Refining area of occupancy to address the modifiable areal unit problem in ecology and conservation

Justin Moat[1,2]*

Steven P. Bachman[1,2]

Richard Field[2]

Doreen S. Boyd[2]

[1]Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK

[2]School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK

*Corresponding author: J.Moat@kew.org

Running header

Refining AOO for extinction risk

Keywords

IUCN Red List, conservation assessments, point data, raster, species distribution, threatened species, Modifiable Areal Unit Problem (MAUP)

Article impact statement

Improvements to minimum-area calculations for IUCN Red List assessments is achieved through rotation and movement of grids.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cobi.13139.

Abstract

The 'modifiable areal unit problem' is prevalent across many aspects of spatial analysis within ecology and conservation. The problem is particularly manifest when calculating metrics for extinction risk estimation, for example, area of occupancy (AOO). Although embedded into the International Union for the Conservation of Nature (IUCN) Red List criteria, AOO is often not used or is poorly applied. Here we evaluate new and existing methods for calculating AOO from occurrence records and present a method for determining the minimum AOO using a uniform grid. We evaluate the grid cell shape, grid origin and grid rotation with both real-world and simulated data, reviewing the effects on AOO values, and possible impacts for species already assessed on the IUCN Red List.

We show that AOO can vary by up to 80% and a ratio of cells to points of 1:1.21 gives the maximum variation in the number of occupied cells. These findings potentially impact 3% of existing species on the IUCN Red List, as well as species not yet assessed. We show that a new method that combines both grid rotation and moving grid origin gives fast, robust and reproducible results and, in the majority of cases, achieves the minimum AOO. As well as reporting minimum AOO, we outline a confidence interval which should be incorporated into existing tools that support species risk assessment. We also make further recommendations for reporting AOO and other areal measurements within ecology, leading to more robust methods for future species risk assessment.

Introduction

The modifiable areal unit problem (MAUP) in conservation and ecology

Any operation using a set of shapes (e.g. grid cells) to aggregate data yields different results depending on size, shape and positioning of those objects. The 'modifiable areal unit problem'

(MAUP; Gehlke & Biehl 1934; Openshaw 1984) is well known in the field of spatial analysis: when artificial reporting units are imposed on continuous geographic phenomena, or on finer-resolution spatial units, part of the resulting spatial pattern or outcome is an artefact of the size (scale MAUP) and shape (zone MAUP) of the reporting units (Heywood et al. 2011). Ecology and conservation deploy such operations commonly, for example in species richness of grid cells or other polygons, environmental data layers created by aggregating or resampling pixel data from remotely sensed imagery, and area of occupancy (AOO) estimated from occurrence records of species. However, the MAUP is rarely acknowledged and in particular, the shape (zone) MAUP has not been fully explored. Failing to address this problem, especially in the context of assessing species and ecosystem extinction risk, has fundamental implications for conservation planning, policy and management.

MAUP and the IUCN Red List(s)

The IUCN Red List of Threatened Species (RLS) (IUCN 2012) is a globally important tool for assessing extinction risk of species. A complementary system has now been developed for ecosystems: the IUCN Red List of Ecosystems (Keith et al. 2013) [RLE]. Both systems employ metrics associated with geographic range that can be affected by the MAUP. The core process of assessing a species' extinction risk is its assignment to one of nine categories based on five criteria: population decline, geographic distribution, small population size, restricted populations and quantitative analysis of extinction risk (IUCN 2012). If thresholds are met for any one of these criteria, it is sufficient to justify a threat category. Criterion B relates to geographic distribution and is the most heavily utilised criterion (~50% of all assessment; Collen *et al.* 2016), reflecting the considerable, and increasing, availability of species occurrence data (e.g. provided by the Global Biodiversity Information Facility).

Either 'extent of occurrence' (EOO) or 'area of occupancy' (AOO), or both, can be used to determine level of threat under Criterion B. For species and ecosystem assessments, AOO must be estimated as less than threshold values (for species: 10, 500 and 2,000 km² and for ecosystems: 200, 2,000 and 50,000 km^2 respectively for Critically Endangered, Endangered and Vulnerable species, see Supporting information; Table S1), in conjunction with known decline, limited number of locations, fragmentation, or fluctuation of range or population size. In addition, declines in estimated AOO or EOO are sufficient for listing under the same three threat categories under Criterion A, and AOO can also be applied to criterion D2. Thus, estimating AOO and EOO are fundamental to the Red Listing process. The use of EOO has recently been discussed (Joppa et al. 2016), but little attention has been given to ensuring consistency in the approaches used for estimating AOO. To ensure ongoing rigour in Red List assessments we address this knowledge gap.

AOO is the geographic range occupied by the species within its EOO (Gaston 1991, 1994; Gaston & Fuller 2009), excluding cases of vagrancy, at a particular scale (IUCN 2012). It is based on known occurrences (usually points), using a repeatable methodology that superimposes a grid, counts the number of cells occupied and, in the case of species RLS assessments, sums their areas RLE assessments use the number of cells). Errors in AOO estimation can occur when sampling intensity is low, or the taxon has low detection probability (IUCN Standards and Petitions Subcommittee 2016). Another major influence on estimating AOO is scale of grid size, where different grid sizes can under- or over-estimate AOO (Willis et al. 2003). Scale in AOO estimation (an example of the scale MAUP) has been explored and documented elsewhere, resulting in various recommendations ranging from up- and down-scaling to applying different scales depending on the threats (Hui 2009; Keil et al. 2013; Marsh 2016; Keith et al. 2017). Choice of cell size influences AOO estimation and IUCN neutralises this problem by recommending a standard cell size (2 × 2 km cells for the RLS (IUCN 2012) and 10 × 10 km cells for the RLE (Keith et al. 2013)). Further flexibility in estimating AOO for the IUCN criteria is allowed with the application of habitat maps, although maps should be rescaled to the IUCN reference scales (IUCN Standards and Petitions Subcommittee 2016).

