ARTICLE

ECOLOGICAL
APPLICATIONS
ECOLOGICAL SOCIETY OF AMERICA

Finding rare species and estimating the probability that all occupied sites have been found

Francois S. Becker[1,2] (ORCID) | **Jasper A. Slingsby**[3,4] (ORCID) | **John Measey**[5] (ORCID) |
Krystal A. Tolley[1,6] (ORCID) | **Res Altwegg**[2,7] (ORCID)

[1]South African National Biodiversity Institute, Kirstenbosch Research Centre, Cape Town, South Africa

[2]Centre for Statistics in Ecology, Environment and Conservation, Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa

[3]Department of Biological Sciences and Centre for Statistics in Ecology, Environment and Conservation, University of Cape Town, Cape Town, South Africa

[4]Fynbos Node, South African Environmental Observation Network, Centre for Biodiversity Conservation, Cape Town, South Africa

[5]Centre for Invasion Biology, Department of Botany & Zoology, Stellenbosch University, Stellenbosch, South Africa

[6]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, Johannesburg, South Africa

[7]African Climate and Development Initiative, University of Cape Town, Cape Town, South Africa

Correspondence
Res Altwegg
Email: res.altwegg@gmail.com

Present address
Francois S. Becker, National Museum of Namibia, Windhoek, Namibia

Funding information
Mandela Rhodes Foundation; National Research Foundation of South Africa, Grant/Award Numbers: 85413, 85802, 118593; National Science Foundation, Grant/Award Number: DEB1120502; South African National Biodiversity Institute

Handling Editor: Beth Gardner

Abstract

Detecting occupied sites of rare species, and estimating the probability that all occupied sites are known within a given area, are desired outcomes for many ecological or conservation projects. Examples include managing all occupied sites of a threatened species or eradicating an emerging invader. Occupied sites may remain undetected because (1) sites where the species potentially occurs had not been searched, and (2) the species could have been overlooked in the searched sites. For rare species, available data are typically scant, making it difficult to predict sites where the species probably occurs or to estimate detection probability in the searched sites. Using the critically endangered Rose's mountain toadlet (*Capensibufo rosei*), known from only two localities, we outline an iterative process aimed at estimating the probability that any unknown occupied sites remain and maximizing the chance of finding them. This includes fitting a species distribution model to guide sampling effort, testing model accuracy and sampling efficacy using the occurrence of more common proxy species, and estimating detection probability using sites of known presence. The final estimate of the probability that all occupied sites were found incorporates the uncertainties of uneven distribution, relative area searched, and detection probability. Our results show that very few occupied sites of *C. rosei* are likely to remain undetected. We also show that the probability of an undetected occupied site remaining will always be high for large unsearched areas of potential occurrence, but can be low for smaller areas intended for targeted management interventions. Our approach is especially useful for assessing uncertainty in species occurrences, planning the required

search effort needed to reduce probability of unknown occurrence to desired levels, and identifying priority areas for further searches or management interventions.

KEYWORDS

detection probability, model-based sampling, occupancy, rarity, species distribution model

INTRODUCTION

We often need to know how many occupied sites of a rare species exist in a certain area, for example to inform conservation efforts if the species is of conservation concern or to direct eradication efforts if the species is an emerging invader. However, the number of occupied sites in a given area is generally uncertain because (1) occupied sites could exist at sites that have not been surveyed given that most areas of interest are too large to be searched comprehensively, and (2) occupied sites may have been overlooked at sites that have been surveyed given that detection is almost never perfect.

The key unknown quantity of interest is the probability that no unknown occupied sites persist in a defined area: $P_{N_A=0}$. In other words, if the area covered by known occupied sites is excluded, we want to know if the number of occupied sites (N), in the remaining area (A), or N_A, is 0. This is the same as the probability that all occupied sites have been found in the larger area of interest. We would like to quantify $P_{N_A=0}$ and increase it as much as possible to be able to target management interventions. If we knew the probability of a site i being occupied, Ψ_i, for all K sites where the species has not been recorded, then $P_{N_A=0} = \prod_{i=1}^{K}(1 - \Psi_i)$. Additionally, $P_{N=0}$ could be increased most effectively by establishing presence or absence at those sites at which Ψ_i is highest. Absence at a particular site cannot be established with certainty unless the probability of detecting the species at an occupied site, p_i, is 1. However, each survey resulting in a nondetection at a site would lower Ψ_i for that site (Wintle et al., 2012).

Ideally, we would like to be able to use occupancy models (MacKenzie et al., 2002, 2017) to estimate Ψ_i and p_i directly. Unfortunately, the site-specific occupancy probabilities Ψ_i are difficult to obtain for very rare species, because there are few data to inform models for the occupancy and detection processes. Occupancy models require the data to be collected in such a way that the detection process can be estimated (Guillera-Arroita et al., 2015), which can be difficult to achieve for a species that is absent nearly everywhere. For such species, information is often only available in the form of opportunistic records that establish presence at some sites (assuming these records can be verified to a degree that excludes false detections), often collected over long time-scales, but does not contain any information about the observation process, such as the spatial pattern of search effort, for example.

Opportunistic presence records can be analyzed using species distribution models that are designed for presence-only data, for example, MaxEnt (Phillips & Dudik, 2008) and MaxLike (Royle et al., 2012). These methods estimate a habitat-suitability index H_i that is not equivalent to Ψ_i (Dorazio, 2012; Yackulic et al., 2013). However, under certain conditions, H_i is related to Ψ_i in a sense that the species is expected to occur at sites with high values of H_i more often than at sites with low H_i. H_i are proportional to Ψ_i if the detection probabilities p_i are constant and if there is no spatial sampling bias (Guillera-Arroita et al., 2015; Merow et al., 2013).

For rare species for which presence records exist, one could therefore increase the probability that all extant occupied sites are found ($P_{N_A=0}$) by fitting species distribution models to these data and then biasing the search toward sites with the highest predicted habitat suitability, H_i. This approach can be successful (e.g., T.C. Edwards et al., 2005; Guisan et al., 2006; Le Lay et al., 2010). However, the reliability of species distribution models depends on the amount of data (Hernandez et al., 2006; Pearson et al., 2007) and no algorithm predicts H_i consistently well when there are small sample sizes (Wisz et al., 2008).

Inference about habitat suitability for rare species can be improved by borrowing occurrence information from closely related or associated species as proxies (T.C. Edwards et al., 2005; McCune, 2016). For instance, T.C. Edwards et al. (2005) stratified their search effort for their rare target species using a species distribution model fitted to occurrence data for a common species that is commonly associated with the target species. McCune (2016) used data on common species associated with the rare target species to assess species distribution model performance even in the absence of additional records of the rare species.

A species distribution model that predicts H_i well may be able to direct searches toward those sites that truly have the highest occupancy probabilities, Ψ_i.

However, nondetections at these sites only establish absence if the detection probability, p_i, is 1. The detection probability can be estimated by repeatedly sampling sites (Guillera-Arroita, 2017) but obtaining useful estimates can be difficult for rare species, as most searches do not lead to any detections. In this case, repeatedly sampling only those sites at which the species has already been detected is better than re-sampling all sites (Specht et al., 2017). In addition, information can be borrowed from more common species with similar detection probabilities (MacKenzie et al., 2005). For instance, Nichols et al. (2000) grouped bird species based on prior experience of observational traits, and estimated detection from data for the group of species, even though some rarer species lacked sufficient observations to estimate species-specific detection probabilities. However, rare species may still have lower detection probabilities induced by lower abundance.

We faced many of these problems associated with rarity when working on the critically endangered Rose's mountain toadlet, *Capensibufo rosei* (SA-FRoG IUCN SSC Amphibian Specialist Group, 2010). Historic records for this species are sparse across both space and time, despite considerable search efforts (Cressey et al., 2015; S. Edwards et al., 2017). Furthermore, the lack of a breeding call in this species (Grandison, 1980) presents an unusual problem for detection of frogs, which are often detected primarily by call (e.g., Measey et al., 2017; Shirose et al., 1997). The known occupied range of *C. rosei* has declined in recent decades, despite most of its habitat being protected (Cressey et al., 2015). A systematic search to determine the number of occupied sites remaining and identify sites for management interventions is therefore a conservation priority (Measey, 2011).

