

RESEARCH ARTICLE

Optimising sampling designs for habitat fragmentation studies

Ellen Bowler¹  | Veronique Lefebvre^{1,2} | Marion Pfeifer^{1,3}  | Robert M. Ewers¹¹Imperial College London, Ascot, UK²Flowminder Foundation, Stockholm, Sweden³School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK

Correspondence

Ellen Bowler

Email: e.bowler@uea.ac.uk

Present address

Ellen Bowler, School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Funding information

FP7 Ideas: European Research Council, Grant/Award Number: 281986; BBSRC, Grant/Award Number: BB/S014586/1

Handling Editor: Nick Isaac

Abstract

1. Habitat fragmentation has become one of the largest areas of research in conservation biology. Empirical studies into habitat fragmentation impacts typically measure ecological responses to metrics describing fragmentation processes, for example 'distance to the nearest forest edge', 'forest fragment area' and 'landscape habitat amount'. However, these studies often fail to sample across representative ranges of fragmentation metrics characterising the study region. They therefore lack the data to account for correlation among multiple fragmentation metrics and for spatial autocorrelation among sample sites, which reduces the strength of derived predictive models.
2. Here, we draw on approaches used in the mining and soil science industry to develop standardised and repeatable protocols for designing optimised sampling schemes of biodiversity in fragmented landscapes that meet three criteria: the distance between sample sites is maximised to reduce spatial autocorrelation, the full range of values of the metrics of interest are sampled and the confounding effects of correlated metrics are minimised.
3. We show that our computational methods can optimise the placement of sample sites in fragmented landscapes to minimise, and in some cases to entirely avoid, over- or under-sampling of fragmentation metrics. Our method is flexible enough to cater for any continuous (e.g. maps of percentage tree cover) or categorical (e.g. habitat and land use types) fragmentation metric, and to simultaneously handle combinations of multiple fragmentation metrics and habitat types. We implement our methods as open-source code which includes options to mask invalid or inaccessible regions, update designs to adapt to unforeseen constraints in the field and suggest optimal numbers of sample sites for given design criteria.
4. Using a case study landscape, we demonstrate how the approach improves on manually generated sampling designs. We also show that the methods are flexible enough to be applied to landscape studies beyond the field of habitat fragmentation. We introduce our package as a novel research tool that is able to streamline the experimental design process for biodiversity sampling and monitoring at landscape scales, leading to improved data quality and representativeness.

KEYWORDS

GIS, habitat fragmentation, open-source software, population ecology, sampling, site selection, study design, survey

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

1 | INTRODUCTION

Empirical research in landscape ecology focusses on establishing relationships between ecological attributes (e.g. species abundance, community composition) and spatial patterns of the study landscape (e.g. habitat amount, road density) to understand ecological processes (Albert et al., 2010; Ewers et al., 2010). A diversity of models has been developed to formalise these relationships using them, for example, to predict the potential distribution of rare species, to identify species' habitat suitability and quantify species abundance responses to habitat fragmentation. A critical assumption in many if not all of these models is that the full range of metrics used to describe spatial landscape patterns has been sampled systematically or randomly (Albert et al., 2010; Pasher et al., 2013). Yet, it is well-documented that logistic constraints (e.g. time, money and site accessibility), and a lack of unified and standardised methodology, can lead to 'convenience sampling' designs (Rosenstock et al., 2002; Schreuder et al., 1999) that may violate key statistical assumptions (Albert et al., 2010; Eigenbrod et al., 2011; McGarigal & Cushman, 2002; Pasher et al., 2013). This can lead to a high degree of model uncertainty, affecting the strength, significance and in some cases the direction of modelled relationships (Eigenbrod et al., 2011; Pasher et al., 2013), ultimately resulting in misplaced guidance for conservation and policy practice.

Spatially biased sampling is particularly prevalent in landscape ecology research focussed on the ecological consequences of habitat loss and fragmentation, where metrics describing both landscape features are typically highly correlated (Didham et al., 2012; Ruffell et al., 2016). Both habitat loss and fragmentation are frequently understood to be the greatest threats to biodiversity in terrestrial ecosystems (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2019), and their effective management requires that we reduce uncertainty in models explaining their relationships with biodiversity metrics (Fletcher et al., 2018). Yet, despite the profusion of literature and data on the subject, a lack of unified standards in methods and metrics used to quantify biodiversity, fragmentation or modelling their relationships has hindered progress (Ewers & Didham, 2006; Fahrig, 2003, 2017; Mortelliti et al., 2010). Fragmented habitats present a complex landscape mosaic encompassing patches of different shape, size and configuration. Data tend to be collected from point locations within wider, more ambiguously defined landscapes, with pre-conceived assumptions on what constitutes a 'habitat' or a 'landscape' for target species (Brudvig et al., 2017; Fardila et al., 2017). Even experimental studies quantify fragmentation using different metrics (e.g. patch size, patch shape, distance to edge and habitat amount; Arroyo-Rodríguez & Dias, 2010; Chalfoun et al., 2002; Fahrig, 2003; Pfeifer et al., 2014). As fragmentation metrics are often correlated (e.g. sites located far from habitat edges can only be found in large fragments), particular values and combinations of metrics are often under-sampled, or at times even omitted (Eigenbrod et al., 2011; Pfeifer et al., 2014).

We suggest research on habitat fragmentation will benefit from the development of robust and standardised sampling protocols

that minimise correlations among landscape metrics when trying to establish relationships with biodiversity. Here, we introduce a tool that can be used for this purpose. We discuss key assumptions which should be met by sampling designs for fragmentation studies, and we outline an automated approach that optimises site placement given these assumptions. Our approach is well-positioned to make a positive contribution to studies designed to address the contentious issues of patch versus landscape and fragmentation versus habitat amount effects (Didham et al., 2012; Fahrig, 2003, 2017; Fletcher et al., 2018), that is studies that rely on landscape metrics reflecting both the characteristics of individual patches and of the wider landscapes in which they are embedded. We note, however, that our general framework can equally be applied to other conceptual models in use in landscape ecology, such as 'continuum' or hybrid models (Brudvig et al., 2017; Manning et al., 2004), as well as other types of studies beyond the fields of fragmentation and landscape ecology such as examining links between geodiversity and biodiversity (Read et al., 2020).

We believe sampling designs need to meet three key criteria to enable robust predictive modelling: account for spatial autocorrelation, capture the full range of predictor variables and minimise correlation between predictor variables. First, spatial autocorrelation (Legendre, 1993) violates the assumption of independent data points, which forms a central assumption in most statistical procedures. Locations which are closer in space tend to be more similar than locations which are further apart (Bridges et al., 2007; F. Dormann et al., 2007; O'Sullivan & Unwin, 2010). Yet, few empirical fragmentation studies account for spatial autocorrelation in either analysis or experimental design (McGarigal & Cushman, 2002; Mortelliti et al., 2010). To minimise the likelihood of collecting non-independent data, sample sites should be spaced at the maximum possible separation in the study region, which has the co-benefit of increasing the likelihood for sampling the full metric space of variables characterising the heterogeneity of the landscape. Second, sample sites should cover the full range of values in the metrics of interest (Albert et al., 2010; Eigenbrod et al., 2011). For example, in a study investigating the relationship between species abundance and area of habitat fragment, sample sites should ideally be spread evenly between small, medium and large fragments (in a simplified non-continuous case). Sampling bias, failure to sample in all ranges of the metric or sampling disproportionately in one part of the range could lead to misclassification of the modelled relationship (Eigenbrod et al., 2011). For instance, a linear increase in abundance may be observed if sampling only medium and large fragments, while additional sites in smaller fragments could reveal a sigmoidal or even exponential relationship. Third, collinearity among fragmentation metrics is widespread (Riitters et al., 1995), but often treated as a nuisance in landscape ecology (Ruffell et al., 2016). If correlations arise due to causal relationships between predictor variables, then accounting for those correlations is crucial to accurately quantify the impact of fragmentation and landscape metrics on measured response variables (Didham et al., 2012; Ruffell et al., 2016). Hence, sites should be selected to minimise (Pasher et al., 2013), or at least

control for (Box & Hunter, 1957), correlations among multiple fragmentation metrics.

In practice, designing sampling schemes for habitat fragmentation studies which fulfil these requirements can be challenging. We implemented a cursory review of fragmentation studies published between 2014 and 2017 (limiting ourselves to studies that collected primary field data and identified through the key words 'habitat fragmentation' and 'sampling'). Of the 25 most cited papers, none described automated methods for optimising sample site placement along ranges and combinations of metric values in the landscape. Automated approaches to site selection have been used to design national biodiversity surveys (Gillespie et al., 2017), but these have been developed in a specific context and do not represent a general solution to sampling design challenges. Our tool provides a solution, offering a standardised, repeatable method that can allocate site placement in sampling designs that account for the three key criteria outlined above.

