The combined effects of scale and conflicting range size methods on patterns of weighted endemism.

Tayla Lawrie (44797836)

School of Biological Sciences,

The University of Queensland

St Lucia, QLD 4072

t.lawrie@uq.net.au

**Abstract**

**Aim:** The most commonly used method of calculating species’ range sizes for weighted endemism is area of occupancy (AOO). An alternative and less utilised method is extent of occurrence (EOO). As each method gives different interpretations of what it means to be ‘restricted’ as an endemic species, we see considerably different patterns of weighted endemism when each method is used. As the difference between AOO and EOO range sizes is inherently scale (i.e. grid cell size) dependent, I explore the combined effect of scale and choice of range size method on patterns of weighted endemism, and the implications this can have in biodiversity conservation.

**Methods:** Using Australian occurrence data of five Myrtaceae genera, I measured the difference between AOO and EOO ranges at four different grid cell sizes. To evaluate the combined effect of scale and choice of range size method on the differences in (1) patterns of weighted endemism, and (2) selection of endemic hotspots, I created two metrics: endemism difference (ED) and hotspot difference (HD), respectively. Both metrics were measured using 31 different grid cell sizes.

**Results:** As grid cell size increased, the percentage of species with AOO ranges larger than EOO ranges increased. Both ED and HD increased with cell size, with the magnitude and pattern of that relationship being highly variable and genera specific.

**Main Conclusions:** The impact of scale on the difference between EOO and AOO range sizes was unexpected. Contradictory to popular belief, as cell size increased, AOO became larger than, not more similar to, EOO, for many species. This range size effect likely contributed to the positive relationship seen between cell size and both ED and HD. The genera specificity and high variability of ED and HD was likely due to differences in species sensitivity to scale as well as the impacts of grid structure allocation over each species distribution. This study highlights the importance of more detailed discussion and investigation of these effects in both the weighted endemism and broader literature, ensuring that if conservation recommendations are made, we can be confident that biodiversity is being represented in the way we intended.

**Keywords:**

area of occupancy, conservation biogeography, endemic, extent of occurrence, grid cell size, range restriction, spatial patterns, spatial scale, species ranges, weighted endemism

**1 | Introduction**

**1.1 | AOO versus EOO**

Historically, the term endemic has been used to describe taxa in two distinct ways; restricted to a location or restricted in range (Anderson, 1994). Here I focus on the latter. Perhaps the most complicated issue with using range to describe endemicity, is that to be ‘restricted’ in range can mean different things depending on how the size of that range is calculated. The most well-known and commonly used method for calculating range size is area of occupancy (AOO), calculated as the number of mapped grid cells the species occupies (Figure 1; Gaston & Fuller, 2009). An alternative, less used method, is extent of occurrence (EOO), often calculated as the area of the minimum convex hull polygon around all occurrences (Figure 1; Gaston & Fuller, 2009). Depending on the distribution of occurrences, AOO and EOO can generate more different or more similar range sizes for different species (Figure 1b and 1a respectively). The use and benefits of each method on a case-specific basis has been extensively explored (Gaston & Fuller, 2009; Gaston, Quinn, Wood & Arnold, 1996). In terms of describing an endemic species, when using AOO, a more endemic species is one whose range is restricted in area (regardless of how far apart its occurrences are), but when using EOO, a more endemic species is one whose range is restricted in extent (all occurrences are closely located to the others) (Guerin, Ruokolainen & Lowe, 2015). As each method gives two very different interpretations of what it means to be ‘restricted’ as an endemic, when we use such methods in measuring and describing patterns of endemism, we see disparity between results.

**1.2 | Weighted endemism and range size method**

One commonly used metric for measuring patterns of endemism is weighted endemism. First introduced by Crisp, Laffan, Linder and Monro (2001), this metric is calculated over an entire region, separated into grid cells. It involves counting the number of species in each cell, additionally weighting each by the inverse of its range size (for the equation see Appendix S1 in Supporting Information; Crisp et al., 2001). As the endemism value of each cell is highly dependent on the range sizes of the species that occur within it, it is clear that when measuring weighted endemism, you can get very different results depending on which range size method, AOO or EOO, is chosen (Figure 2, and Guerin et al., 2015). As mapping and identifying centres of endemism is considered a fundamental step in conservation planning and management (Kier & Barthlott, 2001; Slatyer, Rosauer & Lemckert, 2007), it is alarming that the use of each method can produce such contrasting results, potentially leading to quite different recommendations for protecting biodiversity. To further complicate the issue, as with any biogeographic analysis, a major factor that influences metric outputs, and the patterns observed, is the scale that is used.

