JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 26 (2023), pp. 79–98
doi:10.5311/JOSIS.2023.26.254

RESEARCH ARTICLE

Mapping uncertain spatial object extents from point samples using fuzzy alpha-shapes

Thomas R. Etherington

Manaaki Whenua - Landcare Research, New Zealand

Received: September 8, 2022; returned: November 21, 2022; revised: December 4, 2022; accepted: December 11, 2022.

Abstract:
Mapping the extent of spatial objects from point samples is a fundamental process in geographical analysis. Computational geometry methods are commonly used, and one method that has been proposed is the alpha-shape as it is insensitive to both bias and errors that are common in crowdsourced geographic data and big geographic data more generally. However, many spatial objects are uncertain in nature, with vague boundaries that are not well represented by the current use of discrete alpha-shapes. Fuzzy alpha-shapes are presented as a highly generic and adaptable methodology that can produce maps of spatial objects that recognise the vague and uncertain nature of many geographies. A series of virtual geography experiments demonstrate that fuzzy alpha-shapes avoid the need for binary thresholds, create a model that better represents the uncertain boundaries of some spatial objects, while also retaining the robustness to errors and bias that motivated the original use of alpha-shapes for mapping spatial objects.

Keywords: bias, biogeography, distribution, entity, errors, geographic, geometry, range, region, species

1 Introduction

Mapping the extent of spatial objects that represent geographic patterns, regions, or distributions is a central component of the study of geography [38]. For example, biogeographers have a long history of using a point sample of species location data to delineate a spatial object representing the geographic range (or distribution) of a species [48]. Given spatial locations and objects are core transdisciplinary spatial information concepts [33] analysing point data to produce areal objects is a fundamental process in geographical analysis [44]. Therefore, while this paper will continue to focus on biogeographical examples, this is a generic geographical problem relevant to the study of many geographies. For example, if we consider a two-dimensional Euclidean space $\mathbb{R}^2$, we might want to determine a subset $S \subset \mathbb{R}^2$ defined by a set $P = \{p_1, p_2, p_3, \ldots, p_n\}$ of points in $\mathbb{R}^2$. One framework to achieve this is via computational geometry techniques that have been adopted widely within geographic information system (GIS) software, a simple example of which would be to compute the convex hull (or minimum convex polygon) of P [10]. The region of space within the convex hull would then constitute the spatial object being mapped (Figure 1a). However, convex hulls will be inappropriate models of geographic objects such as the range of a species as the shape of the convex hull does not allow for disjointed or concave shapes that will occur due to islands, bays, and peninsulas [48].

A more complex computational geometry technique called the alpha-shape is another method to delineate the extent of a region of space $S \subset \mathbb{R}^2$ defined by P [13]. Alpha-shapes have been proposed as an improvement on convex hulls for mapping the range of a species from a set of points [5], and alpha-shapes have also been used in other geographical domains too [9]. The alpha-shape is closely related to the Delaunay triangulation [10, 11], and one way to define an alpha-shape (confusingly there are alternative definitions) is to consider the alpha-shape as the union of a subset of simplices formed by a Delaunay triangulation [14]. Each simplex (or triangle) of a Delaunay triangulation has an associated circumcircle that has a circumcentre and a circumradii (Figure 1b). By specifying an α value, a subset of Delaunay triangulation simplices is formed by retaining only those Delaunay triangulation simplices whose circumradii $\leq \alpha$. This subset of Delaunay triangulation simplices is called an alpha-complex, and the union of the alpha-complex defines an alpha-shape. When $\alpha = \infty$ the alpha-complex is equivalent to the Delaunay triangulation and the resulting alpha-shape is equivalent to a convex hull (Figure 1c). Then as α is reduced the alpha-shape becomes smaller (Figure 1d) and smaller (Figure 1e) potentially achieving complex non-contiguous shapes [14].

The major analytical benefit of alpha-shapes for mapping the extent of species ranges is an insensitivity to both bias and errors [5]. Errors relate to the spatial and thematic accuracy of individual data points. For example, errors occur if a point for a spatial object is recorded in the wrong location, or if a point is in the correct location but is for the wrong spatial object. Bias relates to how well a sample of points represents the whole spatial object, and idealised sampling free from bias often follows a random or systematic design. However, biased sampling occurs when some parts of space are sampled more intensely than others. Bias is a particular problem for unstructured big data that has no sampling design, as even if the data are free of errors bias can lead to flawed conclusions. Therefore, the insensitivity of alpha-shapes to both bias and errors [5] is an extremely attractive benefit as the natural history data that is relied upon for mapping species ranges is known to contain errors [43, 60] and significant bias [27, 40]. But again, while focussing here on biogeographical examples, these issues of bias and errors are transdisciplinary and will be relevant for any crowdsourced geographic data [3] or big geographic data more generally [23].

The major analytical challenge for alpha-shapes is specifying an α value. Previous applications of alpha-shapes, including the geographic applications, have involved the pursuit of the correct, or at least optimum, α value [4, 5, 9, 21, 24]. However, use of a single α value may not always be appropriate for geographical applications as the extent of spatial objects may not have the crisp boundaries [33] that result from the use of a single α value

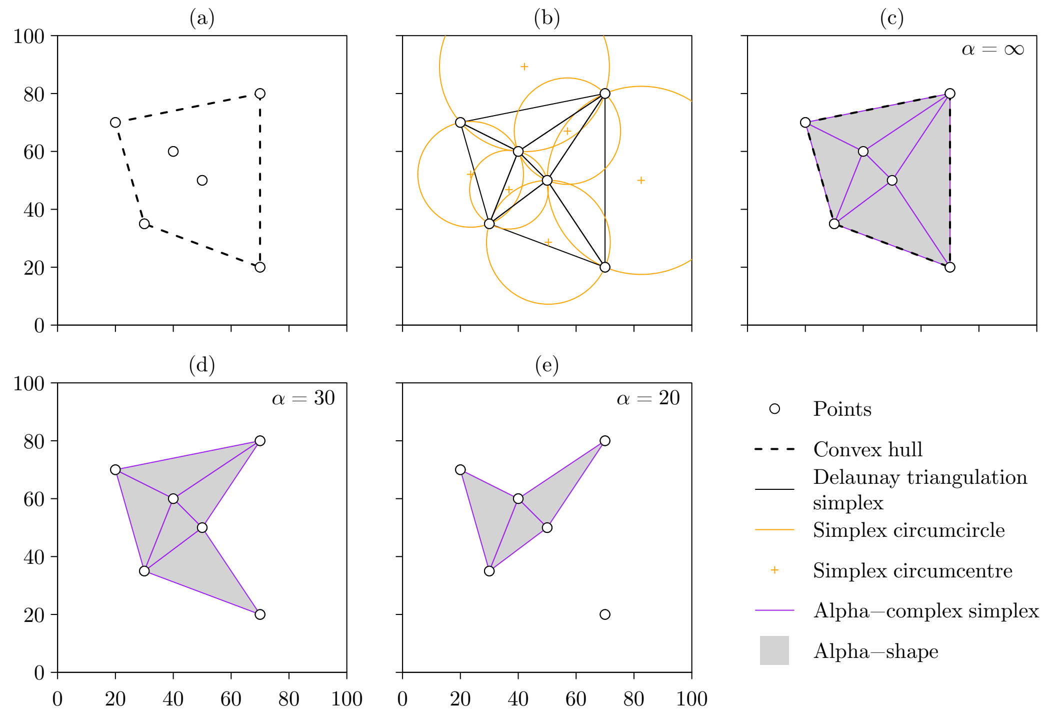

Figure 1: Examples of alpha-complex construction in the context of related computational geometry concepts. (a) Given a set of points a convex hull can define a region of space that encapsulates the points. (b) The same points can be used to construct a Delaunay triangulation for which each simplex (or triangle) has an associated circumcircle and circumcentre. By specifying (c) $\alpha = \infty$, (d) $\alpha = 30$, and (e) $\alpha = 20$, a variety of alpha-complexes can be generated that only retain those simplices with a circumradii $\leq \alpha$ and when combined these alpha-complexes produce alpha-shapes.

and that produce a binary classification of space as either inside or outside an alpha-shape (Figure 1). Many geographies are vague and for these vague geographies the boundaries of spatial objects will be uncertain [18].