MAUP shape/zone

An important consideration that has received little attention so far is whether grid- or cell-based estimates of AOO establish the true minimum value for AOO given the data and scale used. In line with the precautionary principle, the IUCN guidelines (IUCN Standards and Petitions Subcommittee 2016 p. 48) state that "If different grid locations (starting points of the grid) result in different AOO estimates, the minimum estimate should be used", but the consequences of this have not been fully explored. Existing tools such as GeoCAT (Bachman et al. 2011), the equivalent package in R 'rCAT' (Moat 2017), R packages 'ConR' (Gilles et al. 2016) and 'red' (Cardoso 2017), and ArcView extension CATS (Moat 2007), use simple AOO algorithms that can return different results depending on the placement of the grid (Supporting information; Figure S1). Consequently, AOO may frequently be over-estimated and lead to inappropriate (optimistic) Red List assessments. This risks diverting conservation resources away from species or ecosystems most needing them. Keith et al (2017) acknowledged the grid origin problem and adopted an algorithm where grid origin is shifted and minimum AOO is reported when the lowest AOO is not bettered after a number of user defined rounds, see also R package 'redlistr' (Lee & Murray 2017). They noted AOO could be 73% larger or 63% smaller than a mean value (Keith et al 2017). Further work is needed to fully address the impact of shifting grid origin as well as other elements of the zone/shape MAUP such as grid cell shape and grid orientation.

With free and easy-to-use tools for performing AOO estimation, combined with availability of spatial species data, AOO may be applied more frequently. Thus it is important to test different ways of estimating AOO, and provide guidance on how best to achieve the minimum AOO estimate. Our aim here is to develop a fast algorithm that varies both grid cell shape, grid orientation and grid origin with descriptive, reproducible results, whilst achieving minimum AOO most of the time, and accords with the rules of the Red List criteria (IUCN Standards and Petitions Subcommittee 2016). Further, we estimate the number of species globally that may be listed under inappropriate threat categories because of the shortcomings of existing AOO estimation methodologies. This should have the impact of seeing AOO more widely used and rigorously documented in Red List assessments.

Methods

Overview: we built a simulation environment using the software R (R Core Team 2016), to experimentally test different variables such as grid cell shape (hexagons and squares), grid origin and grid orientation. We applied novel algorithms, developed here, to simulated and real-world test data to calculate AOO, identifying the most efficient (processing time) and appropriate method (lowest AOO).

Data

We used five different datasets of occurrence records, two of which were simulated (see supporting information figure S2), and three real-world (Madagascar, Africa, and Caribbean):

- **Square set of random points,** on a simple 0–10 in X and Y co-ordinate system (simulated): small, simple data set for testing multiple iterations and algorithms.
- **Oval set of random points**, with area of 100 units, orientated at 45°: more realistic simulated point dataset, without the obvious geometric constraints of the square set.
- **Legumes of Madagascar.** 19,343 georeferenced records for 761 taxa, with species having from just one record up to 228 records (*Baudouinia fluggeiformis*), constrained to the island (Du Puy et al. 2002).

- ***Coffea* species of Africa.** 2,606 records for 58 taxa, covering a wide range of tropical/sub-tropical environments across many countries (Davis et al. 2006). Includes *Coffea arabica* (395 occurrence records).

- **Selected Caribbean endemics.** 899 records from 10 taxa. A set of Red List-assessed plant species useful for reviewing the outcomes from the different algorithms on a group of very range-restricted, well-sampled taxa. Used to assess possible implications in terms of changes in threat status (Burton 2008; Royal Botanic Gardens Kew 2016).

All the real world datasets had been cleaned removing erroneous or highly inaccurately localities, these datasets were reprojected to the cylindrical equal area projection, in metres, based on the distribution of all of the point data, using the projection wizard in 'rCAT' (Moat 2017).

AOO algorithms

There is little guidance on the geometric needs of AOO measurements for RLS, other than to specify a regularly spaced, equal-area grid with a common origin (i.e. no individual cells floating, with different origins) (Resit Ackacaya, pers. comm. 2017) and a steer towards reporting in terms of 4 km^2 cells. We varied grid cell shape, origin and rotation to test for an optimal approach.

Cell shape: squares and hexagons. The most common cell and survey shape used in ecology is squares (Wheater et al. 2011), followed by circles (often used in forestry; de Vries, 1986) and rectangles. Circles do not tessellate, thus 'waste' area at their margins (21% more area than hexagons; Kershner, 1939) and rectangles are too variable: it would be possible to choose long, thin rectangles (transect-shaped) providing very small and unrealistic AOO for species with few occurrences. Squares give a simple and standardised fit, with straightforward mathematics, and represent the standard method used by many. Squares also match raster imagery's regular grids, which are often used to produce species distribution models (SDM), as well as habitat and vegetation maps. Hexagons are increasingly being used to describe and quantify distributions, their main advantages being to reduce edge/corner effects and having identical distances to neighbours (Birch et al. 2007). Hexagons were calculated with the R packages 'sp' (Pebesma & Bivand 2005) and 'lattice' (Sarkar 2008).

Grid origin: fixed, brute force and optimized. First, we tested the very simple default 'fixed origin' method, where the grid starts at 0,0 (using whatever projection is applied) (Moat 2007; Bachman et al. 2011; Gilles et al. 2016; Cardoso 2017). Second, we tested a brute-force 'movable origin' method that iteratively moves the origin using a fixed number of positions, 1024 here (Schmidt *et al.*, 2017 used a simplified version of this method). Finally, we tested a 'movable origin (optimized)' method designed to produce the smallest possible AOO from a given set of points, from all possible positions of the origin (see Table 1).

Grid rotation: NSEW, brute force and optimised. In mapping/cartography, grids are predominately orientated in north–south/east–west (NSEW), a convention that mainly prevails for aesthetic reasons. We implemented three rotation methods for calculating AOO. Firstly, we tested NSEW - the grids were not rotated from their original orientation. Secondly, the 'multiple rotation (brute force)' method rotated the grid through 1024 iterations. Finally, the 'multiple rotation (optimised)' was designed to give the rotation that minimised the estimated AOO.

To keep results comparable and processing time reasonable, we restricted the number of iterations for the brute force methods to 1024, except for simulations that combined rotation and moving origin which was 1152 iterations (summarised in Table 1; examples shown in Figure 1).

Experimental design

We tested how occupancy density (the balance of cell size and the number of occurrence points) affected the outcome and range of values for AOO calculation. To do this, we ran the 'Movable origin (brute force)' algorithm with varying cell sizes and numbers of points, using 1024 combinations (iterations). For each combination, we recorded the average standard deviation.

To compare processing speeds of each algorithm, we recorded the time taken using 'proc.time' in the core R package (R Core Team 2016). Each method was timed using the same PC (Windows 7 PC, with Xeon 3.3 GHz processor, 8 cores, 64 Gb Ram and a SSD hard disk). We ran all methods except the hexagons across a 5 × 5 grid, using 1–4000 points. As the optimal algorithms run exponentially, we curtailed these once the processing time exceeded 60 seconds.