**1.3 | The importance of scale**

Choice of scale often depends on the specific question being asked, the taxa being studied and/or the patterns being investigated. The importance of considering scale, in this case grid cell size, is well acknowledged in the biological literature, with most papers stating, at least, the scale used their analyses (Du Toit, 2010). From assessing how each method calculates range, it is clear that the difference between AOO and EOO range size is inherently scale dependent (Figure 1). For some species, increasing scale will reduce the difference between AOO and EOO, plus for AOO potentially eliminating the impact of uneven or uncompleted sampling over its distribution (Figure 1b and 1d). For other species, such as those more restricted in extent, and well sampled within their distribution, AOO can increase to match EOO, or in some cases exceed it (Figure 1a and 1c). For the latter, a case could be made that the choice of scale is inappropriate for that species. This highlights a major problem in how we operationalise these methods when applying them to metrics such as weighted endemism, where only a single scale can be chosen, and that scale must be applied to all species, often hundreds, in the same way. Increasing scale may be appropriate for one species (Figure 1b and 1d) but inappropriate for another (Figure 1a and 1c).

In a review of the weighted endemism literature (see Appendix S3 for details), I found that in the 79% of cases that did use AOO, the scales used ranged considerably from 400 square metres up to 20,500 square kilometres (Figure 3). The most commonly used scale being approximately 10,000 square kilometres. Despite this large range, only 44% percent of these studies gave any reasonable explanation for their choice of scale. For those that did, it usually had more to do with how appropriate the cell size was for the region, not taxa, being studied. It appears that fully considering the appropriateness of a scale on the particular species being analysed is lacking in the literature.

**1.4 | Our investigation**

Using five Myrtaceae genera, I first assessed the effect of scale on differences between AOO and EOO range size. I then tested the combined effects of scale and choice of range size method on the differences in (1) patterns of weighted endemism, and (2) selection of endemic hotspots. To the best of my knowledge, this is the first study to investigate these effects. As differences between AOO and EOO depend on the species being assessed, I hypothesised that these scale-method effects are taxa specific. As a particular limitation of EOO is that species with minimal non-polygon forming occurrences must be assigned an alternative range size, I also investigated the impact of the different methods used to deal with this on patterns of weighted endemism. From these analyses I hope to highlight the importance of more detailed discussion and investigation of these effects in the weighted endemism and broader literature. Doing so will ensure we can be confident in making any recommendations related to biodiversity conservation and management.

**2 | Methods**

**2. 1 | Choice of taxa and source of data**

I selected five Myrtaceae genera; *Eucalyptus, Corymbia, Melaleuca, Callistemon* and *Leptospermum*. Australian occurrence data for all genera was retrieved from the Atlas of Living Australia (ALA, 2020; Appendix S4). Due to the possible influence of difficulties in identification and limited sampling effort, hybrids and species without accepted names were excluded for all data sets (see Appendix S5 for summary details on the included data). For this analysis, outliers were removed using the 95% centroid method (Appendix S6).

**2.2 | The effect of scale on AOO versus EOO**

Area of occupancy (AOO) range sizes of all species were calculated using four different grid cell sizes: approximately 1, 110, 11,000 and 20,000 square kilometres. These were then compared to the range sizes calculated using extent of occurrence (EOO).

**2.3 | The joint effect of method and scale on weighted endemism**

Weighted endemism (WE) using each range weight method was calculated for each genus across Australia using a modified version of Guerin et al. (2015) weighted endemism function (see appendix S7) in R version 3.6.3 (R Core Team, 2020). Normalised scaling was applied to all weighted endemism cell values. To measure how endemism differs when calculated using the opposing range weight methods, I created two metrics: endemism difference (ED) and hotspot difference (HD).

ED is the absolute mean difference between all non-zero cell values of weighted endemism maps calculated using AOO and EOO (Equation 1).

, (1)

where WE(AOO) and WE(EOO) are the two gridded WE maps generated using each method, i is the position of the corresponding cells within each WE grid and n is the total number of non-zero grid cells (per grid). This metric is useful in understanding the overall, more general, differences between patterns of WE generated using EOO and AOO.

