Title	Simulating movement-related resource dynamics to improve species distribution models: a case study with oilbirds in northern South America
Author(s)	Holloway, Paul
Publication date	2018-11-16
Original citation	Holloway, P. (2018) 'Simulating Movement-Related Resource Dynamics to Improve Species Distribution Models: A Case Study with Oilbirds in Northern South America', The Professional Geographer, 70(4), pp. 528-540. doi: 10.1080/00330124.2018.1479972
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://dx.doi.org/10.1080/00330124.2018.1479972 Access to the full text of the published version may require a subscription.
Rights	**© 2018 American Association of Geographers. This is an Accepted Manuscript of an article published by Taylor & Francis in The Professional Geographer on 16 Nov 2018, available online: http://www.tandfonline.com/10.1080/00330124.2018.1479972**
Embargo information	Access to this article is restricted until 18 months after publication by request of the publisher.
Embargo lift date	2020-05-16
Item downloaded from	http://hdl.handle.net/10468/9172

TITLE PAGE

SIMULATING MOVEMENT-RELATED RESOURCE DYNAMICS TO IMPROVE SPECIES DISTRIBUTION MODELS: A CASE STUDY WITH OILBIRDS IN NORTHERN SOUTH AMERICA

Paul Holloway, Department of Geography, University College Cork, Cork, Ireland

ABSTRACT

A better understanding of the current and future distributions of organisms is a critical facet of biodiversity conservation, and species distribution models (SDMs) are an important framework for achieving this. Despite the potential of SDMs to address an array of biogeography questions, they are subject to a number of conceptual and methodological uncertainties, such as the role of animal movement processes in determining geographic ranges. Movement processes have only recently been incorporated in SDMs, predominantly conceptualized as broad-scale movement processes (e.g., dispersal), while finer-scale ambulatory movements of mobile animals (e.g., foraging) have been omitted. This research addresses this gap by developing a model that simulates the dynamic relationship between movement and biotic resources (e.g., food sources) for oilbirds (*Steatornis caripensis*) in Venezuela. This simulation represented the sustainability of an oilbird's neighborhood, based on the connectivity, accessibility, and viability of its biotic resources. These dynamic variables improved the accuracy and ecological realism of the SDM projection compared to other commonly applied SDM scenarios. Integration of a Lagrangian (individual-level) form of movement in SDM with step-selection functions to parameterize biased-correlated random walks provides a new empirical framework for applying geographic context to simulation.

Keywords: *movement; spatial simulation; species distribution modelling; step-selection function; Venezuela*

By understanding the factors that define a species' geographic range, more precise estimations of species' resilience, range dynamics and potential for extinction can be made. Species distribution models (SDMs) are a widely accepted framework for studying the geographic distribution of a range of taxa (Franklin 2009; Peterson et al. 2011), and have been used extensively for a variety of applications, including assessing the impact of climate change on species distributions (e.g., Albouy et al. 2014), testing biogeographical hypotheses (e.g., Gavin et al. 2014), and providing recommendations for biodiversity conservation (e.g., Platts et al. 2014). Despite the potential of SDMs to address an array of biogeography questions, they are subject to a number of conceptual and methodological uncertainties.

The consideration of the different environmental factors that could influence a species distribution, and subsequent selection of these variables with which to train the model is an enduring issue (Guisan and Zimmermann 2000; van de Pol et al. 2016). To illustrate the individual and joint effects of the three factors deemed most important in determining species distributions, Soberón and Peterson (2005) developed the heuristic 'BAM' framework. In this framework, biotic factors (**B**) represent interactions with other species (i.e., competition, herbivory), abiotic factors (**A**) represent the physiological tolerances of a species (i.e., temperature, precipitation) and movement (**M**) refers to the area that has been or will be accessible to a species within a certain timeframe (i.e., dispersal, connectivity). While the importance of all three factors is well recognized when projecting the actual distribution, rarely are all three considered together, with the majority of studies calibrating SDMs using only abiotic variables (Saupe et al. 2012).

Biotic interactions can vary over time and among populations, and additionally, these interactions are characteristically superseded by abiotic factors. As such, biotic interactions are typically studied at a local scale; however, research has shown that biotic interactions have affected the distribution of species at regional, continental and even global extents (Wisz et al. 2013). Similarly, movement has been under-utilized in SDM research. When movement has been incorporated in SDM, its almost exclusive conceptualization has been to couple the statistical model (e.g., a model of projected suitable abiotic habitat based on presence-absence observations) with broad-scale movement processes such as dispersal in response to climate change (Franklin 2010; Miller and Holloway 2015). Movement has also been incorporated as a measure of accessibility (e.g., the extent of the study area that is currently accessible to the species through movement) with which to calibrate the model (Barve et al. 2011; Qiao, Soberón, and Peterson 2015). However, the causes of animal movement are vast (Nathan et al. 2008), with movements ranging from broad-scale dispersal into unknown landscapes to finer-scale daily movements such as foraging and homing. To-date, these finer-scale movement behaviors have yet to be incorporated within SDM, despite the important role they play in the maintenance of a mobile species distribution (Holloway and Miller 2017).

Moreover, movement and biotic interactions do not act independently of each other. In a study investigating the range constraints of Swiss stone pine, Neushulz et al. (2017)

found that the upper and lower elevational limits were determined by the intensity of nutcracker herbivory and not abiotic factors as previously thought. While biotic interactions are typically represented as processes such as herbivory and competition, Wisz et al. (2013) argue that these interactions should also include the dependence of animals on food plants, as this has been found to be an important biotic determinant of species distributions (Kissling, Field, and Böhning-Gaese 2008). The distribution of food plants (herein biotic resources) has long been noted as a driver of animal movement (Levin 1992), and such is the relationship between the two that the ‘underlying environment’ is considered one of the four driving influences of movement in the movement ecology paradigm (Nathan et al. 2008).