The mining and soil science industries have experience in the optimisation of sampling designs using mapped data for many years. Their sampling surveys often cover vast areas, and so designing schemes which maximise the amount of information at the minimum expense of time and resources has been critical for business (Clifford et al., 2014). Focus is often placed on using remotely sensed data (e.g. curvature, terrain ruggedness) as environmental covariates for predicting soil properties. This pre-existing data can then be used to generate informed sampling designs. Minasny et al. (2007), for example, proposed *quadtree decomposition* to focus sampling effort towards areas of covariate maps with more variation in values, as well as *Latin hypercube* sampling (Minasny & McBratney, 2006) which could be conditioned to take sample site accessibility into account. Similar methods which optimise the geographic spread of data by minimising the maximum Kriging variance have also been developed (Van Groenigen, 2000; Van Groenigen et al., 1999), these approaches maximise the distance between sample sites while optimising the amount of landscape variation encompassed in the design.

In the mining and soil science industries, a relationship between covariate maps and sampled soil properties is assumed, and often modelled from previous experiments, to inform the design of the sampling strategy. In most fragmentation studies, however, establishing this link (between observed landscape patterns and collected biodiversity data) is the targeted end point of the study rather than the beginning. Despite this apparent difference in objectives, we suggest that similar concepts can be applied to the design of fragmentation studies, using landscape metric maps derived through remote sensing (Ewers et al., 2010; Pfeifer et al., 2016, 2019; Riitters et al., 1995) to inform and optimise the placement of sampling sites in the field. We propose a solution that applies to continuous and categorical landscape metrics and optimises sample placement across multiple metrics and habitat types simultaneously. Our approach also accounts for practical sampling constraints imposed by the study region, such as excluding habitat types that are not of interest for a particular experiment (e.g. a study designed to sample in natural habitats but not in urban areas), and excluding areas that

are inaccessible (such as remote areas in rugged terrain, or private land parcels).

Our approach uses two design-generating algorithms: (a) a spatially uniform design algorithm (SUDA), which aims to evenly distribute sites in geographic space, thereby minimising spatial autocorrelation in data and achieving a proportional subsample of the metric space, and (b) a metric uniform design algorithm (MUDA), which aims to evenly distribute sites throughout the metric space, ensuring that the range of values present in the habitat fragmentation metrics are equally represented in the sampling design. We outline the methods behind the spatial and metric uniform designs and present their implementation in the form of open-source Python code.

We illustrate our approach using a fragmented landscape in New Zealand as a case study, and we show that both algorithms can be used to output optimised biodiversity sampling designs in terms of covering geographic space and the fragmentation metric space. We illustrate their flexibility in being able to account for sampling obstacles, as well as extensions to the methods which allow new information gained in the field to be used to adapt partially implemented designs. This is to cater for the case when some sites have been sampled, but a subsequent location is found to be inaccessible, or to complement previous studies. Finally, as an example of applications beyond habitat fragmentation, we use a digital elevation map to generate a MUDA design which samples an altitudinal gradient. This illustrates how the approach can easily be used to sample across other forms of environmental variation.

2 | MATERIALS AND METHODS

2.1 | Study region

We used the Hope River forest fragmentation landscape in New Zealand to validate the outputs of the sampling design algorithms. In this landscape, Ewers et al. (2007) investigated the effect of distance to forest edge and forest fragment area on beetle community composition using a transect design, where transects extended from the centre of the forest fragment, across the edge and an equivalent distance into the surrounding grassland habitat. We obtained a 30-m resolution invalid area mask (masking out areas above the tree line) and a binary habitat map (forest and grassland) of the study region using data compiled within the BIOFRAG database (Pfeifer et al., 2014). Both landscape maps represent an area of approximately 1,300 km² and were derived from satellite imagery collected in the year the original study was conducted. The binary habitat map was used to compute two continuous fragmentation metrics: the log₂ transformed distance to the nearest edge (m) and log₁₀ transformed fragment area (ha), which match those used in the original design (Ewers et al., 2007; Figure 1). While our case study here is built upon two patch-level fragmentation metrics, the same approach applies equally to landscape-level metrics such as the percentage of habitat cover in a given radius.

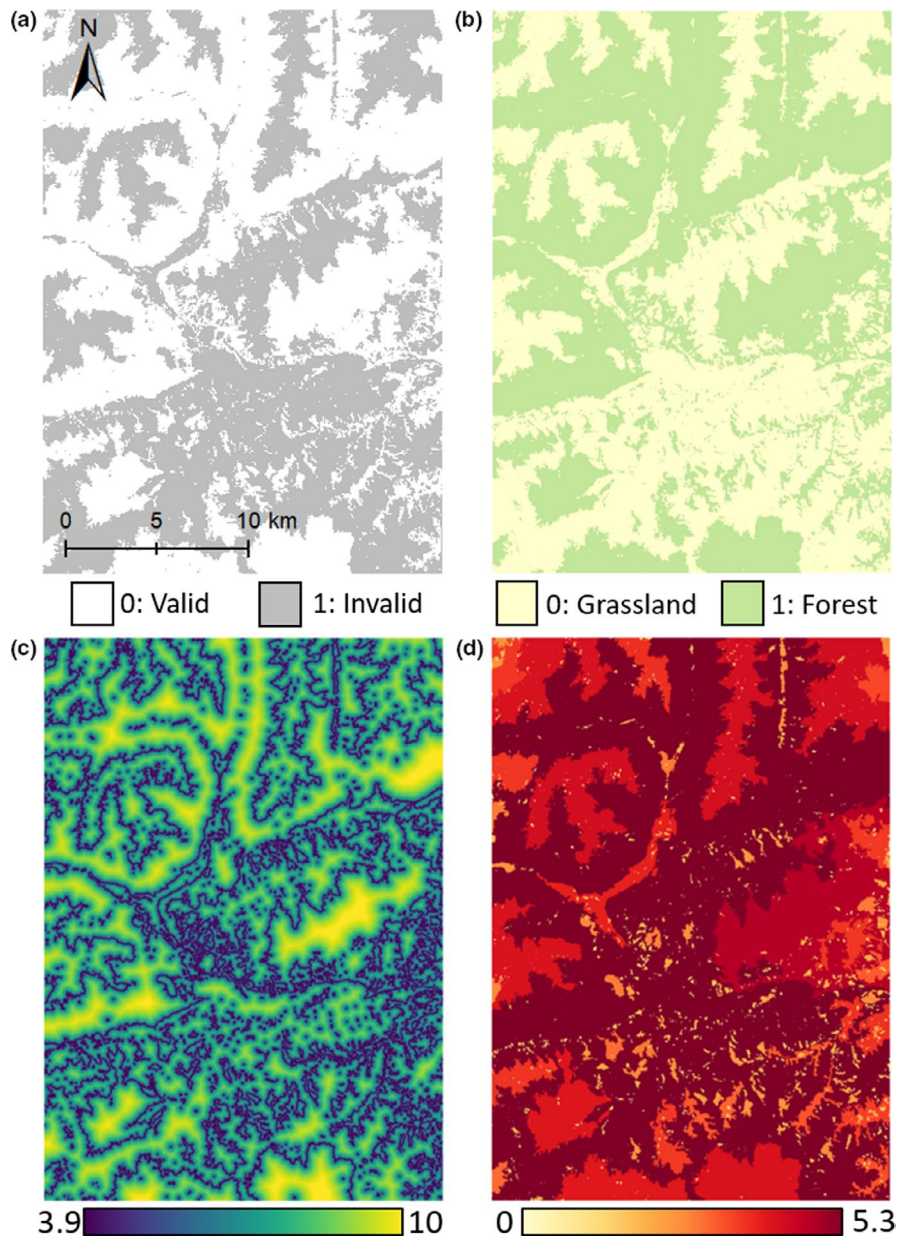


FIGURE 1 Example spatial inputs for the algorithms, using the case study of the Hope River study region (Ewers et al., 2007). (a) An invalid area mask (with 0 = valid and 1 = invalid), in this example, invalid denotes a non-focal habitat (grassland). (b) A categorical habitat map, showing forest and grassland. (c) Log_2 distance to the nearest habitat edge (m, measured positively in both habitats and up to 1,024 m extent). (d) Log_{10} transformed fragment area (ha)

2.2 | Algorithms for generating sampling designs

Our approaches use maps derived from remotely sensed data. These maps fall into three categories:

1. **Invalid mask:** A binary map (with 0 = invalid and 1 = valid) indicating areas of the study region which should not be sampled. Examples include water bodies, roads, inaccessible areas and non-focal habitats such as grasslands and urban areas (Figure 1a).
2. **Habitat map:** A categorical map classifying the study region into land-cover types (Figure 1b). Examples include forest, grasslands, urban areas and plantations.
3. **Metric map(s):** A map(s) which quantifies fragmentation characteristics. Examples include distance from nearest habitat edges

or the size of habitat patches (Figure 1c,d). Metrics can be either categorical or continuous in nature.