Using *C. rosei* as a case study, we outline a search strategy aimed at discovering unknown occupied sites of a rare species and quantify the resulting increase in the probability that all remaining occupied sites in a given area are found, $P_{N_A=0}$. Our search strategy consists of seven steps: (1) Assemble a list of sites at which the species is known to occur or has occurred historically and survey these sites. (2) Fit a species distribution model to these data of known extant or historic presence and use the fitted model to stratify search effort across the area of interest, based on predicted habitat suitability. (3) Carry out searches at the selected sites while using expert knowledge to guide the timing, methods, and survey effort to maximize detection probability. (4) At the same sites, also search for and record detections of more common proxy species which appear to require similar habitat conditions. (5) Estimate detection probability for the focal species where it is known to occur and learn about the observation process from the more common detections of the proxy species; (6) Evaluate the predictive accuracy of the species distribution model using the presence/absence of proxy species and apparently suitable microhabitat for the focal species in the searched grid cells; (7) Estimate the probability that all extant occupied sites have been found across the study area. Steps 2–7 can be repeated until $P_{N_A=0}$ reaches the desired level.

MATERIALS AND METHODS

Study area and study species

Rose's mountain toadlet (*C. rosei*) is endemic to fynbos vegetation on the Cape Peninsula, South Africa (Channing et al., 2017; Tolley et al., 2010). Only two extant breeding sites were known at the beginning of this study, both occurring within Table Mountain National Park: one in the Silvermine section and one in the Cape of Good Hope section (see Figure 1). Each breeding site consists of up to 10 individual breeding pools, typically still-standing water, < 2 m in diameter, < 3 cm deep, usually within a radius of less than ~ 200 m in localized, ephermal seepages within a wetland area (Becker et al., 2018; S. Edwards et al., 2017). Although *C. rosei* has no breeding call and therefore cannot be detected using auditory surveys, it breeds in dense aggregations. Intensive surveys in previous years have established that the toads breed in pools located in small clearings and avoid densely vegetated patches (Becker, 2017). These aggregations are more easily seen than individual adults, and last for a variable period of a few weeks to 50 days, depending on the rainfall (Becker et al., 2018). The distinct and dense masses of eggs and tadpoles are also relatively easy to observe, and usually remain for > 2 months (S. Edwards et al., 2017).

Historic observations and observations from the breeding site monitoring over more than a decade, suggest that breeding sites are highly localized, and may remain roughly in the same location over many years.

We defined our study area as all land containing natural fynbos vegetation, excluding intertidal zones (based on data obtained from the City of Cape Town's Open Data Portal: https://web1.capetown.gov.za/web1/opendataportal/Default), between latitudes: $-33.890°$ and $-34.407°$, and between longitudes: $18.515°$ and $18.280°$, that is, the Cape Peninsula. The 31,462 ha study area is mountainous, ranging from sea level to 1086 m. The weathering-resistant sandstone bedrock that is found across much of the peninsula creates rugged topography, and gives rise to the dominant acidic, nutrient-poor silica soils (Cowling et al., 1996).
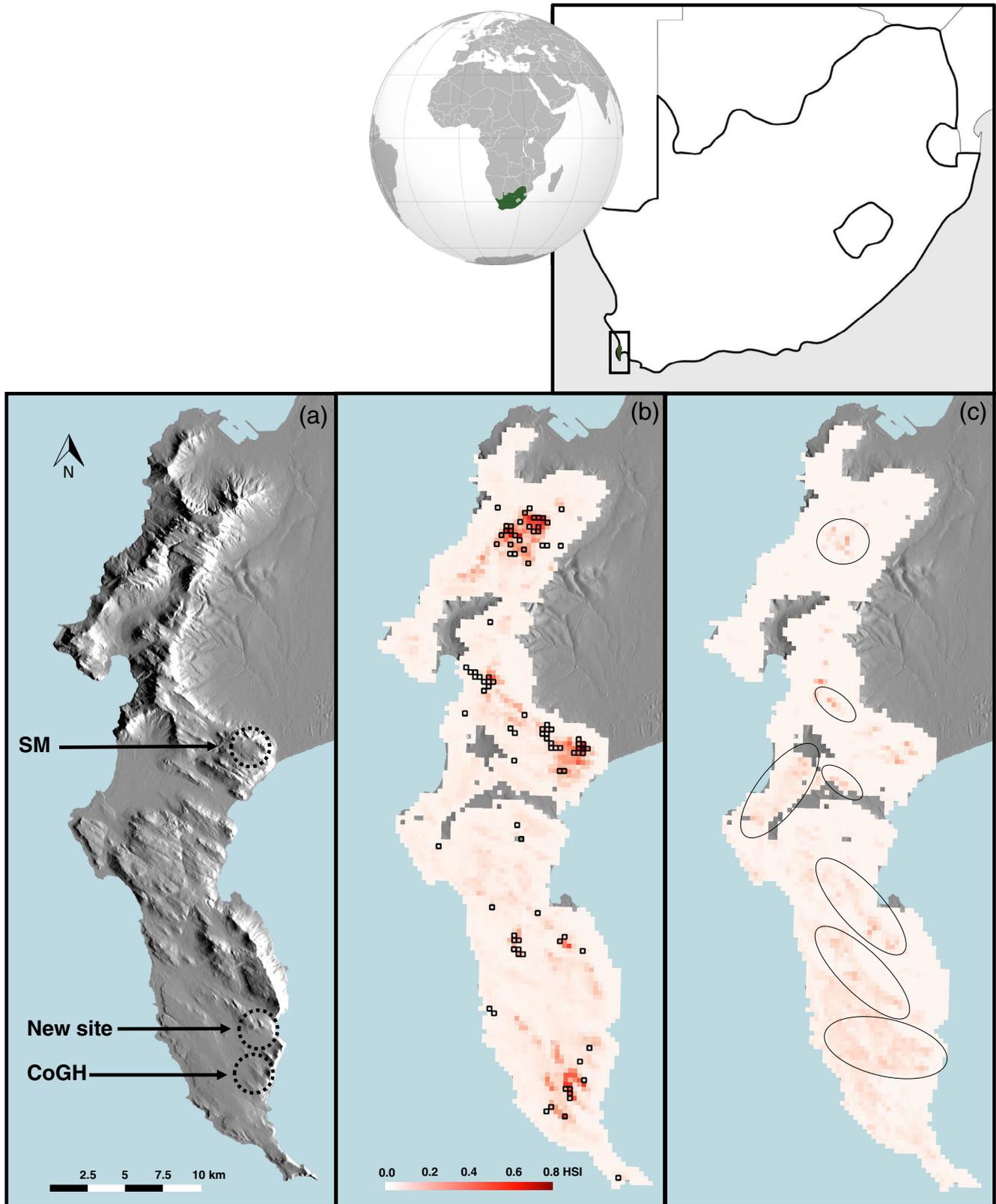
FIGURE 1 Potential range of *Capensibufo rosei* on the Cape Peninsula, South Africa (between latitudes: $-33.890°$ and $-34.407°$, and between longitudes: $18.515°$ and $18.280°$), showing three-dimensional rendering of the topography. (a) Shows the approximate location of the two previously known and one newly discovered breeding sites (SM, Silvermine; CoGH, Cape of Good Hope), because the precise locations are sensitive information. (b) Shows the mean predicted habitat suitability (H_i) per 300×300 m grid cell overlaid on this map, and the 96 searched cells indicated by black squares. (c) Shows the difference in predicted H_i between the first SDM and the second, after the inclusion of the new record. Red indicates areas of higher predicted H_i in the second model, some of which are additional areas to search (circled areas)

Seven-step search strategy

Our aim was to find additional breeding sites of *C. rosei* and estimate the probability that no further unknown occupied sites occur (1) across its entire range and (2) on top of Table Mountain. The latter area is of particular interest because *C. rosei* has not been recorded from this area since 1983, even though it historically used to hold large occupied sites of this species (Cressey et al., 2015).

Step 1: Compile list of extant and historic presences

Cressey et al. (2015) assembled a list of historic presence records for *C. rosei*, searched them and found no active breeding sites except at two locations, the rest were presumed to be no longer extant. There were 17 possible historic localities recorded in Cressey et al. (2015). Of these localities, 11 were precise enough that we could use them for species distribution modeling (spatial error <30 m).

Step 2: Stratify search effort according to habitat suitability

Species distribution modeling

We fitted a species distribution model to 19 *C. rosei* occurrence records.

These included nine historic records of presumed extinct occupied sites (from Cressey et al., 2015) and 10 recent occurrence records from the two extant occupied sites. The latter consisted of seven distantly spaced (75–750 m) breeding pools and three records of adults that were found >30 m away from the breeding pools.