For biogeographers it is certainly the case that the range of a species is a vague concept with uncertain boundaries. MacArthur [36] noted that “Although we have presented discussions of the boundaries of species distributions [ranges] as if these boundaries were immutably fixed . . . these boundaries can be very tenuous” and Gaston [22] “that *the* edge of a geographic range does not exist” both of which recognise the vagueness of a species range as a spatial object.

This description of a spatial object with uncertain boundaries can be better represented by a fuzzy set for which a membership function m is used to define set membership for a set of conditions x as a possibility ranging from zero to one [59]. As with the classical crisp set theory, when $m(x) = 1$ then x is in the set, when $m(x) = 0$ then x is not in the set, but when $0 < m(x) < 1$ then x has partial membership of the set to recognise the uncertainty of set membership.

The use of fuzzy sets in geometry [54] and geography [18, 52] has been advocated for some time. This paper shows how alpha-shapes and fuzzy sets can be integrated in a highly generic and adaptable methodology that can map the extent of spatial objects from point samples that recognise the vague and uncertain nature of many geographies.

2 Fuzzy alpha-complex

The usefulness of fuzzy sets for representing the uncertainty around geometrical objects was introduced some time ago [54], and has even been applied to alpha-hulls [37] that are related to alpha-shapes [13]. In the context of alpha-shapes the concept of fuzzy set membership can be applied easily using a membership function $m(r)$ for the circumradii r of each simplex in an alpha-complex to create a fuzzy alpha-complex in which a simplex can have partial fuzzy membership.

Geographical applications of fuzzy sets have used a wide variety of fuzzy membership functions and the choice should be dictated by the specific application [52]. In this context of alpha-shapes, a membership function is required that has $m(r) = 1$ for small circumradii with that decreases to $m(r) = 0$ for large circumradii.

One way to approach this problem is to use an asymmetric membership function that has two different membership functions around a critical value [7]. In an alpha-shape context this critical value would continue to be α, with $m(r) = 1$ where $r \leq \alpha$. To create a decreasing membership function where $r > \alpha$ one option is to use the shape of a normal distribution. The normal distribution is commonly used as the basis for a membership function in geographical applications, perhaps as the parameters of mean and standard deviation are well understood and are expressed in the same units as the value forming the basis of the membership function [52] – as opposed to parameters that are unitless indices. Using this approach the mean is set as α and a standard deviation σ is selected to control the slope of the membership function where $r > \alpha$. This approach creates a membership function of

$$m(r) = \begin{cases} 1 & r \leq \alpha \\ e^{-\frac{1}{2}\left(\frac{r-\alpha}{\sigma}\right)^2} & r > \alpha \end{cases} \tag{1}$$

that can be adapted to produce a variety of forms (Figure 2a). The effect of applying these different fuzzy membership functions to alpha-complexes from the same point sample demonstrates that where the point sample is denser the alpha-complex simplices become smaller and fuzzy alpha-complex membership increases, and where the point sample is sparser the alpha-complex simplices become larger and fuzzy alpha-complex membership decreases. Therefore, these fuzzy alpha-complexes can represent a spatial object with a gradation of membership that can represent the uncertainty associated with delineating the region of space that belongs to the spatial object from the point sample (Figure 2b–c).

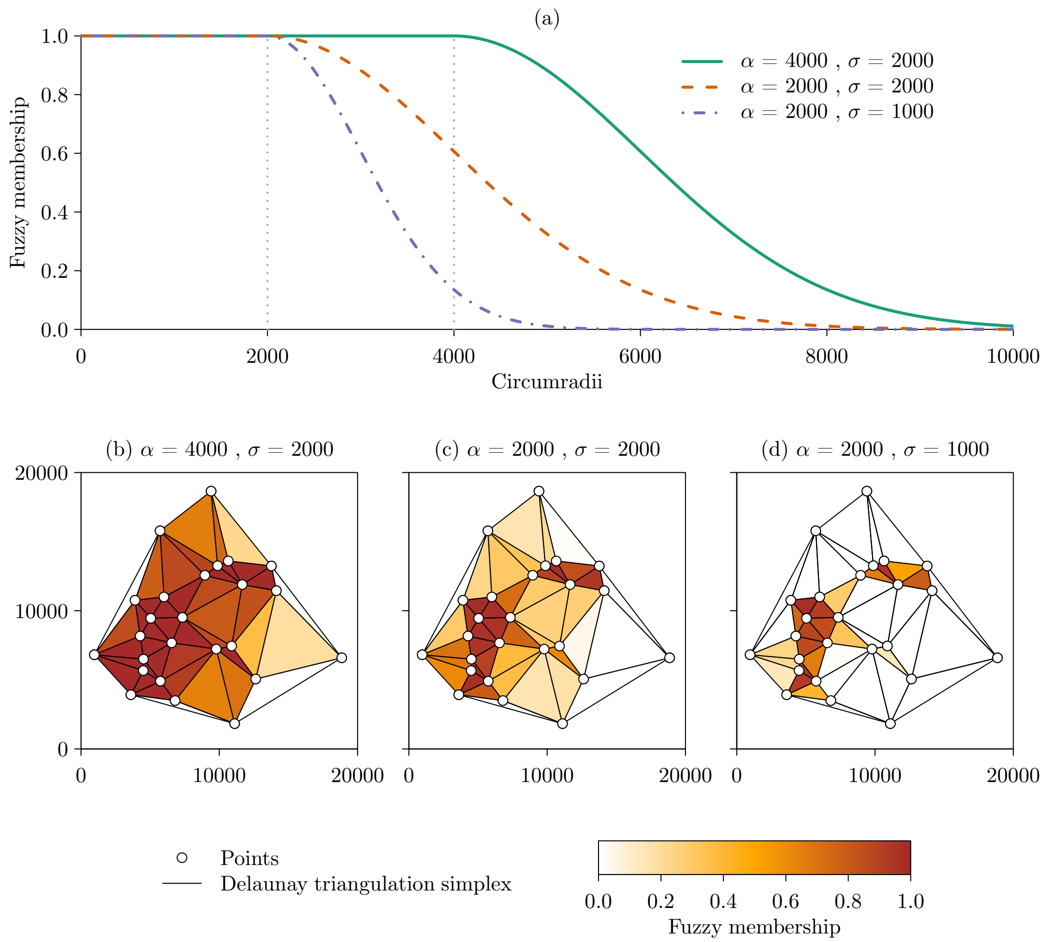

Figure 2: Examples of asymmetric fuzzy membership functions (Equation 1) for creating a fuzzy alpha-complex. The shape of each membership function is controlled by the parameters α that determines a threshold for the membership, and σ as the standard deviation of a normal distribution with smaller values of σ producing steeper curves above the membership threshold.

3 Fuzzy alpha-shape

While the fuzzy alpha-complex gives a gradation of membership, the triangular shapes of the simplices with hard edges does not produce a spatially fuzzy boundary. This can be problematic as using sharp lines to represent geographical boundaries can result in misleading conclusions as a fuzzy approach should convey information about the rate of change in the boundaries [58].

