Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

ESDM: A tool for creating and exploring ensembles of predictions from species distribution and abundance models

Samuel M. Woodman[1]* 🔟 | **Karin A. Forney[2,3]** 🔟 | **Elizabeth A. Becker[1,4]** 🔟 |
Monica L. DeAngelis[5] | **Elliott L. Hazen[6]** 🔟 | **Daniel M. Palacios[7]** 🔟 | **Jessica V. Redfern[1]** 🔟

[1]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA, USA; [2]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Moss Landing, CA, USA; [3]Moss Landing Marine Laboratories, San Jose State University, Moss Landing, CA, USA; [4]Cooperative Institute for Marine Ecosystems and Climate (CIMEC), University of California, Santa Cruz, Santa Cruz, CA, USA; [5]West Coast Regional Office, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Long Beach, CA, USA; [6]Environmental Resource Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Monterey, CA, USA and [7]Marine Mammal Institute and Department of Fisheries and Wildlife, Hatfield Marine Science Center, Oregon State University, Newport, OR, USA

Correspondence
Samuel M. Woodman
Email: sam.woodman@noaa.gov

Present address
Monica L. DeAngelis, Naval Undersea Warfare Center, Newport, RI 02841, USA

Abstract

1. Species distribution models (SDMs) are a valuable statistical approach for both understanding species distributions and identifying potential impacts of environmental changes or management decisions to species, but multiple SDMs for the same species in a region can create confusion in decision-making processes.

2. One solution is to create ensembles (i.e. combinations) of predictions from existing SDMs. However, creating ensembles can be challenging if the predictions were made at different spatial resolutions, using different data sources, or with different prediction value types (e.g. abundance and probability of occurrence).

3. We present ESDM, an R package that allows users to create an ensemble of SDM predictions overlaid onto a single base geometry. These predictions can be evaluated (e.g. through among-model uncertainty or AUC, TSS and RMSE metrics), mapped, and exported. ESDM includes a built-in GUI created using the R package SHINY, which makes the package accessible to non-R users.

4. We provide an overview of ESDM functionality and use ESDM to create an ensemble of predictions from three blue whale *Balaenoptera musculus* SDMs for the California Current Ecosystem.

KEYWORDS
abundance, blue whale *Balaenoptera musculus*, ensemble, model evaluation, model predictions, probability of occurrence, species distribution models, species management

1 | INTRODUCTION

Species distribution models (SDMs; i.e. habitat-based occurrence models or ecological niche models) characterize the relationship between spatially and temporally explicit species observations and environmental data. SDMs are widely used to predict species distribution and abundance based on habitat covariates, and these predictions can be used to make conservation and management decisions (Elith & Leathwick, 2009; Gregr, Baumgartner, Laidre, & Palacios, 2013). The increased use of SDMs worldwide (Guisan et al., 2013) has created new challenges when multiple SDMs for the same species in a single region produce conflicting results (Araújo & New, 2007; Jones-Farrand et al., 2011). Individual SDMs may identify unique ecological niches or suggest different management actions because of the strengths, biases, and limitations of each underlying dataset and model algorithm (Jones-Farrand et al., 2011). These issues are often difficult to reconcile and incorporate into management decision-making.

An ensemble (i.e. a weighted or unweighted average or combination) provides an established method for resolving differences between individual models and estimating uncertainty (Araújo & New, 2007; Marmion, Parviainen, Luoto, Heikkinen, & Thuiller, 2009). For example, model ensembles have been widely used in global climate change assessments to evaluate mean predictions and associated uncertainties (Annan & Hargreaves, 2010; Tebaldi & Knutti, 2007). In addition, ensembles have been successfully used to model species distributions (e.g. Forney, Becker, Foley, Barlow, & Oleson, 2015; Grenouillet, Buisson, Casajus, & Lek, 2011; Oppel et al., 2012; Pikesley et al., 2013; Scales et al., 2016), although these studies each relied upon a single data source. The authors created ensembles by averaging corresponding predictions from SDMs generated using different model algorithms and the original species and environmental data. Several existing software tools implement this method, including R packages (R Core Team, 2019) BIOMOD2 (Thuiller, Georges, Engler, & Breiner, 2019) and SDM (Naimi & Araújo, 2016).

A different approach is needed when multiple data sources exist. Integrated analyses, such as a Bayesian hierarchical framework, can be used to obtain a single, probabilistic assessment of species distributions from several original data sources (e.g. Golding & Purse, 2016; Hefley & Hooten, 2016). However, this approach is not always practical for general use because it requires extensive statistical expertise and is generally time-consuming and computationally challenging. Simpler methods for combining information from multiple data sources exist (e.g. Merow, Wilson, & Jetz, 2017; Pacifici et al., 2017), but still require the original data sources. If original data are unavailable, SDM predictions derived from these original data may be the only accessible information for a particular region. Combining or reconciling these predictions can be difficult, particularly if they were created using different methods or at different spatial resolutions (but see Sansom, Wilson, Caldow, & Bolton, 2018 for methods comparing prediction maps from different sources).

For example, multiple predictions from blue whale *Balaenoptera musculus* SDMs for the California Current Ecosystem (CCE) have been published (Becker et al., 2016; Hazen et al., 2017; Redfern et al., 2017), although some of the underlying datasets are not publicly available. These predictions were created at several spatial resolutions, in various coordinate systems, and using different data sources, habitat covariates, and modelling frameworks. In addition, the SDMs predicted absolute density, habitat preference, and relative density (e.g. density calculated without line transect correction factors; see Redfern et al., 2017), respectively (see Table 1 for model details).

We present ESDM (Ensemble tool for predictions from Species Distribution Models), an R package with a built-in graphical user interface (GUI) for creating ensembles of SDM predictions. ESDM allows users to overlay SDM predictions onto a single base geometry, create ensembles of these overlaid predictions, and evaluate, map, and export predictions. It also provides several options for incorporating or calculating uncertainty. The information provided by this tool can assist users in identifying spatial uncertainties and making informed conservation and management decisions. ESDM (v0.3.0; https://doi.org/10.5281/zenodo.3371754) is available on CRAN, and the GUI can be run locally or accessed online. ESDM uses the R package SF (Pebesma, 2018) for fast processing of spatial data, while the GUI, created using the R package SHINY (Chang, Cheng, Allaire, Xie, & McPherson, 2019), makes the tool accessible to non-R users. In this paper, we provide an overview of ESDM functionality and use the GUI to create and evaluate ensembles of predictions from the three blue whale SDMs (Table 1).