The current methodological standard in SDM is to couple the response data (e.g., presence-absence of a species) with a set of geospatial variables at the same location (e.g., temperature or biotic resources); however, mobile species utilize several patches of biotic resources throughout their home range. Figure 1 illustrates some of the landscape configuration factors that affect the distribution of an individual or population of mobile animals. Individually, these isolated patches of biotic resources do not sufficiently fulfill the requirements of an individual animal, but when considered at a neighborhood scale, they provide suitable sustenance (Figure 1A). Issues arise when the area through which an animal forages is substantially larger than the resolution of the environmental data. In such cases, the use of a local correlative model (coupling the response data with the environmental data at that location -- Figure 1B-C) assumes that an individual can survive in the area based on only the biotic resources found at that location. Such models will over-predict (Figure 1B) or under-predict (Figure 1C) the geographic distribution of a species, leading to incorrect estimations of resilience and extinction (Sinclair, White, and Newell 2010). Model calibration in SDM therefore needs to account for the spatial variation of biotic resources within the distribution of mobile animals. The use of focal correlative models (coupling the response data with the environmental data from a predefined neighborhood of *n* observations -- Figure 1D-E) accounts for the spatial configuration of environmental variables. Studies have begun to utilize focal statistics to describe land cover indicators of habitat suitability in SDM, reporting equal or higher predictive power than local models (Czúcz et al. 2011; Ashcroft and Major 2013; Betts et al. 2014). However, to-date this concept has not been extended to biotic resources, despite the potential for a focal approach to incorporate the relationship between **B** and **M** within SDM predictions.

However, it is not sufficient to consider the relationship between **B** and **M** as static. Biotic resources deplete when organisms move through an area (e.g., herbivory) and subsequently replenish while they are absent (e.g., regeneration), and this dynamic relationship is well documented in the natural environment (e.g., the wolf-elk-aspen interactions of Yellowstone National Park -- Turner, Gardner, and O’Neill 2001). Subsequently, a species distribution could be better explained by an ecological factor that integrates the dynamic relationship between ambulatory movements and biotic factors. The three BAM factors do not operate in isolation, yet a static representation of biotic factors and movement may not reliably account for the complex relationship that exists between them.

For complex geographic systems, such as the dynamic relationship among the BAM factors, modelling is required to link reality with theory in order to explore and understand the underlying processes (Manson 2001; 2007). This epistemological approach to complexity research in geography is widely adopted (O'Sullivan 2004; Manson 2007; Batty 2009), but models (even complex ones) are abstractions of reality. Currently, the BAM framework lacks the detail required to model such complex dynamic relationships, and further research is needed to identify new methods that can do this.

Spatial simulation models are increasingly being used to understand animal movement and plant dispersal due to their potential to incorporate the inherent relationship between movement and the environment (Tang and Bennet 2010), and are a burgeoning GIScience research area (Dodge 2016). Three main conceptualizations of spatial simulation have emerged within the GIScience discipline; spread models (diffusive movement through a neighborhood), random walk (RW) models (simulation of individual movement paths), and agent-based models (ABMs; simulation of movement as the emergent characteristics of the subcomponent relationships) (O'Sullivan and Perry 2013; Holloway and Miller 2018).

The RW framework is a well-established stochastic representation of movement trajectories (Miller 2015); however, within complex systems, RWs are too simplistic to represent the actual trajectories of moving objects, meaning RWs have been advanced to include directional persistence (correlation) and a consistent bias towards a given target (bias) (Codling, Plank, and Benhamou 2008). Simulation models have been successfully implemented to explore the relationship between movement and biotic resources in non-SDM contexts (e.g., South 1999; Van Moorter et al. 2009) and the reliability of such models to accurately simulate movement in response to the underlying dynamic biotic environment has led to biased correlated random walks (BCRWs) being proposed as the most flexible and discrete-step simulation model (Benhamou 2014).

A criticism of BCRWs is the subjective nature of parameter estimation, and subsequently the development of simulations based on quantitative analysis of real movement observations have been proposed (Dodge 2016). Recently, Ahearn et al. (2017) used the Chi-square distribution derived from the relationship between a tiger's Global Positioning System (GPS) location and the underlying environment to determine the next probable move made by an individual, identifying the importance of incorporating geographic context into BCRWs. The use of movement observations to increase the robustness in simulation models is promising, but more methods of incorporating such information need to be explored.

Recently Duchesne, Fortin, and Rivest (2015) showed that BCRW and step-selection function (SSF) provided similar insights into the environmental factors influencing bison movement in Canada. SSF is a spatial statistic that uses telemetry data to analyze movement patterns and resource use based on the underlying environment (Thurfjell, Ciuti, and Boyce 2014). SSF considers the landscape variables associated with a discrete movement step between two successive locations, and statistically compares these to a set of alternatively generated steps that the animal theoretically could have taken (Fortin et al. 2005). SSF has been used to evaluate the influence of the landscape features and biotic factors on a variety of species (Holloway and Miller 2014), with spatial memory recently incorporated into the statistical model (Oliveira-Santos et al. 2016). SSF models the choices actually presented to the animal as it traverses through the landscape, and would be a powerful input in spatial simulation.

The simulations developed herein improve upon diffusion and RW models by adopting a Lagrangian approach that incorporates SSF parameter coefficients to inform geographic context from empirical telemetry data. The incorporation of discrete movement steps to simulate movement at an individual level within an ABM will inform animal-movement decisions. When integrated within the SDM framework, this model will provide researchers with an improved understanding of the relationship between biotic interactions and movement factors (something that has been absent in previous studies) and an enhanced awareness of which factors contribute to a species distribution.

METHODOLOGY

Study Species

The oilbird (*Steatornis caripensis* Humboldt 1817) is a nocturnal avian frugivore that inhabits northern South America, roosting in caves or crevices during the day, and foraging for fruit at night. In a recent study tracking oilbirds in Venezuela, Holland et al.

(2009) found that the birds often roosted near their foraging sites rather than returning nightly to a cave. These findings have had important ecological consequences for the oilbird ecosystem, as the oilbirds are now seen as a primary disperser of many seeds throughout the Amazon. The oilbird has an extensive home range (mean foraging distance of 44km -- Holland et al. 2009), and a varied diet of fruit (from both natural forests and cropland plantations -- Bosque, Ramíerez, and Rodríguez 1995). Telemetry data for 40 individual oilbirds in Venezuela were obtained from Holland et al. (2009) via Movebank (Holland et al. 2012), which resulted in a total of 5,353 locations. Data filtering using the 'best daily location' method (Abecasis, Afonso, and Erzini 2014) was undertaken to account for a number of new conceptual issues related to telemetry data structure and use in SDM, including spatial and temporal autocorrelation and measurement error (see Hebblewhite and Haydon 2010 for a discussion). This resulted in a total of 302 points from 25 oilbirds for use as response data, with each oilbird contributing between 1-6 percent of the total observations.

