RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Quantifying apart what belongs together: A multi-state species distribution modelling framework for species using distinct habitats

Veronica F. Frans[1,2] (iD) | **Amélie A. Augé**[3,4] | **Hendrik Edelhoff**[1] |
Stefan Erasmi[5] | **Niko Balkenhol**[1] | **Jan O. Engler**[1,6]

[1]Department of Wildlife Sciences, University of Göttingen, Göttingen, Germany

[2]Workgroup on Endangered Species, University of Göttingen, Göttingen, Germany

[3]School of Surveying, University of Otago, Dunedin, New Zealand

[4]ARC Center of Excellence for Coral Reef Studies, James Cook University, Townsville, Australia

[5]Institute of Geography, University of Göttingen, Göttingen, Germany

[6]Zoological Research Museum Alexander Koenig, Bonn, Germany

Correspondence
Veronica F. Frans
Email: verofrans@gmail.com

Present addresses
Amélie A. Augé, Department of Conservation, Whangarei, New Zealand

Hendrik Edelhoff, Bavarian State Institute of Forestry, Freising, Germany

Jan O. Engler, Terrestrial Ecology Unit, Ghent University, Ghent, Belgium

Handling Editor: Jana McPherson

Abstract

1. Species distribution models (SDMs) have been used to inform scientists and conservationists about the status and change in occurrence patterns in threatened species. Many mobile species use multiple functionally distinct habitats, and cannot occupy one habitat type without the other being within a reachable distance. For such species, classical applications of SDMs might lead to erroneous representations of habitat suitability, as the complex relationships between predictors are lost when merging occurrence information across multiple habitats. To better account for the spatial arrangement of complementary—yet mandatory—habitat types, it is important to implement modelling strategies that partition occurrence information according to habitat use in a spatial context. Here, we address this issue by introducing a multi-state SDM framework.

2. The multi-state SDM framework stratifies occurrences according to the temporal or behavioural use of distinct habitat types, referred to as "states." Multiple SDMs are then run for each state and statistical thresholds of presence are used to combine these separate predictions. To identify suitable sites that account for distance between habitats, two optional modules are proposed where the thresholded output is aggregated and filtered by minimum area size, or through moving windows across maximum reachable distances.

3. We illustrate the full use of this framework by modelling the dynamic terrestrial breeding habitat preferences of the New Zealand sea lion (NZSL) (*Phocarctos hookeri*), using Maxent and trialling both modules to identify suitable sites for possible recolonization.

4. The Maxent predictions showed excellent performance, and the multi-state SDM framework highlighted 36–77 potential suitable breeding sites in the study area.

5. This framework can be applied to inform management when defining habitat suitability for species with complex changes in habitat use. It accounts for temporal and behavioural changes in distribution, maintains the individuality of each partitioned SDM, and considers distance between distinct habitat types. It also yields one final, easy-to-understand output for stakeholders and managers.

1 | INTRODUCTION

Many mobile species rely on different habitats throughout their lives. Their use can be for different resources, such as shelter and food, and access to multiple habitats is required throughout the day, across seasons or during different life cycle stages (Law & Dickman, 1998). These habitats can aggregate to a mosaic of neighbouring but ecologically distinct patches—each of which is crucial for a species' persistence. Habitat suitability is thus defined by the presence of two or more functionally distinct habitat types, and a lack of one cannot compensate for the other—even if the other is of superior quality. Quantifying suitability for each habitat type, while simultaneously accounting for the distance between them, is therefore a crucial task in defining overall habitat suitability.

Species distribution models (SDMs) (Franklin, 2010) have been applied in a wide range of ecological and evolutionary contexts, including conservation (Guisan et al., 2013; Johnson & Gillingham, 2005; Sousa-Silva, Alves, Honrado, & Lomba, 2014). They contrast environmental conditions at known species presence locations with the surrounding environment and probabilistically estimate potential distribution (Franklin, 2010). SDMs come in the form of various algorithms for different sampling situations and biases (Phillips & Dudík, 2008; Royle, Chandler, Yackulic, & Nichols, 2012; Thuiller, 2003). More and more, their implementation is improved through fine-resolution environmental predictors (see Cord, Meentemeyer, Leitão, & Václavík, 2013 and He et al., 2015 for reviews), and the fine-tuning of occurrences such as those derived from telemetry (e.g. Edrén, Wisz, Teilmann, Dietz, & Söderkvist, 2010; Roever, Beyer, Chase, & Aarde, 2014). Such fine-tuning has highlighted the importance of partitioning species occurrences to model functionally distinct habitats, as variable responses and importance can differ with changing seasons (e.g. Gschweng, Kalko, Berthold, Fiedler, & Fahr, 2012; Zuckerberg, Fink, La Sorte, Hochachka, & Kelling, 2016), behaviours (e.g. Brambilla & Saporetti, 2014; Roever et al., 2014) or life stages (Taboada, von Wehrden, & Assmann, 2013). However, multiple partitions also yield multiple predictions of distribution (e.g. 12 different predictions at the monthly scale; Bombosch et al., 2014), and a clear definition of a site's overall suitability is consequently lost.

To date, there have been few attempts at merging multiple predictions. Simply unifying predictions by calculating the mean across them is possible (Gschweng et al., 2012), but this could lead to misinterpretation, as high values in one location do not necessarily signify suitability across all seasons or behaviours. Averaging suitability also does not consider the availability of other suitable habitat types in nearby locations. Therefore, it would be important to develop a framework that can use multiple SDMs, account for the spatial context of distinct habitats, and provide one output that differentiates suitability across

predictions. Here, we present such a framework, referred to as a *multi-state SDM*, that:

1. accounts for different habitat types of species-specific importance;
2. identifies suitable sites that comprise a mosaic of habitats with distance criteria based on species behaviour or life cycle;
3. maintains the predictions and statistical integrity of each single-state SDM;
4. yields one final, easy-to-understand output for end-users.

We illustrate this framework through a case study on the endangered New Zealand sea lion (NZSL) (*Phocarctos hookeri*), a species that uses distinct terrestrial habitats across three temporal phases during its 2–3 month breeding period. In addition, we show possible extensions to the basic framework that can be useful for conservation and management applications, such as extracting suitable sites of a minimum area size that encompass all distinct habitat types required by a species, and identifying suitable habitats within a species' range of movement. We later list other situations for which multi-state SDMs are useful, and discuss the differences between this framework and others. We also provide a tutorial (Appendix S1) with step-by-step instructions to implement this framework in R (R Core Team, 2015).