We first tested the difference between AOO shape (squares vs hexagons) using a restricted set of algorithms, leading us to eliminate hexagons from further analysis. We then ran all the algorithms using square cells, for all the datasets, to test a wide range of point/cell densities. For both the oval and square sets of points we ran the algorithms with 10, 30 and 80 points across an area of 25 cells, representing sparse, medium (with maximum variability) and saturated sets of points, respectively. We ran a further set with 120 points with an area of 100 cells, to check that scale was not affecting results. For each of the simulated datasets, we ran 100 randomly generated datasets. For the real-world datasets, we do not have an optimal density for each (as they represent multiple species with different areas, multiple collection densities at multiple scales), but we wanted to use these to represent real-world distributions and shapes. For the legumes of Madagascar and the *Coffea* species of Africa, we used cell widths of 2000, 1000, 500 and 2 km (corresponding to 4,000,000, 1,000,000, 250,000 and 4 km^2 cell areas). For the well sampled, narrowly distributed endemic species of the Caribbean we used 2, 4, 8 and 16 km cell widths (4, 16, 64 and 256 km^2 cell areas). Our reasoning for a range of cell sizes (including RLS reference scale of 2 × 2km) was not only to test the actual IUCN recommended cell size, but to review the most optimal algorithms. For most species the sampling will be too low to allow an accurate calculation of AOO. These real-world datasets were included as they have the inherent bias and collection densities of real data, but at the cell sizes chosen we should see increasing variation in AOO values as we approach the optimal ratio of cells to points.

To determine which algorithms worked best, we calculated the number of times an algorithm produced the minimum AOO across all methods. For each, we also recorded the mean change in AOO compared to the simplest algorithm (square, fixed origin, NSEW). To further differentiate the algorithms, they were ranked 1–6 (1 = highest number of times it equalled the minimum AOO) for each dataset and scenario.

To assess the potential impact of the different algorithms on existing IUCN assessments, we reviewed the prevalence of AOO reported in the latest version of the Red List (methods detailed in Supporting Information). We queried the IUCN Red List (IUCN 2016a) for all Critically Endangered (CR), Endangered (EN), Vulnerable (VU), Near Threatened (NT) and Least Concern (LC) species (07/09/2016), using the IUCN API (IUCN 2016b) and the R package 'rredlist' (Chamberlain 2016). Where a range of values or limits was given for AOO, we used a middle value.

Results

Cell size and cell occupancy

Testing the number of points versus cell occupancy revealed that cell occupancy tended to follow a binominal distribution (Figure 2A). The upper bound is where AOO equals all available cells (saturation); the lower bound is where one point can only occupy one cell. We found maximum variability in AOO at the cell to random point ratio of 1:1.21 (e.g. 121 points in 100 cells; Figure 2B) – the ratio used within our point simulations to give the greatest variability (see Methods).

Which algorithms estimated the lowest AOO

Square cells always produced the lowest AOO estimates (Supporting information; Table S2). In the real-world examples, the minimum AOO for hexagons only approaches that for squares when cells become saturated with points, but the opposite happens for the simulated datasets, in which hexagons improve at lower occupancy densities. For grid origin and rotation algorithms, all results (Supporting information; Tables S2 and S3) are summarised in table 2, and figure 1 shows an example graphic. The main findings (percentages quoted in brackets equals the percentage of runs that the algorithm achieved the lowest AOO value) are:

Moving origin brute force vs optimized. The optimal always performed better (69%) than, or as well as, the brute force algorithm (62%).

Rotation brute force vs optimized. The brute force (60%) was as good as, in many cases better than, the optimized (59%). Optimized was only better at higher numbers of points, in which the number of iterations far exceeds the 1024 used for brute force.

Brute force, moving origin vs rotation. The moving origin gave the best results for the square set of points, but only at higher occupation densities (e.g. 80 points in 25 cells). In the other data sets, there was little difference between the two (moving origin 62% vs rotation 60%).

Combined methods: The combined brute force constantly performed well (average rank 1.8; with an average AOO reduction of 17%), often outperforming the combined optimized (average rank 1.5, average AOO with a reduction of 15%) for the average reduction in AOO. It should be noted that the optimized algorithm is restricted to running on datasets with ≤ 19 points, so this does not allow gains from the higher AOO reductions at higher numbers of points.

AOO reductions: In the real-world data, estimated AOO decreased by 3% to 32% compared with the standard method (square, fixed origin, NSEW), with smaller reductions tending to be associated with sparse occupancy. For the well-sampled Caribbean species, average AOO reduction was 22% to 28%.

Processing time

The hexagon algorithms were very slow; the processing time increases exponentially with the number of grids cells. All the square algorithms performed as expected from their number of iterations (Table 1). The processing time of the iterative algorithms increased linearly with the number of points, with the slightly more complex rotation and combined algorithms running a little slowly than the simpler moving origin (Figure 3A). The optimized algorithms became very slow with more than a few occurrence points, exceeding 60 seconds at 266, 190 and 19 points for rotation, moving origin and combined optimized algorithms respectively (Figure 3B).

Impact on existing Red List assessments

We examined 68,574 species on the IUCN Red List website (IUCN 2016a), 23% documented AOO in their extinction risk assessment (Supporting information; Table S4), but only those using the B2 criteria will have explicitly used AOO (A1c and D2 could have used AOO or other geographic metrics). It is very difficult to assess how many species use the AOO metric for their assessment, as not all species have an AOO measurement documented. We detail the results and issues in supporting information (text, table S4 and Figure S3). We estimate that 10,000–14,500 species use AOO within their assessment (15% to 21% of all species assessed). When reviewing species near thresholds (Supporting information; Figure S3), we estimated that approximately 3% of species using AOO within their assessment (~ 300 species), could be uplisted to a higher threat category (more threatened) by applying a minimum AOO algorithm.

Discussion

To-date there has been inconsistency in the method used to calculate the AOO of a species. Here we have shown that the estimated AOO was substantially reduced (by as much as 40%, or even more for small extents, e.g. 80%, from 5 cells to 1 cell) by modifying the origin and rotation of the grid system, compared with the commonly applied algorithm. Reviewing already assessed species we estimate that about 300 species may have been placed in a less-threatened category than they should have during the IUCN Red List assessment procedure. The new algorithm that we propose (combined moving origin and rotational brute force) should overcome this issue, providing more accurate AOO estimates and running quickly on standard computers. The proposed algorithm can also be applied to analogous operations which involve area estimation from point data.

