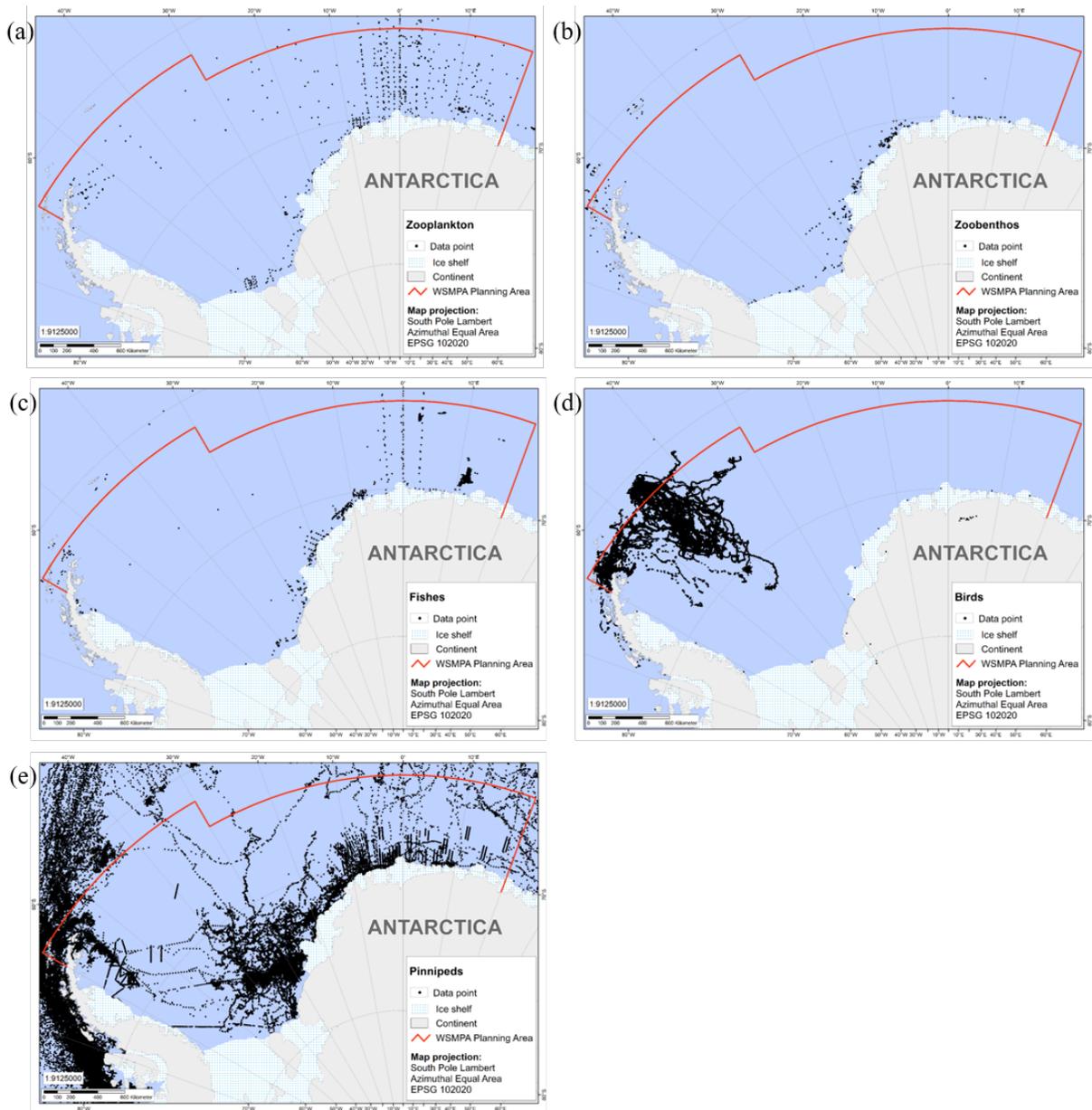


Figure S1. Distribution of data recordings per higher taxonomic group, i.e. zooplankton (a), zoobenthos (b), fishes (c), birds (d) and pinnipeds (e), across the wider Weddell Sea region, which were compiled in the context of the WSMMPA planning initiative.

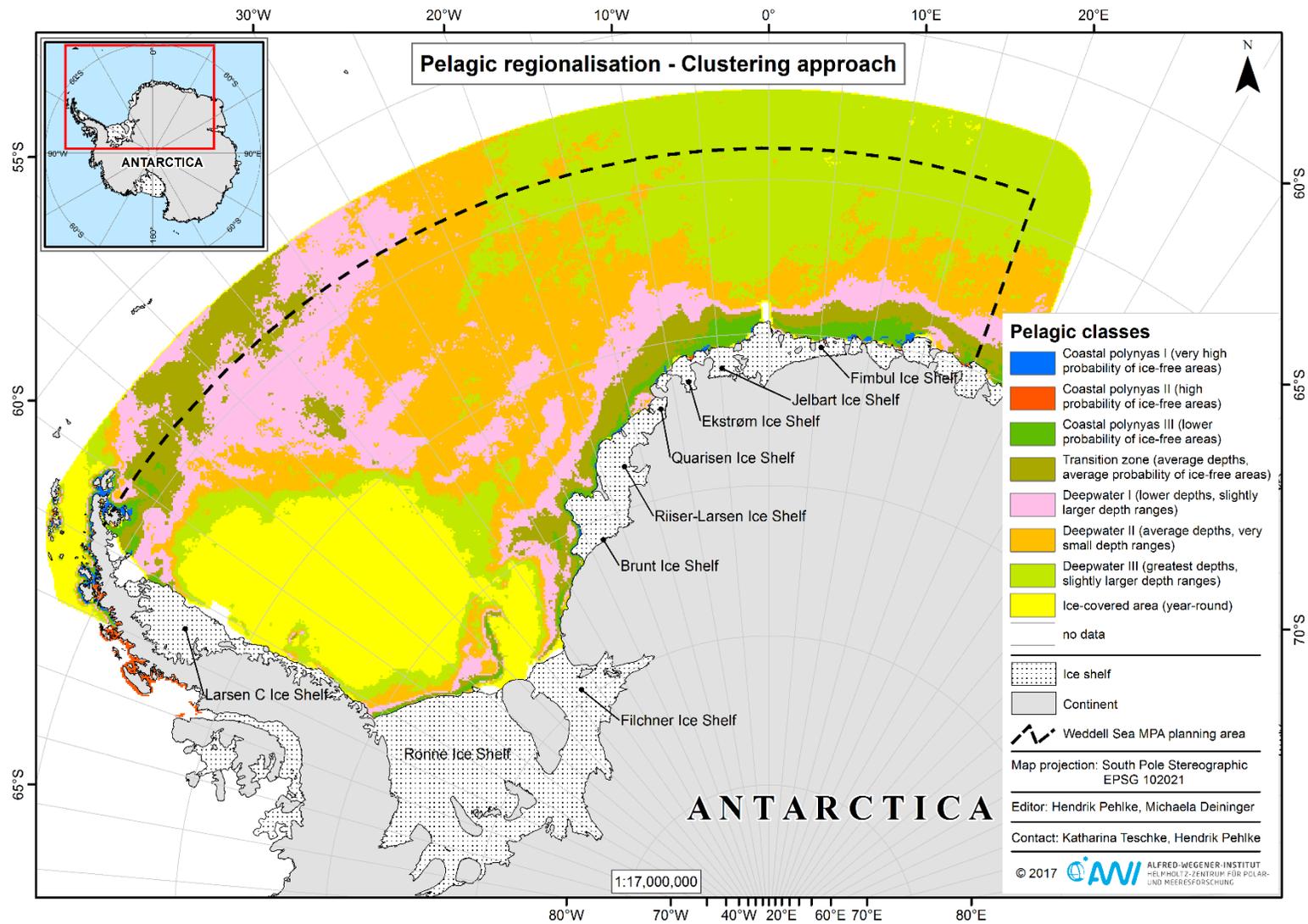


Figure S2. Pelagic regionalisation of the WSMMPA Planning Area.

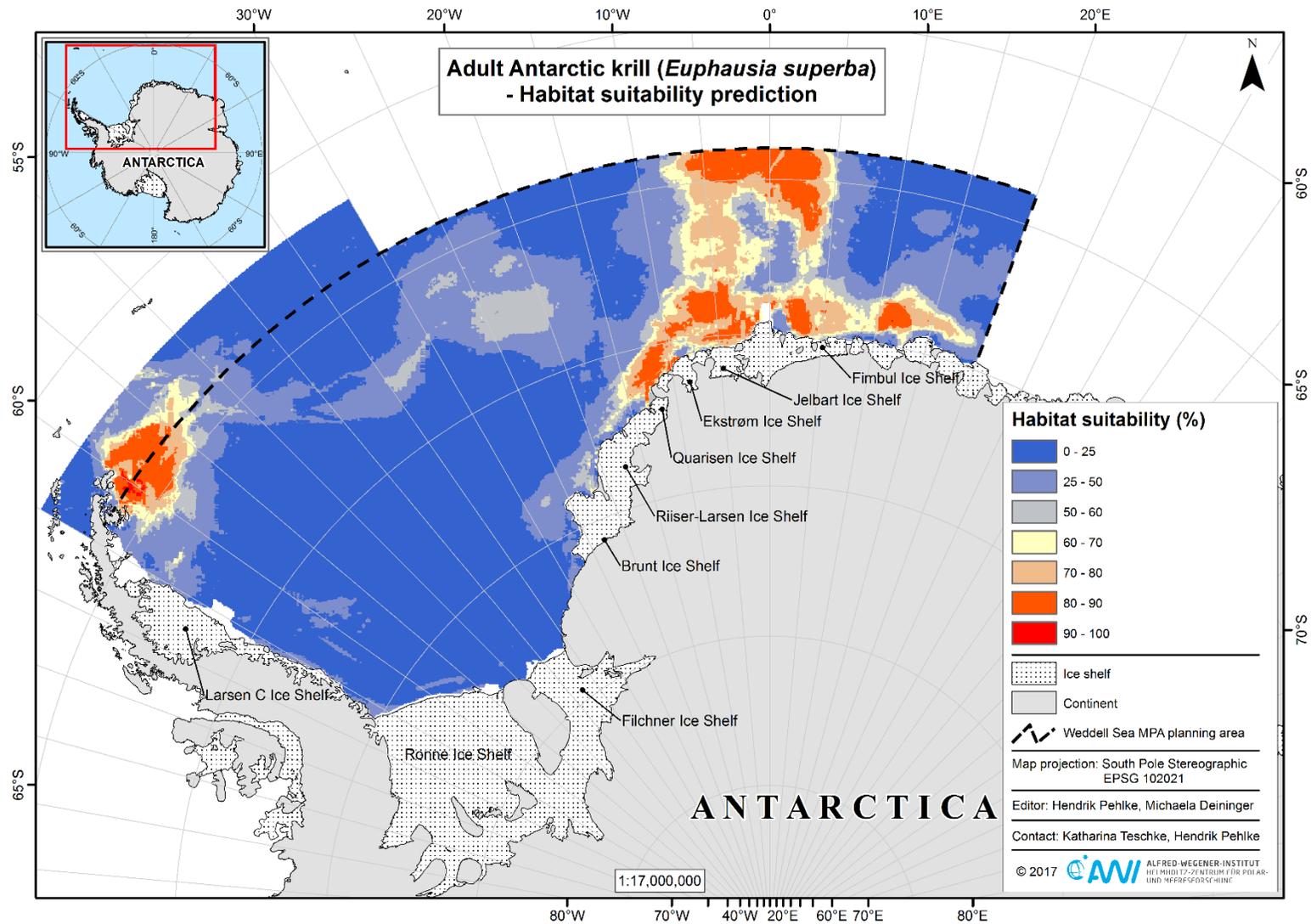


Figure S3. Habitat suitability predictions of adult Antarctic krill (*Euphausia superba*) in the WSMMPA Planning Area.