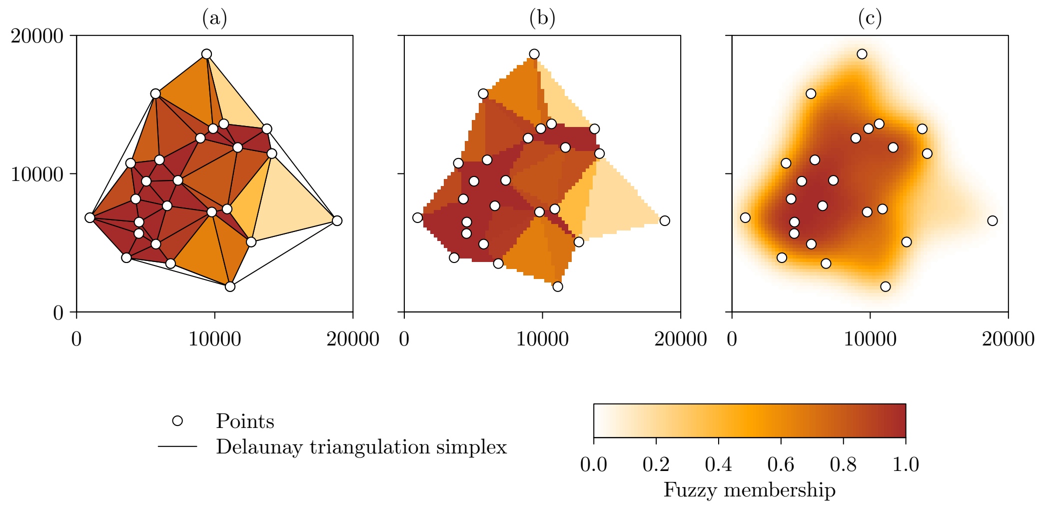

Figure 3: Example of fuzzy alpha-shape construction. (a) A fuzzy alpha-complex generated with $\alpha = 4000$ and $\sigma = 2000$ is (b) digitised such that (c) a Gaussian low-pass filter with $\sigma = 2000$ can be applied to produce a fuzzy alpha-shape.

A first step towards addressing this problem is to adopt a digital geometry approach to producing a fuzzy alpha-complex. Within a digital geometry framework computational geometry algorithms calculate the geometric properties of a grid (or lattice) of points in Euclidean space, and this usually involves the digitisation of discrete geometric objects [55]. In a geographic context digitisation occurs by representing space using a raster data structure, and for which each cell has a coordinate for its centre and a value that can be a crisp binary 0 or 1 membership value, or a fuzzy 0 to 1 membership value to denote if the element belongs to a digitised geometric object [31, 54]. Therefore, it is a very simple process to convert a discrete fuzzy alpha-complex (Figure 3a) into a digital fuzzy alpha-complex (Figure 3b).

Having adopted a digital geometry approach the problem of the crisp geographical boundaries between the fuzzy alpha-complex simplices can be approached through raster generalisation. In this application using fuzzy alpha-shapes to map uncertain spatial objects, the crisp transitions between the fuzzy alpha-complex simplices represent irrelevant information, and the information to enhance is the rate of change in the boundaries between simplices [58]. Generalisation with raster data usually takes the form of some kind of low-pass spatial filter (or moving window) that smooths out the numerical data within a raster by some form of two-dimensional weighted average [39]. One such low-pass filter that is used widely in digital image processing for remote sensing is a Gaussian filter that uses a bivariate normal distribution as the filter to generalise the data within the raster [35]. Having already defined a mean α and standard deviation σ with the fuzzy membership function (Equation 1) a Gaussian low-pass smoothing filter can be easily applied using the same parameter values. This results in the crisp boundaries between the simplices of the digital fuzzy alpha-complex (Figure 3b) being generalised such that the resulting fuzzy alpha-shape (Figure 3c) better fits the requirements for fuzzy boundaries [58] as the amorphous shape and smooth gradient better communicate the uncertain nature of the data. Visualisation of the resulting fuzzy alpha-shape can be achieved using methods such as monochromatic or multi-chromatic colour scales, contours, three-dimensional models [30] or some combination thereof.

4 Virtual geography experiments

One of the original advantages of using alpha-shapes for mapping species ranges was the insensitivity to both bias and errors [5]. Having developed the alpha-shape methodology further to create fuzzy alpha-shapes it is important to ensure that this robustness to bias and errors remains. To assess if these properties are still true a series of virtual geography experiments were undertaken.

4.1 Baseline experiments

The first set of baseline experiments sought to understand how well fuzzy alpha-shapes would represent an uncertain spatial object given idealised sampling conditions. This was done to establish a baseline performance against which to compare the later experiments that introduced error and bias.

Virtual, and therefore known and true, spatial objects with uncertain boundaries were generated using a single three period octave of Perlin noise [15] rescaled with the logistic function

$$f(x) = \frac{1}{1 + e^{-20(x-0.75)}} \tag{2}$$

to produce spatially complex and potentially disjointed distribution patterns that range in value from zero to one along smooth gradients. This process was repeated to produce 100 different true distributions, the fuzzy areas [53] of which covered a median of 15% (range 5% to 28%) of the experimental space (Figure 4).

Sets of data points that represent knowledge about the true distributions were then simulated. Each data point simply consists of a coordinate where the spatial object is present and has no associated value or attributes. An idealised sample was simulated via an inhomogeneous Poisson point process [12] that sampled randomly across the experimental space but only retained data points with a probability equal to the underlying true distribution. What was unknown was the minimum idealised sample size required for a fuzzy alpha-shape to achieve a good estimation of the true distribution. Keeping the parameters of $\alpha = 5$ and $\sigma = 2$ constant, for each of the 100 true distributions a fuzzy alpha-shape was calculated for an idealised sample size of 10, 50 (Figure 5a), 100, 150 (Figure 5b), 200, and 250 (Figure 5c) points.

The similarity between the true virtual distribution and fuzzy alpha-shape from each point sample was measured using the Ružička (or weighted Jaccard) similarity index [56]. The Ružička similarity index R for two 2-dimensional spatial grids (or: arrays, matrices) $\mathbf{A}$ and $\mathbf{B}$ is calculated as

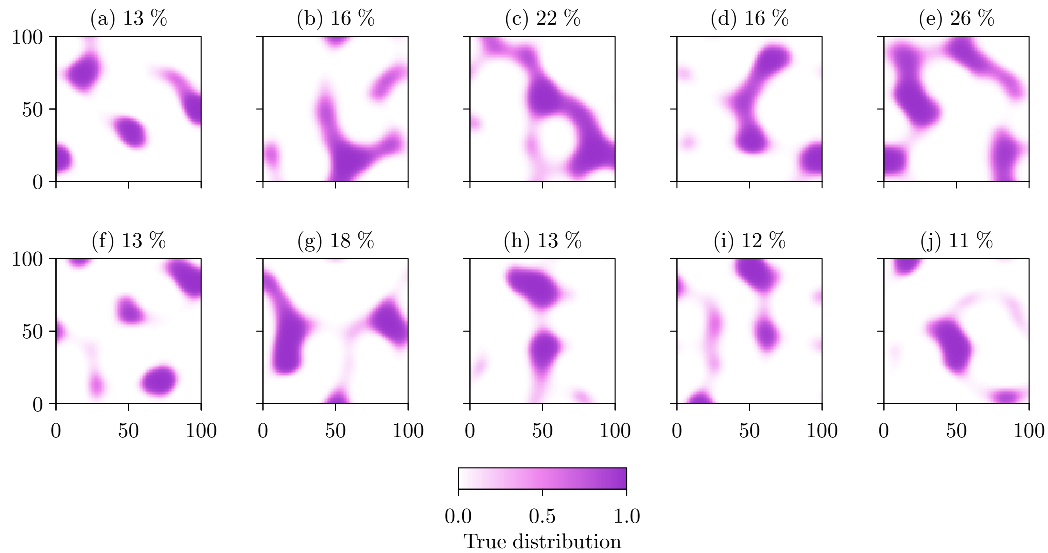

Figure 4: The first 10 examples of 100 virtual true distributions simulated via Perlin noise. In each case the percentage of the experimental space covered by the fuzzy area of the distribution is noted.

$$R(\mathbf{A},\mathbf{B}) = \frac{\sum_x \sum_y \min(\mathbf{A}_{x,y}, \mathbf{B}_{x,y})}{\sum_x \sum_y \max(\mathbf{A}_{x,y}, \mathbf{B}_{x,y})} \tag{3}$$

which produces a value ranging between zero and one, with zero for grids with no similarity and one for grids that are identical.

