Evaluation of the ecological niche model approach in spatial conservation prioritization

Fumiko Ishihama , Akio Takenaka, Hiroyuki Yokomizo, Taku Kadoya

Published: December 20, 2019 • https://doi.org/10.1371/journal.pone.0226971

Abstract

Ecological niche models (ENMs) are widely used in spatial prioritization for biodiversity conservation (e.g. selecting conservation areas). However, it is unclear whether ENMs are always beneficial for such purposes. We quantified the benefit of using ENMs in conservation prioritization, comparing the numbers of species covered by conservation areas selected on the basis of probabilities estimated by ENMs (ENM approach) and those selected on the basis of raw observation data (raw-data approach), while controlling survey range, survey bias, and target size of conservation area. We evaluated three ENM algorithms (GLM, GAM, and random forests). We used virtual community data generated by simulation for the evaluation. ENM approach was effective when survey bias is strong, survey range is narrow, and target size of conservation area is moderate. The percentage of cases in which the ENM approach outperformed the raw-data approach ranged from 0.0 to 33% (GLM), 31% (GAM), and 75% (random forests) depending on conditions. The number of rare species (< 20 presence records) included in the conservation area based on the ENM approach was less than, or the same as, that of the raw-data approach. The unexpectedly limited cases in which the ENM approach was effective in the present research may depend on the conservation target we used (to cover as many species as possible in conservation area). Our results highlight urgent need for evaluating ENM's effectiveness under other conservation targets for wise use of ENM in conservation prioritization.

Citation: Ishihama F, Takenaka A, Yokomizo H, Kadoya T (2019) Evaluation of the ecological niche model approach in spatial conservation prioritization. PLoS ONE 14(12): e0226971. https://doi.org/10.1371/journal.pone.0226971

Editor: Paulo De Marco Júnior, Universidade Federal de Goias, BRAZIL

Received: April 12, 2019; **Accepted:** December 9, 2019; **Published:** December 20, 2019

Data Availability: All relevant data are within the manuscript and its Supporting Information files.

Funding: The author FI, AT and TK were supported by the Environment Research and Technology Development Fund (S9) of the Ministry of the Environment.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Establishment of conservation area networks is central to the conservation of biodiversity, and it is of great consequence to select conservation areas appropriately [1, 2] in the face of the current biodiversity crisis [3]. Sophisticated computational site-selection algorithms have been developed and are widely used to help identify cost-effective conservation area networks that fulfill conservation targets [4]. However, the ability of these algorithms to select appropriately representative conservation areas depends largely on data such as species presence and the extent of vegetation types. Such raw survey data have various limitations, including limited survey ranges, bias in survey ranges, and misclassification of species. A limited survey range and omission errors can reduce the efficacy and options of conservation area networks [5–8], and spatially and taxonomically biased data cause serious problems in conservation prioritization [9, 10]. Without correcting false-positive errors due to misidentification, some species can be excluded from a conservation area even though it was designed to include the species [5].

Considering these limitations of survey data, Margules and Sarkar [1] recommended analyzing or treating raw survey data before using them for conservation prioritization in systematic conservation planning. Using an ecological niche model (ENM) is one of such pretreatments. ENMs generate geographical maps of a species' probability of presence typically based on correlations between the presence, presence/absence, or abundance of species at multiple locations and the relevant environmental conditions[11, 12]. ENMs are generally expected to improve the comprehensiveness and representativeness of conservation areas by estimating species distributions in unsurveyed ranges and by reducing survey bias [13]. It remains unclear, however, under which condition ENMs can resolve the limitations of survey data for use in spatial prioritization, because they have their own limitations in terms of estimation performance, depending on various conditions. In the case of survey range, for example, it would be difficult to construct an accurate ENM when the survey range is too narrow, and when it is sufficiently large, using an ENM does not improve–and can even worsen–prioritization as compared with using raw data alone. Moreover, prediction by using the ENM approach may have greater uncertainty, and it is not clear how the uncertainty affects the results of conservation prioritization [14].

Thus, the effectiveness of ENMs in conservation prioritization should depend on various conditions. To our knowledge, no research has directly aimed to identify those conditions. Some studies have compared the size and structure of selected conservation networks chosen by using ENMs versus raw survey data [15–18]. These studies used distribution data of real organisms, however, and therefore the true distributions are unknown. Thus, these studies could not refer to the true number of species included in the networks as the baseline for the performance evaluation and could not truly assess whether ENMs actually improved conservation area prioritization. Despite the unclear nature of the benefit of using ENMs and the conditions needed for them to be effective in conservation prioritization, ENMs have been frequently used in studies of conservation prioritization [5, 19–24].

Here, we used realistic distribution data generated by simulating several realistic macro-assemblage properties such as species–abundance relationships. This virtual community data for which we knew the true species distribution allowed us to clarify the conditions under which the use of ENMs resulted in better conservation networks than using raw survey data alone. To evaluate the effectiveness of ENMs, we compared the performance with regard to spatial prioritization between two approaches: (1) the raw-data approach, in which we used only survey data for conservation prioritization; and (2) the ENM approach, in which we estimated ENMs based on the same survey data but used only the predicted probability of species presence for conservation prioritization. We used the "maximal coverage" target as the aim of conservation prioritization [25] in the present study: that is, the size of the conservation area was predetermined (hereafter "target size"), and we selected sites so that they covered as many species as possible. Performance here is defined as the number of species covered in the selected conservation area networks. Though there are variations of "maximal coverage" target to improve persistence of biodiversity in conservation areas networks, we used the simplest conventional target. We considered the effects of the most common limitations of raw survey data, namely survey range and degree of spatial survey bias, on the relative performance of the two approaches.

Materials and methods

The flow of the analysis is shown in Fig 1, and we explain the details of each analytical process in the following sections.

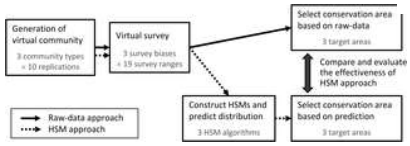

Fig 1. Flow chart of analysis of the effectiveness of the ecological niche model (ENM) approach in conservation prioritization, based on simulated virtual distribution data.
https://doi.org/10.1371/journal.pone.0226971.g001

Generating virtual community structure

We used artificially generated virtual community data to evaluate the performance of ENMs in conservation prioritization. We generated a species distribution dataset on a one-dimensional space (grid) in which there was one environmental variable, and the environmental value increased linearly as the spatial axis value increased. Both the number of sites (grid cells) and the number of species in the overall range were fixed to 1000.

To generate various and realistic community structures on this environmental space, we developed a program named ComGen (S1 supplemental methods), a generator of an virtual community structure written in Perl. The program generates a dataset of the presence and absence of multiple species at multiple sites that satisfies the three constraints; 1. frequency distributions of number of presence sites per species (rank-abundance curves), 2. frequency distributions of number of species per site (richness patterns), and 3. environmental preferences of individual species. Those assemblage patterns are given to the ComGen program as input.

The rank-abundance curve is expected to affect the accuracy of an ENM, because estimating the distributions of rare species is generally difficult and the curve determines the proportion of rare species in the entire species pool. The species richness pattern among sites influences the effectiveness of a conservation area network, because if many species are concentrated within a small number of sites, many species can be covered by a small conservation area network and vice versa. Therefore, we used realistic rank-abundance patterns: that is, log-linear and log-normal curves and patterns extracted from actual distribution data of various taxonomic groups (amphibians and reptiles, birds, butterflies, dragonflies, freshwater fishes, land and freshwater mollusks, and mammals) in Japan obtained from reports of the Fifth National Survey on the Natural Environment [26]. Log-linear and log-normal curves are commonly used to fit rank-abundance curves of real communities [27]. There are few reports on richness patterns in real communities, so we used log-linear and extracted patterns from the actual distribution data of the Fifth National Survey on the Natural Environment.

We assumed that each species had a unimodal environmental preference. The environmental suitability was a logit conversion of a quadratic function of the environment, and the central value of a suitable environment was chosen randomly. We determined presence/absence of each species considering this habitat suitability and above-mentioned rank-abundance curve and richness pattern.

In the generating process of presence/absence of each species, we preset the number of species at each site and the number of presence sites) of each species so that they meet the above-mentioned richness pattern and rank-abundance curve. The most preferable environmental condition for each species is also preset at random, while the width of the preferable environmental condition depends on the abundance of each species.

Then, we generate a tentative distribution of each species by randomly assigning the predetermined number of presence sites, considering the number of species present at each site, but without considering the environmental preference of the species. Next, we adjust the random tentative distribution to reflect the environmental preference of each species. We randomly select two sites, and a pair of species each of which is present in one of the two sites are swapped, if the swapping makes the distribution more consistent with the environmental preferences. This process is repeated until there is no more swapping to improve consistency with the environmental preferences. The richness pattern and rank-abundance curve are kept unchanged through the swapping process. Details of this process are explained in S1 supplemental methods. Absence sites of each species are defined as the sites where the presence of the species is not assigned.

We generated various community structures changing parameter values of rank-abundance and richness patterns. See S1 Table for the names of parameters we controlled. Because the generation procedure includes stochastic processes and the generated community structure fluctuates randomly, we repeatedly generated 10 communities for each parameter setting. We adopted this iteration number considering the balance between variability of results and calculation time. The calculation time was about 15 min on average per simulation run, including all the processes in Fig 1.

Virtual survey

We surveyed the virtual community in survey areas of various sizes and with various levels of spatial survey bias. Survey areas were set in blocks, and we controlled the level of spatial sampling bias by changing the number and size of blocks. Survey bias is greater with a smaller number of larger blocks. Although there are other ways to generate spatial survey bias, we chose the block method because the response of ENM effectiveness to survey bias was largest with this method. To set survey grid cells, we first divided the total area (1000 grid cells) into N equally sized blocks. When the predetermined proportion of the surveyed range is R, $1000 \times R/N$ consecutive grid cells are surveyed in each block. The position of the consecutive survey grid cells in each block is randomly set. When $1000 \times R$ is indivisible by N, the remainder is randomly assigned to survey blocks so that the total number of grid cells surveyed becomes $1000 \times R$. We used 19 R values from 5% (50 grid cells) to 950% (950 grid cells) at 5% intervals, and three levels of survey bias (no bias, weak bias with $N = 5$ survey blocks, and strong bias with $N = 2$ blocks).

Ecological niche model

We assumed that species distribution data were presence/absence data, which is one of the most frequently used data type in ENMs[28], and we used three ENM algorithms including two classical statistical algorithms; generalized linear model (GLM)and generalized additive model (GAM), and one machine learning algorithm; random forests (RF) [29]. We chose these algorithms because of the following reasons. GLM and GAM are one of the most popular and established statistical models, and GAM is able to fit non-linear more complicated function than GLM. RF is one of the most popular machine learning algorithms known for good performance and requires relatively short calculation time including tuning parameters for model complexity.

For GLM, we used logit-link function and binomial error distribution, and linear predictor was a quadratic function, which was the same functional form as used in generating the virtual community. Thus, we assumed that we knew the true functional form of each organism's response to the environmental condition. For GAM, we also used logit-link function and binomial error distribution, and spline curve.

Selecting the conservation area

We selected the conservation area on the basis of complementarity [30, 31] in both the ENM and raw-data approaches. We used the "maximal coverage" target as the aim of conservation prioritization [25], and we selected sites so that they covered as many species as possible.

Another type of conservation target is the "minimum set" target [25], in which as small a number of sites as possible is selected to cover all species or a certain proportion of species, with no limitation on the size of the conservation area. However, the minimum set target was not applicable in our case because we assumed the survey range was limited and we did not know the true number of species in the whole area.

Target size directly affects the potential number of species to be conserved in the area, and thus it is one of the most important factors that influence the results of conservation prioritization. In the evaluation, we used five target sizes, i.e.1.0, 2.5, 5.0, 9.1, and 17% of the total area. The 9.1 and 17% target size corresponds to the percentage area of national parks in Japan and the Aichi Target of the Convention on Biodiversity, respectively.

Evaluation of the ENM approach

We evaluated the effectiveness of the ENM approach in conservation area selection based on the following "effectiveness index": (number of species covered in the conservation area by ENM approach)/(number of species covered in the conservation area by raw-data approach)− 1. An effectiveness index value greater than zero indicates that the ENM approach is beneficial: that is, we could select a conservation area that covers more species by using the ENM approach than by using the raw-data approach.

Selecting conservation areas by using the raw-data approach

In the raw-data approach, we selected the conservation area on the basis of complementarity by using the greedy algorithm. The greedy algorithm is a heuristic algorithm that uses stepwise analysis to select conservation area [32]. In the first step, the cell with the largest species richness is selected, and then additional cells that add the most additional species are selected step by step.

However, survey range is sometimes insufficient to select a predetermined target size of the conservation area, such as when the survey range is smaller than the target size or when very few species occur in the survey range. Thus, we used the following procedure: (1) we selected a conservation area that covered all the species recorded in the surveyed range, and (2) when the surveyed range was narrower than the target size, we added sites to the conservation area by randomly selecting sites from the unsurveyed range until the conservation area equaled the target size.

Selecting conservation areas by using the ENM approach

In the ENM approach, we selected the conservation area on the basis of the estimated probability of the presence of each species over the whole area under consideration for conservation. We ignored species that never appeared in the survey range because we could not construct an ENM for such species. We directly used the presence probability and selected conservation areas to maximize the expected number of species represented in the area by the greedy algorithm, as described by Polasky, Camm [33].

There are other ways to select complementary conservation areas on the basis of the probability of species presence, such as converting probability into presence/absence data by using threshold values [16, 34, 35]. We performed a preliminary analysis in which we converted the probability of presence into presence/absence data by using various threshold values, and the results we obtained were qualitatively the same as those with the direct use of probability.

Results

Among the various community structures we tried, we identified those cases in which the ENM approach performed best in each type of rank-abundance pattern (i.e., type 1: log-linear, type 2: log-normal, and type 3: realistic patterns). The parameters used to generate these three virtual communities are shown in the S1 Table, and the community structures are illustrated in Fig 2.

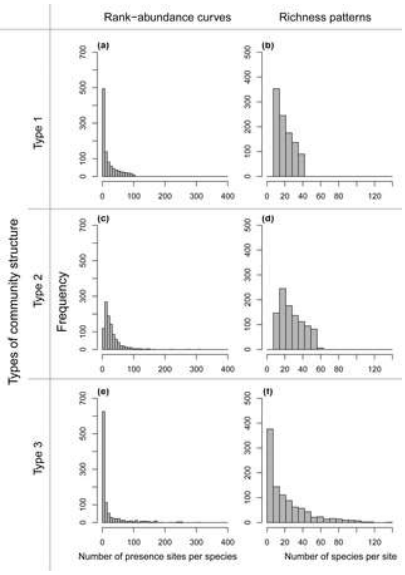

Fig 2. Three combinations of rank-abundance curve and richness pattern used to generate community structure types of virtual distribution data used for the evaluation.
(a), (c), (e) Histograms of number of presence sites per species (rank-abundance curve) and (b), (d), (f) number of species per site (richness pattern) in communities with (a), (b) type 1: log-linear, (c), (d) type 2: log-normal, and (e), (f) type 3: realistic patterns of rank-abundance curves in Japanese freshwater mollusks.
https://doi.org/10.1371/journal.pone.0226971.g002

For each of these three community structure types, we generated 10 virtual communities. Thus, we performed 25650 runs in total resulting from 3 community structure types × 10 iterations × 19 survey ranges × 3 survey biases × 5 conservation targets × 3 ENMs.

Among the three ENM algorithms, the performance of RF was the best, and we show the effectiveness of the ENM approach using RF in these three community structures in Figs 3–5. Effectiveness of ENM approach using GLM and GAM are shown in Figs A-F in S1 Figs. The effectiveness of the ENM approach was measured by effectiveness index which takes a value greater than zero when the ENM approach is beneficial. When RF was used, the percentage of cases in which effectiveness index > 0 was 75% (144 of 190 cases) under the best condition (190 cases = 19 sampling ratios × 10 replications with community type 1, target size of 9.1%, and strong survey bias; Fig 3), and it was 29% (2469 of 8550 cases) on average of all conditions we showed in Figs 3–5. When GLM was used, the percentages dropped to 33% (best) and 6.9% (on average), and when GAM was used, the percentages were 31% and 4.5%, respectively.