HD is the percent difference in selection of endemic hotspots based on weighted endemism maps calculated using AOO and EOO. I identified hotspots through selecting the highest valued cells that make up 5% (384,414 square kilometres) of the total land area being analysed. This percentage is one commonly chosen in regional studies of roughly the same size (Huang et al., 2012; Huang et al., 2016). HD values of zero indicate that every cell chosen using one method was also chosen using the other. This metric is particularly useful in understanding how the method and scale chosen can significantly impact identification of hotspots and therefore recommendations for conservation planning and management.

To test the impact of scale, I calculated ED and HD using 31 different grid cell sizes ranging from 1 to 20,000 square kilometres. This range encompassed most scales observed in the literature review described earlier.

**2.4 | Methods for dealing with minimal occurrence record species**

For genera containing species with occurrence records that cannot generate a polygon (relevant for EOO only), I tested four different minimal occurrence methods: removal from data set, assigning a set area per record, distance between records and record jittering (see Table 1 for details on each method). The number of species in which these methods were applied was eight in *Eucalyptus*, three in *Melaleuca* and one in *Leptospermum*, approximately one percent of total species in all three cases. The impact of using each method was assessed only on HD.

**3 | Results**

**3.1 | The effect of scale on AOO versus EOO**

In general, the percent of species for which extent of occurrence (EOO) was larger than area of occupancy (AOO) decreased with increasing grid cell size (Table 2). At the largest cell size analysed (20,000 square kilometres), more than 50 percent of species had EOO ranges smaller than AOO ranges in all genera excluding *Corymbia* (Table 2).

**3.2 | The joint effect of method and scale on weighted endemism**

Both endemism difference (ED) and hotspot difference (HD) increased as cell size increased for all genera (Figure 4). The magnitude off this effect was however highly variable and genera specific (Figure 4). Notably, at larger cell sizes, Eucalyptus had considerably larger HD values, up to 70 percent, generating considerable differences in which cells were chosen as endemic hotspots using each method (Figure 4a, Figure 5).

**3.3 | Methods for dealing with minimal occurrence record species**

Choice of minimal occurrence method had no considerable impact on HD across all cell sizes in each of the three relevant genera.

**4 | Discussion**

**4.1 | The effect of scale on AOO versus EOO**

Extent of occurrence (EOO) and area of occupancy (AOO) are two fundamentally distinct methods of calculating a species range size, each serving different purposes and with different limitations (Gaston & Fuller, 2006). Our analysis of the impact of scale on AOO and EOO range sizes generated surprising results. By definition, and throughout the literature, AOO is considered to be equal or less than EOO (Gaston & Fuller, 2006). Based on that, it would be intuitive to predict that as scale increases, AOO would become more similar to EOO. Although inherently true at miniscule scales, for some species studied here, AOO exceeded EOO even at relatively small cell sizes (Table 2). As mentioned earlier, scales that produce such effects would be considered unsuitable for said species. This criticism may be appropriate for a single species analysis, but when applying such methods to measurements of biodiversity, where only a single cell size can be applied, such criticism is less supported. Using *Melaleuca lophocoracorum* as an example, a species known only to occur in a small area of the Ravenshoe State Forest (QLD), even at cell sizes of 1 square kilometre AOO exceeded EOO. By contrast, for species such as *M. sericea*, known to have quite a wide distribution in northern WA and NT, AOO was less than EOO for most cell sizes. Here, choice of a very small scale suitable for *M. lophocoracorum* would be inappropriate for *M. sericea* who due to sampling issues, may have considerable unevenness in the distribution of its occurrences. For such cases, choosing a single cell size presents a somewhat win-lose situation. Despite this, I found no evidence of discussion on such effects in the weighted endemism literature. In future studies, perhaps final selection of a scale or method should be one that minimises some of these unwanted, misrepresenting, effects.

