Spatial dynamics of *Pygoscelis* penguin coloniality

A Dissertation Presented

by

Philip McDowall

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Ecology and Evolution

Stony Brook University

May 2018

Stony Brook University

The Graduate School

Philip McDowall

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Heather Lynch – Dissertation Advisor
Associate Professor, Department of Ecology and Evolution, Stony Brook University

Catherine Graham - Chairperson of Defense
Professor, Department of Ecology and Evolution, Stony Brook University

Christian Luhmann
Associate Professor, Department of Psychology, Stony Brook University

Tom Hart
Research Fellow, Department of Zoology, Oxford University

This dissertation is accepted by the Graduate School

Charles Taber

Dean of the Graduate School

Abstract of the Dissertation

Spatial dynamics of *Pygoscelis* penguin coloniality

by

Philip McDowall

Doctor of Philosophy

in

Ecology and Evolution

Stony Brook University

2018

Organisms typically inhabit a two-dimensional manifold embedded in three-dimensional space, but the spatial methods that we use to study them are often applied to projections in lower dimensional spaces. Additionally, the resolution of datasets used to understand the space-use preferences and spatial dynamics of organisms is often dictated by data availability rather than suitability to the primary research question. I provide a novel method of collecting very high spatial-resolution data, which captures the three-dimensional (3D) structure of terrain and the organisms that inhabit it simultaneously. I explore the implications of projections to lower dimensional spaces on bias in statistical techniques, demonstrating that failure to account for 3D structure can lead to incorrect type I error rates in null hypothesis testing. Using these high-resolution 3D data collection tools, I study the spatial dynamics of nesting *Pygoscelis* penguins, modeling the effects of both terrain and conspecific interactions on choice of nest location. Using an individual-based model, I demonstrate that nest site fidelity in a stochastically changing population can lead to fragmentation of a colony, which feed back into population level changes through high predation rates at the edges of colonies. Finally, I demonstrate a novel method for collecting spectral data from an Unmanned Aerial System (or 'drone'), which provides a better understanding of how 3D structure relates to function in ecological systems.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my sincere gratitude to my advisor, Heather Lynch, for her continuous support of my research, and for her outstanding mentorship throughout my time at Stony Brook University. I gratefully acknowledge assistance from the US National Science Foundation Office of Polar Programs (Award No. NSF/OPP-1255058), the Institute of Advanced Computational Science at Stony Brook University, and Oceanites, Inc. Funding for the Danger Islands Expedition, during which much of the data presented here was collected, was provided by the Dalio Foundation.

Introduction

Ecological systems exist in a three-dimensional physical space, and evolve over the fourth dimension representing time. Organisms that fly or swim live in a three-dimensional world, but most organisms inhabit a two-dimensional manifold embedded in three-dimensional space. In this research, which is largely focused on the ecology of the *Pygoscelis* penguins, I study the spatial dynamics of nesting penguins and examine the role of conspecific attraction and fine scale habitat preferences. To address these complex topics, I develop novel methods for spatial ecology based on recent technological advances in computer vision, satellite imagery, and unmanned aerial systems. These methods allow us to collect spatial data in three dimensions at very high resolutions. By applying these techniques to nesting *Pygoscelis* penguins, I explore how we can link spatial pattern to ecological and behavioral processes to better understand how spatial dynamics influence penguin population dynamics. Further, I explore two scenarios in which the projection of a high-dimensional system to lower dimensions can be misleading. I show that the naive projection of three-dimensional space (x,y,z) to two dimensions (x,y) by simply dropping the z-coordinate can bias our estimates of spatial point patterning and, consequently, our understanding of animal interactions and habitat preferences. I also show that the projection from four-dimensional spatiotemporal state (x,y,z,t) to a one-dimensional (t) time series of abundance may mask important ecological processes and reduce our ability to forecast abundance into the future.

The Pygoscelis penguins

For nine months of the year, individual *Pygoscelis* spp. penguins are widely dispersed, foraging across large distances at the edges of the Antarctic sea ice. However, during the short austral spring and summer breeding season, these birds congregate in densely packed colonies that may range in size from tens of birds up to hundreds of thousands (Ainley 2002). While at the colony, pairs build and maintain stone nests in which two eggs are laid. Parents share the task of incubation and chick rearing, before eventually leaving the chicks to crèche. Individual birds display a high level of site-fidelity often returning to the same nest location year after year, and rarely moving between colonies.

Pygoscelis colonies represent a self-organizing aggregation of independently acting individuals. Such aggregations are overwhelmingly common in nesting marine birds, with over 95% of species adopting this breeding strategy (Wittenberger 1985). Despite the prevalence of this breeding strategy, its role, evolution, and implications are not fully understood. Individuals are likely to experience density-related costs to fitness through resource sharing, both in terms of nest site availability and resource availability within the limited foraging range that is utilized during breeding (Ainley et al. 1995). Additionally, nesting in close proximity to conspecifics is likely to cause higher rates of parasite transmission (Tella 2002, Brown and Brown 2004) and may result in increased presence of predators who are attracted to these large aggregations (Emslie et al. 1995), all of which carry additional costs for the fitness of an individual. Despite these probable costs, this breeding strategy is almost ubiquitous in marine birds and, as such, it is likely that the benefits outweigh the costs.

The most obvious reason for the coloniality of *Pygoscelis* penguins lies in the predator defense that is achieved by being in a group. Group defense against predation has often been

cited as a reason for colonial existence, with increased predation observed on the peripheries of colonies of many Antarctic bird species (Weidinger 1998). The most prolific predators of the *Pygoscelis* penguins are the South-Polar Skua (*Stercorarius maccormicki*), and Brown Skua (*Stercorarius antarcticus*) (Young 2005). While a lone penguin may be unable to defend its nest from a pair of Skuas, a group of penguins acting together can collectively dissuade the Skua from attacking. There are differing opinions on the strength of edge effects in *Pygoscelis* penguins, although there appears to be a consensus that predation rates are higher, and therefore reproductive success lower, on the peripheries of a colony (Barbosa et al. 1997). Attacks from Skuas may come either from the ground or from the air but are, in either case, more likely to be successful when attacking nests on the edges of a colony (Young 2005). Observational studies have reported a risk of predation for birds nesting on the margins of a colony that is 4-8 times higher than for those birds nesting at the interior (Emslie et al. 1995). Interior nesting sites may also be preferable to those on the margins of the colony due to different rates of disturbance (Tenaza 1971).

While collaborative defense gives a rationale for the existence of colonial breeding, it also raises some interesting questions. For one, it implies that the costs and benefits of coloniality are not shared equally by all individuals within a colony and are in fact likely to be highly spatially structured, such that those individuals nesting in interior nests have much reduced costs compared to those on the exterior. This suggests that the correct unit for consideration in density-dependent effects is the sub-colony (smaller groups of contiguously nesting birds) rather than the colony as a whole, which is the spatial unit at which population dynamics have usually been considered. Secondly, the fitness of an individual cannot be based upon the decisions that they make alone and must be a result of the decisions made by all other

individuals within an aggregation. We frequently assume that individuals in populations adopt behavioral strategies that maximize their fitness. However, when there are interactions between organisms, the optimal strategy for one individual is often a function of the strategy selected by the other individuals, whose strategy is, in turn, a reaction to the strategies adopted by all other individuals

Coloniality

Coloniality is a challenging area to study as the processes involved in coloniality operate at a wide range of spatial scales. At the largest scales, resource availability is likely to drive the location of colonies. However, there are also processes that occur within the colony, often down to the sub-meter scale. While low resolution, high spatial coverage data has been available for the consideration of the largest spatial scale processes, and has demonstrated that factors such as sea surface temperatures are important in dictating the locations of colonies (Cimino et al. 2013), the data required to address within-colony processes have been lacking.

Traditionally, the census of nesting *Pygocelis* populations has involved the manual counting of individual nests in the field, resulting in a one-dimensional time series of abundance for each accessible colony. Due to the logistical difficulties of working in the Antarctic, these datasets are often patchy, with some sites having many years of missing data between completed surveys. This manual counting methodology has been supplemented using remotely-sensed data, which allows us to overcome some of the logistical difficulties of working in the field, but until recently has required us to accept less precise estimates of abundance (LaRue et al. 2014). Advances in remote sensing technology and changes in legislation have led to the availability of

imagery at sub-meter resolutions, allowing us to refine data on both the locations of nesting sites, our estimates of colony size, and improve the spatial covariates that are available to build models of spatial usage. This high-resolution imagery has demonstrated that *Pygoscelis* colonies are highly spatially structured, with sub-colonies forming patterns of spots and stripes and labyrinths across the terrain (See Fig. 3-1, Chapter 3). Given the spatial structuring of the benefits of coloniality, we might expect that these colonies should take on more regular shapes that minimize the proportion of nests that occur on the peripheries and maximize the number of birds with relatively safe interior nest locations. To understand the processes driving the observed spatial pattern, and the mechanisms by which these patterns feed back into the biological processes controlling abundance, we need to understand the dynamics of the colony at the level of the individual. Unfortunately, the improved resolution available with the current generation of satellites is still at a courser scale than the world as perceived by an individual bird and so does not provide the individual-level information required to answer questions about the interactions occurring between individuals within an aggregation.

Thesis summary

In Chapter 1 of this thesis, I present a novel application of a computer vision technique, Structure-from-Motion (SfM), which allows us to collect centimeter-scale data on the topography of a landscape and the individuals within that landscape. This method allows the collection of occupancy and topography datasets simultaneously, ensuring no temporal or spatial mismatch between the topographic dataset and the locations of the individuals. SfM algorithms are based on triangulating matched points between sets of overlapping images, allowing the estimation of three-dimensional structure from two-dimensional image sequences. This method differs from stereo-matching algorithms traditionally used to create DEMs, and from more

common photogrammetric methods, in that the camera position and scene geometry are solved simultaneously, negating the need for data on camera position, or three-dimensional control points. This is achieved using algorithms that track multiple correspondences between multiple overlapping, offset images. While the three-dimensional model that is created is scaleless and arbitrarily orientated the addition of supplementary information such as known positions within the scene or the locations of the cameras in a real-world coordinate system, allows us to reconstruct a scaled, georectified three-dimensional model of the scene.

When SfM is combined with the capabilities of modern Unmanned Aerial Systems (UAS), we have a tool to rapidly census large populations of nesting birds and to collect spatial data that is simultaneously very high resolution and able to cover an entire colony that may span hundreds of hectares. However, with the increased resolution of these datasets available through UAS surveys, we may encounter statistical issues which, at the courser scales we have been used to, have not previously been considered.

In Chapter 2 of this thesis, I focus on the statistical biases introduced when projecting location data using the common but incorrect projection of a three-dimensional (3D) dataset to the two-dimensional (2D) plane. There are a range of tools that we might use to model the presence/absence, abundance, or density of a species, but many of these models represent variations on the Poisson point process model (Renner and Warton 2013). Tools for applying point process models (and their derivatives) to one- and two-dimensional datasets are well developed, and there are numerous ways to estimate the parameters of such a model (Cressie 1993, Lieshout and Baddeley 1996, Baddeley and Turner 2005). However, when we deal with ecological datasets in lower dimensional space, we have assumed that the projection from the 3D world to a 2D coordinate system is a valid one and does not introduce bias. This step of re-

projection to a lower dimensional space is often made implicitly, without the realization that it is occurring, or that it may introduce significant bias in subsequent analysis. We tend to collect data on the location of species in some two-dimensional coordinate system, either as the latitude or longitude at which the observation was made, or in a coordinate system defined specifically for a study in which individual records are the position, in x-y coordinates, of the observation. The act of recording these observations as a pair of coordinates is an orthographic projection of the real locations in three-dimensional space into a two-dimensional space. In Chapter 2, I demonstrate that when point processes occur on a two-dimensional manifold embedded in three-dimensional space, simple geometric considerations imply that the orthographic projection does not maintain the key parts of the spatial pattern that are used in point process modelling, resulting in biased estimates that are most significant in landscapes in which there is large variance in the slope of terrain.