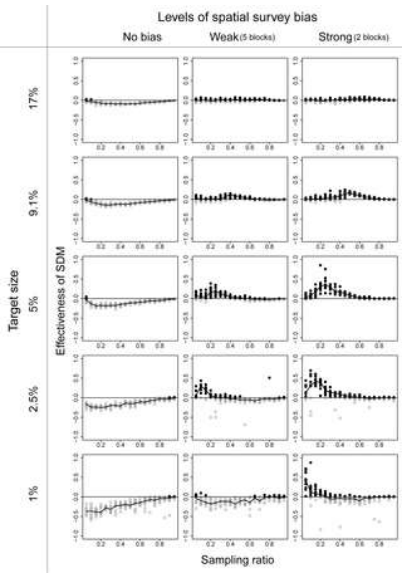

Fig 3. Performance of the ecological niche model (ENM) approach using RF in the community structure type 1 illustrated in Fig 2A and 2B.
Black filled circles are values when the ENM approach was beneficial, and gray open circles were the other case. Line chart shows median value in each condition. Target sizes of 9.1% and 17% correspond to the percentage area of national parks in Japan and the Aichi Target of the Convention on Biodiversity, respectively.
https://doi.org/10.1371/journal.pone.0226971.g003

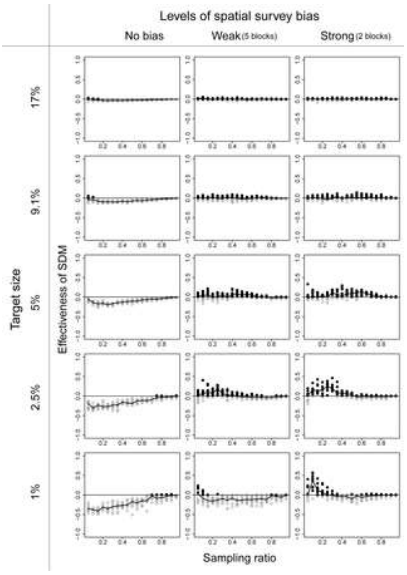

Fig 4. Performance of the ecological niche model (ENM) approach using RF in the community structure type 2 illustrated in Fig 2C and 2D. See Fig 3 for more details.
https://doi.org/10.1371/journal.pone.0226971.g004

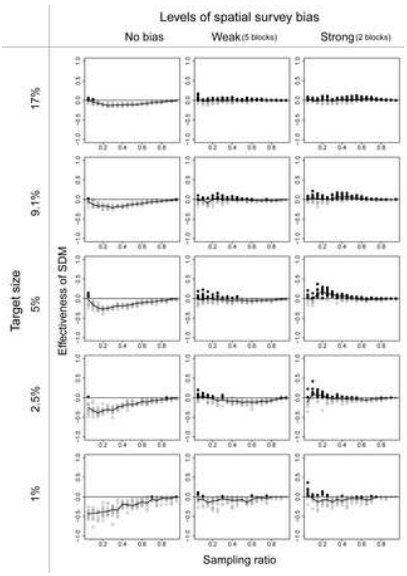

Fig 5. Performance of the ecological niche model (ENM) approach using RF in the community structure type 3 illustrated in Fig 2E and 2F. See Fig 3 for more details.
https://doi.org/10.1371/journal.pone.0226971.g005

When we used RF, the conditions ENM approach to be beneficial were that survey bias was large, survey range was narrow to medium, and target size of the conservation area was small to medium. The parameter value of the surveyed range in which the ENM approach became beneficial varied depending on the target size of the conservation area, and the parameter value became smaller with smaller target size. When all three conditions were met, the effectiveness index varied from 0.0 to 1.0. In other conditions, the number of species covered by the selected conservation area based on ENM approach decreased in comparison with the raw-data approach when bias is small and/or target size was small (lower left panels of Figs 3–5). When bias was large and target size was large (upper right panels of Figs 3–5), the numbers of species covered were almost similar (the values of effectiveness index were near 0) between the two approaches.

When we used GAM or GLM, the general tendencies were similar to the results of RF, except that the values of effectiveness index were smaller than 0 even when bias was large and target size was large (upper right panels of S1 Figs and S2 Figs).

In the conservation areas selected by the ENM approach, the ratio of rare species (species with <20 presence records) covered tended to be smaller than that of the raw-data approach, and this tendency was especially marked when the ENM approach was effective, irrespective of ENM algorithms (Fig 6A–6C). The number of rare species covered by ENM approach using RF tended to be similar to or slightly smaller than that of the raw-data approach (Fig 6D). On the other hand, the number of rare species covered by ENM approach using GLM or GAM was generally smaller than that of the raw-data approach, and this tendency was especially marked when the ENM approach was not effective (Fig 6E and 6f). When the effectiveness of ENM was measured on the basis of

the number of rare species by substituting “number of species” in the effectiveness index with “number of rare species,” the parameter ranges for the ENM approach to be beneficial became narrower, and the percentages of cases in which effectiveness index > 0 for rare species using RF, GLM, and GAM were 20, 2.1, and 2.5%, respectively, on average of all conditions for each algorithm (Figs A-I in S3 Figs).

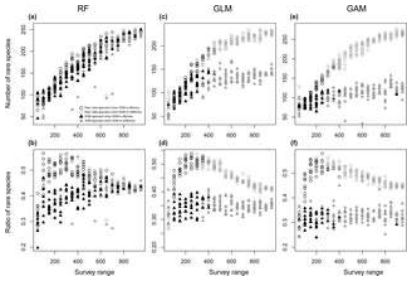

Fig 6.
Ratios of rare species (< 20 presence records) among species in conservation area networks (a), (b), (c), and numbers of rare species covered in conservation area networks (d), (e), (f). The ecological niche model (ENM) approach using RF is considered effective when the total number of species in the selected conservation area networks is larger than that by using the raw-data approach. The values are in the community structure type 1 illustrated in Fig 2A and 2B, with strong survey bias and a 2.5% target size: that is, in one of the best-performing cases of the ENM approach.
https://doi.org/10.1371/journal.pone.0226971.g006

Discussion

The conditions under which the ENM approach outperformed the raw-data approach in conservation area selection with a target to cover as many species as possible included: (1) large survey bias; (2) narrow to medium survey range; and (3) small to medium target size of conservation area. All of these conditions must be satisfied for the ENM approach to be beneficial from the viewpoint of number of species covered, and the parameter ranges of these conditions were not wide.

The reason for the conditions under which ENMs to be beneficial in conservation prioritization

Two inherent properties of ENMs are likely to be responsible for the conditions under which ENM-approach to be beneficial in conservation area selection by complementality with the target to cover as many species as possible: (1) ENMs require a certain number of presence records per species for good estimation and cannot properly estimate the distribution of rare species unless the survey range is sufficiently large; and (2) ENMs cannot estimate the distribution range of species that never occur in the survey range.

The parameter conditions under which the ENM approach became beneficial was closely related to the ENMs’ property (1). In the cases of large survey bias (condition 1) and narrow or medium-sized survey range (condition 2), only a few species appeared in the survey range (Fig 7), because only species in limited environmental conditions were sampled under such conditions. There is a trade-off between the total number of species in the survey range and presence records per species, and fewer total species results in increased presence records per species (Fig 7). The trade-off exists because the total number of presence records (i.e., the sum of species richness at all surveyed sites) in survey ranges of the same size is expected to be constant irrespective of their levels of spatial bias, because the species richness at each site was assigned in a spatially random manner in the present simulation (Fig D in S1 Supplemental Methods). Therefore, under conditions 1 and 2, a small number of species frequently appeared in the survey range, and ENMs could be constructed for a larger proportion of the species. In the raw-data approach, however, a reduction in the number of species in the survey range directly resulted in a reduction of conservation efficiency. Thus, under conditions 1 and 2, the ENM approach outperformed the raw-data approach.

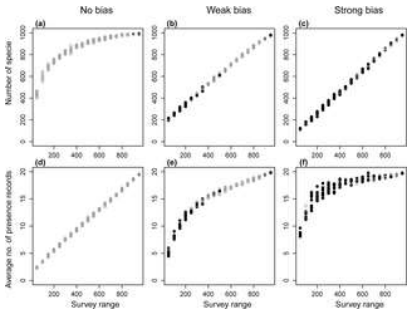

Fig 7.
(a), (b), (c) Numbers of species in surveyed sites and (d), (e), (f) average number of presence records of each species observed at surveyed sites. With (a), (d) no, (b), (e) moderate (five survey blocks), and (c), (f) strong (two survey blocks) survey bias. The values are in the community structure illustrated in Fig 2A and 2B, with a 2.5% target size. Black filled circles are values when the ENM approach using RF was effective, and gray open circles were the other case.
https://doi.org/10.1371/journal.pone.0226971.g007

When the target size was small to medium (condition 3), the ability of ENMs to predict common species’ distributions was fully utilized. To increase the number of species included in a small conservation area, we had to find sites of high species richness. The species included needed not be rare species, but could be common species, given that we could conserve only a small number of

species in a small conservation area. Finding sites where common species are concentrated is exactly the condition under which ENMs work well.

Conditions 2 (narrow to medium survey range) and 3 (small to medium target size) were interrelated. With a smaller target size, the survey range in which a ENM was beneficial became narrower. This is because in a small to medium conservation area, only a few species can potentially be covered, and even a narrow survey range is sufficient to fill such a small conservation area.

The fact that ENMs cannot estimate the distribution range of species that never occur in the survey range is obvious but implies that they have a limited ability to complement spatial information. The main reasons why ENMs were expected to have some benefit in conservation prioritization are that they can complement distribution information in unsurveyed ranges and that they can correct the survey bias. However, a ENM corrects the distribution information of only species that appear in the survey range. The benefit of this correction would be small under the target to cover as many species as possible, because the raw-data approach is also able to cover observed species when the target size is sufficiently large. A more serious problem caused by a biased survey range is that species that live only in unsurveyed range are not considered in conservation prioritization, and an ENM is not able to solve this problem.

Among the three ENM algorithms, RF performed considerably better than GLM and GAM. Though the above-mentioned three conditions for ENM approach to perform better than raw-data approach were common to all three ENM algorithms, RF resulted in negative effectiveness value much less frequently than the other two algorithms especially when target size is large. The superiority of RF is thought to be attributable to its ability to predict distribution of rare species relatively well. The number and ratio of rare species covered in the conservation selected by using RF were notably larger than those of GLM and GAM, especially when survey range was wide (Fig 6), and prediction accuracies were higher when RF was used than when GLM or GAM was used (S4 Fig).

Previous researches compared characteristics of conservation areas selected based on raw-data and ENMs using real distribution data [14–16]. These studies used "minimal area" target, which seeks for minimal conservation area that satisfies a target such as covering all the species. These studies found common characteristics that conservation areas based on ENMs tend to be narrower than that based on raw-data. This is because ENMs estimate wider habitats than known limited presence points, and more frequent overlaps among distribution ranges of different species occur. These studies could not compare effectiveness of raw-data and ENM approaches, because there was no perfect distribution data for the real communities they studied. Our study is the first study which evaluate the conditions for ENM approach to be effective in comparison to raw-data approach using virtual community data as far as we know.

We used ENM for each species, which is the most popular type of distribution models used in conservation prioritization. In contrast, Arponen, et al. [36] compared the efficiency of various combinations of community-level modelling, which model multiple species at once, and maximization algorithms for conservation prioritization by using simulated community data. They showed that maximization of complementary richness–a procedure that accounts for gradients in species richness and non-constant turnover rates of community composition in environmental space–outperforms other approaches including direct selection based on raw data, irrespective of target size they tested. They used virtual community composed of 3000 species in 160 × 160 grid cells with gradient of two environmental variables, biased and fixed narrow survey range (200 sites, randomly scattered within the 25% of total environmental range), and narrow target size (2–64 sites). Direct comparison between their results and our study is difficult because of the difference in the type of modelling, but we can find common point that modelling is effective when survey range is relatively narrow and environmentally biased, and target size is small.

Implication for future research

It is possible that the unexpectedly narrow parameter ranges in which ENM approach is effective in the present result is dependent on the target of conservation prioritization. We used the "maximal coverage" target to cover as many species as possible. To achieve this target, accuracy of distribution estimation of rare species is important, and this situation is unfavorable to the ENM approach. There are other possible targets for conservation prioritization [25]. For example, when the target is to better cover the distribution range of well-known species in the conservation area, it would be just the case in which ENM-approach is very useful. Although it is beyond the scope of the current study, it is promising avenue for future research to examine relationships between target types and performance of ENMs and to look for the target that has higher affinity with ENM approach.

In addition, the setting in the current analysis to make up for the shortage of conservation area by randomly selecting sites from unsurveyed range when the survey range is insufficient in the raw-data approach may be unrealistic. In reality, we would not select a conservation area in such a way. Another way may be to use a hybrid approach that combines an ENM with raw data. For example, one could select a conservation area on the basis of the raw-data approach and then make up for the shortage by selecting additional sites based on prediction by ENMs. Or, one could select a conservation area on the basis of the raw-data for rare species or species of low ENM accuracy, and ENMs for other species. We performed a preliminary analysis of such a hybrid approach, but among all our attempts there was no case in which the hybrid approach performed better than the raw-data approach. This is probably because an ENM cannot provide any profitable information as compared with random selection of sites owing to its limited ability to complement spatial information (an ENM's second property). However, there are various possible ways to combine ENMs and raw data, and further analysis of such hybrid approaches may be warranted.

There would be room to improve reality of virtual community structure. We generated virtual communities which satisfy three constraints, i.e. rank-abundance curve, richness pattern, and habitat suitability of each species. We did not consider environmental or spatial structures in richness patterns, but in reality, existence of such structures is well known (e.g. latitudinal gradient in species diversity). In addition, we considered only one environmental variable in this research. Developing algorithm to generate virtual community which satisfy spatio-environmental structures in richness pattern and having multiple environmental variables is challenging, but it would be very interesting to validate the effect of these factors on the performance of ENM approach.

Our present analysis is the first step to understand the characteristics of ENM approach in conservation prioritization, and the present results highlight the importance of future researches in the above-mentioned perspectives for wise use of ENM in conservation prioritization.

Supporting information

S1 Supplemental Methods. Generating virtual community structure by using the ComGen package.
https://doi.org/10.1371/journal.pone.0226971.s001
(PDF)

S1 Table. Settings of the ComGen package used to generate virtual community structures for the analysis.
https://doi.org/10.1371/journal.pone.0226971.s002
(PDF)

S1 Figs. The effectiveness of ENM approach using GLM.
https://doi.org/10.1371/journal.pone.0226971.s003
(PDF)

S2 Figs. The effectiveness of ENM approach using GAM.
https://doi.org/10.1371/journal.pone.0226971.s004
(PDF)

S3 Figs. The effectiveness of ENM approach using Random Forests measured on the basis of the number of rare species in the selected conservation area.
https://doi.org/10.1371/journal.pone.0226971.s005
(PDF)

S4 Figs. Accuracy of each ENM measured by AUC (area under the curve).
https://doi.org/10.1371/journal.pone.0226971.s006
(PDF)

S1 Data. The presence/absence distribution data of each species in the virtual community used for the analysis in Figs 3–5.
The data is "list" format of the statistical software R, and the list is composed of four hierarchies (1. the Figure in which the data used, 2. target size, 3. level of the spatial survey bias, and 4. 10 iterations).
https://doi.org/10.1371/journal.pone.0226971.s007
(ZIP)

Acknowledgments

The author FI, AT and TK were supported by the Environment Research and Technology Development Fund (S9) of the Ministry of the Environment.

References

1. Margules CR, Sarkar S. Systematic conservation planning. New York: Cambridge University Press; 2007. 270pp p.
2. Fuller RA, McDonald-Madden E, Wilson KA, Carwardine J, Grantham HS, Watson JEM, et al. Replacing underperforming protected areas achieves better conservation outcomes. Nature. 2010;466(7304):365–7. pmid:20592729
View Article • PubMed/NCBI • Google Scholar
3. Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, et al. The biodiversity of species and their rates of extinction, distribution, and protection. Science. 2014;344(6187):1246752. pmid:24876501.
View Article • PubMed/NCBI • Google Scholar
4. Sarkar S, Pressey RL, Faith DP, Margules CR, Fuller T, Stoms DM, et al. Biodiversity Conservation Planning Tools: Present Status and Challenges for the Future. Ann Rev Env Resour. 2006;31(1):123–59.
View Article • Google Scholar
5. Rondinini C, Wilson KA, Boitani L, Grantham H, Possingham HP. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol Lett. 2006;9(10):1136–45. pmid:16972877
View Article • PubMed/NCBI • Google Scholar
6. Freitag S, Jaarsveld ASV. Sensitivity of selection procedures for priority conservation areas to survey extent, survey intensity and taxonomic knowledge. Proc R Soc Lond, Ser B: Biol Sci. 1998;265(1405):1475–82. PMC1689332.
View Article • Google Scholar
7. Gladstone W, Davis J. Reduced survey intensity and its consequences for marine reserve selection. Biodivers Conserv. 2003;12(7):1525–36. WOS:000182721700015.
View Article • Google Scholar
8. Gaston KJ, Rodrigues ASL. Reserve Selection in Regions with Poor Biological Data. Conserv Biol. 2003;17(1):188–95.
View Article • Google Scholar
9. De Ornellas P, Milner-Gulland EJ, Nicholson E. The impact of data realities on conservation planning. Biol Conserv. 2011;144(7):1980–8.
View Article • Google Scholar
10. Andelman SJ, Willig MR. Alternative configurations of conservation reserves for paraguayan bats: Considerations of spatial scale. Conserv Biol. 2002;16(5):1352–63. WOS:000178183600022.
View Article • Google Scholar
11. Elith J, Leathwick JR. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics. 2009;40(1):677–97.
View Article • Google Scholar
12. Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, et al. Predicting species distributions for conservation decisions. Ecol Lett. 2013;16(12):1424–35. WOS:000326114000011. pmid:24134332

View Article • PubMed/NCBI • Google Scholar

13. Elith J, Leathwick J. The contribution of species distribution modelling to conservation prioritization. In: Moilanen A, Wilson KA, Possingham HP, editors. Spatial conservation prioritization. New York: Oxford University Press; 2009. p. 70–93.