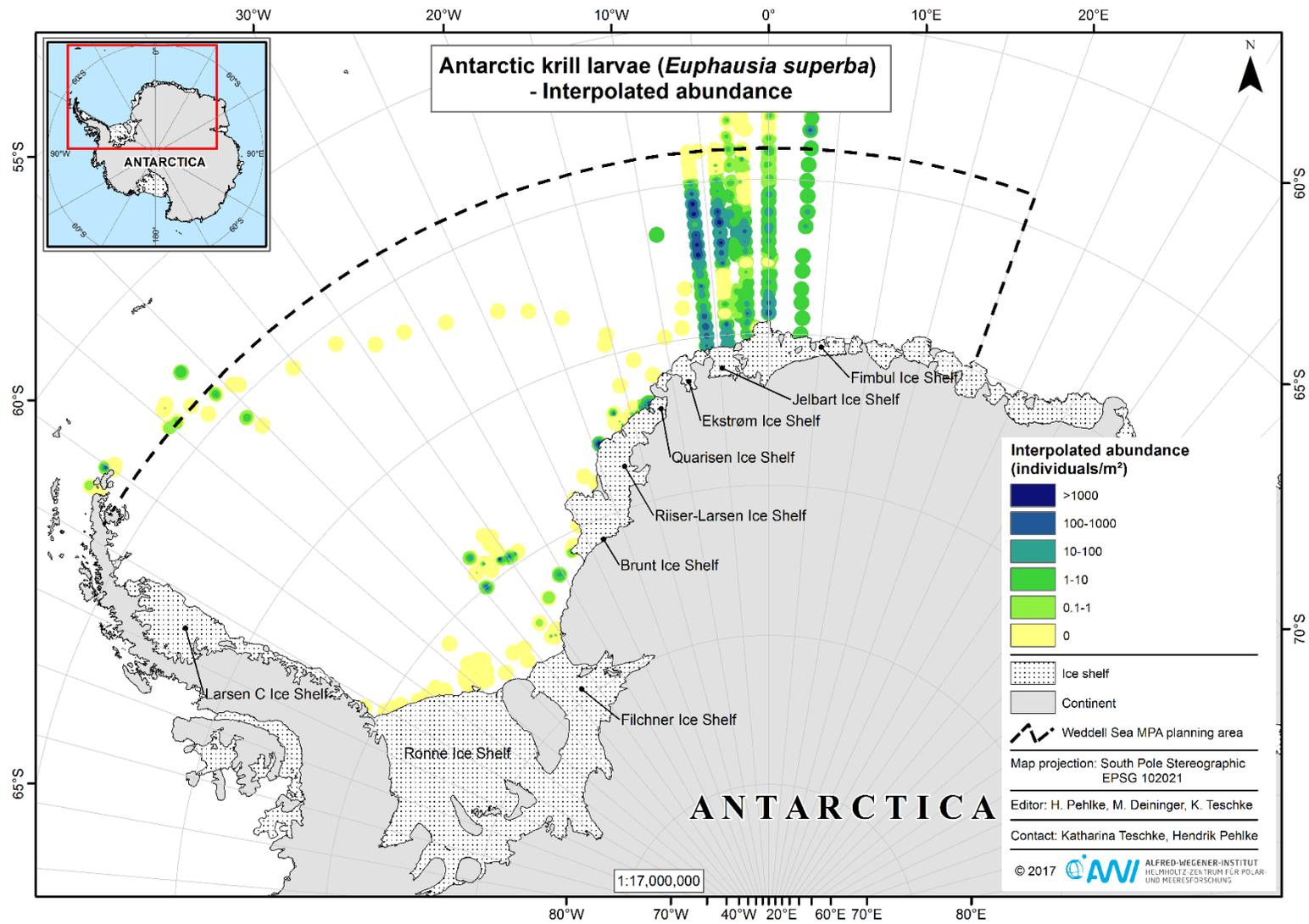


Figure S4. Interpolated abundances of Antarctic krill larvae (*Euphausia superba*) in the WSMMA Planning Area.

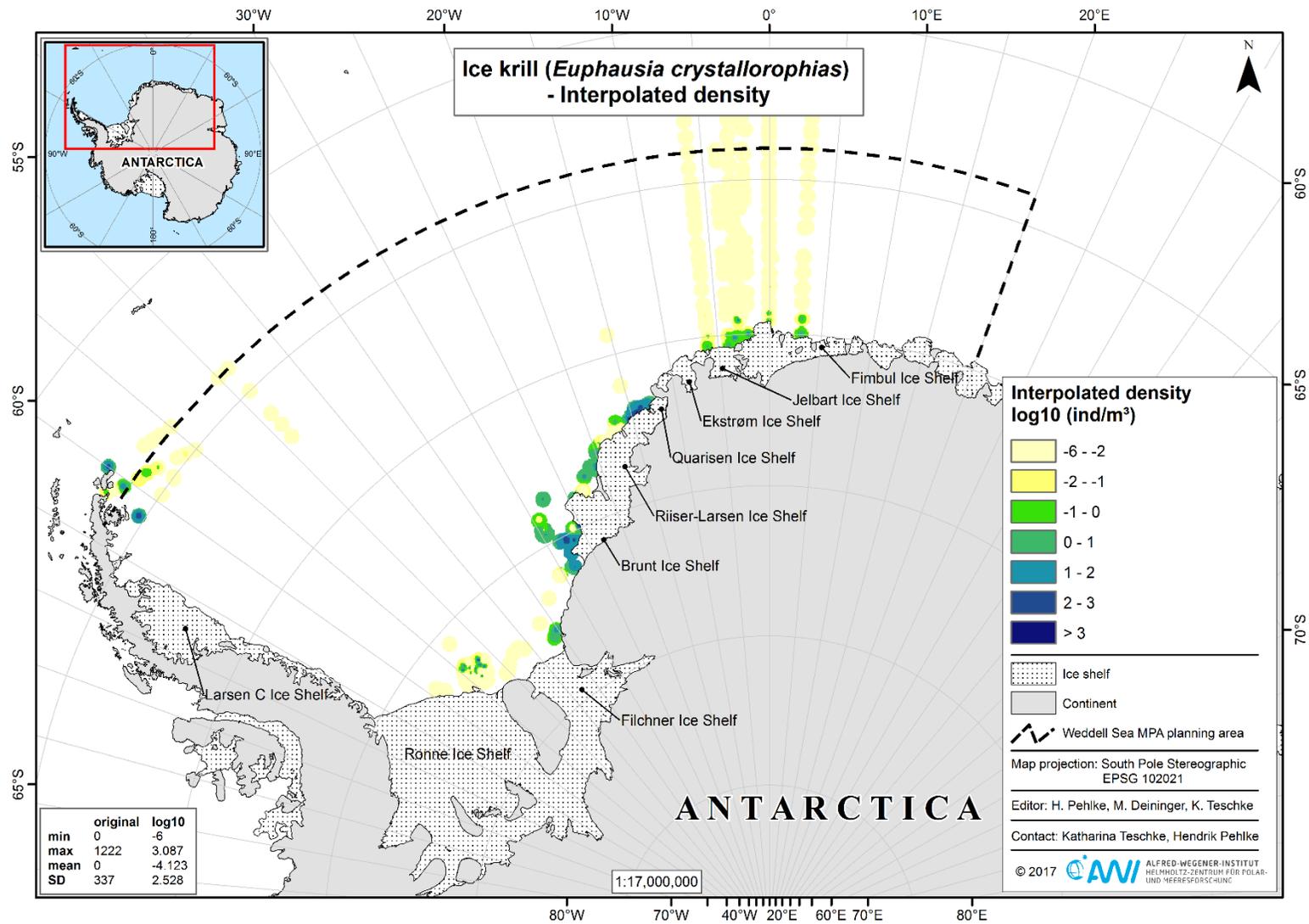


Figure S5. Interpolated abundances of ice krill (*Euphausia crystallorophias*) in the WSMPA Planning Area.