The first algorithm, the SUDA, is solely based on spreading sites as far apart as possible in any valid area of the study region (specified as area outside the 'invalid mask') to minimise the effect of spatial autocorrelation. The second algorithm, the MUDA, extends upon the first algorithm, focusing on placing an equal number of samples sites along gradients of the metric maps.

2.2.1 | The spatially uniform design algorithm

The SUDA iteratively spaces sites at the maximum possible geographic separation using the Euclidean distance transform, an operation which

takes a binary image (with 0 = non-feature and 1 = feature) as an input and outputs an image where every pixel is labelled with the distance to the nearest feature pixel. Euclidean distance from a non-feature pixel of spatial coordinates (x, y) to the nearest feature pixel (x_f, y_f) is:

$$D = \sqrt{(x - x_f)^2 + (y - y_f)^2}. \quad (1)$$

This results in a complete distance map, D_{map} .

The SUDA distributes samples in a naïve manner with respect to fragmentation metrics, which are not included as inputs for the algorithm. The only inputs required for the SUDA are an invalid mask (I_m) and an integer number of sample sites (n). The algorithm consists of four main operations (Figure 2a):

1. Use a pseudo-random number generator to select a site in the valid region of I_m .
2. Compute the Euclidean distance transform from the selected site(s), and apply the mask to exclude invalid regions.
3. Select the next site as the pixel with the maximum value in the masked Euclidean distance transform (i.e. the furthest away from

already placed sites). If there is more than one pixel with the same maximum value, then one is selected at random.

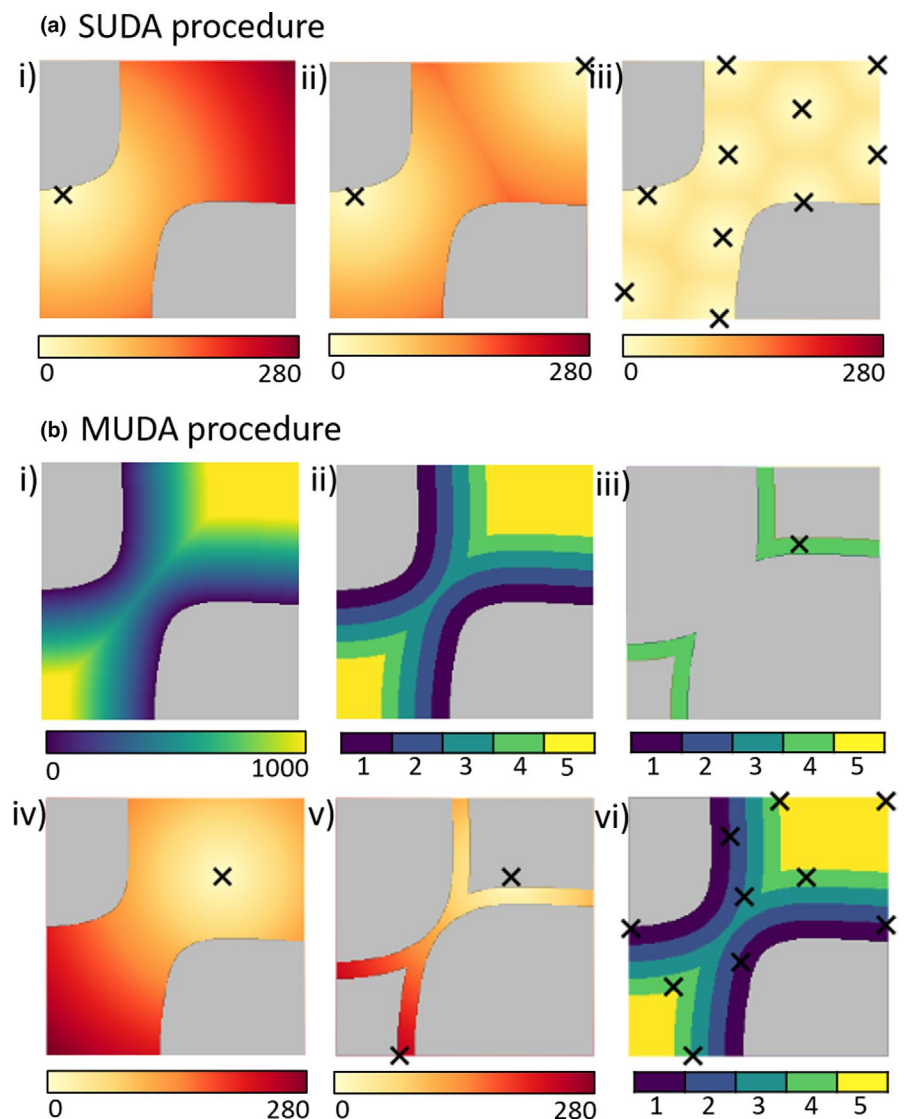
4. Check number of placed sites is lower than n and repeat steps (2) and (3).

The procedure ends when all n sites have been placed.

2.2.2 | Assessing variation in the SUDA outputs

In the SUDA, the initial sample site is selected at random in the valid region of I_m , hence each run of the algorithm will produce a different design. A further source of variation is introduced in step (3) when selecting a random maximum value pixel if multiple equivalent maximums exist. To compare the final designs emerging from multiple runs of the SUDA, and to assess how evenly sites are spread in geographic space, we implemented a scoring criterion based on minimising the maximum distance. This requires computing the Euclidean distance transform of the final design, applying the invalid mask and extracting the maximum distance value from the valid region as the

FIGURE 2 The (a) spatially uniform design algorithm (SUDA) procedure and (b) metric uniform design algorithm (MUDA) procedure on a 200×200 test grid, with invalid areas represented in grey. (a, i) Select a site at random in the valid region, compute the Euclidean distance transform from the chosen site and apply the mask. (ii) Select the pixel with the highest distance value as the next site. Recompute the Euclidean distance transform and apply the mask. (iii) Continue this procedure until all sample sites have been placed, resulting in the final design. (b, i) Distance to the nearest edge as a continuous metric, up to a maximum of 1,000 m. (ii) The ID image, with the continuous distance to edge categorised into five unique ID values representing 200-m intervals. (iii) Isolate the first ID in $list_{ID}$ (e.g. ID = 4) and select the first site at random. (iv) Compute the distance from the chosen site. (v) Isolate the next ID in $list_{ID}$ (e.g. ID = 3) and mask all values not corresponding to this region in the Euclidean distance transform image. Select the maximum distance value as the next site. (vi) Update the Euclidean distance transform and continue iterating through $list_{ID}$ until all sites have been placed



score. Designs which score a smaller maximum distance will have a more uniform spread, whereas those with the same number of sites but an element of clustering will have higher scores. This method is based on principles proposed by Van Groenigen et al. (1999), who used minimisation of the maximum Kriging variance as a congruent measure.

Using the Hope River landscape, we ran 10,000 simulations of the SUDA to examine the extent of variation in the final sampling designs and compared them to 10,000 simulations in which the same number of sites ($n = 45$) were randomly distributed. We found that the algorithm returned consistently optimised results, with small variation between final scores relative to those of a random sampling design (Figure 3). The distributions were significantly different, with no overlap between the maximum SUDA score (2.49 km) and the minimum random design score (2.79 km). In addition, the range of scores in the SUDA designs was small (0.59 km) compared to that of the random designs (7.55 km). For this reason, the variation emerging among repeated applications of the SUDA was considered to be at a reasonable level.

2.2.3 | The metric uniform design algorithm

The primary aim of the MUDA is to place an equal number of sample sites across gradients of the chosen fragmentation metrics, thereby uniformly subsampling the metric space. Within this constraint, sites are also maximally separated into geographic space using an adapted version of the SUDA described above. Similar to the SUDA, the methods take an invalid mask (I_m) and an integer number of sample sites (n) as inputs. Additionally, the user enters k metric maps of interest ($M_1, \dots, M_k$) and a corresponding integer number of bins ($B_1, \dots, B_k$). These bins are used to specify the number of fixed width intervals the range of each metric should be broken into. For example, specifying 10 bins for fragment area ranging between 0 and 10 km² would result in sampling intervals of 0–1, 1–2, ... 9–10 km². We first describe the methods for a 'one metric case ($k = 1$)' (Figure 2b). In the following section, we detail how this extends to multiple metrics and bins. We describe the algorithm in three main steps.