14. Moilanen A. Reserve selection using nonlinear species distribution models. The American Naturalist. 2005;165(6):695–706. pmid:15937749.
View Article • PubMed/NCBI • Google Scholar

15. Freitag S, Nicholls AO, Jaarsveld AS. Nature reserve selection in the Transvaal, South Africa: what data should we be using? Biodivers Conserv. 1996;5(6):685–98.
View Article • Google Scholar

16. Wilson KA, Westphal MI, Possingham HP, Elith J. Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biol Conserv. 2005;122(1):99–112.
View Article • Google Scholar

17. Carvalho SB, Brito JC, Pressey RL, Crespo E, Possingham HP. Simulating the effects of using different types of species distribution data in reserve selection. Biol Conserv. 2010;143(2):426–38.
View Article • Google Scholar

18. Loiselle BA, Howell CA, Graham CH, Goerck JM, Brooks T, Smith KG, et al. Avoiding Pitfalls of Using Species Distribution Models in Conservation Planning. Conserv Biol. 2003;17(6):1591–600.
View Article • Google Scholar

19. Esselman PC, Allan JD. Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica. Freshwat Biol. 2011;56(1):71–88.
View Article • Google Scholar

20. Peralvo M, Sierra R, Young K, Ulloa C. Identification of Biodiversity Conservation Priorities using Predictive Modeling: An Application for the Equatorial Pacific Region of South America. Biodivers Conserv. 2007;16(9):2649–75.
View Article • Google Scholar

21. Pawar S, Koo MS, Kelley C, Ahmed MF, Chaudhuri S, Sarkay S. Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. Biol Conserv. 2007;136(3):346–61. WOS:000246947300002.
View Article • Google Scholar

22. Leathwick J, Moilanen A, Francis M, Elith J, Taylor P, Julian K, et al. Novel methods for the design and evaluation of marine protected areas in offshore waters. Conserv Lett. 2008;1(2):91–102. WOS:000207586900006.
View Article • Google Scholar

23. Linke S, Pressey RL, Bailey RC, Norris RH. Management options for river conservation planning: condition and conservation re-visited. Freshwat Biol. 2007;52(5):918–38. WOS:000245987200012.
View Article • Google Scholar

24. Ortega Huerta MA. Fragmentation patterns and implications for biodiversity conservation in three biosphere reserves and surrounding regional environments, northeastern Mexico. Biol Conserv. 2007;134(1):83–95. WOS:000243630200009.
View Article • Google Scholar

25. Cabeza M, Moilanen A. Design of reserve networks and the persistence of biodiversity. Trends Ecol Evol. 2001;16(5):242–8. pmid:11301153
View Article • PubMed/NCBI • Google Scholar

26. Reports on 5th National Basic Survey on Natural Environment [Internet]. Biodiversity Center of Japan, Ministry of the Environment. [cited August, 2012]. Available from: http://www.biodic.go.jp/ne_research_e.html#id03.

27. McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecol Lett. 2007;10(10):995–1015. pmid:17845298
View Article • PubMed/NCBI • Google Scholar

28. Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, et al. Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecol Biogeogr. 2015;24(3):276–92. WOS:000349391600002.
View Article • Google Scholar

29. Breiman L. Random forests. Machine Learning. 2001;45(1):5–32. WOS:000170489900001.
View Article • Google Scholar

30. Wilson KA, Cebeza M, Klein CJ. Fundamental concepts of spatial conservation prioritization. In: Moilanen A, Wilson KA, Possingham HP, editors. Spatial conservation prioritization. 1 st ed. New York: Oxford University Press; 2009. p. 16–27.

31. Kirkpatrick JB. An iterative method for establishing priorities for the selection of nature reserves: An example from Tasmania. Biol Conserv. 1983;25(2):127–34. http://dx.doi.org/10.1016/0006-3207(83)90056-3.
View Article • Google Scholar

32. Mehri A, Salmanmahiny A, Mirkarimi SH, Rezaei HR. Use of optimization algorithms to prioritize protected areas in Mazandaran Province of Iran. J Nat Conserv. 2014;22(5):462–70. https://doi.org/10.1016/j.jnc.2014.05.002.
View Article • Google Scholar

33. Polasky S, Camm JD, Solow AR, Csuti B, White D, Ding R. Choosing reserve networks with incomplete species information. Biol Conserv. 2000;94(1):1–10.
View Article • Google Scholar

34. Sarkar S, Pappas C, Garson J, Aggarwal A, Cameron S. Place prioritization for biodiversity conservation using probabilistic surrogate distribution data. Divers Distrib. 2004;10(2):125–33.
View Article • Google Scholar

35. Tole L. Choosing reserve sites probabilistically: A Colombian Amazon case study. Ecol Model. 2006;194(4):344–56.
View Article • Google Scholar

36. Arponen A, Moilanen A, Ferrier S. A successful community-level strategy for conservation prioritization. J Appl Ecol. 2008;45(5):1436–45.
View Article • Google Scholar

Check List 10(3): 689–691, 2014

ISSN 1809-127X (available at *www.checklist.org.br*)

Check List
Journal of species lists and distribution

Using species distribution models as a tool to discover new populations of *Phaedranassa brevifolia* Meerow, 1987 (Liliopsida: Amaryllidaceae) in Northern Ecuador

Nora H. Oleas [1,2,3*], Alan W. Meerow [4], Kenneth J. Feeley [1,3], Jennifer Gebelein [5] and Javier Francisco-Ortega [1,3]

1 Universidad Tecnológica Indoamérica, Centro para el estudio de la Biodiversidad y Cambio Climático, Machala y Sabanilla, CP. EC170103, Quito, Ecuador.
2 Florida International University, Department of Biological Sciences, 11200 SW 8th St., Miami, Florida, 33199, USA.
3 Fairchild Tropical Botanic Garden, Kushlan Tropical Science Institute, 10901 Old Cutler Road, Coral Gables, Miami, Florida, 33156, USA.
4 USDA-ARS-SHRS, National Germplasm Repository, 13601 Old Cutler Rd., Miami, Florida 33158, USA.
5 Florida International University, Department of Earth and Environment, 11200 SW 8th St., Miami, Florida, 33199, USA.
* Corresponding author: e-mail: *noraoleas@gmail.com*

Abstract: *Phaedranassa brevifolia* Meerow (1987) is an endangered plant species endemic to Ecuador. Until recently, this species was only known by the type location and four other adjacent subpopulations in a restricted area of less than four km^2. We combined geo-referenced occurrences of this species and bioclimatic variables to generate a prospective species distribution model. Using the resulting map as a guide for field work, we found three new populations of *P. brevifolia,* increasing its geographic distribution to 16 km^2.

DOI: 10.15560/10.3.689

Phaedranassa brevifolia was first described from a single specimen collected in 1978 in Northern Ecuador (Meerow 1987) in a habitat classified as Dry Montane Scrub, characterized by thorny dry forest (Valencia *et al.* 1999). The above-ground parts of this bulbous species consist of only one or two leaves approximately 12 cm in length (Meerow 1990), which are hard to locate among other shrubs. The reproductive biology of *P. brevifolia* is unknown, but the general phenological pattern of the genus is to flower once annually, before the wet season (N. Oleas, personal observations). Above-ground leaves are absent before flowering, which makes spotting individuals especially difficult during that time of the year. At flowering, each plant is a 39-56 cm stem or scape, with five to seven funnel-shaped to ventricose flowers (Meerow 1990) (Figure 1A). After an extensive survey in 1999, the species was found in the type location and another four nearby subpopulations located in an area of less than four km^2 (Figure 1B) (Oleas and Pitman 2003). One of the difficulties for finding new records of *P. brevifolia* is the small size of the plant.

Species Distribution Models (SDMs) can be used to identify areas where species might be potentially found (de Siqueira *et al.* 2009). The SDMs are a general set of simulations that forecast the ranges of species based on the relation between real field records and a set of environmental predictors (Guisan and Zimmermann 2000). The resulting model can be used to judge habitat suitability and hence predict the species' potential distribution, across large landscapes (Elith and Leathwick 2009), making them a valuable tool in biogeography and conservation biology. One of the challenges for modeling the distribution of poorly-known species is that the models generally require a minimum of 20 records (Hernandez *et al.* 2006; Wisz *et al.* 2008) or more (Feeley and Silman 2011) in order to produce accurate models. However, for rare species, even SDMs generated with less than the optimal number of records may still be valuable if they can suggest areas that may have suitable habitats, and thereby help direct collecting efforts (de Siqueira *et al.* 2009). Ideally, new collections can then be incorporated back into the SDM in an iterative process such that each new record helps to improve the accuracy of range predictions and increase the likelihood of finding new populations (Guisan *et al.* 2006).

We used the geo-referenced data of the five known locations of *P. brevifolia* to develop a preliminary distribution model for the species (Figure 1C). The model was generated through a maximum entropy approach as implemented in the program MAXENT ver. 3.3.3a (Phillips *et al.* 2006) with the default settings, which have been shown to provide good results (Syfert *et al.* 2013). The settings included: 10000 maximum number of background points, crossvalidated replicate run, 500 maximum iterations, 0,00001 convergence threshold. MAXENT is appropriate for presence/absence data (Newbold *et al.* 2010) and is one of the most commonly-used method to generate SDMs. MAXENT has shown superior prediction accuracy compared to other methods (Elith *et al.* 2006) and is less sensitive to sample size than other SDMs (Hernandez *et al.* 2006; Pearson *et al.* 2007; Hernandez *et al.* 2008; Wisz *et al.* 2008). For our model, we used 19 bioclimatic variables derived from monthly temperature and rainfall values (www.worldclim.org, spatial resolution of 30 arc

second or ~ 1 km^2) (Hijmans *et al.* 2005) as the underlying environmental surfaces. The bioclimatic variables include annual trends like mean annual temperature and annual precipitation, seasonality that takes into consideration the annual range of precipitation and temperature and extreme environmental conditions as temperature of the coldest and warmest month (Hijmans *et al.* 2005). The output of MAXENT is a continuous probability field which we transformed into a predicted suitable vs. unsuitable map based on thresholding with the cutoff set as the point of maximum training sensitivity plus specificity as suggested by Jiménez-Valverde *et al.* (2008). Maximum training sensitivity refers to minimize false negatives and specificity aims to reduce false positives. We assessed model performance with the receiver operating characteristic analysis (ROC) as the Area under the curve (AUC) as implemented in MAXENT (Phillips *et al.* 2006). Our model had an AUC value of 0.9995, which indicates a better than random model performance (Manel *et al.* 2001; Franklin, 2009).

In July–December 2009, we searched previously unexplored areas predicted as suitable for the occurrence of the species according to the SDM map. We need to point out finding new records of *P. brevifolia* is not an easy task. Each sterile individual aboveground is only one leave of 12 cm, which is difficult to distinguish, especially during the rainy season, when the vegetation is fuller. Furthermore, *P. brevifolia* habitat is located in one of the most deforested areas in Ecuador (Valencia *et al*. 1999), where agricultural practice includes setting up fire to clear the sites for agricultural purposes (N. Oleas, personal observations). Our species distribution model, based on the few available

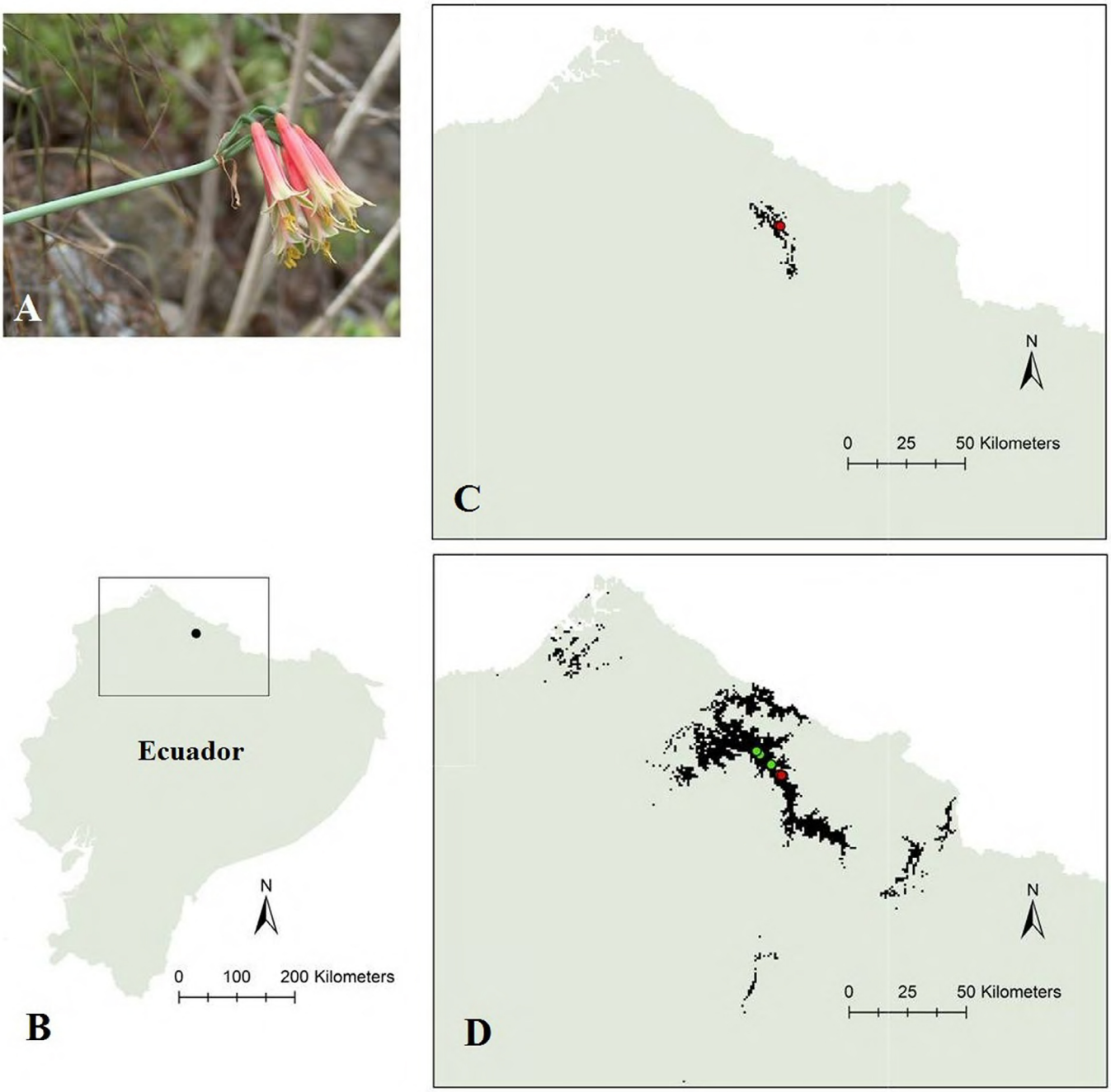

Figure 1. New records of *Phaedranassa brevifolia* (Amaryllidaceae). (A) *P. brevifolia* in the wild, (B) Map of Ecuador showing *P. brevifolia* distribution, (C) Species Distribution Model (black area) estimating potentially suitable areas based on the five previously known registers (red dots), (D) New records of the species (green dots) and Species Distribution Model of potentially suitable areas estimated with the combined dataset of new and old records (black area).

records before this study, suggested that the species could be found in adjacent areas to the west (Fig. 1C). The SDM increase the altitudinal range of *P. brevifolia* from 1200-1500 m to 800 to 1800 m. The habitat of *P. brevifolia* then would extend from Dry Montane Scrub habitats at the eastern portion of the geographic range to Wet Montane Scrub at the western part of the range. Within this area, we found three new locations expanding the known distribution of *P. brevifolia* to 16 km^2 to the west of the previously known range (Figure 1B). Botanical vouchers for these locations are deposited at the National Herbarium (QCNE) in Quito, Ecuador (Oleas # 1010, 1011, 1012). The conservation status of *P. brevifolia*, even with these new populations does not change. The previously reported five populations in Oleas and Pitman (2003) correspond to a very small area, with each record separated by less than two km^2. Because of this, they can be considered one location or at most two, so that the total number of locations for *P. brevifolia* after incorporating the new populations is still just five. It is likely that there are more populations of *P. brevifolia* in the wild, but additional fieldwork is needed to find new populations. Thus, *P. brevifolia* is Endangered (EN) under the IUCN criteria B1ab (iii); C2a (i) (IUCN, 2001). Figure 1D shows a SDM including the new records found at the field; information that can be used to located new records of the species in the future.

Our experience supports the potential application of SDMs as a tool to guide the search for new populations of rare species, even when models are based on limited sample sizes. Indeed, the use of SDMs has the potential to decrease time and effort to direct field efforts

Acknowledgments: We thank the Ministerio del Ambiente for collecting permit FLORA 04-09-IC-FLO-DNB/MA. We thank our fieldwork assistants: Teresa Gonzalez, Maria, Roberto and Luciana Melo, Patricio Vallejo, Susana and Patricio Oleas. This study was financed by the Dissertation Evidence Acquisition Fellowship at Florida International University and by a grant from The South Florida Chapter of the Explorers Club to N.O. We thank the suggestions of Augusto Francener Nogueira Gonzaga and one anonymous reviewer for their comments to improve this paper. This is contribution 278 from the Tropical Biology Program of FIU.

Literature Cited

de Siqueira, M.F., G. Durigan, P.J. de Marco and A.T. Peterson. 2009. Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *Journal for Nature Conservation* 17(1): 25–32 (doi: 10.1016/j.jnc.2008.11.001).

Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC..M. Overton, A.T. Peterson, S.J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz and N.E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129–151 (doi: 10.1111/j.2006.0906-7590.04596.x).

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

Feeley, J.K. and M.R. Silman. 2011. Keep collecting: accurate species distribution modeling requires more collections than previously thought. *Diversity and Distributions* 17(6): 1132–1140 (doi: 10.1111/j.1472-4642.2011.00813.x)

Guisan, A. and N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135(2–3): 147–186 (doi: 10.1016/S0304-3800(00)00354-9)

Guisan, A., O. Broennimann, R. Engler, M. Vust, N.G. Yoccoz, A. Lehmann and N.E. Zimmermann. 2006. Using niche-based models to improve the sampling of rare species. *Conservation Biology* 20(2): 501–511 (doi: 10.1111/j.1523-1739.2006.00354.x).