2 | ESDM OVERVIEW

Creating ensemble predictions using ESDM requires three major steps: (a) importing original SDM predictions, (b) overlaying the

TABLE 1 A summary of the individual SDMs that predicted the blue whale distributions used in the example analysis

Model citation	Becker et al. (2016)	Hazen et al. (2017)	Redfern et al. (2017)
Abbreviation	Model_B	Model_H	Model_R
Whale data source	1991–2009 shipboard line-transect surveys	1994–2008 satellite telemetry data	1991–2009 shipboard line-transect surveys
Modelling framework	Generalized additive model (GAM)	Generalized additive mixed model (GAMM)	Generalized additive model (GAM)
Spatial resolution	0.09° × 0.09° (~10 × 10 km^2) grid	0.25° × 0.25° (~25 × 25 km^2) grid	10 × 10 km^2 equal-area grid
Prediction value type	Absolute density	Habitat preference	Relative density

original predictions onto a single base geometry, and (c) creating ensemble predictions via a weighted or unweighted average of rescaled overlaid predictions. Additional steps may include evaluating, mapping, or exporting predictions. Validation data can be read from comma-separated value (CSV) and GIS files (shapefiles and file geodatabase feature classes) as either binary (i.e. species presence/absence) or count data. ESDM allows users to calculate several common evaluation metrics: area under the receiver operating characteristic curve (AUC; Fielding & Bell, 1997), true skill statistic (TSS; Allouche, Tsoar, & Kadmon, 2006), and root-mean-square error (RMSE). AUC and TSS measure the discriminatory ability of an SDM, and can be calculated with predictions of any value type. RMSE, a scale-dependent measure that requires count validation data, evaluates both the discriminatory ability and calibration of an SDM. Users can export ensemble predictions to calculate other metrics. Uncertainty associated with predictions (e.g. standard error values) can be imported, mapped, and used to weight predictions in an ensemble or calculate uncertainty values

TABLE 2 Brief descriptions of ESDM functions; see package documentation for more details

| Function	Description
ESDM _ GUI	Launch the ESDM GUI
ENSEMBLE _ CREATE	Create a weighted or unweighted ensemble of SDM predictions, and calculate associated uncertainty values
ENSEMBLE _ RESCALE	Rescale predictions using the abundance or sum to one method
EVALUATION _ METRICS	Calculate AUC, TSS, or RMSE of SDM predictions using validation data
MODEL _ ABUNDANCE	Calculate the predicted abundance using SDM density predictions and the area of the corresponding prediction polygons
OVERLAY _ SDM	Overlay SDM predictions onto a base geometry
PTS2POLY _ VERTICES	Create polygon(s) from a data frame containing the longitude and latitude coordinates of the polygon vertices
PTS2POLY _ CENTROIDS	Create polygons from a data frame containing the longitude and latitude coordinates of a regular grid of polygon centroids
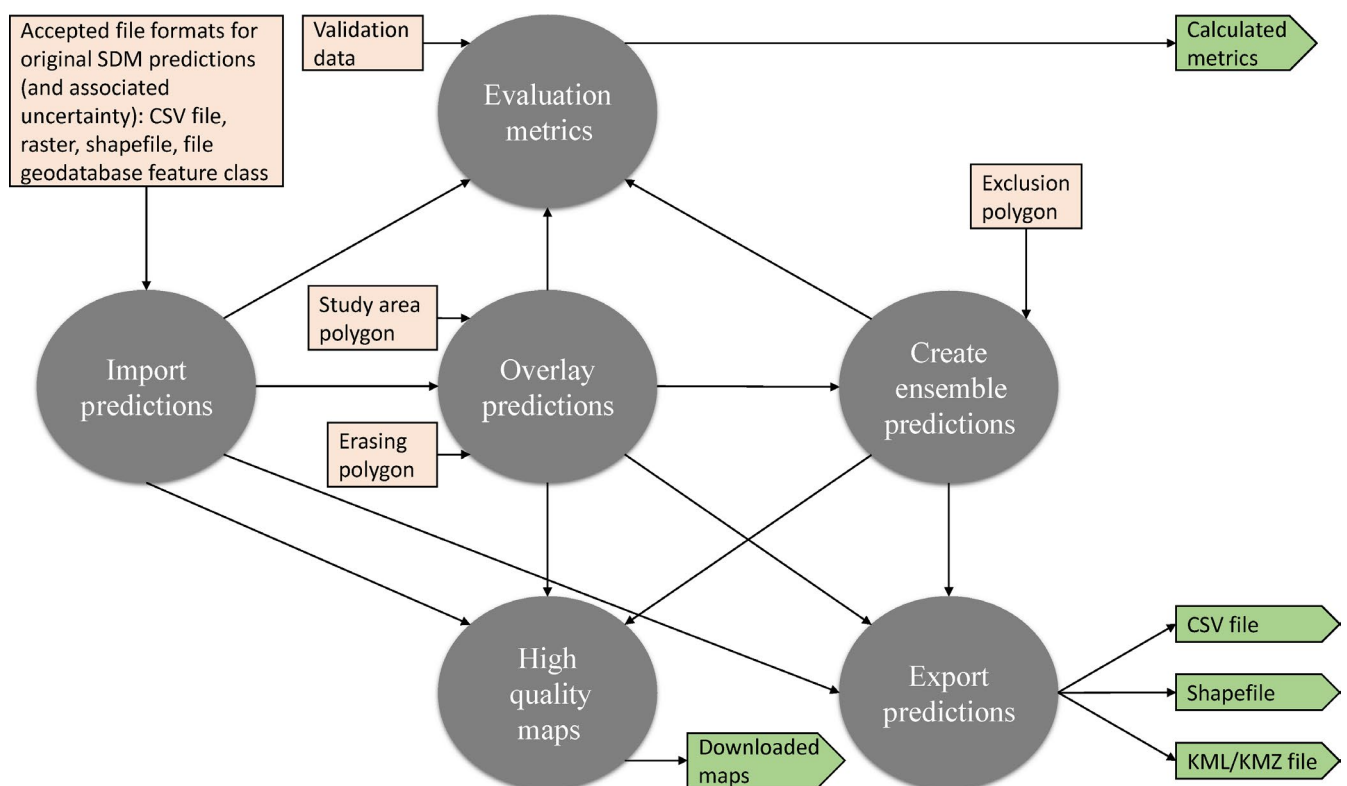
FIGURE 1 Flowchart detailing the workflow of the ESDM GUI, i.e. the order in which users can access and use sections of the GUI. Grey ovals represent tabs within the GUI, orange squares represent files imported by users, and green arrow boxes represent files exported by users. Users can also load a saved GUI workspace rather than re-importing and processing predictions during each session. In the 'Create Ensemble Predictions' tab, the user can use ensemble weights based on user inputs or metrics calculated in the 'Evaluation Metrics' tab. Additional details are provided in the text and the ESDM GUI manual

for the ensemble predictions. Ensemble uncertainty can also be assessed using the among-model variance. In addition, the GUI allows users to create maps of predictions and additional objects, such as validation data or areas of human use (e.g. shipping lanes).