2 | THE MULTI-STATE SDM FRAMEWORK

In its broadest sense, a multi-state SDM is a three-step approach: first, states are defined based on temporal or behavioural parameters of a mobile species; then, SDMs are calculated separately for each habitat type used by the species; lastly, these different SDMs are combined to identify sites where the basic ecological requirements for occurrence are met (i.e. the overall suitability across multiple suitable habitat types). In the following, we describe basic data needs and guide through the three main analytical steps. Figure 1 illustrates the overall framework and the steps detailed in later sections.

2.1 | Step 1: State occurrences and data requirements

The "multi-state" aspect of this framework refers to the occupancy of multiple distinct habitats over time or for different types of use by a species. An occurrence is therefore defined as a confirmed location of an animal at a recorded time, with or without behavioural information. In designing a multi-state SDM, occurrences are separated according to time or behaviour, into what we hereafter refer to
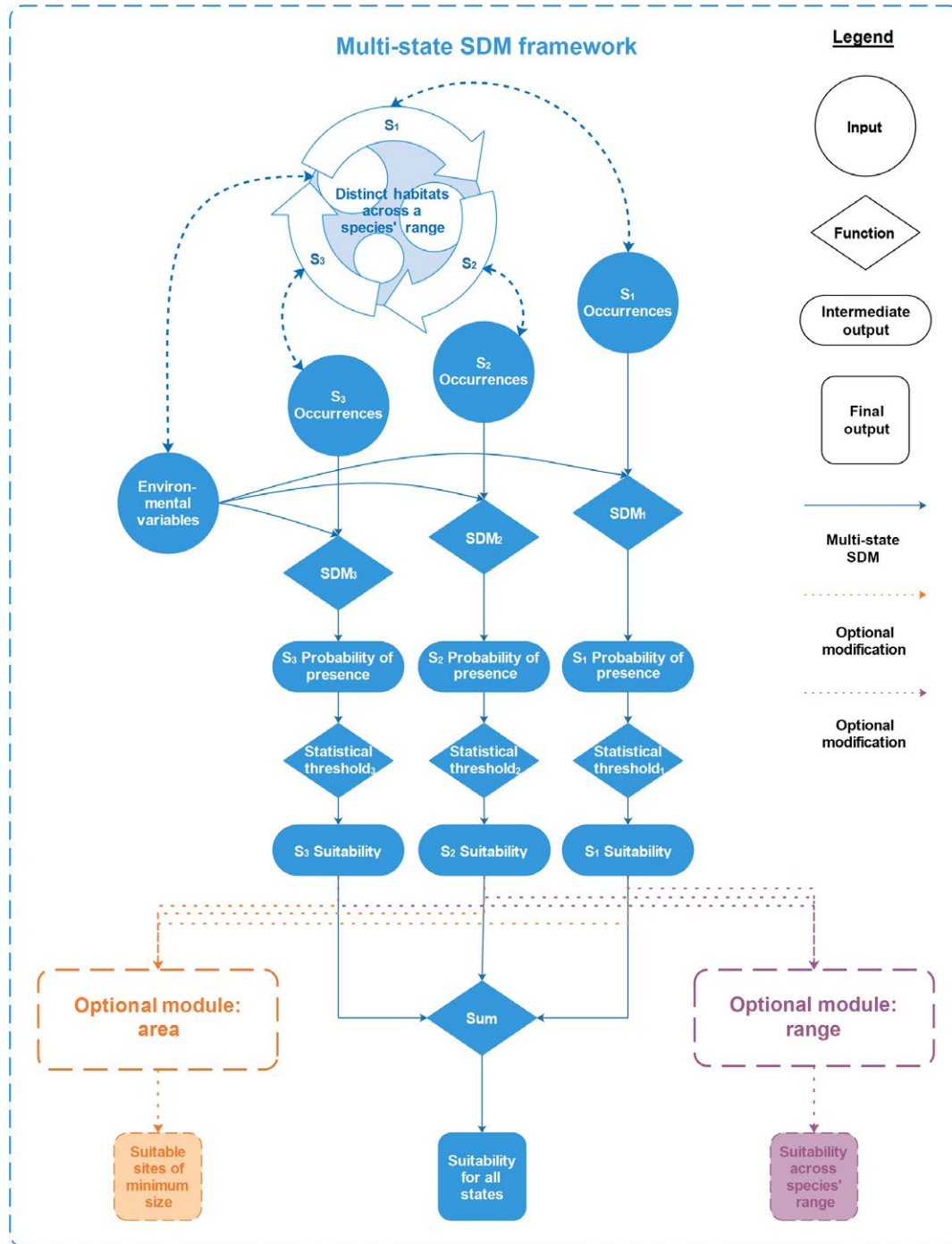
FIGURE 1 Illustration of the multi-state SDM framework (blue), with optional modules for site selection by minimum area size (Area Module; orange) and the maximum reachable distance between habitat types (Range Module; purple). The state (S_n) is defined as the unit through which species occurrences are partitioned (i.e. time interval or behaviour). The steps for the optional modules are further expanded in the Supporting Information (Appendix S1, Figure S1)

as *states* (Patterson, Thomas, Wilcox, Ovaskainen, & Matthiopoulos, 2008). Although the term, *state*, has been used in other SDM frameworks to describe habitat condition (e.g. source and sink habitat suitability; Naves, Wiegand, Revilla, & Delibes, 2003; Nielsen, Stenhouse, & Boyce, 2006), we adopted this term from the state-space modelling definition, where a state describes an "attribute" of a system (Patterson et al., 2008), which in this case is behaviour or time.

High-quality spatio-temporal occurrence information is a key component in multi-state SDMs. For species occupying different habitats over time (e.g. different phases of a breeding cycle, or during night-time), occurrence data would simply require temporal information. If, however, habitat use is related to certain behaviours (e.g. feeding, nesting), behavioural information is required. This information may be derived from detailed analyses of bio-logging data (e.g. Edelhoff,

Signer, & Balkenhol, 2016; Patterson et al., 2008; Roever et al., 2014) or from manual records if occurrences originated from on-the-ground surveys (e.g. Augé, Chilvers, Mathieu, & Moore, 2012).

Multi-state SDMs require that the spatio-temporal resolution of occurrence data is fine enough to distinguish the states of interest (variable by species), or that behavioural information is already included with the localities. The records should therefore have a grain that is smaller than the species' average home-range, but is also fine enough to model the use of distinct habitats within it, as finer grains increase predictive performance (Guisan, Graham, Elith, & Huettmann, 2007). The ability to model state occurrences also depends on the quality of the environmental data used. Quality refers to the resolution of the environmental data in relation to the occurrence information (e.g. Mitchell, Monk, & Laurenson, 2017), the effects of resolution on the prediction (e.g. Cord & Rödder, 2011; Filz, Schmitt, & Engler, 2013; He et al., 2015), and which environmental variables are used. The choice of environmental variables should be reflective of the ecology of the species and what influences their occurrence across multiple habitat types (van Gils, Westinga, Carafa, Antonucci, & Ciaschetti, 2014).