We used environmental predictor variables (Appendix S1: Table S1) for which the spatial pattern is unlikely to have changed substantially (Roubicek et al., 2010) over the period in which the occurrence data were recorded (1950–2014). Some variables focused on the broader scale habitat context (such as vegetation and small-scale climatic variation), while some aimed to target the likely areas in which suitable breeding pools would form (e.g., topography variables related to drainage patterns). We did not have access to any remote-sensing variables that were capable of directly detecting suitable breeding pools, mostly due to their small size (often <0.2 m^2 and <2 cm deep) and varying vegetation thickness.

These variables consisted of spatial data at 30×30 m resolution, which we derived from a 10 m digital elevation model and map of vegetation types sourced from the City of Cape Town (https://web1.capetown.gov.za/web1/opendataportal/AllDatasets).

We used the vegetation data, excluding transformed land (we classified land as "transformed" if none of the natural vegetation types currently occur), to determine the boundaries of the study area. We aggregated the digital elevation model to 30 m resolution and used it to calculate slope, topographic position and two components of aspect (north–south = cosine of aspect in degrees; east–west = sine of aspect in degrees). These layers were used to calculate total solar radiation received in January and July using the "r.sun" function in GRASS (https://grass.osgeo.org/grass72/manuals/r.sun.html). Mean maximum January and mean minimum July temperature surfaces were provided by the South African Environmental Observation Network (SAEON). These were interpolated by applying a Bayesian kriging approach (Finley et al., 2015) to data from a network of 100 loggers sampling temperature variation across the range in slope, aspect, elevation, and distance to sea across the peninsula (Slingsby & Ackerly, unpublished).

We fitted species distribution models using MaxEnt (Phillips et al., 2004), which has a predictive accuracy typically superior to other methods (Merow et al., 2013; Pearson et al., 2007), especially with few available occurrence records and no reliable absence data (Hernandez et al., 2006; Wisz et al., 2008), as is the case with *C. rosei*. MaxEnt has been used to aid the discovery of new occupied sites or areas of occurrence for several rare and threatened species (Fois et al., 2015; Le Lay et al., 2010; Mizsei et al., 2016; Rebelo & Jones, 2010), with as few as five occurrence records for model fitting (Oleas et al., 2014).

We used MaxEnt's default settings (autofeatures setting in MaxEnt version 3.3.3k, October 2011), which are appropriate for most purposes (Phillips & Dudik, 2008; Warren & Seifert, 2011).

We used "bootstrap replicated run type" with maximum iterations set to 5000.

As an initial test of the predictive accuracy of our model with the limited data available, we fitted five replicates of each model, and included some of the records for model fitting and others for model testing: we used Random Seed setting, choosing 80% of records for fitting and 20% for testing with each iteration. We then used the resulting mean area under the [receiver–operator] curve (AUC) values to select the best model. AUC is a measure of the model's predictive accuracy, based only on the ranking of test locations (Merow et al., 2013). In this case, the random 20% of the occurrence records excluded from model fitting were contrasted against randomly selected background points as pseudo-absences. We initially fitted a model with all variables,

then removed variables systematically from lowest to highest relative variable importance. Where variable pairs were correlated with $r > 0.8$, we removed the less important one (according to permutation importance from MaxEnt output) because collinearity may affect model predictions (Merow et al., 2013).

Stratification

We used the logistic output from the species distribution model with the highest AUC score as a measure of habitat suitability (H_i). We fitted species distribution models at 30 m resolution, because the habitat patches (i.e., ephemeral seepages) occupied by amphibians on this landscape are often small and we were concerned that suitable habitat would be missed by the model if we adopted a coarser scale.

After model fitting, we aggregated the 30×30 m grid cells into 300×300 m grid cells as potential sites to search for biological and field-logistical reasons. We expected the frogs to have an activity radius of a few hundred meters, and so could be found in the vicinity of suitable sites, and we were confident that we could thoroughly search grid cells of 300×300 m, therefore maximizing the searched area for the fixed number of grid cells we could visit. We used function aggregate in package *raster* (Hijmans, 2019) in program R (R Core Team, 2019) to construct the 300 m stratification landscape, aggregating by maximum H_i.

We then used stratified random sampling to select 45 grid cells to be searched. We randomly selected nine grid cells out of those with predicted $H_i \leq 0.95$ and $H_i > 0.85$ (there were no grid cells with $H_i > 0.95$), eight with $H_i \leq 0.85$ and > 0.75, seven with $H_i \leq 0.75$ and > 0.65, and so on. This procedure strongly biased the inclusion probability toward grid cells with high suitability, but included some grid cells with low suitability.

Additional sites to search

The species distribution model (SDM) provides objective estimates of habitat suitability, but its predictive ability is imperfect, due to several sources of error: (1) lack of habitat variation due to few input locations; (2) the model may not include all environmental variables/layers that are important for the species; (3) error in the environmental variables/layers used; (4) error in the species locality data; (5) SDM may not be fitted at the most appropriate spatial scale. In this situation, it may make sense to search extra locations that appear suitable based on less easily quantifiable criteria. Furthermore, searches for rare species often involve some experts, which may have a sense of what constitutes a suitable site for the target species. Expert knowledge is often the only criterion used for guiding searches, and rare species searches may be dependent on the buy-in from such experts. Using only sites selected by an SDM to guide searches discards such potentially useful information, and may disenfranchise interested parties on which the research is dependent. We therefore suggest including expert opinion in site selection. Such sites would nonetheless contribute data for testing SDM performance, while not wasting effort on sites where the absence of the target species is near certain.

In our case, the criteria for choosing such a site included: (1) the presence of shallow, ephemeral seepages for breeding; (2) less dense vegetation as *C. rosei* is only known to breed in pools that are in small clearings; and (3) the abundance of *Arthroleptella lightfooti* and/or *Strongylopus bonaespei* calls, as these two species were the most abundant proxy species at both known active breeding sites. The abundance of *Amietia fuscigula* was also sometimes a good indicator of shallow seepage, but these were usually more abundant in overgrown sites, which may not be suitable for *C. rosei* breeding. We included data from a total of 96 grid cells: 44 chosen by random-stratified sampling; 25 near or at historic breeding sites (often a good starting point, as these sites contained suitable habitat, at least historically); and 27 chosen by expert opinion, due to presence of suitable seepages.

Step 3: Perform field searches

We thoroughly searched the selected sites by walking a series of parallel line transects from the one edge of the grid cell to the other and using search methods designed to maximize detection of *C. rosei*. Expert knowledge and previous studies (Becker et al., 2018; Cressey et al., 2015; S. Edwards et al., 2017)) showed that *C. rosei* is most detectable during the breeding season when adults are more active and aggregate densely in the small, open breeding puddles. The species can also easily be found shortly after the breeding season when the dense clusters of eggs and/or tadpoles are visible. We therefore focused on locating breeding pools, and carried out the searches on foot from when spawn first became evident (17 August 2015) until the last tadpoles metamorphosed at the known occupied sites (7 October 2015). We searched suitable pools and seepages for *C. rosei* breeding particularly carefully (see Study area and study species), but also inspected deeper pools and streams, and dry areas. We conducted searches during the day and used a high-powered torch (225 lumen, LED Lenser) to further increase the visibility of small amphibian adults, eggs, and larvae.

Step 4: Searching for habitat proxy species

We expected to find few or no new *C. rosei* occurrences on our search. We therefore also searched for more common species that we regarded as habitat proxy species. From these species, we expected to gain additional information on whether the variables, modeling method, and the spatial scale modeled at (30 × 30 m, scaled up to 300 × 300 m), were appropriate for predicting *C. rosei* occurrence. We chose four anuran species as habitat proxy species: *Amietia fuscigula*, *Arthroleptella lightfooti*, *Strongylopus bonaespei*, and *Strongylopus grayii*. These species commonly co-occur with *C. rosei* on a macrohabitat scale. We suggest choosing at least one highly similar, or several relatively similar proxy species for this purpose. We measured niche similarity between the proxy-species SDMs and the *C. rosei* SDM using Schoener's D and Hellinger's I statistics (Warren et al., 2008), using the package *dismo* (Hijmans et al., 2017), program R (R Core Team, 2019). We also tested for statistically significant niche convergence or divergence (based on the D and I statistics) between the proxy species and *C. rosei* using the background similarity test (Warren et al., 2008) in the package *phyloclim* (Heibl & Calenge, 2018) in program R.