Step 1: Identify metric bin combinations and create ID image

The invalid mask I_m is applied to the metric map (M_j). The total range of the remaining values in the metric is then divided equally into B_j bins, with the width of each bin (w_j) being calculated as:

$$w_j = \frac{\max(M_j) - \min(M_j)}{B_j}. \quad (2)$$

Continuous metrics are converted to discrete values by replacing all values which fall within a certain bin a unique value corresponding to that bin. For example, if M_j was distance to the nearest edge with a maximum distance of 1 km and if B_j was set to 5, values between 0 and 200 m would be labelled as 1, 200–400 m as 2, 400–600 m as 3, 600–800 m as 4 and 800–1,000 m as 5. For categorical metrics, such

as habitat maps, B_j should be equal to the number of unique categories present in the metric. In some cases, bins may be 'empty'; for example, if M_j represented fragment area in a landscape which contained only small and large fragments, then bins relating to medium-sized fragments would not be present in the landscape. These empty bins are removed from consideration, and the remaining non-empty bins are assigned a unique identification number (ID). This generates the final ID image (ID_{im}) which contains an integer number of unique IDs (denoted as N_{ID}).

Step 2: Assign sample sites to IDs

The aim is to place an equal number of sample sites in each ID. The optimum number of sample sites in each ID, S_{opt} , is calculated as:

$$S_{opt} = \frac{n}{N_{ID}}. \quad (3)$$

Since the number of sample sites must be discrete, this can only be achieved in practice if S_{opt} is an integer. Users may experiment with adjusting the number of bins for each metric, which will alter N_{ID} and bring the optimal sample sites within their desired range. In the case of non-integer S_{opt} values, the lower and upper bounds for n are calculated as:

$$n_{lower} = n \times \text{floor}(S_{opt}), \quad (4)$$

$$n_{upper} = n \times \text{ceiling}(S_{opt}). \quad (5)$$

One of these two values can be selected to replace the existing value of n , resulting in an integer S_{opt} . If the user wishes to continue with the original value of n , then n_{lower} sites will be placed in each ID and the remainder ($n - n_{lower}$) will be randomly distributed between the IDs.

Step 3: Place sample sites in geographic space within each ID

As we now know the number of sample sites required in each ID, we need to select specific site locations in the study landscape. To do this, we first generate a list where each ID is repeated S_{opt} times. For example, if $n = 10$ and $N_{ID} = 5$ (and so $S_{opt} = 2$ using Equation 3), then the ID list would be 1, 1, 2, 2, 3, 3, 4, 4, 5, 5 (as in Figure 2b). The order of this list is randomly permuted, resulting in the final ID list (hereafter referred to as $list_{ID}$).

The algorithm continues in a similar way to the SUDA; however, in this case, we iterate through $list_{ID}$ and constrain sites to fall in specific regions of ID_{im} . These regions can be isolated by applying the following function to each pixel, p , in ID_{im} ,

$$p = ID_{im}(x, y),$$

$$f(p) = \begin{cases} \text{valid, } p = ID \\ \text{invalid, } p \neq ID \end{cases}, \quad \text{for the ID selected in } list_{ID}. \quad (6)$$

This creates an ID mask (ID_{mask}). We iterate through the values in $list_{ID}$ and implement the following procedure (illustrated in Figure 2b):

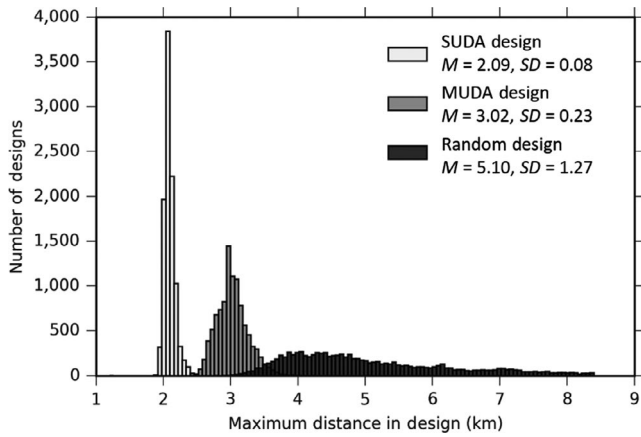


FIGURE 3 Comparing the distributions of maximum distance scores for 10,000 spatially uniform design algorithm (SUDA), metric uniform design algorithm (MUDA) and random designs, using the Hope River study region and 45 sample sites. In this example, the MUDA designs were generated using $\log_2$ distance to edge with seven bins and $\log_{10}$ fragment area with 10 bins

1. Select the first ID in $list_{ID}$ and create ID_{mask} . Use a pseudo-random number generator to select a pixel in the valid region.
2. Compute the Euclidean distance transform image from the selected site(s). Isolate the next ID in $list_{ID}$ and create ID_{mask} . Apply ID_{mask} to the Euclidean distance transform image to mask invalid regions.
3. Select the next site as the pixel with the maximum value in the masked Euclidean distance transform. If there are multiple pixels with the same maximum value, then one is selected at random.
4. Repeat steps (2) and (3) until all IDs in $list_{ID}$ have been selected.

2.2.4 | Extending the MUDA to work with k multiple metrics

Since all metrics ($M_1, \dots, M_k$) represent the same landscape, each pixel has k different values; one for each of the k fragmentation metrics. Every pixel in the landscape can therefore be plotted in k -dimensional metric space. By breaking the ranges of each metric into their specified number of bins ($B_1, \dots, B_k$), as in Equation 2, we can assign ID values based on the combinations of bins in the metric space (Figure 4: an example with two metrics, M_1 and M_2). The total number of bins present in the metric space can be calculated as:

$$N_{bins} = \prod_{i=1}^k B_i \quad (7)$$

In real landscapes, it is unlikely that all the bin combinations in the metric space exist in the landscape image (e.g. empty bins in Figure 4a). These bins are removed from consideration, and non-empty bins are allocated a unique ID. All pixels are then assigned the value of their corresponding bin, thus generating the ID image (ID_{im}) described in Stage

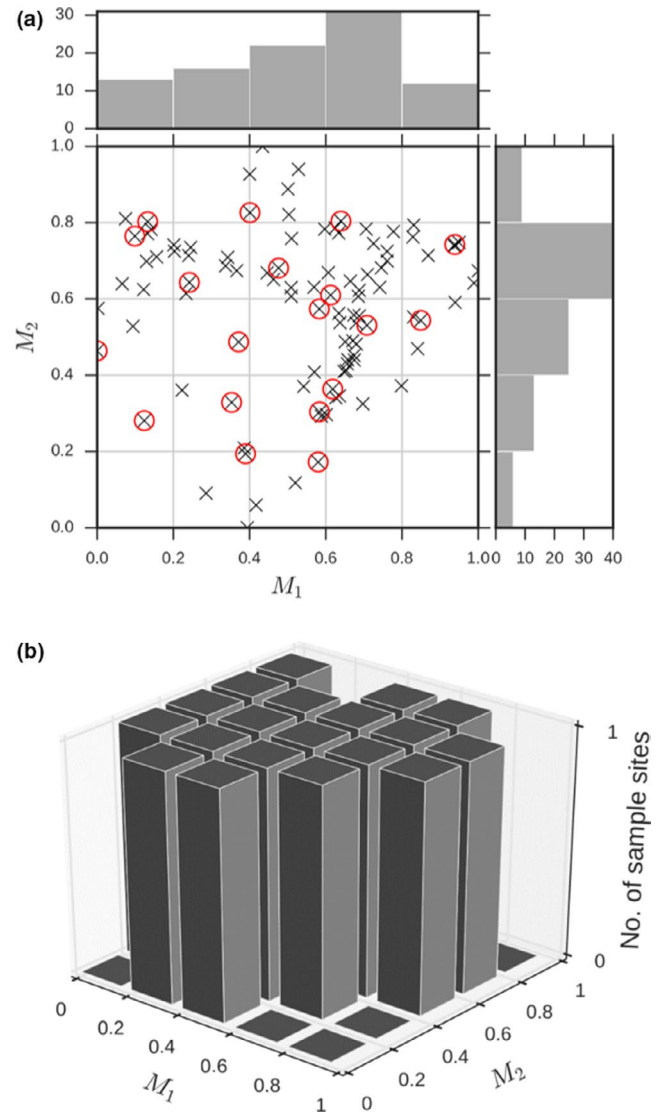


FIGURE 4 (a) Example of the distribution of pixels in metric space for two randomly generated metrics M_1 and M_2 . These metrics are 10×10 images (100 pixels) containing values between 0 and 1, with distributions broken into five bins each ($B_1 = B_2 = 5$). Crosses represent the location of all pixels in the two-dimensional metric space. To uniformly sample this space, an equal number of samples must be drawn from every non-empty bin (ID) in the metric space (in this case $N_{ID} = 19$), for example those circled in red. (b) The distribution of selected sample sites as a 3D histogram, showing that the number of sample sites selected from each ID forms a uniform distribution

1 of the MUDA procedure. From this point, the algorithm proceeds as detailed above in Stages 2 and 3.