The results of these baseline experiments indicate that under these experimental conditions similarity between the virtual true distributions and the fuzzy alpha-shapes increased as sample size increased, but with a plateau in performance once the sample size reached 150 (Figure 5g).

This is important to understand as it means that for an idealised sample of 150 points we could add up to 100 additional points that are either errors or biased and if they have no effect then we would expect the performance to remain the same as it does with additional idealised sampling. Therefore, any observed changes in performance from the addition of points that are either errors or biased will not be a function of increasing sample size, but rather the inclusion of errors or bias within the extended sample.

4.2 Error experiments

The error experiments used the same approach as the baseline experiments in that they were also based on 100 virtual true distributions each with 150 idealised sample points all simulated in the same manner as in the baseline experiments. However, the idealised sample was then corrupted with errors that were generated as a homogeneous Poisson

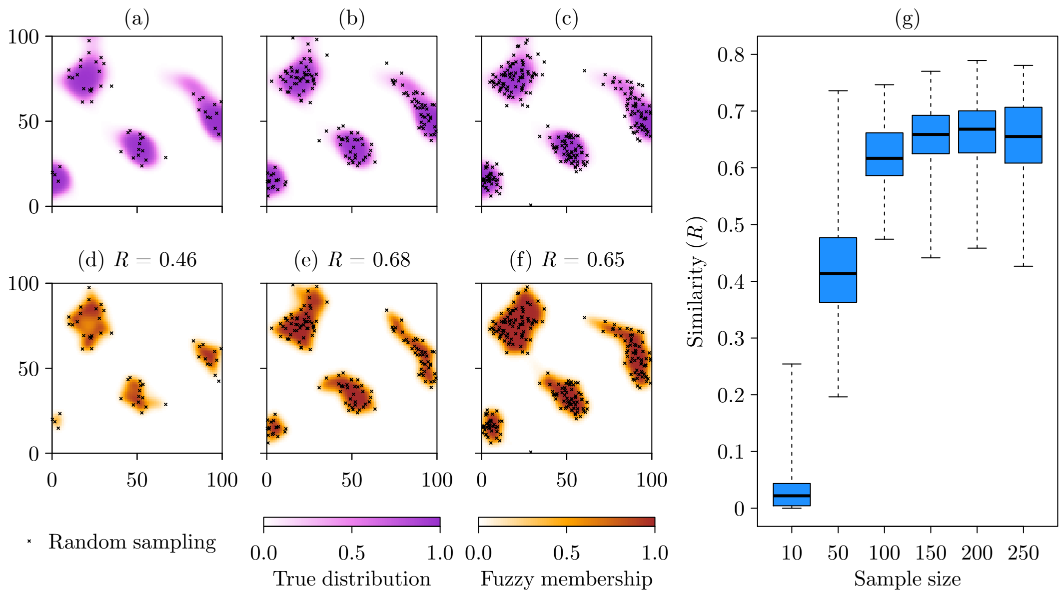

Figure 5: Examples and results of baseline virtual geography experiments used to establish the performance of fuzzy alpha-shapes with idealised sampling of differing sample sizes. For the same virtual true distribution increasing sample size from (a) 50, (b) 150, and (c) 250 points resulted in denser clustering of points concentrated within the true distribution. Similarity between the true distribution and fuzzy membership values of fuzzy alpha-shapes are measured using the Ružička index (R) that ranges from zero to one. Increasing the sample size from (d) 50 to (e) 150 resulted in improvement in the fuzzy alpha-shape, while no improvement was seen by increasing the sample size from (e) 150 to (f) 250. Replicating this process 100 times produces (g) distributions showing the median, inter-quartile range, and range of the similarity values across all 100 experiments.

point process [12] and hence were simply random locations within the experimental space that ignored the underlying true distribution.

The range of error size explored were 0, 5 (Figure 6a), 10, 20 (Figure 6b), 50, and 100 (Figure 6c) points, with increasing errors having clear effects on the resulting fuzzy alpha-shape (Figure 6d–f) compared to the idealised sample alone (Figure 5e). As with the baseline experiments the similarity between the true virtual distributions and fuzzy alpha-shapes was measured using the Ružička similarity index (R), and the effect of the errors was measured as ΔR that was the difference between R for the idealised sampling alone and for the idealised sampling plus the errors.

The examples indicate that when errors occur either within an existing cluster of correct data points, or occur as isolated points distant from other points the errors have minimal effects (Figure 6d–e). This held true across the 100 replications, as while the errors always tended to produce negative effects, even with 10 errors making up 9.5% of the data, the median $\Delta R = -0.03$ which is a negligible change for an index ranging zero to one (Figure

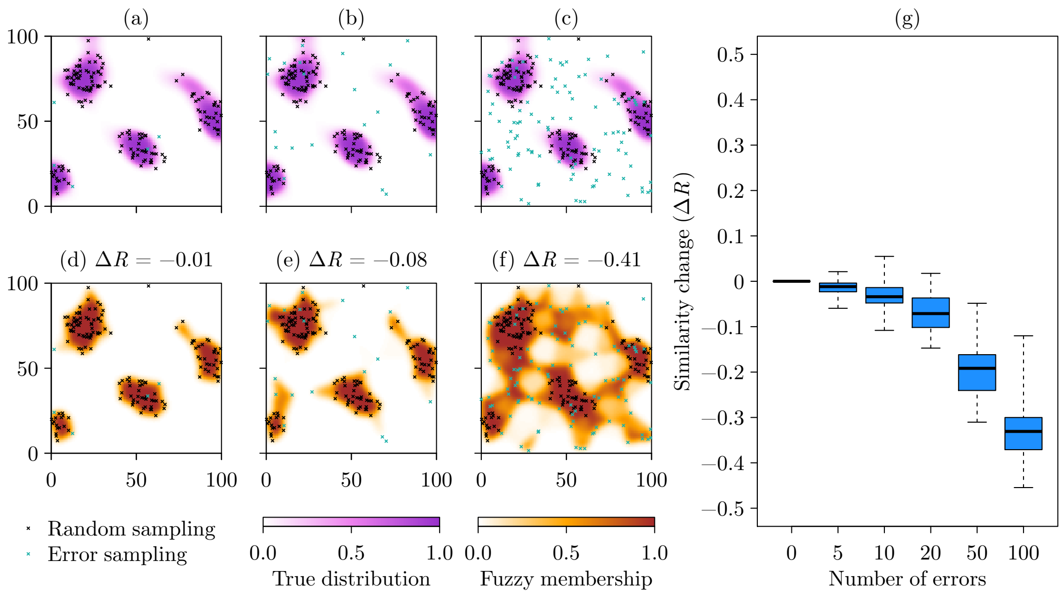

Figure 6: Examples and results of virtual geography experiments used to establish the performance of fuzzy alpha-shapes with idealised sampling contaminated with errors. The same virtual true distribution and 150 idealised samples were contaminated with (a) 5, (b) 20, and (c) 100 errors. The effect of contamination was measured as ΔR that was the difference between the Ružička similarity index for the idealised sampling alone and for the idealised sampling plus the errors. Errors consistently had a negative effect on similarity that increased as the number of errors increased from (d) 5 to (e) 20 to (f) 100 errors. Replicating this process 100 times produces (g) distributions showing the median, inter-quartile range, and range of the similarity change across all 100 experiments.