In the subsequent chapters of this thesis, I study the mechanisms that drive pattern formation in penguin colonies, using a dataset created using SfM applied to high-resolution UAV imagery captured at Beagle Island, in the Danger Islands chain at the northern extent of the Western Antarctic Peninsula. Due to the spatial structuring of predation risks within a colony, we may expect colonies to be arranged in such a way as to minimize the number of individuals on the peripheries. In the absence of other factors, we expect colonies should be contiguous and roughly circular. However, real colonies are highly spatially structured and often consist of many small "sub-colony" units. Complicating the spatial dynamics of penguin colonies is the underlying terrain on which these colonies are built, which controls habitat suitability independent of conspecifics and would, *a priori*, be expected to influence the distribution of penguin nests within the colony. The locations of individual nests reflect, therefore, a

convolution of preference for particular terrain and a desire to nest in proximity to conspecifics. I use a Bayesian auto-logistic model (Besag 1972) to estimate the strength of each of these components of nest site selection, allowing us to separate abiotic and biotic processes in nest site selection. Given this information I can then consider whether the patterns we see in the field might represent an optimal arrangement in a spatially structured landscape. In Chapter 3, I apply a meta-heuristic approach, simulated annealing (Van Laarhoven and Aarts 1987) to artificially optimize the locations of penguins on this landscape to demonstrate that the current arrangement observed at Beagle Island is in fact sub-optimal. This technique demonstrates an important point for the spatial dynamics of penguin nest locations within a colony; a high level of nest site fidelity in *Pygoscelis* penguins means that the arrangements of individuals in each year is highly dependent on the previous year's arrangement. These rearrangements must occur at the individual level, with no coordination between individuals, and as such there is little scope for the population to reorganize to an optimal arrangement. As I discuss in Chapter 3, there may even be 'stable' configurations of nests in which no individual acting alone can improve their position, forming an analog of the Nash Equilibrium from game theory (Nash 1951). The fact that the sub-optimal condition of the colony in revealed in a static snapshot of time is in a sub-optimal condition raises the questions of how the colonies end up in this state, and what the implications of this sub-optimality and iterative rearrangement of the configuration are for the population.

In Chapter 4, I attempt to address these questions by constructing a dynamic individual-based model to simulate how penguin colonies grow and decline over time, how spatial pattern might emerge and what the implications of this spatial patterning might be for the population. Individual-based models are well suited to the task of modelling local interactions between

individuals and exploring properties that might emerge at the system level as a consequence of these interactions (DeAngelis and Grimm 2014). In this model, I provide each individual with a simple set of rules about when to move, where to move to, and the probability of breeding success in a given location. From these simulations we see that the spatial patterns observed in *Pygoscelis* colonies are most likely formed during periods of population decline. As there is a lag in an individual's ability to respond to a changing landscape of conspecifics, these declines tend to form holes (areas not occupied by nesting penguins) in otherwise contiguous colonies that can fragment colonies into smaller disconnected sub-colony units. These declines increase the number of individuals nesting on the periphery, which in turn reduces the average reproductive success of the colony. By measuring the responses of simulated colonies in different configurations to perturbation, we observe that the rate at which populations recover to their original size is a function of the level of aggregation of the original population, with more fragmented colonies recovering more slowly. This slowed recovery is often regarded as signal of an impending 'critical collapse' (Dakos et al. 2011, Scheffer et al. 2012), in which a population suddenly declines to extinction. The ability to create such dynamics in this four-dimensional system represent yet another weakness of projection to a lower dimensional space. When colonies are considered as a one-dimensional system in which change in the population at time t+1 is a function of only the population at time t, we are discarding the additional dimensions of spatial configuration. As a result, we may not anticipate such critical transitions. Importantly, the assumption that population trends act as a Markov Chain, with simple probabilistic transitions, is violated. The spatial configuration and state of a colony is very much dependent upon the history of that colony and its previous configurations. I find that the spatial properties of a colony are

highly correlated with its population trajectory over a period of around a decade, and that high levels of fragmentation result from periods of decline.

In the final chapter of this thesis, I explore how ecologists can leverage the increasing capabilities, and rapidly decreasing cost, of unmanned aerial systems (UASs). These systems are now able to carry a variety of sensing instruments such as high-resolution cameras and spectrometers to collect additional data which, when integrated into full three-dimensional scenes, allows us to greatly increase the scale and resolution of the questions that we can answer. I demonstrate this through the application of a novel method for combining 3D scene reconstruction and readings collected using a high-resolution non-imaging spectrometer that allow us to explore how structure and function are interconnected.

Chapter 1 - Ultra-Fine Scale Spatially-Integrated Mapping of Habitat and Occupancy Using Structure-From-Motion[1]

Abstract

Organisms respond to and often simultaneously modify their environment. While these interactions are apparent at the landscape extent, the driving mechanisms often occur at very fine spatial scales. Structure-from-Motion (SfM), a computer vision technique, allows the simultaneous mapping of organisms and fine scale habitat, and will greatly improve our understanding of habitat suitability, ecophysiology, and the bi-directional relationship between geomorphology and habitat use. SfM can be used to create high-resolution (centimeter-scale) three-dimensional (3D) habitat models at low cost. These models can capture the abiotic conditions formed by terrain and simultaneously record the position of individual organisms within that terrain. While coloniality is common in seabird species, we have a poor understanding of the extent to which dense breeding aggregations are driven by fine-scale active aggregation or limited suitable habitat. We demonstrate the use of SfM for fine-scale habitat suitability by reconstructing the locations of nests in a gentoo penguin colony and fitting models that explicitly account for conspecific attraction. The resulting digital elevation models (DEMs) are used as covariates in an inhomogeneous hybrid point process model. We find that gentoo

[1] McDowall, Philip, and Heather J. Lynch. "Ultra-Fine Scale Spatially-Integrated Mapping of Habitat and Occupancy Using Structure-From-Motion." PLoS ONE 12.1 (2017): e0166773.

Author Contributions: (PM – Philip McDowall, HJL – Heather Lynch) Conceptualization: PM. Data curation: PM. Formal analysis: PM. Methodology: PM HJL. Project administration: HJL. Supervision: HJL. Validation: PM. Visualization: PM HJL. Writing – original draft: PM HJL. Writing – review & editing: PM HJL.
Thanks to Tom Hart and penguinlifelines.org for the images used in this analysis, and Ben Weinstein, Catherine Graham, Cecilia O'Leary and Casey Youngflesh for discussions regarding this manuscript.
Funding provided by the National Science Foundation Office of Polar Programs (NSF/OPP-1255058).

penguin nest site selection is a function of the topography of the landscape, but that nests are far more aggregated than would be expected based on terrain alone, suggesting a strong role of behavioral aggregation in driving coloniality in this species. This integrated mapping of organisms and fine scale habitat will greatly improve our understanding of fine-scale habitat suitability, ecophysiology, and the complex bi-directional relationship between geomorphology and habitat use.

Introduction

Habitat suitability models for plants and animals often focus on course-grained abiotic habitat characteristics at the expense of microhabitat factors and/or biotic interactions that can also be important for structuring the use of space (Guisan and Thuiller 2005, Campomizzi et al. 2008). Despite their ubiquity and importance for spatial ecology, the scale and extent of data used to explore relationships between organisms and the space they occupy are often dictated by the availability of environmental data rather than the ecology and physiology of the organism under consideration (Jones 2001). Moreover, since data on environmental conditions and the presence/absence of the organism are usually recorded independently, there can be considerable spatial and temporal alignment errors between data types, making it difficult to infer the true relationship between them (Guisan and Thuiller 2005, Barry and Elith 2006). Issues of scale or spatiotemporal registration errors are relatively minor when key environmental covariates vary slowly (e.g., elevation) or when the spatial scale of occupancy is large (e.g., an island), but they can be highly problematic when modeling habitat suitability or space use at much smaller spatial

scales, where occupancy may hinge on the detailed hydrology of the site or subtle variations in exposure to wind scour or solar irradiation (Bennie et al. 2008).

Microclimate is likely to be critical in driving occupancy and abundance of a range of plants and animals because the environment as experienced by an individual can be very different from the average background condition measured at much larger spatial scales (Suggitt et al. 2011). When high-resolution topographic maps are combined with climatic modeling, we may begin to detect fine scale variations that drive habitat associations. Seabra et al. (Seabra et al. 2011) demonstrated that limpets at equal tidal heights and separated by less than two meters may experience significantly different incident solar radiation depending on which side of a rock they inhabit. The micro-scale differences in resulting temperature, which can be critical for the organisms as they perceive their environment, are far smaller than could be mapped using traditional air or water temperature datasets derived from satellite-based sensors. Terrestrial ecology also provides many such examples; different species of Anoline lizard occupy distinct microhabitats based on shade availability (Rand 1964), the distribution of saxicolous lichens in the Rocky Mountains appear to be driven by fine scale variability in snow cover across the faces of boulders (John 1990), and thermally constrained butterflies select topographically-driven microhabitat based on fine scale variations in temperature (Lawson et al. 2014). Topography can shape microclimate and provide fine-scale habitat that falls within the physiological tolerances of a species in an otherwise apparently unsuitable landscape. While these microhabitat drivers may not be as valuable as regional scale variables in predicting occupancy over large geographic ranges, they may be key to understanding small-scale interactions that structure the use of space by animals, and may improve model performance when used in conjunction with these regional variables (Mcclure et al. 2012, Tattoni et al. 2012).

Biotic interactions can also influence how animals and plants use space, making it difficult to infer the strength of abiotic habitat associations. Positive associations between conspecifics, either caused by active behavioral aggregation such as colonial breeding or the more passive dispersal limitations often seen in plant distributions, should lead to an increased density of individuals. Alternatively, negative associations, such as competition for resources, territorial behavior, or allelopathy should result in lower densities of individuals that are more regularly spaced across the landscape. When these biotic interactions are excluded from habitat suitability models, we risk erroneously assigning this variance to some landscape factor, and may find that our models perform poorly at predicting the spatial distribution of a species (Campomizzi et al. 2008). Similarly, if we do not explicitly include abiotic landscape heterogeneity into models of aggregation then we may incorrectly attribute patterns of aggregation to complex biotic interactions.

The use of topography as an explanatory variable in distribution modeling has a long history in the field of gradient analysis, whereby the abundance of a plant species is related to environmental gradients such as elevation. While most gradient analysis studies focus on the landscape scale, important fine scale details can be lost when data are collected at such large scales (Lookingbill and Urban 2005) as small scale heterogeneity in the landscape has been shown to promote species richness and beta diversity in plant communities (Opedal et al. 2015). In addition to information lost due to inappropriate scale, ecological datasets almost always project a three-dimensional landscape onto a two-dimensional raster, preventing a complete consideration of covariates that require a three-dimensional understanding of habitat (e.g., terrain, canopy structure, etc.) (Opedal et al. 2015). The additional information provided by digital elevation models (DEMs) can often be used as proxies for environmental condition and

permit direct modeling of many processes, such as hydrological flow, critical to a site's suitability. To fully understand the spatial ecology of a species, it is important to map both organisms and the fine scale three-dimensional details of their landscape simultaneously. These maps, which we refer to as 'integrated' terrain and occupancy maps, are likely to be useful across a range of applications including ecology, ecophysiology, and biogeomorphology.

Mapping fine-scale terrain

One of the basic structuring elements of habitat is elevation, however the resolution of available elevation data is highly variable across the globe. Within the United States, the USGS National Elevation Dataset provides low-resolution elevation data with relatively high spatial coverage, and a smaller number of high-resolution data products derived from aerial photogrammetry (with horizontal resolution up to 3 m). While similar data sets are available from several national agencies and commercial providers, global coverage is limited and high-quality data are often hard to find and/or may be prohibitively expensive. Many of the more readily available products are derived from sources such as RadarSat and yield elevation datasets with horizontal resolutions in the range of hundreds or even thousands of meters. At these scales it is unlikely that recorded environmental variables accurately reflect the environment as experienced by an organism, as variation in microhabitat can lead to very different conditions existing at the smallest scales. Stereo pairs of imagery from commercial satellite images can be used to construct digital elevation models with spatial resolutions on the order of meters, but imagery at this resolution can be prohibitively expensive and cloud cover and shadows can create holes in the imagery that must be imputed.

LiDAR technology is capable of recording high-resolution 3D landscape structure (Vierling et al. 2008) and it has been shown that such high resolution information on vertical structure can improve habitat suitability modeling studies (Tattoni et al. 2012, Garabedian et al. 2014, Questad et al. 2014), but LiDAR surveys entail equipment costs and often flight times that can be economically or logistically prohibitive. Structure-from-Motion (SfM), a computer vision technique, can be used to rapidly and economically produce detailed 3D information on the structure of the landscape (Fonstad et al. 2013) and simultaneously record the location of organisms within that landscape. SfM is simple enough to deploy in any landscape (Lavy et al. 2015), and provides data on fine scale habitat characteristics that are otherwise unavailable. While this survey method has been used to create digital models of artifacts in archeology (Green et al. 2014), measure canopy cover (Dandois and Ellis 2010, 2013, Zahawi et al. 2015) and record coral morphology (Lavy et al. 2015) in ecology, this work represents the first time the lower cost methodology of SfM has been used for modeling habitat suitability.