The ESDM GUI provides ESDM functionality through a user-friendly, web-based interface. Alternatively, users familiar with R can incorporate ESDM functions in their own code (see Table 2 for function descriptions). Here we present a flowchart of the GUI workflow (Figure 1) and describe the major steps of creating ensemble predictions.

2.1 | Importing predictions

The ESDM GUI accepts SDM predictions in several common formats (Figure 1) and processes them to create a 'prediction polygon' for each individual prediction value. These prediction polygons make up the 'geometry' of a set of predictions, similar to how individual cells make up a raster. When importing predictions from a CSV file, the provided coordinates must be WGS 84 geographic coordinates (i.e. decimal degrees) and represent the centroids of a regular grid of prediction polygons. The GUI can also read and process predictions from GIS files (rasters, shapefiles, and file geodatabase feature classes), which have already-defined geometries and coordinate systems. Those writing their own R code can use ESDM function PTS2POLY_CENTROIDS to convert centroid coordinates to prediction polygons and

functions from the RASTER (Hijmans, 2019) and SF packages to import GIS files.

The GUI accepts 'Abundance', 'Absolute density', or 'Relative density' as prediction value types. Users should select 'Relative density' for value types that are proportional to density but do not represent an absolute abundance or density (e.g. probability of occurrence or habitat preference; see Aarts, Fieber, & Matthiopoulos, 2012). The GUI allows the user to rescale these values if needed (described in Section 2.3 below).

2.2 | Overlaying predictions

The overlay function, OVERLAY_SDM, is the backbone of ESDM. It overlays SDM predictions onto a single base geometry, transforming all predictions to the same spatial resolution and coordinate system (Figure 2). Within the GUI, users can choose which of the imported predictions to use as the base geometry and specify the coordinate system in which the overlay will be performed. They can also import polygons to clip or erase portions of the base geometry, such as to specify a study area or erase land from marine predictions.

The overlay function intersects the prediction polygons from an original SDM with the prediction polygons from the user-selected base geometry (i.e. base geometry polygons). It then calculates the percentage of each base geometry polygon that overlaps with these intersected polygons, ignoring intersected polygons that have missing (i.e. 'NA') prediction values. If this percentage
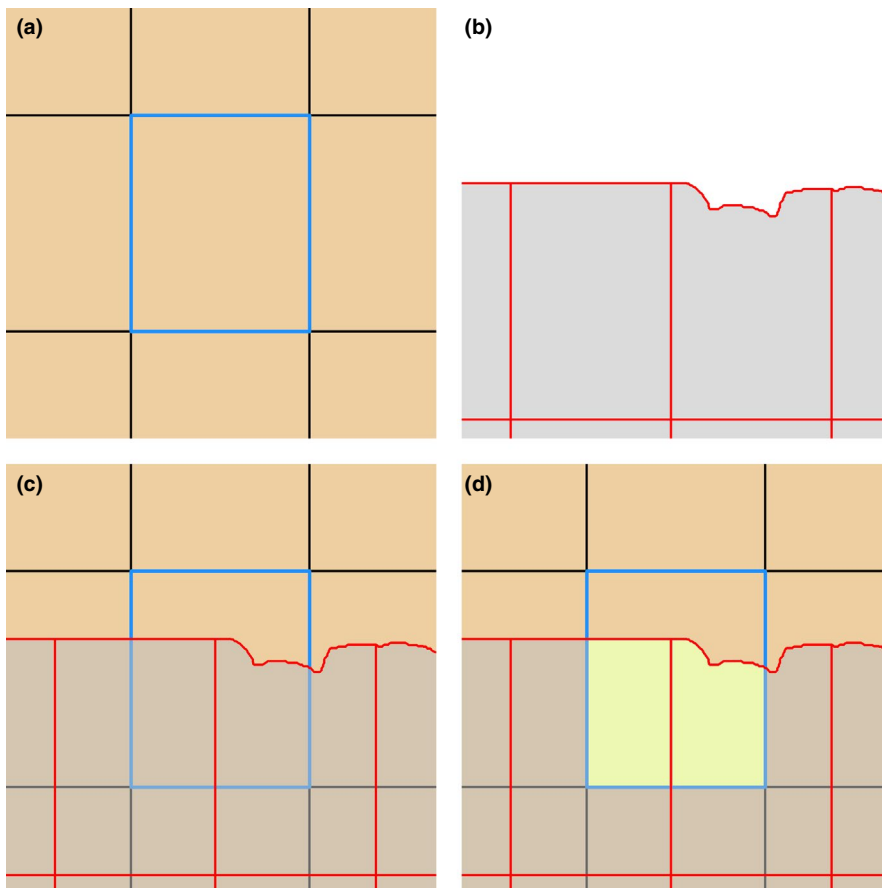

FIGURE 2 Schematic illustration: (a) The base geometry, with the blue outline indicating the current base geometry polygon. (b) The geometry of the SDM predictions being overlaid. (c) The SDM predictions overlaid onto the base geometry. (d) Same as (c), with the intersection between the overlaid polygons and the current base geometry polygon (i.e. intersected polygons) coloured yellow

meets or exceeds the user-specified percent overlap threshold, the function calculates the overlaid prediction as an area-weighted average of the predictions of the intersected polygons (i.e. areal interpolation; Goodchild & Lam, 1980). Otherwise, the function assigns that base geometry polygon an overlaid prediction of 'NA', thereby excluding it from any ensembles. Associated uncertainty values and weights are also overlaid using an area-weighted average.

2.3 | Creating ensemble predictions

2.3.1 | Rescaling different prediction value types

Overlaid predictions that have different prediction value types (e.g. absolute density vs. probability of occurrence), should be rescaled to ensure predictions do not contribute disproportionately to the ensemble. Users can either rescale predictions to a specified total abundance within the study area or, if they do not have an abundance estimate, rescale predictions to sum to one. These rescaling methods are inherently similar and result in ensembles with similar distribution patterns. However, only the abundance rescaling method results in an ensemble with a meaningful abundance estimate. If another rescaling method is desired, users can rescale predictions before importing them, or export and rescale overlaid predictions.