2.2 │ Step 2: Multiple SDMs

Occurrences from different states are used to run separate, single-state SDMs. Consistent model settings are used to yield comparable, state-specific predictions (Figure 1). We suggest that the settings remain the same for each model because this framework assumes that the occurrences and environmental data are fine enough for the SDM to generate variable responses solely from the single-state occurrences. These SDMs yield multiple predictions of occurrence for each respective state.

2.3 │ Step 3: Combining SDMs to have a multi-state output

Once the probability of presence is determined from each state, the predictions are combined into one map of overall suitability. This is done using occurrence thresholds to combine the most suitable habitats (Liu, Berry, Dawson, & Pearson, 2005). Thresholds can be calculated from presence-only or presence-absence data. They evaluate true and false positives (presences) and true and false negatives (absences) to yield a "cutoff" value for presence and absence, or, suitability and unsuitability (Liu et al., 2005). These thresholds are specific to each SDM run, and hence to each state, thereby allowing for suitable sites to be identified on a state-by-state basis.

Many thresholds exist (e.g. *10th percentile training presence*, *mean predicted value*, *maximizing kappa*), and their use depends on the SDM algorithm and the type of occurrence information used. Guidelines for their selection can be found in Liu et al. (2005) and Liu, White, and Newell (2013). Once an appropriate threshold is selected and applied to each model, the predictions of suitability are combined by reclassifying suitable pixels according to a reclassification scheme exemplified in Table 1. The reclassified layers are then summed to yield a final output of overall suitability for the pixel. This summation shows the degree of suitability for each pixel (e.g. an area's suitability for one, none or all states), which, as opposed to having multiple separate predictions, allows for all predictions to be evaluated at once (Table 1).

2.4 │ Optional modules: Identifying suitable sites by size or range

Supplementary to this framework, optional modules could incorporate the distance between suitable habitats for each state. In Step 3 of the framework, overlap is emphasized by adding the thresholded predictions together. However, it is not necessarily overlapping suitability, but rather the distance between other (available) functional habitat types that can be important for some species, and whether or not they can be reached. Distance could thus be incorporated in one of two ways: (1) defining minimum area size (Area Module), or (2) moving windows analysis across an individuals' maximum range (Range Module; Figure 1).

For the Area Module, minimum area size refers to an area's capacity for n individuals; as long as the suitable habitats for each state are contiguous within x units of distance, the site is suitable for the species. It is assumed that under those conditions, the species is easily able to reach these habitats across the landscape or seascape (i.e. there is no specific path or cost to movement). To generate this, unsuitable pixels are reclassified to null (NA) for each state, and contiguous pixels of the remaining suitable values are aggregated in a Geographic Information System (GIS). Next, a minimum mapping unit (MMU) is defined by calculating the minimum area required for n individuals per state, S, and converting it into a minimum count of contiguous cells to be aggregated. The MMU is calculated as:

$$MMU_S = \frac{(a/n_1) \times n_2}{r},$$

where a/n_1 is the average density of individuals at one time, calculated from occurrence data, with n_1 representing the number of individuals per given area, a; n_2 is the minimum number of individuals set by the researcher; and r is the resolution of the raster layer, in the same unit as a.

TABLE 1 An example of reclassified pixel values for thresholded predictions of a multi-state SDM of three states (S_1, S_2 and S_3). The sum of these values provides, in this case, eight suitability combinations, indicating whether a pixel is suitable or unsuitable for one, two or all states

S_1	S_2	S_3	Sum	State suitability
0	0	0	0	None
1	0	0	1	S_1 only
0	10	0	10	S_2 only
0	0	100	100	S_3 only
1	10	0	11	S_1 and S_2
1	0	100	101	S_1 and S_3
0	10	100	110	S_2 and S_3
1	10	100	111	All states

Aggregated areas ≥MMU are then extracted to yield suitable sites for n individuals for one state. Aggregations are made for subsequent states, and then all state MMUs are combined, aggregated once again and filtered to a minimum total area size across all states (i.e. $MMU_1 + MMU_2 + MMU_3$, and so on). The final map consists of suitable sites expected to encompass enough areas of each suitable habitat type across states to host n_2 individuals.

Additionally, for cases where functionally distinct habitats do not necessarily need to be overlapping or contiguous, the Range Module can be applied. Unlike the Area Module, here, non-contiguous suitable pixels are included in site selection (i.e. patches of unsuitable pixels are expected). Access to these habitats may therefore be restricted, but this is not calculated. In the Range Module, moving windows are used to define sites where suitable state habitats are found within a maximum reachable distance (e.g. Downs, Gates, & Murray, 2008). Moving windows gather information from surrounding pixels across a defined distance (range) to make calculations (e.g. mean, minimum, maximum). Here, moving windows are used to count the number of unique states per window (variety).

To calculate variety, the width of each state-to-state window is first defined by dividing the maximum reachable distance from S_1 to S_2, S_2 to S_3, S_3 to S_4 and so on, by the resolution of the pixel, and rounding the results to the nearest uneven integer (the focal pixel needs to be centered). The thresholded state values (e.g. S_1 and S_2; see Table 1) are summed, and variety is calculated from this raster. This value is then reclassified to extract pixels with a variety count ≥2, indicative of sites comprising both state habitats within the range. This value is then summed with the next thresholded state (e.g. S_3), using the window width of S_2 and S_3 for the next variety calculation, and a count ≥2 is once again extracted to represent variety for all three states. This procedure forms a nested moving windows analysis, and continues with reclassifications and extractions from state to state (S_3 to S_4, S_4 to S_5, and so on) and for each window width. The maximum number of windows is thus one less than the total number of modelled states, due to the first pairing. The minimum area size of the resulting sites will be no smaller than the smallest window width among the states, as a result of the extractions.

The framework for these modules is further explained in Appendix S1.

3 | APPLICATION: THE NEW ZEALAND SEA LION

We illustrate the applicability of the proposed multi-state SDM framework using the endangered NZSL as a case study. Once found throughout the mainland (Childerhouse & Gales, 1998), breeding colonies are now only left in two of New Zealand's Subantarctic Islands (Robertson & Chilvers, 2011). Conservation priorities for this species aim at increasing population growth and distribution (Department of Conservation, 2009). Although a rare event, recolonization on the mainland is possible (Lalas & Bradshaw, 2003); if suitable sites can be identified, proactive management and education can be used to facilitate the recolonization

process. Recently, analyses on species' habitat preferences and GIS-based multi-criteria analysis were used to try identifying such sites (MacMillan, Moore, Augé, & Chilvers, 2016), but multi-state SDMs could provide more in-depth modelling and further support conservation actions (Fourcade, Engler, Besnard, Rödder, & Secondi, 2013), especially if the species' distinct shifts in terrestrial habitat preferences are accounted for when modelling habitat suitability.