6g). With 100 errors making up 40.0% of the data points, performance was extremely poor as the errors became spatially dense enough to be incorporated into the fuzzy alpha-shape (Figure 6f), and while there was a wide range of effects across the replications a median $\Delta R = -0.35$ is a sizeable effect (Figure 6g). While it is perhaps unlikely that errors will form such a large proportion of a dataset, it is still useful to see that the fuzzy alpha-shape can perform poorly as this indicates that the virtual geography experiments are using sample sizes that provide a robust test of the methodology.

4.3 Bias experiments

The bias experiments followed a similar approach to the error experiments, with 100 virtual true distributions each with 150 idealised sample points all simulated in the same manner as in the baseline experiments. Similarly to the error experiments, these idealised samples were then corrupted with biased sampling. As it is not the number of biased data, but rather the level of bias that is the problematic factor, all bias experiments used a biased

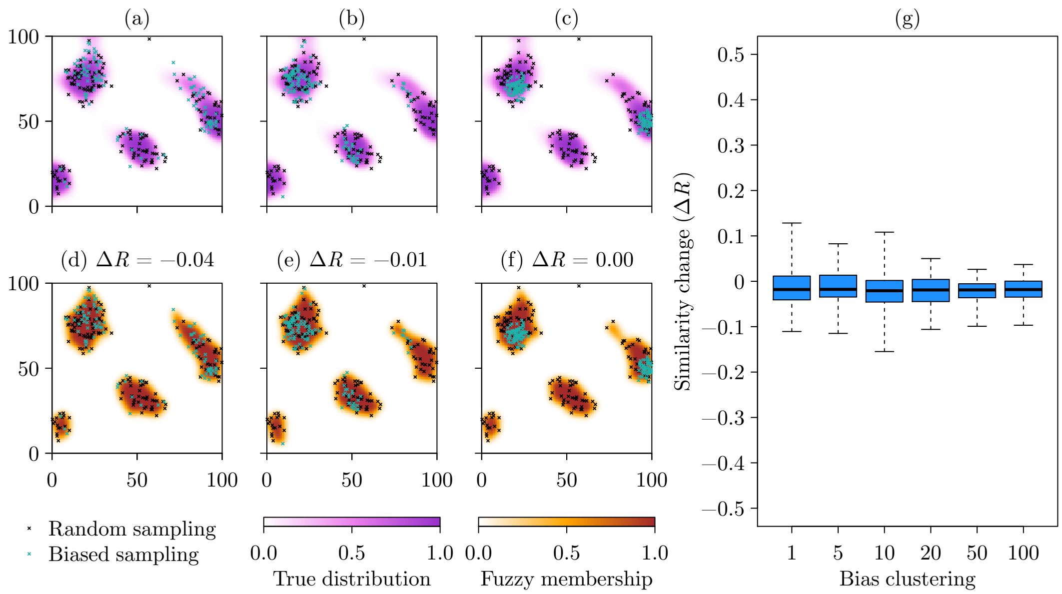

Figure 7: Examples and results of virtual geography experiments used to establish the performance of fuzzy alpha-shapes with idealised sampling contaminated with biased sampling. The same virtual true distribution and 150 idealised samples were contaminated with 100 biased samples generated from Poisson cluster point processes with (a) 5, (b) 20, and (c) 100 offspring points. The effect of contamination was measured as ΔR that was the difference between the Ružička similarity index for the idealised sampling alone and for the idealised sampling plus the biased sampling. Biased sampling consistently had an extremely minimal negative effect on similarity regardless of whether the Poisson cluster point process had (d) 5 to (e) 20 to (f) 100 offspring. Replicating this process 100 times produces (g) distributions showing the median, inter-quartile range, and range of the similarity change across all 100 experiments.

sample of 100 points, for which the spatial clustering of these points systematically varied. This was done using a Poisson cluster point process [12] in which a random location within the experimental space was chosen to act as a parent point from which to produce a cluster of offspring points based on a bivariate normal distribution with a standard deviation of 10. Each of the offspring points was then assessed in turn and became part of the biased sample with a probability equal to the underlying true distribution. Therefore, while each biased sampling point in itself is an accurate and correct record of the underlying true distribution, the biased sample as a whole provides more data for some parts of the experimental space. This process was continued until 100 biased points had been generated. The cluster sizes explored were 1 (that in essence is equivalent to the idealised sampling), 5 (Figure 7a), 10, 20 (Figure 7b), 50, and 100 (Figure 7c) that resulted in the biased sample become increasing concentrated into localised areas of the experimental space.

As with the error experiments the effect of biased sampling was measured as ΔR that was the difference between R for the idealised sampling alone and for the idealised sampling plus the biased sampling. The examples indicate that as biased data was correct in relation to the underlying true distribution, the biased data largely occurred within existing clusters of idealised sampling and as such the effect of biased sampling was minimal (Figure 7d–e). This outcome was consistent across all 100 replications with levels of biased sampling producing extremely minor reductions in performance on average (Figure 7g).

4.4 Sensitivity experiments

The experiments thus far have kept the fuzzy alpha-shape parameters constant at $\alpha = 5$ and $\sigma = 2$. However, it is important to understand how sensitive the methodology is to parameter variation. A series of sensitivity experiments were conducted that used the same approach as the baseline experiments in that they were also based on 100 virtual true distributions. For each true distribution a fuzzy alpha-shape was created for 600 combinations of α and σ values, with the α values ranging from 0.5 to 15 and the σ values ranging from 0.5 to 10 both in steps of 0.5. The performance of the fuzzy alpha-shape for each parameter combination was measured as the mean Ružička similarity index $\bar{R}$ for all 100 virtual true distributions. To examine if sensitivity varies with sample size this approach was repeated for idealised samples of 50, 150, and 250 points.

The sensitivity results show that there is some variation in the optimum α and σ amongst the different sample sizes (Figure 8). However, once a reasonable sample size was achieved this variation was minimal there was little difference between the sensitivity analyses for idealised samples of 150 (Figure 8b) or 250 points (Figure 8c). For all sample sizes there was a smooth and gradual response to changes in the α and σ parameters (Figure 8). Therefore, while there will always be an optimum combination of parameter values, the performance of the fuzzy alpha-shape does not appear to be highly sensitive to slight variations around this optimum, and is still capable of producing reasonable models of the true distribution if the chosen parameters are close to the optimum parameters.

5 Discussion

It is unlikely that there is a single universal best method to map distributions [20], but where the questions and data are well suited the fuzzy alpha-shape should provide a useful approach. In comparison to the discrete alpha-shape approach to distribution mapping [5] the fuzzy alpha-shape avoids the need to choose a singular 'best' value of α and allows for the inherent uncertainty around the extent of a distribution to be incorporated into the distribution model and visualised for a map reader.

The virtual geography experiments demonstrate clearly that the fuzzy alpha-shape method is extremely robust to biased sampling (Figure 7), which is an important issue for natural history data used to map species distributions [43] and for crowdsourced and big geographic data more generally [3, 23]. Where biased sampling is not an issue and the density of point samples is a reliable indication of the greater possibility of a spatial object, then there are existing density based methods for fuzzy mapping of spatial objects, such as density peak clustering [25] and density-based spatial clustering with noise [45], that could also be considered.