AOO Algorithms

In isolation, changing rotation (average rank 3.6) and changing origin (average rank 3.5) had similar effects on AOO estimation, except that in some of the real-world data, rotation produced a small advantage. This is likely due to the rounding of latitudes and longitude (to the nearest minute or decimal place), which would place more points along the latitude and longitude axes, where a diamond square (square rotated 45° from NSEW), would have more 'reach' to gather points. Combining rotation and moving origin gave the best results (average rank of 1.8), even with the same number of iterations. As expected, the optimized algorithm for the moving origin always gave better or the same results as the brute force algorithm, and we can be confident that this is the optimal solution for moving origin. We applied the same logic to optimizing rotation, rotating the grid to all the possible point pairs, but we found that this did no better than the brute force solution. This suggests that our optimization for rotation can be improved; this would benefit from further research. So, would the issue of the quality of input data (in our case the rounding and accuracy of point data) and its influence on AOO and EOO.

Given unlimited time and processing power, combining optimizing algorithms for moving origin and rotation should produce optimal results. However, for any algorithm to be useful, it should be responsive (run quickly), robust and reproducible. The optimizing algorithms for rotation and moving origin quickly become untenable, even with small numbers of points, producing a vast number of iterations (1 million computations for 38 points in the combined optimization). We quantitatively compared results for datasets with relatively low numbers of points. For those, the combined iterative ('brute force') approach with 1152 iterations achieved the minimum AOO 83% of the time across all our test data and 96% for the real-world datasets (Supporting information, Table S3), and was almost always very close. This is despite the large shortfall in number of iterations (1152 for brute force vs. up to the equivalent of 65,000 iterations for optimal). The Caribbean dataset was the only dataset with sufficiently well-sampled point data to fully test the influence of algorithm choice on real AOO values, the combined brute force algorithm with 1152 iterations only achieved the minimum AOO for 50% of the taxa, compared to the optimal algorithms. To further investigate this using a fairer test, we randomly placed 10 points in 8 cells and ran both the optimization algorithm (which required the equivalent of 4500 iterations) and the combined brute force algorithm with 4500 iterations. We ran this equal-effort procedure 100 times; the brute force algorithm achieved minimum AOO 98% of the time, compared to 84% for the optimization algorithm.

We therefore recommend using the combined brute-force algorithm, with minimum of 1152 iterations, as a good compromise between time taken and quality of results. If minimum AOO is critical (particularly where a taxon is close to a threshold), then we recommend an increase in iterations – and suggest the upper limit is would approach 100,000. This approach has the additional advantage that AOO can be fully quantified by giving the spread of AOO values (mean, mode and maximum) as well as the minimum, in line with IUCN guidelines to document uncertainty (IUCN Standards and Petitions Subcommittee 2016).

Hexagon grid shapes never out-performed squares and ran slowly with even small numbers of cells. Square cells have the advantage of requiring very simple mathematics, being highly applicable to ground surveys and being directly analogous to raster data. For the application of minimum AOO, square cells are a clear winner.

For our simulations, we have assumed that these data represent well-sampled taxa or ecosystems. This assumption will make little/no difference for our testing but will be critical for actual species assessment using AOO. It is in these cases that the impact of previous calculations of AOO (or their non-usage) is particularly important. Overlooking the modifiable areal unit problem has important implications for the conservation of many species.

Impact of findings on the IUCN Red List of Species

We estimated from documented AOO values on the IUCN website that 3% of species may be affected by improving AOO estimation methodology. This is a reasonable, if slightly high, estimate because additional sub-criteria such as number of locations also need to be fulfilled to warrant a category change (IUCN Standards and Petitions Subcommittee 2016). Of greater significance is the impact AOO estimation methods will have on the unknown, but very large number of species, yet to be assessed using the IUCN criteria. For our well-sampled Caribbean species, applying the new algorithms reduces estimated AOO by 22% to 28%, giving some indication of the likely implications of applying the changed methodology to the present Red List for species with AOO of about 250 km^2 or less (which would qualify for the EN or CR categories). It would seem prudent to give a range of values for AOO (as practically, at the thresholds, they are very sensitive to small changes in number of cells which would lead to changes in extinction risk assessment). Such a range allows the user to determine how far a threshold has been exceeded. We also suggest that automated techniques, such as described here, can be used as a first pass and then updated/adjusted by experts, particularly if close to a threshold.

To demonstrate how critical the AOO methods of measurement can be, we applied the combined brute force algorithm to the Caribbean plant species Spermacoce capillaris (Figure 4). This species is endemic to the Turks and Caicos Islands, has a very small geographic range (both AOO and EOO), and is experiencing continued decline due to habitat loss/degradation. It is assessed as Endangered under B1 and B2 criteria (Endangered B1ab (ii,iii,v)+2ab (ii,iii,v)) (Barrios & Manco 2014). Its documented range (AOO), calculated using GeoCAT (Bachman et al. 2011), is 16–52 km^2, where the higher estimate uses both inferred and observed occurrences of the species (total = 19) and the lower only uses the observed (and verified) records. Applying the combined brute force algorithm (Figure 4) reduces the lower AOO to 12 km^2 (25% lower) and the higher one to 36 km^2 (33% lower than the documented value). In this example, the reduction in AOO estimate does not result in a change in the rating, but it does bring the minimum estimate much closer to the threshold for 'Critically Endangered' (10 km^2). Clearly, in many cases an equivalent change would change a threat rating.

Moreover, using the 19 observed and inferred occurrences, the range of AOO values across the iterations is 36–60 km^2 (Figure 4A). Thus, some arrangements of cells of the same size give AOO 66% higher than others, highlighting how important it is to run algorithms to minimize the AOO estimate. Equivalently, if AOO is calculated over different time points, for example to calculate change over time (RLS criteria A1-4c; IUCN 2012), the vagaries of grid cell placement could easily give a false indication of reduction or expansion of an AOO purely as an artefact of methodology.