The NZSL's shifting terrestrial habitat preferences over the breeding period occur at three temporal phases: the breeding phase, transition phase and dispersion phase (Augé, Chilvers, Moore, Mathieu, & Robertson, 2009). In the first phase, females remain on sandy beaches up to 100 m from the coast during a period slightly before and after the pups are born (Augé et al., 2012). The females then begin to move into grassy areas behind the beach, representing the transition phase. By the dispersion phase, mother and pup are found from 1,100 m to 2,000 m inland in the forest. As the phases occur in distinct habitat types within certain distances of each other, the NZSL serves as a good example for the multi-state SDM approach, where each phase is a state. The Area and Range Modules can be illustrated with this case study to define potential sites for recolonization.

3.1 | Study area

The chosen study area to illustrate the framework is a southern portion of South Island New Zealand, the small islands that surround it, and Stewart Island (Figure 2). The NZSL has been found as far as 2,000 m from the coast (McNally, Heinrich, & Childerhouse, 2001). We therefore considered all areas 2,500 m inland from the coast for the SDM, covering an area of c. 5,863 km^2.

3.2 | Occurrence information

Female NZSL presence records (occurrences) were collected over two consecutive breeding periods at one of the species' remaining breeding colony sites in the Auckland Islands (50º28′S, 165º52′E; see Augé et al., 2009 for details). NZSL occurrences were taken from daily, on-the-ground surveys, with positions recorded on a handheld Garmin 12 GPS (Garmin International, USA), with an average accuracy of 7 m. These occurrences were filtered to represent successful breeding females only (c.f. Augé et al., 2012).

As the occurrences are spatio-temporally independent, taken once a day (Augé et al., 2009), NZSL presence was assessed at the population level. We therefore separated the occurrences into three states by median date at which the NZSL's spatial behaviour and habitat preferences changed (see Augé et al., 2009 for how this was evaluated): breeding (S_1; December 6 to January 18; 2,247 occurrences), transition (S_2; January 19 to February 18; 1,333 occurrences) and dispersion (S_3; February 19 to March 21; 293 occurrences; Figure 2).

3.3 | Environmental data

Following analyses and descriptions of female NZSL breeding habitat requirements in McNally et al. (2001) and Augé et al.
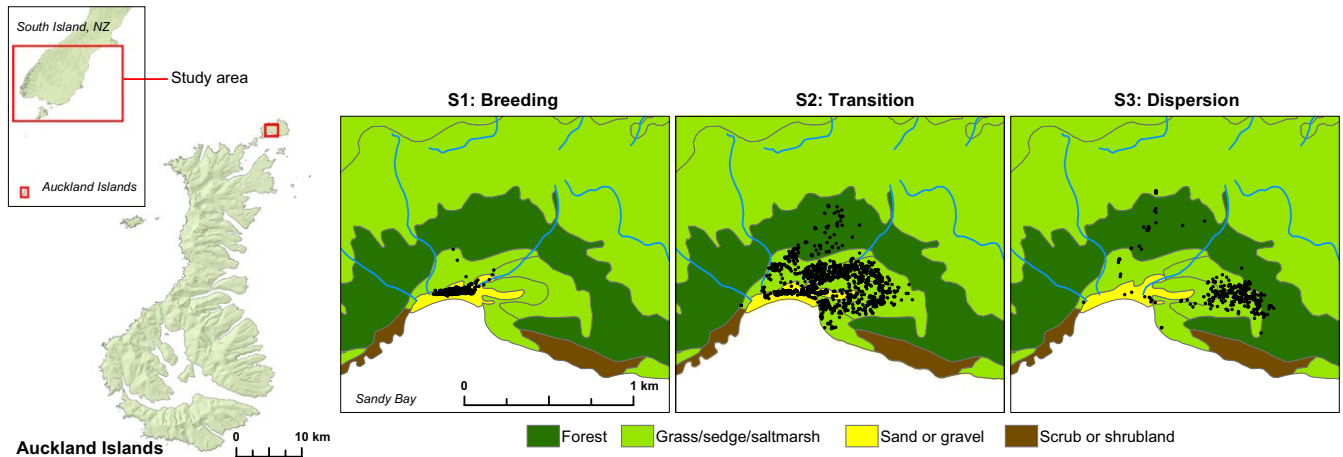
FIGURE 2 Study area (top left) and NZSL occurrences (black dots) across states on Sandy Bay (Auckland Islands), during the 2001/02 and 2002/03 breeding seasons (December to March)

(2009, 2012), the following eight environmental variables were used to model habitat preferences: land cover, slope, cliff edges and Euclidean distances from the coastline, inland water bodies, sand, grass and forest (see Table S1 for source and data preparation). All variables were prepared at 25 m resolution using ArcGIS 10.2 (ESRI, Redlands, CA) in New Zealand Transverse Mercator projection (EPSG 2193). This was the finest resolution available for the digital elevation model from which the slope was derived. Additionally, this resolution was at the closest possible scale to the occurrence resolution and its spatial error (±7 m), and allowed us to account for abrupt landscape changes important for the species (e.g. presence or absence of cliffs, steep slopes; Augé et al., 2012), while also covering a large study area. Generally, locational uncertainty is expected to affect models that rely on fine-scale predictors (e.g. Mitchell et al., 2017). Here, we expect this effect to be minimal, as the location error is lower than the resolution of the environmental variables.

3.4 | Species distribution modelling

We generated SDMs for each state using Maxent v. 3.3.3k (Phillips & Dudík, 2008), a statistical machine-learning algorithm based on the principle of maximum entropy (Elith et al., 2011). It is amongst the top-performing approaches and can be used under various conditions (Hernandez, Graham, Master, & Albert, 2006; Wisz et al., 2008). We ran Maxent from R (version 3.2.0; R Core Team 2015) using the DISMO package (Hijmans, Phillips, Leathwick, & Elith, 2013) under default settings, except we used only hinge and categorical features to help smooth the variable responses and reduce noise (Elith et al., 2011; Merow, Smith, & Silander, 2013). We ran Maxent with 100 iterations for each state by randomly selecting 100 state occurrences for training and 33 occurrences for testing, and projecting the single runs to the projection area. From this, we calculated the arithmetic mean for each state, with clamping (Merow et al., 2013) and a logistic output (Phillips & Dudík, 2008) enabled. We assessed the performance of each single-state SDM by first calculating the mean AUC value

(Area under the Receiver Operating Characteristic Curve; Hosmer & Lemeshow, 2000). We then extracted and calculated mean variable response curves, jackknife of regularized training gain, and percent contribution and permutation importance, which were generated by Maxent.