Franklin, J. 2009. *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge, New York. 338 pp.

Hernandez, P.A., C.H. Graham, L.L. Master and D.L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29(5): 773–785 (doi: 10.1111/j.0906-7590.2006.04700.x).

Hernandez, P.A., I. Franke, S.K. Herzog, V. Pacheco, L. Paniagua, H.L. Quintana, A. Soto, J.J. Swenson, C. Tovar, T.H. Valqui, J. Vargas, and B.E. Young. 2008. Predicting species distributions in poorly-studied landscapes. *Biodiversity and Conservation* 17(6): 1353–1366 (doi: 10.1007/s10531-007-9314-z).

Hijmans R.J., J.D. Cameron, J.L. Parra, P.G Jones and A. Jarvis. 2005. Very high resolution interpolated climate surface for global land areas. *International Journal of Climatology* 25(15): 1965–1978 (doi: 10.1002/joc.1276).

IUCN. 2001. *IUCN Red List Categories and Criteria: Version 3.1.* , Gland, Switzerland and Cambridge, UK: IUCN Species Survival Commission.30 pp.

Jiménez-Valverde, A., J.M. Lobo and J. Hortal. 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions* 14(6): 885–890 (doi: 10.1111/j.1472-4642.2008.00496.x).

Manel S., H.C. Williams and S.J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38(5): 921–931 (doi: 10.1046/j.1365-2664.2001.00647.x).

Meerow, A.W. 1987. New species of *Phaedranassa* and *Eucharis* (Amaryllidaceae). *Sida* 12(1): 29–49.

Meerow, A.W. 1990. Amaryllidaceae; pp. 1–54, in: G. Harling and L. Andersson (ed.). *Flora of Ecuador*. Volume 41. Goteborg, Stockholm, Quito: University of Gotenborg; Riksmuseum, Pontificia Universidad Católica del Ecuador.

Newbold, T., T. Reader, A. El-Gabbas, W. Berg, W.M. Shohdi, S. Zalat, S. Baha El Din and F. Gilbert. 2010. Testing the accuracy of species distribution models using species records from a new field survey. *Oikos* 119(8): 1326–1334 (doi: 10.1111/j.1600-0706.2009.18295.x).

Oleas, N. and N. Pitman. 2003. *Phaedranassa brevifolia*. In: IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2. Accessible at http://www.iucnredlist.org. Captured on 09 January 2012.

Pearson, R.G., C.J. Raxworthy, M. Nakamura and A.P. Townsend. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34(1): 102–117 (doi: 10.1111/j.1365-2699.2006.01594.x).

Phillips, S. J., R.P. Anderson and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3-4): 231–259 (doi: 10.1016/j.ecolmodel.2005.03.026).

Syfert, M.M., M. J. Smith and D.A. Coomes. 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PlosOne* 8(2): e55158 (doi: 10.1371/journal.pone.0055158).

Valencia, R., C. Cerón,W. Palacios and R. Sierra. 1999. *Las formaciones naturales de la sierra del Ecuador*; pp. 79–108, in: R. Sierra (ed.). *Propuesta preliminar de un sistema de clasificación de vegetación para el Ecuador continental*. Proyecto INEFAN/GEF-BIRF y EcoCiencia.

Wisz, M.S., R.J. Hijmans, J. Li, A.T. Peterson, C.H. Graham, A. Guisan and Group N.P.S.D.W. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14(5): 763–773 (doi: 10.1111/j.1472-4642.2008.00482.x).

Received: February 2012
Accepted: March 2014
Published online: July 2014
Editorial responsibility: Frederico Augusto Guimarães Guilherme

Marxan and relatives: Software for spatial conservation prioritization

Article · January 2011

Source: OAI

CITATIONS
963

READS
11,098

3 authors, including:

Hugh P Possingham
University of Queensland
1,114 PUBLICATIONS **87,572** CITATIONS

SEE PROFILE

Matthew Earl Watts
University of New England
60 PUBLICATIONS **8,836** CITATIONS

SEE PROFILE

CHAPTER 14

Marxan and Relatives: Software for Spatial Conservation Prioritization

Ian R. Ball, Hugh P. Possingham, and Matthew E. Watts

14.1 Introduction: The philosophy and history of Marxan

The original intent of Marxan and its predecessors was to solve a version of the minimum set reserve design problem (Cocks and Baird 1989; Chapter 3). In this problem, conservation targets are set for a number of biodiversity features (e.g. three populations of each species), and Marxan selects planning units that represent these targets for a minimum total cost, while allowing for more or less emphasis on spatially clustering the selected planning units. Marxan is now used to solve a range of spatial prioritization problems beyond the selection of reserves. The Marxan software is freely available and interacts with a variety of geographical information system (GIS) tools. It designs clumped reserve systems that make sense to a policy maker or planner.

One of the distinguishing features of Marxan is its use of simulated annealing to find multiple alternative good solutions to the impossibly huge minimum set problem (Kirkpatrick et al. 1983). Why did we choose simulated annealing as the basic optimization algorithm (see Chapter 5 for a discussion of algorithms)? We investigated a number of other methods for solving the minimum set problem including Genetic Algorithms and Branch and Bound methods (Taha 1992). We found that simulated annealing provided good answers quickly and was very flexible. Simulated annealing works well on problems of very different sizes. Non-linearities and other complexities can be added to the problem without having to change the formulation of the optimization algorithm (Cocklin 1989). In short simulated annealing proved to be relatively fast, relatively simple, and robust to changes in the size and type of the problem.

Marxan has evolved over more than a decade. The philosophy and ideas behind Marxan were developed in the original version of the software, titled Siman. It was written by Ian Ball as part of his PhD thesis under the supervision of Hugh Possingham (Ball 1996) at The University of Adelaide, Australia. SPEXAN, which stands for 'spatially explicit annealing' was the next stage in the development of Marxan (funded by Environment Australia). The Nature Conservancy (TNC) funded, through UC Santa Barbara, a project where Ian linked SPEXAN with ARCVIEW, resulting in SITES. Marxan is a modified version of SPEXAN, partially funded by the Great Barrier Reef Marine Park Authority in Australia and the US National Marine Fisheries Service. Marxan stands for 'marine reserve design using spatially explicit annealing'; however, it is just as applicable to terrestrial conservation planning problems.

The software is changing and developing in a variety of directions. Marxan with zones is the most recent development; it incorporates multiple zones and multiple costs (Watts et al. submitted). The blueprint for Marxan with zones was designed and implemented by the authors of this chapter in collaboration with their colleagues (Watts et al. submitted). Its development was partially supported by Ecotrust and the University of California. Other emerging versions and features of Marxan include:

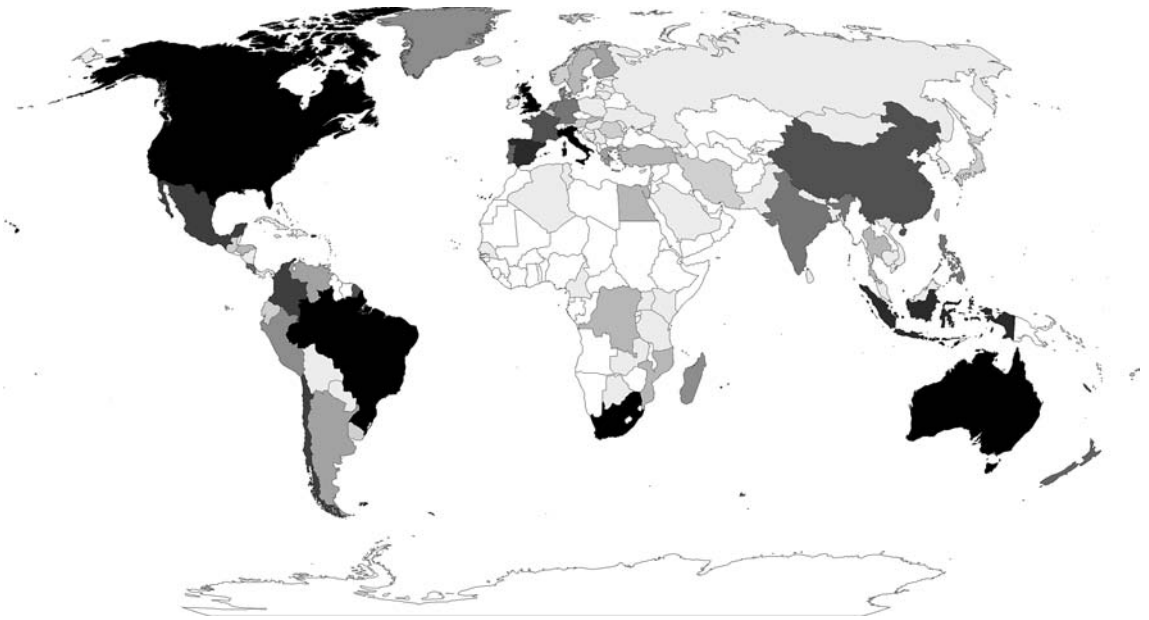

Figure 14.1 The world of Marxan: darker colours indicate more Marxan users, and white indicates no Marxan users. Marxan is very widely used and is possibly the most widely used conservation planning software in the world.

- introducing programming efficiencies to improve speed and the capacity to deal with large data sets,
- dealing with probabilistic treatments of threats,
- using cluster analysis to find good, but very different, Marxan solutions, and
- handling asymmetric connectivity for more sophisticated management of connectivity.

One of the most high profile and successful applications of Marxan was the rezoning of the Great Barrier Reef (Fernandes et al. 2005). TNC use Marxan for most of their spatial prioritization work in the USA and around the world. A partial list of Marxan applications – successful, purely academic, and pending – can be found on the Marxan web site (www.ecology.uq.edu.au/marxan). There are over 1,700 individuals in over 1,200 organizations from more than 100 countries that have used Marxan for marine and terrestrial conservation planning. Figure 14.1 shows the distribution of Marxan users around the world.

Marxan has some strengths as a tool for conservation Planning. First, Marxan uses spatially variable cost data to compute efficient solutions; each planning unit can be assigned a separate cost, which can be a complex combination of financial, opportunity, and social costs. Second, it uses a powerful optimization technique, simulated annealing, to generate near-optimal solutions. Third, it can generate many solutions quickly, allowing comprehensive data exploration and analysis. Fourth, Marxan solves a well-defined and explicit mathematical problem, as we shall see in the next section. Lastly, because of its widespread use, Marxan is well tested in a range of situations, and training programmes and free online support are provided to make sure that users get the best use out of the software.

Flexibility is probably the key to the success of the package. Due to the simplicity and power of its mathematical formulation, clever problem definition allows it to compute efficient solutions to a range of well-defined conservation planning problems.

14.2 What conservation planning problem does Marxan find solutions to?

Marxan finds good solutions to a mathematically well-specified problem which means that there is

no ambiguity about what the software is trying to achieve. The goal in Marxan is to minimize a combination of the cost of the reserve network and the boundary length of the entire system, whilst meeting a set of biodiversity targets. A larger boundary length is considered to be bad for several reasons which include: management costs, increased edge effects, and reduced connectivity. The mathematical problem to which Marxan finds good solutions is:

$$\text{minimize} \sum_{i}^{N_s} x_i c_i + b \sum_{i}^{N_s} \sum_{h}^{N_s} x_i (1 - x_h) cv_{ih} \qquad (14.1)$$

subject to the constraint that all the representation targets are met

$$\sum_{i}^{N_f} x_i r_{ij} \geq T_j \quad \forall\, j \qquad (14.2)$$

and x_i is either 0 or 1

$$x_i \in \{0,1\} \quad \forall\, i$$

where r_{ij} is the occurrence level of feature j in site i, c_i is the cost of site i, N_s is the number of sites, N_f is the number of features, and T_j is the target level for feature j. The control variable x_i has value 1 for sites selected for the reserve network, and value 0 for sites not selected (see Equation 3.1*a*).

The first term in Equation 14.1 is a penalty associated with the cost of all the sites (planning units) that are in the system. The second term in Equation 14.1 is a penalty associated with configuration (or shape). A reserve network with a smaller boundary length (or boundary cost) has a more compact configuration. The more fragmented the reserve system, the greater its boundary length. While it is typical for users to set cv_{ih} to be the length of the physical boundary between sites, and hence favour solutions that reduce the length of open boundaries, it is also possible to use this parameter more innovatively to improve the connectivity of a reserve system. In a more general sense, this second term of Equation 14.1 allows us to introduce a cost (or benefit) for including a particular site and any other site to which that particular site is 'connected'. It is another mechanism that allows the reserve network to be more than the sum of its parts, in addition to complementarity.

The parameter b in Equation 14.1 is the boundary multiplier which determines the cost of the reserve system (the first term in Equation 14.1) relative to the penalty for its spatial configuration (the second term in Equation 14.1). The matrix CV is the connectivity matrix with elements cv_{ih} which reflects the cost of the connection (such as a boundary) shared by planning units i and h. If one site is in the reserve system, and the other is not, then the connection cost must be paid. If both sites are out or in, the connection cost is not paid. In general we encourage users to think about assigning these connectivity costs according to the extra benefits of having any pair of sites in the reserve system. For example, cv_{ih} could be a quantitative measure of the flow of propagules from sites i to h where the sites may be separated by some distance. In this case, Marxan will try to find solutions that maximize the tendency for propagules generated in the reserve network to be retained within the reserve network. The use of a connectivity matrix allows connections between sites that are not adjacent, which means the concept of connectivity within the system can be quite complex (Figure 14.2).

The targets for a given conservation feature, T_j, can be an amount of that feature which is to be

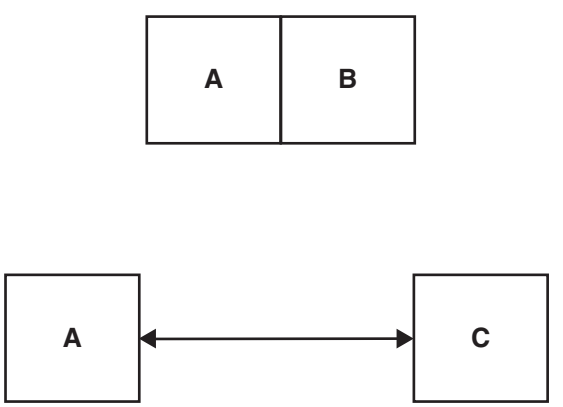

Figure 14.2 Two types of connectivity. Sites A and B have a shared boundary. Marxan typically represents this connection as a boundary between the sites. Planning units A and C do not have a shared boundary; however, they are connected due to propagule dispersal between the two sites. Marxan can represent this non-boundary connection in the connectivity matrix.

included within the reserve configuration, or it can be a number of occurrences of that feature within the system. For example we may wish to conserve 30 % of the original extent of every vegetation type or 1,000 adult females of every species. For smaller problems it can also include rules about the contiguity of occurrences of the conservation feature within the system or the separation of representations of the conservation feature within the system (www.uq.edu.au/marxan/). These software features can be used to ensure a minimum reserve size, or ensure that several representations of each conservation feature are adequately separated in space, the idea of risk spreading. Targets for contiguity and separation features can slowdown the operation of the software substantially if the number of planning units is in the high thousands or greater. If several separate representations of a specific conservation feature are required, e.g. we want a single vegetation type conserved in three separate areas on a continent, it may be best to simply treat that vegetation type as three separate conservation features.

Targets in Marxan are specific to the conservation features and not for other configuration characteristics, such as the minimum size of areas zoned for conservation, or the number of distinct areas zoned for conservation.

Equations 14.1 and 14.2 are the basic formulation for the problem. Marxan solves the problem by placing the objectives (Equation 14.1) and the constraints (Equation 14.2) together into an objective function by transforming the constraints into an additional penalty term. This means that a configuration of planning units which does not meet all of its conservation targets can still be given a value, which is of practical use in the annealing process.

A penalty is given to each conservation feature to either enforce or encourage it to be included in the final configuration to a certain level of representation. If the value of the penalty is less than 1, then a planning unit which only meets that conservation feature's target will tend not to be included, but one which meets other conservation targets at the same time will be included. If the value is greater than 1, Marxan should ensure that the target is met. The penalty is given as a proportion of the target that is not met multiplied by an internally derived penalty value for the conservation feature. This derived penalty indicates how costly it is to meet the features' target in a reserve specifically for that conservation feature. This is calculated with a greedy algorithm for simple conservation targets and using a variation on the iterative improvement method for more complex target requirements.

14.3 How does Marxan work exactly?

A number of different optimization techniques were considered to drive the optimization of the Marxan software. These included the exact optimization method of integer linear programming as well as approximating optimization methods such as genetic algorithms and simulated annealing (see Chapters 4 and 5). The serial selection heuristics which were traditionally applied to reserve selection, including the useful greedy algorithm, were also tested.

Integer programming can guarantee an optimal solution to a problem of the form given in Equations 14.1 and 14.2, and some have argued that these classical approaches are best for conservation planning problems (Underhill 1994; Rodrigues and Gaston 2002). However, we found two drawbacks of these more rigorous classical methods – first and foremost they failed to solve extremely large problems; and second, for practical and political reasons, finding the single best solution is not that useful in conservation planning.

The problem described in Equations 14.1 and 14.2 and those in Chapter 3 are in the mathematical category of NP-Complete. Optimal solution times for problems of this category rise faster than linearly with the size of the problem, where the problem size is the number of planning units and number of conservation features (each of which creates a constraint). Our initial explorations had long solution times for problems with 2,000 planning units and 250 conservation features and simple planning requirements (i.e. no spatial configuration requirements, Pressey et al. 1997). Integer programming was considered too slow to deal with problems of increasing complexity and size, where the spatial configuration is important and uncertainty in data and parameters meant there was a need for extensive sensitivity analysis. Increases in computational speed continue, but larger non-linear conservation

planning problems are still insoluble in reasonable time using classical operations research methods.