Concluding remarks and recommendations

We have presented some very simple, yet powerful, methods and algorithms for calculating AOO based on a set of points. As computing power increases, and with improved algorithms, optimized solutions may become a more viable method for estimating AOO, which would be desirable for conservation. For at least the present, we recommend using the combined brute force method presented here. We have also shown that hexagons rarely give optimal (minimum) results and are computationally demanding; we advise against using them for AOO calculations. We tested our methods with example species (relevant to RLS) but not ecosystems (RLE); from preliminary analysis using moist Afromontane forest in Ethiopia at 30 m resolution (Supporting information, Figure S4), which shows as RLS a binomial distribution, we expect our findings to be as applicable to RLE as to RLS. We demonstrate that a small percentage (3%) of species presently assessed on the IUCN Red List may have AOO estimates that are too high. It would seem needless to reassess most of these species, but we recommend the following procedure for future application of AOO with both RLS and RLE:

1. Apply the suggest algorithm: combined moving origin and rotational brute force algorithms with a default of 1,152 iterations. This can be increased if the AOO is close to a threshold or if a minimum value AOO is critical.
2. Record summary statistics across the iterations of AOO estimates (minimum, maximum, mean, mode) and the method and grid shape used to calculate AOO, as well as the original occurrence data.
3. Record the angle of rotation and origin for the minimum AOO. This is important for reproducibility, but also for any analysis looking at temporal changes in AOO.
4. For species recorded in latitude and longitude, use an equal area projection (Moat 2017) and record the mapping projection, shift and datum used (Wieczorek et al. 2012).

Using *Spermacoce capillari*s (Figure 4) as an example, our recommendation for reporting is:

Spermacoce capillaris AOO: 12 km^2 (using 1152 iterations for verified records, minimum = 12 km^2, maximum = 28 km^2, mean = 18.8 km^2, mode = 20 km^2, method= combined moving origin and rotational brute force using square cells. For minimum AOO estimate, rotation = -78.3873 ^{o}, grid origin = (0,0), map projection = Cylindrical Equal Area with central meridian: longitude -71.51022, latitude 21.53104, Datum=WGS84.)

The method we present here has been tested with point data and, in the main, assessed for RLS, but the findings and algorithms can also be used for the Red List of Ecosystems, easily transferred to MAUP and other analogous situations (e.g. SDM and raster imagery). We have shown that our combined brute force AOO algorithm is quick and robust, and hope that our recommendations will be implemented in the IUCN RLS and RLE guidelines. We would like to encourage the increased uptake of AOO within the community and beyond.

Acknowledgments

We would like to thank to Tom Reader and Christopher Brignell from University of Nottingham for their help with some of the optimising algorithms. Our gratitude goes out to all scientists who make their data available, for this publication we are grateful to Aaron Davis for the *coffea* data of Africa and Sara Barrios for the Caribbean species data sets.

Code

All scripts in R and simulated data are freely available in the lead author's GitHub account: https://github.com/#####

Literature cited

Bachman S, Moat J, Hill AW, de Torre J, Scott B. 2011. Supporting Red List threat assessments withGeoCAT: geospatial conservation assessment tool. ZooKeys **126**:117–26.

Barrios S, Manco BN. 2014. Spermacoce capillaris. The IUCN Red List of Threatened Species 2014. Available from http://dx.doi.org/10.2305/IUCN.UK.2014-1.RLTS.T56500592A56503910.en (accessed June 13, 2017).

Birch CPD, Oom SP, Beecham JA. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological Modelling **206**:347–35.

Burton FJ. 2008. Threatened plants of the Cayman Islands : the Red List. Royal Botanic Gardens, Kew.

Cardoso P. 2017. red: IUCN Redlisting Tools. Available from https://cran.r-project.org/package=red.

Chamberlain S. 2016. rredlist: “IUCN” Red List Client. R package version 0.1.0. Available from https://cran.r-project.org/package=rredlist.

Collen B et al. 2016. Clarifying misconceptions of extinction risk assessment with the IUCN Red List. Biology letters **12**:1424–1442. Available from

Davis AP, Govaerts R, Bridson DM, Stoffelen P. 2006. An annotatated taxonomic of the genus coffea (Rubiaceae). Botanical Journal of the Linnean Society **152**:465–512. Blackwell Publishing Ltd Oxford.

de Vries PG. 1986. Sampling with Circular Plots. Pages 212–222 Sampling Theory for Forest Inventory. Springer Berlin Heidelberg, Berlin, Heidelberg.

Du Puy DJ, Labat JN, Rabevohitra R, Villiers JF, Bosser J, Moat J. 2002. The Leguminosae of Madagascar. Royal Botanic Gardens, Kew.

Gaston KJ. 1991. How Large Is a Species' Geographic Range? Oikos **61**:434–438.

Gaston KJ. 1994. Measuring geographic range sizes. Ecography **17**:198–205.

Gaston KJ, Fuller RA. 2009. The sizes of species' geographic ranges. Journal of Applied Ecology **46**:1–9.

Gehlke CE, Biehl K. 1934. Certain effects of grouping upon the size of the correlation coefficient in census tract material. Journal of the American Statistical Association **29**:196–1700.

Gilles A, Maintainer D, Dauby G. 2016. ConR: Computation of Parameters Used in Preliminary Assessment of Conservation Status. Available from https://cran.r-project.org/web/packages/ConR/ConR.pdf (accessed June 5, 2017).

Heywood DI, Cornelius S, Carver S. 2011. An introduction to geographical information systems. Prentice Hall.

Hui C. 2009. On the scaling patterns of species spatial distribution and association. Journal of Theoretical Biology **261**:481–487.

IUCN. 2012. IUCN Red List Categories and Criteria, 2nd edition. IUCN, Gland, Switzerland.

IUCN. 2016a. The IUCN Red List of Threatened Species. Available from http://www.iucnredlist.org/.

IUCN. 2016b. IUCN Red List API - V3. Available from http://apiv3.iucnredlist.org/api/v3/docs (accessed September 22, 2017).

IUCN Standards and Petitions Subcommittee. 2016. Guidelines for Using the IUCN Red List Categories and Criteria. Version 12. Available from http://www.iucnredlist.org/documents/RedListGuidelines.pdf.

Joppa LN et al. 2016. Impact of alternative metrics on estimates of extent of occurrence for extinction risk assessment. Conservation Biology **30**:362–370.

Keil P, Belmaker J, Wilson AM, Unitt P, Jetz W. 2013. Downscaling of species distribution models: A hierarchical approach. Methods in Ecology and Evolution **4**:82–94.

Keith DA et al. 2013. Scientific Foundations for an IUCN Red List of Ecosystems. PLoS ONE **8**:e62111.

Keith DA, Akçakaya HR, Murray NJ. 2017. Scaling range sizes to threats for robust predictions of risks to biodiversity. Conservation Biology:1–27.

Kershner R. 1939. The Number of Circles Covering a Set. American Journal of Mathematics **61**:665.

Lee C, Murray NJ. 2017. redlistr: Tools for the IUCN Red List of Ecosystems and Species. Available from https://cran.r-project.org/package=redlistr.