2.2.5 | Assessing variation in the MUDA outputs

Since the number of sites placed in each ID is predetermined, the MUDA will always return designs which uniformly sample the metric space, with the maximum difference between the number of

sites in any two IDs being 1. Only the spatial arrangement of sites will differ between final layouts. In this case, variation is derived from the order in which the IDs appear in $list_{ID}$, which has been randomised, as well as the placement of the initial site and subsequent maximum Euclidean distance transform selections. As with the SUDA (Section 2.3.2), we computed 10,000 sampling designs using the MUDA and the Hope River landscape, in this case setting M_1 to be $\log_2$ distance to edge with $B_1 = 7$ and M_2 as $\log_{10}$ fragment area with $B_2 = 10$. The maximum distance scoring criteria showed an intermediate spatial spread of sites for MUDA designs ($M = 3.02$ km, $SD = 0.23$ km) compared to randomly generated designs ($M = 5.10$ km, $SD = 1.27$ km), suggesting a very highly significant difference between the distributions.

2.3 | Adapting SUDA and MUDA designs

Options to adapt partially completed sampling designs were developed for both the SUDA and MUDA. In this scenario, a researcher would like to update the entire study design but does not want to sacrifice the effort already expended to collect data. The methods take an updated invalid area mask (which includes the new inaccessible area), as well as the geographic coordinates corresponding to the original design, as inputs. The coordinates for the design are tagged as either 'sampled', 'unsampled' or 'inaccessible' (outlined in the online documentation). This approach allows researchers to 'force' the sampling design to factor in sites which have been sampled in the past, or for which the researcher already has data available. Rather than computing an initial Euclidean distance transform from a randomly selected site, the adapted SUDA and MUDA methods compute the Euclidean distance transform from all sites which have been marked as 'sampled'. Remaining sites will then be placed using the same iterative procedures but taking the new constraints into account.

3 | VALIDATION OF THE OPTIMISED SAMPLING DESIGN APPROACHES IN A CASE STUDY

3.1 | Case study design generation

We compared our algorithmically generated designs to a sampling design previously implemented in a fragmentation study (Ewers et al., 2007). The sampling design deployed in the study was designed with the express goal of sampling metric space equally, a goal directly analogous to the targeted output of the MUDA algorithm. The original field study sampled a total of 233 sites distributed across 15 forest fragments of different area, and with sites located along linear transects placed at up to 11 distances (0, 2, 4, 8, 16, 32, 64, 128, 256, 512 and 1,024 m) from the edge to the interior of the forest. Edge distance was $\log_2$ transformed and fragment area was $\log_{10}$ transformed in design and analysis. Transects were also

extended into the adjacent grassland at the same range of distances; however, the area of the grassland fragments was not considered.

To compare the original design to those output by the SUDA and MUDA, two simplifications were made: (a) due to the resolution of the maps (with one pixel representing 30 m² on the ground), plots separated by small edge distances (0 to 16 m) were merged to a single site; and (b) only sample sites within forest fragments were considered, since the area of grassland fragments was not incorporated into the original design. This resulted in a reduced set of 45 sites, covering seven edge distances and 15 fragment sizes.

We generated a MUDA design using $\log_2$ edge distance and $\log_{10}$ fragment area maps to match the choices made in the original study. The MUDA can be run equally on transformed or untransformed metrics, and this user choice should reflect their planned final analysis. For edge distance, we used seven bins, to match the original design. Using 15 bins for $\log_{10}$ fragment area would result in an optimum MUDA design of 55 sample sites, so we reduced the bin number to 10 to generate a design with 45 sample sites to ensure the sample size was the same as in the simplified original study. The invalid area mask was updated to include all grassland areas as well as the outer 5 km of the study region, which was not sampled in the original study. We ran both the SUDA and MUDA 20 times each and selected the best scoring designs as the final outputs. For comparison, we also computed a random design which consisted of 45 sample sites placed in any valid pixel with equal probability. In this case, sample site placement is random in both geographic space and in metric space.

3.2 | Measures for comparing study designs

To compare how successfully the case study designs meet the three key sample design criteria, we use four measures. For spatial autocorrelation, we use the minimum of the maximum distance criterion, as outlined in Section 2.3.2. Designs with the lowest maximum distance are more uniform in geographic space, which lowers the chance of autocorrelation (we note that the eventual realised level of autocorrelation can still be checked at a later stage in the study). To assess the coverage of metric space we report the percentage of metric bins sampled, with higher percentages indicating more successful coverage. In addition, we assess how uniformly sample sites are spread between metric bins by taking the standard deviation of the number of sample sites in each metric bin. A lower standard deviation equates to more uniform coverage. Finally, to assess correlations between metrics, we use Pearson's product-moment correlation (r), with lower r values indicating reduced correlation.

4 | RESULTS AND DISCUSSION

4.1 | Sampling geographic space

We present four different sampling design layouts on the Hope River landscape, along with their respective maximum distance

scores (Figure 5). The SUDA produced the best scoring design under the minimum of maximum distance criterion, with a maximum of 2.07 km (Figure 5a(i)) compared to 3.45 km achieved in the randomly generated example (Figure 5c(i)). The MUDA (Figure 5b(ii)) scored higher than the SUDA, as expected, with a maximum distance of 3.22 km. The original design naturally scores worst under this measure (8.69 km), due to the large separation between transects.

When possible, the MUDA is designed to seek out pixels which fall in the same ID, but that are at the maximum distance from all other points geographically. When there is just one suitable region in the landscape, then clusters of sites can occur (Figure 5b(ii)). In our case study, several fragment area bins were comprised of just a single fragment in the landscape, meaning all samples needed to capture the distance to edge combinations must be drawn from a small subset of the full study landscape. Such a limitation represents an unavoidable constraint in using natural landscapes for studies of habitat fragmentation effects, in much the same way that non-existent combinations of fragmentation metrics also represent a limitation on what can be analysed (Betts et al., 2019). Different

iterations of the MUDA will produce different configurations, however, so running multiple designs and manually selecting the best can mitigate this to some extent. We also imposed a threshold in the algorithm to ensure there was at least a 25 m minimum separation between sample sites in this case study. For different sampling methods and focal species, this threshold could easily be adjusted to tailor to the specific requirements of a given study.

The minimum of maximum distance scores for the SUDA and MUDA designs were selected after 20 repeat runs of the SUDA and MUDA. This was used as a method for obtaining the best sampling design from a set of candidate outputs. Increasing the number of repeats will increase the chance of obtaining better scoring designs; however, this will also lead to a rise in computation times. These times will also depend on the size and resolution of the landscape image, on the number of sample sites and on computer processing speed. Testing the time to generate a single design for a specific study can be used to gauge how many repeats are feasible if this method is used. To ensure that the global minimum is met, these designs could also be used as initial conditions in an automated