Step 5: Estimate detection probability and learn about the observation process

We searched the 300 × 300 m grid cells where we knew *C. rosei* occurred three times each, using the standard search protocol (see Step 3), including additional presence grid cells discovered during our search (five grid cells total). These included sites with low expected detection probability due to dense vegetation and low apparent density of toads (indicated by distance from central breeding site and the absence of breeding pools). We then estimated the detection probability per 300 × 300 m grid cell, p_i, as the proportion of surveys during which we detected *C. rosei*, and calculated the confidence interval of this estimate using profile likelihood (Data S1: [DS1EstimatingProfileLikelihoodOfDetectionProbabilityPerGridCell.R]) in program R 3.6.1 (R Core Team, 2019).

We further attempted to learn about the observation process from examining detections of the proxy species. We detected the presence of proxy species in three main ways: eggs/tadpoles, adults/subadults, and calls. Because *C. rosei* does not call (Grandison, 1980), we calculated the decrease in p_i that resulted when call detections were omitted for each proxy species, as: the number of searched grid cells where the species was detected by call only, divided by the total number of searched grid cells where the species was detected.

The variation in complex life histories of these proxy species was expected to influence detection properties considerably: *A. lightfooti* does not have free-living tadpoles, but breeds terrestrially (Rose, 1950); *A. fuscigula* may remain in the tadpole life stage for more than a year, while *S. bonaespei* and *S. grayii* tadpoles develop much more rapidly; *S. bonaespei* appeared to call much less frequently after July/early August, and fewer tadpoles were observed by late September, while the other species appeared to call consistently throughout the search period; *A. fuscigula* usually calls infrequently; *S. grayii* calls relatively frequently and *A. lightfooti* calls consistently throughout the day.

Step 6: Evaluate the species distribution model

The surveys provide a new sample of data on which the performance of the SDM can be tested. However, with no or few new detections such a test may not be particularly informative. We therefore also tested the ability of the model to predict presence/absence of two indicators for suitable *C. rosei* habitat.

The first indicator was whether ≥ 5% of a 300 × 300 m searched grid cell contained shallow seepage habitat similar to known *C. rosei* breeding puddles (*shallow wetland*). The second measure was the presence of two out of the three proxy species: *A. fuscigula*, *A. lightfooti*, and *S. bonaespei* (2/3 *closest spp.*), in high enough abundances to be detected on a single search. We did not consider the fourth proxy species, *S. grayii*, which turned out to have less similar habitat preferences than the other three. We chose this as a direct measure of *C. rosei* habitat occurrence because two of these species were always present in considerable abundance at all current and historic *C. rosei* breeding grid cells, and in most suitable-looking *C. rosei* breeding habitat. These three species were fairly similar in terms of predicted habitat suitability to *C. rosei* (see Table 1).

As a further test of the predictive ability of SDMs in our system, we fitted separate models to each of the proxy anuran species using occurrence records from several databases (see Appendix S1: Table S2), and using the same approach and initial variable set as for *C. rosei* (see Step 2). We then tested the models' predictive performance for these proxy species on the independent presence/absence data set from the searches.

As a measure of predictive ability of each SDM, we calculated the AUC using the package *AUC* (Ballings & Van den Poel, 2013) and manipulated the spatial files

Proxy species/measure	AUC	n	D	I
Amietia fuscigula	0.84	72	0.531	0.814
Arthroleptella lightfooti	0.82	72	0.566	0.842
Strongylopus bonaespei	0.85	72	0.552	0.842
Strongylopus grayii	0.66	72	0.425	0.737
2/3 *closest spp.*	0.81	72	0.587	0.862
Shallow wetland	0.82	86		

using the package *raster* (Hijmans, 2019) in program R (R Core Team, 2019).

This analysis and all subsequent analyses involving habitat suitability were based on 300×300 m grid cells using the mean of all 30×30 m grid cells contained within it.

We based our calculation of AUC on the 300×300 m resolution grid cells because the validation data were collected at that resolution for biological and logistical reasons (see Step 3 above).

Step 7: Estimate the probability that all extant occupied sites have been found

After completing the surveys, we calculated whether we found all occupied sites that existed in the search area, that is, an estimate of the probability that no further unknown occupied sites existed $P_{N_A=0}$ and by how much our search had increased that probability. We carried out this calculation as follows: if we knew the probability of each grid cell i to be occupied, Ψ_i, before we started the surveys, we could calculate $P_{N_A=0} = \prod_{i=1}^{K} (1 - \Psi_i)$ for the K grid cells in which the species had not been found. Grid cells in which the species was found were assumed to have a $\Psi_i = 1$, and were therefore excluded from the area of interest, A. We do not have Ψ_i but we had estimates of habitat suitability H_i from the SDM, which are proportional to Ψ_i if (1) the detection probabilities p_i are constant, (2) if there is no spatial sampling bias and (3) if

the SDM successfully captures variation in H_i (Guillera-Arroita et al., 2015). In our case, p_i appeared to be close to 1 despite varying conditions. Spatial sampling bias in historical presence records is also likely to be small as the potential range of *C. rosei* was small and accessible. In addition, we thoroughly checked the SDM (see Step 6 above) to increase our confidence in its ability to estimate H_i.

If it is reasonable to assume that $\Psi_i \propto H_i$, the next step is to estimate a conversion factor c so that $\Psi_i = c \times H_i$. We can get an estimate of c based on the expected values of the two distributions, $E(\Psi) = c \times E(H)$, that is, $c = \frac{E(\Psi)}{E(H)}$. We calculated $E(H)$ as the mean of the H_i values predicted by the SDM. We did not have a direct estimate of $E(\Psi)$, but we were able to give a lower and an upper bound on this quantity. We know that the species occurs in two grid cells. Assuming there is at least one additional unknown occupied site, the minimum for the mean occupancy probability is $E(\Psi)_{min} = \frac{3}{n}$ where n is the total number of grid cells in the study area. Given the known decline of the species (Cressey et al., 2015) and its small potential range which is relatively well explored, we think it unlikely that more than seven total occupied sites remain (half the maximum number of breeding sites ever recorded). We therefore used $E(\Psi)_{max} = \frac{7}{n}$ as an upper bound and calculated $P_{N_A=0}$ using these two values for $E(\Psi)$. For the entire study area, $n = 3{,}883$ excluding cells without data, which were in the ocean or city; $E(H) = 0.0388$; $E(\Psi)_{min} = 0.000772$; $E(\Psi)_{max} = 0.00180$; $c_{min} = 0.0199$; $c_{max} = 0.0464$. For the top of Table Mountain (a smaller area of interest): $n = 296$; $E(H) = 0.1046$; $E(\Psi)_{min} = 0.00208$; $E(\Psi)_{max} = 0.00485$ (no c calculated; rather, expected values of Ψ were calculated as the mean of the Ψ_i values for the Table Mountain area).

For the grid cells that we included in the search, the outcome of the survey changed Ψ_i. If the species is found in a grid cell, then the postsurvey occupancy probability $\Psi_i' = 1$ for that grid cell is excluded when calculating $P_{N_A=0}$. For a grid cell where the species is not found, the postsurvey occupancy probability can be calculated using Bayes' theorem:

$$\Psi_i' = \frac{\Psi_i \times (1 - p_i)}{\Psi_i \times (1 - p_i) + (1 - \Psi_i)}$$

where p_i is the probability of detecting the species if it is present (Peterson & Bayley, 2004; Wintle et al., 2012). We assumed p_i to be constant across all searched grid cells and used the estimate obtained in Step 5, above. For grid cells that we did not search, $\Psi_i' = \Psi_i$, as we obtained no new information on occupancy in these grid cells.

We then calculated the probability that all extant occupied sites had been found after the search as

$P'_{N_A=0} = \prod_{i=1}^{K}(1 - \Psi'_i)$. We calculated the factor by which our search increased the probability that all occupied sites are found, $\frac{P_{N_A=0}}{P'_{N_A=0}}$, for (1) the entire potential range of *C. rosei* and (2) the top of Table Mountain.

RESULTS

Search results

We searched 96 grid cells in total: 44 chosen by random-stratified sampling; 25 near or at historic breeding sites; 27 were searched either during this study due to presence of suitable wetlands, or inadvertently searched shortly before this study (see Figure 1b). We discovered one new *C. rosei* breeding site, within ∼3 km of a previously known breeding site (Figure 1a). The proxy species were all much more common and widespread in the study area than *C. rosei*. Out of the 72 grid cells adequately searched for all the proxy species *Amietia fuscigula* was observed in 29, *Arthroleptella lightfooti* in 45, *Strongylopus bonaespei* in 20, and *Strongylopus grayii* in 20 grid cells.