3.5 | Building a multi-state SDM

As the most suitable NZSL breeding habitats are those that contain suitable locations for all three states, we combined the final predictions of single-state suitability. In order to combine them, the predictions were first thresholded by maximizing the sum of sensitivity and specificity (*maxSSS*; Liu et al., 2013). Through Maxent's standard output, we extracted these values from each run and calculated the mean and standard deviation (*SD*) for each state.

We applied the thresholds for each state by reclassifying the predictions as 0 for all unsuitable pixels, and 1, 10 and 100 for suitable pixels of each subsequent state (Table 1). We then took the sum of the three thresholded state predictions to yield a grid of eight possible combinations of suitability (Table 1).

3.6 | Area Module: Identifying suitable habitats of minimum area size

We extracted contiguous pixels representing suitable areas ≥802 cells (0.50 km^2) in size from the dataset as our total MMU. This MMU is the size of a breeding site that would provide enough area of each habitat for the different states for 50 breeding females (this number represents an established sustainable breeding colony). Female NZSL densities during the breeding phase are as high as 85 individuals per 100 m^2, which lowers to 30 individuals per 100 m^2 by the transition phase, and reaches as low as 0.01 females per 100 m^2 by the end of the dispersion phase (Augé et al., 2009). Using these densities, the MMU for each state was calculated as 1 cell for S_1 and S_2, and 800 cells for S_3 (see Appendix S1 for calculations).

3.7 | Range Module: Identifying suitable habitats within a defined range

Another option to search for areas that were not necessarily contiguously suitable or of a minimum area size was the Range Module. As three states were modelled, a total of two moving windows were generated, from which the variety, or, count of unique suitability values per window, was calculated. Representing the movement from S_1 to the end of S_2, first, the sum of the first two thresholded state SDMs were taken and then a moving window width of ~620 m (25 pixels) was used (Augé et al., 2009). From this first neighbourhood raster, we then extracted and reclassified all pixels with a count of ≥2 (i.e. areas with suitable sites for two states within a range of 620 m) to 2, and added this to the third thresholded state SDM. From S_2 to S_3, the maximum inland movement of a breeding female NZSL was 1,100 m (Augé et al., 2012), so the second window width was calculated as 19 pixels (480 m; 1,100 m minus 620 m). We then reclassified the results to retain variety values of at least 2 to yield locations containing all three state habitats within a total range of 1,100 m.

4 | RESULTS

4.1 | SDM performance and evaluation

All three models had high mean AUC_{Test} (± SD) scores for S_1, S_2 and S_3 at 0.9995 ± 0.0018, 0.9986 ± 0.0004 and 0.9987 ± 0.0005, respectively. Variable responses differed across states, as expected of the species' shifting habitat preferences. The largest differences were seen in the S_1 model compared to S_2 and S_3, while S_2 and S_3 shared similar variable responses (see Table S2).

MaxSSS values (± SD) for S_1, S_2 and S_3 were 0.35 ± 0.17, 0.04 ± 0.03 and 0.12 ± 0.04, respectively. Areas above these thresholds covered varying amounts of the study area (0.14%, 1.01% and 1.68% respectively for each state).

4.2 | MMUs and sites of minimum area size

After applying the Area Module, 36 potential suitable breeding colony sites comprising enough habitats to hold 50 females for all three states were found (Figure 3). Suitable sites ranged from 0.51 km^2 to 12.95 km^2, with average sizes of 2.26 ± 2.46 km^2.

4.3 | Moving windows and range

From the Range Module, 77 sites with suitable habitats for all three states were found (Figure 3). These sites ranged from 0.01 to 12.05 km^2, with average sizes of 1.35 ± 1.83 km^2.

5 | DISCUSSION

Quantifying functionally distinct habitats is important but challenging when predicting species distributions for conservation management

using SDMs. The multi-state SDM framework presented here addresses this issue by maintaining multiple fine-tuned predictions through the use of statistically sound thresholds and their combination. In cases where suitable habitat types do not need to overlap, but rather can be within a certain distance, the Area and Range Modules offer additional, enhanced, outputs. The modules allow for defining suitable sites of a minimum area or within a species' maximum range of movement, which are crucial for management decisions.

As demonstrated in our case study of the NZSL, the multi-state SDM framework was an efficient way to identify suitable breeding sites for possible recolonization and improve management. Defining minimum area and habitat type contiguity through the Area Module was also beneficial for prioritizing amongst suitable sites. The Range Module can provide insight on habitat availability if unsuitable patches are acceptable within a range, dependent on the species' movement. Further evaluation of the sites identified through this framework and its application throughout the coasts of New Zealand could be beneficial for current management efforts aimed at facilitating the spread of the NZSL on the mainland (Department of Conservation, 2009).

5.1 | Future applications

This framework could be applied to model multiple functional habitats occupied by mobile species across temporal or behavioural states. For example, temporal partitions could be used to model suitable habitats for cetaceans across seasons (e.g. Edrén et al., 2010; Table 2). Behavioural partitions could be used to model foraging and nesting habitats of other species, such as woodpeckers (Brambilla & Saporetti, 2014). Identifying lek and breeding or foraging sites for grouse species could also involve modelling behavioural states, as these sites are distinct and their habitat requirements differ (Connelly, Schroeder, Sands, & Braun, 2000; Knick, Hanser, & Preston, 2013). This can be seen with migrating birds (Zuckerberg et al., 2016) and land species exhibiting sedentary and dispersal behaviours, as well. Having such large ranges across habitat types, moving windows could also be incorporated in models for these species. Future research could also try to incorporate landscape resistance for species movement between functionally distinct habitat types (see Kindlmann & Burel, 2008; McRae, 2006). Further, nesting and foraging sites could be modeled for sea turtles as seascape and landscape SDMs and unified through the exemplified Range Module. Such a combination of seascape and landscape SDMs could be conducted for the NZSL to evaluate the proximity between identified suitable breeding sites and suitable marine habitats, using foraging locations (Augé, Chilvers, Moore, & Davis, 2011), as well as marine environmental variables. These and other examples are listed in Table 2 (see Law & Dickman, 1998).