Species Distribution Modelling

Figure 2 is a conceptual diagram that highlights the main steps involved within this SDM framework. Five different BAM scenarios were modelled (expanded upon below). Each scenario incorporates different methodologies to derive inputs and subsequently distribution projections (Figure 2). Maximum entropy (MaxEnt) was chosen as the statistical method to model the species-environment relationships for all five scenarios. MaxEnt selects suitable environmental conditions by measuring how well they delineate recorded presence and pseudo-absence observations, and choosing the split that records the purest distribution of response variables. Other presence-only methods, such as genetic algorithms for rule prediction (GARP) and ecological niche factor analysis (ENFA) were explored, but MaxEnt provided the most accurate preliminary results. Furthermore, MaxEnt has been used extensively in SDM applications, and has been found to generate accurate predictions of species distributions using presence-only data (Elith and Graham 2009). The maximum number of iterations was set to 5,000 to allow model convergence; the number of pseudo-absences was set at 10,000 (following Barbet-Massin et al. 2012), and the model incorporated only linear and quadratic features to avoid over-fitting the model. SDMs were generated at a 10km resolution (the same scale many bird atlases are collected at, e.g., Balmer et al. 2013). As Holland et al. (2009) deployed GPS tags on individuals from the *Cueva del Guácharo* in Northeastern Venezuela, there was unequal survey coverage of oilbirds across the entire country, which could violate the assumption of independence in the response data and increase the likelihood of type I errors (Fourcarde et al. 2014). Following Elith, Kearney, and Phillips (2010), a bias grid was created with the method outlined in Supplementary Information 1.

Scenario A

Abiotic factors represent the physiological tolerances of species (i.e., temperature, precipitation) meaning climate variables best represent these factors. WorldClim has interpolations of observed data from ~1950-2000, with 19 bioclimatic variables derived from the monthly temperature and rainfall values (see Hijmans et al. 2005 for a list). The 19 bioclimatic variables were the abiotic factors used in the statistical model.

Scenario B

Biotic factors represent interactions with other species (i.e., competition, herbivory). The oilbird is a frugivore, suggesting that the most important biotic interactions are those with the fruit bearing trees that make up its diet. Bosque, Ramíerez, and Rodríguez (1995) identified five widespread (naturally occurring and/or cultivated) trees whose fruit made up a substantial part of the oilbird's diet: *Euterpe precatoria*, *Nectandra membranacea*, *Ocotea floribunda*, *Persea caerulea* and *Prestoea acuminata.* The distribution of these five species was predicted using observed locations from the Global Biodiversity Information Facility (GBIF 2015a; 2015b; 2015c; 2015d; 2015e), the 19 bioclimatic variables (Hijmans et al. 2005), and projected using MaxEnt with default settings (with the exception of 5,000 minimum iterations to allow model convergence -- see Phillips, Anderson, and Schapire 2006). Maximum sensitivity plus specificity was used as the threshold for presence (the probability value above which a case is predicted to be positive), and distribution maps were generated at a 1km resolution. Categorical variables representing presence-absence of the five species within the 10km grid were used as the biotic factors in the statistical model.

Scenario BA

This scenario projects the invadable distribution in the BAM framework. The 19 bioclimatic variables and the five presence-absence predictions of the food species were used as the input variables in the statistical model.

Scenario Classic BAM (cBAM)

Movement factors refer to the area that has been or will be accessible to a species within a certain timeframe. Accessibility to biotic resources was identified by using a distance buffer. The mean foraging distance of oilbirds is 44.4km (Holland et al. 2009). Therefore, presence-absence of the five fruit-bearing tree species was calculated within a focal neighborhood of 50km for each 10km grid. This was incorporated into the statistical model as a binary presence-absence value with the resolution held consistent at 10km.

Scenario Dynamic BAM (dBAM)

A variable which summarizes the dynamic relationship between movement and biotic resources within the 50km focal area was created using spatial simulation. Movement was simulated in NetLogo 5.05 (Wilensky 1999) as a BCRW parameterized on an SSF model with factors including land cover, presence of resources, and memory (see Supplementary Information 2). Identifying the survival of individual agents in spatial simulation by the amount of energy consumed is common (e.g., Ahearn et al. 2001; Silby et al. 2013), and captures the dynamic relationship between movement and biotic resources. Individuals move in response to the spatial configuration of biotic resources parameterized by the SSF model, consuming resources and energy as they move. Biotic resources subsequently deplete if consumed. If an individual does not obtain enough energy, it dies. The agent-based model was run three times for a one-year period with 1,000 oilbirds (the 90th percentile of oilbird population size in Venezuela -- Herrera 2003). The mean survival rate (see Supplementary Information 2 for details) of the three simulations was the input along with the abiotic factors into the statistical model. A high value represents a landscape with a large amount of accessible and sustainable biotic resources, while a low value indicates insufficient or inaccessible resources. In following with the current standard for modelling, a thorough account of the design of the model is outlined using the Overview, Design Concepts and Details (ODD) protocol (Grimm et al. 2006; 2010) in Supplementary Information 2, and the model code for the SSF parameterized BCRW is presented in Supplementary Information 3.

Model Evaluation

Over thirty unique oilbird presence observations in Venezuela (GBIF 2016) were used as the independent test dataset. Multiple evaluation metrics should be used when reporting the evaluation of SDM projections due to the potential differences reported among metrics (Liu et al. 2005; Franklin 2009). The lowest possible threshold (LPT) is the value that would result in zero omission errors. The minimum predicted area (MPA) is the threshold value that minimizes omission errors, but at the same time minimizes the area of the map predicted to be suitable (Franklin 2009). The MPA represents the threshold value that encompasses a predefined proportion of observed species occurrences (in this case 90 percent). The Boyce Index is based on a moving window analysis across the range of predicted values, and uses Spearman rank correlation coefficient to measure the monotonic increase in the predicted-to-expected frequency ratio with increasing habitat suitability (Boyce et al. 2002). The Boyce Index was implemented using the *ecospat* package (Broenniman et al. 2014) in R 3.3.0 (R Development Core Team 2008), and ranges from -1 to 1, with 0 indicating predictions indifferent from a random model.

RESULTS

Step-Selection Function

The final SSF model recognized several significant variables that identified the landscape, resource and memory factors associated with oilbird movement (Table 1). Positive coefficients represent the probability of selecting steps into that habitat compared to reference habitats (for binary variables this is the opposite of what is modelled and for continuous variables this is compared to the current value), while negative coefficients represent the probability of selecting the reference habitat. Oilbirds preferentially chose movement paths into landscapes which they have previously visited and knew to have resources (Memory*Resources), and also favorably selected steps into unknown habitat rather than returning to a location which they had visited, but knew not to have resources. Oilbirds also favored movement steps into evergreen broadleaf forests and croplands over movement steps into other land cover classes, while selecting steps away from rivers. The results of the SSF model (Table 1) parameterized a BCRW. The BCRW trajectory consisted of several smaller movement steps, followed by a few longer steps, with a high return rate to previous locations where the interaction between memory and resources positively influenced the step. Furthermore, the difference in the spatial configuration of the underlying landscape also influenced the movement path, with oilbirds preferentially choosing evergreen forests (and to a lesser extent, croplands) over other land cover types.