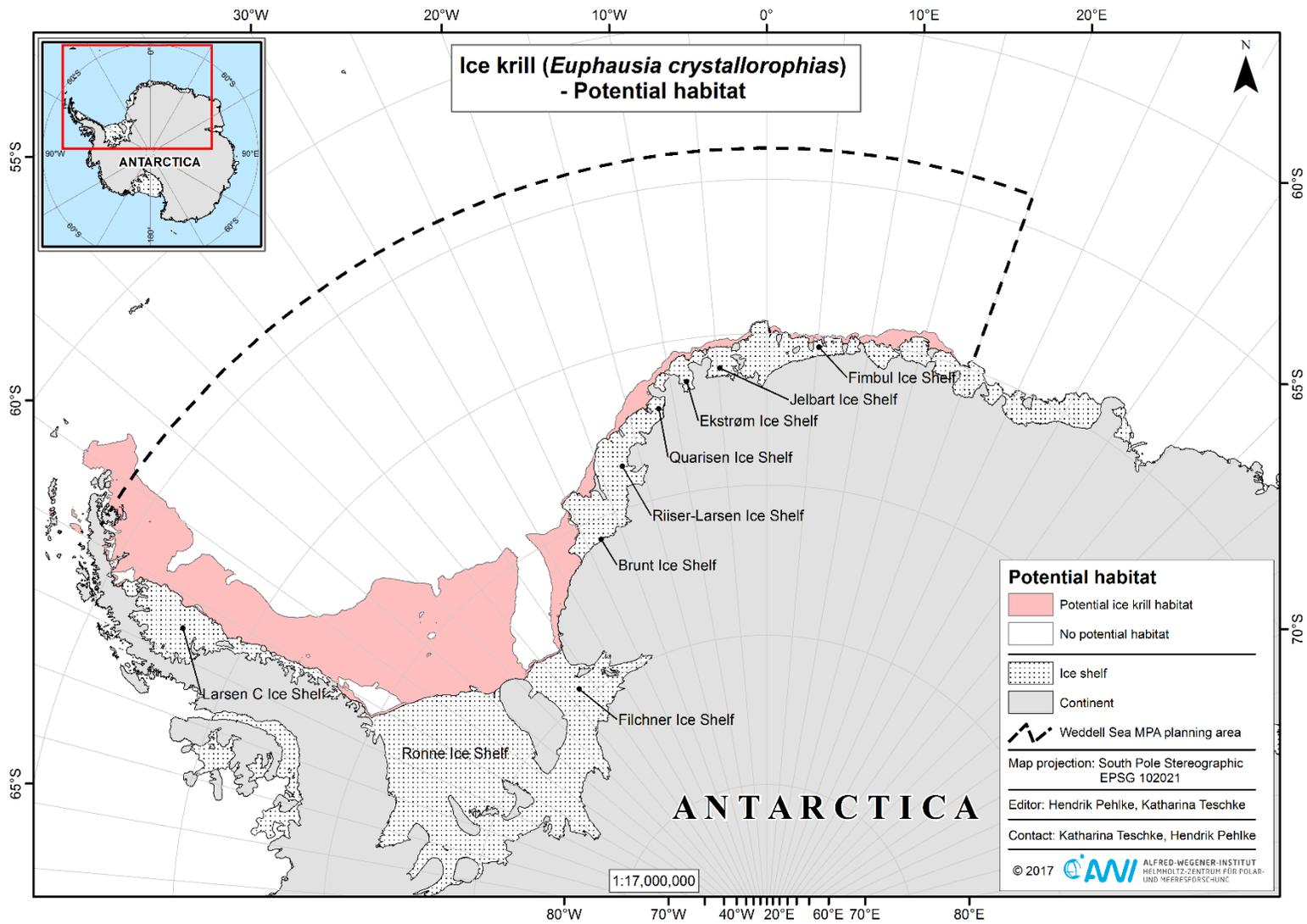


Figure S6. Potential habitat of ice krill (*Euphausia crystallorophias*) in the WSMPA Planning Area.

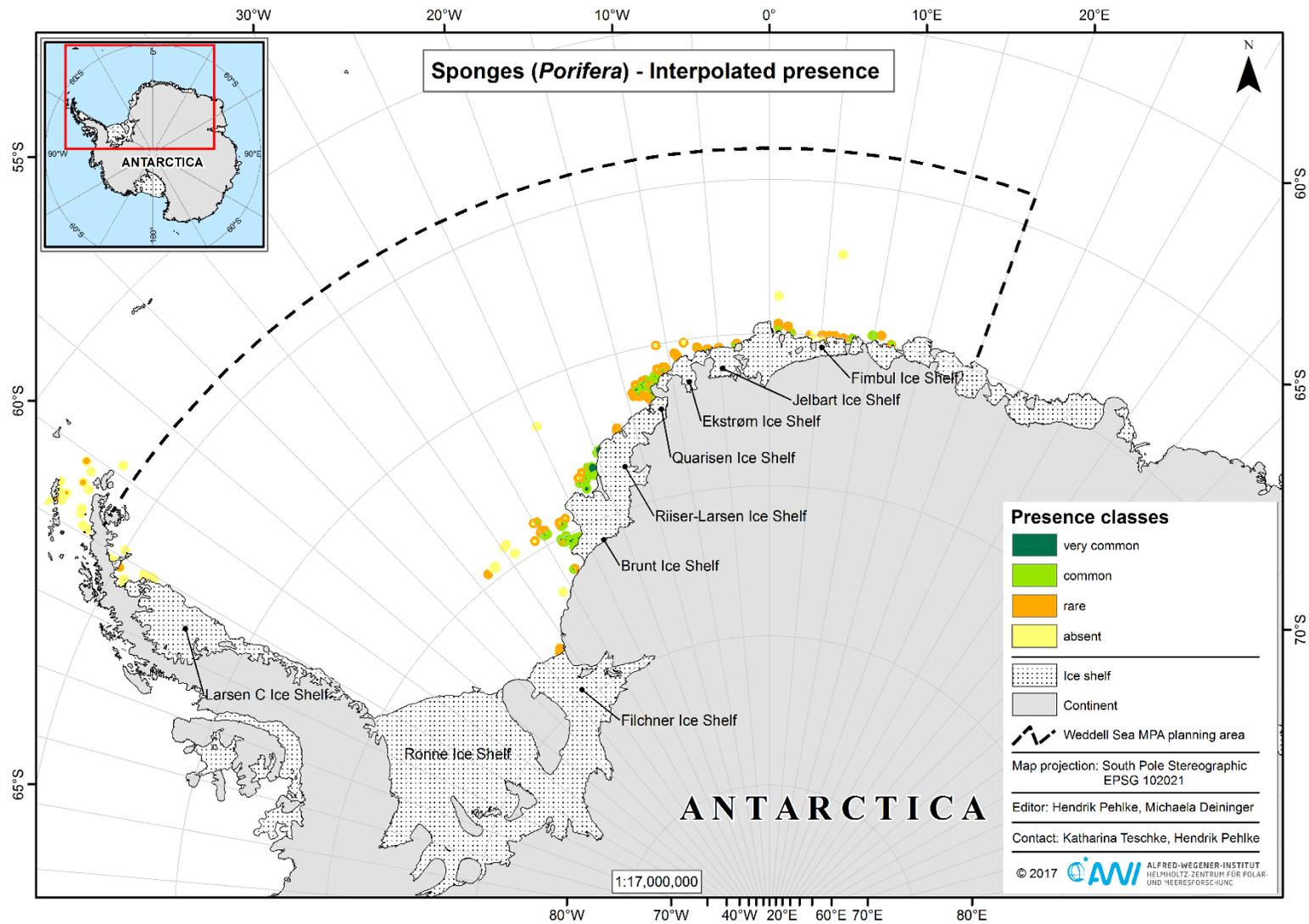


Figure S7. Interpolated occurrences of sponges in the WSMPA Planning Area.

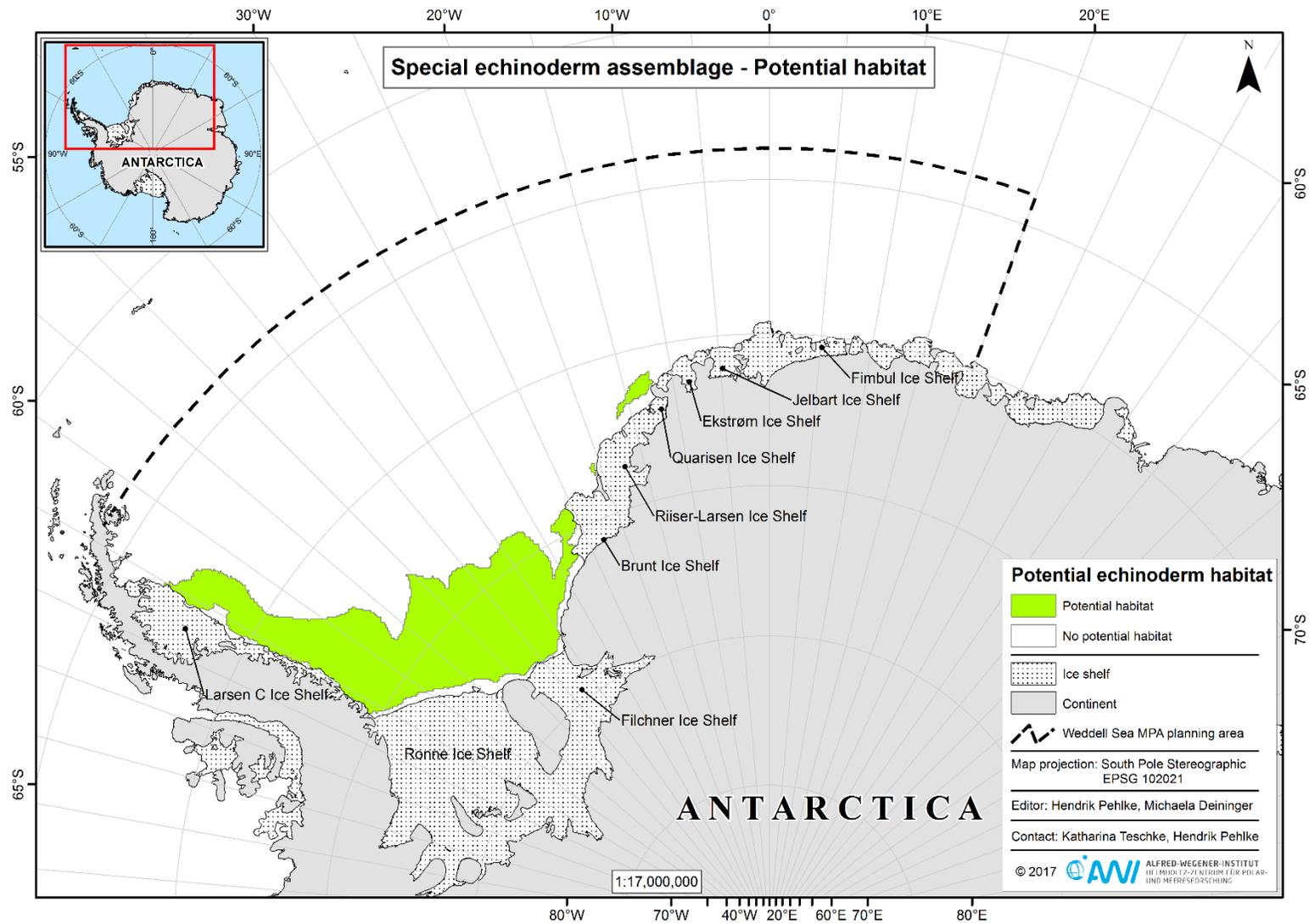


Figure S8. Potential habitat of a special echinoderm assemblage in the WSMMA Planning Area.