Marsh C. 2016. Package "downscale" 1.2-2. Available from https://cran.r-project.org/web/packages/downscale/index.html.

Moat JF. 2007. Conservation assessment tools extension for ArcView 3.x, version 1.2. Available from http://www.kew.org/gis/projects/cats/catsdoc.pdf.

Moat JF. 2017. rCAT: Conservation assessment tools and functions in R. Available from https://cran.r-project.org/web/packages/rCAT/index.html.

Openshaw S. 1984. The Modifiable Areal Unit Problem. Geobooks, Norwich.

Pebesma EJ, Bivand RS. 2005. Classes and methods for spatial data in R. R News 5.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.r-project.org/.

Royal Botanic Gardens Kew. 2016. UKOTs Online Herbarium. Available from http://herbaria.plants.ox.ac.uk/bol/ukot (accessed September 1, 2017).

Sarkar D. 2008. Lattice: Multivariate Data Visualization with R. Available from https://cran.r-project.org/web/packages/lattice/index.html

Schmidt M et al. 2017. Diversity, distribution and preliminary conservation status of the flora of Burkina Faso. Phytotaxa **304**:1–215.

Wheater CP, Bell JR, Cook PA. 2011. Practical field ecology: a project guide. Wiley.

Wieczorek J, Bloom D, Guralnick R, Blum S, Doring M, Giovanni R, Robertson T, Vieglais D. 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE **7**:e29715.

Willis F, Moat J, Paton A. 2003. Defining a role for herbarium data in Red List assessments : a case study of Plectranthus from eastern and southern tropical Africa. Biodiversity & Conservation **12**:1537–1552.

Tables

Table 1. Methods used to calculate AOO, number of iterations and algorithm description. * n = number of points in the test dataset or number of points representing the species. NSEW = north–south/east–west orientation.

Shape, Method	Number of iterations	Algorithm
Square, fixed origin, NSEW	1	Unique(floor(SetofPoints/CellSide))
Square, moving origin (brute force), NSEW	1024	As above moving points dx and dy. Where dx and dy = 0 to cell width with 32 steps (total iterations of 32^2 = 1024)
Square, moving origin (optimized), NSEW	n^2	As above, but where dx = difference between each point and every other point

		in dx and dy = difference between each point and every other point in y and x
Square, fixed origin, brute force rotation	1024	Rotating points between 0:1.5708 (radians) with 1024 steps
Square, fixed origin, optimized rotation	$\frac{(n^2 - \mathrm{n})}{2}$	As above, but rotations are between each point and every other point
Square, brute force origin and rotation	1152	Combining rotation and moving origin, with 6 shifts in x and y and 32 rotations (6^2 x 32 = 1152)
Square, optimized origin and rotation	$\frac{(n^4 - n^3)}{2}$	Optimal rotation and moving origin algorithms combined
Hexagon, fixed origin, NSEW	1	Build hexagons across region. Overlay points and hexagons, unique count of hexagons
Hexagon, moving origin (brute force), NSEW	1024	As above moving hexagons dx and dy. Where dx and dy = 0 to cell width with 32 steps (total iterations of 32^2 = 1024)

Table 2. Summary ranking of Algorithms lowest AOO achieved. * Ranking of Algorithms from best (1) to worst performing (5 or 6). For details see supporting information table S3. NB we allow the optimal algorithms to run up to 60 seconds, which gives the equivalent of ~ 62,000-71,000 iterations vs brute force algorithms of 1,024 or 1,152 iterations.

Data type, scenario	Number of runs	Moving origin: brute force	Rotation: brute force	Combined: brute force	Moving origin: optimal	Rotation: optimal	Combined: optimal
Simulated: square, area of 25 cells: 10 points	100	5	3	2	4	6	1
Simulated: square, area of 25 cells: 30 points	100	3	3	1	2	5	
Simulated: square, area of 25 cells: 80 points	100	2	5	3	1	4	
Simulated square, area of 100 cells: 120 points	100	5	4	1	2	3	
Simulated: oval, area of 25 cells: 10 points	100	4	5	2	3	6	1
Simulated: oval, area of 25 cells: 30 points	100	3	4	1	2	5	
Simulated: oval, area of 25 cells: 80 points	100	3	5	1	2	4	
Simulated: oval, area of 100 cells: 120 points	100	3	5	1	2	4	
Legumes of Madagascar: cell size 4 km^2	1435	5	3	2	4	6	1
Legumes of Madagascar: cell size 250,000 km^2	1435	5	3	1	3	6	2
Legumes of Madagascar: cell size 1,000,000 km^2	1435	4	4	1	3	6	2
Legumes of Madagascar: cell size 4,000,000 km^2	1435	1	1	1	1	6	1
Coffea of Africa: cell size 4 km^2	58	6	2	3	5	3	1
Coffea of Africa: cell size 250,000 km^2	58	1	6	4	1	5	3
Coffea of Africa: cell size 1,000,000 km^2	58	1	6	3	1	5	3
Coffea of Africa: cell size 4,000,000 km^2	58	1	5	1	1	6	4
Caribbean taxa: cell size 4 km^2	10	6	2	4	5	3	1
Caribbean taxa: cell size 16 km^2	10	6	3	4	5	1	1
Caribbean taxa: cell size 64 km^2	10	6	4	1	5	1	1
Caribbean taxa: cell size 256 km^2	10	5	1	1	6	1	1
Summary (total, mean)	6812	3.6	3.5	1.8	2.7	4	1.5