Species Distribution Modelling

The five BAM scenarios differed substantially in terms of the area predicted present (Figure 3A-E) and their accuracy when evaluated against an independent test dataset (Table 2). The general trend observed was that an increase in the number of BAM factors incorporated in the SDM reduced the area predicted as present, with Scenarios B (Figure 3B) and dBAM (Figure 3E) predicting 394,100km^2 and 211,100km^2 respectively. The dBAM scenario removes the distribution projected within the center of the country by scenarios A, B, BA, and cBAM, and projects a more fragmented distribution around the boundary of the country compared to the continuous distribution projected by scenarios A, BA, and cBAM. The B scenario projected the largest distribution of 394,100km^2 with a relatively continuous surface of oilbird presence around the whole country (Figure 3B), and this is a likely explanation for the higher accuracy metrics of LPT and MPA (Table 2) as there is no penalization for over-prediction.

Precipitation of the coldest quarter was the environmental variable that contributed the largest percentage to all scenarios that included abiotic factors with the exception of the dBAM scenario, where it was the second highest contributor in the MaxEnt selection process. For dBAM, the spatial simulation representing the dynamic relationship between **B** and **M** accounted for the highest relative contribution (36.7 percent) to the delineation between oilbird presence and pseudo-absences (almost double precipitation of the coldest quarter). The use of an ABM to simulate this complex relationship identified a number of locations of biotic resources that are actually unsustainable compared to modelling presence-absence at a focal neighborhood (Figure 3F-H). The difference between the presence of the five tree species at a 10km resolution (used in Scenarios B and BA -- Figure 3F) compared to when presence-absence is viewed at a 50km neighborhood scale at a 10km resolution (Scenario cBAM -- Figure 3G) is apparent. Representing presence-absence at a 50km focal area overestimates the distribution across Venezuela, particularly in the south of the country, and it is this methodological artefact that accounts for the higher Boyce Index score recorded for BA compared to cBAM (Table 2). When the results of the ABM are explored (Figure 3H), many locations of biotic resources are unsustainable for 1,000 oilbirds, with large areas of the projected tree species distributions from Figures 3F-G disappearing. This better represents the dynamic relationship between biotic resources and movement, and does well to identify core areas of oilbird suitability, sustainability and accessibility.

DISCUSSION

The aim of this research was to investigate how spatial simulation could be used to represent the relationship between movement and biotic factors. The inclusion of these simulated data, which account for the dynamic relationship between movement and biotic resources, improved the accuracy (when measured as the Boyce Index) of the SDM (Table 2), as well as better explaining the factors that affect the distribution of oilbirds in Venezuela. The Boyce Index is considered the most robust presence-only accuracy metric due to its ability to measure the monotonic increase in the predicted-to-expected frequency ratio with increasing habitat suitability, with a higher value representative of a more valid model (Hirzel et al. 2006). The Boyce Index was highest for the dBAM model, suggesting that incorporating a variable that measured the relationship between **M** and **B** was advantageous to the SDMs projection capacity. This argument is furthered by the fact that this layer had the highest relative percentage contribution of all the BAM variables used in delineating between the environmental and geographic distribution of oilbird presence and pseudo-absence in the final statistical model (37.6 percent). The use of a dynamic environmental layer of biotic resources and movement removed the assumption of homogeneity across the distribution of food species, and delineated between accessible and sustainable areas of biotic resources from isolated and unsustainable areas.

As the distribution of the five plant species was based on the 19 bioclimatic variables, it could be argued that the B scenario was actually representative of the BA scenario. However, by its definition in the BAM framework, **A** represents the physiological tolerances of the studied species, meaning that the abiotic conditions that define the food species may not be the same as those of the oilbird. Nevertheless, this is still an important aspect of the BAM framework that should be acknowledged. None of the three factors are completely independent of each other. While this study focused on the dynamic relationship between **M** and **B**, interactions among all three BAM factors occur, and it may be theoretically difficult to parse out the influence of each individual factor. For example, abiotic factors have been found to influence the movement of animals, with Delattre, Vernon, and Burel (2013) finding that the dispersal distances of meadow brown butterflies were greater across fragmented landscapes at cooler temperatures, but greater across continuous landscapes at warmer temperatures. Therefore, future BAM research should not only focus on the individual and combined effects of the three factors, but also identify a clear methodology with which to successfully model the dynamic relationships among all three factors. The results from this study suggest that spatial simulation could be an effective framework to achieve this.

Oilbirds are particularly appropriate for studying issues related to incorporating geographic context into a Lagrangian simulation of movement. The specialized habitat and feeding preferences of the oilbird, coupled with the high metabolic rate incurred from large movements between nesting and feeding locations made it conceptually easier to model than more generalist species. Moreover, oilbirds feed on fruit from a variety of trees including those naturally occurring in the Amazon, as well as commercial plantations of oil palm. Multiple food sources found across different land cover types meant that the SSF model and subsequent BCRW did not favor only one land cover type, and more realistic movement choices were presented to the oilbirds. Despite this, such a modelling framework should scale from specialist to generalist species, although each model should be underpinned by domain experts.

While the SSF and ABM developed have increased our understanding of the oilbird’s spatial ecology, there is still a lot we do not yet understand about this elusive animal. Edge habitat is an important driver of animal movement with species using the border between two land cover types to maximize foraging opportunities, and has been used in previous SSF studies (Holloway and Miller 2014). While edge habitat was not used in the calibration of this SSF model due to differences in the spatial resolution, it could have a large effect on the movement patterns of oilbirds. Given the preferential disposition for oilbirds to move into both evergreen forest and cropland, simulations appeared to favor edge habitat (results not shown); however, whether this preference is actually a driver of oilbird movement or a methodological artefact warrants further investigation. Moreover, cropland and evergreen forest edge habitat represents a new biotic interaction between humans and oilbirds. In reality, oilbirds will most likely interact with multiple other species confounding the difficulty to parse out the individual influence of one biotic factor on the projected distribution. This further reiterates the importance of spatial simulation to model the dynamic and complex relationships that exist among all BAM factors.