2.3.2 | Ensemble method

Ensembles can be created using a weighted or unweighted average of the rescaled predictions. Weights can be based on evaluation metrics (i.e. evaluation metric values, rescaled to sum to one, of the overlaid predictions), the inverse of the variance of the overlaid predictions, or assigned by users either for the entire study area or for each prediction polygon. Users can also regionally exclude predictions from the ensemble if they have some a priori reason to do so (e.g. known biases in a specific region). ESDM calculates uncertainty for the ensemble predictions using either the user-specified prediction uncertainties or the among-model variance.

3 | EXAMPLE ANALYSIS

Predictions from cetacean SDMs can be used to assess the risk of entanglements and ship-strikes (e.g. Redfern et al., 2013), which represent the largest sources of anthropogenic injury or mortality for blue whales in the CCE (Carretta et al., 2018). Becker et al. (2016), Hazen et al. (2017), and Redfern et al. (2017) developed models of blue whale distributions in this region (henceforth Model_B, Model_H, and Model_R, respectively) that can provide information for risk assessments. However, the predictions from these models differ in some areas (Figure 3), making them challenging to use for management purposes. We use the ESDM GUI to perform an example analysis that explores differences between the blue whale SDM

predictions and creates an ensemble of the predictions, with associated uncertainty.

The three blue whale models differ in multiple ways (Table 1). Model_B predicted absolute whale densities using line-transect survey data and 8-day composites of predictor variables in a generalized additive model (GAM) framework. The predictions were made at a 0.09° (approximately 10 km) spatial resolution for August–November. Model_H used whale presences and pseudoabsences, derived from telemetry data, in a generalized additive mixed model (GAMM) framework to predict monthly whale probability of occurrence at a 0.25° (approximately 25 km) spatial resolution. Hazen et al. (2017) scaled these predictions by an independent abundance estimate; we follow the terminology used in Hazen et al. (2017) and refer to the scaled predictions as habitat preference. We averaged the Model_H predictions from August to November to match the other predictions. Model_R predicted relative whale densities using line-transect survey data and predictor variables, averaged from late July to early December, in a GAM framework at a 10-km spatial resolution.

We followed the methods of Becker et al. (2016) and used the mean of the summer/fall blue whale predictions for 2001, 2005 and 2008 (the years with both line transect surveys and satellite tracks) from each model in the example analysis. Interannual variability has been shown to be the greatest source of uncertainty for cetacean SDMs in this region (Becker et al., 2016), and thus we calculated standard errors (SEs) using the three yearly predictions from each model. We imported these mean predictions into the GUI and created maps to compare prediction values, uncertainty, and distribution patterns (Figure 3). All three models predicted high blue whale densities in the Southern California Bight and along the central California coast. However, the Model_H predictions also had high values north of 40°N, where shipboard survey sightings and telemetry records of blue whales have been infrequent (Barlow & Forney, 2007; Becker et al., 2018; Irvine et al., 2014).

To create overlaid predictions, we imported a study area polygon that spanned the CCE and loaded the GUI-provided land polygon as the erasing polygon. We selected the equal area geometry of the Model_R predictions as the base geometry because polygon intersection and area calculations are most accurate in appropriate equal area coordinate systems. We also specified an overlap threshold of 50 percent, meaning if less than half of a base geometry polygon intersected with original predictions, the polygon was excluded from the ensemble. The different prediction value types of the overlaid SDMs were rescaled using an abundance estimate of 1648 blue whales, which was the mean of the Model_B predicted study area abundance for 2001, 2005, and 2008. For the Model_H predictions, this rescaling follows the method used in Hazen et al. (2017) to relate the predictions from the GAMMs to an independent study area abundance.

We evaluated SDM performance by calculating AUC and TSS using several binary validation datasets: (a) species presence/absence points derived from survey transects (Becker et al., 2016; 71
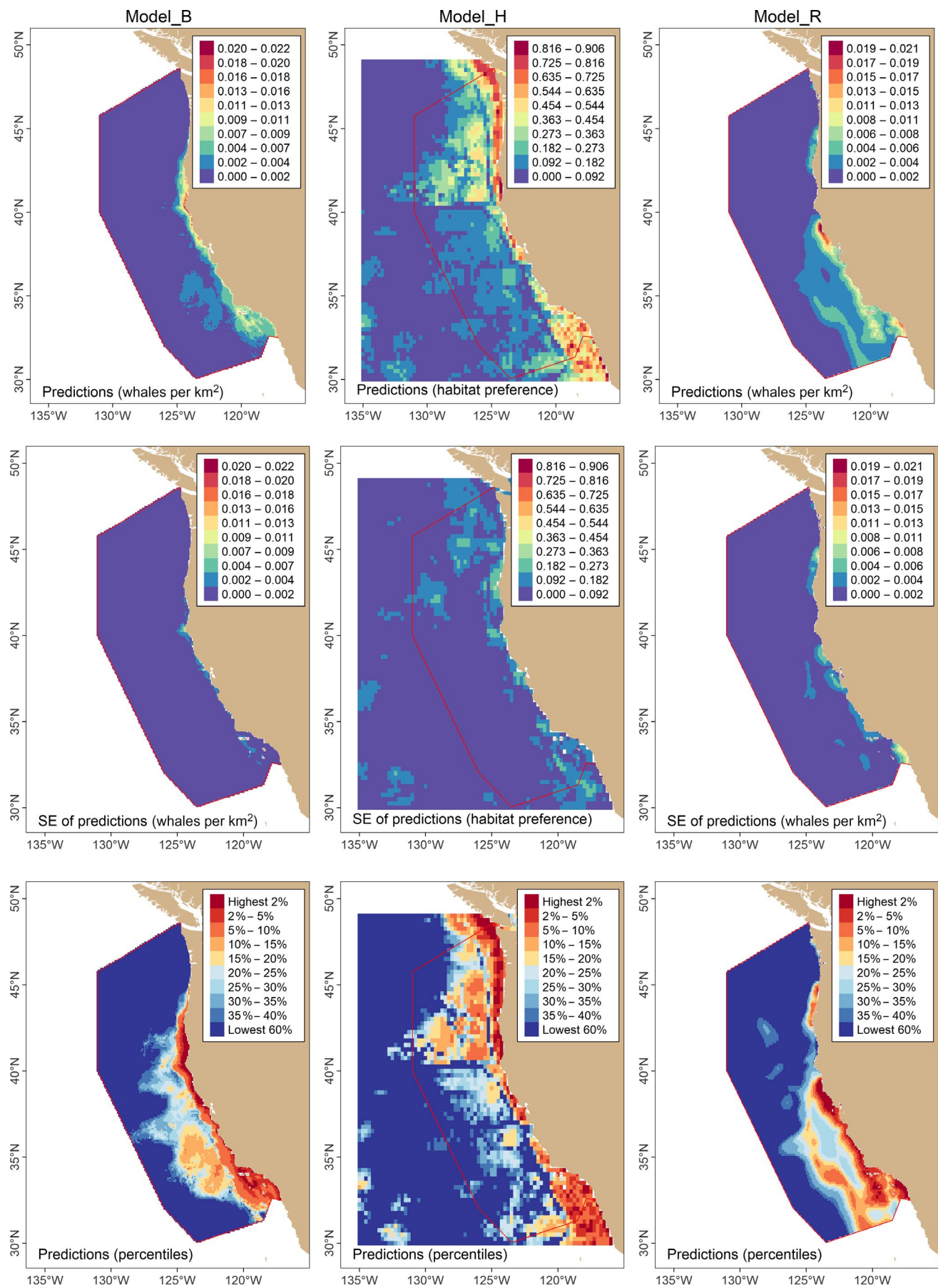
FIGURE 3 Maps of the original predictions and associated uncertainty for Model_B (Becker et al., 2016), Model_H (Hazen et al., 2017) and Model_R (Redfern et al., 2017). In the top row, predictions are colour-coded using the numerical prediction value for each SDM. In the middle row, the original standard error values have the same colour-coding as the top row. In the bottom row, the original predictions are colour-coded using relative percentages (i.e. percentiles). In all maps, the red line is the California Current Ecosystem study area boundary while the tan area represents the erasing polygon (i.e. the U.S. West Coast). For the top and middle rows, the units are whales per km^2, (left and right panels) and habitat preference (centre panels)