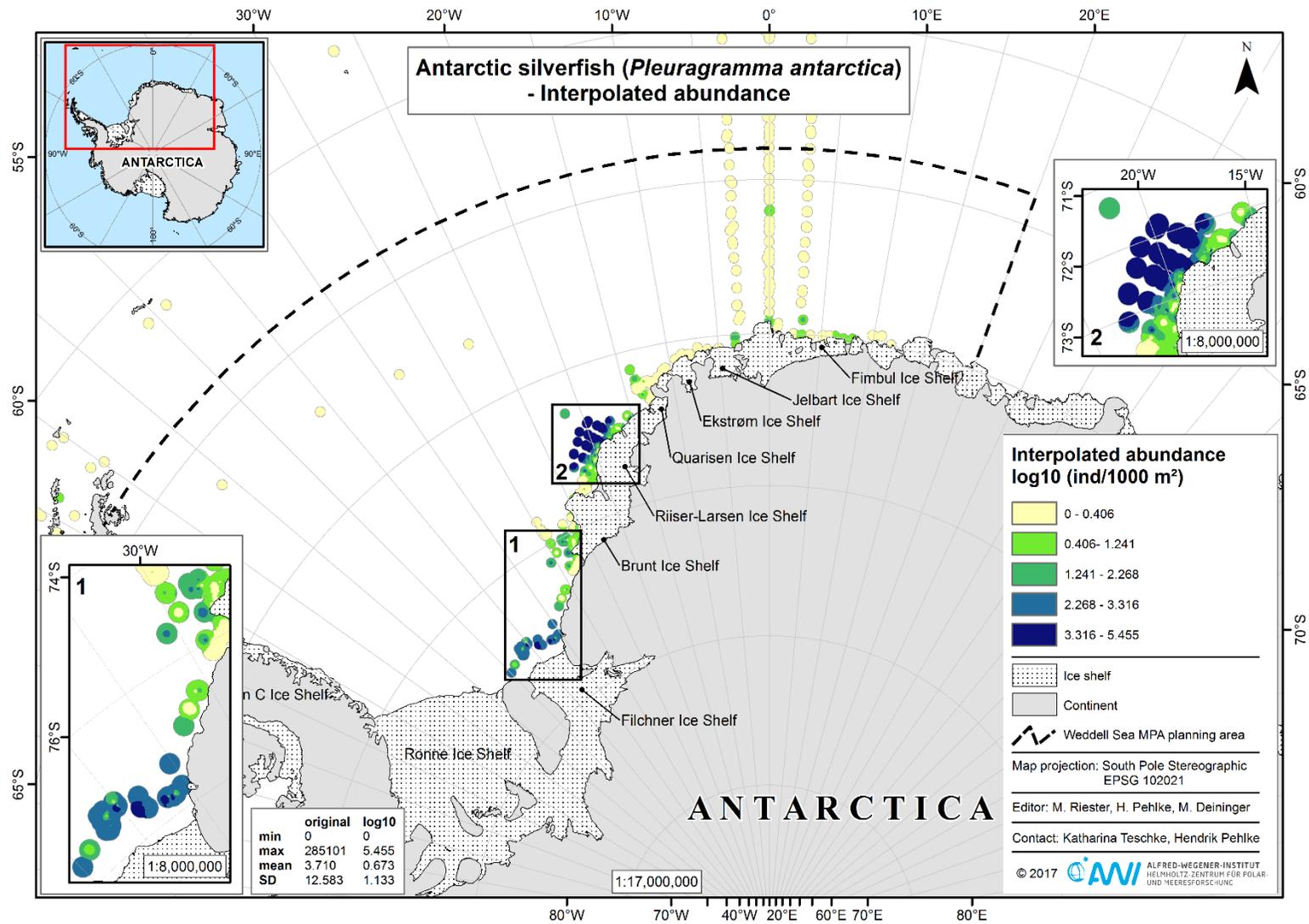


Figure S9. Interpolated abundances of Antarctic silverfish (*Pleuragramma antarctica*) in the WSMPA Planning Area.

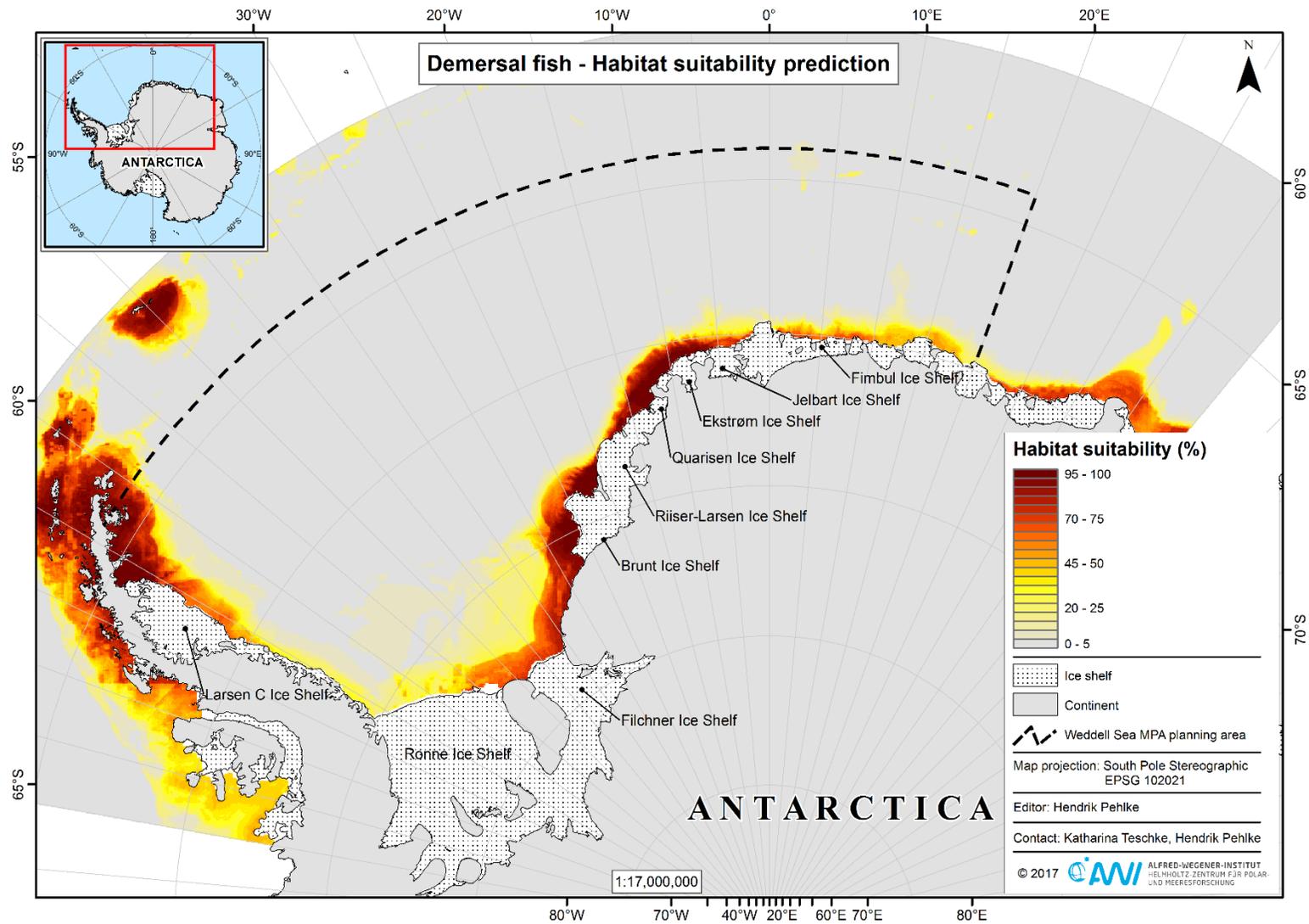


Figure S10. Habitat suitability predictions of demersal fishes in the WSMMA Planning Area.

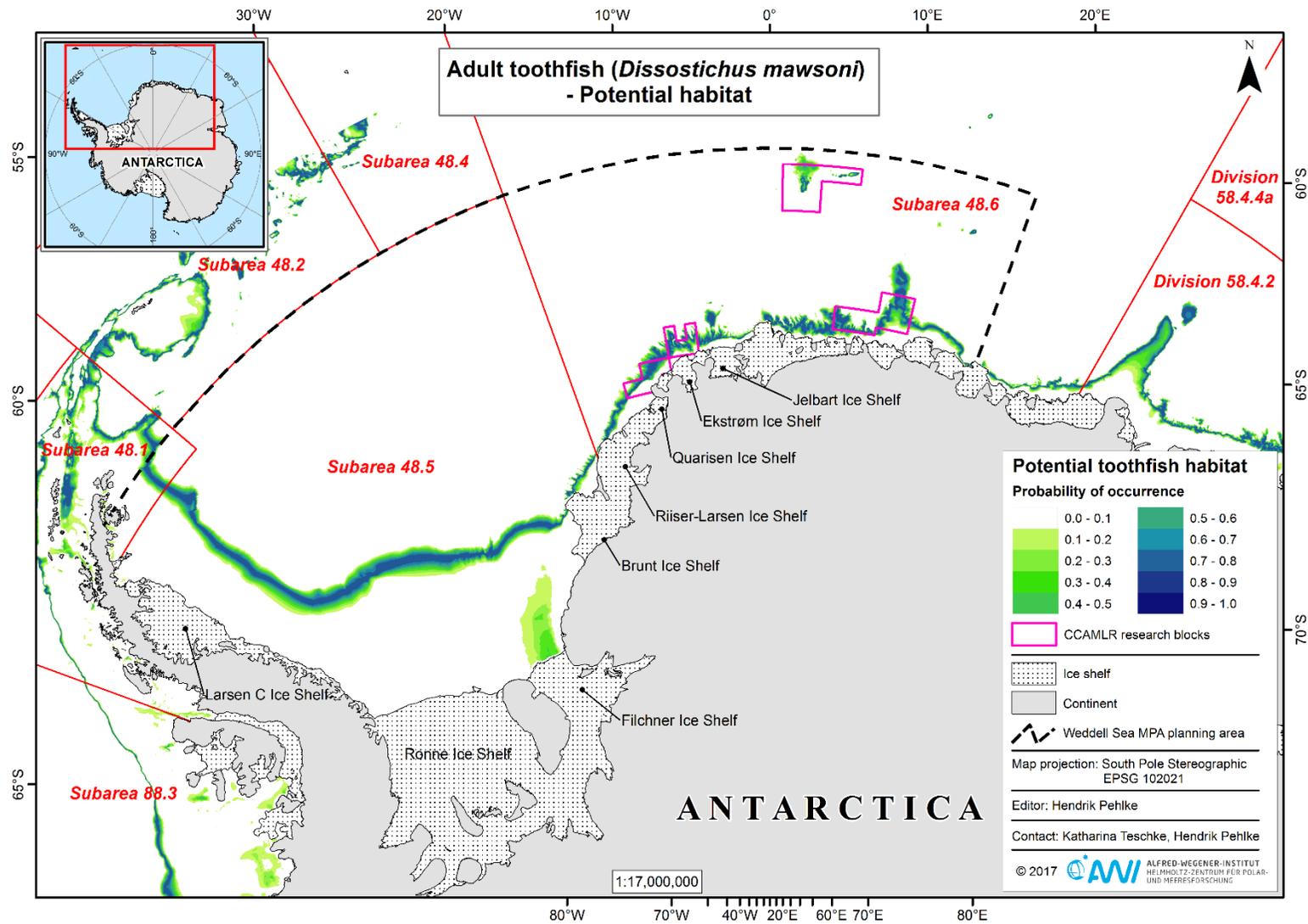


Figure S11. Probability model for the potential habitat of Antarctic toothfish (*Dissostichus mawsoni*) in the WSMPA Planning Area.