Detection probability

Multiple *C. rosei* adults, eggs, and tadpoles were detected on all 15 searches of the five known-presence grid cells, some of which were only discovered during this search and included a few grid cells adjacent to the main breeding sites. Assuming that the detection process follows a binomial distribution, the maximum likelihood estimate for the detection probability was 1. The 95% confidence interval was 0.88 to 1, based on profile likelihood.
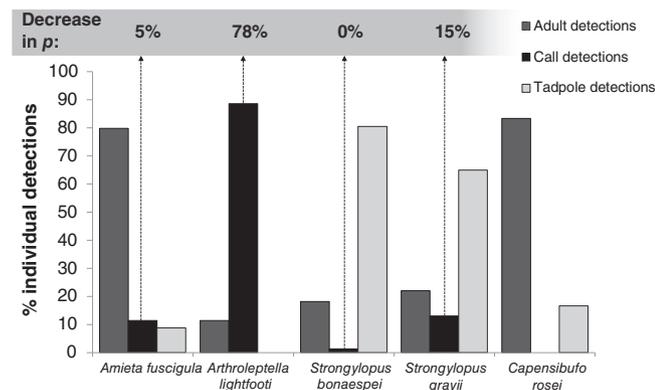
Examining the detections of the proxy species, we found that calls were the dominant form of detection for *A. lightfooti*, but constituted a much smaller percentage of detections for *A. fuscigula*, *S. grayii*, and *S. bonaespei*; eggs/tadpoles were the dominant detection mode for both *S. bonaespei* and *S. grayii*, while adult/subadult detections were dominant for *C. rosei* and *A. fuscigula* (Figure 2). The omission of call detections resulted in a 0.78 reduction in detection probability per searched grid cell of *A. lightfooti*, but was much less severe for the other species: 0.15 for *S. grayii*, 0.05 for *A. fuscigula* and no effect on *S. bonaespei* detection (Figure 2).

Species distribution model predictive accuracy

The final chosen SDM for *C. rosei* showed good model fit using the presence-only data, with AUC = 0.974. The proxy-species SDMs also fitted well to the available data and all preliminary (presence-only) AUC values were >0.8 (Appendix S2: Table S1). The SDM fitted to the *C. rosei* data predicted high suitability, H_i, only for a small proportion of the study area (Figures 1b and 3). Median H_i was 0.016. The newly discovered breeding site was found in a grid cell with H_i of 0.080. This site was chosen based on expert opinion: the site contained open, shallow wetland, and many *S. bonaespei* and *A. lightfooti* were heard calling there. The site was next to the road, and was identified en route from the southernmost breeding site during a standard monitoring visit.

The models fitted to three of the proxy species and both habitat measures had a high out-of-sample
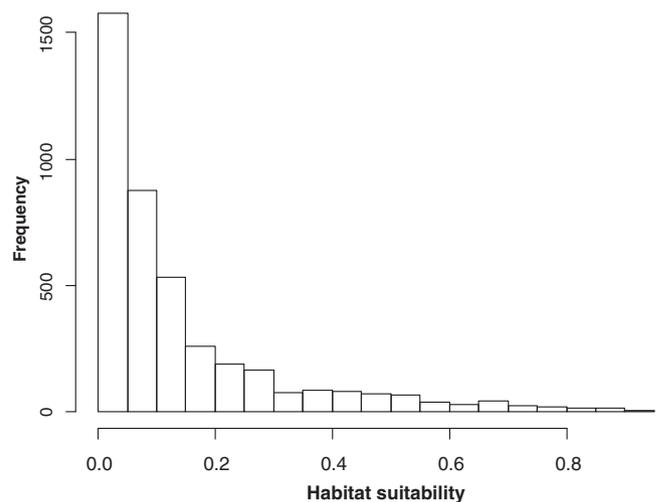

FIGURE 2 Percentage of total individual detections (NOT *n* sites detected) per detection mode, for each of the four proxy species and *Capensibufo rosei*. The percentage reduction in detection probability per searched site (p_i) that resulted from excluding all call detections from the search, are also displayed per species

FIGURE 3 Frequency distribution of SDM-predicted habitat suitability (H_i) of all cells for *Capensibufo rosei* across the study area

TABLE 2 Probability of no more unknown occupied sites of *C. rosei* occurring ($P_{N_A=0}$) in the study area and on Table Mountain, after the search, based on two prevalence scenarios ($E(\Psi) = \frac{3}{n}$ and $E(\Psi) = \frac{7}{n}$); the factor by which the search has improved $P_{N_A=0}$

Area covered	Value	$E(\Psi) = \frac{3}{n}$	$E(\Psi) = \frac{7}{n}$
Study area	$P_{N_A=0}$	0.072	0.002
	$P_{N_A=0}$ factor improved	1.416	2.258
Table Mountain	$P_{N_A=0}$	0.623	0.331
	$P_{N_A=0}$ factor improved	1.155	1.403

predictive accuracy (AUC > 0.8; Table 1). For one further proxy species (*S. grayii*), the SDM showed only moderate predictive accuracy (Table 1). This species was also the least similar to *C. rosei* in terms of predicted habitat suitability, with similarities from greatest to least being (Table 1): *A. lightfooti*, *S. bonaespei*, *A. fuscigula* and *S. grayii*. According to the background similarity tests, none of the proxy species' predicted niches were significantly more or less similar to *C. rosei* than expected by chance.

Probability that all extant occupied sites are known

The probability that we discovered all extant occupied sites in the searched area, $P_{N_{A,searched}=0}$, was 1 (95% confidence interval = $0.920 - 1$ or $0.965 - 1$, using the lower and upper bound on the estimate of $\Psi^{'}$). For the entire area, across all grid cells where *C. rosei* remained undetected, $P_{N_A=0}$ remained low (Table 2). Our search improved the certainty that all occupied sites had been found by a factor of ~ 1.8, dependent on $\Psi^{'}$ (Table 2). The estimated probability that no more occupied sites occurred on Table Mountain was ~ 0.5 after the search, which improved by a factor of ~ 1.3 compared with before the search (Table 2).

Guiding the next search

Our search has improved our knowledge in several ways that could guide further searches. First, the newly discovered occupied site broadened our knowledge of *C. rosei*'s preferred habitat as we found it in a location for which our SDM predicted relatively low habitat suitability. Moreover, the observed habitat at several historic locations, which were often imprecise, did not contain suitable-looking habitats (shallow wetland for *C. rosei* breeding), whereas some nearby areas did. We used a location in these nearby wetland patches, in addition to the new occupied site, to re-fit the SDM, that is, using improved knowledge of the habitat from our search.

The habitat suitability per grid cell of our newly fitted SDM based on these methods, was similar to the initial SDM ($r = 0.85$), but did highlight several additional priority areas to be searched (Figure 1c). The new SDM showed an increased predictive performance in the searched grid cells, for both proxy measures of *C. rosei* habitat: 2/3 *closest spp.* AUC = 0.827 (was 0.818) and *shallow wetland* AUC = 0.844 (was 0.825).

DISCUSSION

Rare species tend to be particularly important for conservation. They also tend to be difficult to assess quantitatively because data on them tend to be sparse (Thompson, 2004). We needed to know whether any undiscovered occupied sites of the rare Rose's mountain toadlet, *C. rosei*, remained in addition to the two occupied sites that were known at the beginning of the study. We followed a seven-step procedure designed to make the best use of the information we could get—both on *C. rosei* and auxiliary information—to guide a search and quantify the probability that no further occupied site exists.

We used a combination of H_i from the SDMs and visible criteria based on expert opinion, to guide site selection. This approach was useful for maintaining the involvement of experts, and may be particularly useful in the early phases of the search, when few locations are available to fit the SDM. The value of including expert opinion in selecting sites here resulted in the discovery of an additional *C. rosei* breeding site (Figure 1a).

Even though we concentrated our search on areas that were most likely to be occupied by *C. rosei*, the probability of at least one undiscovered *C. rosei* occupied site remaining within its potential range, $P_{N_A=0}$, was still high due to the proportionally small area that we could search (see Figure 1b). Our approach for using a SDM to estimate occupancy probabilities was, however, useful for assessing the search results in smaller areas of importance. We contend that *C. rosei* is most likely locally extinct from Table Mountain. This was a historic stronghold for the species in a world-renowned tourist

destination of conservation importance (Picker & Samways, 1996; Standish et al., 2004).