The incorporation of SSF within a spatial simulation model is novel, and the movement paths of the individual oilbirds (Supplementary Information 3) were distinctive of home range movements (e.g., Lévy walk -- Plank and Edwards 2011). The benefit of using SSF model coefficients (Table 1) to parameterize a BCRW over subjective implementations is that it includes the results of conditional logistic regression. With simulated movement underpinned by a statistical model generated from empirical observations, more realistic movement behaviors are computed compared to a model specified solely on subjective assumptions. This means that any subsequent results better illustrate the influences of the environment on actual movements (Duchesne, Fortin, and Rivest 2015). The SSF framework would also allow further environmental variables to be included. For example, a variable representing locations of caves across the Amazon could be incorporated in the SSF and may improve the predictive ability of these models; however, to-date no dataset exists that accurately describes such a phenomenon.

A recent criticism of SSF has been the lack of methodology with which to incorporate a measure of memory (Fagan et al. 2013; Duchesne, Fortin, and Rivest 2015). To ignore the effects of memory when analyzing movement decisions made by animals can lead to biased estimates of resource-selection coefficients (Oliveira-Santos et al. 2016). The results in this study agree with Oliveira-Santos et al. (2016); memory is an important factor to consider within an SSF (Table 1). Furthermore, the importance of considering memory of biotic resources rather than simply spatial memory irrespective of the landscape should now be considered. Memory does not operate independently of information or events, and when memory was the only factor considered, oilbirds avoided moving into habitats that they had previously visited, but when modelled as an interaction with awareness of biotic resources, oilbirds conversely selected locations they had previously visited and knew had resources. Memory can be modelled in a number of ways, and the assumption of memory in this study relied on the theory that time spent in an area contributed to the spatial memory of that individual. While this description of memory is a simplified approximation, it does illustrate that this is an important process which should be accounted for when simulating movement, and as such should help foster future SSF studies which look to incorporate spatial memory.

CONCLUSION

The three ‘BAM’ factors used in projecting species distributions do not operate in isolation, meaning a traditional static representation does not reliably account for the complex dynamic relationships that exist among them. Simulations that adopted a Lagrangian approach (individual-level) to movement, while incorporating SSF parameter coefficients to integrate geographic context from empirical telemetry data, informed oilbird-movement decisions and improved the accuracy and ecological realism of oilbird SDM projections compared to other widely used BAM scenarios. The use of SSF as an input into the spatial simulation was novel, and an interaction term between memory and biotic resources aided in the interpretation of how oilbirds were selecting movement steps. The application of empirical movement data in informing spatial simulation is a promising avenue of GIScience research, and the results of this study have highlighted how it improves our understanding of both the geographic distribution and movement ecology of the biogeographically important oilbird.

REFERENCES

Abecasis, D., P. Afonso, and K. Erzini. 2014. Combining multispecies home range and distribution models aids assessment of MPA effectiveness. *Mar Ecol Prog Ser*, 513: 155-169.

Ahearn, S. C., J.L.D. Smith., A.R. Joshi., and J. Ding. 2001. TIGMOD: an individual-based spatially explicit model for simulating tiger/human interaction in multiple use forests. *Ecological Modelling, 140*(1): 81-97.

Ahearn, S.C., S. Dodge, A. Simcharoen, G. Xavier, and J.L.D. Smith. 2017. A context-sensitive correlated random walk: a new simulation model for movement. *International Journal of Geographical Information Science, 31*(5): 867-883.

Albouy, C., L. Velez., M. Coll., F. Colloca., F. Loc'h., D. Mouillot., and D. Gravel. 2014. From projected species distribution to food-web structure under climate change. *Global change biology*, *20*(3): 730-741.

Ashcroft, M. B., and R.E. Major. 2013. Importance of matrix permeability and quantity of core habitat for persistence of a threatened saltmarsh bird. *Austral Ecology*, 38(3): 326-337.

Balmer, D.E., S. Gillings., B.J. Caffrey., R.L. Swann., I.S. Downie., and R.J. Fuller. 2013. *Bird Atlas 2007–11: the Breeding and Wintering Birds of Britain and Ireland*. Thetford, UK: BTO.

Barbet-Massin, M., F. Jiguet., C.H. Albert., and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many?. *Methods in Ecology and Evolution*, *3*(2): 327-338.

Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S.P. Mayer, A.T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling,* 222: 1810-1819.

Batty, M. 2009. Urban Modeling. *International Encyclopedia of Human Geography.* Oxford, UK: Elsevier.

Benhamou, S. 2014. Of scales and stationarity in animal movements. *Ecology letters*, *17*(3): 261-272.

Betts, M. G., L. Fahrig., A.S. Hadley., K.E. Halstead., J. Bowman., W.D. Robinson., J.A. Wiens., and D.B. Lindenmayer. 2014. A species-centered approach for uncovering generalities in organism responses to habitat loss and fragmentation. *Ecography*, 37(6): 517-527.

Bosque, C., R. Ramírez., and D. Rodríguez. 1995. The diet of the Oilbird in Venezuela. *Ornitologia Neotropical*, *6*: 67-80.

Boyce, M.S., P.R. Vernier., S.E. Nielsen., and F.K. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological modelling*, *157*(2): 281-300.

Broenniman, O., B. Petitpierre., C. Randin., R. Engler., F. Breiner., M. D’Amen., L. Pellissier., J. Pottier., D. Pio., R. Garcia Mateo, et al. 2014. ecospat: Spatial ecology miscellaneous methods. R package version 1.0.

Codling, E. A., M.J. Plank., and S. Benhamou. 2008. Random walk models in biology. *Journal of the Royal Society Interface*, *5*(25): 813-834.

Czúcz, B., A. Csecserits., Z. Botta-Dukát., G. Kröel-Dulay., R. Szabó., F. Horváth., and Z. Molnár. 2011. An indicator framework for the climatic adaptive capacity of natural ecosystems. *Journal of Vegetation Science*, *22*(4): 711-725.

Delattre, T., P. Vernon., and F. Burel. 2013. An agri-environmental scheme enhances butterfly dispersal in European agricultural landscapes. *Agriculture, ecosystems & environment*, 166: 102-109.

Dodge, S. 2016. From observation to prediction: the trajectory of movement research in GIScience. In: Onsrud, H. and Kuhn, W. (eds) *Advancing geographic information science: the past and next twenty years*. Needham, MA: GSDI Association Press. 123-136.

Duchesne, T., D. Fortin., and L.P. Rivest. 2015. Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. *PloS one*, *10*(4): e0122947.

Elith, J., and C.H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32(1): 66-77.