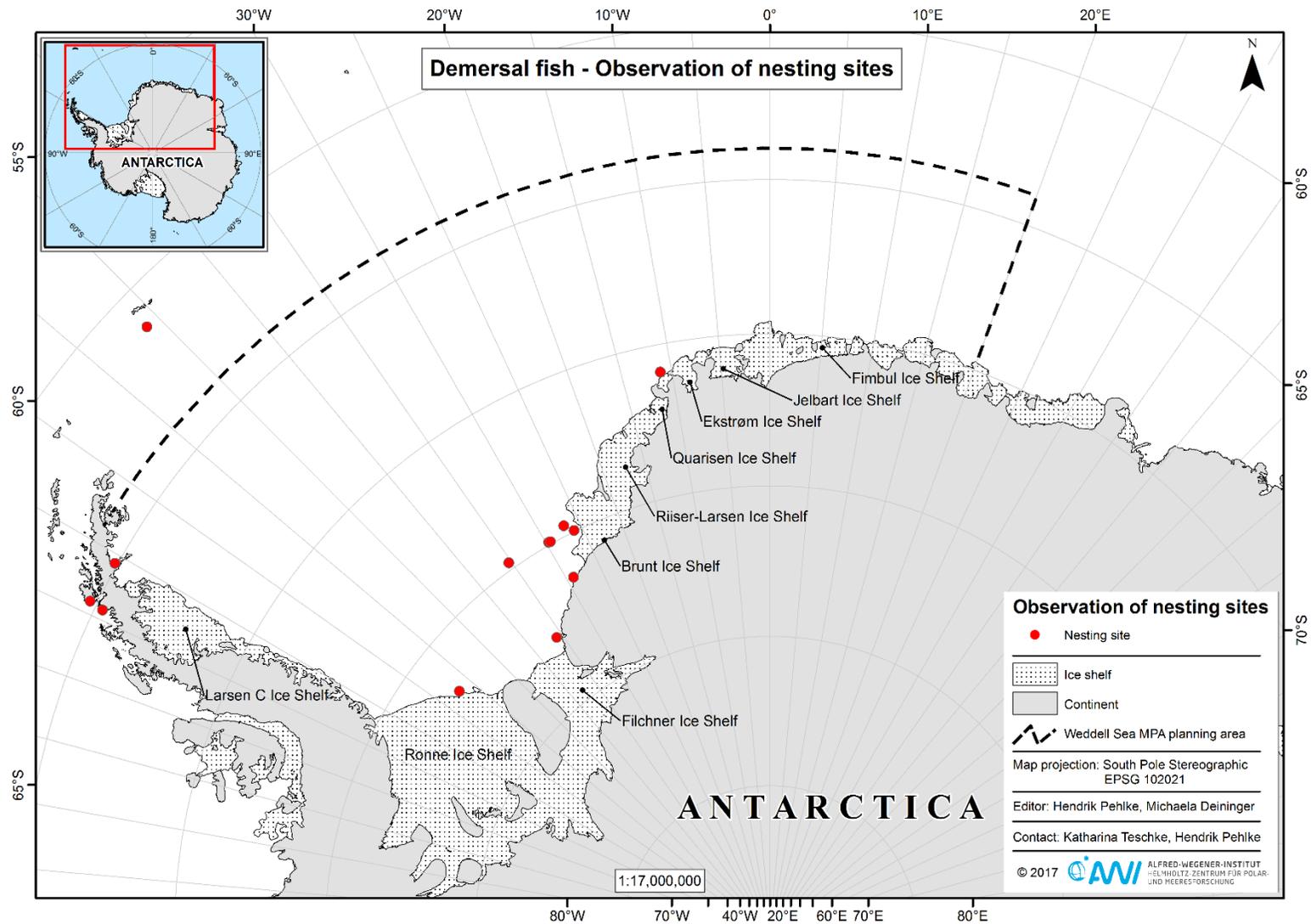


Figure S12. Nesting sites of demersal fish observed in the WSMMPA Planning Area.

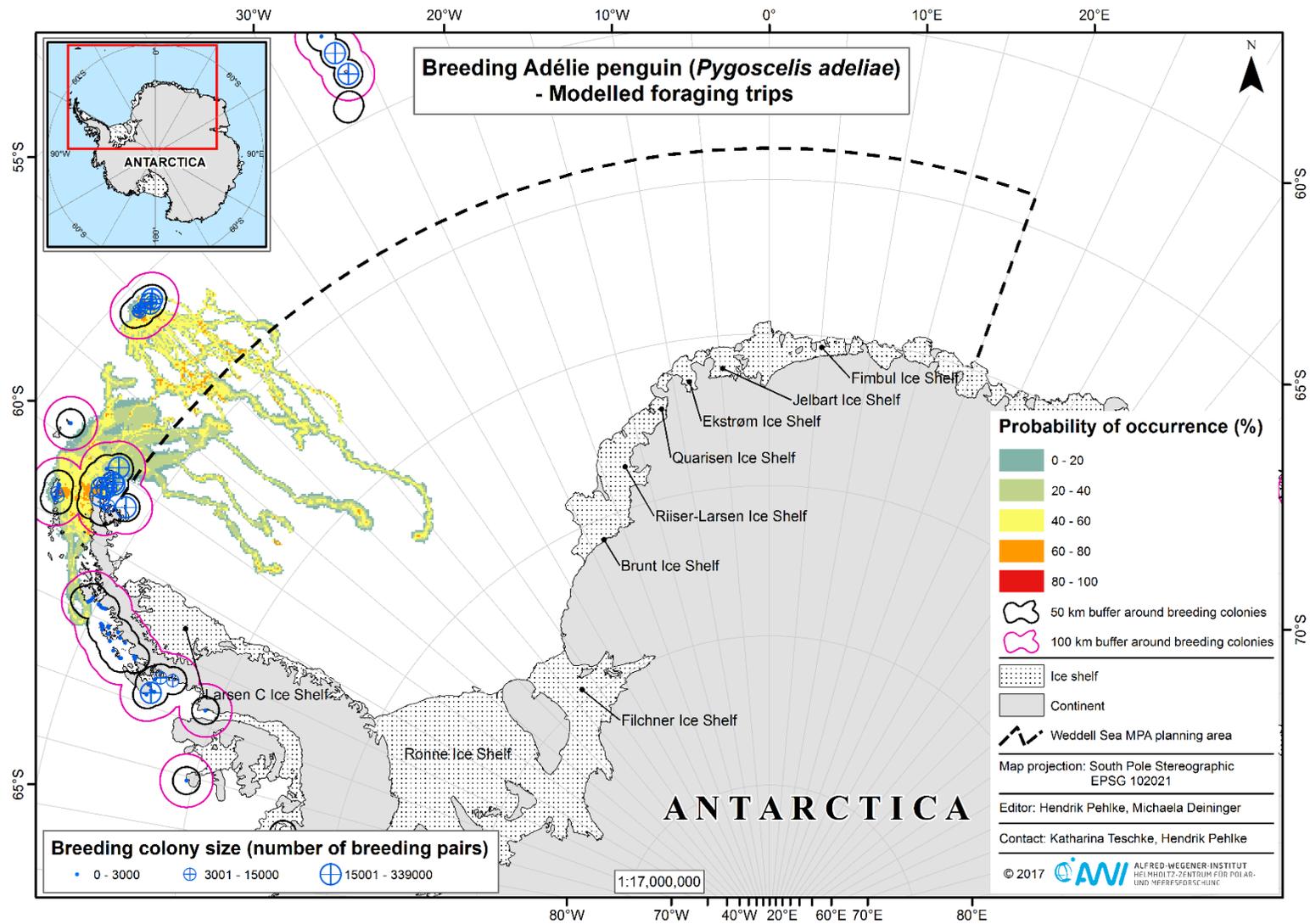


Figure S13. Modelled probability of the occurrence of breeding Adélie penguins (*Pygoscelis adeliae*) during foraging in the WSMPA Planning Area.

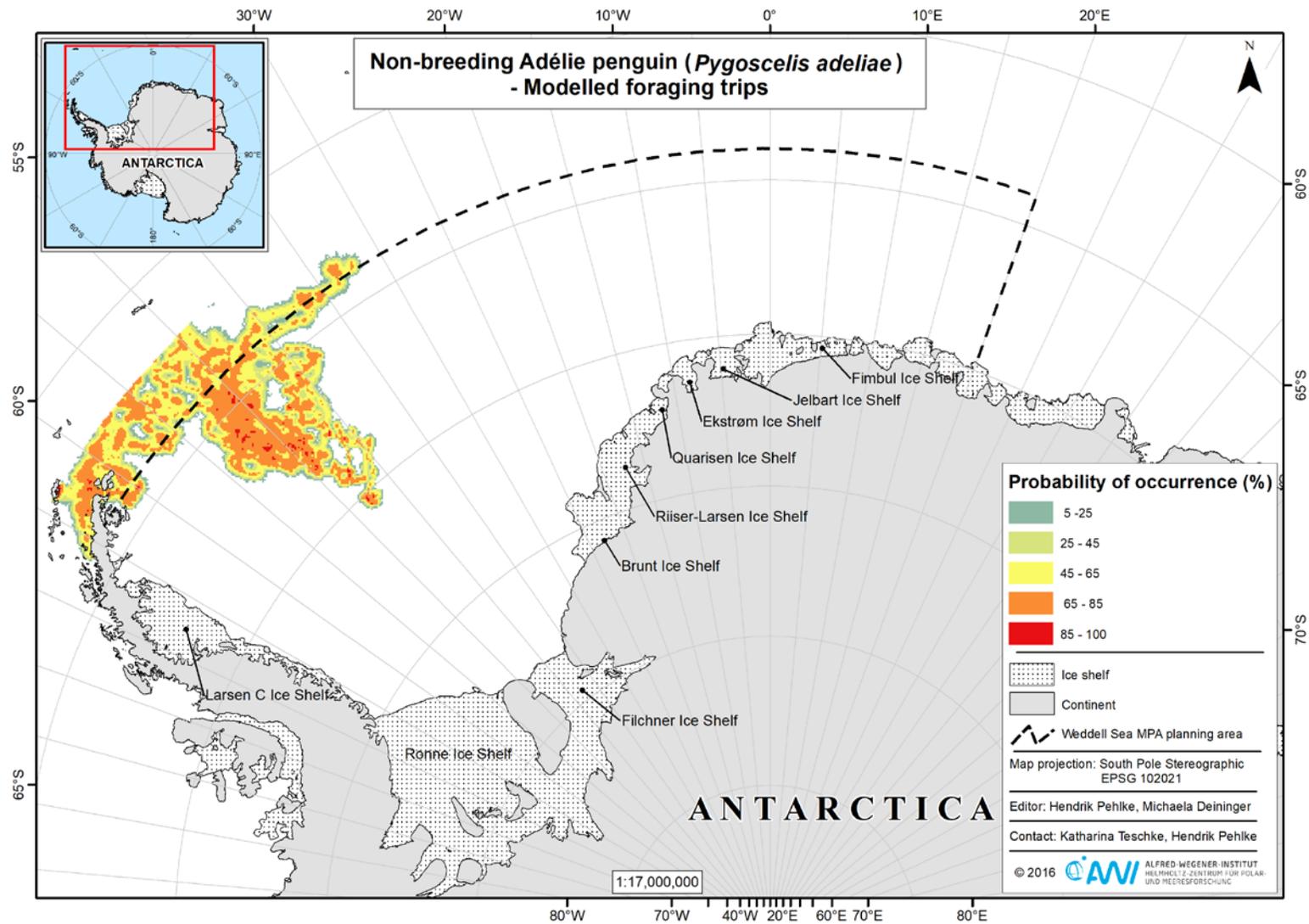


Figure S14. Modelled probability of the occurrence of non-breeding Adélie penguins (*Pygoscelis adeliae*) during foraging in the WSMMPA Planning Area.