**4.2.1 | The joint effect of method and scale on weighted endemism (ED)**

As you may intuitively predict that AOO becomes more similar to EOO with increasing cell size, you would also predict the same for the patterns of weighted endemism (WE) generated using each method. As the former prediction is not so straightforward in this case, it is not surprising that we did not see endemism difference (ED) converge towards zero as cell size increased. Here, as cell size increased, and more species had AOO exceeding EOO, ED generally increased. Essentially, these species were de-weighted, and became less ‘endemic’ compared to their EOO equivalents as scale increased. This pattern was taxa specific, likely due to some genera having more or less scale sensitive species. In addition to taxa specificity, there was also considerable variation in ED from one cell size to the next. A likely explanation for this is in how the grid structure, critical to AOO calculations, is allocated over each species distribution. If you imagine four cells of equal 1 square kilometre size, coming together to form a grid, there is a single point of intercept at the centre of the grid block. If this point of intercept occurs within the centre of a species distribution, occurrences would occupy each cell, and the AOO range would be 4 square kilometres (Figure 6a). Now let’s say we increase the size of those grids to 3 square kilometres, and the single point of intercept shifts to just outside the species range, meaning all of its occurrences now sit within a single grid cell, the AOO range would be 3 square kilometres (Figure 6b). In this case, an increase in cell size decreased AOO. The extreme variation in ED for each genus across cell size could be due to this grid allocation effect, as some species had AOO ranges increase at one cell size (diverging from EOO) but decrease in the next (converging towards EOO). Future studies should further explore this effect with real species distributions, in R and other software commonly used.

**4.2.2 | The joint effect of method and scale on weighted endemism (HD)**

Logically, the effects of cell size on ED would be similar for hotspot difference (HD) and here we see evidence of that. Although highly taxa specific, the percent difference in the selection of endemic hotspots increased with cell size. As many AOO range sizes increasingly surpassed those of EOO, de-weighting and up-weighting certain cells, there would be a considerable shift in which cells are chosen as hotspots from each AOO and EOO WE maps. In assessing this scale-method effect on HD for *Eucalyptus* specifically, we see that less cells are chosen by both methods at larger scales. In a situation where these analyses are used to make recommendations for conservation or management, the high disparity between AOO and EOO selected cells, even at intermediate cell sizes, is confronting. Although a considerable limitation of EOO is with range size allocation to non-polygon occurring species, as only a small percentage of all species had this problem, the effects of scale on HD were not further influenced by the method chosen to deal with these species.

**4.3 | Implications and future directions**

To the ignorance of all patterns described above, AOO is used almost exclusively throughout the weighted endemism literature, and in most cases, without consideration of the appropriateness of scale on the study taxa. As it is evident that the relationship between AOO and EOO at increasing scales is not as simple as originally thought, and that this can have considerable impacts on patterns of weighted endemism and selection of hotspots, moving forward it is critical that we more closely examine this relationship for all species being analysed. In doing this we can make more informed decisions about the scale and range size method we use and better understand what this really means for the type of ‘restriction’ in range we are trying to explore. Although this paper has focussed on the idea of using just one method depending on the question asked, future studies could investigate the viability and benefits of using both AOO and EOO methods together, in much the same way that groups such as the International Union for Conservation of Nature (IUCN) use both methods in classifying threatened species on their Red List (IUCN Standards and Petitions Subcommittee, 2019). To conclude, this study highlights the importance of more thorough discussion and investigation of these scale-method effects, ensuring that if any conservation recommendations are made, we can be confident that biodiversity is being represented in the way we intended.

**Acknowledgments**

I would like the acknowledge the contribution and guidance of my supervisors Professors Richard Fuller and Margaret Mayfield.

**References**

ALA (2020). Atlas of Living Australia (ALA). Available at: [http://www.ala.org.au](about:blank) (accessed June 2020).

Anderson, S. (1994). Area and endemism. *Quarterly Review of Biology, 69*(4), 451–471. [https://doi.org/10.1086/418743](about:blank)

Crisp, M. D., Laffan, S., Linder, H. P., & Monro, A. (2001). Endemism in the Australian flora. *Journal of Biogeography, 28*(2), 183–198. [https://doi.org/10.1046/j.1365-2699.2001.00524.x](about:blank)

Du Toit, J. T. (2010). Considerations of scale in biodiversity conservation. *Animal Conservation, 13*(3), 229–236. [https://doi.org/10.1111/j.1469-1795.2010.00355.x](about:blank)

Gaston, K. J., & Fuller, R. A. (2009). The sizes of species’ geographic ranges. *Journal of Applied Ecology, 46*(1), 1–9. [https://doi.org/10.1111/j.1365-2664.2008.01596.x](about:blank)

Gaston, K. J., Quinn, R. M., Wood, S., & Arnold, H. R. (1996). Measures of geographic range size: the effects of sample size. *Ecography, 19*(3), 259–268. [https://doi.org/10.1111/j.1600-0587.1996.tb01253.x](about:blank)