Elith, J., M. Kearney., and S. Phillips. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1: 330–342

Fagan, W. F., M.A. Lewis., M. Auger-Méthé., T. Avgar., S. Benhamou., G. Breed., L. LaDage., U.E. Schlagel., W. Tang., Y.P. Papastamatiou., et al. 2013. Spatial memory and animal movement. *Ecology letters*, *16*(10): 1316-1329.

Fortin, D., H.L. Beyer., M.S. Boyce., D.W. Smith., T. Duchesne., and J.S. Mao. 2005. Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. *Ecology*, *86*(5): 1320-1330.

Fourcade, Y., J.O. Engler., D. Rödder., and J. Secondi. 2014 Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias, *PLoS One,* 9(5): e97122.

Franklin, J. 2009. *Mapping species distributions: spatial inference and prediction*. Cambridge, UK: Cambridge University Press.

Franklin, J. 2010. Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions,* 16 (3): 321-330.

Gavin, D.G., M.C. Fitzpatrick., P.F. Gugger., K.D. Heath., F. Rodríguez-Sánchez., S.Z. Dobrowski., A. Hampe., F.S. Hu., M.B. Ashcroft., P.J. Bartlein., et al. 2014. Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, *204*(1): 37-54.

GBIF 2015a GBIF.org (8th October 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.xlvctu

GBIF 2015b GBIF.org (8th October 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.ksm9he

GBIF 2015c GBIF.org (8th October 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.sphwvu

GBIF 2015d GBIF.org (8th October 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.s95nga

GBIF 2015e GBIF.org (8th October 2015) GBIF Occurrence Download http://doi.org/10.15468/dl.mqsyq4

GBIF 2016 GBIF.org (9th February 2016) GBIF Occurrence Download http://doi.org/10.15468/dl.ahjffh

Grimm, V., U. Berger., F. Bastiansen., S. Eliassen., V. Ginot., J. Giske., J. Goss-Custard., T. Grand., S.K. Heinz., G. Huse., et al. 2006. A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, *198*(1): 115-126.

Grimm, V., U. Berger., D.L. DeAngelis., J.G. Polhill., J. Giske., and S.F. Railsback. 2010. The ODD protocol: a review and first update. *Ecological modelling, 221*(23): 2760-2768.

Guisan, A., and N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological modelling, 135*(2): 147-186.

Hebblewhite, M., and D.T. Haydon. 2010. Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 365: 2303-2312.

Herrera, F. F. 2003. Distribución actualizada de las colonias de guacharos (Steatornis caripensis) en Venezuela [Updated distribution of the oilbird (Steatornis caripensis) in Venezuela]. *Bol Soc Venezol Espeleol,* 37: 31-40.

Hijmans, R.J., S.E. Cameron., J.L. Parra., P.G. Jones., and A. Jarivs. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* 25: 1965-1978.

Hirzel, A.H., G. Le Lay., V. Helfer., C. Randin., and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. E*cological Modelling*, *199*(2): 142-152.

Holland, R. A., M. Wikelski., F. Kümmeth., and C. Bosque. 2009. The secret life of oilbirds: new insights into the movement ecology of a unique avian frugivore. *PLoS One*, *4*(12): e8264.

Holland, R.A., M. Wikelski., F. Kümmeth., and C. Bosque. 2012. The secret life of oilbirds: new insights into the movement ecology of a unique avian frugivore. Movebank Data Repository. doi: 10.5441/001/1.35fs26kq

Holloway, P., and J.A. Miller. 2014. Uncertainty Analysis of Step-Selection Functions: The Effect of Model Parameters on Inferences about the Relationship between Animal Movement and the Environment. *Lecture Notes in Computer Science,* 8728: 48-63.

Holloway, P., and J.A. Miller. 2017. A quantitative synthesis of the movement concepts used in species distribution modellimg. *Ecological Modelling,* 356: 91-103.

Holloway, P. and J.A. Miller. 2018. Analysis and Modeling of Movement. *Comprehensive Geographic Information Systems. Vol. Principles and Technical Designs of GIS,* ed B. Huang, T. Cova and M. Tsou, 162-180. Oxford, UK: Elsevier.

Kissling, W.D., R. Field., and K. Böhning-Gaese. 2008. Spatial patterns of woody plant and bird diversity: functional relationships or environmental effects?. *Global Ecology and Biogeography*, *17*(3): 327-339.

Levin, S.A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, *73*(6): 1943-1967.

Liu, C., P.M. Berry., T.P. Dawson., and R.G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography, 28* (3): 385-393.

Manson, S. M. 2001. Simplifying complexity: a review of complexity theory. *Geoforum* 32(3): 405-414.

Manson, S. M. 2007. Challenges in evaluating models of geographic complexity. *Environment and Planning B: Planning and Design* 34: 245-260.

Miller, J.A. 2015. Towards a better understanding of dynamic interaction metrics for wildlife: a null model approach. *Transactions in GIS, 19*(3): 342-361.

Miller, J. A., and P. Holloway. 2015. Incorporating movement in species distribution models. *Progress in Physical Geography*, *39*(6): 837-849.

Nathan, R., W.M. Getz., E. Revilla., M. Holyoak., R. Kadmon., D. Saltz., and P.E. Smouse. 2008. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences,105*(49): 19052-19059.

Neuschulz, E.L., D. Merges, K. Bollmann, F. Gugerli, and K. Böhning-Gaese. 2017. Biotic interactions and seed deposition rather than abiotic factors determine recruitment at elevational range limits of an alpine tree. *Journal of Ecology.*

Oliveira-Santos, L.G.R., J.D. Forester., U. Piovezan., W.M. Tomas., and F.A. Fernandez. 2016. Incorporating animal spatial memory in step selection functions. *Journal of Animal Ecology* 85(2): 516-524.

O'Sullivan, D. 2004. Complexity science and human geography. *Transactions of the Institute of British Geographers* 29(3): 282-295.

O'Sullivan, D., and G.L.W. Perry. 2013. *Spatial Simulation: Exploring Pattern and Process*. Oxford, UK: Wiley-Blackwell.

Peterson, A. T., J. Soberón., R.G. Pearson., R.P. Anderson., E. Martínez-Meyer., M. Nakamura., and M.B. Araújo. 2011. *Ecological niches and geographic distributions (MPB-49)*. New Jersey, USA: Princeton University Press.

Phillips, S.J., R.P. Anderson, and R.C. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190: 231-259.

Plank, J.A., and A.M. Edwards. 2011. Assessing Lévy walks as models of animal foraging. *Journal of the Royal Society Interface*, *8*(62): 1233-1247.