A criticism which is occasionally directed at the configurations produced by Marxan is that none of them are necessarily the optimal solution. However, it was considered more important to be able to rapidly produce a range of solutions which are near optimal to enable decision makers to negotiate and make choices amongst a good range of options. The data inputs into the problem are often imprecise, including estimates of the abundance of conservation features, or their occurrence pattern, the costs associated with planning unit, and the cost associated with the connectivity values between each pair of planning units. Also there is often an arbitrariness associated with each of the conservation targets. This means finding the single best system is somewhat academic (Pressey et al. 1996*b*). If possible it can be instructive to see how suboptimal Marxan solutions are for problems which can be solved using classical methods (Fischer and Church 2005).

Simulated Annealing is described in detail in Chapter 5. In order for it to work, one needs an objective function which can be evaluated explicitly for any valid configuration of planning units. By combining the constraint of Equation 14.2 as a penalty factor appearing in Equation 14.1, it is possible to give a value to any configuration of planning units whether they satisfy all the conservation targets or not. This is not just a convenience for the application of the optimization algorithm, but it is generally useful in allowing targets which are unreachable, or targets which only need to be partially fulfilled, or in evaluating the state of alternative configurations which do not meet all the targets.

In our implementation of simulated annealing, an initial potential solution is created either from a user-defined starting point or by randomly selecting some proportion of planning units (which might be all or none of them). New trial solutions are generated iteratively by randomly changing the status of a single planning unit and assessing the new configuration in terms of an improved or worsened objective function value. For example, if the planning unit was not previously part of the proposed configuration and its random inclusion improves the system, then it is kept in, if not it is removed. Similarly, if it was in the original solution and its random exclusion improves the system it is excluded, if not it remains included.

The heuristics that we tested tended to produce inferior results. Also, they produce only a single result, although they run quickly and are useful for a rough exploration of different scenarios. However, some of the Greedy Algorithms (described in Chapter 5) can also be robust and may prove a suitable alternative for some problems, especially those without complex spatial constraints (Pressey et al. 1997). Hence, they can be run within Marxan.

14.4 The Marxan user interface and output

Marxan is a command line program written in the C programming language. The version available for download is compiled and tested on the Microsoft Windows operating system, although it can also be compiled for other platforms (e.g. Linux). The data input files for Marxan are simply constructed text files which can be edited with any text editor. The program supports a number of types of simple formatting such as comma-delimited values or tab-delimited values which make it easier to edit in programs such as Microsoft Excel, OpendOffice.org Calc, Arc-View, S, or R. It can be used as a stand-alone program or as a plug-in for another decision support system (DSS) such as C-Plan, CLUZ, PANDA, or Vista. These DSS provide graphical outputs (maps and charts) and facilitate interpretation and manipulation of Marxan inputs and outputs. Hence they considerably enhance the ability of Marxan to help managers and stakeholders make decisions.

There are two standard Marxan outputs. First, the best solution file lists the reserve network with the lowest score from all the good reserve networks generated. Second, the summed solution file records the selection frequency for sites across all the good reserve networks generated.

In addition to being used to generate a single configuration, or a number of alternative configurations, the software can be used to produce nested configurations by iteratively running it with increasing representation targets. In this way it can produce core areas and then partially protected areas built around those core areas, a kind of zoning.

The selection frequency of a planning unit is a measure of how important it is to get that planning unit to meet all the targets. It is calculated by the fraction of the alternative solutions, derived from the same data and targets, in which a planning unit is selected. The selection frequency map indicates which areas are more often included in configurations and which are not. This is often used as a surrogate for irreplaceability (Ferrier et al. 2000). The planning units are infrequently selected if there are a range of equally good alternatives. If they are truly irreplaceable, then they will appear in every solution. Of particular interest are those planning units which are frequently included but have a high cost associated with them. Highlighting these can focus the attention of the user on areas where hard decisions need to be made.

Marxan has been constructed for use within a range of different graphical front-ends. First, GIS software changes a lot and different people use different packages. For example CLUZ is a front end for Marxan that runs on ArcView 3.X, a widely used commercial GIS package that will eventually be replaced by other GIIS software (Smith 2004). Second, this method encourages other programmers to develop front ends for Marxan that can be tailored to the needs of their organization.

Detailed information about how to use Marxan can be found in a new user manual (Game and Grantham 2008) while solid advice on tricks and pitfalls can be found in the new Marxan Good Practices Handbook (http://www.uq.edu.au/marxan/).

14.5 The Great Barrier Reef: An example of how Marxan has been used

Marxan was used as one of the decision support tools for rezoning the Great Barrier Reef, Australia. The Great Barrier Reef was rezoned on July 1 2004, and it represents the largest successful, and most complex, real-world application of systematic conservation planning principles (Fernandes et al. 2005). The rezoning took the Great Barrier Reef Marine Park Authority (GBRMPA) a decade to complete, highlighting the social and political complexities of conservation planning at this huge scale. Here, we give a very brief overview of how Marxan was used, the kinds of principles and data that underpinned its application, and some of the pros and cons of its application in this situation.

The Great Barrier Reef is a huge area off the north-east coast of Australia. It is a multi-use marine park that before 2004 had less than 5 % of its extent in no-take (no-fishing) zones. The rezoning raised the percentage that was in no-take zones to over 33 %. The target for each feature type was 20 %, and this was met for all but a couple of feature types. The use of Marxan for this exercise was interesting for a number of ecological and technical reasons:

- The region was divided into over 17,000 planning units where each reef, and a buffer around each reef, was a planning unit. The area between reefs was divided into hexagons of around 12 km^2.
- The cost of each planning units was derived from a linear combination of commercial fishing interests, recreational fishing interests, indigenous interests, and interests of people who would prefer particular places to be inside the reserve system (a negative cost).
- The most important biodiversity feature layer was the 'bioregions'. The whole system was divided into 70 bioregional types. These bioregions were derived from extensive statistical modelling and expert analysis of ecological and biophysical data. Targets of 20 %, and higher for rare reef types, were set for each feature.
- Reserves were preferentially set to be bigger than 400 km^2 where possible and each bioregion was meant to be represented in at least three well-separated reserves of at least this size. The minimum reserve size and minimum separation facilities of Marxan were used to achieve this.

It is instructive to note that GBRMPA initially ignored social and economic information and tried to get Marxan to find a biologically 'pure' conservation plan, a plan with only biological information. This is frequently done, or aspired to, but is fraught with problems. First, area is usually assumed to be cost, so that the problems of economic and social considerations were ignored, and the area conserved was simply minimized while meeting all conservation targets. This failed to deliver a system that was socially and economically viable.

Second, Marxan can be very indecisive if there are many planning units with roughly equal costs and equal amounts of different conservation features. It produced selection frequency outputs where most planning units have a selection frequency of around 20 % (the target set by GBRMPA) which is not very useful as decision support for planning.

While Marxan contributed to the selection of no-take areas, the final zoning places each part of the region into one of seven zones. Our new software, Marxan with zones, would be ideally suited to similar problems of marine park zoning.

14.6 Marxan applied to the Australian continent

Marxan has been used to do a conservation prioritization for the entire Australian continent (Carwardine et al. 2008; Klein et al. 2009 in press). This application of Marxan involved a few innovations that we would like to highlight.

First, the planning units used for the whole continent were sub-catchments (Stein 2005, 2006). By using sub-catchments (n = 62,630) and the knowledge of which sub-catchments were connected by the flow of water (e.g. the share the same river or stream), priority reserve networks were preferentially grouped along waterways, as illustrated in Figure 14.3. This has been perceived to be a need for many people interested in freshwater conservation planning and incorporating ecological processes into conservation planning (Chapters 2 and 7).

Second, a large number and array of conservation features were included:

- 1,763 vegetation types, identified by intersecting 62 native vegetation subgroups (The National Vegetation Information System 2001) with 85 bioregions (Australian Government Version 6.1),
- 563 bird species distributions modelled using Birds Australia location data (Carwardine et al. 2006),
- 1,222 threatened species distributions (Commonwealth of Australia 1999), and
- 151 environmental domains, represented by an environmental classification based on a set of key climatic, topographic, and substrate conditions that characterize the landscape (Mackey et al. 2008).

Third, with the aim of protecting a selection of ecological and evolutionary processes that maintain and sustain biodiversity, two types of refugia were targeted:

- Drought refugia: Areas which are critical for the persistence of many species during harsh climatic conditions in arid Australia. We identified drought refugia, areas of high and regular herbage production in arid and semi-arid Australia, using estimates of gross primary productivity modelled from high-resolution satellite data, and spatially interpolated climate data (Figure 2.3).
- Evolutionary refugia: Areas which are important for maintaining and generating unique biota during long-term climatic changes. Evolutionary refugia were identified by experts as areas in Australia important for generating species (Morton et al. 1995).

Fourth, areas were prioritized for two different actions, land acquisition for the national reserve system and stewardship arrangements on private land. Many users think that software like Marxan is just about building a network of protected areas. However, if we assume the cost of a planning unit is the cost of entering into a stewardship or other conservation arrangement on private land, then it can be used to prioritize places for these sorts of conservation action.

Finally large intact areas in good condition were preferentially targeted rather than fragmented landscapes (Klein et al. 2009 in press). Using data on the condition (wilderness quality) of each sub-catchment, we either choose to meet targets preferentially in planning units in good condition or choose large clusters where adjacent sites in good condition were well connected.

The aim of this study was to guide the Australian government in identifying spatially explicit priorities for biodiversity conservation using fundamental concepts of spatial conservation prioritization (Chapter 2) at a continental scale. An example of a final prioritization, using a target of 30 % of each feature and with a bias towards large intact landscapes is shown in Figure 14.3.

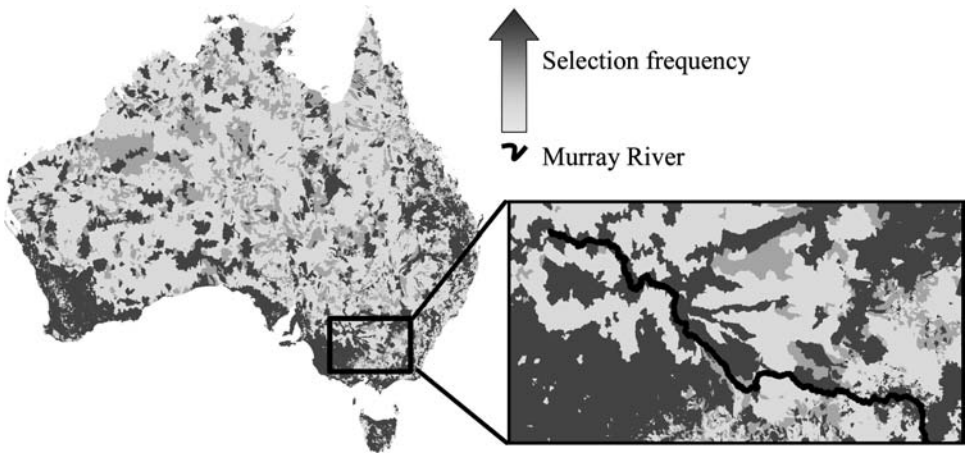

Figure 14.3 A majority of the sub-catchments along the Murray River were prioritized for conservation investment when the connectivity of sub-catchments was considered. Darker colours indicate a higher selection frequency or priority for investment.

14.7 Applicability and limitations of Marxan

Marxan runs from a command line interface and therefore has no graphical user interface of its own. This is a strength of Marxan as its architecture allows use by any DSS as a plug-in, and it does not become outdated from being tied to a particular DSS, GIS, or operating system.

There is inherent complexity in setting up input files and interpreting output files, which are both a strength and a weakness. It is a strength as it requires users to think through their problem definition in detail to manage these complexities, so the user is more likely to understand the problem they are solving. It is a weakness as the management of files is not a simple point and click exercise, leading to many users becoming frustrated.

Indeed most problems using Marxan arise when input files do not conform to the standard data definition. These problems can be avoided by careful application of the input-file specifications laid out in the Marxan user, or by using one of the many DSSs that support the building of Marxan input files. Improved error checking and data diagnostics would reduce this limitation of the software.

Marxan does not readily deal with the thorny issue of conservation adequacy. Probably the most sensible approach is to set targets for key species using outputs from viability analyses (see Chapters 8 and 9). In more recent applications surrogates for conservation adequacy have been used where ecosystem processes, vegetation condition, and evolutionary refugia are included in the formulation of the problem.

There is the challenge of including even more complex problems like ecological processes, social dynamics, and continuous benefit functions (Chapter 2). The desire to include more functionality and improve the software should always be tempered by consideration of whether important complexities are better handled within Marxan, or outside of Marxan in a DSS. This is particularly true when we consider constructing the cost of each planning unit – a parameter that may be built up from a great deal of other information as in the Great Barrier Reef rezoning case study.

Marxan is designed to be open-ended with regard to the size of the problems to which it can be applied. Users have successfully run problems with over 2,000,000 planning units. The software is designed to be efficient in memory usage so that the number of conservation feature records is one of the key memory components. A sparse matrix of occurrences of conservation features can require little memory even if there are very large numbers of planning units or conservation features. Typically the factor which slows the software the most is the number of planning

units – there tends to be a limit to the number of conservation features which can be usefully considered, but the number of planning units can always be increased by finer subdivisions of space. The ideal number of iterations to use in the system is largely a function of the number of planning units so that increasing them increases both the memory requirements and running time of the software.

14.8 Common misconceptions and misapplications of Marxan

In this section, we describe some of the misconceptions about Marxan and how it can be applied inappropriately. Some of these misunderstandings are common to all conservation planning software.

First we discuss the 'black box' issue. Some users ask questions about why different planning units were chosen and how the algorithm goes about selecting planning units for inclusion into the configuration. Because Marxan does not serially select planning units the way a designer might do it by hand, it appears that the selection mechanism is a mysterious 'black box'.

This concern arises partially because of a lack of understanding of how the simulated annealing algorithm works, and mathematical optimization in general. Its operations are viewed as complex and mysterious, and many users feel uncomfortable with the technical description of the algorithm as given in the manual and supporting papers. There are a variety of descriptions of how simulated annealing works associated with Marxan and CLUZ, including an online game.

The other problem comes from focussing on the wrong part of the system. Historically the focus of systematic planning was on the process of serially selecting planning units to build a reserve network and software was designed to replicate the manual serial selection of planning units. However, the key to understanding the system is to understand the problem formulation: the objective function and the choices which are made in terms of setting targets, differentially weighting different conservation features, planning units, connections or boundary lengths, and the relative weightings of the cost term and the connectivity term. If we understand the objective function that is being optimized, then we know what problem is being solved (see Section 14.2). The optimization method can then well be a black box without the user being disadvantaged; in the same way that the user of a calculator need not know the internal method by which multiplications are computed. Indeed the process of serially selecting planning units can be quite misleading, because it assumes myopic algorithms are useful, and it tends to ignore the fact that the whole reserve system has a value, not individual planning units – in reserve system design the whole is more than the sum of the parts.

It is still useful for the user to understand the relative strength and weakness of the optimization methods (covered in Chapters 4 and 5) as well as having a practical understanding of setting different values for the parameters that control the optimization method.

Some users will only apply the in-built heuristics such as the richness, or greedy, algorithm. This can produce quick results and is a fast and useful way of exploring trial solutions. Because it is implemented as a greedy algorithm, it can both add and remove planning units from a configuration. However, even though these heuristics, and a number of the other serial heuristics, are included in Marxan, it is a mistake to use these as a complete substitute for simulated annealing. The serial heuristics produce solutions which are inferior to the simulated annealing method with the greedy algorithm for example being driven by the distribution of the planning units with the highest density of conservation feature representation. An ordering of planning unit selection and deletion is produced which can be useful in the planning process, but it misses out on the power of simulated annealing to find solutions to the more complex and interesting nonlinear versions of this problem.

The production of alternative configurations of conservation areas, particularly through the selection frequency measure, is one of the strengths of Marxan, but some have complained that this is an indication that the solutions are not optimum. Although the authors would recommend against it, the user has the option of saving only the best configuration from an arbitrary number of repeats or of producing only a single solution using a much longer annealing trial. At the end, the main thing is that it produces significantly better solutions than could be

obtained without using a computational algorithm (Klein et al. 2008*b*; Leathwick et al. 2008*c*).

One use of Marxan on which there is divided opinion is in the creation of cost layers. One way to set the cost associated with each planning unit in the configuration is to combine the costs from a number of different sources into a single value by differentially weighting these sources. For example there might be abstract social costs for including planning units in the network, actual setup costs, and opportunity costs to a variety of users of the system such as recreational and commercial fishermen. If these are combined into a single cost then a weighting for each type of cost must be determined, and it is not always clear how this weighting should be determined in an objective manner.

The alternative, which some of the developers recommend, is to produce alternative configurations each using a separate cost layer. This can help focus attention in the design process on those areas which are most different in their impact on different user groups.

Another complaint that is often directed at Marxan is that if the problem definition is not correct then invalid results will be obtained. This is, of course, as obvious as it is important when considering the results from Marxan or any other systematic planning software. The alternative is to use some informal method in which the problem definition is never made explicit, which in our view is poor planning. Good problem definition lies at the heart of good conservation (Possingham et al. 2001).

14.9 The future of Marxan

Marxan is continually changing as we work with partners to provide new functions and more flexibility. In this section, we briefly introduce three of the most substantive new functionalities for Marxan.