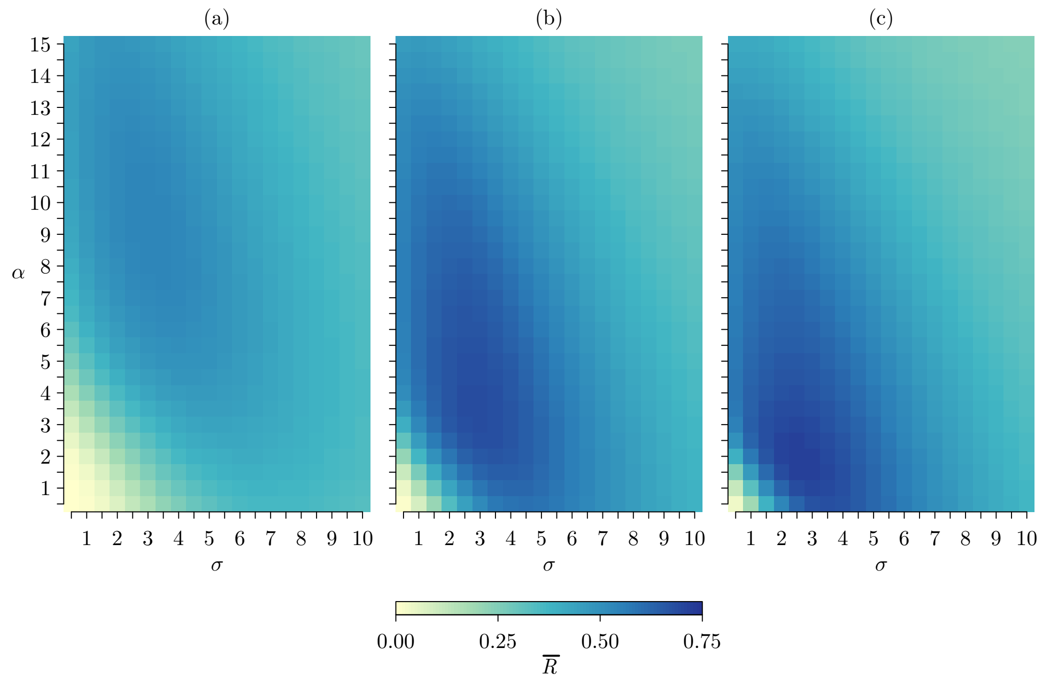

Figure 8: Results of fuzzy alpha-shape sensitivity analysis for which the performance of a range of α and σ parameter values were explored. For each combination of α and σ the same 100 virtual true distributions were used to generate fuzzy alpha-shapes from idealised samples of (a) 50, (b) 150, and (c) 250 points. The performance of each combination of α and σ was calculated as the mean Ružička similarity index $\bar{R}$ between the 100 true virtual distributions and the fuzzy alpha-shapes.

The virtual geography experiments also demonstrate that the fuzzy alpha-shape method is robust to low levels of error, with virtually no effect on performance when errors consisted of around 3% of the sample, and minimal effects on performance when errors consisted of around 6% of the sample (Figure 6). It is of course very hard to know exactly how many errors are contained within a sample, as obviously for the errors to be counted they need to be identified at which point they can be corrected or removed. At least for natural history data there is some evidence to indicate that error levels of around 4% could be expected [60] and so in this context at least the fuzzy alpha-shape method would be expected to be reasonably robust. Clearly in other contexts an analyst needs to make a decision about the quality of their data as based on the virtual geography experiments the fuzzy alpha-shape method will begin to perform poorly when errors consist of around 25% of the sample (Figure 6). Of course, under such error levels it would be questionable if any analysis of the data would be recommended.

In summary, by developing the discrete alpha-shape approach into the fuzzy alpha-shape approach to distribution mapping we have avoided the need for binary thresholds, created a model that better represents the uncertain boundaries of some spatial objects, while also retaining the robustness to errors and bias that motivated the original proposal of alpha-shapes for distribution mapping.

The fuzzy alpha-shape method is also highly adaptive as there are many fuzzy membership functions, and filters with which to generalise. So while fuzzy alpha-shapes have been presented here based on parameters from the mean and standard deviation of a normal distribution for both the fuzzy membership function (Equation 1) and the raster generalisation (Figure 3b–c), many other approaches could be equally justifiable given the data and question at hand. The flexibility of the fuzzy membership and generalisation functions does mean that fuzzy alpha-shape parameterisation provides a challenge, but equally this challenge also presents an opportunity to modify the method to best suit the data and question at hand. Such a challenge is also not a novel problem in geographical analysis as kernel density estimation is a widely used analytical method that also requires choices of both a kernel shape and bandwidth parameter that can be somewhat arbitrary [44].

Regardless of the fuzzy membership function and generalisation filter chosen, there will be potential problems in specifying the values of required parameters. The sensitivity analyses indicate that fuzzy alpha-shapes are reasonably robust to slight variations around the optimum parameters (Figure 8), but careful consideration clearly needs to be given to parameter specification. So how should researchers go about choosing parameters such as the α and σ values used within the fuzzy membership function and generalisation filter used here? It is possible that there could be quantitative data available that could guide these decisions. For a biogeographical application, if the likely dispersal distance of a species is known, then it may be possible to use this kind of information to determine the spatial distance over which presence of a species between sampling points could be more safely assumed. A similar approach could be taken in other geographical contexts if data on the mobility or spatial autocorrelation of the spatial object being modelled is known.

In the absence of any data to guide parametrisation, formal expert opinion elicitation methods could be used. For example, the Delphi method [42] could be used to form some consensus amongst experts around the necessary parameters. Such approaches have been used before in a geographic context by querying a range of opinions about the extent of an uncertain spatial boundary to create a fuzzy spatial object [41] and have even been automated via the use of human-machine interaction [50, 51]. To try and illustrate how this could be approached in the context of fuzzy alpha-shapes, remembering that an alpha-complex consists of a series of triangular simplices that are included based on their circumradii distance (Figure 1), a question such as "Given known presences at the corners of an equilateral triangle, at what distance would you be comfortable that we can assume presence at the centre of the triangle?" could be posed. This would then elicit a range of potential responses, and the mean and standard deviation of the responses could then be used as the α and σ respectively to create a fuzzy membership function (Equation 1).

Once established, fuzzy alpha-shapes could form the basis of a variety of analytical questions about the spatial objects they represent. For example, simple spatial measures such as area [19, 53], distance, and direction [1] can be calculated for or between fuzzy spatial objects. The union or intersection of fuzzy sets [59] when applied in a spatial context also allows for multi-criteria evaluation analysis of fuzzy spatial objects to be undertaken [6, 34]. Given these analytical options fuzzy spatial objects have been applied to study a range of geographic domains such as soils [6], climates [34], and urban areas [41]. Given soil cores, weather stations, and social media data commonly create point data in these geographic domains respectively, there is the potential for fuzzy alpha-shapes to be of use in a wide variety of geographic domains.

In terms of future development of this approach, there are inevitable uncertainties associated with species occurrence data relating to the available sample and locational accuracy [20] that should also ideally also be incorporated into a species distribution estimate. One way that this could be easily achieved is through the use of the simple, flexible, and scalable Monte Carlo methods [32]. For example, if the available data points had a locational uncertainty, then a Monte Carlo method could be applied whereby data points are shifted randomly in space by their spatial uncertainty before developing a fuzzy alpha-shape. This process could then be repeated numerous times, with the output being the average across all Monte Carlo simulations. The original development of alpha-shapes was for planar geometry [13] and so in a geographical context the method is only suitable for projected coordinate data where planar geometry assumptions around distances and angles hold reasonably true. This requirement to meet the assumptions of planar geometry within a projected coordinate system will likely limit application of the fuzzy alpha-shape to the study of distributions at national and possibly continental extents. However, it should be possible to extend the fuzzy alpha-shape approach to geographic coordinate data using spherical geometry, as Delaunay triangulations that form the basis of an alpha-shape can be computed for a sphere [29, 49].

To enable the use and development of the proposed fuzzy alpha-shapes technique, functions to produce fuzzy alpha-shapes have been developed in the R [47] and Python [46] programming languages and are available under permissive licences from https://doi.org/10.7931/kg7v-k118. The R packages spatstat [2], compGeometeR [17], and extrafont [8] and the Python packages NumPy [26], SciPy [57], matplotlib [28], and NLMpy [16] were used in the development of the code and to generate the examples used within this paper.