Platts, P.J., R.A. Garcia., C. Hof., W. Foden., L.A. Hansen., C. Rahbek., and N.D. Burgess. 2014. Conservation implications of omitting narrow-ranging taxa from species distribution models, now and in the future. *Diversity and distributions*, *20*(11): 1307-1320.

Qiao, H., J. Soberón., and A.T. Peterson. 2015. No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution*, *6*(10): 1126-1136.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org.

Saupe, E. E., V. Barve., C.E. Myers., J. Soberón., N. Barve., C.M. Hensz., A.T. Peterson., H.L. Owens., and A. Lira-Noriega. 2012. Variation in niche and distribution model performance: the need for a priori assessment of key causal factors. *Ecological Modelling*, 237*:* 11-22.

Sibly, R. M., V. Grimm., B.T. Martin., A.S. Johnston., K. Kułakowska., C.J. Topping., P. Calow., J. Nabe-Neilsen., P. Thorbek., and D.L. DeAngelis. 2013. Representing the acquisition and use of energy by individuals in agent-based models of animal populations. *Methods in Ecology and Evolution, 4*(2): 151-161

Sinclair, S.J., M.D. White., and G.R. Newell. 2010. How useful are species distribution models for managing biodiversity under future climates. *Ecology and Society*, 15(1): 1-8.

Soberón, J., and A.T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2: 1-10.

South, A. 1999. Extrapolating from individual movement behaviour to population spacing patterns in a ranging mammal. *Ecological Modelling*, 117(2): 343-360.

Tang, W., and D.A. Bennett. 2010. Agent-based Modeling of Animal Movement: A Review. *Geography Compass, 4*(7): 682-700.

Thurfjell, H., S. Ciuti., and M.S. Boyce. 2014. Applications of step-selection functions in ecology and conservation. *Movement Ecology, 2*(4).

Turner, M.G., R.H. Gardner., and R.V. O'Neill. 2001. *Landscape ecology in theory and practice: pattern and process*. New York, USA: Springer Science & Business Media.

van de Pol, M., L.D. Bailey., N. McLean., L. Rijsdijk., C.R. Lawson., and L. Brouwer. 2016.. Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution* 7(10): 1246-1257.

Van Moorter, B., D. Visscher, D., S. Benhamou, L. Börger, M.S. Boyce, and J.M. Gaillard. 2009. Memory keeps you at home: a mechanistic model for home range emergence. *Oikos*, *118*(5): 641-652.

Wilensky, U. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer Based Modeling, Northwestern University, Evanston, IL.

Wisz, M. S., J. Pottier., W.D. Kissling., L. Pellissier., J. Lenoir., C.F. Damgaard., C.F. Dormann., M.C. Forchhammer., J.A. Grythnes., A. Guisan., et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews*, *88*(1): 15-30.

ACKNOWLEDGEMENTS

I would like to thank Jennifer A. Miller for initial comments and support on this research, as well as Cody Schank, Molly Polk, Barney Warf, Robin Maier, and the anonymous reviewers for their comments and suggestions. I would also like to thank the Nystrom committee for their feedback. Furthermore, I would like to thank Richard Holland, Martin Wikelski, Franz Kümmeth and Carlos Bosque for making the oilbird telemetry data

available through Movebank, as well as the data providers who made their data openly available on GBIF.

AUTHOR BIO

PAUL HOLLOWAY is a lecturer in geographic information science and systems in the Department of Geography at University College Cork, Ireland. Email: paul.holloway@ucc.ie. His research and teaching interests include using GIScience and spatial analysis to address a suite of ecological, environmental and geographic issues. His current research addresses the long-standing issue of how to incorporate movement at different spatial and temporal extents into species distribution models.

FIGURES

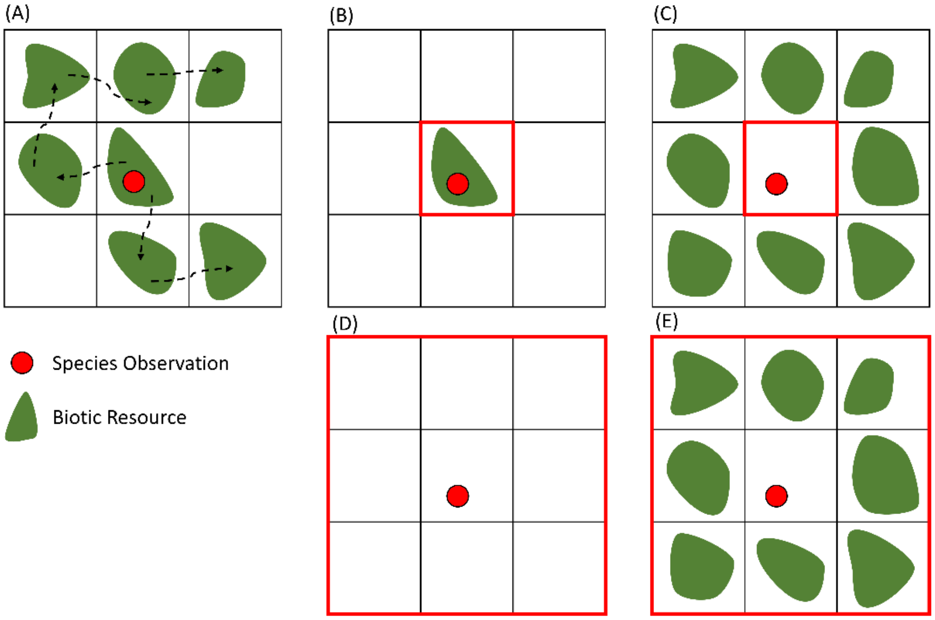

Figure 1: Conceptual diagram to illustrate the importance of the surrounding landscape configuration when developing variables to represent biotic resources in SDM. A) represents a sustainable and connected landscape for a hypothetical species; B-C) illustrate a local model approach, whereby the species observation is correlated with the environmental data within only the corresponding grid cell represented in red; D-E) illustrate a focal model approach, whereby the species observation is correlated with the environmental data within a predefined neighborhood represented in red (n.b., habitat suitability is still projected at the original resolution of the central grid).