TABLE 3 Evaluation metrics, area under the receiver operating characteristic curve (AUC) and true skill statistic (TSS), for all example analysis predictions. The first two columns ('AUC' and 'TSS') contain metrics calculated using the combined validation dataset. Columns 'AUC-LT' and 'TSS-LT' contain metrics calculated using only the line transect validation dataset, while columns 'AUC-HR' and 'TSS-HR' contain metrics calculated using only the home range validation dataset. See the example analysis section for additional details

| Predictions	AUC	TSS	AUC-LT	TSS-LT	AUC-HR	TSS-HR
Becker et al. (2016) original	0.912	0.717	0.732	0.374	0.963	0.824
Hazen et al. (2017) original	0.734	0.414	0.620	0.284	0.772	0.471
Redfern et al. (2017) original	0.919	0.756	0.684	0.290	0.980	0.882
Becker et al. (2016) overlaid	0.916	0.742	0.732	0.380	0.967	0.856
Hazen et al. (2017) overlaid	0.735	0.406	0.620	0.286	0.772	0.460
Redfern et al. (2017) overlaid	0.919	0.756	0.684	0.290	0.980	0.882
Ensemble – unweighted	0.915	0.772	0.699	0.345	0.972	0.888
Ensemble – AUC-based weights	0.917	0.777	0.703	0.349	0.973	0.893
Ensemble – TSS-based weights	0.920	0.785	0.708	0.352	0.975	0.900
Ensemble – variance-based weights	0.888	0.670	0.713	0.344	0.936	0.764
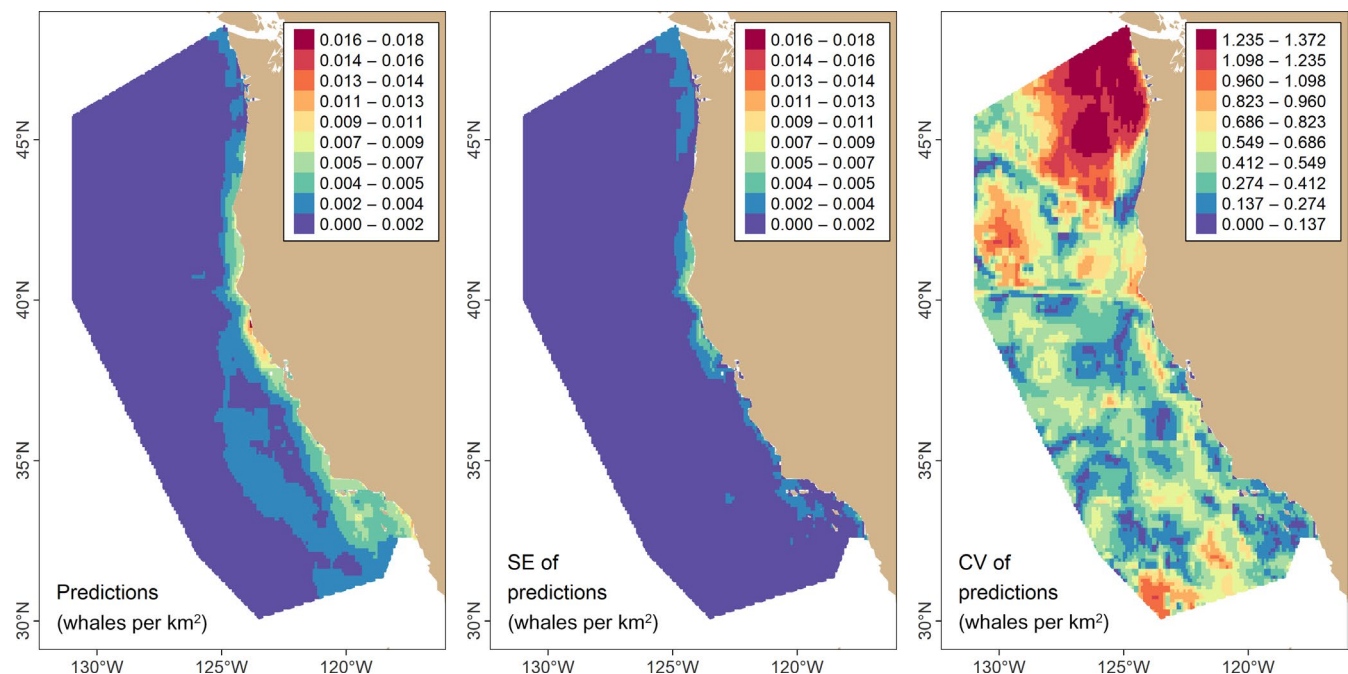
FIGURE 4 Maps of an unweighted ensemble of the overlaid predictions, the among-model standard error (SE) of the ensemble predictions, and the associated coefficient of variation (CV). The tan area represents the U.S. West Coast. The units are whales per km^2

presence and 7,368 absence points), (b) home ranges (90% isopleths) derived from 171 satellite-tagged blue whales (Irvine et al., 2014; 328 presence and 10,386 absence points), and (c) a combination of these two datasets. These validation data are not independent data, as the survey transects were used in Model_B and Model_R and the satellite telemetry data were used in Model_H. However, we are not

aware of any independent validation datasets for blue whales that span the CCE. Combining these data resulted in validation data with at least some novel presence and absence points for all predictions.