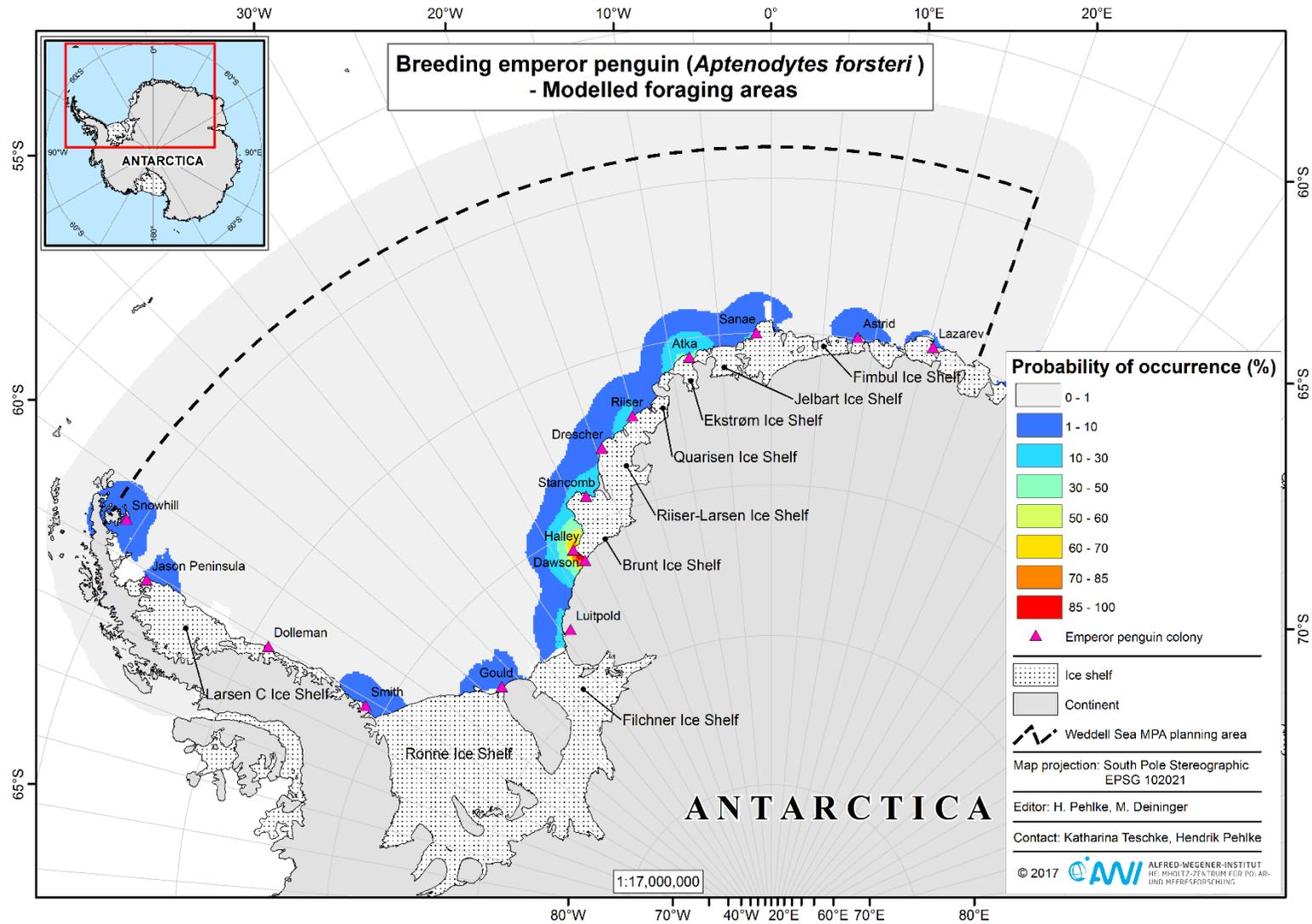


Figure S15. Modelled probability of the occurrence of Emperor penguins (*Aptenodytes forsteri*) during foraging in breeding season.

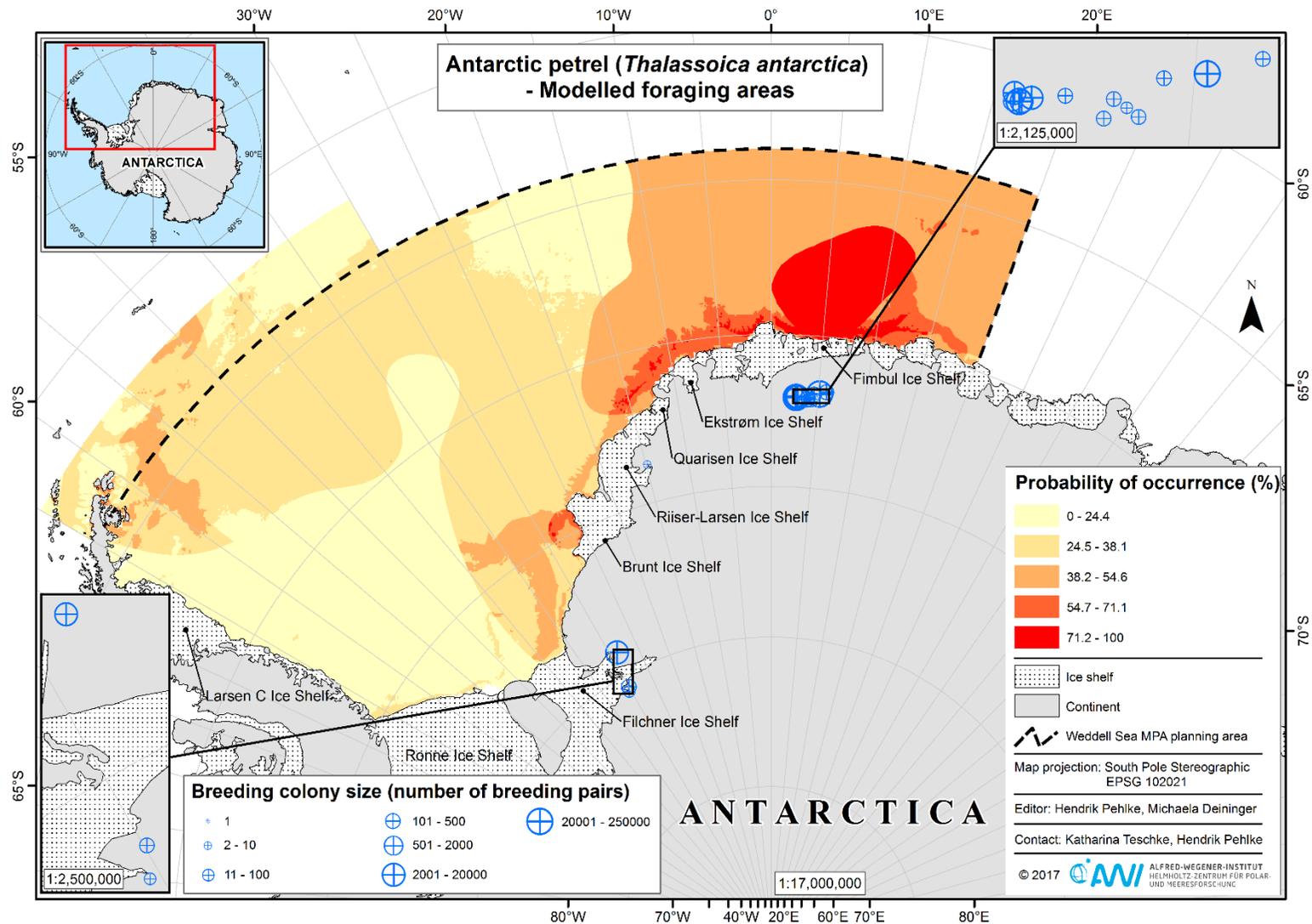


Figure S16. Probability model for the potential habitat of Antarctic petrel (*Thalassoica antarctica*) in the WSMPA Planning Area.

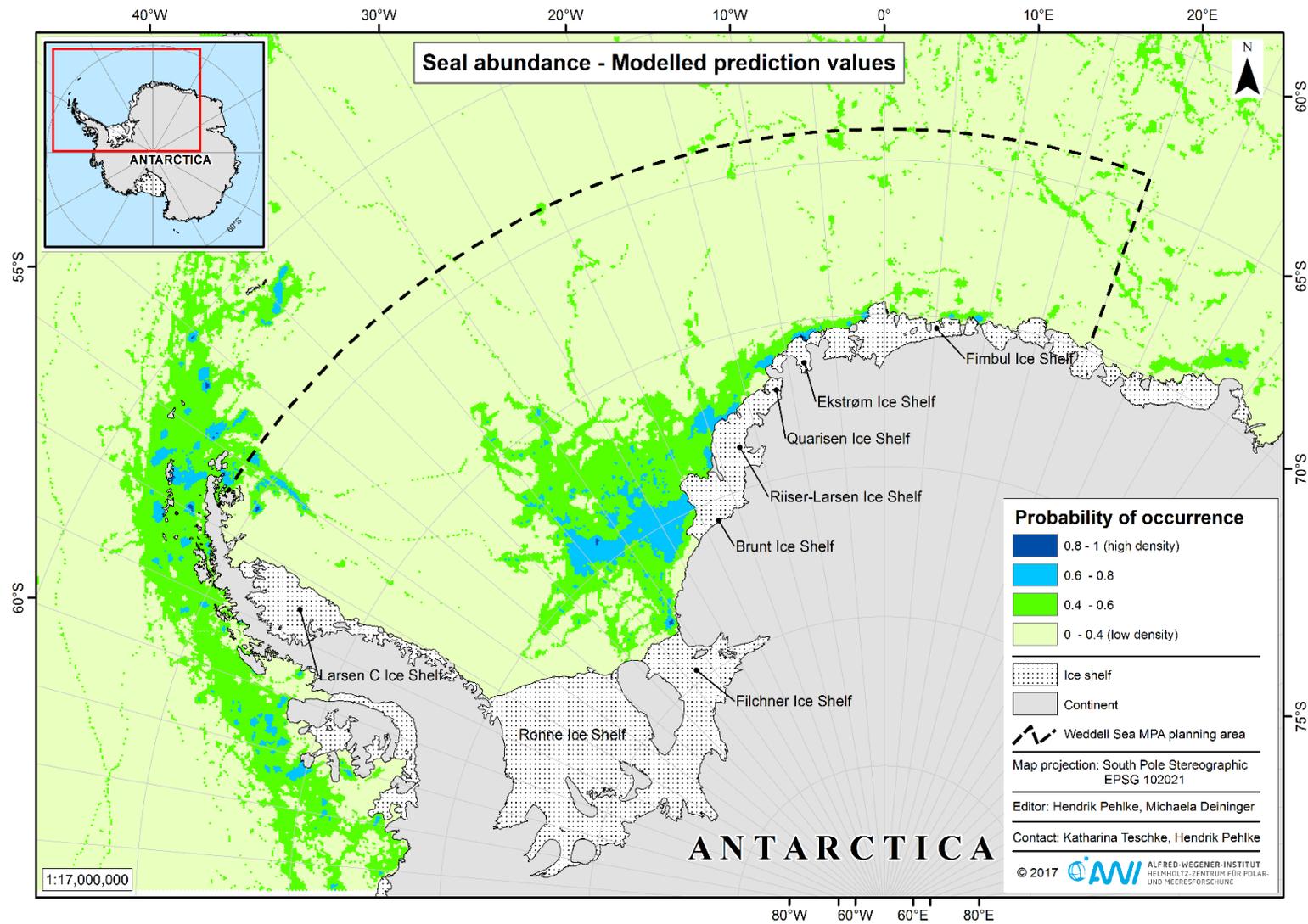


Figure S17. Modelled probability of seal occurrence in the WSMPA Planning Area.