Figures

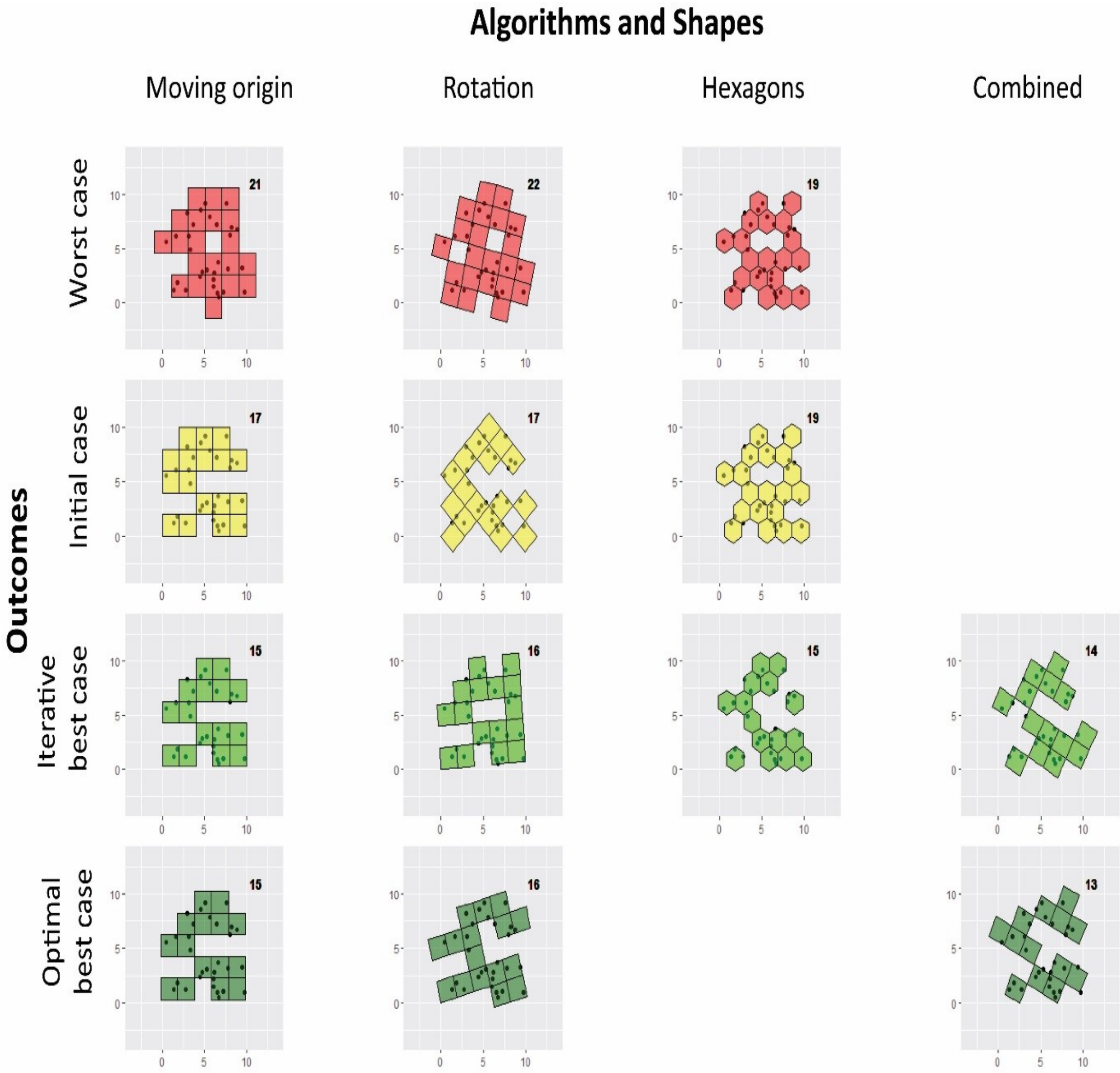

Figure 1. AOO example shapes and algorithms, using 30 random points in a square (of size 10 x 10 km in the x and y direction) and cell size of 4 km². Number of occupied cells is given in the top-right of each example. Algorithms (left hand side) are: worst case = highest number of occupied cells when running the brute force algorithm; initial case = simplest algorithm (square, fixed origin, NSEW); iterative best case = lowest number of occupied cells for the brute force algorithm; optimal best case = lowest number of occupied cells using the optimizing algorithm. Details of the algorithms are given in Table 1

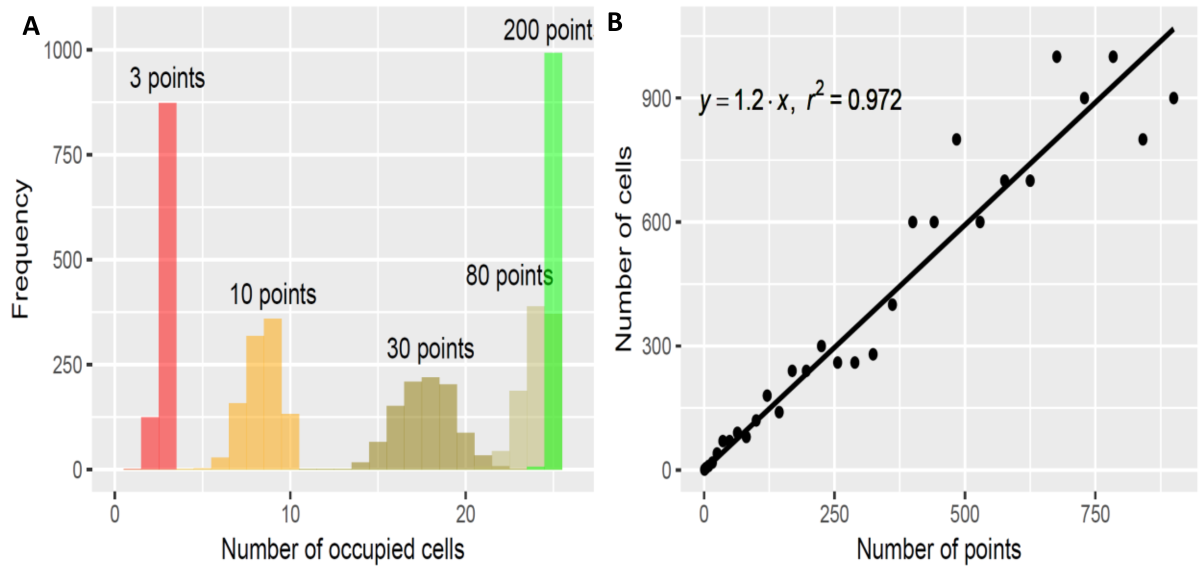

Figure 2. A. Example histograms for numbers of cells occupied, with 1024 iterations, for each of 3, 10, 30, 80 and 200 occurrence points spread over 25 cells. B. For each number of points shown, the number of cells for which the mean Standard Deviation was highest (from 100 runs of the brute force algorithm with 1024 iterations) is plotted.