'Marxan with Zones' is a decision support tool developed under contract to Ecotrust. It was developed to support the design of marine protected areas along California's coast as part of California's Marine Life Protection Act (Klein et al. 2008*b*). Marxan with Zones is an extension of the Marxan software, developed by Ian Ball, Matthew Watts, and Hugh Possingham. It has the same functionality as Marxan but is also able to set priorities for multiple zones (i.e. marine protected areas of various protection levels) and incorporate multiple costs into a systematic design framework. The purpose of Marxan with Zones is to assign each planning unit in a study region to a particular zone while meeting a number of biodiversity targets at a minimum total cost. This software clearly has many advantages and is ideally suited to many current marine zoning problems as well as terrestrial problems where more than one land-use change is under consideration. A preliminary output from Marxan with zones applied to the area around Rottnest Island is shown in Figure 14.4.

We will not discuss Marxan with Zones in detail here but emphasize two important considerations. First, this software requires a lot of additional data – for example one must specify the cost of placing a planning unit into any one of the different zones, the benefit to each biodiversity feature of being placed into a particular zone, and the relative merits of having each zone type juxtaposed with another zone type. Figure 14.1 shows that clever use of the latter concept enables us to create a zoning map where highly protected areas can be buffered by less protected areas. Second, Marxan with Zones has a strong focus on the interaction between different zones. Instead of a simple boundary between planning units which are in the configuration and those which are not, every planning unit can be in one of many different types of zones, some of which have a variety of conservation uses and others which are set aside for specifically non-conservation uses. The boundaries or connectivity between each planning unit is even more important than the simpler case and the matrix of different types of zone connectivity adds more control over the system as well as more complexity.

In Section 3.5, we discussed the fact that we are often unsure exactly whether or not a feature is present in any particular planning unit, and if so, how much of it is there. This uncertainty may arise from the habitat modelling package or population survey used to create feature distribution maps or the possibility that the feature has or may appear or disappear from the planning unit (through habitat change, catastrophes, or population dynamics). Indeed, there are many reasons why we may wish to assume that the representation matrix (Chapter 3) is filled with probabilities and not absolute numbers. A new version of Marxan, currently undergoing testing, will be able to solve problems where there is uncertainty about feature presence. In particular, it

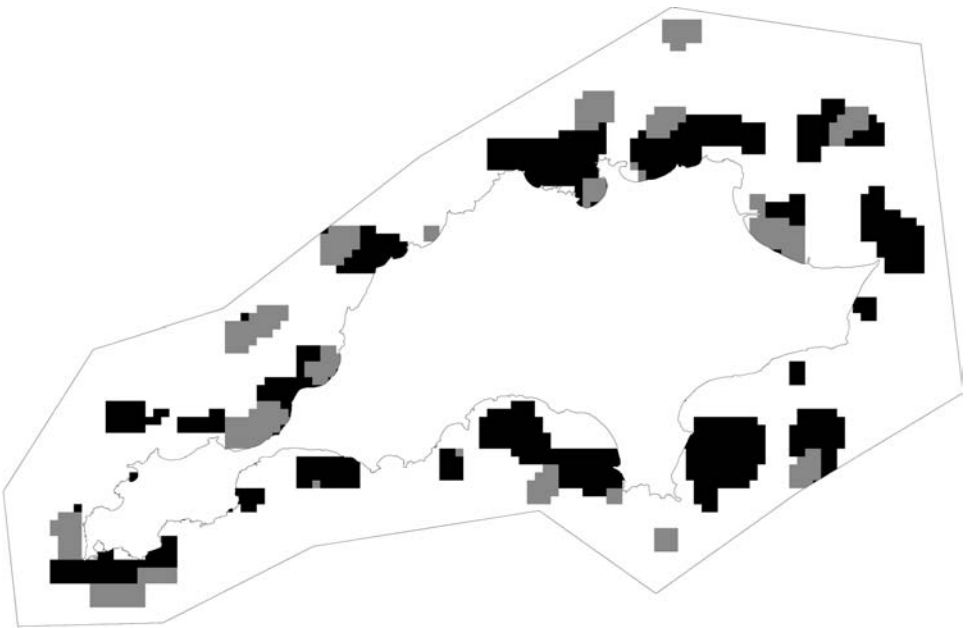

Figure 14.4 A zoning plan of Rottnest Island (Western Australia) created with Marxan with Zones showing two kinds of zone. The black zone is a proposed no access zone, grey is a proposed recreation zone with no-fishing, and white is a proposed recreational fishing zone. You can see how the grey zone tends to buffer the black zone, creating a recreational no-fishing zone around the no-access zone. This is generated by placing a high cost on boundaries that connect the most incompatible zones – no-access and recreational fishing.

will allow users to set not just a feature target, T_j, but also the confidence with which we want to meet that target. For example, we may wish to be 90 % sure that there are 2,000 ha of seagrass bed in our marine reserve. This idea is to build reserve networks that are robust to uncertainties pertaining to catastrophic events, and changes in feature distribution through disturbance and succession (Game et al. 2008; Drechsler et al. 2009 in press).

We have implemented an automated cluster analysis of solutions to explore the many different solutions generated by Marxan. Typically, the best solution (which is simply the solution with the lowest cost) and the summed solution (the selection frequency of planning units over a number of solutions) are the only Marxan outputs used. We use the R statistics package to perform hierarchical cluster analysis and hence a selection of good but different solutions. This is particularly useful where planners want a range of credible options to consider.

New front ends are being developed for Marxan and new functionality is being added to existing front ends to make them easier to use. A version of Marxan that runs natively in the Linux operating system has recently been developed and tested. We encourage other researchers to cooperate with us in the development of this free tool – its development has always been very much a cooperative and international venture.

Acknowledgements

Marxan has developed over many years with advice and input from many people – too many to name. However we would like to particularly thank: Carissa Klein, Josie Carwardine, Jeff Ardron, Bob Smith, Natalie Ban, Doug Fischer, Sandy Andelman, Kerrie Wilson, Romola Stewart, Lindsay Kirchner, and Dan Segan.

Marxan would not have existed without some financial support. While it is free software, and its support is funded through the University of Queensland and the spare time of generous individuals – several groups have paid for important developments and these include: the Environment Department (Australia), the Great Barrier Reef Marine Park Authority, Ecotrust (USA), the Nature Conservancy, and the US National Marine Fisheries Service.

Biological Conservation 214 (2017) 168–175

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Designing protected area networks that translate international conservation commitments into national action

Jake E. Bicknell[a,*], Murray B. Collins[b,k], Rob S.A. Pickles[c,d], Niall P. McCann[e], Curtis R. Bernard[f], Damian J. Fernandes[g], Mark G.R. Miller[h], Samantha M. James[i], Aiesha U. Williams[j], Matthew J. Struebig[a], Zoe G. Davies[a], Robert J. Smith[a]

[a] Durrell Institute of Conservation & Ecology (DICE), School of Anthropology and Conservation, University of Kent, Canterbury CT2 7NR, UK
[b] School of Geosciences, University of Edinburgh, Edinburgh EH8 9XP, UK
[c] Panthera, 8 West 40th Street, 18th Floor, New York, NY 10018, USA
[d] Department of Biology, Trent University, Peterborough, Ontario K9J 7B8, Canada
[e] Cardiff School of Biosciences, Museum Avenue, Cardiff CF10 3AX, UK
[f] Conservation International Guyana, 91 Robb Street, Bourda, Georgetown, Guyana
[g] Guyana Protected Areas Commission, National Park, Thomas Lands, Georgetown, Guyana
[h] College of Marine and Environmental Sciences, James Cook University, Cairns, Queensland 4870, Australia
[i] Iwokrama International Centre for Rainforest Conservation and Development, 77 High Street, Georgetown, Guyana
[j] World Wildlife Fund Guianas, 285 Irving Street, Queenstown, Georgetown, Guyana
[k] Space Intelligence, Edinburgh Centre for Carbon Innovation, High School Yards, Edinburgh EH1 1LZ, UK

ARTICLE INFO

Keywords:
Convention on biological diversity
Local communities
Marxan
REDD+
Systematic conservation planning
Tropical forest conservation

ABSTRACT

Most countries have committed to protect 17% of their terrestrial area by 2020 through Aichi Target 11 of the Convention on Biological Diversity, with a focus on protecting areas of particular importance for biodiversity. This means national-scale spatial conservation prioritisations are needed to help meet this target and guide broader conservation and land-use policy development. However, to ensure these assessments are adopted by policy makers, they must also consider national priorities. This situation is exemplified by Guyana, a corner of Amazonia that couples high biodiversity with low economic development. In recent years activities that threaten biodiversity conservation have increased, and consequently, protected areas are evermore critical to achieving the Aichi targets. Here we undertake a cost-effective approach to protected area planning in Guyana that accounts for in-country conditions. To do this we conducted a stakeholder-led spatial conservation prioritisation based on meeting targets for 17 vegetation types and 329 vertebrate species, while minimising opportunity costs for forestry, mining, agriculture and urbanisation. Our analysis identifies 3 million ha of priority areas for conservation, helping inform government plans to double the current protected area network from 8.5 to 17%. As part of this, we also develop a new technique to prioritise engagement with local communities whose lands are identified as important to conservation. Our study both provides a scientifically robust, politically acceptable protected area expansion strategy for Guyana, and illustrates the importance of conservation planning at the country-scale to translate international commitments into national action.

1. Introduction

Protected areas form the cornerstone of global biodiversity conservation efforts, and today there are > 200,000 terrestrial protected areas worldwide (Bruner et al., 2001; Chape et al., 2005; UNEP-WCMC and IUCN, 2016). In recognition of this, signatories to the United Nations Convention on Biological Diversity (CBD) have committed through Aichi Target 11 to ensure that 17% of the terrestrial realm is protected by 2020, with a focus on establishing protected areas and other effective area-based conservation measures (OECMs). Implementing this commitment involved each country setting a national target, with most adopting 17%. However, with less than three years until 2020, only 14.8% of global land area is protected, representing a total shortfall of 3.1 million km^2 (UNEP-WCMC and IUCN, 2016), an area nearly the size of India. This shortfall is because over half of countries are yet to reach their national targets (World Bank, 2017), and while between 1990 and 2012 the area of the global conservation estate grew rapidly, progress has since plateaued (UNEP-WCMC and

* Corresponding author.
E-mail address: J.E.Bicknell@kent.ac.uk (J.E. Bicknell).

http://dx.doi.org/10.1016/j.biocon.2017.08.024
Received 25 January 2017; Received in revised form 20 July 2017; Accepted 14 August 2017
Available online 21 August 2017

IUCN, 2016). Consequently, for countries falling short of their target, up-to-date protected area expansion plans are needed.

Signatories to the CBD have recognised that for protected areas to be effective, they must be well-connected, ecologically representative and conserve areas of particular importance for biodiversity (CBD, 2010). This is against the backdrop that many existing protected areas are biased towards locations that are less important for biodiversity and/or on remote and economically unproductive land (Brooks, 2014; Joppa and Pfaff, 2009; Venter et al., 2017). Therefore, the Aichi targets have created an opportunity for the conservation science community to guide protected area expansion, as there is a real need to develop evidence-based plans that prioritise biodiversity (Watson et al., 2016).

Several global spatial conservation prioritisations have been conducted (e.g. Butchart et al., 2015; Larsen et al., 2011; Pollock et al., 2017; Venter et al., 2014), and these provide broad insights into the optimal locations for future protection at the international scale. However, as Aichi Target 11 is implemented at the national-level (CBD, 2010), and as this is the scale most relevant for land-use policy development and delivery, national-scale spatial conservation prioritisations are needed. To undertake these, under the CBD, government agencies must develop National Biodiversity Strategy and Action Plans (NBSAPs), which, where necessary, include a roadmap to achieving the Aichi targets, including "…integrating biodiversity into spatial planning exercises through the mapping of biodiversity ecosystem services and through systematic conservation planning" (CBD, 2010). Systematic conservation planning is one of the most transparent and robust methods for informing spatial planning, as it aims to maximise conservation benefits while minimising impacts on other stakeholders (Margules and Pressey, 2000). In addition to global analyses, systematic conservation planning has been extensively used at the local, regional and landscape level (e.g. Smith et al., 2008; Venter et al., 2013). However, despite national CBD targets, it is less commonly applied at the national-level (Di Minin et al., 2017), even though this is generally the scale most relevant for the government agencies charged with delivering CBD targets.

To illustrate the benefits of such country-wide analyses, here we describe a national-scale systematic conservation planning process for Guyana, which was led by the main government agency for protected areas in collaboration with a range of stakeholders and conservation scientists. Our plan sought to identify priority areas for protected area network development in Guyana, to adequately represent biodiversity while accounting for other land-uses. Guyana forms part of Amazonia and combines economic poverty with some of the highest global levels of biodiversity (Jenkins et al., 2013), and lowest deforestation rates (Hansen et al., 2013). Over 80% of the land area is covered with tropical forest. However, as in many parts of South America, deforestation rates have risen over the last decade, primarily as a result of gold mining (Fig. 1; Howard et al., 2011; Laing, 2015). This was partly because forests produced little government revenue compared with mining. This situation changed in 2009, when Norway committed up to $250 million to Guyana over an initial five-year period for Reducing Emissions from Deforestation and forest Degradation (REDD+) (Gutman and Aguilar-Amuchastegui, 2012). The expectation was that the funds can shift the economy away from a reliance on resource extraction towards a more sustainable and low-carbon model (Office of the President, 2013). Therefore, under the REDD+ agreement, Guyana committed to fulfilling its CBD obligations, through the implementation of a national conservation planning process. Both the REDD+ agreement and Aichi Target 11 stipulate that protected areas should be established and managed in close collaboration with indigenous and local communities (CBD, 2010; Gutman and Aguilar-Amuchastegui, 2012; Office of the President, 2013), and this is highly relevant in Guyana because community lands cover c. 15% of the country (Fig. 1), most of which are owned by indigenous Amerindians.

The existing protected areas in Guyana were not selected systematically, representing just 8.5% of the land area, of which 3.1% is a community conservation area. In 2016, the President of Guyana pledged an additional 2 million ha of protected area would be established across the country, thereby addressing both the shortfall in the 17% Aichi Target, and making an important contribution to the reduction in deforestation required to receive performance-related REDD + payments. To guide this process, we formed a group of stakeholders from Government of Guyana agencies, academia, and Non-Governmental Organisations, and used a systematic conservation planning approach. We identify priority areas to achieve conservation targets for 329 species and 17 vegetation types, while minimising opportunity costs (i.e. the choice of the best lower cost alternative) from the forestry, mining and agricultural industries (Margules and Pressey, 2000; Venter et al., 2013). Given the importance of local communities, we also developed a method to identify the most important community lands for meeting conservation targets. This provides a technique to help prioritise the engagement process for free prior and informed consent during the creation of new protected areas. Our study serves as a benchmark for countries looking to undertake national-scale spatial conservation prioritisations to expand their protected area networks.

2. Methods

The study was initiated by the Government of Guyana's Protected Areas Commission in collaboration with academics who joint-led the systematic conservation planning process. Our team quickly grew to consist of representatives from all of the non-governmental conservation organisations in Guyana, including Conservation International, WWF, and the Iwokrama International Centre for Conservation and Development. We consulted with stakeholders and policy makers during every stage of the planning process to ensure the spatial prioritisation remained relevant (Smith et al., 2009). This began with a workshop formed of all government agencies and stakeholders responsible for forestry, mining, natural resource management, land-use planning, environmental protection, and indigenous peoples as well as our study team, and initial feedback was given on preliminary analyses. Recommendations from these consultations were that the conservation prioritisation should: i) focus on Guyana's habitats and biodiversity, explicitly including threatened species; ii) incorporate opportunity costs; and iii) consider the role of community lands. The stakeholders also agreed that due to data availability, species distribution maps would need to be developed, and that the planning analysis should use Marxan, a software package designed to identify sets of priority areas that meet quantitative targets for specified conservation features, while minimising costs and maintaining connectivity (Ball et al., 2009). All stakeholders were kept up-to-date and remained involved as the spatial conservation prioritisation was developed and completed.