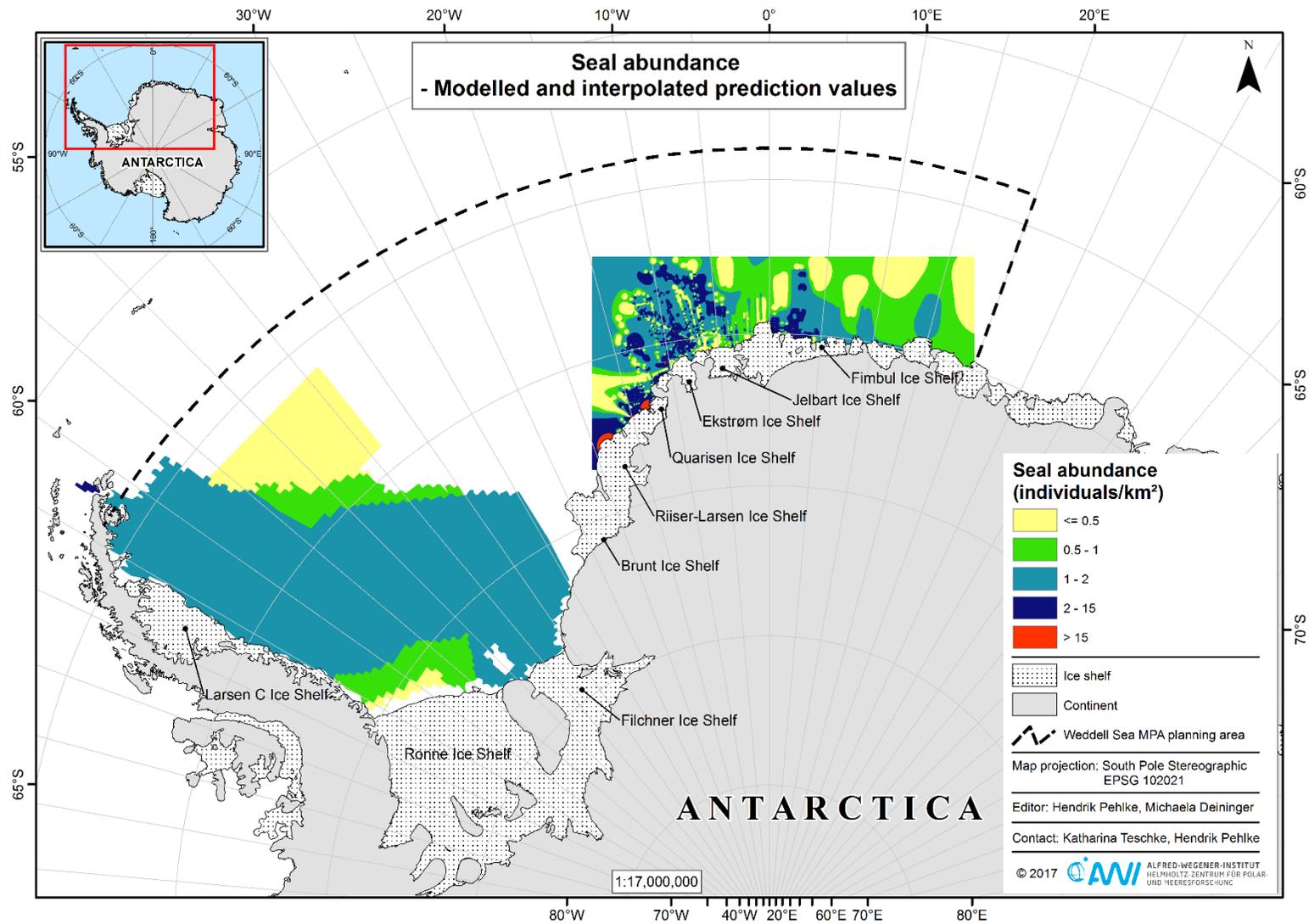


Figure S18. Modelled and interpolated seal abundances in the WSMPA Planning Area.

S3 Tables

Table S2. Detailed list of adult Antarctic krill (*Euphausia superba*) data with survey name, station number per survey and respective source of data, which were used from the database KRILLBASE within the WSMPA planning initiative.

Survey name	Station	Data source
bas1985sib	4a and 5	UK data (SIBEX cruise)
epa1993saf	211-216	South African data
epa1995saf	1-14	South African data
epa1998saf	27	South African data
epa1995bon	189	South African data
epa1996xxx	5 and 6	South African data
epa1989ikt	7, 8, 10, 11	Soviet data
epa1990mkx	149-173	Soviet data
epa1989smt	5, 6, 13, 14	Soviet data
epa1989ikt	7, 8, 10, 11	Soviet data
his1928dis-top	46	UK historical data
his1931wsc-top	548-552	UK historical data
his1932dis-str	813a, 815a, 816a, 822a	UK historical data
his1932dis-top	814, 818-820	UK historical data
his1933dis-str	1148a, 1150a, 1151a, 1153a	UK historical data
his1933dis-top	1149, 1152	UK historical data
his1935dis-str	1519a	UK historical data
his1935dis-top	1513-1528	UK historical data
his1935wsc-top	892 and 893	UK historical data
his1937dis-str	2004a	UK historical data
his1937dis-top	1998-2000, 2002, 2003, 2006-2012	UK historical data
his1937dis-ver	2010a, 2012a	UK historical data
his1939dis-str	2596a, 2598a, 2600a, 2606a-2608a, 2010a	UK historical data
his1939dis-top	2543-2563, 2595, 2599, 2601, 2603, 2604, 2609	UK historical data
his1939dis-ver	2543a, 2545a, 2547a	UK historical data
pol1984wit	165 and 168	Polish data; Witek et al. (1985)
vsi2004lak	1021, 1023-1025, 1028-1031, 1033-1037, 100-1042, 1044-1046, 1048-1050, 1053-1056, 1059-1060, 1062, 1065, 1066, 1068-1071, 1075-1078, 1080, 1084, 1086, 1087, 1090-1092, 1095-1097, 1100-1102, 1116-1117	German data (LAKRIS cruise 2004)
vsi1980pol	20, 21, 23, 24, 27-29, 31, 33-35, 51, 52, 54-79, 82-85, 87-101, 105-111	German data (MV Polarsirke cruise 1979/80)
vsi1981pol	89, 91, 94, 97-99, 101, 106, 108, 110, 111, 126-128, 130, 132, 134, 136, 138, 143, 178, 179, 185, 186, 191, 193, 195, 197, 203, 205, 208-211, 213-215, 217, 219, 224, 226-235, 237, 282	German data (MV Polarsirke cruise 1980/81)
vsi1976xxx	161, 162, 176, 183, 185, 186, 189, 190	German data (RV Walther Herwig cruise 1975/76)
vsi1978xxx	330-337, 339, 340a, 341b, 342c, 343d, 344e, 345	German data (RV Walther Herwig cruise 1977/78)

S4 References

Ainley, D. G., Jacobs, S. S.: Seabird affinities for ocean and ice boundaries in the Antarctic. *Deep-Sea Res. PT I*, 28, 1173-1185, [https://doi.org/10.1016/0198-0149\(81\)90054-6](https://doi.org/10.1016/0198-0149(81)90054-6), 1981.

Ainley, D. G., O'Connor, E. F., and Boekelheide, R. J.: The marine ecology of birds in the Ross Sea, Antarctica. *Ornithol. Monogr.*, 32, 97 pp., <https://doi.org/10.2307/40166773>, 1984.

Ainley, D. G., Ribic, C. A., and Fraser, W. R.: Ecological structure among migrant and resident seabirds of the Scotia-Weddell Confluence region. *J. Anim. Ecol.*, 63, 347-364, <https://doi.org/10.2307/5553>, 1994.

Arndt, J. E., Schenke, H. W., Jakobsson, M., Nitsche, F.-O., Buys, G., Goleby, B., Rebesco, M., Bohoyo, F., Hong, J., Black, J., Greku, R., Udintsev, G., Barrios, F., Reynoso-Peralta, W., Morishita, T., and Wigley, R.: The International Bathymetric Chart of the Southern Ocean (IBCSO) Version 1.0 - A new bathymetric compilation covering circum- Antarctic waters, *Geophys. Res. Lett.*, 40, 3111 –3117, <https://doi.org/10.1002/grl.50413>, 2013.

Baylis, A. M. M., Tierney, M., Orben, R. A., Warwick-Evans, V., Wakefield, E., Grecian, W. J., Trathan, P., Reisinger, R., Ratcliffe, N., Croxall, J., Campioni, L., Catry, P., Crofts, S., Dee Boersma, P., Galimberti, F., Granadeiro, J. P., Handley, J., Hayes, S., Hedd, A., Masello, J. F., Montevecchi, W. A., Pütz, K., Quillfeldt, P., Rebstock, G. A., Sanvito, S., Staniland, I. J., and Brickle P.: Important at-sea areas of colonial breeding marine predators on the Southern Patagonian Shelf, *Sci. Rep.*, 9, 8517, <https://doi.org/10.1038/s41598-019-44695-1>, 2019.

Bester, M. N. and Odendaal, P. N.: Abundance and distribution of Antarctic pack ice seals in the Weddell Sea, in: *Antarctic Ecosystems: Models for Wider Ecological Understanding*, edited by Davison, W., Howard-Williams, C., and Broady, P., Caxton Press, Christchurch, 51-55, 2000.

Bester, M. N., Erickson, A. W., and Ferguson, J. W. H.: Seasonal change in the distribution and density of seals in the pack ice off Princess Martha Coast, Antarctica, *Antarct. Sci.*, 7, 357-364, <https://doi.org/10.1017/S0954102095000502>, 1995.

Bester, M. N., Ferguson, J. W. H., and Jonker, F. C.: Population densities of pack ice seals in the Lasarev Sea, Antarctica, *Antarct. Sci.*, 14, 123-127, <https://doi.org/10.1017/S0954102002000676>, 2002.