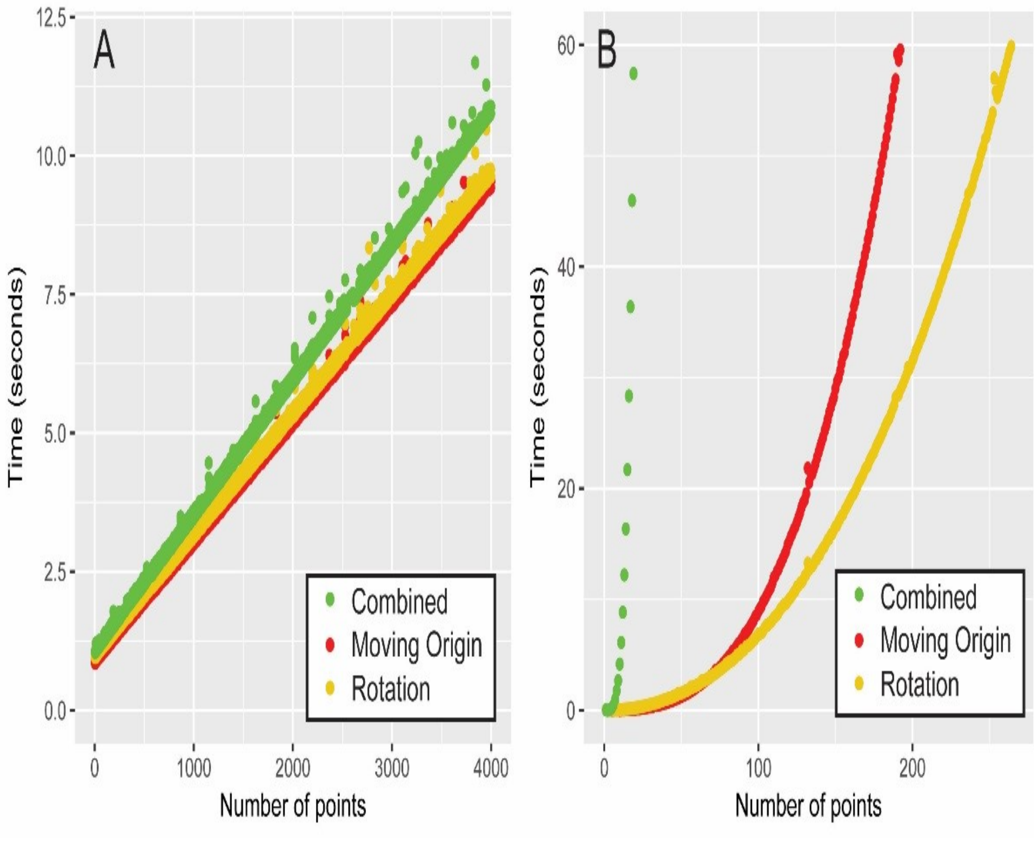

Figure 3. Algorithm time taken vs number of points in test dataset. A. Iterative 'brute force' algorithms; B. Optimizing algorithms. Note the different scales in A and B, on both axes.

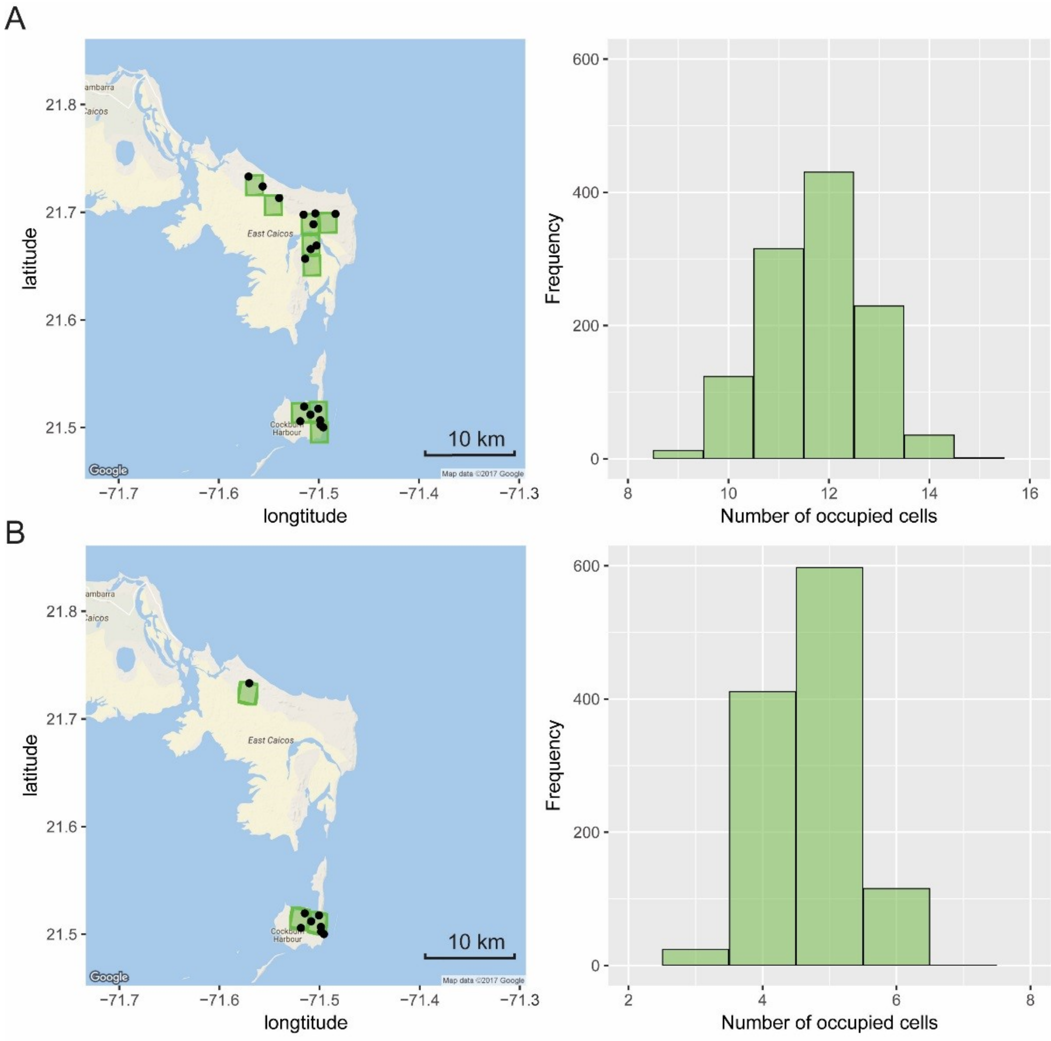

Figure 4. Maps and AOO ranges for Spermacoce capillaris*, Turks and Caicos Islands. A = All records (inferred and known), mode of 1152 iterations using combined brute force algorithm = 12 cells (48 km²), mean = 11.74 (47 km²), maximum= 15 (60 km²) and minimum = 9 (36 km²). B= All known records, mode of the 1152 iterations = 5 cells (20 km²), mean = 4.7 (19 km²), maximum= 7 (28 km²) and minimum = 3 (12 km²).*