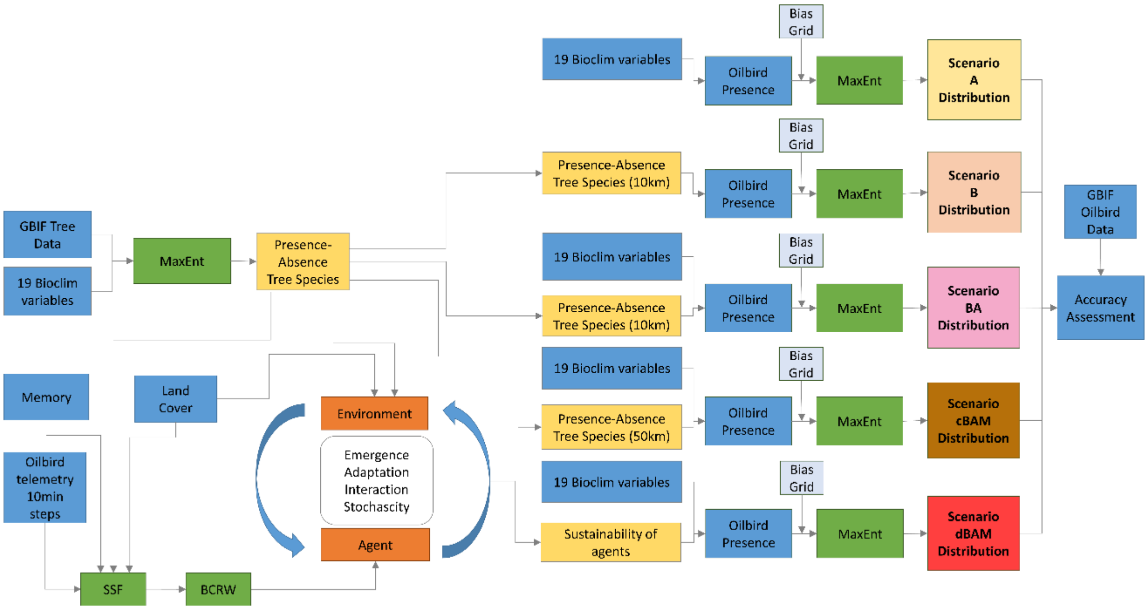

Figure 2: Conceptual diagram of the main steps in data analysis.

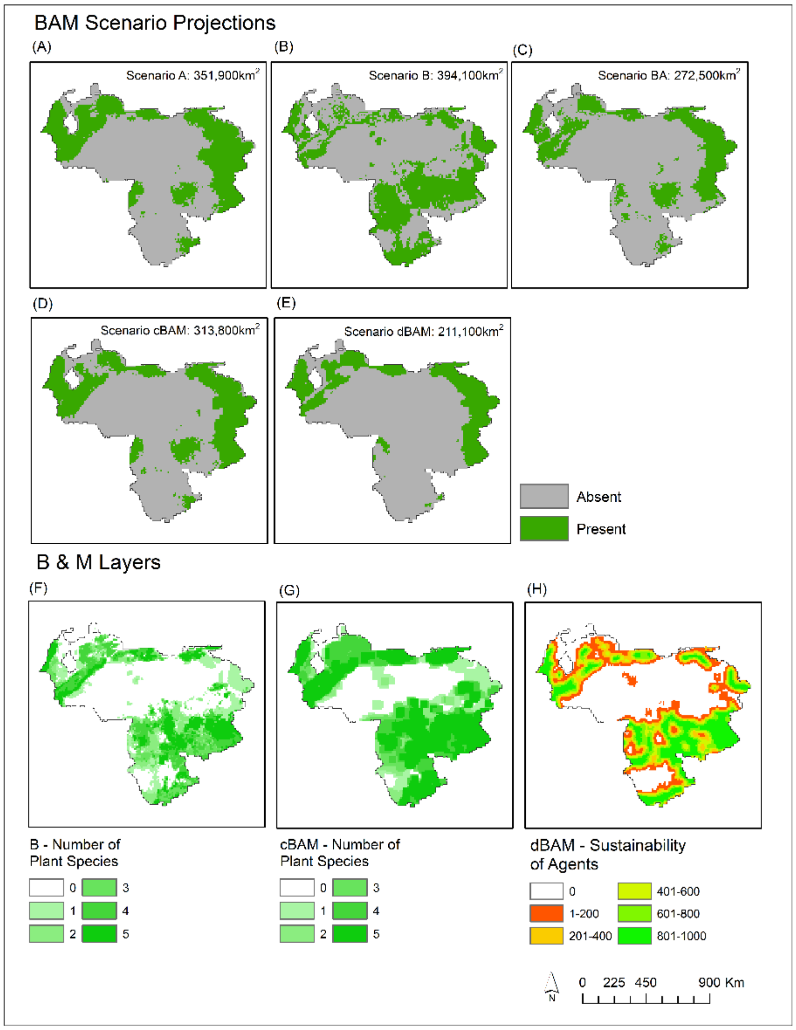

Figure 3: Spatial distribution of oilbirds in Venezuela projected from the five BAM scenarios and the data layers that represent the biotic resources and movement factors. Different scenarios are A) Scenario A, B) Scenario B, C) Scenario BA, D) Scenario cBAM, and E) Scenario dBAM. A threshold of 0.0302 delineated between presence and absence. This represents the proportion of presences compared to pseudo-absences. The **B** and **M** layers are represented by F) the number of plant species present within a 10km resolution,

G) the number of plant species present within a 50km focal area at a 10km resolution, and H) the number of oilbirds that can be sustained within a 50km focal area at a 10km resolution (the dynamic **B** and **M** layer).

Table 1: The results of the conditional logistic regression model used for step-selection function scores. Positive coefficients represent the probability of selecting steps into that habitat compared to reference habitats (e.g., for binary variables this is the opposite of what is modelled and for continuous variables this is compared to the current value), while negative coefficients represent the probability of selecting the reference habitat. * represents significant at α < 0.1, and ** represents significant at α < 0.05.

Variables	***Coefficient (Standard Error)***
Evergreen Forest	0.906 (0.466)*
Cropland	0.792 (0.547)
Rivers	< -0.000 (0.000)**
Memory	-3.115 (0.973)**
Resources	-1.340 (0.504)**
*Memory*Resources*	3.393 (0.987)**

Table 2: The three accuracy metrics of the five BAM scenarios using an independent test dataset: Lowest possible threshold (LPT), minimum predicted area (MPA), and the Boyce index (BI). Values in bold highlight the highest score that was recorded across the five BAM scenarios for each accuracy metric.

	LPT	**MPA**	**BI**
A	0.0002	0.0197	0.8210
B	**0.0172**	**0.0296**	0.0580
BA	0.0001	0.0256	0.7600
cBAM	0.0004	0.0220	0.6350
dBAM	0.0002	0.0271	**0.8920**