5.2 | Framework limitations and comparison with other approaches

Although applicable to many species, limitations to this framework exist. As previously mentioned, the multi-state SDM approach necessitates occurrence information that includes a way to define
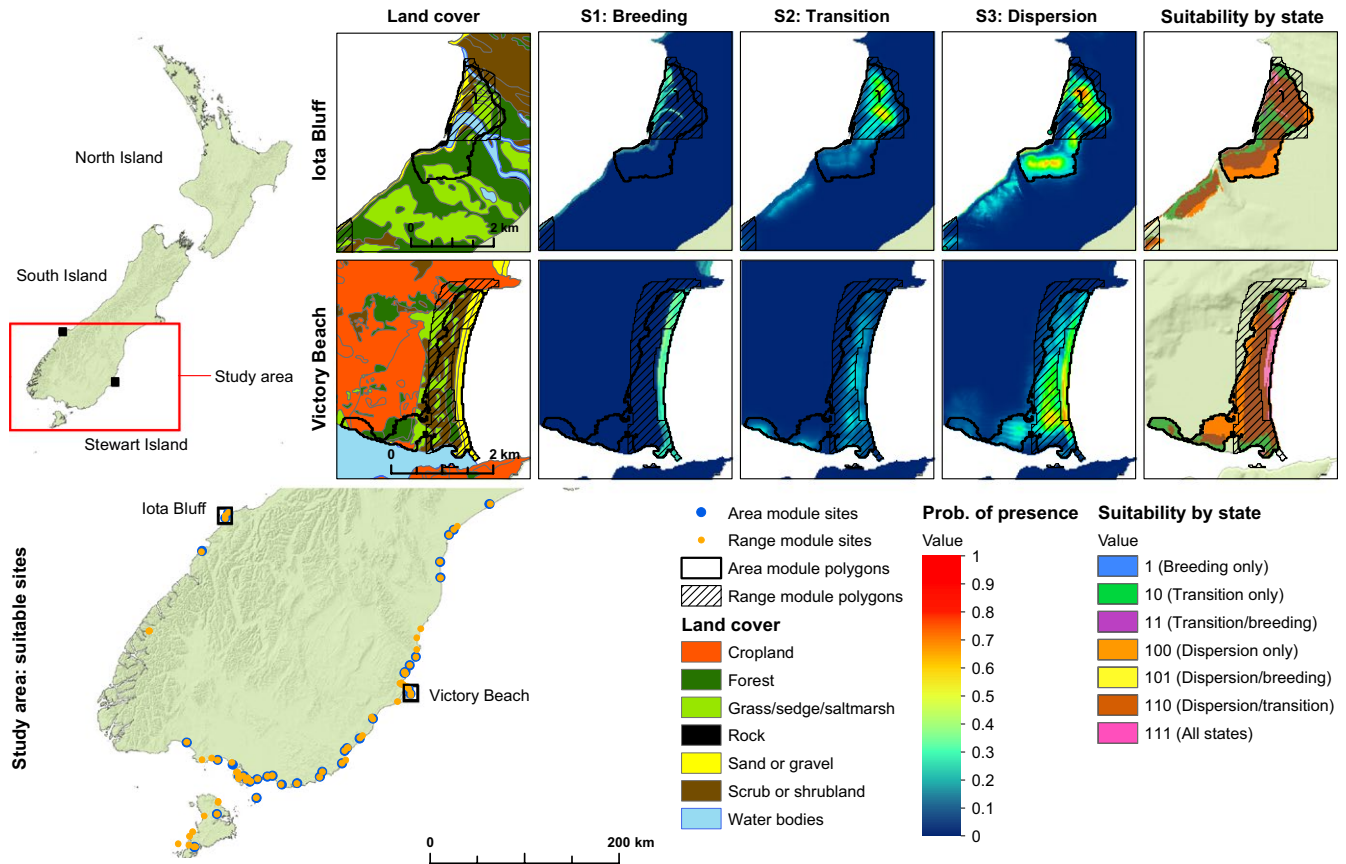
FIGURE 3 Locations of 36 potential suitable breeding habitats for the NZSL derived from the Area Module (blue), and 77 suitable habitats calculated from the Range Module (orange; bottom left). Examples of the state predictions and overall state suitability are shown for two selected sites as an illustration

behavioural or temporal states. Telemetry data, for example, provide occurrences from which behavioural information or resource selection functions can be derived (Abrahms et al., 2017; Patterson et al., 2008; Roever et al., 2014). However, inferences on the population (i.e. the SDM as a whole) are then made from individual-based data; to solve this, the occurrences must therefore be spatio-temporally independent (see e.g. Edrén et al., 2010).

Using fine-grained occurrences requires that predictors be of similar resolution (or, representative of the species' interaction with the environment)—a general means to study species-environment relationships and minimize uncertainty in modelling. Uncertainty affects predictions when the spatial error is high, the sample size is small, or an inappropriate algorithm is used (Graham et al., 2008; Mitchell et al., 2017). Here, with the NZSL example, we minimized the effects of such uncertainty by (1) ensuring that the spatial error is lower than the resolution of the predictors; (2) selecting records from a large sample size; and (3) choosing Maxent as our algorithm, which is less sensitive to locational error (see Graham et al., 2008). We suggest that similar precautionary actions be made based on an assessment of data type and quality before the framework is applied.

In terms of the NZSL, quite simply, one could argue that the habitat types are so spatio-temporally adjacent that perhaps a model with

no partitions would have rendered similar results. However, despite their proximity, these habitat types and the spatial behaviour of the NZSL vastly differ between states, so partitioned SDMs allowed for fine-scaled modelling of these habitat preferences, as exemplified by the varying variable responses across states (Table S2). Also, high values of suitability do not necessarily imply overall suitability, as training occurrences can be dominated by one state—especially if more abundant or highly clustered—as seen with occurrences from S_1 vs. those of S_3. Nevertheless, comparing outputs, variable responses and importance with and without state partitions is helpful in validating the use of this framework. This is exemplified in the supplementary tutorial (Appendix S1).

If other, more simplified ways to model multi-state species distributions are possible, we encourage further exploration. SDM techniques have been advancing alongside technology, and more and more, it is being proven that multiple, state-by-state algorithms are indeed necessary (e.g. Gschweng et al., 2012; Jaberg & Guisan, 2001; Zuckerberg et al., 2016); single, unpartitioned SDMs cannot appropriately account for changing species-environment relationships over time or behaviours.