2.1. Habitat and species distributions

The conservation features we used in the analysis were 17 vegetation types, as classified in the Guyana national vegetation map (ter Steege, 2001), and all of Guyana's vertebrates for which range maps were available or could be developed. Faunal communities in many parts of Guyana have not been extensively studied, so we generated species distribution models to fill these gaps. We assessed data availability for all the c. 1000 terrestrial vertebrate species known to occur in Guyana and produced a species distribution model if ≥ 15 spatially referenced records had been collected. Species locality data were obtained from the Global Biodiversity Information Facility and published studies and rapid biodiversity assessments (Appendix Table A1). To increase the sample size for each species and, therefore, the reliability of our models (Elith et al., 2006; Hernandez et al., 2006), we widened

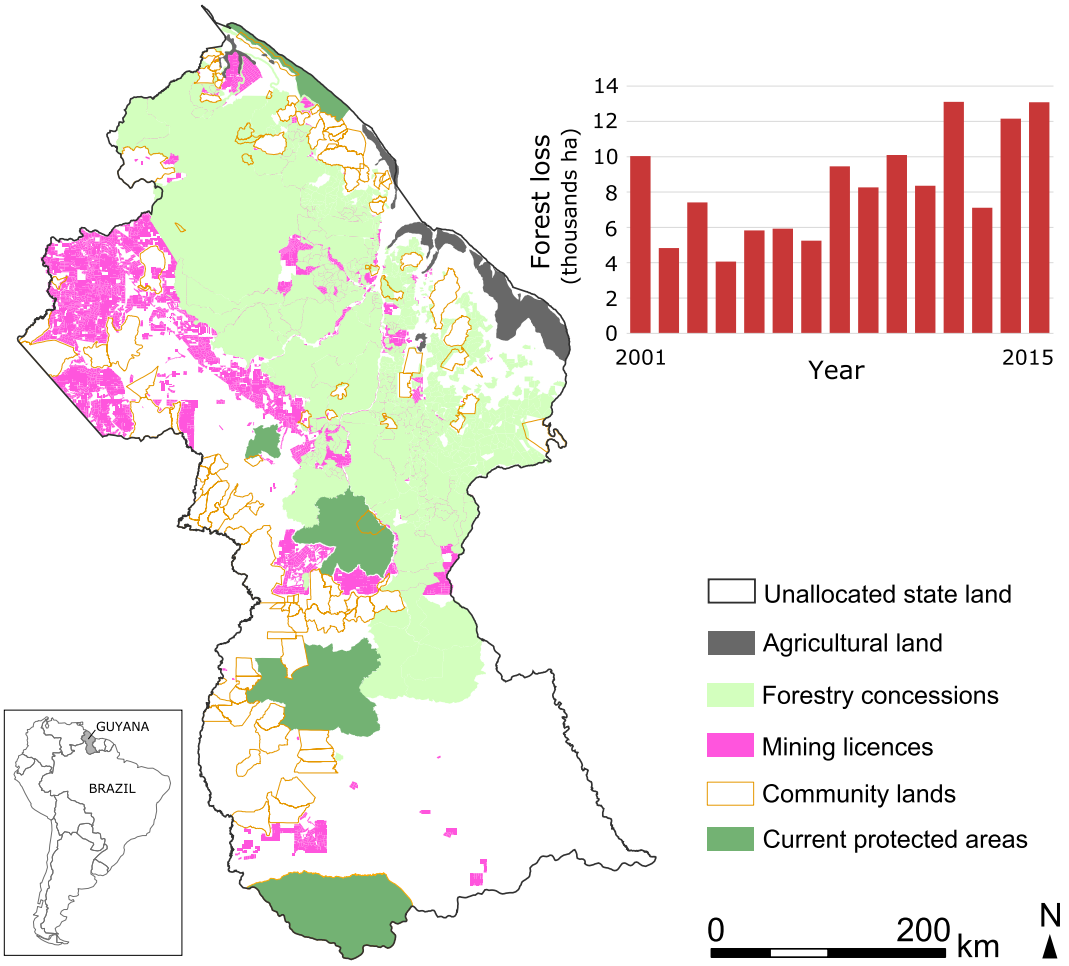

Fig. 1. Map shows current land-use in Guyana. Forestry concessions are indicated on top of mineral licenses, but many forestry areas also have mineral licenses granted within them. Forestry concessions and mineral licenses may have been previously exploited, in active use, or allocated for future extraction (data from various governmental departments in Guyana). Graph shows forest loss for the period 2001–2015 (Hansen et al., 2013). Inset shows the location of Guyana in South America.

the geographic area over which we obtained locality data to include a 200 km buffer around Guyana. We generated species distribution models using MaxEnt (Phillips et al., 2006), a widely used modelling package. MaxEnt has often been used to provide inputs for spatial conservation prioritisations, particularly in regions where data are sparse (Esselman and Allan, 2011; Moore et al., 2016). Our inputs to MaxEnt were species locality, and topographic and meteorological data (Table A3). Survey data are usually spatially biased and can cause inaccuracies if this bias is not properly accounted for when inputted into distribution modelling (Phillips et al., 2009). Therefore, to improve model predictions and reduce errors associated with survey effort bias (Syfert et al., 2013), we constructed species-specific bias grids in R (R Core Team, 2015). We ran the models by splitting each species dataset into separate training (70%) and testing (30%) components and measured performance using the Area Under Curve (AUC), based on 10 repetitions per species. Species were dropped from further analyses if their model had an AUC score of < 0.5 when predicting to the test dataset (data not used in model construction), making it a more robust measure than predicting to the training dataset (Fielding and Bell, 1997). We then split the resulting probability maps into binary presence/absence using the MinROCdist threshold, which has been shown to yield the most appropriate value based on predictive performance (Liu et al., 2005).

We also included threatened species (IUCN Red List: CR, EN, VU) even if we could not model their distributions. This was not only because of stakeholder feedback but also because omitting these species can result in a failure to recognise the irreplaceability of some sites (Platts et al., 2014). We did this by using their IUCN distribution maps. However, many of the maps were not suitable (e.g. the resolution was too coarse), so that only 13 of the 31 CR, EN, and VU species could be integrated into the spatial conservation prioritisation, the majority of which were highly localised endemics. One species, the red siskin (*Carduelis cucullata*) is endangered and has an extremely restricted range in Guyana, but no IUCN range map has been drawn. We therefore used a distribution published elsewhere (Robbins et al., 2004). The final dataset used in the conservation planning consisted of 329 species of terrestrial vertebrate (236 birds, 58 mammals, 35 reptiles and amphibians; Table A1, A2, A3), and the 17 vegetation types.

2.2. Spatial conservation prioritisation

Next, we developed the Guyana conservation planning system, by first dividing the country into planning units as required by Marxan. These planning units consisted of a series of hexagonal 1000 ha land parcels that we combined with the boundaries of the protected areas and community lands. We then calculated the amount of each conservation feature found in each unit. We collated data from various government agencies in Guyana to estimate the opportunity costs (US

dollars) associated forestry, current mining concessions, agriculture, urbanisation and areas with bedrock predicted to be rich in gold deposits (which in Guyana is used to allocate mining licenses; Fig. A2 and Table A4). For forestry, mining and cultivated land, we calculated opportunity costs from the contribution to national Gross Domestic Product (GDP), divided by the area allocated to its production. To estimate the opportunity cost of areas predicted to be rich in gold outside of current mining concessions, using the geological map of Guyana, we calculated half the value of current mining concessions in areas dominated by greenstone (which is associated with gold). For urban land, we used the market area-based value for housing. We then converted these data into the Marxan format using the CLUZ plugin for QGIS.

We set targets for each conservation feature based on their geographic coverage. We did this separately for the vegetation types and species, using a set target percentage for the feature with the smallest and largest range, and calculating the targets for remaining features using linear interpolation between the two extremes (Maiorano et al., 2006). Informed by sensitivity analyses, and in collaboration with the stakeholders, we set a 20% target for the smallest range and 1% target for the largest range (Fig. A3), helping to ensure a viable amount of each feature would be protected and producing results that did not select an unfeasibly large proportion of the country. We then measured the extent to which these targets were met in the current protected area network.

Marxan analyses involve multiple runs to identify near-optimal portfolios of planning units that meet targets, while minimising opportunity costs and boundary lengths. Thus, the most effective portfolios meet the targets while containing large patches of low-cost planning units (Ball et al., 2009). Marxan then produces two main outputs: the 'best' portfolio, which is the one with the lowest cost and the 'selection frequency' output, which counts the number of times each planning unit appeared in one of the portfolios. Each of our Marxan analyses consisted of 100 runs of 50 million iterations. After sensitivity analyses, we used a Boundary Length Modifier value of five to calculate the boundary cost, based on the total external edge of the portfolio, which is subsequently used to select viable patch sizes that are not too fragmented (Ball et al., 2009).

In our first baseline analysis, we specified that the existing protected areas should be automatically included in every Marxan portfolio, and any of the other planning units could be selected if needed. The results confirmed the conservation importance of community lands in Guyana, as planning units within each of 50 different community lands had high selection frequencies. Consequently, we then measured the relative importance of each of these community lands by excluding them sequentially in 50 additional analyses, each time calculating the number of targets met, the total planning unit cost, external edge and median patch area of the best portfolio compared with the baseline analysis. In this way, we: (a) determined whether excluding specific community lands from the analysis affected target attainment and the ecological integrity of the portfolio compared to the baseline analysis; (b) could measure the extent to which excluding each community land increased the opportunity costs of protecting the set of priority areas identified by Marxan.

3. Results

3.1. Representation in the current protected area network

Guyana's current protected area network covers c. 1.8 million ha but meets the representation targets for only 48% of vertebrate species (60%, 24% and 11% for bird, mammal, reptile and amphibian species respectively) and just five of the 17 vegetation types (mangrove, marsh forest, mixed lowland forest and white sand forest in southern Guyana) (Fig. 2). Grasslands, highlands and wetlands are largely missing from the network. Of the threatened species, eight are completely absent

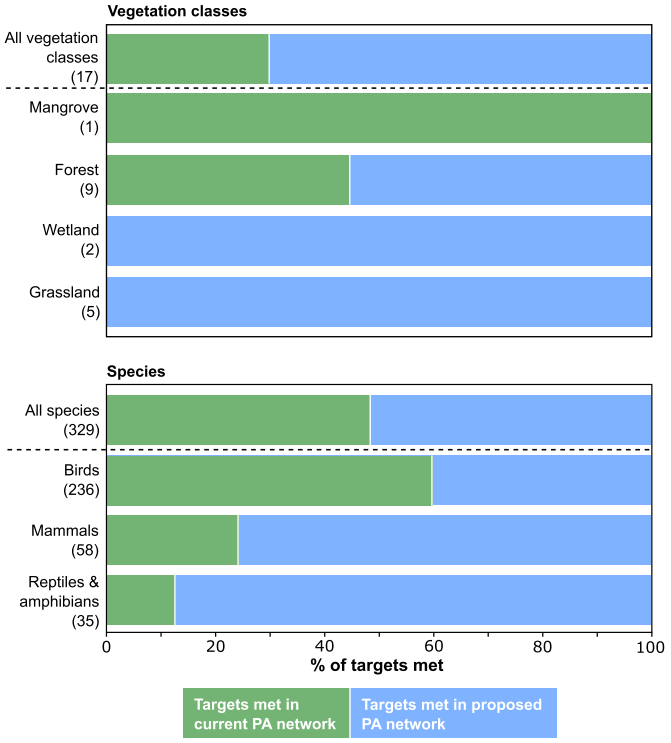

Fig. 2. Percent of conservation feature targets met in Guyana's the current protected area (PA) network in green, and the percent of targets that would be met by the conservation plan in blue (i.e. the proposed network meets all targets). Vegetation classes (top panel) are partitioned into broad categories of mangrove, forest, wetland, and grassland. Species (bottom panel) are partitioned into birds, mammals, and reptiles and amphibians. Number of species/vegetation types associated with each bar are show in parentheses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from the existing protected areas. These are, four birds: Rio Branco antbird (*Cercomacra carbonaria*; CR), hoary-throated spinetail (*Synallaxis kollari*; CR), Red Siskin (*Carduelis cucullata*; EN), white-bellied piculet (*Picumnus spilogaster*; VU); two amphibians: MacConnell's bush toad (*Oreophrynella macconnelli*; VU), Pebas stubfoot toad (*Atelopus spumarius*; VU); and two mammals: Reig's opossum (*Monodelphis reigi*; VU), and Venezuelan fish-eating rat (*Neusticomys venezuelae*; VU).

3.2. Priority areas for conservation outside of the current protected area network

To meet the CBD Aichi Target of 17%, Guyana needs to double the extent of its protected area network with an additional 8.5% (1.8 million ha) by 2020. Based on meeting representation targets for biodiversity and vegetation, our conservation planning analysis identified approximately 20 priority areas for protection (Fig. 3a). In order to meet all the targets, our baseline analysis shows that an additional 14% (3 million ha) of Guyana's terrestrial area would be required, bringing the extent of the protected area network to 22.5% (4.8 million ha) of the country. Of the additional priority area required for this, 8.8% (1.9 million ha) is state-owned land, and 5.2% (1.1 million ha) is community land. The analysis showed that to meet representation targets, approximately 750,000 ha would be required in the highlands; a little over 1 million ha in the south-western grasslands; 660,000 ha in the north-eastern mixed grasslands, forests and wetlands; 200,000 ha in north western wetlands; and 190,000 ha in the southern forests, with

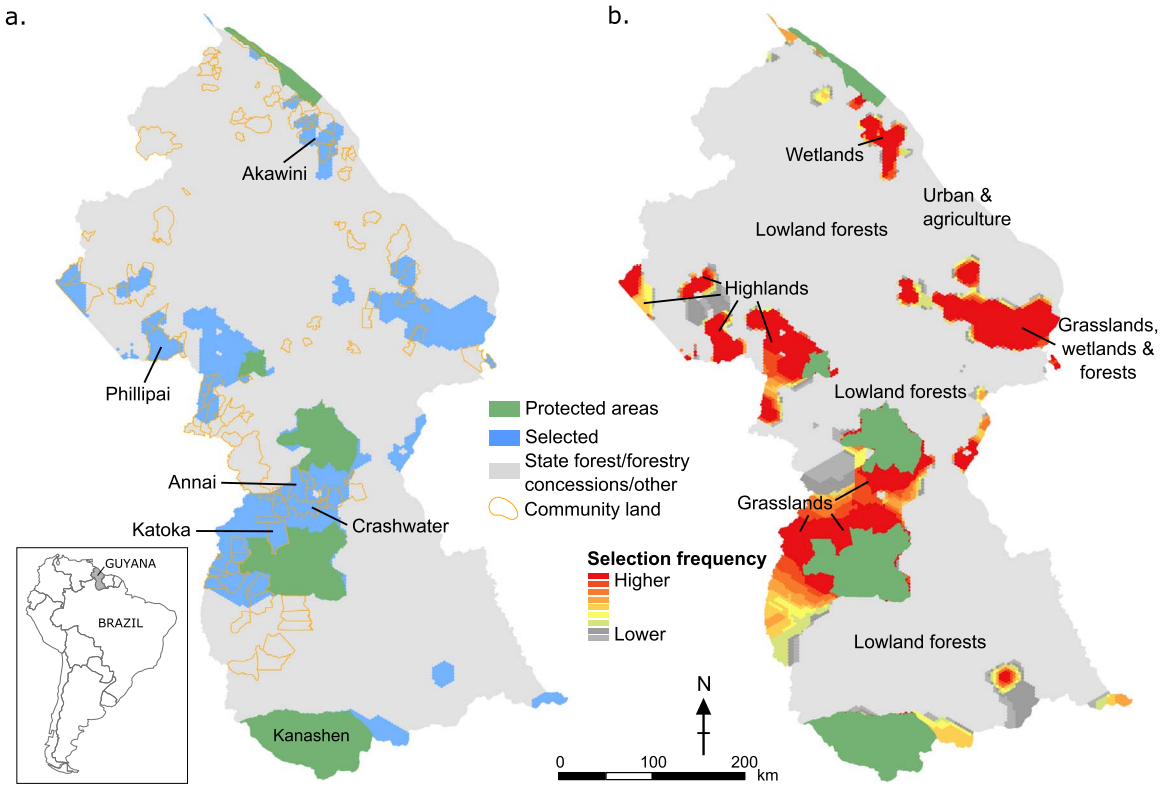

Fig. 3. A. shows the results of the baseline systematic conservation plan for Guyana. Blue areas are selected in the best solution of the analysis. B. indicates the selection frequency of each planning unit from 100 Marxan runs, with the darkest red showing those areas selected in ≥ 90 runs. Existing protected areas are shown in green. The unselected grey areas include unallocated state land, forestry concessions, mining concessions, agricultural land and urban areas (which are almost exclusively along the coastal belt) (Fig. 1). Areas and community lands mentioned in the text are labelled. See Fig. A1 for detailed habitat map. Inset shows the location of Guyana in South America. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Attributes of the baseline conservation plan for Guyana, where no community lands were excluded, and the five most important communities for meeting conservation targets with greatest ecological integrity (e.g. largest patches sizes, less edge, decreased opportunity costs). See also Fig. 4.

Analysis	Mean selection frequency in baseline portfolio (%)	Median patch size (ha)	Number of patches	Total edge (m)	Total opportunity cost (US$)	Portfolio area (ha)	Portfolio area (% of country)	Percentage of portfolio in community lands (excluding Kanashen)
Baseline	–	70,832	18	3,392,979	12,869,228	4,768,386	22.5	5.2
Excluded community								
Phillipai	85.38	58,798	18	3,530,605	13,578,233	4,668,840	22.0	4.8
Akawini	83.26	48,000	21	3,442,013	13,382,968	4,669,679	22.0	4.7
Katoka	99.92	51,050	22	3,491,226	13,060,845	4,595,027	21.7	4.4
Crash water	92.25	69,510	20	3,515,180	13,107,425	5,147,217	24.3	5.1
Annai	88.38	65,138	21	3,436,697	13,137,224	4,578,050	21.6	4.4

the remaining approximately 200,000 ha distributed in smaller patches throughout the country. To meet just the 17% target, the top priority areas are the largest patches with highest selection frequency scores, i.e. those selected in ≥ 90% of runs which are at least as big as the current smallest protected area in Guyana (Fig. 3b). These are two areas of highlands in the west, grasslands in the south-west that join two existing protected areas, mixed grasslands, forests and wetlands in the north-east, and wetlands in the north-west. Protecting these areas would mean every species and vegetation type would at least be represented within a protected area in Guyana.