The home ranges represent areas of high use for blue whales, as identified by a long-term satellite tracking dataset (1994–2008; Irvine et al., 2014). To translate the home ranges into binary validation data,

we assumed that greater home range overlap indicates a higher likelihood of whale presence. The home ranges for all whales spanned most of the CCE, making it unrealistic for individual home ranges to indicate presence. Consequently, we used cut-off values for the number of overlapping home ranges to define presence and absence points. We performed a sensitivity analysis to identify cut-off values that maximized the AUC values of the overlaid SDM predictions. We defined the centroid of each base geometry polygon as a presence if it intersected with the home ranges of at least twenty whales, and an absence for the home ranges of nine or fewer whales. Points that intersected with ten to nineteen home ranges were not included in the validation data.

We calculated evaluation metrics using all three validation datasets to determine whether different predictions performed better with different validation datasets (i.e. the line transect or satellite telemetry data). The AUC and TSS values for the original and overlaid predictions were similar across validation datasets (Table 3), confirming that the overlay conserved the predicted distributions. Model_B and Model_R predictions had higher AUC and TSS values than Model_H predictions for all validation datasets (Table 3). However, the metrics indicated fair performance for the Model_H predictions and these predictions were included in all ensembles.

We created ensembles using several weighting methods: equal weights (i.e. unweighted), AUC-based weights (as in Oppel et al.,
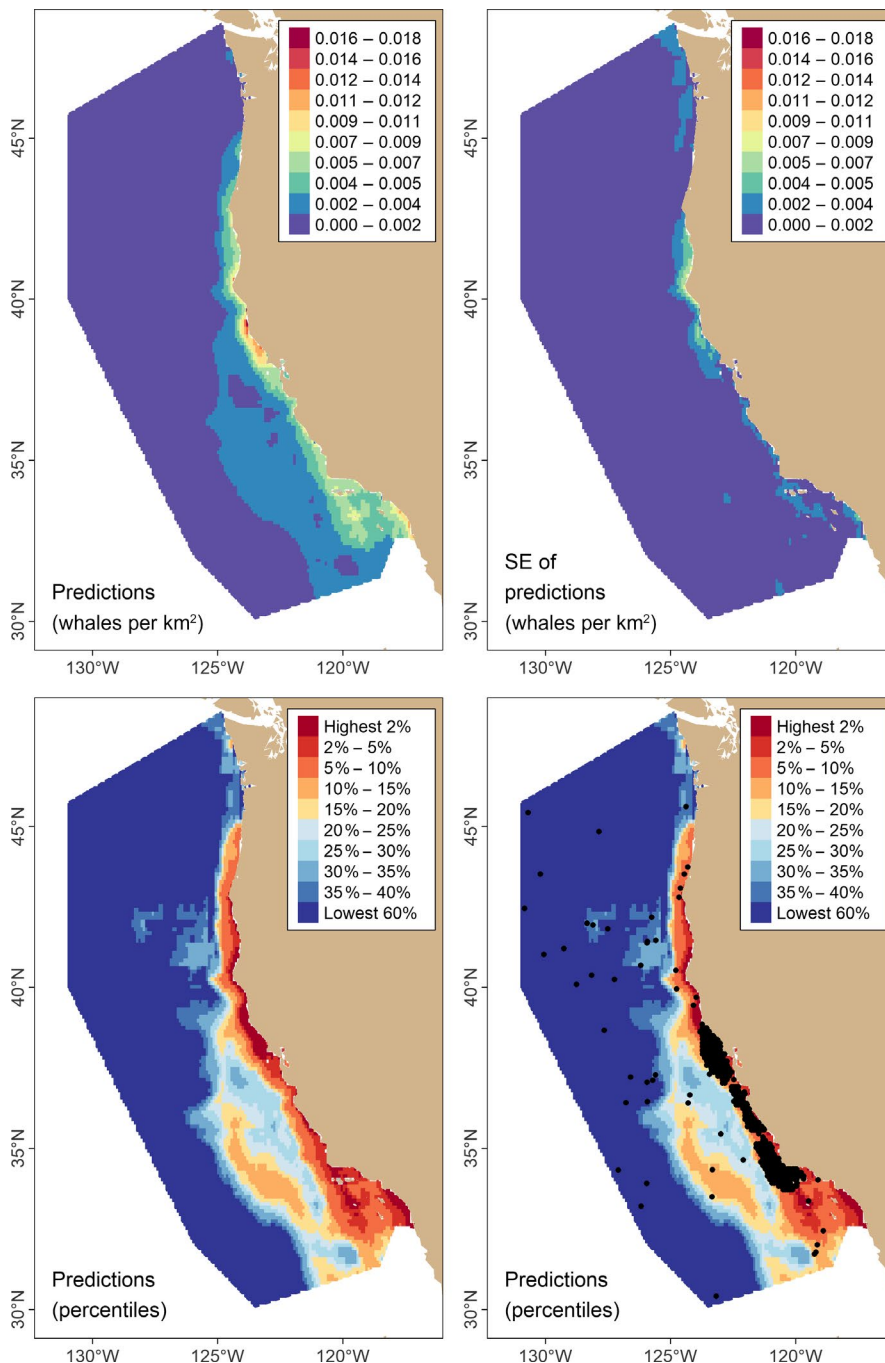

FIGURE 5 Maps of the ensemble created with weights based on TSS and associated uncertainty. In the top row, the prediction and standard error (*SE*) values are both colour-coded using the same numerical scale (whales per km^2). In the bottom row, the predictions are colour-coded using relative percentages (i.e. percentiles) and the right-most map includes the presence points from the combined validation dataset as black dots. The tan area represents the U.S. West Coast

2012), TSS-based weights (as in Scales et al., 2016), and weights calculated as the inverse of the prediction variance. The among-model uncertainty of the unweighted ensemble allowed us to examine spatial agreement between the predictions (Figure 4). We found relative agreement between the overlaid predictions south of 40°N, particularly in areas of high prediction values along the California coast and in the Southern California Bight. However, the ensemble uncertainty values were greater north of 40°N where only the Model_H predictions were high, suggesting that the northern ensemble predictions should be used with caution. The ensemble created using TSS-based weights had the highest evaluation metrics of the ensemble predictions, and mostly higher AUC and TSS scores than the original predictions (Table 3). Its distribution patterns also visually matched known blue whale habitat (Calambokidis et al., 2015; Figure 5), and thus we considered it the 'best' ensemble for this example analysis.