To our knowledge, attempts at unifying multiple predictions have been minimal. Previous studies have unified multi-temporal SDMs by calculating the mean (e.g. Gschweng et al., 2012; see

TABLE 2 Examples of mobile species for which the multi-state SDM framework could be applied, and the types of partitions that could be implemented

Partition	S₁	S₂	S₃	S₄	Example
Behavioural	Encamped	Exploratory	–	–	Elephants (Roever et al., 2014)
	Feeding	Refuge	–	–	Kangaroos (Coulson, 1993)
	Haul-out	Foraging	–	–	Pinnipeds
	Lekking	Breeding/Foraging	–	–	Grouse (Connelly et al., 2000)
	Nesting	Foraging	–	–	Bats (Law, 1993); woodpeckers (Brambilla & Saporetti, 2014)
	Sedentary	Dispersal	–	–	Wild dogs (Abrahms et al., 2017)
Temporal	Breeding	Foraging	–	–	NZSL, as per this study (Augé et al., 2012)
	Day	Night	–	–	Wallabies (Southwell & Fletcher, 1988)
	Predispersal	Dispersal	Post-dispersal	–	Lynx (Palomares et al., 2000)
	Winter	Spring	Summer	Fall	Cetaceans (Edrén et al., 2010)
	Winter	Summer	–	–	Salamanders (Lunghi, Manenti, & Ficetola, 2015)

Appendix S1) or have opted to maintain multiple, separate predictions for each temporal or behavioural state, which can often be numerous (e.g. 12 months; Bombosch et al., 2014). In both cases, a clear definition of overall habitat suitability is consequently lost. The use of statistical thresholds in our framework is thus one solution to simplify multiple predictions, while at the same time allowing for an end-user to differentiate results across pixels. The habitat definition is then, however, threshold-dependent, as the cumulative or relative probability of presence values are replaced. State predictions should therefore be combined using an additional, second threshold for comparison.

Lastly, other SDM approaches have yet to emphasize the availability of each suitable habitat type within reachable ranges. We were able to incorporate this via the optional modules, which can add critical information for the management of mobile, multi-state species.

5.3 | Concluding remarks

In sum, the proposed framework is applicable to a wide range of circumstances in wildlife conservation management, as long as data availability allows for the analysis of species distributions and different states. A strong benefit of the approach is its easy and intuitive applicability, using existing software solutions that are widely accepted, open access and of high power. This also makes it simple to replace certain tools by new releases or developments of next generation tools while keeping the general logic of the whole proposed framework intact. With the increasing availability of fine-scale species occurrences, as well as environmental data, we see an increasing potential in future applications of our proposed framework. This will be of particular importance since the demand for management solutions in biodiversity conservation will be further increased in our anthropocentric future.

AUTHORS' CONTRIBUTIONS

V.F.F., J.O.E. and A.A.A. conceived the ideas and designed the methodology, with additional input by N.B. and H.E.; V.F.F. and A.A.A. collated and prepared the data; V.F.F. analyzed the data and drafted the manuscript, with critical revisions by J.O.E. and A.A.A.; V.F.F. drafted the supplementary tutorial; V.F.F., A.A.A., H.E., S.E., N.B. and J.O.E. all contributed to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

The environmental data used in this publication are available from Land Information New Zealand (http://data.linz.govt.nz) (datasets: *NZ Mainland Topo 50* – https://data.linz.govt.nz/layer/767-nz-topo50-maps/; *NZ Auckland Islands Topo 50* – https://data.linz.govt.nz/layer/861-nz-auckland-island-topo50-maps/) and Landcare New Zealand (https://lris.scinfo.org.nz) (datasets: *Land Cover Database v. 4.0* – https://lris.scinfo.org.nz/layer/423-lcdb-v41-land-cover-database-version-41-mainland-new-zealand/; *NZ DEM South Island 25 m* – https://lris.scinfo.org.nz/layer/127-nzdem-southisland-25-metre/). The NZSL location data can be accessed on the Dryad Digital

REFERENCES

Abrahms, B., Sawyer, S. C., Jordan, N. R., McNutt, J. W., Wilson, A. M., & Brashares, J. S. (2017). Does wildlife resource selection accurately inform corridor conservation? *Journal of Applied Ecology*, *54*, 412–422.

Augé, A. A., Chilvers, B. L., Mathieu, R., & Moore, A. B. (2012). On-land habitat preferences of female New Zealand sea lions at Sandy Bay, Auckland Islands. *Marine Mammal Science*, *28*, 620–637.

Augé, A. A., Chilvers, B. L., Moore, A. B., & Davis, L. S. (2011). Foraging behaviour indicates marginal marine habitat for New Zealand sea lions: Remnant versus recolonising populations. *Marine Ecology Progress Series*, *432*, 247–256.

Augé, A. A., Chilvers, B. L., Moore, A., Mathieu, R., & Robertson, B. C. (2009). Aggregation and dispersion of female New Zealand sea lions at the Sandy Bay breeding colony, Auckland Islands: How unusual is their spatial behaviour? *Behaviour*, *146*, 1287–1311.

Bombosch, A., Zitterbart, D. P., van Opzeeland, I., Frickenhaus, S., Burkhardt, E., Wisz, M. S., & Boebel, O. (2014). Predictive habitat modelling of humpback (*Megaptera novaeangliae*) and Antarctic minke (*Balaenoptera bonaerensis*) whales in the Southern Ocean as a planning tool for seismic surveys. *Deep Sea Research Part I: Oceanographic Research Papers*, *91*, 101–114.

Brambilla, M., & Saporetti, F. (2014). Modelling distribution of habitats required for different uses by the same species: Implications for conservation at the regional scale. *Biological Conservation*, *174*, 39–46.

Childerhouse, S., & Gales, N. (1998). Historical and modern distribution and abundance of the New Zealand sea lion *Phocarctos hookeri*. *New Zealand Journal of Zoology*, *25*, 1–16.

Connelly, J. W., Schroeder, M. A., Sands, A. R., & Braun, C. E. (2000). Guidelines to manage sage grouse populations and their habitats. *Wildlife Society Bulletin*, *28*, 967–985.

Cord, A. F., Meentemeyer, R. K., Leitão, P. J., & Václavík, T. (2013). Modelling species distributions with remote sensing data: Bridging disciplinary perspectives. *Journal of Biogeography*, *40*, 2226–2227.

Cord, A., & Rödder, D. (2011). Inclusion of habitat availability in species distribution models through multi-temporal remote-sensing data? *Ecological Applications*, *21*, 3285–3298.

Coulson, G. (1993). Use of heterogeneous habitat by the western grey kangaroo, *Macropus fuliginosus*. *Wildlife Research*, *20*, 137–149.

Department of Conservation. (2009). *New Zealand sea lion species management plan: 2009–2014*. Wellington, New Zealand: Department of Conservation. http://www.doc.govt.nz/documents/science-and-technical/sap251entire.pdf [accessed 9 January 2015].

Downs, J. A., Gates, R. J., & Murray, A. T. (2008). Estimating carrying capacity for sandhill cranes using habitat suitability and spatial optimization models. *Ecological Modelling*, *214*, 284–292.

Edelhoff, H., Signer, J., & Balkenhol, N. (2016). Path segmentation for beginners: An overview of current methods for detecting changes in animal movement patterns. *Movement Ecology*, *4*, 21.