Modeling conspecific interactions

While integrated terrain and occupancy models would find utility across a number of fields in ecology, we demonstrate its use by applying it to the study of conspecific attraction in colonial seabirds. Occupancy can be modeled either as a binomial process on a two-dimensional grid or as an inhomogeneous spatial point process; both approaches are made feasible using SfM, however we use a point process approach to demonstrate the use of SfM in a case study of nest site selection in the gentoo penguin (*Pygoscelis papua*). As in many colonially nesting birds, it is difficult to determine to what extent gentoo penguin nests are clumped due to active behavioral aggregation or whether they are simply responding independently to patchiness in suitable

nesting habitat. High resolution terrain models integrated with information on patch occupancy allow us to quantitatively estimate the strength of these two competing hypotheses and provide an ecologically important case study for the use of this integrated mapping technique.

Methods

Structure-from-Motion

Structure-from-Motion, a computer vision technique in which 3D structure is estimated from a set of overlapping images of the landscape, was used to produce a high-resolution 3D model of the study site with embedded information on nest locations (Fig. 1-1). No prior information on the position from which images are captured is required, as the SfM algorithm is able to estimate the position of the cameras independent of the unknown 3D scene and, consequently, the structure of the 3D terrain.

We use a commercial product, Photoscan Professional Edition (Photoscan 2016), for the photogrammetric workflow. Educational licensing is available for this product at a reduced rate, and in 2016 was priced at $59 for the standard edition and $549 for the professional edition. Alternative freely available software include visualsfm (Wu et al. 2011, Wu 2013), tools in the python module OpenCV, and a selection of free-to-use web based services such as AutoDesk Catch 123D. All of these options use similar processes and workflows, but have different strengths and weaknesses (Kersten and Lindstaedt 2012). Having tested many of these options, we think Photoscan provides the most complete and user-friendly set of features for working with spatial data and producing georeferenced output from 3D reconstruction.

Feature matching algorithms such as Scale Invariant Feature Transformations (Wu 2007) are used to automatically identify corresponding points observed in multiple images and solve for the position of pairs of camera locations independently of the unknown structure of the 3D scene. Given the positions and orientation of the cameras from which the images were taken the software can project each matched point back into 3D space to yield a vector along which this point must lie. For each matched point we expect the intersection of the projections from multiple images to converge, giving the location of that point in 3D space. The estimated camera positions are then passed to a multi-view stereo algorithm to create a dense point cloud representing the 3D surface (Fig. 1-2a). This dense point cloud is used as the basis for the generation of a polygon mesh, a 3D surface consisting of polygons that interpolates between the points in the dense point cloud (Fig. 1-2). Though the mesh is in an arbitrary coordinate system, this mesh can be transformed into a real-world coordinate system through the use of additional information collected during the survey and an appropriate georectification method. When using cameras with a built-in GPS system, the metadata attached to each image can be used to estimate the position of the cameras. These estimated locations can then be used to transform the mesh to a real-world coordinate system, while providing estimates of errors in camera locations. It should be noted, however, that these estimated errors convolve errors in the GPS positions and in the reconstructed mesh. Alternatively, information on the real-world position of points within the reconstructed scene can be used to georectify the model. Coded machine-readable targets, which can be produced automatically by Photoscan, can be placed into the scene prior to surveying and their positions recorded via a GPS (or differential GPS) unit. These targets can be automatically detected by the Photoscan software and encode an identifying number allowing unique markers to be identified. The model can then be georectified to these known points. These markers, when

visible in multiple overlapping images, can also aid in the alignment of images and estimation of camera positions. If coded targets are not available, any recognizable point within the scene, or identifiable non-coded markers, can be used, although this requires the user to manually identify the targets within images and set their real-world coordinates. Once a 3D mesh has been produced and georectified, it can be converted into a raster DEM for convenient use in GIS software.

If no positional information is available at the time of surveying it is possible to export the model in an arbitrary coordinate system to GIS software where it can be georectified to satellite imagery as long as corresponding features, such as the corners of buildings, can be identified in both the 3D model and the satellite imagery. If this is the case, as it was in this study, it is important that the data are rescaled in the z-axis as georectification outside the SfM software will leave the elevation values in the original unscaled coordinate system. We exported an orthorectified image to ArcGIS and georectified it to existing 0.5 m resolution panchromatic satellite images using a linear transformation.

The ideal approach to SfM is to produce images using an aerial survey either from aircraft or unmanned aerial systems (UASs). However, due to cost considerations, frequent high winds, and risk of negative interactions between wildlife and UASs, we use a GoPro Hero 3 mounted on a 2.7 m pole that is carried around the site, with the camera capturing images of the terrain every ten seconds. While these oblique images are not ideal for SfM, tests showed that the elevation provided by the pole system was sufficient to allow reconstruction of the terrain. In an ideal scenario, a systematic transect survey would be planned in advance to ensure complete coverage; however, the opportunistic nature of field site access in the Antarctic and the need to avoid disturbing potentially sensitive nesting birds made this impossible for our application. To

compensate for the lack of a systematic survey plan we collected many more images than necessary to ensure that we had complete coverage and sufficient views of all terrain. Despite the challenges of three-dimensional mapping across such complex terrain we managed to reconstruct an area of over 50,000 m^2 in with only around 40 minutes of survey effort.

Occupancy

After estimating the camera locations, we can project any given point in an image back into 3D space, resulting in a vector along which that point must lie. If we calculate the 3D coordinate of the intersection of this vector and the 3D mesh, we can place the point into the 3D landscape.

We selected a subset of the images that contained multiple projections of every occupied nest at the study site. The viewpoint afforded by the 2.7 m pole was sufficient to ensure that there was no occlusion of nests, and while we cannot always distinguish between a resting non-breeder (juvenile) and a nest with eggs being incubated by a breeding penguin, the use of expert interpretation of images minimizes false positives. In each of the images the pixel coordinates of each nest were used to project back to an intersection point on the 3D mesh. This step is possible directly in the Photoscan software package through marker placement on an image after generation of the mesh. This approach results in a set of points on the 3D surface corresponding to the estimated locations of all penguin nests in the site. Due to errors in alignment and the fact that different points within each nest are selected and projected onto the 3D mesh from different viewpoints, nests are occasionally recorded as a cluster of points. We filter these points assuming a minimum distance (10 cm) between the centers of adjacent nests, visually check the images to

ensure each cluster is in fact an artifact of using multiple projections for a single nest, and then reduce these clusters to a single point at the center of the cluster. We can also use the total number of nests counted on the ground during surveying as a check on the number of clusters identified. The latitude and longitude of each point is then exported to GIS software.

Point process modeling

Ripley's K-function (Haase 1995), K(r), a measure of the average number of points occurring within a radius 'r' of any other point, was used to provide a visual representation of pattern present. The value of Ripley's K can be assessed over a range of values of 'r' and compared against a theoretical value expected under complete spatial randomness to understand scale dependent patterns occurring within the point pattern.

Mapping individual penguin nest locations also allows us to model the location of nests as the outcome of a spatial point process. We hypothesize that the observed clumping of *Pygoscelis* penguin individuals (both at the scale of the colony and sub-colony units) is a convolution of preference for auto-correlated terrain (specifically, well-draining areas at the top of local peaks in elevation) and significant levels of conspecific attraction. We model nest locations as the outcome of a hybrid Gibbs point process (Baddeley and Turner 2005). This model allows for interactions among points even as the intensity of the point process varies according to the underlying abiotic landscape features mapped using SfM. The hybrid model has three components; hard-core repulsion that prevents points from occurring within a distance h of each other, a Strauss interaction in which points separated by a distance between h and a radius r_1 contribute a factor γ_1 ($\gamma_1 < 1$), to the probability density, resulting in a decreased probability

of inter-point distances being found within this range, and a Strauss interaction in which points separated by a distance between r_1 and a radius r_2 contribute a factor γ_2 ($\gamma_2 > 1$), resulting in an increased probability of points within this distance. The probability density for the hybrid process is:

$$f(\mathrm{Z}) = \alpha \left[\prod_{i=1}^{n} \beta(\mathrm{Z}_i) \right] \left[\prod_{i<j} \psi(\mathrm{Z}_i, \mathrm{Z}_j) \right]$$

where $\mathrm{Z} = (\mathrm{Z}_1, \ldots, \mathrm{Z}_n)$ is the set of n points in the observed point process, α is a normalizing constant, and $\psi(\mathrm{Z}_i, \mathrm{Z}_j)$ represents the pairwise interaction between points that depends on the distance $\|\mathrm{Z}_i - \mathrm{Z}_j\|$ between the points

$$\psi(\mathrm{Z}_i, \mathrm{Z}_j) = \begin{cases} 0, & \|\mathrm{Z}_i - \mathrm{Z}_j\| \le h \\ \gamma_1, h < \|\mathrm{Z}_i - \mathrm{Z}_j\| \le r_1 \\ \gamma_2, r_1 < \|\mathrm{Z}_i - \mathrm{Z}_j\| \le r_2 \\ 1, r_2 < \|\mathrm{Z}_i - \mathrm{Z}_j\| \end{cases}$$

with γ > 1 representing attraction. The function $\beta(\mathrm{Z}_i)$ is related to the first order intensity of the point pattern at the point locations Z_i,

$$\log(\beta(\mathrm{Z}_i)) = \mu + \sum_{k=1}^{l} \rho_k x_{k,i}$$

where μ is an intercept term, ρ_k are the coefficients for the set of l environmental covariates, and $x_{k,i}$ are the values of the k^{th} environmental covariate at the point location Z_i.

In the case of gentoo penguins, we hypothesize that the hard-core repulsion at small distances is driven by the physical size associated with each individual nest, while at slightly longer length scales territorial behavior increases average inter-nest distance, and at larger scales

defense against aerial predators creates positive attraction. The main drivers of nest site selection (i.e. the inhomogeneous intensity of the point process $\beta(Z_i)$) are likely to be associated with the hydrology of the site, as waterlogging of the nest may prevent important gas exchange over the egg shell (Portugal et al. 2014), and cause hypothermia in chicks that have not yet reached thermal independence (Fargallo et al. 2001). Covariates associated with the underlying suitability of terrain which were selected for inclusion in the statistical model of nesting include elevation, flow accumulation, and a travel cost metric that combines distance to the coast (where penguins haul out of the water after foraging) and slope along their commute back to the nest. While cost-weighted distance and elevation are generally inversely correlated, cost-weighted distance also accounts for those areas in which there is no direct path from the coast to the nest. Flow accumulation uses the aspect of each cell to determine the sum of cells likely to contribute to water-flow into any given cell. The calculation of flow accumulation is dependent on the scale used for analysis; for this reason, flow accumulation was calculated at a variety of scales and model selection used to determine which scale(s) should be retained in the best-fitting model.

Point process models were fitted using the 'ppm' function in the 'statspat' R package (Baddeley and Turner 2005). Appropriate interaction distances are estimated via a model comparison method in which models for all combinations of parameters h and r are fitted and the model with the lowest Akaike Information Criterion (AIC) score selected. While artificially introduced boundaries in a point pattern can affect model fitting, the point pattern's boundary in this case is created by the coastline of the island and is thus not an artifact of the sampling; correspondingly, no edge correction was used in fitting the models.