Guerin, G. R., Ruokolainen, L., & Lowe, A. J. (2015). A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution, 6*(7), 845–852. [https://doi.org/10.1111/2041-210X.12361](about:blank)

Huang, J., Chen, B., Liu, C., Lai, J., Zhang, J., & Ma, K. (2012). Identifying hotspots of endemic woody seed plant diversity in China. *Diversity and Distributions, 18*(7), 673–688. [https://doi.org/10.1111/j.1472-4642.2011.00845.x](about:blank)

Huang, Jihong, Huang, Jianhua, Liu, Canran, Zhang, Jinlong, Lu, Xinghui, & Ma, Keping. (2016). Diversity hotspots and conservation gaps for the Chinese endemic seed flora. *Biological Conservation, 198*, 104–112. [https://doi.org/10.1016/j.biocon.2016.04.007](about:blank)

IUCN Standards and Petitions Subcommittee (2019). Guidelines for Using the IUCN Red List Categories and Criteria (Version 14). Prepared by the Standards and Petitions Committee. Retrieved from: [http://www.iucnredlist.org/documents/RedListGuidelines.pdf](about:blank)

Kier, G., & Barthlott, W. (2001). Measuring and mapping endemism and species richness: a new methodological approach and its application on the flora of Africa. *Biodiversity & Conservation, 10*(9), 1513–1529. [https://doi.org/10.1023/A:1011812528849](about:blank)

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: [https://www.R-project.org/](about:blank)

Slatyer, C., Rosauer, D., & Lemckert, F. (2007). An assessment of endemism and species richness patterns in the Australian Anura. *Journal of Biogeography, 34*(4), 583–596. [https://doi.org/10.1111/j.1365-2699.2006.01647.x](about:blank)

**Tables**

**Table 1** Descriptions of four alternative minimal occurrence methods used for calculation of extent of occurrence (EOO) in non-polygon occurring species.

|  |  |
| --- | --- |
| Minimal occurrence method | Description |
| Removal from data set | Species that cannot generate a convex hull polygon are removed from the data set and excluded from the analysis. |
| Assigning a set area per record | Species are assigned a range size of x area multiplied by the number of its records. For this analysis, 3 values of x were explored: 0.1, 1 and 10 km2. |
| Distance between records | Species with two or more records are assigned a range size of the distance between the two furthest records in kilometres. Species with one record are assigned a range size of 1 km2. |
| Record jittering | For species with three or more occurrences, one randomly chosen record is jittered by a factor of 0.1 in either longitude or latitude. Species with two occurrences are assigned a range size of the distance between them in kilometres. Species with one record are assigned a range size of 1 km2. |

**Table 2** The percent of all species for which extent of occurrence (EOO) was larger than area of occupancy (AOO), for five Myrtaceae genera at four different grid cell sizes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Grid cell size (km2) | | | |
| Genera | 1 | 110 | 11,000 | 20,000 |
| *Eucalyptus* | 98 (702) | 92 (660) | 40 (289) | 44 (312) |
| *Callistemon* | 93 (41) | 82 (36) | 39 (17) | 34 (15) |
| *Corymbia* | 100 (93) | 100 (93) | 82 (76) | 69 (64) |
| *Melaleuca* | 99 (100) | 93 (229) | 34 (84) | 30 (74) |
| *Leptospermum* | 98 (84) | 85 (73) | 27 (23) | 34 (29) |

Notes: values in parentheses indicate number of species.

**Figures**

A close up of a map

Description automatically generated

**Figure 1** Schema representing the ranges of *Corymbia blakei* (a and c) and *Corymbia papuana* (b and d) calculated using extent of occurrence (blue – minimum convex hull polygon area) and area of occupancy (red – total area of all cells occupied) using both smaller (a and b; approximately 5300 km2) and larger (c and d; approximately 18,300 km2) cell sizes. Green points indicate occurrence records for each species following outlier removal using the 95% centroid method.

A close up of a map

Description automatically generated

**Figure 2** Regional maps of weighted endemism for *Eucalyptus* calculated using area of occupancy (a and b) and extent of occurrence (c and d) on both normalised (a and c) and logarithmic (b and d) scales. For all calculations, cell size was set to approximately 10,800 km2 (1° resolution) and only Australian occurrence records were included following outlier removal using the 95% centroid method. For detailed cell-based comparisons of log values see appendix S2.