Edrén, S., Wisz, M. S., Teilmann, J., Dietz, R., & Söderkvist, J. (2010). Modelling spatial patterns in harbour porpoise satellite telemetry data using maximum entropy. *Ecography*, *33*, 698–708.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*, 43–57.

Filz, K. J., Schmitt, T., & Engler, J. O. (2013). How fine is fine-scale? Questioning the use of fine-scale bioclimatic data in species distribution models used for forecasting abundance patterns in butterflies. *European Journal of Entomology*, *110*, 311–317.

Fourcade, Y., Engler, J. O., Besnard, A. G., Rödder, D., & Secondi, J. (2013). Confronting expert-based and modelled distributions for species with uncertain conservation status: A case study from the corncrake (*Crex crex*). *Biological Conservation*, *167*, 161–171.

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction*. New York, NY: Cambridge University Press.

Frans, V. F., Augé, A. A., Edelhoff, H., Erasmi, S., Balkenhol, N., & Engler, J. O. (2017). Data from: Quantifying apart what belongs together: A multi-state species distribution modeling framework for species using distinct habitats. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.14mt7

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Peterson, A. T., Loiselle, B. A., & NCEAS Predicting Species Distribution Working Group (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, *45*, 239–247.

Gschweng, M., Kalko, E. K. V., Berthold, P., Fiedler, W., & Fahr, J. (2012). Multi-temporal distribution modelling with satellite tracking data: Predicting responses of a long-distance migrant to changing environmental conditions. *Journal of Applied Ecology*, *49*, 803–813.

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

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.

He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., … Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, *1*, 4–18.

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*, 773–785.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2013) DISMO: species distribution modeling: R package version 0.9-3. http://CRAN.R-project.org/package=dismo [accessed 9 January 2015]

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*, 2nd ed. New York, NY: John Wiley and sons.

Jaberg, C., & Guisan, A. (2001). Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology*, *38*, 1169–1181.

Johnson, C. J., & Gillingham, M. P. (2005). An evaluation of mapped species distribution models used for conservation planning. *Environmental Conservation*, *32*, 117–128.

Kindlmann, P., & Burel, F. (2008). Connectivity measures: A review. *Landscape Ecology*, *23*, 879–890.

Knick, S. T., Hanser, S. E., & Preston, K. L. (2013). Modeling ecological minimum requirements for distribution of greater sage-grouse leks: Implications for population connectivity across their western range, U.S.A. *Ecology and Evolution*, *3*, 1539–1551.

Lalas, C., & Bradshaw, C. J. A. (2003). Expectations for population growth at new breeding locations for the vulnerable New Zealand sea lion (*Phocarctos hookeri*) using a simulation model. *Biological Conservation*, *114*, 67–78.

Law, B. S. (1993). Roosting and foraging ecology of the Queensland blossom bat (*Syconycteris australis*) in north-eastern New South Wales: Flexibility in response to seasonal variation. *Wildlife Research*, *20*, 419–431.

Law, B. S., & Dickman, C. R. (1998). The use of habitat mosaics by terrestrial vertebrate fauna: Implications for conservation and management. *Biodiversity & Conservation*, *7*, 323–333.

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*, 385–393.

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*, 778–789.

Lunghi, E., Manenti, R., & Ficetola, G. F. (2015). Seasonal variation in microhabitat of salamanders: Environmental variation or shift of habitat selection? *PeerJ*, *3*, e1122.

MacMillan, H., Moore, A. B., Augé, A. A., & Chilvers, B. L. (2016). GIS-based multi-criteria analysis of breeding habitats for recolonising species: New Zealand sea lions. *Ocean & Coastal Management*, *130*, 162–171.

McNally, N., Heinrich, S., & Childerhouse, S. (2001). Distribution and breeding of New Zealand sea lions *Phocarctos hookeri* on Campbell Island. *New Zealand Journal of Zoology*, *28*, 79–87.

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*, 1551.

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*, 1058–1069.

Mitchell, P. J., Monk, J., & Laurenson, L. (2017). Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution*, *8*, 12–21.

Naves, J., Wiegand, T., Revilla, E., & Delibes, M. (2003). Endangered species constrained by natural and human factors: The case of brown bears in northern Spain. *Conservation Biology*, *17*, 1276–1289.

Nielsen, S. E., Stenhouse, G. B., & Boyce, M. S. (2006). A habitat-based framework for grizzly bear conservation in Alberta. *Biological Conservation*, *130*, 217–229.

Palomares, F., Delibes, M., Ferreras, P., Fedriani, J. M., Calzada, J., & Revilla, E. (2000). Iberian Lynx in a fragmented landscape: Predispersal, dispersal, and postdispersal habitats. *Conservation Biology*, *14*, 809–818.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology & Evolution*, *23*, 87–94.

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, *31*, 161–175.

R Core Team. (2015) *R: A language and environment for statistical computing*, Vienna, Austria. Retrieved from http://www.R-project.org/. Accessed January 9, 2015.

Robertson, B. C., & Chilvers, B. L. (2011). The population decline of the New Zealand sea lion *Phocarctos hookeri*: A review of possible causes. *Mammal Review*, *41*, 253–275.

Roever, C. L., Beyer, H. L., Chase, M. J., & Aarde, R. J. (2014). The pitfalls of ignoring behaviour when quantifying habitat selection. *Diversity and Distributions*, *20*, 322–333.

Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. (2012). Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, *3*, 545–554.

Sousa-Silva, R., Alves, P., Honrado, J., & Lomba, A. (2014). Improving the assessment and reporting on rare and endangered species through species distribution models. *Global Ecology and Conservation*, *2*, 226–237.

Southwell, C. J., & Fletcher, M. S. (1988). Diurnal and nocturnal habitat utilisation by the whiptail wallaby, *Macropus parryi*. *Wildlife Research*, *15*, 595–603.

Taboada, A., von Wehrden, H., & Assmann, T. (2013). Integrating life stages into ecological niche models: A case study on tiger beetles. *PLoS ONE*, *8*, e70038.

Thuiller, W. (2003). BIOMOD-optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, *9*, 1353–1362.

van Gils, H., Westinga, E., Carafa, M., Antonucci, A., & Ciaschetti, G. (2014). Where the bears roam in Majella National Park, Italy. *Journal for Nature Conservation*, *22*, 23–34.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*, 763–773.

Zuckerberg, B., Fink, D., La Sorte, F. A., Hochachka, W. M., & Kelling, S. (2016). Novel seasonal land cover associations for eastern North American forest birds identified through dynamic species distribution modelling. *Diversity and Distributions*, *22*, 717–730.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.