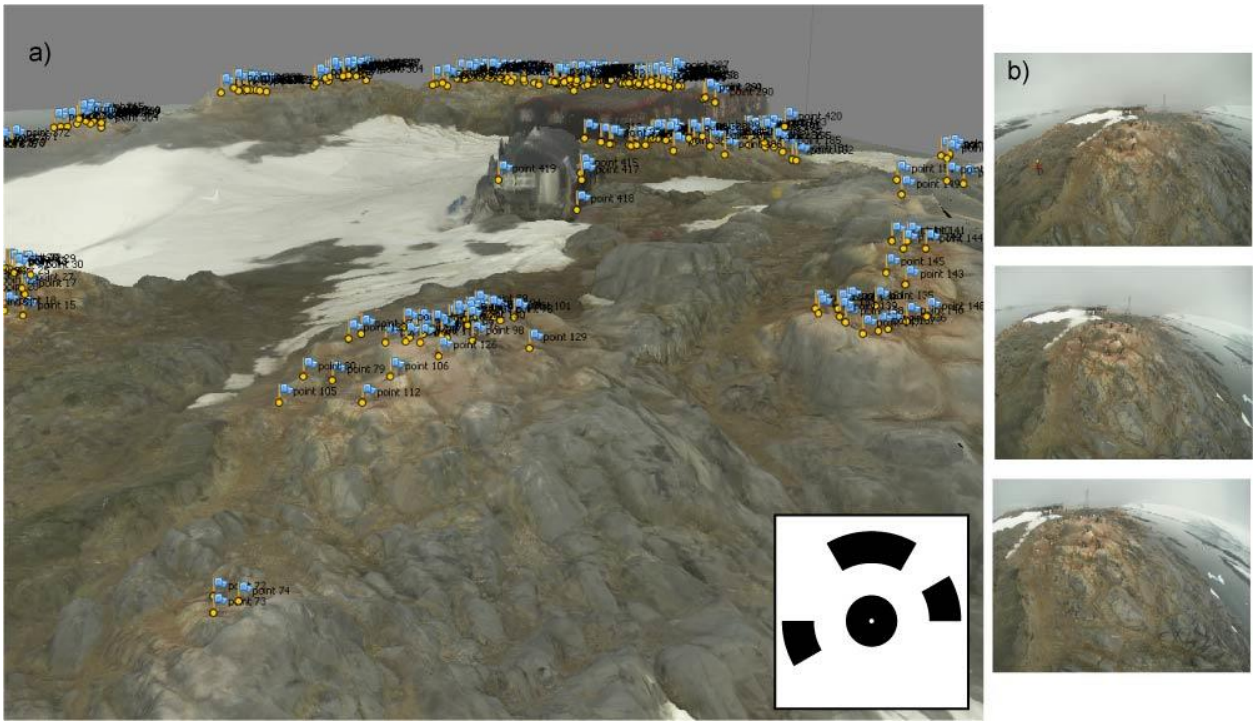

Figure 1-1. (a) Textured 3D mesh of Port Lockroy, Antarctica created from 493 images (b) Sample of images captured using a GoPro 3. Locations of occupied penguin nests are marked with yellow dots and blue flags.

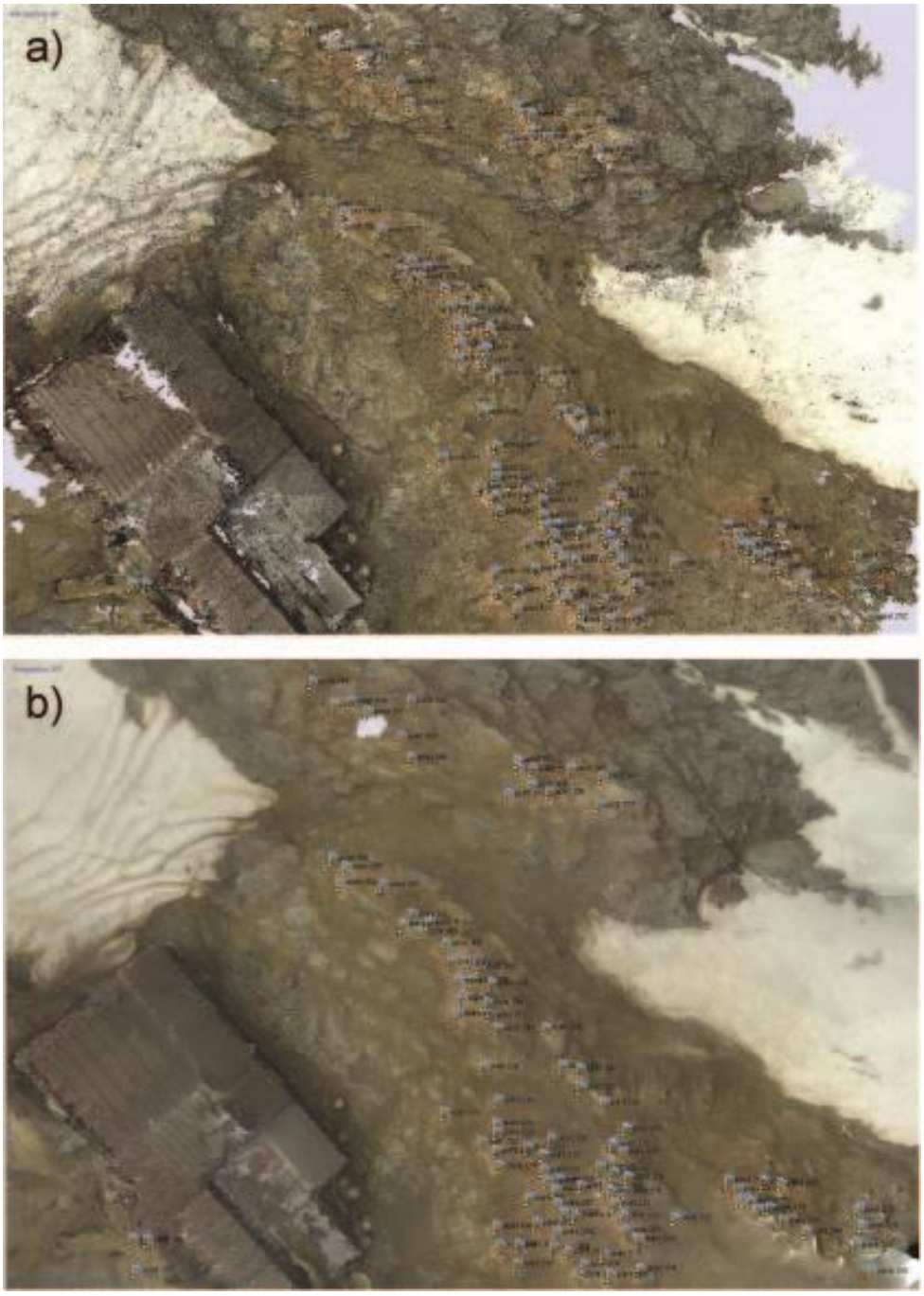

Figure 1-2. (a) Dense point cloud of Port Lockroy containing 113,338,579 points produced using 493 images processed with Photoscan (Agisoft). (b) Textured mesh containing 2,495,043 vertices fitted to dense point cloud.

Results

Of the 493 images available for Port Lockroy, the Antarctic penguin breeding site considered for this analysis, 459 were successfully aligned, providing at least 9 views of all areas of the island for generation of the 3D point cloud. The point cloud consisted of 113,338,579 points in 3D space (5,667 points per square meter over an area of 66,395.90 m2), which were generalised to a mesh containing 2,495,043 vertices and 4,988,477 faces. This mesh was georectified and converted to a raster with a resolution of ~6 mm.

Ripley's K indicated significant under-dispersion, or clustering of points, at distances greater than 0.3 m while at smaller distances the points were over-dispersed relative to a stationary Poisson point process (Fig 1-3). Model selection via AIC indicated that the estimated hard-core distance for the point process model (h) was 0.28 m while the Strauss interaction radii (r_1, r_2) were 0.5 m and 1.86 m, respectively.

Elevation, flow accumulation, and cost distance were found to be statistically significant ($p < 0.001$) for the occurrence of nests even after allowing for the interaction of points. Flow accumulation was found to be significant at multiple scales ($p < 0.001$). Nest densities were higher in areas at greater elevations and in those locations unlikely to become water–logged (Fig 4, Table 1-1). The interaction coefficients (Table 1-1) are the natural logarithm of the estimated interaction parameters γ_1 and γ_2. Our estimates of $\widehat{\gamma_1} = -1.13$ (95^{th} percentile CI = [-1.36,-0.88]) and of $\widehat{\gamma_2} = 0.56$ (95^{th} percentile CI = [0.54,0.58]) indicate strong negative interaction at short length scales ($\leq$ 0.5m), and positive interactions at longer scales (0.5 m - 1.86 m), leading to a higher density of nesting than would be expected based on first-order inhomogeneity in habitat suitability.

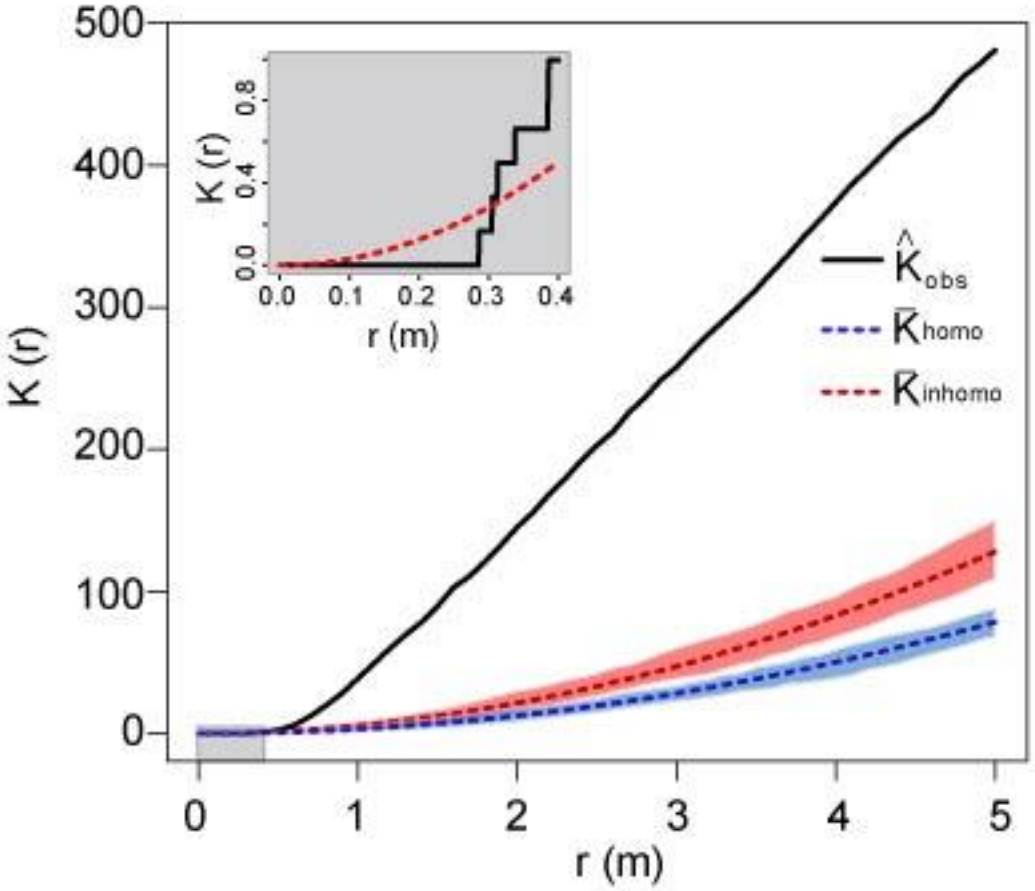

Figure 1-3. Ripley's K-function (mean number of points within radius r from any point) for the observed point pattern (black), theoretical value under complete spatial randomness (blue), and value under an inhomogeneous Poisson process with no inter-point interaction (red). Confidence intervals generated through 1,000 simulations of point processes. Gentoo nests show over-dispersion (fewer points than expected) at short scales (inset) and under-dispersion (more points than expected) at larger scales.