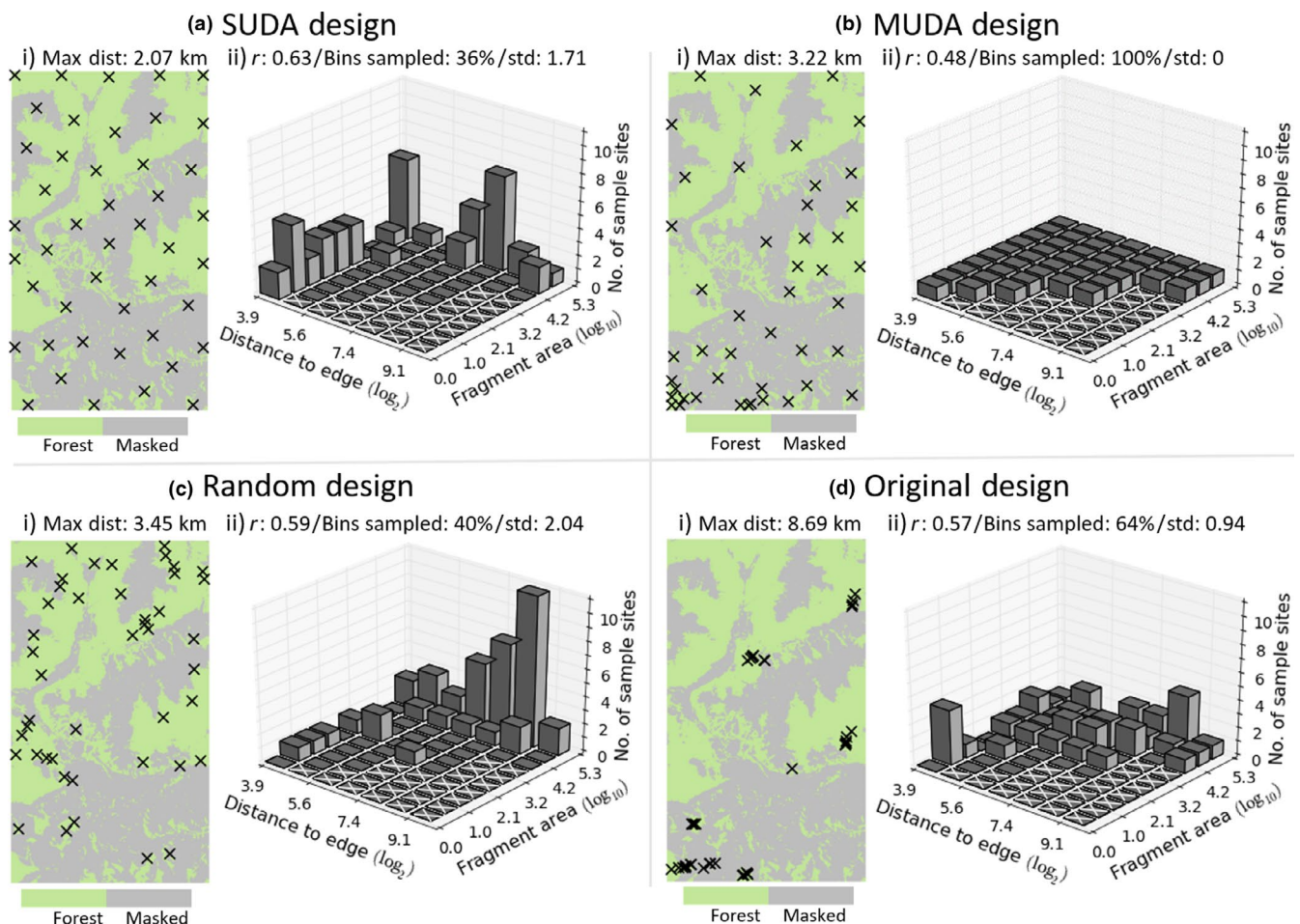


FIGURE 5 Comparing the (a) spatially uniform design algorithm (SUDA) and (b) metric uniform design algorithm (MUDA) design outputs to a (c) random design and (d) a simplified version of the original transect design, using the Hope River landscape. All designs were generated using $\log_2$ distance to edge (with seven bins), $\log_{10}$ fragment area (with 10 bins) and 45 sample sites. Panels show (i) the distribution of sample sites geographically, with the maximum distance score; (ii) the distribution of sample sites in the metric space, with the Pearson's product-moment correlation coefficient (r), the percentage of metric space bins sampled and the standard deviation of the distribution (std)

optimisation procedure. For example, van Groenigen and Stein (1998) used spatial simulated annealing to optimise the spread of points by minimising the maximum Kriging variance. This method, however, was not adapted to suit sampling in specific ranges of the metric space, and so would not represent a simple extension to the MUDA implementation. Given the speed of the SUDA, generating multiple designs and selecting the one with the minimum score is a viable alternative if assurance of a local minimum is sought.

4.2 | Sampling the metric space

Using the MUDA, we were able to generate a design which uniformly subsampled the metric space, covering a wider range of metric values more evenly in comparison to the original sampling design. Comparing the standard deviations of the distributions presented in Figure 5, we achieve a perfect uniform distribution ($\text{std} = 0$) for the MUDA design (Figure 5b(ii)) compared to the original ($\text{std} = 0.94$, Figure 5d(ii)). These are both improvements on the random ($\text{std} = 2.04$, Figure 5c(ii)) and SUDA ($\text{std} = 1.71$, Figure 5a(ii)) design outputs. To compare metric coverage, we assessed the number of bins sampled by each design. Since high distance to edge values cannot occur in small fragments, these combinations do not exist in the landscape and are necessarily absent from the distributions. Despite this, the MUDA was able to identify sampling locations in all 45 bin combinations which did exist (100%), compared to just 30 (67%) that were sampled in the original transect design. Sampling distributions using the random (18 bin combinations, 40%) and the SUDA (16, 36%) approaches were even worse, representing very poor coverage of fragmentation metric combinations. Sampling in all metric combinations, and in particular sampling the full range of all metrics (as present in the landscape), was shown to be an important factor in obtaining robust inferences by Eigenbrod et al. (2011). Sampling only a proportion of the potential range was shown to have the greatest effect on the strength of inferences, in some cases altering the sign of relationships between predictor variables and species responses. The MUDA proved to be a viable method for ensuring that these maximum ranges are encompassed in study designs. It was also able to produce more efficient designs, with the same number of sites being able to span the metric space more evenly compared to the original.

Distributing the sites in this manner has the additional benefit of minimising the correlation between metric values at the sample locations (Pasher et al., 2013). Pearson's product-moment correlations between distance to edge and fragment area showed that the MUDA had a lower correlation between metrics ($r = 0.48$, $p < 0.001$) compared to the original design ($r = 0.57$, $p < 0.001$). The SUDA ($r = 0.63$, $p < 0.001$) and random ($r = 0.59$, $p < 0.001$) designs had the highest correlation scores, which is unsurprising given these sampling designs were generated in a naïve manner with no consideration of the metrics. There is a strong level of correlation among many fragmentation metrics (Riitters et al., 1995), exemplified in our case study by the fact that high distance to edge and small fragment

area combinations cannot exist. While no sampling design can compensate for this feature of real-world landscapes, the MUDA proved an effective method of minimising the confounding effect that such correlations would have on subsequent analyses of field data.

4.3 | Adapted designs

We test the adapted design method, we added an area of private land (in this example covering 12% of the study region) to the invalid area mask after designating eight out of 45 sample sites as being already sampled. The adapted SUDA was able to keep the eight previously sampled sites fixed and reallocate the remaining 37 sample sites to find a new optimal layout. We achieved a maximum distance score of 1.95 km, which was roughly the same as in the original SUDA design in Figure 5a(i). Similarly, the adapted MUDA (max dist = 3.05 km) was able to relocate the inaccessible sample sites, in this case to locations with equivalent ID values which fell elsewhere in the landscape. In both cases, coverage of the metric space and the correlation between metrics were the same as the original versions of the SUDA and MUDA designs (Figure 5a(ii),b(ii)).

In some cases, it may be that the new invalid region masks out an entire ID, meaning that no other locations are available. If this occurs, the sample site corresponding to that ID will be removed from the design, resulting in either an empty bin in the metric space or an imperfect uniform distribution if the ID has been previously sampled. Since such IDs are likely to represent only a very small proportion of the total landscape area, omitting to sample in these outlying regions is unlikely to have a great effect.

4.4 | Further examples

So far, we have presented examples using two fragmentation metrics (distance to edge and fragment area) and sampling in just one habitat (forest), but the MUDA is flexible to cater to different design requirements. Figure 6 shows examples of designs covering different combinations of metrics and habitat types, which can all be generated using the same MUDA procedure. Single metric designs were produced for both one (Figure 6a) or two (Figure 6c) habitat types, with Figure 6b,d showing how this can also be achieved in the multiple metric case. The MUDA can generate designs for more than two metrics and habitats. However, such designs may require large numbers of sample sites to match the increasing numbers of ID combinations. This can be compensated for by reducing the number of bins per metric, or accepting suboptimal uniform distributions if resources limit the maximum number of sites for the study.