Acknowledgements

This research was funded by internal investment by Manaaki Whenua - Landcare Research and by the New Zealand Ministry of Business, Innovation and Employment via the Beyond Myrtle Rust (#C09X1806) research programme.

References

[1] ALTMAN, D. Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems 8*, 3 (1994), 271–289. doi:10.1080/02693799408902000.

[2] BADDELEY, A., RUBAK, E., AND TURNER, R. *Spatial Point Patterns: Methodology and Applications with R.* Chapman and Hall/CRC Press, London, 2015.

[3] BASIRI, A., HAKLAY, M., FOODY, G., AND MOONEY, P. Crowdsourced geospatial data quality: challenges and future directions. *International Journal of Geographical Information Science 33*, 8 (2019), 1588–1593. doi:10.1080/13658816.2019.1593422.

[4] BRASSEY, C. A., AND GARDINER, J. D. An advanced shape-fitting algorithm applied to quadrupedal mammals: improving volumetric mass estimates. *Royal Society Open Science 2*, 8 (2015), 150302. doi:10.1098/rsos.150302.

[5] BURGMAN, M. A., AND FOX, J. C. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Animal Conservation 6*, 1 (2003), 19–28. doi:10.1017/S1367943003003044.

[6] BURROUGH, P. A. Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science 40*, 3 (1989), 477–492. doi:10.1111/j.1365-2389.1989.tb01290.x.

[7] BURROUGH, P. A., AND MCDONNELL, R. A. *Principles of geographical information systems*. Oxford University Press, Oxford, 1998.

[8] CHANG, W. *extrafont: Tools for using fonts*, 2014. R package version 0.17.

[9] CHEN, M., ARRIBAS-BEL, D., AND SINGLETON, A. Understanding the dynamics of urban areas of interest through volunteered geographic information. *Journal of Geographical Systems 21*, 1 (2019), 89–109. doi:10.1007/s10109-018-0284-3.

[10] DE BERG, M., CHEONG, O., VAN KREVELD, M., AND OVERMARS, M. *Computational Geometry*, 3rd ed. Springer, Berlin, 2008. doi:10.1007/978-3-540-77974-2.

[11] DELAUNAY, B. Sur la sphére vide. *Bulletin de l'Académie des Sciences de l'URSS, Classe des Sciences Mathématiques et Naturelles 6* (1934), 793–800.

[12] DIGGLE, P. J. *Statistical analysis of spatial point patterns*, 2nd ed. Arnold, London, 2003.

[13] EDELSBRUNNER, H., KIRKPATRICK, D., AND SEIDEL, R. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory 29*, 4 (1983), 551–559. doi:10.1109/TIT.1983.1056714.

[14] EDELSBRUNNER, H., AND MÜCKE, E. P. Three-dimensional alpha shapes. *ACM Transactions on Graphics 13*, 1 (1994), 43–72. doi:10.1145/174462.156635.

[15] ETHERINGTON, T. R. Perlin noise as a hierarchical neutral landscape model. *Web Ecology 22*, 1 (2022), 1–6. doi:10.5194/we-22-1-2022.

[16] ETHERINGTON, T. R., HOLLAND, E. P., AND O'SULLIVAN, D. NLMpy: a Python software package for the creation of neutral landscape models within a general numerical framework. *Methods in Ecology and Evolution 6*, 2 (2015), 164–168. doi:10.1111/2041-210X.12308.

[17] ETHERINGTON, T. R., AND OMONDIAGBE, O. P. compGeometeR: an R package for computational geometry. *OSF Preprints* (2021). doi:10.31219/osf.io/b4zvr.

[18] FISHER, P. Sorites paradox and vague geographies. *Fuzzy Sets and Systems 113*, 1 (2000), 7–18. doi:10.1016/S0165-0114(99)00009-3.

[19] FONTE, C. C., AND LODWICK, W. A. Areas of fuzzy geographical entities. *International Journal of Geographical Information Science 18*, 2 (2004), 127–150. doi:10.1080/13658810310001620933.

[20] FORTIN, M.-J., KEITT, T. H., MAURER, B. A., TAPER, M. L., KAUFMAN, D. M., AND BLACKBURN, T. M. Species' geographic ranges and distributional limits: pattern analysis and statistical issues. *Oikos 108*, 1 (2005), 7–17. doi:10.1111/j.0030-1299.2005.13146.x.

[21] GARDINER, J. D., BEHNSEN, J., AND BRASSEY, C. A. Alpha shapes: determining 3d shape complexity across morphologically diverse structures. *BMC Evolutionary Biology 18*, 1 (2018), 184. doi:10.1186/s12862-018-1305-z.

[22] GASTON, K. J. *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, Oxford, 2003.

[23] GORMAN, S. P. The danger of a big data episteme and the need to evolve geographic information systems. *Dialogues in Human Geography 3*, 3 (2013), 285–291. doi:10.1177/2043820613513394.

[24] GRUSON, H. Estimation of colour volumes as concave hypervolumes using α-shapes. *Methods in Ecology and Evolution 11*, 8 (2020), 955–963. doi:10.1111/2041-210X.13398.

[25] GUO, J., AND DU, S. Building fuzzy areal geographical objects from point sets. *Transactions in GIS 25*, 6 (2021), 3067–3087. doi:10.1111/tgis.12808.

[26] HARRIS, C. R., MILLMAN, K. J., VAN DER WALT, S. J., GOMMERS, R., VIRTANEN, P., COURNAPEAU, D., WIESER, E., TAYLOR, J., BERG, S., SMITH, N. J., KERN, R., PICUS, M., HOYER, S., VAN KERKWIJK, M. H., BRETT, M., HALDANE, A., DEL RÍO, J. F., WIEBE, M., PETERSON, P., GÉRARD-MARCHANT, P., SHEPPARD, K., REDDY, T., WECKESSER, W., ABBASI, H., GOHLKE, C., AND OLIPHANT, T. E. Array programming with NumPy. *Nature 585*, 7825 (2020), 357–362. doi:10.1038/s41586-020-2649-2.

[27] HUGHES, A. C., ORR, M. C., MA, K., COSTELLO, M. J., WALLER, J., PROVOOST, P., YANG, Q., ZHU, C., AND QIAO, H. Sampling biases shape our view of the natural world. *Ecography 44*, 9 (2021), 1259–1269. doi:10.1111/ecog.05926.

[28] HUNTER, J. D. Matplotlib: a 2D graphics environment. *Computing in Science & Engineering 9*, 3 (2007), 90–95. doi:10.1109/MCSE.2007.55.

[29] JACOBSEN, D. W., GUNZBURGER, M., RINGLER, T., BURKARDT, J., AND PETERSON, J. Parallel algorithms for planar and spherical Delaunay construction with an application to centroidal Voronoi tessellations. *Geoscientific Model Development 6*, 4 (2013), 1353–1365. doi:10.5194/gmd-6-1353-2013.

[30] JIANG, B. Visualisation of fuzzy boundaries of geographic objects. *Cartography 27*, 2 (1998), 41–46. doi:10.1080/00690805.1998.9714275.

[31] KLETTE, R., AND ROSENFELD, A. *Digital Geometry*. Morgan Kaufmann, San Francisco, 2004. doi:10.1016/B978-1-55860-861-0.X5000-7.

[32] KROESE, D. P., BRERETON, T., TAIMRE, T., AND BOTEV, Z. I. Why the Monte Carlo method is so important today. *WIREs Computational Statistics 6*, 6 (2014), 386–392. doi:10.1002/wics.1314.

[33] KUHN, W. Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science 26*, 12 (2012), 2267–2276. doi:10.1080/13658816.2012.722637.

[34] LEUNG, Y. On the imprecision of boundaries. *Geographical Analysis 19*, 2 (1987), 125–151. doi:10.1111/j.1538-4632.1987.tb00120.x.