A screenshot of a cell phone

Description automatically generated

**Figure 3** Frequency distribution of grid cell sizes used in 23 studies measuring weighted endemism using cell-based range size methods (i.e. area of occupancy). The remaining six studies investigated (29 total) used non cell-based methods such as extent of occurrence and other operational geographical units.

A close up of a map

Description automatically generated

**Figure 4** Measurements of (a) hotspot difference (HD) and (b) endemism difference (ED) calculated over increasing grid cell size for five Myrtaceae genera. For all taxa and calculations, only Australian occurrence records were included following outlier removal using the 95% centroid method. For all species with minimal occurrence records, EOO was set using the ‘distance between records’ method.

A close up of a map

Description automatically generated

**Figure 5** The selection of *Eucalyptus* endemic hotspots based on weighted endemism cell values calculated using extent of occurrence and area of occupancy at cell sizes of (a) 5300 and (b) 20,000 km2. Hotspot difference (HD) was 45 and 60 percent for (a) and (b) respectively. Total area selected in both (a) and (b) is approximately 384,414 km2 (five percent of Australia’s total land area).

A picture containing bird

Description automatically generated

**Figure 6** Schema representing the impact of two alternative scale-dependent grid allocations on the area of occupancy (AOO) range size for a hypothetical species. The point of cell intercept could occur (a) within a species distribution or (b) outside a species distribution.

**Supporting information**

**Appendix S1**

Weighted endemism (WE) is expressed as:

,

where n is the number of species occurring in the grid cell, and Ri is the range weighting of species i.

**Appendix S2**

A screenshot of a cell phone

Description automatically generated

**Figure S2** Scatterplot of logarithmically scaled weighted endemism scores for *Eucalyptus* calculated using extent of occurrence (EOO) versus area of occupancy (AOO) for 1360 cells. For all calculations, cell size was set to approximately 10,800 km2 (1° resolution) and only Australian occurrence records were included following outlier removal using the 95% centroid method. See Figure 2b in the main text for map representation.

**Appendix S3**

Twenty-nine studies analysing patterns of weighted endemism were investigated for their choice of range size method and scale used. The articles were found using the search term ‘weighted endemism’ in the Web of Science database (available at: [https://webofknowledge.com/](about:blank)).

**Appendix S4**

Details on occurrence record downloads.

Eucalyptus: Atlas of Living Australia occurrence download at [https://biocache.ala.org.au/occurrences/search?q=lsid%3Ahttps%3A%2F%2Fid.biodiversity.org.au%2Ftaxon%2Fapni%2F51302291](about:blank) accessed on Fri May 8 2020.

*Callistemon*: Atlas of Living Australia occurrence download at [https://biocache.ala.org.au/occurrences/search?q=lsid%3Ahttps%3A%2F%2Fid.biodiversity.org.au%2Ftaxon%2Fapni%2F51288949](about:blank) accessed on Fri May 8 2020.

*Melaleuca*: Atlas of Living Australia occurrence download at [https://biocache.ala.org.au/occurrences/search?q=lsid%3Ahttps%3A%2F%2Fid.biodiversity.org.au%2Ftaxon%2Fapni%2F51290434](about:blank) accessed on Fri May 8 2020.

*Corymbia*: Atlas of Living Australia occurrence download at [https://biocache.ala.org.au/occurrences/search?q=lsid%3Ahttps%3A%2F%2Fid.biodiversity.org.au%2Ftaxon%2Fapni%2F51289369](about:blank) accessed on Fri May 8 2020.

*Leptospermum*: Atlas of Living Australia occurrence download at [https://biocache.ala.org.au/occurrences/search?q=lsid%3Ahttps%3A%2F%2Fid.biodiversity.org.au%2Ftaxon%2Fapni%2F8499083](about:blank) accessed on Fri May 8 2020.

**Appendix S5**

**Table S5** Summary information of study taxa data sets prior to outlier removal.

|  |  |  |
| --- | --- | --- |
| Taxa | Number of species | Number of records |
| *Eucalyptus* | 715 | 723627 |
| *Melaleuca* | 245 | 113047 |
| *Corymbia* | 93 | 68269 |
| *Callistemon* | 44 | 18562 |
| *Leptospermum* | 86 | 83379 |

**Appendix S6**

The 95% centroid method removes 5% of all records furthest from the centre of the convex hull polygon surrounding all records.

**Appendix S7**

All R code and functions used can be found at [https://github.com/tlawrie19/Endemism\_Project](about:blank)