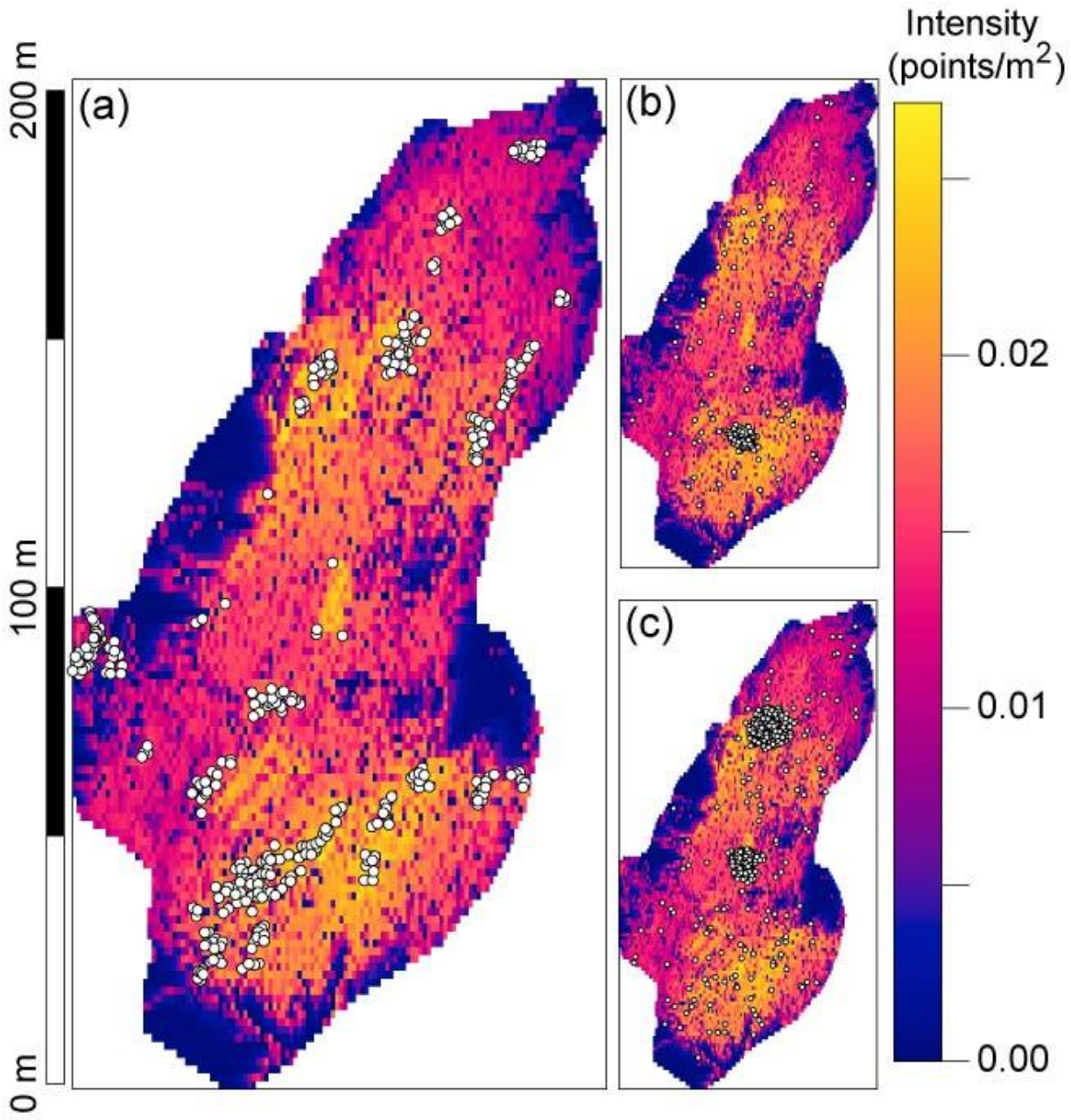

Figure 1-4. (a) Estimated intensity of inhomogeneous point process of nest locations driven by underlying terrain characteristics. White circles represent the locations of nests in the observed point pattern. (b) A stochastic realization simulated from the fitted Strauss hard-core point process.

	Coefficient	Lower 95% CI	Upper 95% CI	p Value
Intercept (μ)	-4.01	-4.18	-3.84	< 0.001
Elevation (ρ_1)	0.11	0.09	0.13	< 0.001
Flow Accumulation (ρ_2)	-0.00195	-0.00269	-0.00018	< 0.001
Flow Accumulation 4x scale (ρ_3)	-0.0082	-0.0095	-0.0008	< 0.001
Flow Accumulation 16x scale (ρ_4)	-0.0305	-0.0323	-0.0025	< 0.001
Cost Distance (ρ_5)	-0.0012	-0.0014	-0.0010	< 0.001
Interaction (γ_1)	-1.13	-1.36	-0.88	< 0.001
Interaction (γ_2)	0.56	0.54	0.58	< 0.001

Table 1-1. Parameter estimates for the fitted hybrid Gibbs process model of gentoo nest locations at Port Lockroy, Antarctica. There is zero probability of points existing within 0.28m (h) of each other. From 0.28m to 0.5m (r_1) the probability of occurrence is reduced by e^{γ_1}, and from 0.5m - 1.86m (r_2) the probability of occurrence is increased by e^{γ_2}.

Discussion

While the datasets required to study occupancy at large spatial scales are readily available through remote sensing, the environment experienced by individual plants and animals is usually quite localized and may depend on idiosyncratic features not apparent in remotely sensed imagery (Questad et al. 2014). Species may occupy sites that at a landscape scale appear unsuitable, but at fine scales contain topographic features that produce microclimates vastly different from the regional average and fall well within physiological requirements. SfM provides a means to understand the interactions between organisms (particularly sessile organisms such as plants or nesting birds) and their environment, and provides data on occupancy and abundance that can be used in a range of spatially-explicit modeling frameworks. While SfM is capable of producing highly detailed 3D models, the standard workflows associated with habitat suitability or point process modeling require data in a planar geometry, causing us to collapse our dataset back to a 2D representation of the environment for analysis (Fig. 1-5). The derived metrics that describe key factors of the microclimate such as hydrology or hillshade may be estimated directly from the high-resolution 3D information collected through SfM. This high resolution topographic information has been shown to improve habitat suitability studies, although previous efforts have utilized LiDAR systems, and have recognized that the associated costs are high. While lower resolution datasets may be constructed through intensive point sampling and interpolation, SfM offers the ability to collect this 3D information of comparable quality to LiDAR at a fraction of the cost.

Recommendations for best results

To produce high quality 3D models, it is important to have sufficient overlap between images (between 60-70%) to allow an adequate number of points to match between images. While camera alignment is theoretically possible with as few as eight matched points, in practice the number of matched points should be orders of magnitude larger. Images should be captured from different locations around the site of interest rather than from an in-situ rotation of the camera. Additionally, an ideal camera should be at least 12 megapixels with a 50 mm film equivalent focal length. Photoscan also supports a fish-eye lens camera model in addition to the standard frame camera model, enabling the use of cameras such as the Go-Pro3 in the SfM pipeline. While results obtained using cameras with a fish-eye lens are of lower quality, the trade-off of lower camera weight may often inform the choice of camera particularly when aerial images are required. The parameters of the camera model, a representation of the transformation of light rays from lens to sensor, are estimated at the same time as the scene structure; however, it is also possible to calibrate cameras in a separate process using a checkerboard pattern image.

The density of reconstructed points is a function of the number of matched points between overlapping images and the distance of the point from the camera. This results in a variable density of points across the reconstructed scene. In fact, there may be portions of the scene that cannot be reconstructed in cases where there is insufficient overlap or coverage in images, occlusion of an area, or insufficient surface texture. While interpolation of the mesh can be used to fill these holes (and can be carried out in Photoscan simultaneously with mesh creation), it is preferable to consider image capture paths prior to surveying to minimize the need for interpolation. In this study, there were several small areas that required interpolation due to the lack of matchable features in snow banks, however these represented a very small proportion

of the total surface reconstructed (< 7%) and these areas did not contain penguin nests. It should be noted that the proportion of the surface that was not constructed was far smaller than in most stereo-derived satellite DEMs where cloud cover can result in large holes, particularly in coastal areas.

While it is impossible to calculate errors in the point cloud in the absence of a reference set of measurements of known accuracy, previous studies have compared the accuracies of the SfM approach and traditional LiDAR systems and found the two systems to have similar accuracies (Fonstad et al. 2013). In some instances, greater point densities are achievable through SfM, however it should be noted that errors may not be constant across the scene and may be affected by scene-dependent factors such as the distance of the camera from the scene (Fonstad et al. 2013). Reconstruction may fail completely or result in systematic errors when there is insufficient overlap between images, or when the scene being reconstructed lacks sufficient texture for feature matching. The result may be either a model based on a small subset of the available images with those images that could not be aligned excluded, or a model in which cameras have been incorrectly aligned and erroneous points included. These points can often be manually identified by their position and the projection of the incorrectly aligned camera can be reset. Manual placement of control points between images may then be used to attempt to correct the alignment issues.

Lessons for seabird ecology

In our demonstration of SfM as applied to gentoo penguin nesting, we find strong evidence of intraspecific interactions that are, in fact, more important to the probability of occupancy than the underlying terrain of the nest site. Simulations from the fitted model show

exaggerated clustering, a well-known problem in aggregative point process models (Diggle 1983). It may be the case that the interactions among individuals are more complex than we have assumed here and that additional repulsive forces would stabilize the point pattern. The smallest inter-nest distance in our dataset was 0.28 m and, accordingly, the hard-core distance in our model was estimated to be 0.28 m. However, the average nearest-neighbor distance is 1 m, consistent with previous estimates that have suggested inter-nest distances of around 1 m (Stonehouse 1975). The smallest inter-nest distances found in our dataset could be caused by the reduction from three dimensions to two, with the z component in inter-nest distance lost during the projection onto the planar surface. While the 'spatstat' package (Baddeley and Turner 2005) provides the tools to visualize and summarize point patterns in 3D, the tools for fitting point process models on a 3D surface have not yet been developed.

The interaction effects that we observe may also be due partially to some spatially autocorrelated abiotic feature not considered in our model. The use of SfM to produce a virtual representation of the scene allows us to return to the dataset and derive new explanatory metrics that describe additional features of the landscape in a way not be possible with a traditional field survey. In this way, SfM provides an opportunity for reanalysis if new biological hypotheses arise after the survey and initial analysis.

Finally, the realized point process recorded by SfM in the field is likely to be a sub-optimal arrangement of nests that reflects, in part, the residual influence of the initial colonization process. Simulated nesting patterns highlight deviations between the observed point process and the classic Strauss hard-core process and suggest the potential importance of initial conditions, though a more complete analysis of 'optimal' nesting strategies and non-equilibrium dynamics is required.

Structure-from-Motion offers a cheap alternative to LiDAR to produce high-resolution 3D information on a landscape (Dandois and Ellis 2013)). While the software used in this study (Photoscan) is a commercial product, with the use of free software, such as visualSfM, it is now feasible to completely integrate organism and habitat mapping for the price of a suitable camera. Data collected at this resolution are much closer to the scale that is relevant to behavioral choices or dispersal limitations of individual organisms than most datasets currently being used in either habitat suitability or range modeling (Questad et al. 2014). The ability to characterize micro-habitat, which can be highly heterogeneous at very small spatial scales, will provide ecologists a much better understanding of the niche requirements of a species.

Additional ecological applications of SfM

In addition to the production of high-resolution data at scales suitable for individual level habitat suitability modeling, SfM technology could be applied to a range of other ecological problems. SfM has already been demonstrated as a means of extracting morphological information from individual objects such as corals, fossils or skeletons, enabling researchers to record metrics such as the volume of individual regions of an object, as well as providing a means to store and share the 3D structure of an object without the need for access to the original sample (e.g. Vierling et al. (2008)). This technology also shows potential for rapid, opportunistic abundance surveys for static organisms, such as plants, nesting seabirds, and hauled out seals to name just a few. Traditional panoramic photography has long been used for the census of organisms (Pickard 2002) but complex topography often makes it difficult to align the perspectives of each overlapping image to identify portions of the scene that may have been missed. Through the use of SfM to reconstruct the scene, the uncertainty in overlap between

images can be identified and the coverage of the survey estimated. Incomplete surveys can then be extrapolated as needed.

While we have demonstrated this technology with oblique imagery captured from the ground, results may be further improved, both in terms of consistency of point density and coverage, by using orthogonal imagery captured from an aerial platform, such as a plane, kite, or UAS. With the rapidly decreasing price and increasing usability and autonomy of UASs such as quadrocopters, this technology offers the potential for mapping relatively large areas at high spatial resolution, and producing both DEMs and orthorectified imagery of a location at much higher spatial resolutions than commercially available satellite imagery. All of these technologies will create new opportunities for understanding the fine-scale spatial ecology of organisms.