Suitability of the species distribution model

The SDM predicted many more low-suitability sites than high-suitability sites (Figure 3). Given the limited number of sites we could search in total, we felt that it was unlikely that we could learn much about the model from focusing a considerable part of our effort on the low-suitability sites. However, we did spend some effort on sites of low predicted suitability, to explore the possibility of uncovering substantial occupied sites in areas thought to be unsuitable (as in McDonald, 2004).

We checked the fitted SDM for its ability to predict known current and historic occurrences, and for its ability to predict apparently suitable habitats, despite the absence of *C. rosei*. In this case, this was indicated by the presence/absence of shallow wetlands, and the concurrent presence/absence of at least two of the three most similar (in terms of predicted H_i) habitat proxy species.

The *C. rosei* SDM had a high discriminatory capacity with an AUC value (0.97) that is often interpreted as indicative of very high performance. However, high values of AUC do not guarantee that the model describes variation in habitat suitability well, especially with the large number of grid cells with a low habitat suitability across the landscape that we observed (see Figure 3, Lobo et al., 2008; Jiménez-Valverde, 2012). The single new occupied site we found was in a grid cell that had a relatively low predicted suitability of 0.08, even though this was considerably higher than the median predicted suitability. Due to the extreme rarity of *C. rosei*, its naturally overdispersed distribution and/or poor dispersal ability (Tolley et al., 2010; Channing et al., 2017), and in light of recent enigmatic declines (Cressey et al., 2015), it is expected that most sites predicted as suitable will be unoccupied. The presence/absence of shallow seepages and the concurrent presence/absence of the two most similar habitat proxy species, indicated that the SDM predicted key habitats relatively well (see Table 1: *shallow wetland*; 2/3 *closest spp.*). However, the discovery of *C. rosei* in a site of low predicted H_i indicated that patches of good habitat were missed by our SDM, and substantiates the notion of including sites chosen by other criteria.

We also used proxy species to obtain a sense of how well the modeling approach worked in our environment. The high out-of-sample predictive ability—judged by AUC—of the models for the three proxy species with the most similar habitat requirements as *C. rosei* (Table 1)

suggested that the habitat variables and scale we used were appropriate for predicting wetland anuran occurrence in the study area. As none of the proxy species' predicted niches were significantly more similar to *C. rosei* than expected by chance, these contributions to directly testing the *C. rosei* SDM's accuracy, were limited.

The choice of the type of SDM for our context is an important consideration. We chose to use MaxEnt (Phillips & Dudík, 2008) which has some limitations (Dorazio, 2012; Yackulic et al., 2013), some of which are shared by other methods that use presence-background data (Merow et al., 2013). The key role of the SDM in our workflow was to provide a best possible description of the variation in habitat suitability across the study area, given the available data. Other model choices may be more sensible in different situations. For example, we would have preferred to use methods that estimated occupancy probabilities directly while accounting for the observation process (e.g., MacKenzie et al., 2017). With only two known occupied sites and no prior data on detection probability, however, this was not possible.

Detection probability

We never failed to find the species during our repeat visits to sites we knew were occupied, suggesting a high detection probability for searches carried out during the breeding season and the tadpole season. There is a chance that detection probabilities at sites of known occurrence are higher than at other sites, either because the observer is unconsciously guided by prior knowledge of where to find individuals or because the known occupied sites may be the larger and denser occupied sites that were more readily discovered in the first place. We always started our searches on the outskirts of the breeding sites, where *C. rosei* densities were lower and vegetation was often thicker, which would reduce detection probability. Despite this, we always detected multiple adults, eggs, or tadpoles within minutes on each repeat-search. The newly discovered breeding site was also spotted within a few minutes of the first visit to that grid cell. We are confident that the high detection probabilities are real.

We further examined the detections of the proxy species to assess how the lack of an advertisement call in *C. rosei* might affect its detection. Call detections were disproportionately important for the detection of *A. lightfooti* due to its small size, cryptic behavior and terrestrial breeding (Rose, 1950). By contrast, the omission of call detections had little impact on us finding the other proxy species (see Figure 2). These species were typically detected either as tadpoles or as subadults, while

searching through puddles and wetland patches. Where a proxy species was only detected by call in a particular site, this was usually the result of a very low density (or absence) of either tadpoles or subadults, or because adults were calling from deeper pools, which were not as closely inspected (not ideal microhabitat for *C. rosei*). Due to the aggregate breeding habits of *C. rosei*, we find it unlikely that very low density populations would persist, and therefore be missed. The proxy-species data also indicated that detection probabilities and properties affecting them can be highly species specific, as reported by several other studies (Bailey et al., 2004; Kroll et al., 2008; Haynes et al., 2013; Smith et al., 2014). Closely related species may have more similar detection properties to the target species, and would be useful for "borrowing" detection data. The lack of a well studied congener species in this case may have reduced the utility of borrowing detection probability data.

The detection process consists of multiple components related to suitable habitat being available, animals being present in the habitat at the time of the survey, animals being detectable at the time of the survey, and detectable animals actually being detected (Nichols et al., 2009). Our estimated detection probability describes the last component, but it may often be useful or even necessary to collect data that allow one to estimate (some of) these components separately (e.g., Amundson et al., 2014). By concentrating our searches on a time window when some life stages (eggs, tadpoles) were confined to small puddles, we ensured that the species was present and detectable. The process of detecting suitable habitat can be broken down into two components. First, one needs to know what constitutes suitable habitat for the species in question and second, one needs to detect this habitat when it is present. While suitable-looking habitats were easy to detect in our relatively small grid cells that could be searched thoroughly, the notion of what constitutes suitable habitat was necessarily based on the relatively few sites where this species was found currently and historically. The rarer a species, the less is generally known about the suite of habitats in which that species can persist. For such species, habitat suitability needs to be periodically re-evaluated as new information becomes available.

Habitat suitability and occupancy probability

To estimate the probability that no undiscovered occupied site remains, $P_{N_A=0}$, we needed an estimate of the probability that individual sites were occupied, Ψ_i. Ideally, these site-specific occupancy probabilities should be estimated directly, for example, using occupancy models

(MacKenzie et al., 2017). Occupancy probabilities are difficult to estimate for rare species for which few data are available, even though recent advances in data integration allow for ever more efficient use of data (Koshkina et al., 2017; Miller et al., 2019; Renner et al., 2019).

With only two known breeding sites, each represented by one or two occupied sites (300×300 m) at the start of our study, we did not have enough data to estimate occupancy probabilities directly. Instead, we used methods that estimate a relative measure of habitat suitability, H_i, based on all the information we could obtain about *C. rosei* occurrence. We assumed that these habitat-suitability estimates were proportional to occupancy probabilities (Guillera-Arroita et al., 2015) and used upper and lower bounds on prevalence as an estimate of mean occupancy across the entire range to convert H_i to Ψ_i. The proportionality assumption could be relaxed somewhat if multiple estimates of occupancy probabilities could be obtained, either for some of the sites or for groups of sites. In that case, the distribution of H_i could be matched more closely to the distribution of Ψ_i, for example, by moment matching, so that both mean and variance would be preserved.

A next step would be to explore whether the data we collected during these surveys would be sufficient to support a SDM that integrated these data with the presence-background data that were available at the start of our study and would yield direct estimates of Ψ_i (Koshkina et al., 2017).

Are there any undiscovered *C. rosei* occupied sites left?

Our searches almost doubled the estimated probability that no further undiscovered occupied sites existed, which was nevertheless still low after the search ($P_{N_A=0} \sim 0.04$, see Table 2). The probability that at least one undiscovered occupied site remained, is therefore still high. With the large number of sites that had low predicted habitat suitability, the possibility of further existing occupied sites across the entire potential range of this species will be difficult to rule out. However, the method can be used to estimate the expected effectiveness of different search strategies in increasing $P_{N_A=0}$ and so guide searches.

We view the search for very rare species as an iterative process guided by the approach we suggest here. The next step would be to select the next set of sites to be searched, guided by the updated habitat-suitability estimates (see Figure 1c), and based on priorities. For instance, one reasonable next step would be to focus

search efforts on sites for which estimated habitat suitability increased most after refitting. This would direct the search to include the new information gained (on additional suitable habitat) from the previous search. The anticipated gain from different search options could be compared in terms of the expected change in $P_{N_A=0}$ one could achieve.