[35] LILLESAND, T. M., AND KIEFER, R. W. *Remote Sensing and Image Interpretation*, 4th ed. John Wiley & Sons, New York, 2000.

[36] MACARTHUR, R. H. *Geographical Ecology: patterns in the distribution of species*. Harper & Row, New York, 1972.

[37] MANDAL, D. P., AND MURTHY, C. A. Selection of α for α-hull and formulation of fuzzy α-hull in $\mathbb{R}^2$. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 3*, 4 (1995), 401–417. doi:10.1142/S0218488595000207.

[38] MATTHEWS, J. A., AND HERBERT, D. T. *Geography: a very short introduction*. Oxford University Press, Oxford, 2008. doi:10.1093/actrade/9780199211289.001.0001.

[39] MCMASTER, R. B., AND MONMONIER, M. A conceptual framework for quantitative and qualitative raster-mode generalization. *GIS/LIS'89 Proceedings 2* (1989), 390–403.

[40] MEYER, C., WEIGELT, P., AND KREFT, H. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters 19*, 8 (2016), 992–1006. doi:10.1111/ele.12624.

[41] MONTELLO, D. R., GOODCHILD, M. F., GOTTSEGEN, J., AND FOHL, P. Where's Downtown?: behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation 3*, 2-3 (2003), 185–204. doi:10.1080/13875868.2003.9683761.

[42] MUKHERJEE, N., HUGE, J., SUTHERLAND, W. J., MCNEILL, J., VAN OPSTAL, M., DAHDOUH-GUEBAS, F., AND KOEDAM, N. The Delphi technique in ecology and biological conservation: applications and guidelines. *Methods in Ecology and Evolution 6*, 9 (2015), 1097–1109. doi:10.1111/2041-210X.12387.

[43] NEWBOLD, T. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography 34*, 1 (2010), 3–22. doi:10.1177/0309133309355630.

[44] O'SULLIVAN, D., AND UNWIN, D. J. *Geographic Information Analysis*, 2nd ed. John Wiley & Sons, Hoboken, 2010.

[45] PARKER, J. K., AND DOWNS, J. A. Footprint generation using fuzzy-neighborhood clustering. *GeoInformatica 17*, 2 (2013), 285–299. doi:10.1007/s10707-012-0152-0.

[46] PYTHON CORE TEAM. *Python: A dynamic, open source programming language*. Python Software Foundation, 2021.

[47] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[48] RAPOPORT, E. H. *Areography: geographical strategies of species*. Pergamon Press, Oxford, 1982.

[49] RENKA, R. J. Algorithm 772: STRIPACK: Delaunay triangulation and Voronoi diagram on the surface of a sphere. *ACM Transactions on Mathematical Software 23*, 3 (1997), 416–434. doi:10.1145/275323.275329.

[50] ROBINSON, V. B. Interactive machine acquisition of a fuzzy spatial relation. *Computers & Geosciences 16*, 6 (1990), 857–872. doi:10.1016/0098-3004(90)90008-H.

[51] ROBINSON, V. B. Individual and multipersonal fuzzy spatial relations acquired using humanâĂŞmachine interaction. *Fuzzy Sets and Systems 113*, 1 (2000), 133–145. doi:10.1016/S0165-0114(99)00017-2.

[52] ROBINSON, V. B. A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. *Transactions in GIS 7*, 1 (2003), 3–30. doi:10.1111/1467-9671.00127.

[53] ROSENFELD, A. The diameter of a fuzzy set. *Fuzzy Sets and Systems 13*, 3 (1984), 241–246. doi:10.1016/0165-0114(84)90059-9.

[54] ROSENFELD, A. Fuzzy geometry an updated overview. *Information Sciences 110*, 3-4 (1998), 127–133. doi:10.1016/S0020-0255(98)10038-5.

[55] ROSENFELD, A., AND MELTER, R. A. Digital geometry. *The Mathematical Intelligencer 11*, 3 (1989), 69–72. doi:10.1007/BF03025195.

[56] RUŽIČKA, M. Anwendung mathematisch-statistischer methoden in der geobotanik. *Biologia, Bratislava 13* (1958), 647–661.

[57] VIRTANEN, P., GOMMERS, R., OLIPHANT, T. E., HABERLAND, M., REDDY, T., COURNAPEAU, D., BUROVSKI, E., PETERSON, P., WECKESSER, W., BRIGHT, J., VAN DER WALT, S. J., BRETT, M., WILSON, J., MILLMAN, K. J., MAYOROV, N., NELSON, A. R. J., JONES, E., KERN, R., LARSON, E., CAREY, C. J., POLAT, I., FENG, Y., MOORE, E. W., VANDERPLAS, J., LAXALDE, D., PERKTOLD, J., CIMRMAN, R., HENRIKSEN, I., QUINTERO, E. A., HARRIS, C. R., ARCHIBALD, A. M., RIBEIRO, A. H., PEDREGOSA, F., VAN MULBREGT, P., VIJAYKUMAR, A., BARDELLI, A. P., ROTHBERG, A., HILBOLL, A., KLOECKNER, A., SCOPATZ, A., LEE, A., ROKEM, A., WOODS, C. N., FULTON, C., MASSON, C., HAGGSTROM, C., FITZGERALD, C., NICHOLSON, D. A., HAGEN, D. R., PASECHNIK, D. V., OLIVETTI, E., MARTIN, E., WIESER, E., SILVA, F., LENDERS, F., WILHELM, F., YOUNG, G., PRICE, G. A., INGOLD, G.-L., ALLEN, G. E., LEE, G. R., AUDREN, H., PROBST, I., DIETRICH, J. P., SILTERRA, J., WEBBER, J. T., SLAVIA, J., NOTHMAN, J., BUCHNER, J., KULICK, J., SCHÖNBERGER, J. L., DE MIRANDA CARDOSO, J. V., REIMER, J., HARRINGTON, J., RODRÍGUEZ, J. L. C., NUNEZ-IGLESIAS, J., KUCZYNSKI, J., TRITZ, K., THOMA, M., NEWVILLE, M., KEMMERER, M., BOLINGBROKE, M., TARTRE, M., PAK, M., SMITH, N. J., NOWACZYK, N., SHEBANOV, N., PAVLYK, O., BRODTKORB, P. A., LEE, P., MCGIBBON, R. T., FELDBAUER, R., LEWIS, S., TYGIER, S., SIEVERT, S., VIGNA, S., PETERSON, S., MORE, S., PUDLIK, T., AND OSHIMA, T. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods 17*, 3 (2020), 261–272. doi:10.1038/s41592-019-0686-2.

[58] WANG, F., AND HALL, G. B. Fuzzy representation of geographical boundaries in GIS. *International Journal of Geographical Information Systems 10*, 5 (1996), 573–590. doi:10.1080/02693799608902098.

[59] ZADEH, L. A. Fuzzy sets. *Information and Control 8*, 3 (1965), 338–353. doi:10.1016/S0019-9958(65)90241-X.

[60] ZIZKA, A., ANTUNES CARVALHO, F., CALVENTE, A., ROCIO BAEZ-LIZARAZO, M., CABRAL, A., COELHO, J. F. R., COLLI-SILVA, M., FANTINATI, M. R., FERNANDES, M. F., FERREIRA-ARAÚJO, T., GONDIM LAMBERT MOREIRA, F., SANTOS, N. M. C., SANTOS, T. A. B., DOS SANTOS-COSTA, R. C., SERRANO, F. C., ALVES DA SILVA, A. P., DE SOUZA SOARES, A., CAVALCANTE DE SOUZA, P. G., CALISTO TOMAZ, E., VALE, V. F., VIEIRA, T. L., AND ANTONELLI, A. No one-size-fits-all solution to clean GBIF. *PeerJ 8* (2020), e9916. doi:10.7717/peerj.9916.