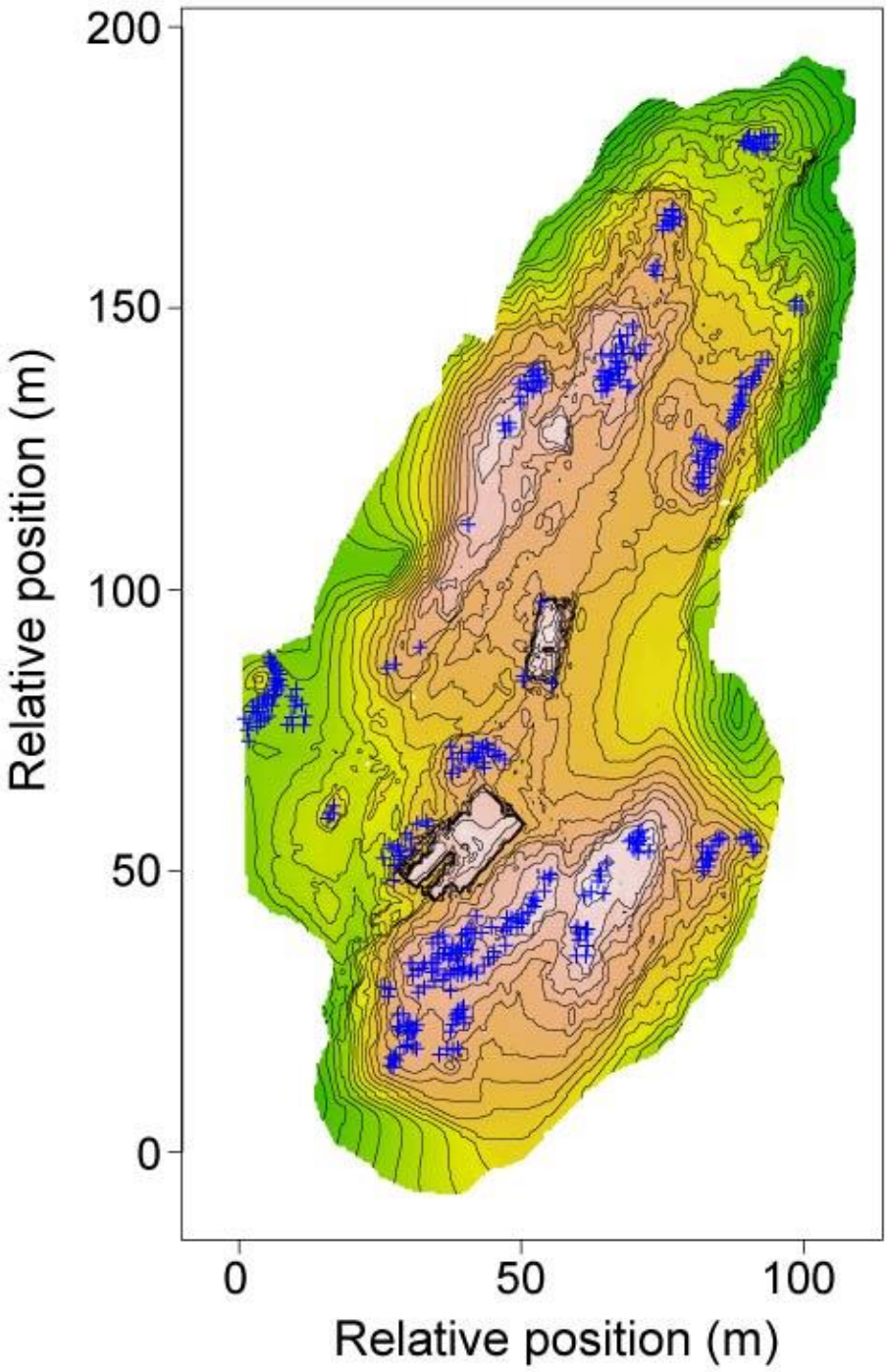

Figure 1-5. Raster and contour map derived from 3D model produced by Structure-from-Motion.

Chapter 2 - The importance of topographically corrected null models for analyzing ecological point processes[2]

Abstract

Analyses of point process patterns and related techniques (e.g., MaxEnt) make use of the expected number of occurrences per unit area and second order statistics based on the distance between occurrences. Ecologists working with point process data often assume that points exist on a two-dimensional x-y plane or within a three-dimensional volume, when in fact many observed point patterns are generated on a two-dimensional surface existing within three-dimensional space. For many surfaces, however, such as the topography of landscapes, the projection from the surface to the x-y plane preserves neither area nor distance. As such, when these point patterns are implicitly projected to and analyzed in the x-y plane, our expectations of the point pattern's statistical properties may not be met. When used in hypothesis testing, we find that the failure to account for the topography of the generating surface may bias statistical tests that incorrectly identify clustering and, furthermore, may bias coefficients in inhomogeneous point process models that incorporate slope as a covariate. We demonstrate the circumstances under which this bias is significant, and present simple methods that allow point processes to be

[2] McDowall, Philip, and Heather J. Lynch. "The importance of topographically corrected null models for analyzing ecological point processes." Ecology (2017) 98: 1764-1770. doi:10.1002/ecy.1877.

Author Contributions: (PM – Philip McDowall, HJL – Heather Lynch) Conceptualization: PM. Data curation: PM. Formal analysis: PM. Investigation: PM HJL. Methodology: PM HJL. Project administration: HJL. Supervision: HJL. Validation: PM. Visualization: PM HJL. Writing – original draft: PM HJL. Writing – review & editing: PM HJL.

Funding provided by the National Science Foundation Office of Polar Programs (NSF/OPP-1255058) and the Institute of Advanced Computational Science at Stony Brook University.

simulated with corrections for topography. These point patterns can then be used to generate 'topographically corrected' null models against which observed point processes can be compared.

Introduction

Spatial point patterns reflecting the distribution of events in time or space are used in a wide range of fields including geology, epidemiology, forestry, and ecology (Diggle 1983, Møller and Waagepetersen 2007). In ecology, point process models have been applied to a variety of problems such as the distribution of plants within a patch (e.g., Stoyan and Penttinen (2000)), the static distributions of animals across the landscape (e.g. Funwi-Gabga and Mateu (2012)), spatial interactions among species (marked point process models; e.g., Illian et al. (2009)), and can be used to make inference about a range of ecological processes (for an extensive review see Velázquez et al. (2016)). Recently, the links between spatial point process modeling and species distribution modelling frameworks such as MaxEnt and logistic regression have been clarified (Renner and Warton 2013, Renner et al. 2015), highlighting their utility for modelling habitat suitability using presence only data. As such, it is likely that point process models will see even greater application within spatial ecology.

A spatial point pattern is a realization of a stochastic point process, in which point locations across a study window are generated according to a stochastic generating mechanism (Cressie 1993). The simplest type of point pattern is the homogeneous Poisson point process (alternatively referred to as ‘complete spatial randomness’ or CSR) in which points are distributed randomly and independently of each other. This type of point process is defined by its intensity λ, which is the expected number of points per unit area across the window 'W' with area

A. Under CSR, points are independent and therefore the expected number of points N in area A is distributed as Poisson(λA). An inhomogeneous Poisson point process extends this model, allowing the intensity of the process to vary with some spatially-structured covariate. A simple ecological model might, for example, treat the locations of trees as a realization of an inhomogeneous Poisson process and their density as a function of covariates such as soil type or elevation.

There are two fundamental types of summary statistics used in the study of univariate point patterns. First order statistics characterize the intensity λ of the underlying process while second order statistics quantify the strength of correlation among points. Points closer together than expected under CSR would be considered clustered, while points further apart than expected under CSR would be considered dispersed. A common metric of second order structure is Ripley's K (Ripley 1976). $K(r)$ is the number of points expected within a distance r from any other point and its sample based estimate, $\widehat{\mathrm{K(r)}}$, is $\lambda^{-1} \sum_{i \neq j} I(d_{ij} < r) / n$, where d_{ij} is the distance between the i^{th} and j^{th} point and I( ) is an indicator function that takes the value 1 (0) when the operand is true (false). For an observed point pattern $\widehat{\mathrm{K(r)}}$ can be evaluated at a range of r values, and deviations between $\widehat{\mathrm{K(r)}}$ and $\lambda \pi r^2$ (the expected number of points within a distance r under CSR) used to indicate inter-point interaction at distance r. It is important to note that inhomogeneity in the intensity of the point process may generate second order properties consistent with clustering even when points are independent of one another. While the theoretical value of $K(r)$ can also be calculated for an inhomogeneous point process (Baddeley et al. 2000), Monte Carlo methods are often used to identify deviations between an observed statistic of the point pattern and its distribution as derived by simulations under a null model. There are many alternative test statistics to $K(r)$, such as the pair correlation function $g(r)$, which also might be

used to understand non-random occurrences of points, and the basic methods available for hypothesis testing are largely shared among metrics.

Software is available (e.g. `spatstat` R package (R Core Development Team 2013,(Baddeley et al. 2015)) for analyzing linear point patterns occurring in $\mathbf{R}^1$ (points generated along a one-dimensional line segment), planar point processes occurring in $\mathbf{R}^2$ (points generated in the two-dimensional x-y plane), and even 3D point processes occurring in $\mathbf{R}^3$ (points generated within a 3D volume). However, the generating processes in many ecological examples, particularly in the terrestrial environment, would not fit comfortably into any of these categories. Ecological data are often recorded and reported as a set of coordinates on a two-dimensional plane, either an x-y pair in a locally-defined spatial reference system or as latitude and longitude in a geographic reference system. The positions of these points represent a projection from a two-dimensional surface embedded in three-dimensional space (the landscape) onto a two-dimensional plane. The projection used, an orthographic projection which converts (x,y,z) to (x,y) by simply dropping the z component, is neither conformal nor authalic (area preserving) and, as such, may only approximately preserve the first and second-order characteristics of the point pattern. If a Poisson point pattern is homogeneously generated on a non-planar surface, S, then it may become inhomogeneous when projected to W with distances and areas in W being less than their original form in S. This results in a bias strictly towards increased intensity, which would bias our inference if the null model is specified as CSR on W rather than S. Failing to account for the transformation to the plane will lead to apparent clustering or, in the case of fitted models, will incorrectly identify covariates as important to the process that are merely correlated with the difference between the terrain and its projection. In an ecological context, terrain slope is the most obvious covariate correlated with the projection to

the x-y plane, though the potential for bias extends to any covariate correlated with slope, such as hydrological flow, soil type, or sun exposure.

While the effects of projecting surfaces to a plane have been widely considered both in cartography and GIS through orthorectification of topographic surfaces, these issues have not been widely acknowledged in the analysis of spatial data in ecology (but see (Baddeley et al. 2015) pages 175-176). It is not unusual for topographically derived information to be included in models of non-stationary Poisson point processes, such that the intensity of points becomes a function of some feature such as elevation or the norm of elevational gradients. However, a topographically correct null model has not been widely considered in the analysis of spatial point pattern data. To address the issues of projecting from S to W, we describe a set of simple correction methods to facilitate the use of topographically corrected null models for point pattern analysis.

Methods

We present several methods for simulating point processes that account for the topography of the generating surface. The first of these methods deals with surfaces that can be written in a parametric form with coordinates $(x, y, f(x, y))$. While surfaces of this nature are unlikely to be found in ecological studies, the algorithm forms the basis of methods that approximate the exact rejection sampling method available for parametric surfaces[3].

[3] R and Python code implementations of these methods (and methods for additional data formats) are included in the Appendix.

Parametric surfaces

If a surface can be represented parametrically in the form

$$S = \{(x, y, f(x, y)) \in R^3 : (x, y) \in W\}$$

where $f(x, y)$ is a differentiable function and W is a bounded domain in R^2, then the surface area can be calculated as the double integral

$$Surface\ area = \iint_W \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dA \qquad (1)$$

where W is the domain of the surface S as projected to the x-y plane. It is this projection of the surface, from S to W, that is typically considered in a point pattern analysis. It is worth noting that we can construct many functions which project to the same region W, any of which might generate a different area of the surface S.

To simulate a homogeneous Poisson point process over the surface S we do not need to compute the full double integral in Eq. 1. Instead we only need a function that can be used to select points with probability proportional to the ratio of the area of the surface and its projection to W, i.e. S_i/W_i . This can be used to construct a rejection sampling-based algorithm to generate points randomly distributed across the surface (modified from (Melfi and Schoier 2004)). The algorithm proceeds as follows.

1. Define a function $g(x, y)$ such that $g(x, y) = \lambda\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$, where λ is the intensity of the point process on the surface.
2. Generate a set of points uniformly randomly over the domain W with intensity equal to the maximum of $g(x, y)$ over the region
3. Calculate $g(x_i, y_i)$ for each point
4. Accept points with probability equal to $\frac{g(x_i, y_i)}{max_{i \in W}\, g(x_i, y_i)}$

The value calculated in Step #4 is the correction factor between the true surface area and that same area projected onto W. This approach can be extended to inhomogeneous point process models by replacing λ with a function that describes the intensity of the point process as a function of spatially-structured covariates (e.g., slope, aspect, elevation). While the approach for parametric surfaces is appealing, it hinges on finding an analytical function f that closely approximates the true surface. While we may not be able to find a suitable function, we can still measure and approximate the surface to allow for topographic correction of the null model.