CONCLUSIONS

We present a systematic approach aimed at estimating the probability that no undiscovered occupied site remained in an area. While the presence of undiscovered occupied sites remains difficult to rule out across large areas, our approach is most useful in smaller target areas for particular management interventions. In this case, it appears that *C. rosei* is locally extinct from Table Mountain, a historic stronghold for this species (Cressey et al., 2015). Our method may also be used to identify potential sites for species re-introductions, based on predicted suitable habitat and probable local extinctions, provided the initial cause of extinction is being managed. Furthermore, the spatial pattern of high versus very low predicted habitat suitability (Figure 1b), suggested that the remaining suitable habitat for *C. rosei* is rare and fragmented. Natural recolonization of these habitat patches may be highly unlikely.

In contrast with previous methods, our approach takes into account multiple sources of uncertainty when estimating the probability that all occupied sites of a rare species are known: uneven distribution, detection probability, modeling accuracy, and available ecological information.

It also makes efficient use of limited occurrence and observational data, all that is normally available for very rare species that are hard to sample (Thompson, 2004).

Searches for rare species regularly fail, and it is unknown whether such a result is evidence of true absence, or if it is worth spending more resources on increased search effort.

Using our method (and the resulting metrics) it is possible to make objective decisions about whether or not to repeat searches, and what is the relative likelihood of achieving the desired outcome. Our approach can help to quantify the amount of resources and money to be spent on searches proportionate to the rarity of the target species. As an added benefit, our approach also allows the gathering of potentially new and important data on non-target species, through targeted spending on rare species.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data (Altwegg et al., 2021) are available on ZivaHub at https://doi.org/10.25375/uct.14609838.v1.

ORCID
Francois S. Becker https://orcid.org/0000-0003-3874-9183
Jasper A. Slingsby https://orcid.org/0000-0003-1246-1181
John Measey https://orcid.org/0000-0001-9939-7615
Krystal A. Tolley https://orcid.org/0000-0002-7778-1963
Res Altwegg https://orcid.org/0000-0002-4083-6561

REFERENCES
Altwegg, R., J. Slingsby, F. Becker, J. Measey and K. Tolley. 2021. "Finding Rare Species and Estimating the Probability that all Occupied Sites Have Been Found: Dataset." University of Cape Town, Dataset.

Amundson, C.L., J.A. Royle, and C.M. Handel. 2014. "A Hierarchical Model Combining Distance Sampling and Time Removal to Estimate Detection Probability during Avian Point Counts." *The Auk: Ornithological Advances* 131: 476–94.

Bailey, L.L., T.R. Simons, and K.H. Pollock. 2004. "Spatial and Temporal Variation in Detection Probability of Plethodon Salamanders Using the Robust Capture-Recapture Design." *The Journal of Wildlife Management* 68: 14–24.

Ballings, M. and D. Van den Poel. 2013. "AUC: Threshold Independent Performance Measures for Probabilistic Classifiers." R package version 0.3.0. https://CRAN.R-project.org/package=AUC.

Becker, F. S. 2017. "Estimating the Population Size of a Rare and Elusive Species: The Case of Rose's Mountain Toadlet." MSc thesis, University of Cape Town.

Becker, F.S., K.A. Tolley, G.J. Measey, and R. Altwegg. 2018. "Extreme Climate-Induced Life-History Plasticity in an Amphibian." *The American Naturalist* 191: 250–8. https://doi.org/10.1086/695315

Channing, A., G.J. Measey, A.L. De Villiers, A.A. Turner, and K.A. Tolley. 2017. "Taxonomy of the *Capensibufo Rosei* Group

(Anura: Bufonidae) from South Africa." *Zootaxa* 4232(2): 282–92. https://doi.org/10.11646/zootaxa.4232.2.11

Cowling, R.M., I.A.W. Macdonald, and M.T. Simmons. 1996. "The Cape Peninsula, South Africa: Physiographical, Biological and Historical Background to an Extraordinary Hot-Spot of Biodiversity." *Biodiversity and Conservation* 5: 527–50.

Cressey, E.R., G.J. Measey, and K.A. Tolley. 2015. "Fading Out of View: The Enigmatic Decline of Rose's Mountain Toad *Capensibufo Rosei*." *Oryx* 49(3): 521–8. https://doi.org/10.1017/S0030605313001051

Dorazio, R.M. 2012. "Predicting the Geographic Distribution of a Species from Presence-Only Data Subject to Detection Errors." *Biometrics* 68: 1303–12.

Edwards, T.C., D.R. Cutler, N.E. Zimmermann, L. Geiser, and J. Alegria. 2005. "Model-Based Stratifications for Enhancing the Detection of Rare Ecological Events." *Ecology* 86: 1081–90.

Edwards, S., K.A. Tolley, and G.J. Measey. 2017. "Habitat Characteristics Influence the Breeding of Rose's Dwarf Mountain Toadlet *Capensibufo Rosei* (Anura: Bufonidae)." *Herpetological Journal* 27: 287–98.

Finley, A.O., S. Banerjee, and A.E. Gelfand. 2015. "spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models." *Journal of Statistical Software* 63: 1–28.

Fois, M., G. Fenu, A. Cuena Lombraña, D. Cogoni, and G. Bacchetta. 2015. "A Practical Method to Speed up the Discovery of Unknown Populations Using Species Distribution Models." *Journal for Nature Conservation* 24: 42–8.

Grandison, A. 1980. "A New Genus of Toad (Anura: Bufonidae) from the Republic of South Africa with Remarks on its Relationships." *Bulletin of the British Museum (Natural History) Zoology* 39: 293–8. https://doi.org/10.5962/p.13275

Guillera-Arroita, G. 2017. "Modelling of Species Distributions, Range Dynamics and Communities under Imperfect Detection: Advances, Challenges and Opportunities." *Ecography* 40: 281–95.

Guillera-Arroita, G., J.J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P.E. Lentini, M.A. McCarthy, R. Tingley, and B.A. Wintle. 2015. "Is My Species Distribution Model Fit for Purpose? Matching Data and Models to Applications." *Global Ecology and Biogeography* 24: 276–92.

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: 501–11.

Haynes, T.B., A.E. Rosenberger, M.S. Lindberg, M. Whitman, and J. A. Schmutz. 2013. "Method- and Species-Specific Detection Probabilities of Fish Occupancy in Arctic Lakes: Implications for Design and Management." *Canadian Journal of Fisheries and Aquatic Sciences* 1062: 1055–62.

Heibl, C. and C. Calenge. 2018. "phyloclim: Integrating Phylogenetics and Climatic Niche Modeling." R package version 0.9.5. https://CRAN.R-project.org/package=phyloclim

Hernandez, P.A., C.H. Graham, L.L. Master, D.L. Albert, and A.D. L. The. 2006. "The Effect of Sample Size and Species Characteristics on Performance of Different Species Distribution Modeling Methods." *Ecography* 29: 773–85.

Hijmans, R. J. 2019. "raster: Geographic Data Analysis and Modeling." R package version 2.8-19. https://CRAN.R-project.org/package=raster.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. "dismo: Species Distribution Modeling." R package version 1.1-4. https://CRAN.R-project.org/package=dismo.

Jiménez-Valverde, A. 2012. "Insights into the Area under the Receiver Operating Characteristic Curve (Auc) as a Discrimination Measure in Species Distribution Modelling." *Global Ecology and Biogeography* 21: 498–507.

Koshkina, V., Y. Wang, A. Gordon, R.M. Dorazio, M. White, and L. Stone. 2017. "Integrated Species Distribution Models: Combining Presence-Background Data and Site-Occupancy Data with Imperfect Detection." *Methods in Ecology and Evolution* 8: 420–30.

Kroll, A.J., K. Risenhoover, T. McBride, E. Beach, B.J. Kernohan, J. Light, and J. Bach. 2008. "Factors Influencing Stream Occupancy and Detection Probability Parameters of Stream-Associated Amphibians in Commercial Forests of Oregon and Washington, USA." *Forest Ecology and Management* 255: 3726–35.

Le Lay, G., R. Engler, E. Franc, and A. Guisan. 2010. "Prospective Sampling Based on Model Ensembles Improves the Detection of Rare Species." *Ecography* 33: 1015–27.

Lobo, J.M., A. Jiménez-valverde, and R. Real. 2008. "AUC: A Misleading Measure of the Performance of Predictive Distribution Models." *Global Ecology and Biogeography* 17: 145–51.

