Incorporating point location uncertainty in habitat map evaluation

Methods to generate statistics for habitat map evaluation using species point occurrences

1. For each species, assemble point location data with accuracy measures

Additionally, if there are very few or no points with accuracy measures, it would be possible to ascribe a “standard” accuracy for point locations (1000m, 5000m?) based on predefined criteria. Such a “standard” would most likely be different for different species or taxonomies (genera, families, etc…) and/or geographies and/or data sources.

1. Generate circular buffers around each point using the accuracy distance (meters) as the circle radius
2. Overlay each point buffer polygon with the species’ habitat map to generate a table showing, for each point buffer, the percentage of habitat and non-habitat cells within the buffer, e.g.:

|  |  |  |
| --- | --- | --- |
| Buffer radius (m) | % Habitat Cells | % Non-habitat Cells |
| 500 | 1.56 | 98.44 |
| 750 | 1.85 | 98.15 |
| 400 | 6.45 | 93.55 |
| 1100 | 18.23 | 81.77 |
| 5000 | 7.22 | 92.78 |
| 150 | 2.50 | 97.50 |
| 6500 | 12.78 | 87.22 |
| 500 | 0.00 | 100.00 |
| 85 | 0.00 | 100.00 |
| 50 | 0.00 | 100.00 |

Because there are no absence locations for species in this type of evaluation, the evaluation of the map as a predictor of habitat for the species across its range will need to be a comparison with a random “habitat” landscape. That is, testing whether the suitable habitat cells predicted by the habitat model and map are NOT distributed randomly with respect to species occurrences.

1. Generate a random “habitat” landscape raster with the same ratio of habitat to non-habitat cells as the species tested. For example, if species X has 2.4% of its range predicted to be habitat, generate a random raster with 2.4% 1’s (habitat) and 97.6% 0’s (non-habitat).
2. Repeat step 3 multiple times (100?) using the simulated landscape as the “habitat map”

**There are 4 metrics that could potentially be used to assess the habitat map relative to a landscape of randomly oriented habitat:**

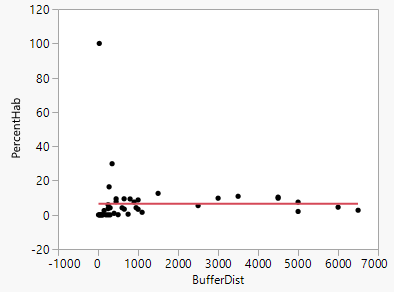
1. The percentage of point buffer locations with and without habitat cells

In the table example above, 7 of the 10 point buffers have predicted habitat cells and 3 do not. That is, 70% of point buffers have predicted habitat. The presumption here is that a random landscape will most likely produce a 50-50 ratio of point buffers with and without habitat cells. Therefore, it could be assumed, a better than random habitat map will have > 50% point buffers with habitat cells and a worse than random map will have < 50% point buffers with habitat cells.

Unfortunately, it appears the 50-50 assumption is incorrect. Using a sample of 25 simulated habitat landscapes intersected with 50 mocked-up occurrence locations, on average > 77% of point buffers of the mocked-up occurrences have simulated habitat. By comparison, a GAP generated salamander species map intersected approximately 65% of those same mocked-up point occurrence buffers. It appears to be a function of random assignment of cells across the landscape. The process evenly distributes “habitat” cells throughout the extent whereas the modeled habitat is more clumped. The dispersed nature of simulated habitat appears to make it more likely that any given location across the extent encounters at least one “habitat” cell.

1. The relationship of the percent habitat within a point buffer and radius of the point buffer’s circle

Theoretically, a random landscape of habitat cells will illustrate a relationship between the area (circle radius) of the point buffer and the percentage of habitat cells within that buffer. That is, the larger the circle radius, the greater the examined area, the higher the probability that habitat cells will overlay with that buffer. If, however, the evaluated habitat map shows NO relationship between point radius/buffer area and percent habitat cells, it could be reasonable to assume the evaluated species habitat map is non-random.



This is not true when generating a simulated landscape using the ratio of habitat to range area as a threshold for delineating habitat cells (see step 4 above). As a matter of fact, because the random raster function in ArcPy distributes cells evenly across the simulated landscape, the mean percent habitat within point buffers tends to approach the applied threshold (i.e. 2.4% from the example in step 4). That is, because the cells are uniformly present throughout the extent, any extracted portion of that extent regardless of size will have an equivalent ratio of 0’s and 1’s.

1. A comparison of the percent habitat within a point buffer to the frequency distribution of percent habitat across all random landscape simulations for each point buffer

Here the presumption is multiple iterations of random habitat landscapes will generate normally distributed frequencies (or any other continuous function) of percent habitat within each point buffer. A “better than random” habitat map would produce percent habitat values in the upper end of that simulated normal distribution, while a “worse than random” habitat map would produce percent habitat values in the lower end of the normal distribution.

Upper end of simulated landscape habitat percent distribution = “better than random”

Lower end of simulated landscape habitat percent distribution = “worse than random”

x

1. Comparison between the percent habitat in range (akin to “model prevalence” in Rondinini et al. 2011) and the mean percent habitat in point occurrence buffers

This assumes (and rightly so – see explanation of why metric 2 does not work) a random habitat landscape would, at most, predict a mean percent habitat within point occurrence buffers (regardless of size) equal to the model predicted percent habitat throughout the entire range. In fact, the greater the point occurrence inaccuracy (10,000 meters, 50,000 meters, etc…) the larger the point buffer area, the more closely representative the percent habitat within the extent given a simulated landscape of evenly distributed “habitat” cells using the species’ percent habitat in range as a threshold. Plotting mean percent habitat in point occurrence buffers vs. percent habitat in range (model prevalence) for each species, should produce a graph similar to figure 1 in Rondinini et al. 2011 where a random model would plot along the 50-50 x-y line, species models ‘worse than random’ would occur in the lower right, and species models ‘better than random’ would occur in the upper left. Note that while percent habitat in range is very similar to Rondinini et al.’s model prevalence, mean percent habitat in point buffers is NOT similar to their point prevalence measure. Rondinini et al.’s point prevalence is based on a static area of ‘influence’ for point occurrences (1km) and is simply a ratio of the number of 1km cells with at least one 300m habitat cell and the total number of point occurrences for the given species. It does not incorporate the point occurrence accuracy because it assumes all points have an accuracy of 1km.

Random

Species models 'worse than random'