Raster data

A raster DEM encodes information about elevation in a two-dimensional grid. If we treat the raster DEM as an approximation of the surface S, we can apply the same algorithm described above to simulate topographically corrected point patterns. In this case, we replace the double integral in Eq. 1 with a Riemann sum over the K cells each of area A,

$$Surface\ area = \sum_1^K \sqrt{1 + \left(\frac{\partial f(x_i,y_i)}{\partial x}\right)^2 + \left(\frac{\partial f(x_i,y_i)}{\partial y}\right)^2}\, A. \qquad (2)$$

The quality of this approximation depends on the resolution of the raster, with higher resolutions (smaller values of A) allowing for a better approximation to the true surface. While we might not have a differentiable function $f(x, y)$ in the case of real terrain, we can use the fact that elevation values in a cell's neighborhood provide an approximation of the gradient of the surface within a cell, and from this we can estimate the area of the surface of S that is projected onto each grid cell. There are several algorithms available for estimating the three-dimensional surface area of grid cells in a raster DEM; we use an 8-point algorithm (Jenness 2004) which approximates the area of a cell by constructing three-dimensional triangles between the center of the cell and its 8 neighbors and then summing the area of triangles that fall within the focal cell. This improves

accuracy over methods using only 4 neighbors, and the implementation in the `sp` R package (Bivand et al. 2008) is fast enough that even large datasets are computationally feasible. Because the area calculation is dependent upon the surrounding cells, the area of cells falling on the boundary of the elevation raster will be poorly defined and it is preferable, therefore, to ensure that the raster extends at least one column or row beyond the domain W. Given a raster of surface areas we must then divide by the area of each cell, giving a correction factor S_i/W_i for the projection that can be used in a rejection sampling algorithm. Generating a rasterized correction factor for the projection allows us to use existing tools to simulate topographically corrected point patterns, perform Monte-Carlo based hypothesis tests using a corrected null model, and fit point process models to data.

By using a null model which is an inhomogeneous point process with intensity given by $\lambda S_i/W_i$ we can simulate a point process that approximates a point process operating on the surface S. Additionally we can explore models of inhomogeneous point processes by including $\log(S_i/W_i)$ as an offset term when fitting models (Baddeley et al. 2015). An offset term allows for a covariate to be included in the models with a parameter value fixed at 1. By including the logarithm of the correction factor in the model specification as an offset we can fit the model $\lambda = S_i/W_i e^{\beta x}$, which is corrected for the projection from S to W. While it is possible to estimate the parameters of a point process with interaction terms while including this offset term, it should be noted that while the intensity of the point process should be topographically correct, interpoint distances and areal statistics used in the interaction term are still computed on the x-y plane with no correction for topography, which may lead to a positive bias in the interaction terms of the model.

Terrain driven bias when testing for CSR

To demonstrate how projecting point patterns to the x-y plane can bias statistical tests for CSR, we provide several examples of the kinds of false inference that may result from failing to correct for topography. For a set of synthetic surfaces (Fig. 2-1) we define two point processes: one simulated directly on the surface S using our algorithm for parametric surfaces (referred to as the observed point pattern) and the other a naïve null model of CSR simulated over the projection to W. For each of the surfaces we calculate Ripley's $K(r)$ and the pair correlation function $g(r)$ for both the observed point patterns and the CSR null models to investigate how the projection from S to W can bias second-order metrics of spatial structure commonly used to identify attraction or repulsion among points. We perform two Monte-Carlo based tests using $K(r)$ and $g(r)$, the Diggle, Cressie, Loosmore and Ford test (dclf) (Loosmore and Ford 2006, Baddeley et al. 2014) and the Maximum Absolute Deviance test (MAD) (Ripley 1977), and one parametric test on $g(r)$ (Analytical Global Envelopes [AGE] (Wiegand et al. 2016)), to identify deviations between the observed and null model. These tests address the issue of multiple comparisons, required due to the testing of hypotheses at many values of r (Baddeley et al. 2014), in different ways. The MCMC methods test for deviations based on the ranking of an observed test statistic amongst the same test statistic generated from multiple simulations of a null model. In each MCMC test we used 999 simulations from the null model. The intensity λ was standardized between surfaces, such that the expected number of points was equal between surfaces, to provide approximately equal power in each test. The AGE test assumes $g(r)$ is normally distributed and independent at each value of r, and that for certain null hypotheses (e.g., CSR) the mean and variance of the test statistic can be analytically derived (Wiegand et al. 2016).

We performed each of these tests (MCMC using $K(r)$) and $g(r)$, and AGE using $g(r)$) 10,000 times for each surface and recorded the number of times that we would reject the null hypothesis at a significance level of 0.05.

In order to demonstrate that terrain is likely to affect real world results and to explore the sensitivity of the bias to point intensity, we also simulate homogeneous Poisson point processes with varying point intensities on the surfaces of three published topographic datasets and perform null hypothesis testing (CSR null on W) using the MAD test based on $K(r)$

Finally, we fit models with and without topographic corrections to the 'bei' dataset accompanying the spatstat package to explore how parameter estimation can be biased by the surface on which points are generated. The 'bei' dataset contains the point locations of 3605 trees (*Beilschmiedia pendula*) in the tropical rainforest of Barro Colorado Island. In addition to the point locations, the dataset also provides topographic information on a 5 m resolution grid. Following Baddeley et al. (2015) we include slope as a covariate x_i for the intensity λ

$$log(\lambda) = \beta x_i \quad (3)$$

and compare it to the model including the offset term for topographic correction

$$log(\lambda) = log(S_i/W_i) + \beta x_i, \quad (4)$$

where S_i/W_i is the correction factor associated with the projection from S to W.

Results

Monte Carlo Tests

While the realized Type I error rates for our planar surface (Fig. 2-1a) were, as expected, approximately 0.05 (Table 2-1), non-planar surfaces (Fig. 2-1b-d) resulted in the rejection of the null at a much higher rate. For the most extreme of the surfaces (the exponential function [Fig. 2-

1-d]), the null was rejected in all 10,000 tests when using tests based on $K(r)$. Type I error rates in tests based on $g(r)$ were also higher than expected, however these methods do not have the same power to detect departures from CSR as those based on the cumulative $K(r)$. The plots of $K(r)$ and $g(r)$ show that for all non-planar surfaces (Fig. 2-1-b-d;iii) the envelopes for the observed point patterns fall above the envelopes for the naïve null model, incorrectly suggesting the points are clustered. In contrast, the envelope for the planar surface (Fig. 2-1-a-iii), which serves as a check on our methodology, overlaps with that of the naïve null model as expected, indicating that the null model and observed point pattern have equivalent statistical properties. Simulations over real topographic surfaces indicate that inflated type I error rates are possible for realistic topographies and densities of points, with error rates ranging from the expected 5% up to 13% in the most extreme cases (Fig. 2-2). A comparison of the distribution of correction factors in these real topographies and synthetic surfaces is presented in the Appendix.

Model fitting

While both slope and elevation remain statistically significant in both models of the distribution of *B. pendulu,* the inclusion of topographic correction alters the estimates of the parameters for both slope (Uncorrected: 5.84 [SE 0.26]; Corrected: 5.79 [SE 0.26]) and intercept (Uncorrected: -8.56 [SE 0.34]; Corrected: -9.55 [SE 0.34]). While the differences are small it should be noted that these parameter estimates are on the log scale; the effect on estimated intensity (on S) is substantially larger, as much as 17%.

Surface	MAD Test k(r) Rejection Rate	dclf Test k(r) Rejection Rate	MAD Test g(r) Rejection Rate	dclf Test g(r) Rejection Rate	AGE Test g(r) Rejection Rate
Planar (a)	0.04	0.04	0.03	0.03	0.07
Sine function (b)	0.09	0.06	0.05	0.05	0.08
Step function (c)	0.76	0.76	0.08	0.06	0.10
Exponential function (d)	1.00	1.00	1.00	1.00	1.00

Table 2-1. Rates of rejection of Poisson point process null-hypothesis for patterns generated on 3D surfaces and projected to a 2D plane. Null-hypothesis testing was carried out on homogeneous Poisson point patterns generated on three surfaces using the Diggle, Cressie, Loosmore and Ford Test (dclf) and Maximum Absolute Deviance (MAD) test.

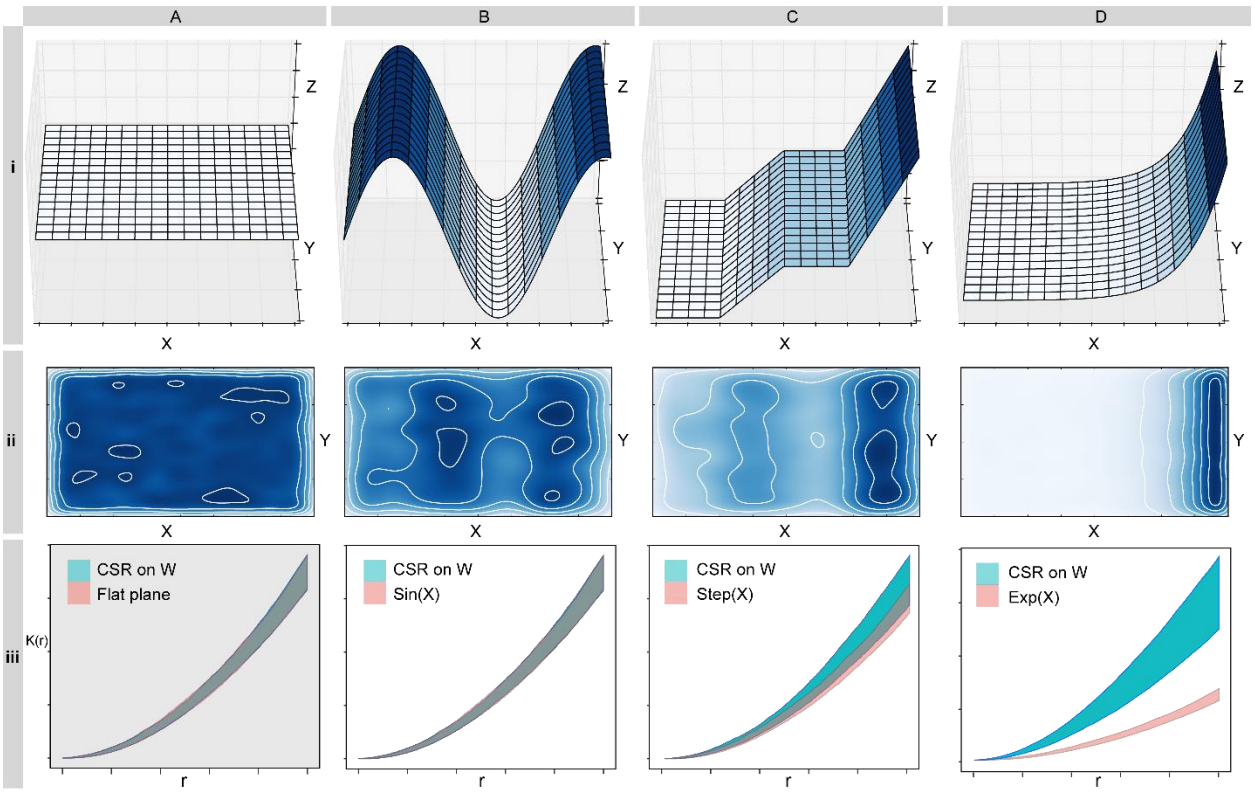

Figure 2-1 – i) Homogeneous Poisson point patterns generated on three 3D parametric surfaces projecting to the same region W. a) z=sin(x); b) stepwise function (note that while this function is not itself differentiable, it is piecewise differentiable which is sufficient); c) z=ex/100. ii) Kernel density estimate of a realization of 10,000 points on the surface from i. iii) Ripley's K for 10000 simulated point patterns on each surface (assuming homogeneity in W)/ The maximum and minimum values of this test statistic are shown in red, while blue denotes the corresponding envelopes from CSR generated in W.