Finally, in Figure 6d, we show an example of how the MUDA can be used with metrics not specific to fragmentation. In this case, we use altitude (m a.s.l.), which was derived from a digital elevation map (DEM) of the selected study region, along with fragment area as used previously. The number of bins for each metric was set to 10, resulting in an 83-site design. As is evident from the histogram (Figure 6d),

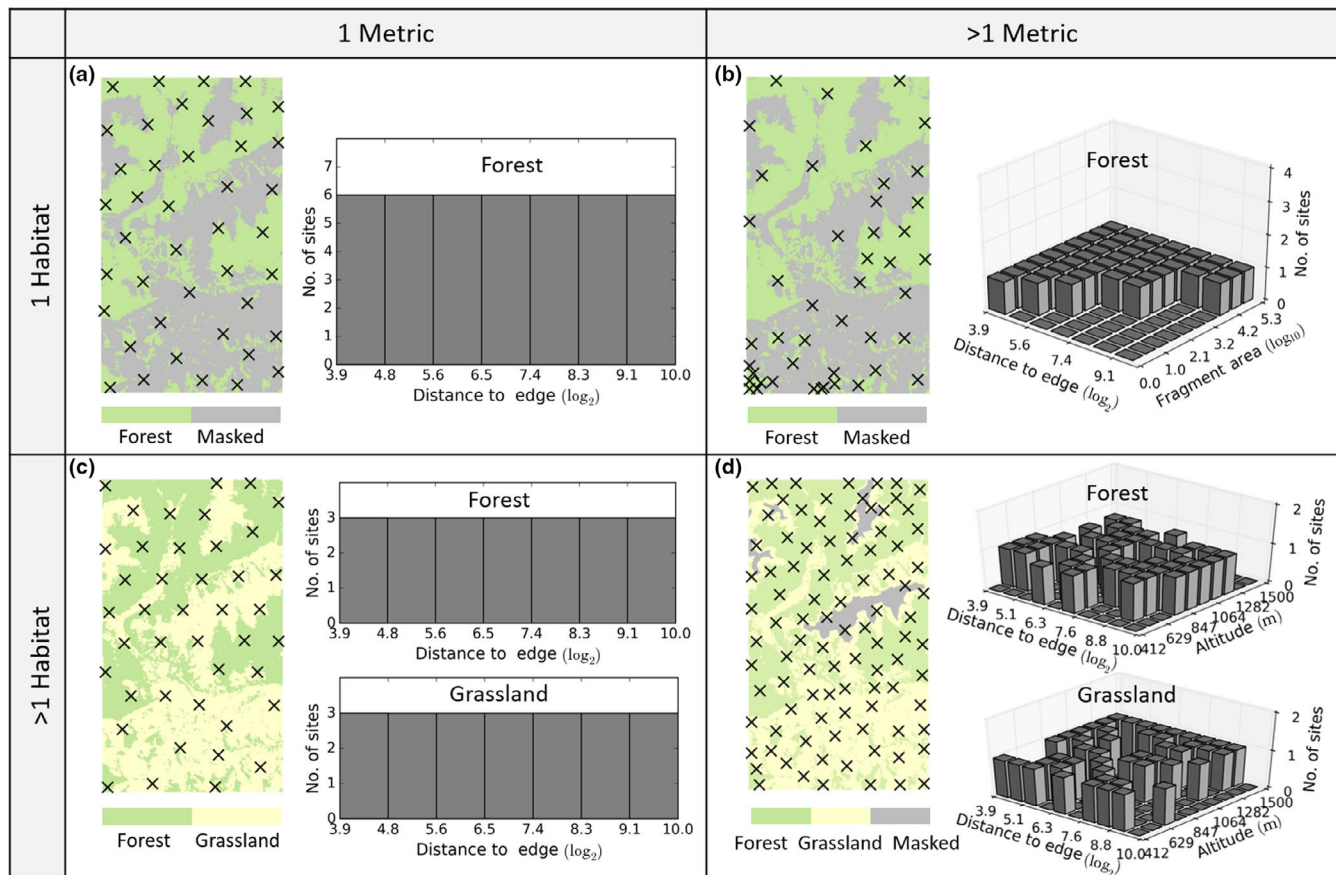


FIGURE 6 Metric uniform design algorithm designs for different combinations of metrics and habitats, using the Hope River landscape. Maps show the distribution of sampling sites geographically, histograms show the distribution of sampling sites in the metric space. We present (a) a 42-site forest only design using distance to edge; (b) a 45-site forest only design with distance to edge and fragment area; (c) a 42-site design using distance to edge in forest and grassland habitats and (d) an 83-site design using distance to edge and altitude (taken from a digital elevation model) in both habitats, masking areas >1,500 m a.s.l.

altitude is not strongly correlated with fragment area, so the sample sites are spread more evenly in the metric space compared to Figure 6b, for example. For the invalid mask, we used the same DEM to exclude all areas above 1,500 m, as an example of how different inaccessible regions can be masked from designs.

5 | CONCLUSIONS

The algorithms developed in this study provide a means of addressing key pitfalls in fragmentation sampling designs, such as accounting for spatial autocorrelation (Mortelliti et al., 2010), sampling the full range of the metrics (Eigenbrod et al., 2011) and reducing the collinearity among multiple metrics (Ruffell et al., 2016). We have shown that using maps derived from remotely sensed data can provide a robust method for assessing the global configuration of metrics in the landscape, with the MUDA providing more complete and more even coverage of the metric space compared to traditional, non-automated methods of developing sampling designs. Our case studies showed that the methods are flexible in being able to optimise site placement given: (a) any combination of continuous and/

or categorical metric maps; (b) multiple metrics and habitat types in combination; and (c) bounds or limitations imposed by the study region and site accessibility. We propose that these methods provide a practical means of streamlining the sampling design process not only for fragmentation experiments, but in other landscape ecology studies. This can help to ensure that key statistical assumptions are met and allow the most robust conclusions to be drawn from collected data.

ACKNOWLEDGEMENTS

R.M.E. and M.P. were supported by European Research Project number 281986; M.P. is now supported by BBSRC award BB/S014586/1; E.B. conducted work as part of an MSc thesis supervised by the co-authors.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

AUTHORS' CONTRIBUTIONS

E.B., V.A.L., M.P. and R.M.E. contributed to the conceptualisation and implementation of the study; E.B. and V.A.L. developed the

computational methods; E.B. led the writing of the manuscript with all authors providing revisions.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13731>.

DATA AVAILABILITY STATEMENT

Source code, example datasets and documentation are available at <https://doi.org/10.5281/zenodo.5503985> (Bowler, 2021).