MacKenzie, D.I., J.D. Nichols, G.B. Lachman, S. Droege, J.A. Royle, and C.A. Langtimm. 2002. "Estimating Site Occupancy Rates When Detection Probabilities Are Less than One." *Ecology* 83: 2248–55.

MacKenzie, D.I., J.D. Nichols, N. Sutton, K. Kawanishi, and L.L. Bailey. 2005. "Improving Inferences in Popoulation Studies of Rare Species That Are Detected Imperfectly." *Ecology* 86: 1101–13.

MacKenzie, D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, L. Bailey, and J.E. Hines. 2017. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence.* London, UK: Elsevier.

McCune, J.L. 2016. "Species Distribution Models Predict Rare Species Occurrences Despite Significant Effects of Landscape Context." *Journal of Applied Ecology* 53: 1871–9.

McDonald, L.L. 2004. "Sampling Rare Populations." In *Sampling Rare Orelusive Species*, edited by W.L. Thompson, 11–42. Washington, DC: Island Press.

Measey, G.J. 2011. *Ensuring a Future for South Africa's Frogs: A Strategy for Conservation Research. Biodiversity Series 19.* Pretoria: SANBI. https://www.sanbi.org/wp-content/uploads/2018/04/biodiver19bfrogs.pdf

Measey, G.J., B.C. Stevenson, T. Scott, R. Altwegg, and D.L. Borchers. 2017. "Counting Chirps: Acoustic Monitoring of Cryptic Frogs." *Journal of Applied Ecology* 54: 894–902. https://doi.org/10.1111/1365-2664.12810

Merow, C., M.J. Smith, and J.A. Silander. 2013. "A Practical Guide to MaxEnt for Modeling species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36: 1058–69.

Miller, D.A., K. Pacifici, J.S. Sanderlin, and B.J. Reich. 2019. "The Recent Past and Promising Future for Data Integration Methods to Estimate Species Distributions." *Methods in Ecology and Evolution* 10: 22–37.

Mizsei, E., B. Üveges, B. Vági, M. Szabolcs, S. Lengyel, W.P. Pfliegler, Z.T. Nagy, and J.P. Tóth. 2016. "Species Distribution

Modelling Leads to the Discovery of New Populations of One of the Least Known European Snakes, Vipera Ursinii Graeca, in Albania." *Amphibia-Reptilia* 37: 55–68.

Nichols, J.D., J.E. Hines, J.R. Sauer, F.W. Fallon, J.E. Fallon, and P. J. Heglund. 2000. "A Double-Observer Approach for Estimating Detection Probability and Abundance from Point Counts." *Auk* 117: 393–408.

Nichols, J.D., L. Thomas, and P.B. Conn. 2009. "Inferences about Landbird Abundance from Count Data: Recent Advances and Future Directions." In *Modeling Demographic Processes in Marked Populations* 201–35. New York, NY: Springer.

Oleas, N.H., A.W. Meerow, K.J. Feeley, J. Gebelein, and J. Francisco-Ortega. 2014. "Using Species Distribution Models as a Tool to Discover New Populations of Phaedranassa Brevifolia Meerow, 1987 (Liliopsida: Amaryllidaceae) in Northern Ecuador." *Check List* 10: 689–91.

Pearson, R.G., C.J. Raxworthy, M. Nakamura, and A. Townsend Peterson. 2007. "Predicting Species Distributions from Small Numbers of Occurrence Records: A Test Case Using Cryptic Geckos in Madagascar." *Journal of Biogeography* 34: 102–17.

Peterson, J.T., and P.B. Bayley. 2004. "A Bayesian Approach to Estimating Presence When a Species Is Undetected." In *Sampling Rare or Elusive Species*, edited by W.L. Thompson, 173–88. Washington, DC: Island Press.

Phillips, S.J., and M. Dudik. 2008. "Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation." *Ecography* 31: 161–75.

Phillips, S.J., M. Dudik, and R.E. Shapire. 2004. "A Maximum Entropy Approach to Species Distribution Modeling." In *Proceedings of the Twenty-First International Conference on Machine Learning* 655–62. New York, NY: Association for Computing Machinery.

Picker, M., and M. Samways. 1996. "Faunal Diversity and Endemicity of the Cape Peninsula, South Africa a First Assessment." *Biodiversity and Conservation* 5: 591–606.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/

Rebelo, H., and G. Jones. 2010. "Ground Validation of Presence-Only Modelling with Rare Species: A Case Study on Barbastelles Barbastella Barbastellus (Chiroptera: Vespertilionidae)." *Journal of Applied Ecology* 47: 410–20.

Renner, I. W., J. Louvrier, and O. Gimenez. 2019. "Combining Multiple Data Sources in Species Distribution Models While Accounting for Spatial Dependence and Overfitting with Combined Penalised Likelihood Maximisation." *bioRxiv* 615583.

Rose, W. 1950. *The Reptiles and Amphibians of Southern Africa*. Cape Town, South Africa: Maskew Miller.

Roubicek, A.J., J. VanDerWal, L.J. Beaumont, A.J. Pitman, P. Wilson, and L. Hughes. 2010. "Does the Choice of Climate Baseline Matter in Ecological Niche Modelling?" *Ecological Modelling* 221: 2280–6. https://doi.org/10.1016/j.ecolmodel.2010.06.021

Royle, J.A., R.B. Chandler, C. Yackulic, and J.D. Nichols. 2012. "Likelihood Analysis of Species Occurrence Probability from Presence-Only Data for Modelling Species Distributions." *Methods in Ecology and Evolution* 3: 545–54.

SA-FRoG IUCN SSC Amphibian Specialist Group. 2010. "Capensibufo rosei." https://doi.org/10.2305/IUCN.UK.2010-3.RLTS.T3776A10070549.en .

Shirose, L.J., C.A. Bishop, D.M. Green, J. Cameron, R.J. Brooks, N. J. Helferty, L.J. Shirose, et al. 1997. "Validation Tests of an Amphibian Call Count Survey Technique in Ontario, Canada." *Herpetologica* 53: 312–20.

Smith, D.H.V., B. Jones, L. Randall, and D.R.C. Prescott. 2014. "Difference in Detection and Occupancy between Two Anurans: The Importance of Species-Specific Monitoring." *Herpetological Conservation and Biology* 9: 267–77.

Specht, H.M., H.T. Reich, F. Iannarilli, M.R. Edwards, S.P. Stapleton, M.D. Weegman, M.K. Johnson, B.J. Yohannes, and T.W. Arnold. 2017. "Occupancy Surveys with Conditional Replicates: An Alternative Sampling Design for Rare Species." *Methods in Ecology and Evolution* 8: 1725–34.

Standish, B., A. Boting, H. Van Zyl, T. Leiman, and J. Turpie. 2004. "The Economic Contribution of Table Mountain National Park." Cape Town, South Africa: South African National Parks.

Thompson, W. 2004. *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters*. Washington, DC: Island Press.

Tolley, K.A., A.L.D.E. Villiers, M.I. Cherry, and G.J. Measey. 2010. "Isolation and High Genetic Diversity in Dwarf Mountain Toads (Capensibufo) from South Africa." *Biological Journal of the Linnean Society* 100: 822–34.

Warren, D.L., and S.N. Seifert. 2011. "Ecological Niche Modeling in Maxent: The Importance of Model Complexity and the Performance of Model Selection Criteria." *Ecological Applications* 21: 335–42.

Warren, D.L., R.E. Glor, and M. Turelli. 2008. "Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution." *Evolution* 62: 2868–83.

Wintle, B.A., T.V. Walshe, K.M. Parris, and M.A. Mccarthy. 2012. "Designing Occupancy Surveys and Interpreting Non-detection When Observations Are Imperfect." *Diversity and Distributions* 18: 417–24. https://doi.org/10.1111/j.1472-4642.2011.00874.x

Wisz, M.S., R.J. Hijmans, J. Li, A.T. Peterson, C.H. Graham, A. Guisan, J. Elith, et al. 2008. "Effects of Sample Size on the Performance of Species Distribution Models." *Diversity and Distributions* 14: 763–73.

Yackulic, C.B., R. Chandler, E.F. Zipkin, J.A. Royle, J.D. Nichols, E. H. Campbell Grant, and S. Veran. 2013. "Presence-Only Modelling Using Maxent: When Can We Trust the Inferences?" *Methods in Ecology and Evolution* 4: 236–43.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.