4 | DISCUSSION

Using the ESDM GUI, we successfully created an ensemble of mean blue whale predictions from Becker et al. (2016), Hazen et al. (2017), and Redfern et al. (2017) despite their different spatial resolutions, data sources, and prediction value types. The best ensemble predictions identified known blue whale habitat in the CCE, while generally improving evaluation metrics and minimizing biases associated with any single SDM. Researchers are frequently updating and improving SDMs (e.g. new blue whale models have been published by Becker et al., 2018, Abrahms et al., 2019, and Palacios et al., 2019 since we undertook our example analysis). Consequently, we do not intend our results to be considered the current best set of predictions for blue whales in the CCE. Instead, we present ESDM as a tool for creating and evaluating ensembles of SDM predictions for any species in a timely, straightforward and robust manner. This tool can allow managers and practitioners to avoid potentially ambiguous choices between models, and instead make more informed, science-based decisions.

The example analysis demonstrates the utility of ESDM and provides a framework and guidelines for ESDM users. These guidelines are important because ensemble predictions are not inherently better than the original predictions; ensemble quality is dependent on sensible inputs and informed user choices when creating the ensemble (Araújo & New, 2007). For example, ensembles can minimize the biases of individual SDMs by averaging predictions across SDMs. However, creating an ensemble of predictions with similar biases will result in a biased ensemble, and thus an ensemble should incorporate predictions from SDMs that rely on different methods and data sources. In addition, ESDM provides several ensemble methods because there is no consensus best method (Araújo & New, 2007; see Dormann et al., 2018 for an in-depth discussion of weighting schemes). An unweighted average is useful when determining reasonable weights is impractical, such as in a data-poor region. A weighted average allows users that know biases a priori, e.g. through evaluation metrics or expert knowledge, to specify the contribution of each set of predictions to the ensemble.

When used properly, ensembles reduce implicit uncertainty (e.g. model type or data source) by averaging predictions made using different model types or data sources (Jones-Farrand et al., 2011). However, ESDM also offers several ways to incorporate explicit uncertainty (e.g. the standard error of model predictions) when creating an ensemble. For instance, ensemble weights based on original prediction uncertainty reduce the contribution of predictions with high uncertainty to an ensemble. However, this feature should only be used with comparable uncertainty values; if a model underestimates uncertainty, then its predictions will contribute disproportionality to an ensemble. In addition, ESDM users can estimate among-model uncertainty to identify areas of spatial agreement and disagreement between the predictions, which can indicate regions of an ensemble with higher or lower levels of precision.

Conservation and management decisions often have short timelines, making it difficult to conduct new studies. ESDM allows decision-makers to quickly create an ensemble of SDM predictions using simple methods (Gregr, Palacios, Thompson, & Chan, 2019; Ward, Holmes, Thorson, & Collen, 2014). To create a meaningful ensemble, users must choose sensible original predictions and an appropriate ensemble method. For less obvious decisions, such as choosing a base geometry or deciding between AUC-based and TSS-based weights, ESDM provides a user-friendly tool for examining the sensitivity of an ensemble to user decisions. While it is important that all choices be realistic and ecologically sound, these sensitivity analyses enable users to better understand the underlying uncertainties in species distribution patterns and allow for informed decision-making.

AUTHORS' CONTRIBUTIONS

K.A.F., E.A.B., M.L.D., E.L.H., D.M.P. and J.V.R. conceived the project. S.M.W. wrote the ESDM R package and led the writing of the manuscript with help from K.A.F. and J.V.R. E.A.B., E.L.H., J.V.R. and D.M.P. provided data used in the example analysis. All co-authors provided feedback on both the ESDM package and manuscript and gave approval for publication.

DATA AVAILABILITY STATEMENT

Instructions for installing ESDM and accessing the GUI, along with code for creating applicable figures, are at https://github.com/smwoodman/eSDM. ESDM (https://doi.org/10.5281/zenodo.3371754) contains the example analysis data (https://doi.org/10.5281/zenodo.3365744) and a vignette performing the example analysis in R.

ORCID

Samuel M. Woodman (iD) https://orcid.org/0000-0001-6071-8186

Karin A. Forney (iD) https://orcid.org/0000-0002-9195-4701

Elizabeth A. Becker (iD) https://orcid.org/0000-0002-8858-122X

Elliott L. Hazen (iD) https://orcid.org/0000-0002-0412-7178

Daniel M. Palacios (iD) https://orcid.org/0000-0001-7069-7913

Jessica V. Redfern (iD) https://orcid.org/0000-0003-4042-3668

REFERENCES

Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, *3*, 177–187. https://doi.org/10.1111/j.2041-210X.2011.00141.x

Abrahms, B., Welch, H., Brodie, S., Jacox, M. G., Becker, E. A., Bograd, S. J., … Hazen, E. L. (2019). Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. *Diversity and Distributions*, *25*(8), 1182–1193. https://doi.org/10.1111/ddi.12940

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Annan, J. D., & Hargreaves, J. C. (2010). Reliability of the CMIP3 ensemble. *Geophysical Research Letters*, *37*, 1–5. https://doi.org/10.1029/2009gl041994

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*, 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Barlow, J., & Forney, K. A. (2007). Abundance and population density of cetaceans in the California Current ecosystem. *Fishery Bulletin*, *105*, 509–526.

Becker, E. A., Forney, K. A., Fiedler, P. C., Barlow, J., Chivers, S. J., Edwards, C. A., … Redfern, J. V. (2016). Moving towards dynamic ocean management: How well do modeled ocean products predict species distributions? *Remote Sensing*, *8*, 149. https://doi.org/10.3390/rs8020149

Becker, E. A., Forney, K. A., Redfern, J. V., Barlow, J., Jacox, M. G., Roberts, J. J., & Palacios, D. M. (2018). Predicting cetacean abundance and distribution in a changing climate. *Diversity and Distributions*, *25*, 626–643. https://doi.org/10.1111/ddi.12867

Calambokidis, J., Steiger, G. H., Curtice, C., Harrison, J., Ferguson, M. C., Becker, E., … Van Parijs, S. M. (2015). 4. Biologically important areas for selected cetaceans within U.S. waters – West coast region. *Aquatic Mammals*, *41*, 39–53. https://doi.org/10.1578/am.41.1.2015.39

Carretta, J. V., Helker, V., Muto, M. M., Greenman, J., Wilkinson, K., Lawson, D., …Jannot, J. (2018). Sources of human-related injury and mortality for U.S. Pacific West Coast marine mammal stock assessments, 2012–2016. U.S. Department of Commerce. NOAA Technical Memorandum NMFS-SWFSC-601. https://doi.org/10.7289/V5/TM-SWFSC-601

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2019). SHINY: Web application framework for R. R package version 1.3.0. Retrieved from https://CRAN.R-project.org/package=shiny (accessed April 1, 2019).

Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., … Hartig, F. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, *88*, 485–504. https://doi.org/10.1002/ecm.1309

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

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*, 38–49. https://doi.org/10.1017/S0376892997000088

Forney, K. A., Becker, E. A., Foley, D. G., Barlow, J., & Oleson, E. M. (2015). Habitat-based models of cetacean density and distribution in the central North Pacific. *Endangered Species Research*, *27*, 1–20. https://doi.org/10.3354/esr00632

Golding, N., & Purse, B. V. (2016). Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Methods in Ecology and Evolution*, *7*, 598–608. https://doi.org/10.1111/2041-210x.12523

Goodchild, M. F., & Lam, N.-S.-N. (1980). Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*, *1*, 297–312.

Gregr, E. J., Baumgartner, M. F., Laidre, K. L., & Palacios, D. M. (2013). Marine mammal habitat models come of age: The emergence of ecological and management relevance. *Endangered Species Research*, *22*, 205–212. https://doi.org/10.3354/esr00476

Gregr, E. J., Palacios, D. M., Thompson, A., & Chan, K. M. (2019). Why less complexity produces better forecasts: An independent data evaluation of kelp habitat models. *Ecography*, *42*, 428–443. https://doi.org/10.1111/ecog.03470

Grenouillet, G., Buisson, L., Casajus, N., & Lek, S. (2011). Ensemble modelling of species distribution: The effects of geographical and environmental ranges. *Ecography*, *34*, 9–17. https://doi.org/10.1111/j.1600-0587.2010.06152.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435. https://doi.org/10.1111/ele.12189

Hazen, E. L., Palacios, D. M., Forney, K. A., Howell, E. A., Becker, E., Hoover, A. L., … Bailey, H. (2017). WhaleWatch: A dynamic management tool for predicting blue whale density in the California Current. *Journal of Applied Ecology*, *54*, 1415–1428. https://doi.org/10.1111/1365-2664.12820

Hefley, T. J., & Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, *1*, 87–97. https://doi.org/10.1007/s40823-016-0008-7

Hijmans, R. J. (2019). RASTER: Geographic data analysis and modeling. R package version 2.8-19. Retrieved from https://CRAN.R-project.org/package=raster (accessed April 1, 2019).

Irvine, L. M., Mate, B. R., Winsor, M. H., Palacios, D. M., Bograd, S. J., Costa, D. P., & Bailey, H. (2014). Spatial and temporal occurrence of blue whales off the U.S. West Coast, with implications for management. *PLoS ONE*, *9*, e102959. https://doi.org/10.1371/journal.pone.0102959

Jones-Farrand, D. T., Fearer, T. M., Thogmartin, W. E., Thompson, F. R. III, Nelson, M. D., & Tirpak, J. M. (2011). Comparison of statistical and theoretical habitat models for conservation planning: The benefit of ensemble prediction. *Ecological Applications*, *21*, 2269–2282. https://doi.org/10.1890/10-1047.1

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, *15*, 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Merow, C., Wilson, A. M., & Jetz, W. (2017). Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography*, *26*, 243–258. https://doi.org/10.1111/geb.12539

Naimi, B., & Araújo, M. B. (2016). SDM: A reproducible and extensible R platform for species distribution modelling. *Ecography*, *39*, 368–375. https://doi.org/10.1111/ecog.01881

Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A. F., Miller, P. I., & Louzao, M. (2012). Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation*, *156*, 94–104. https://doi.org/10.1016/j.biocon.2011.11.013

Pacifici, K., Reich, B. J., Miller, D. A. W., Gardner, B., Stauffer, G., Singh, S., … Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology*, *98*, 840–850. https://doi.org/10.1002/ecy.1710

Palacios, D. M., Bailey, H., Becker, E. A., Bograd, S. J., DeAngelis, M. L., Forney, K. A., … Mate, B. R. (2019). Ecological correlates of blue whale movement behavior and its predictability in the California Current Ecosystem during the summer-fall feeding season. *Movement Ecology*, *7*, 1–21. https://doi.org/10.1186/s40462-019-0164-6

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, Retrieved from https://journal.r-project.org/archive/2018/RJ-2018-009/ (accessed April 1, 2019).

Pikesley, S. K., Maxwell, S. M., Pendoley, K., Costa, D. P., Coyne, M. S., Formia, A., … Witt, M. J. (2013). On the front line: Integrated habitat mapping for olive ridley sea turtles in the southeast Atlantic. *Diversity and Distributions*, *19*, 1518–1530. https://doi.org/10.1111/ddi.12118

R Core Team. (2019). R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Redfern, J. V., Mckenna, M. F., Moore, T. J., Calambokidis, J., Deangelis, M. L., Becker, E. A., … Chivers, S. J. (2013). Assessing the risk of ships striking large whales in marine spatial planning. *Conservation Biology*, *27*, 292–302. https://doi.org/10.1111/cobi.12029

Redfern, J. V., Moore, T. J., Fiedler, P. C., de Vos, A., Brownell, R. L., Forney, K. A., … Ballance, L. T. (2017). Predicting cetacean distributions in data-poor marine ecosystems. *Diversity and Distributions*, *23*, 394–408. https://doi.org/10.1111/ddi.12537

Sansom, A., Wilson, L. J., Caldow, R. W. G., & Bolton, M. (2018). Comparing marine distribution maps for seabirds during the breeding season derived from different survey and analysis methods. *PLoS ONE*, *13*, e0201797. https://doi.org/10.1371/journal.pone.0201797

Scales, K. L., Miller, P. I., Ingram, S. N., Hazen, E. L., Bograd, S. J., Phillips, R. A., & Thuiller, W. (2016). Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. *Diversity and Distributions*, *22*, 212–224. https://doi.org/10.1111/ddi.12389

Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*, 2053–2075. https://doi.org/10.1098/rsta.2007.2076

Thuiller, W., Georges, D., Engler, R., & Breiner, F. (2019). BIOMOD2: Ensemble platform for species distribution modeling. R package version 3.3-7.1. Retrieved from https://CRAN.R-project.org/package=biomod2 (accessed April 1, 2019).

Ward, E. J., Holmes, E. E., Thorson, J. T., & Collen, B. (2014). Complexity is costly: A meta-analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos*, *123*, 652–661. https://doi.org/10.1111/j.1600-0706.2014.00916.x