ORCID

Ellen Bowler  <https://orcid.org/0000-0002-9681-1355>

Marion Pfeifer  <https://orcid.org/0000-0002-6775-3141>

REFERENCES

- Albert, C. H., Yoccoz, N. G., Edwards, T. C., Graham, C. H., Zimmermann, N. E., & Thuiller, W. (2010). Sampling in ecology and evolution - bridging the gap between theory and practice. *Ecography*, 33, 1028–1037. <https://doi.org/10.1111/j.1600-0587.2010.06421.x>
- Arroyo-Rodríguez, V., & Dias, P. A. D. (2010). Effects of habitat fragmentation and disturbance on howler monkeys: A review. *American Journal of Primatology*, 72, 1–16. <https://doi.org/10.1002/ajp.20753>
- Betts, M. G., Wolf, C., Pfeifer, M., Banks-Leite, C., Arroyo-Rodríguez, V., Ribeiro, D. B., Barlow, J., Eigenbrod, F., Faria, D., Fletcher, R. J., Hadley, A. S., Hawes, J. E., Holt, R. D., Klingbeil, B., Kormann, U., Lens, L., Levi, T., Medina-Rangel, G. F., Melles, S. L., ... Ewers, R. M. (2019). Extinction filters mediate the global effects of habitat fragmentation on animals. *Science*, 366, 1236–1239. <https://doi.org/10.1126/science.aax9387>
- Bowler, E. (2021). Data from: EllieBowler/optimising-sample-designs: First release for publication. *Zenodo*, <https://doi.org/10.5281/zenodo.5503985>
- Box, G. E. P., & Hunter, J. S. (1957). Multi-factor experimental designs for exploring response surfaces. *The Annals of Mathematical Statistics*, 28, 195–241. <https://doi.org/10.1214/aoms/1177707047>
- Bridges, L. M., Crompton, A. E., & Schaefer, J. A. (2007). Landscapes as gradients: The spatial structure of terrestrial ecosystem components in southern Ontario, Canada. *Ecological Complexity*, 4, 34–41. <https://doi.org/10.1016/j.ecocom.2006.05.001>
- Brudvig, L. A., Leroux, S. J., Albert, C. H., Bruna, E. M., Davies, K. F., Ewers, R. M., Levey, D. J., Pardini, R., & Resasco, J. (2017). Evaluating conceptual models of landscape change. *Ecography*, 40, 74–84. <https://doi.org/10.1111/ecog.02543>
- Chalfoun, A. D., Thompson, F. R., & Ratnaswamy, M. J. (2002). Nest predators and fragmentation: A review and meta-analysis. *Conservation Biology*, 16, 306–318. <https://doi.org/10.1046/j.1523-1739.2002.00308.x>
- Clifford, D., Payne, J. E., Pringle, M. J., Searle, R., & Butler, N. (2014). Pragmatic soil survey design using flexible Latin hypercube sampling. *Computers and Geosciences*, 67, 62–68. <https://doi.org/10.1016/j.cageo.2014.03.005>
- Didham, R. K., Kapos, V., & Ewers, R. M. (2012). Rethinking the conceptual foundations of habitat fragmentation research. *Oikos*, 121, 161–170. <https://doi.org/10.1111/j.1600-0706.2011.20273.x>
- Eigenbrod, F., Hecnar, S. J., & Fahrig, L. (2011). Sub-optimal study design has major impacts on landscape-scale inference. *Biological Conservation*, 144, 298–305. <https://doi.org/10.1016/j.biocon.2010.09.007>
- Ewers, R. M., & Didham, R. K. (2006). Confounding factors in the detection of species responses to habitat fragmentation. *Biological Reviews*, 81. <https://doi.org/10.1017/S1464793105006949>
- Ewers, R. M., Marsh, C. J., & Wearn, O. R. (2010). Making statistics biologically relevant in fragmented landscapes. *Trends in Ecology and Evolution*, 25, 699–704. <https://doi.org/10.1016/j.tree.2010.09.008>
- Ewers, R. M., Thorpe, S., & Didham, R. K. (2007). Synergistic interactions between edge and area effects in a heavily fragmented landscape. *Ecology*, 88, 96–106. [https://doi.org/10.1890/0012-9658\(2007\)88\[96:SIBEEA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2007)88[96:SIBEEA]2.0.CO;2)
- F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30, 609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>
- Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, & Systematics*, 34, 487–515. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132419>
- Fahrig, L. (2017). Ecological responses to habitat fragmentation per se. *Annual Review of Ecology, Evolution, and Systematics*, 48, 1–23. <https://doi.org/10.1146/annurev-ecolsys-110316-022612>
- Fardila, D., Kelly, L. T., Moore, J. L., & McCarthy, M. A. (2017). A systematic review reveals changes in where and how we have studied habitat loss and fragmentation over 20 years. *Biological Conservation*, 212, 130–138. <https://doi.org/10.1016/j.biocon.2017.04.031>
- Fletcher, R. J., Didham, R. K., Banks-Leite, C., Barlow, J., Ewers, R. M., Rosindell, J., Holt, R. D., Gonzalez, A., Pardini, R., Damschen, E. I., Melo, F. P. L., Ries, L., Prevedello, J. A., Tschamtkte, T., Laurance, W. F., Lovejoy, T., & Haddad, N. M. (2018). Is habitat fragmentation good for biodiversity? *Biological Conservation*, 226, 9–15. <https://doi.org/10.1016/j.biocon.2018.07.022>
- Gillespie, M. A. K., Baude, M., Biesmeijer, J., Boatman, N., Budge, G. E., Crowe, A., Memmott, J., Morton, R. D., Pietravalle, S., Potts, S. G., Senapathi, D., Smart, S. M., & Kunin, W. E. (2017). A method for the objective selection of landscape-scale study regions and sites at the national level (J. Travis, Ed.). *Methods in Ecology and Evolution*, 8, 1468–1476. <https://doi.org/10.1111/2041-210X.12779>
- Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services*.
- Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, 74(6), 1659–1673. <https://doi.org/10.2307/1939924>
- Manning, A. D., Lindenmayer, D. B., & Nix, H. A. (2004). Continua and Umwelt: Novel perspectives on viewing landscapes. *Oikos*, 104, 621–628. <https://doi.org/10.1111/j.0030-1299.2004.12813.x>
- McGarigal, K., & Cushman, S. A. (2002). Comparative evaluation of experimental approaches to the study of habitat fragmentation effect. *Ecological Applications*, 12, 335–345. [https://doi.org/10.1890/1051-0761\(2002\)012\[0335:CEOEAT\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2002)012[0335:CEOEAT]2.0.CO;2)
- Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, 32, 1378–1388. <https://doi.org/10.1016/j.cageo.2005.12.009>
- Minasny, B., McBratney, A. B., & Walvoort, D. J. J. (2007). The variance quadtree algorithm: Use for spatial sampling design. *Computers and Geosciences*, 33, 383–392. <https://doi.org/10.1016/j.cageo.2006.08.009>
- Mortelliti, A., Amori, G., Capizzi, D., Rondinini, C., & Boitani, L. (2010). Experimental design and taxonomic scope of fragmentation studies on European mammals: Current status and future priorities. *Mammal Review*, 40, 125–154. <https://doi.org/10.1111/j.1365-2907.2009.00157.x>
- O'Sullivan, D., & Unwin, D. J. (2010). *OSU2: The pitfalls and potential of spatial data* (pp. 33–53). Geographic Information Analysis.
- Pasher, J., Mitchell, S. W., King, D. J., Fahrig, L., Smith, A. C., & Lindsay, K. E. (2013). Optimizing landscape selection for estimating relative

- effects of landscape variables on ecological responses. *Landscape Ecology*, 28, 371–383. <https://doi.org/10.1007/s10980-013-9852-6>
- Pfeifer, M., Boyle, M. J. W., Dunning, S., & Olivier, P. I. (2019). Forest floor temperature and greenness link significantly to canopy attributes in South Africa's fragmented coastal forests. *PeerJ*, 7. <https://doi.org/10.7717/peerj.6190>
- Pfeifer, M., Kor, L., Nilus, R., Turner, E., Cusack, J., Lysenko, I., Khoo, M., Chey, V. K., Chung, A. C., & Ewers, R. M. (2016). Mapping the structure of Borneo's tropical forests across a degradation gradient. *Remote Sensing of Environment*, 176, 84–97. <https://doi.org/10.1016/j.rse.2016.01.014>
- Pfeifer, M., Lefebvre, V., Gardner, T. A., Arroyo-Rodriguez, V., Baeten, L., Banks-Leite, C., Barlow, J., Betts, M. G., Brunet, J., Cerezo, A., Cisneros, L. M., Collard, S., D'Cruze, N., da Silva Motta, C., Duguay, S., Eggermont, H., Eigenbrod, F., Hadley, A. S., Hanson, T. R., ... Ewers, R. M. (2014). BIOFRAG—a new database for analyzing BIOdiversity responses to forest FRAGmentation. *Ecology and Evolution*, 4, 1524–1537. <https://doi.org/10.1002/ece3.1036>
- Read, Q. D., Zarnetske, P. L., Record, S., Dahlin, K. M., Costanza, J. K., Finley, A. O., Gaddis, K. D., Grady, J. M., Hobi, M. L., Latimer, A. M., Malone, S. L., Ollinger, S. V., Pau, S., & Wilson, A. M. (2020). Beyond counts and averages: Relating geodiversity to dimensions of biodiversity. *Global Ecology and Biogeography*, 29, 696–710. <https://doi.org/10.1111/geb.13061>
- Riitters, K. H., O'Neill, R. V., Hunsaker, C. T., Wickham, J. D., Yankee, D. H., Timmins, S. P., Jones, K. B., & Jackson, B. L. (1995). A factor analysis of landscape pattern and structure metrics. *Landscape Ecology*, 10, 23–39. <https://doi.org/10.1007/BF00158551>
- Rosenstock, S. S., Anderson, D. R., Giesen, K. M., Leukering, T., & Carter, M. F. (2002). Landbird counting techniques: Current practices and an alternative (F. Thompson III, Ed.). *The Auk*, 119, 46–53. <https://doi.org/10.1093/auk/119.1.46>
- Ruffell, J., Banks-Leite, C., & Didham, R. K. (2016). Accounting for the causal basis of collinearity when measuring the effects of habitat loss versus habitat fragmentation. *Oikos*, 125, 117–125. <https://doi.org/10.1111/oik.01948>
- Schreuder, H. T., Hansen, M., & Kohl, M. (1999). Relative costs and benefits of a continuous and a periodic forest inventory in Minnesota. *Environmental Monitoring and Assessment*, 59, 135–144. <https://doi.org/10.1023/a:1006137914405>
- Van Groenigen, J. W. (2000). The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma*, 97, 223–236. [https://doi.org/10.1016/S0016-7061\(00\)00040-9](https://doi.org/10.1016/S0016-7061(00)00040-9)
- Van Groenigen, J. W., Siderius, W., & Stein, A. (1999). Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, 87, 239–259. [https://doi.org/10.1016/S0016-7061\(98\)00056-1](https://doi.org/10.1016/S0016-7061(98)00056-1)
- van Groenigen, J. W., & Stein, A. (1998). Constrained Optimization of Spatial Sampling using Continuous Simulated Annealing. *Journal of Environmental Quality*, 27, 1078–1086. <https://doi.org/10.2134/jeq1998.00472425002700050013x>

How to cite this article: Bowler, E., Lefebvre, V., Pfeifer, M., & Ewers, R. M. (2022). Optimising sampling designs for habitat fragmentation studies. *Methods in Ecology and Evolution*, 13, 217–229. <https://doi.org/10.1111/2041-210X.13731>