CCAMLR Conservation Measure 22-09: Protection of registered vulnerable marine ecosystems in subareas, divisions, small-scale research units, or management areas open to bottom fishing, available at: <https://www.ccamlr.org/en/measure-22-09-2012>, 1p., 2012.

Cuzin-Roudy, J., Irisson, J.-O., Penot, F., Kawaguchi, S., and Vallet, C.: Southern Ocean Euphausiids, in: *Biogeographic Atlas of the Southern Ocean*, edited by: De Broyer, C., Koubbi, P., Griffiths, H. J., Raymond, B., d'Udekem d'Acoz, C., Van de Putte, A., Danis, B., David, B., Grant, S., Gutt, J., Held, C. Hosie, G., Huettmann, F., Post, A., and Ropert-Coudert, Y., Scientific Committee on Antarctic Research, Cambridge, 309-320, 2014.

Delignette-Muller, M. L., Dutang, C., Pouillot, R., and Denis, J.-B.: *Fitdistrplus*: help to fit of a parametric distribution to non-censored or censored data, R package version 1.0-2, <https://cran.r-project.org/src/contrib/Archive/fitdistrplus/>, 2014.

Descamps, S., Tarroux, A., Cherel, Y., Delord, K., Godø, O. R., Kato, A., Krafft, B. A., Lorentsen, S.-H., Ropert-Coudert, Y., Skaret, G., Varpe, Ø.: At-sea distribution and prey selection of Antarctic petrels and commercial krill fisheries. *PLoS ONE*, 11, e0156968, doi:10.1371/journal.pone.0156968, 2016.

Elith, J. and Graham, C. H.: Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models, *Ecography*, 32, 66-77, <https://doi.org/10.1111/j.1600-0587.2008.05505.x>, 2009.

Environmental Systems Research Institute (ESRI): *ArcGIS*, Release 10.2, Redlands, CA, 2013.

Flores, H., Haas C., van Franeker, J. A., and Meesters, E.: Density of pack-ice seals and penguins in the western Weddell Sea in relation to ice thickness and ocean depth, *Deep-Sea Res. PT. II*, 55, 1068-1074, <https://doi.org/10.1016/j.dsr2.2007.12.024>, 2008.

Forcada, J., Trathan, P. N., Boveng, P. L., Boyd, I. L., Burns, J. M., Costa, D. P., Fedak, M., Rogers, T. L., and Southwell, C. J.: Responses of Antarctic pack-ice seals to environmental change and increasing krill fishing, *Biol. Cons.*, 149, 40-50, <https://doi.org/10.1016/j.biocon.2012.02.002>, 2012.

Fretwell, P. T., LaRue, M. A., Morin, P., Kooyman, G. L., Wienecke, B., Ratcliffe, N., Fox, A. J., Fleming, A. H., Porter, C., and Trathan, P. N.: An emperor penguin population estimate: the first global, synoptic survey of a species from space, *PLoS ONE*, 7(4), e33751, <https://doi.org/10.1371/journal.pone.0033751>, 2012.

Fretwell, P. T., Trathan P. N., Wienecke, B., and Kooyman, G. L.: Emperor penguins breeding on iceshelves, *PLoS ONE*, 9, e85285, <https://doi:10.1371/journal.pone.0085285>, 2014.

Han, J., Kamber, M., and Pei, J.: Cluster analysis: Basic concepts and methods, in: *Data mining: concepts and techniques*, edited by: Han, J., Kamber, M., and Pei, J., Morgan Kaufmann, Waltham, MA, USA, 443-494, 2011.

Hao, T., Elith, J., Guillera-Arroita, G., and Lahoz-Monfort, J. J.: A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD, *Divers. Distrib.*, 25, 839-852, <https://doi.org/10.1111/ddi.12892>, 2019.

Hedley, S. L., and Buckland, S. T.: Spatial models for line transect sampling. *J. Agr. Biol. Envir. St.*, 9, 181-199, <https://doi.org/10.1198/1085711043578>, 2004.

Hewitt, J. E., Thrush, S. F., Legendre, P., Funnell, G. A., Ellis, J., and Morrison, M.: Mapping of marine soft-sediment communities: integrated sampling for ecological interpretation, *Ecol. Appl.*, 14, 1203-1216, <https://doi.org/10.1890/03-5177>, 2004.

Jerosch, K., Kuhn, G., Krajnik, I., Scharf, F., and Dorschel, B.: A geomorphological seabed classification for the Weddell Sea, Antarctica, *Mar. Geophys. Res.*, 37, 127-141, <https://doi.org/10.1007/s11001-015-9256-x>, 2016.

Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W.: Continuous-time correlated random walk model for animal telemetry data, *Ecology*, 89, 1208-1215, <https://doi.org/10.1890/07-1032.1>, 2008.

Johnson, D.: Crawl: fit continuous-time correlated random walk models to animal movement data, R package version 1.5, <https://cran.r-project.org/src/contrib/Archive/crawl/>, 2015.

Legendre, P., Ellingsen, K. E., Bjørnbom, E., and Casgrain, P.: Acoustic seabed classification: improved statistical method, *Can. J. Fish. Aquat. Sci.*, 1085-1089, <https://doi.org/10.1139/f02-096>, 2002.

Lu, G.Y., and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, *Comput. Geosci.-UK*, 34, 1044-1055, <https://doi.org/10.1016/j.cageo.2007.07.010>, 2008.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., and Roudier, P.: Cluster: cluster analysis extended Rousseeuw et al., R package version 1.15.2., <https://cran.r-project.org/src/contrib/Archive/cluster/>, 2014.

Neukermans, G., Reynolds, R. A., and Stramski, D.: Optical classification and characterization of marine particle assemblages within the western Arctic Ocean, *Limnol Oceanogr*, 61, 1472-1494, <https://doi.org/10.1002/lno.10316>, 2016.

QGIS Development Team: QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>, 2015.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, 2014.

Raymond, B.: Pelagic Regionalisation, in: Biogeographic Atlas of the Southern Ocean, edited by: De Broyer, C., Koubbi, P., Griffiths, H. J., Raymond, B., d'Udekem d'Acoz, C., Van de Putte, A., Danis, B., David, B., Grant, S., Gutt, J., Held, C., Hosie, G., Huettmann, F., Post, A., and Ropert-Coudert, Y., Scientific Committee on Antarctic Research, Cambridge, 418-421, 2014.

SC-CAMLR-XXXV: Domain 3 and 4 - Weddell Sea, in: Report of the thirty-fifth meeting of the Scientific Committee, Hobart, Australia, 17 - 21 October 2016, available at: <https://www.ccamlr.org/en/sc-camlr-xxxv>, 51-54, 2016.

Soimasuo, J., Neteler, M., Blazek, R., Landa, M., Metz, M., and Cho, H.: GRASS GIS package v.distance, <http://grass.osgeo.org/grass70/manuals/v.distance.html>, 1994.

Thuiller, W.: BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change, *Glob. Change Biol.*, 9, 1353-1362, <https://doi.org/10.1046/j.1365-2486.2003.00666.x>, 2003

Thuiller, W., Lafourcade, B., Engler, R., and Araújo, M. B.: BIOMOD - A platform for ensemble forecasting of species distributions., *Ecography*, 32, 369-373, <https://doi.org/10.1111/j.1600-0587.2008.05742.x>, 2009.

Thuiller, W., Lafourcade, B., and Araújo, M.: Presentation Manual for BIOMOD, available at: <https://pdfs.semanticscholar.org/5522/2d3a8501206807a9a0a32a2b7e62260f53af.pdf>, 35 pp., 2010.

Thuiller, W., Georges, D., and Engler, R.: Biomod2: Ensemble platform for species distribution modeling, R package version 3.1-64, <https://cran.r-project.org/src/contrib/Archive/biomod2/>, 2014.

Tukey, J. W.: *Exploratory data analysis.*, Addison-Wesley Publishing Company, Reading, MA, USA, 1977.

Van Franeker, J. A., Gavriolo, M., Mehlum, F., Veit, R. R., and Woehler, E. J.: Distribution and abundance of the Antarctic Petrel, *Waterbirds*, 22, 14-28, <https://doi.org/10.2307/1521989>, 1999.

Verfaillie, E., Degraer, S., Schelfaut, K., Willems, W., and Van Lanckerad, V.: A protocol for classifying ecologically relevant marine zones, a statistical approach, *Estuar. Coast. Shelf Sci.*, 83, 175-185, <https://doi.org/10.1016/j.ecss.2009.03.003>, 2009.

Ward, J. H.: Hierarchical grouping to optimize an objective function, *J. Amer. Statist. Ass.*, 58, 236-244, <https://doi.org/10.1080/01621459.1963.10500845>, 1963.

Warwick-Evans, V., Ratcliffe, N., Lowther, A. D., Manco, F., Ireland, L., Clewlow, H. L., and Trathan, P. N.: Using habitat models for chinstrap penguins *Pygoscelis antarctica* to advise krill fisheries management during the penguin breeding season, *Divers. Distrib.*, 24, 1756-1771, <https://doi.org/10.1111/ddi.12817>, 2018.

Warwick-Evans, V., Downie, R., Santos, M., and Trathan, P. N.: Habitat preferences of Adélie *Pygoscelis adeliae* and Chinstrap Penguins *Pygoscelis antarctica* during pre-moult in the Weddell Sea (Southern Ocean), *Polar Biol.*, 42, 703-714, <https://doi.org/1007/s00300-019-02465-9>, 2019.

Weise, M. J., Harvey, J. T., and Costa, D. P.: The role of body size in individual-based foraging strategies of a top marine predator, *Ecology*, 91, 1004-1015, <https://doi.org/10.1890/08-1554.1>, 2010.

WG-EMM-15/42: Report of the Second International Workshop for identifying Marine Protected Areas (MPAs) in Domain 1 of CCAMLR (Palacio San Martín, Buenos Aires, Argentina, 25 to 29 May 2015), 42 pp., 2015.

WG-EMM-16: MPA Planning Domains 3 and 4 - Weddell Sea, in: Report of the Working Group on Ecosystem Monitoring and Management, Bologna, Italy, 4 - 15 July 2016, available at: <https://www.ccamlr.org/node/88723>, 252-254, 2016.

Witek, Z., Kittel, W., Czykieta, H., Żmijewska, M. I., and Presler, E.: Macrozooplankton in the southern part of Drake Passage and in the Bransfield Strait during BIOMASS-SIBEX (December 1983 - January 1984), *Pol. Polar Res.*, 6, 95-115, 1985.

Zharikov, Y., Skilleter, G. A., Loneragan, N. R., Taranto, T., and Cameron, B. E.: Mapping and characterising subtropical estuarine landscapes using aerial photography and GIS for potential application in wildlife conservation and management, *Biol. Conserv.*, 125, 87-100, <https://doi.org/10.1016/j.biocon.2005.03.016>, 2005.

Zimmer, I., Wilson, R.P., Gilbert, C., Beaulieu, M., Ancel, A., and Plötz, J.: Foraging movements of emperor penguins at Pointe Géologie, Antarctica, *Polar Biol.*, 31, 229-243, <https://doi.org/10.1007/s00300-007-0352-5>, 2008.