3.3. Identifying the most important community lands for expanding the protected area network

The 50 analyses that excluded each community land in turn showed that no single community land was essential for meeting targets. However, excluding these community lands did have an impact on the opportunity costs and fragmentation levels of the portfolios, compared to the baseline analysis where all community land was available for selection. The five most important community lands were Phillipai, Akawini, Katoka, Crashwater and Annai, as their omission caused opportunity costs to rise by between 1.5% and 5.5% (Table 1; Fig. 4), and

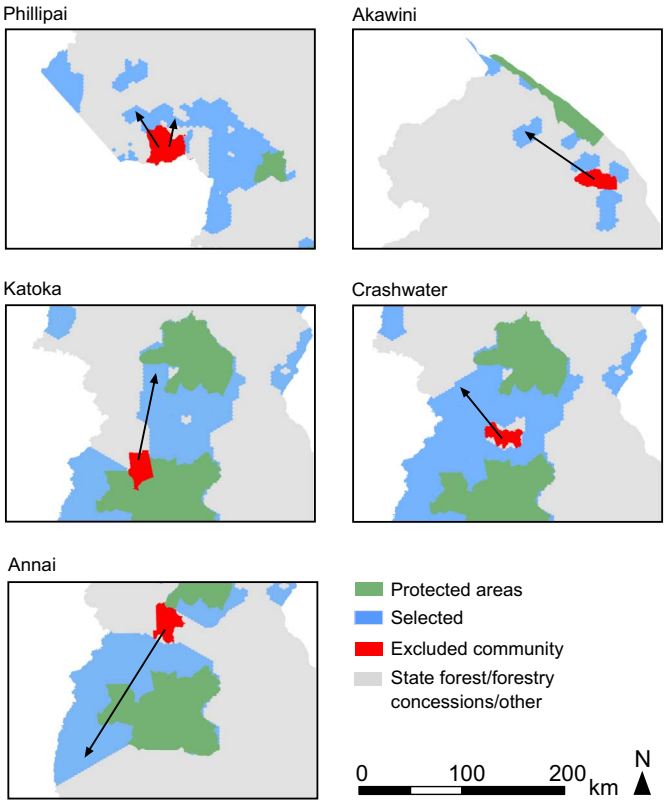

Fig. 4. Results of the spatial prioritisation when each of the 5 most important community lands are excluded from the analysis. The arrows show the areas where the analysis has selected to meet the targets in lieu of the community land being unavailable. See Table 1 for further details of each analysis.

resulted in the selection of areas that in the baseline analysis had comparatively low selection frequency scores (Fig. 3b). Likewise their omission increased the boundary edge of the sets of priority areas by between 49,000 m and 140,000 m, and decreased median patch size by between 1.9% and 32.2%, compared to the baseline analysis.

4. Discussion

With over half of CBD parties yet to meet their commitments under Aichi Target 11, the coming three years should represent the fastest terrestrial protected area expansion rate ever seen. To contribute to this, our study shows that the systematic conservation planning approach is suitable for national-scale prioritisations because it is based on a set of principles that are scientifically sound but flexible enough to adapt to national conditions (Margules and Pressey, 2000). Therefore, the value of these techniques depends on accounting for the implementation context (Knight et al., 2011), ensuring the results are relevant for guiding policy. For Guyana, this entails establishing new protected areas to fulfil Aichi commitments, and additionally to contribute to avoided deforestation targets under the country's REDD+ agreement with Norway. As such, our study illustrates a cost-effective, stakeholder-led and collaborative initiative that is now guiding conservation action because it was driven by national priorities, as well as biodiversity conservation. This not only provides a conservation plan to represent all biodiversity and vegetation types in this part of Amazonia, but also shows the value of national-scale and agency-led spatial conservation prioritisations.

In the near-term, Guyana has committed to expand its protected area network by 2 million ha (to 17%). To direct this, our analyses highlight several biomes that are currently unprotected, such as the forested highlands that are home to the unique, long isolated steep-sided mountains known as Tepuis. Their flat peaks are rich in restricted range endemic species (McPherson, 2008), and connect with the Canaima and Mount Roraima National Parks in Venezuela and Brazil. These highlands are shown to be critical to achieving conservation targets in Guyana, because they exhibit high selection frequency, and in addition, they are shown to be important in global spatial conservation prioritisations (Pollock et al., 2017; Venter et al., 2014). So alongside affirming the biodiversity value of these areas, our study provides maps at a suitable scale for use by policy makers to delineate protected area boundaries.

Our spatial conservation prioritisation also demonstrates that it is not just forests that need protection, as the biodiversity-rich grasslands are poorly represented in Guyana's protected area network. In particular, our analyses identified the Rupununi Savannas, which contain a range of under-protected habitats and species, and the Rio Branco Endemic Bird Area. This area additionally illustrates the importance of community lands in optimising the biodiversity value of future land-use strategies, as the Rupununi Savannas are predominantly community-owned, and encompass three of the five most critical community lands as identified by our analysis (Table 1; Fig. 4). If the communities in the Rupununi opted for protected area expansion (and this is currently being assessed), their lands would connect two existing protected areas to the Raposa Serra do Sol reserve in Brazil, and establish the largest protected area expanse in Guyana.

Our spatial prioritisation additionally contributes directly to Guyana's commitments under the REDD+ agreement with Norway, by incorporating biodiversity into Guyana's land-use planning process, which is centred on reducing deforestation rates and carbon emissions via sustainable economy initiatives. While our approach does not directly target areas of increased deforestation risk, country-wide deforestation targets will be delivered through policies that combine effectively managed protected areas (which are selected for biodiversity) with low-level resource extraction outside protected areas. If well managed, these extraction areas can be OECMs, which are recognised in the Aichi targets as playing a similarly important role as protected areas (Angelsen and Rudel, 2013; CBD, 2010). Indeed, most forestry operations in Guyana adopt low-intensity reduced-impact logging, which local and global studies have shown to maintain an almost full complement of tropical forest biodiversity (e.g. Bicknell et al., 2014, 2015; Roopsind et al., 2017). For this reason, the Government of Guyana is strengthening sustainable land management in general, and this can also be guided by our priority area map, which identified a further 1.2 million ha beyond the land needed to meet the Aichi Target. In addition, community areas may contribute to both protected area expansion and OECMs (Nepstad et al., 2006), and together with reduced-impact logging, could form a spatial conservation network aimed at reconciling development with the maintenance of high conservation values. As a national strategy, these might also help to shift the economy away from mineral mining, which is the principal source of forest loss in Guyana, towards a more environmentally sustainable and low-carbon model. Doing so would contribute to the reduction in deforestation required to receive performance-related REDD+ payments (Office of the President, 2013).

Given the importance of community lands in Guyana, successfully expanding the protected area network will involve working together with landowners, a process that underpins the principle of free prior and informed consent required by the CBD and REDD+ agreement. About 15% of Guyana's land is under the ownership of local communities, so support from these people is crucial for meeting the targets. Previous analyses have shown that incorporating data on landowner willingness to engage in conservation prioritisations is a powerful way to increase support for the identified priority areas (Game et al., 2011; Guerrero et al., 2010). However, because community lands in Guyana are large and remote, it would be prohibitively costly to measure willingness beforehand. Instead, we used our scenario analyses to identify the relative importance of each community, thus helping

prioritise where consultations should first take place. We found no single community land was vital for meeting targets, and this flexibility means policy makers can make alternative plans if one or several communities choose not to participate. However, if any community decided not to engage with the conservation plan, this would have negative impacts on opportunity costs and levels of fragmentation in the resulting priority areas.

5. Conclusions

Our study is a stakeholder-led spatial conservation prioritisation process for Guyana, informed by science and underpinned by the systematic conservation planning approach. This has helped both to ensure that biodiversity is adequately represented in protected area expansion, and embed the results into broader land-use decision-making, including working with local communities. Once implementation is completed, Guyana will be a major contributor to long-term conservation in this part of Amazonia, alongside demonstrating exemplary accomplishment of its Aichi Target 11 commitments. As such, our work shows the relevance of using systematic conservation planning to design protected area networks that translate international conservation commitments into national action.

Acknowledgements

We would like to thank the following people for contributing to the spatial conservation prioritisation: H. Persaud, A. Roberts, P. Rampersaud, C. Hutchinson, C. Bernard, B. ter Welle, D. Singh, R. Singh, N. Nasir, A. Foo, C. Sookdeo, M. Paul, M. Kalamandeen, G. Agard, M. Saywack, R. Smith, D. Singh, P. Williams, S. Rose, R. Thomas, K. Mosley, A. Mohamed, T. Yussuff, J. Breems, A. Bayney, C. Harilal, A. Adams, O. Davis, V. Rajkumar. Thanks also to K. Metcalfe for Marxan guidance, and P. Mitchell and T. Laing for advice. Data used in the analyses are freely available on request from the Guyana Conservation Network (www.guycon.org). J. Bicknell, M. Struebig and Z. Davies are supported by NERC grant number NE/K016407/1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.biocon.2017.08.024.

References

Angelsen, A., Rudel, T.K., 2013. Designing and implementing effective REDD + policies: a forest transition approach. Rev. Environ. Econ. Policy 7, 91–113.
Ball, I.R., Possingham, H.P., Watts, M., 2009. Marxan and relatives: software for spatial conservation prioritisation. In: Moilanen, A., Wilson, K.A., Possingham, H.P. (Eds.), Spatial Conservation Prioritisation: Quantitative Methods and Computational Tools. Oxford, Oxford University Press.
Bicknell, J.E., Struebig, M.J., Edwards, D.P., Davies, Z.G., 2014. Improved timber harvest techniques maintain biodiversity in tropical forests. Curr. Biol. 24, 1119–1120.
Bicknell, J.E., Struebig, M.J., Davies, Z.G., 2015. Reconciling timber extraction with biodiversity conservation in tropical forests using reduced-impact logging. J. Appl. Ecol. 52, 379–388.
Brooks, T.M., 2014. Conservation: mind the gaps. Nature 516, 336–337.
Bruner, A.G., Gullison, R.E., Rice, R.E., da Fonseca, G.A.B., 2001. Effectiveness of parks in protecting tropical biodiversity. Science 291, 125–128.
Butchart, S.H., Clarke, M., Smith, R.J., Sykes, R.E., Scharlemann, J.P., Harfoot, M., Buchanan, G.M., Angulo, A., Balmford, A., Bertzky, B., 2015. Shortfalls and solutions for meeting national and global conservation area targets. Conserv. Lett. 8, 329–337.
CBD, 2010. COP 10 Decision X/2: Strategic Plan for Biodiversity 2011–2020. www.cbd.int/decision/cop/?id=12268.
Chape, S., Harrison, J., Spalding, M., Lysenko, I., 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. Philos. Trans. R. Soc., B 360, 443–455.
Core Team, R., 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria Available at. http://www.R-project.org/.
Di Minin, E., Soutullo, A., Bartesaghi, L., Rios, M., Szephegyi, M.N., Moilanen, A., 2017. Integrating biodiversity, ecosystem services and socio-economic data to identify priority areas and landowners for conservation actions at the national scale. Biol. Conserv. 206, 56–64.
Elith, J.H., Graham, C.P., Anderson, R., Dudík, M., Ferrier, S., Guisan, A.J., Hijmans, R., Huettmann, F.R., Leathwick, J., Lehmann, A., Li, J.G., Lohmann, L.A., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, McC M., Townsend, J., Peterson, A.J., Phillips, S., Richardson, K., Scachetti-Pereira, R.E., Schapire, R., Soberón, J., Williams, S.S., Wisz, M.E., Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.
Esselman, P.C., Allan, J.D., 2011. Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica. Freshw. Biol. 56, 71–88.
Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.
Game, E.T., Lipsett-Moore, G., Hamilton, R., Peterson, N., Kereseka, J., Atu, W., Watts, M., Possingham, H.P., 2011. Informed opportunism for conservation planning in the Solomon Islands. Conserv. Lett. 4, 38–46.
Guerrero, A.M., Knight, A.T., Grantham, H.S., Cowling, R.M., Wilson, K.A., 2010. Predicting willingness-to-sell and its utility for assessing conservation opportunity for expanding protected area networks. Conserv. Lett. 3, 332–339.
Gutman, P., Aguilar-Amuchastegui, N., 2012. Reference levels and payments for REDD +: lessons from the recent Guyana–Norway agreement. World Wildlife Fund, USA Available at: http://assets.panda.orf/downloads/rls_and_payments_redd_lessons.pdf/.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853.
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.
Howard, J., Trotz, M.A., Thomas, K., Omisca, E., Chiu, H.T., Halfhide, T., Akiwumi, F., Michael, R., Stuart, A.L., 2011. Total mercury loadings in sediment from gold mining and conservation areas in Guyana. Environ. Monit. Assess. 179, 555–573.
Jenkins, C.N., Pimm, S.L., Joppa, L.N., 2013. Global patterns of terrestrial vertebrate diversity and conservation. Proc. Natl. Acad. Sci. 110, E2602–E2610.
Joppa, L.N., Pfaff, A., 2009. High and Far: biases in the location of protected areas. PLoS One 4, e8273.
Knight, A.T., Cowling, R.M., Boshoff, A.F., Wilson, S.L., Pierce, S.M., 2011. Walking in STEP: lessons for linking spatial prioritisations to implementation strategies. Biol. Conserv. 144, 202–211.
Laing, T., 2015. Rights to the forest, REDD + and elections: mining in Guyana. Res. Policy 46, 250–261.
Larsen, F.W., Londoño-Murcia, M.C., Turner, W.R., 2011. Global priorities for conservation of threatened species, carbon storage, and freshwater services: scope for synergy? Conserv. Lett. 4, 355–363.
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.
Maiorano, L., Falcucci, A., Boitani, L., 2006. Gap analysis of terrestrial vertebrates in Italy: priorities for conservation planning in a human dominated landscape. Biol. Conserv. 133, 455–473.
Margules, C.R., Pressey, R.L., 2000. Systematic conservation planning. Nature 405, 243–253.
McPherson, S., 2008. Lost Worlds of the Guiana Highlands. Redfern Natural History Productions Ltd.
Moore, C.H., Radford, B.T., Possingham, H.P., Heyward, A.J., Stewart, R.R., Watts, M.E., Prescott, J., Newman, S.J., Harvey, E.S., Fisher, R., Bryce, C.W., Lowe, R.J., Berry, O., Espinosa-Gayosso, A., Sporer, E., Saunders, T., 2016. Improving spatial prioritisation for remote marine regions: optimising biodiversity conservation and sustainable development trade-offs. Sci Rep 6, 32029.
Nepstad, D., Schwartzman, S., Bamberger, B., Santilli, M., Ray, D., Schlesinger, P., Lefebvre, P., Alencar, A., Prinz, E., Fiske, G., Rolla, A., 2006. Inhibition of Amazon deforestation and fire by parks and indigenous lands. Conserv. Biol. 20, 65–73.
Office of the President, 2013. Transforming Guyana's Economy While Combating Climate Change. Strategy update, Low Carbon Development.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.
Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197.
Platts, P.J., Garcia, R.A., Hof, C., Foden, W., Hansen, L.A., Rahbek, C., Burgess, N.D., 2014. Conservation implications of omitting narrow-ranging taxa from species distribution models, now and in the future. Divers. Distrib. 20, 1307–1320.
Pollock, L.J., Thuiller, W., Jetz, W., 2017. Large conservation gains possible for global biodiversity facets. Nature 546, 141–144.
Robbins, M.B., Braun, M.J., Finch, D.W., 2004. Discovery of a population of the endangered Red Siskin (*Carduelis cucullata*) in Guyana. Auk 120, 291–298.
Roopsind, A., Caughlin, T.T., Sambhu, H., Fragoso, J., Putz, F.E., 2017. Logging and indigenous hunting impacts on persistence of large Neotropical animals. Biotropica 49, 565–575.
Smith, R.J., Easton, J., Nhancale, B.A., Armstrong, A.J., Culverwell, J., Dlamini, S.D., Goodman, P.S., Loffler, L., Matthews, W.S., Monadjem, A., Mulqueeny, C.M., Ngwenya, P., Ntumi, C.P., Soto, B., Leader-Williams, N., 2008. Designing a transfrontier conservation landscape for the Maputaland centre of endemism using biodiversity, economic and threat data. Biol. Conserv. 141, 2127–2138.
Smith, R.J., Verissimo, D., Leader-Williams, N., Cowling, R.M., Knight, A.T., 2009. Let the locals lead. Nature 462, 280–281.

ter Steege, H., 2001. Mapping Forest and Vegetation of Guyana at Regional and National Level.
Syfert, M.M., Smith, M.J., Coomes, D.A., 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PLoS One 8, e55158.
UNEP-WCMC, IUCN, 2016. Protected Planet Report 2016. UNEP-WCMC and IUCN: Cambridge UK and Gland, Switzerland.
Venter, O., Possingham, H.P., Hovani, L., Dewi, S., Griscom, B., Paoli, G., Wells, P., Wilson, K.A., 2013. Using systematic conservation planning to minimize REDD plus conflict with agriculture and logging in the tropics. Conserv. Lett. 6, 116–124.
Venter, O., Fuller, R.A., Segan, D.B., Carwardine, J., Brooks, T., Butchart, S.H., Di Marco, M., Iwamura, T., Joseph, L., O'Grady, D., 2014. Targeting global protected area expansion for imperiled biodiversity. PLoS Biol. 12, e1001891.
Venter, O., Magrach, A., Outram, N., Klein, C.J., Marco, M.D., Watson, J.E.M., 2017. Bias in protected-area location and its effects on long-term aspirations of biodiversity conventions. Conserv. Biol. http://dx.doi.org/10.1111/cobi.12970.
Watson, J.E., Darling, E.S., Venter, O., Maron, M., Walston, J., Possingham, H.P., Dudley, N., Hockings, M., Barnes, M., Brooks, T.M., 2016. Bolder science needed now for protected areas. Conserv. Biol. 30, 243–248.
World Bank, 2017. Available at: http://databank.worldbank.org/data/reports.aspx?source=2&series=ER.LND.PTLD.ZS&country=.