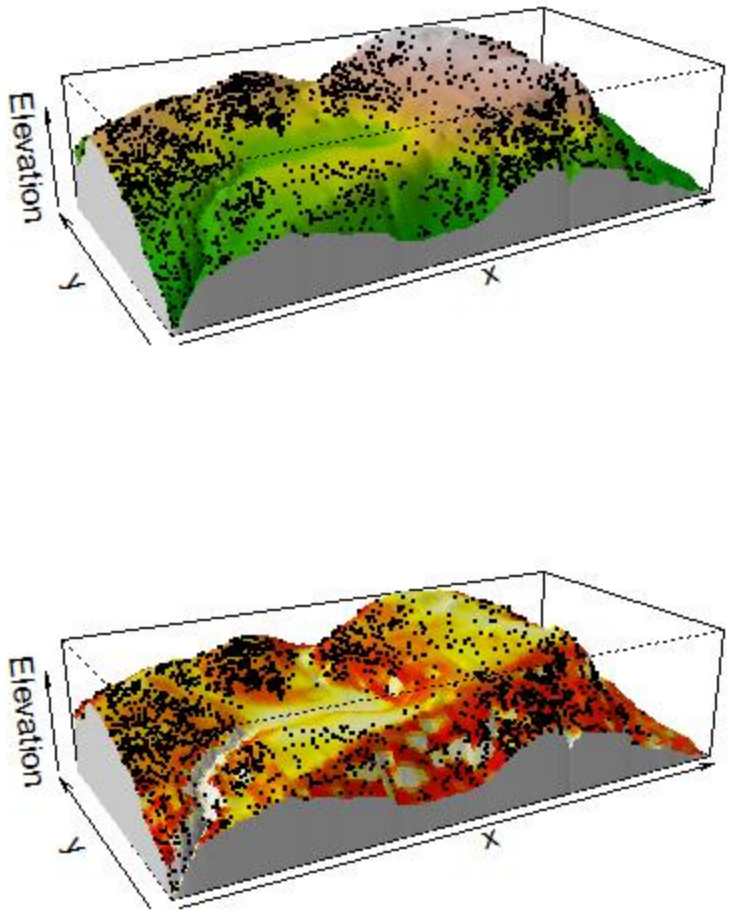

Figure 2-2 – Topographic correction of the ‘bei’ point process dataset. i) The 'bei' dataset from spatstat includes the point locations of 3605 trees and associated topography. ii) Proportional difference in estimated intensity of points between models with and without topographic correction.

Discussion

While planar projections allow for simple analysis, the appropriate null model must account for the non-planar surface on which the generating process operates. When a hypothesis test for CSR is performed in the x-y plane, the intention is often to determine whether there is some structure to the observed points resulting from inter-point interactions or spatial inhomogeneity of the generating process. It is important, however, to recognize that there is a distinction between rejecting an unsuitable null model of CSR and assuming that points are non-randomly distributed.

We have demonstrated that the statistical properties of the point pattern projected onto W are a function of the surface S on which they were created, and that the difference between a naïve null model and a topographically correct model are non-trivial and may lead to false inference. As illustrated by plots of Ripley's K, use of a naïve null model will bias tests towards detecting clustering in point patterns even when points are independent but, due to the projection from S to W, inhomogeneous. In some cases, the properties of the topographical surface are such that the correction factor S_i/W_i is uniformly unity, it may be unnecessary to correct for topography. However, even when the projection to the plane maintains both the first- and second-order properties of the point pattern, these surfaces still exhibit bias in estimates of the intensity of the point process, with λ being consistently overestimated. In the majority of cases, consistent Type I error rates and unbiased parameter estimation require corrections for topography.

Even a cursory review of the ecological literature reveals papers in which it was concluded that slope is an important factor in driving point patterns. Without correcting for the non-area preserving projection, we would expect *a priori* that slope (and covariates correlated with slope)

would have a positive effect on the intensity of the point pattern due to the projection to W; any apparent association between intensity and slope might be a geometric artifact, and so the appropriate null model for concluding an ecologically-driven association requires careful consideration. If we redefine the study area to be a smaller subset of W and defined a local coordinate system on an approximately tangent plane to the surface, we may find that the intensity would be inconsistent with our expectations derived across W.

Recent publications have highlighted the equivalence of point process modelling and more widely adopted species distribution model methods such as MaxEnt and logistic regression (Renner and Warton 2013). These techniques estimate the probability that a point occurs in the i^{th} cell of a regular grid across the study area. Assuming the generating process operates locally on the terrain, the probability of a point falling within the i^{th} cell of a regular grid in W will be a function of the ratio between the area of the surface projected onto the i^{th} cell and the area of the cell. Both MaxEnt and logistic regression, therefore, are likely to be affected by the topographic projection, and may be prone to bias when covariates in the model are correlated with slope.

While we have demonstrated approximate corrections for the effects of projection on summary statistics in the plane, more complex models involving interpoint interactions need further development. In most cases, the biologically meaningful distance between points should be calculated on S rather than W. By weighting interactions by the surface corrected line segment between two points, it should be possible to construct second-order point process models that are also topographically corrected.

Constructing topographically correct null models is straightforward; methods are available for a range of topographic data types and the rejection sampling algorithms required to simulate from these null models are already implemented in various software packages. While CSR is

often the first choice of null model, prior knowledge of a system may suggest that a more nuanced null is appropriate. While we have dealt with corrections to CSR based null models, the same approach can be used to topographically correct any inhomogeneous null model, by applying the same correction factors as calculated for CSR. It should be noted however that correction of non-Poisson nulls (i.e. those that include interpoint interactions) would require further work to calculate interpoint distances across a surface.

Ecologists should consider using these methods in the analysis of point pattern data, and should use topographically correct null models before seeking biological or ecological drivers for clustering. The increased availability of high resolution topographic datasets will provide ecologists unprecedented capacity to understand static spatial distributions of animals, and it is now time to upgrade our tools to consider the full three-dimensional nature of the environment in spatial ecology.

Chapter 3 - When the "selfish herd" becomes the "frozen herd": Sub-optimality in colonial seabirds and its consequences for persistence under decline[4]

Abstract

Hamilton's 'selfish herd' hypothesis suggests that aggregations may be driven by an individual's effort to minimize their risk of predation by surrounding themselves with conspecifics. In fluid, highly mobile aggregations, individuals are constantly moving in response to changing environmental conditions, the locations of predators, or the movements of conspecifics. However, when the ability to rearrange is limited and spatial reconfiguration occurs on slower time scales than changes in population size, systems may become trapped in sub-optimal arrangements. We use simulated annealing to demonstrate that *Pygoscelis* spp. penguin colonies are frozen in sub-optimal spatial arrangements and discuss some hypotheses as to why such sub-optimal arrangements might persist.

[4] Author Contributions: (PM – Philip McDowall, HJL – Heather Lynch) Conceptualization: PM. Data curation: PM. Formal analysis: PM. Funding acquisition: HJL PM. Investigation: PM HJL. Methodology: PM HJL. Project administration: HJL. Supervision: HJL. Validation: PM. Visualization: PM HJL. Writing – original draft: PM HJL. Writing – review & editing: PM HJL.

Thanks to Catherine Graham for her suggestions during the preparation of this manuscript, and to Thomas Sayre-McCord for his work in collecting the imagery used as a basis for this study.

Funding provided by the National Science Foundation Office of Polar Programs (NSF/OPP-1255058) and the Institute of Advanced Computational Science at Stony Brook University.

Introduction

The theory of the selfish herd, proposed by Hamilton (1971), suggests that aggregations of individuals arise when individuals seeking to minimize their 'domain of danger' surround themselves with conspecifics. This attempt to optimize one's individual spatial configuration leads to larger, more compact aggregations in which the risk of predation varies spatially through the aggregation. Studies have identified a number of systems that appear to exemplify the 'selfish herd', particularly among highly mobile organisms such as flocking birds or schooling fish (McKaye et al. 1992, Viscido and Wethey 2002, Couzin and Krause 2003, King et al. 2012). Highly mobile individuals and the aggregations they form can rapidly reorganize in response to changes in the shape of the cost function, such as the arrival of a predator (Parrish and Edelstein-Keshet 1999). However, when reorganization of a spatial configuration is impeded or occurs on a slower time scale, the ability of an aggregation to rearrange may be limited. One manifestation of this phenomenon occurs in colonially nesting seabirds, where the spatial aggregation of nests is driven by predation at the edge of the colony (Schreiber and Burger 2001). *Pygoscelis* spp. penguin colonies are approximately hexagonally-packed aggregations of nests distributed across bare rock areas in coastal Antarctica and represent an excellent case study for understanding the dynamics of the 'selfish herd' when reorganization is impeded. The most prolific predators of the *Pygoscelis* penguins, the Skua (*Stercorarius sp.*), often initiate predation from the ground on the peripheries of a colony and, as a result, reproductive success on the edge of a colony can be up to eight times lower than in the center (Emslie et al. 1995, Hahn and Peter 2003). Accordingly, individuals should seek to minimize their risk by searching for a nest site that maximizes the occupancy of their neighborhood by conspecifics. The associated Allee effect may slow the

establishment of new colonies and is likely to be important for the spatial patterning of colonies that do form.

Given a free choice of location across a homogeneous landscape and sequential filling of available nest sites, we might expect that colonies would grow outwards in a circular fashion as individuals select nest sites that maximize their neighborhood. However, recent advances in the use of high-resolution satellite imagery have revealed that penguin colonies are often highly fragmented, with nesting areas forming spots, stripes, and labyrinthine patterns (Fig. 3-1).

We sought to understand why apparently sub-optimal arrangements would be nearly universal in *Pygoscelis* colonies. We hypothesized that these temporally dynamic arrangements of individuals within a colony are the result of individual level decisions (the 'selfish' part of the selfish herd) that seek to maximize the individual's fitness by providing the best chance of successful reproduction. However, when the optimality of an individual choice is dependent on the choices of all other individuals, the ability to find a globally optimal solution may be impacted. In a homogeneous landscape, the optimal configuration of penguins is simple to derive, however, terrain creates a heterogeneity in 'suitable' habitat that complicates the search for a globally optimal arrangement. To understand the benefits of colonial breeding, and the associated spatially structured Allee effects, we must consider habitat suitability and conspecific interactions simultaneously. Identifying how spatial patterning and its temporal dynamics might influence vital rates is essential for understanding the population dynamics of colonial seabirds. Spatial patterning of animal aggregations is inherently interesting as a driver and reflection of animal behavior. Understanding such patterns takes on additional urgency, however, when they either feed into population dynamics or can be used as a proxy for population health.

Methods

We focused our analysis on the nesting patterns of *Pygoscelis adelie* (Adélie penguin) at Beagle Island, a 180 ha island at the northern extent of the Western Antarctic Peninsula. Using high-resolution 3D models of terrain and nests, produced through photogrammetry (McDowall and Lynch 2017), we produced a set of five environmental covariates for 1,893,597 1.2m diameter hexagonal grid cells, each of which may be occupied by a single penguin nest. The use of a hexagonal grid allows interactions between nearest neighbors to be isotropic, in contrast to a more traditional square grid in which the centers of diagonal nearest neighbors are more distant than the cardinal neighbors.

We hypothesized that the spatial arrangement of penguins on the island represents a convolution of the optimal configuration for predator avoidance and the spatial configuration of the landscape itself (specifically, well-draining areas at the top of local peaks in elevation). To separate the effects of habitat quality and intraspecific interactions, we constructed Bayesian auto-logistic use-availability models for *Pygoscelis* penguins on our hexagonally-gridded landscape. The auto-logistic model is a Markov random field model that extends the traditional logistic model often used to model binary occupancy over a two dimensional grid by incorporating an additional term, which allows for spatial dependence between the response variable, while simultaneously modelling the effect of spatially structured covariates (Hughes et al. 2011). This model is of the form:

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta X + \gamma\left(\sum_{neighbors} Y_i\right),$$

where Y_i is the 0-1 coded occupancy of the i^{th} cell, p_i is the probability that the i^{th} cell is occupied, α is an intercept term, β and X are, respectively, vectors of parameters and covariates relating to the abiotic conditions at the i^{th} cell, γ is the strength of nearest neighbor interaction, and $\sum_{neighbors} Y_i$ is the number of the i^{th} cells neighbors that are occupied.

We parameterized this model using variables relevant to individual penguins making decisions about where to nest, while simultaneously modeling the interactions among conspecifics (Fig 3-2 and Fig 3-3b). Water damage to eggs is an important factor in breeding failure, and *Pygoscelis* penguins invest a lot of energy into building and maintain a large pebble nest to protect their eggs. We therefore included 'flow accumulation', a hydrological metric, in the model. Other variables included were elevation, slope, and distance to the shoreline cost-weighted with the slope of terrain that must be crossed. Additionally, for each cell we calculated the occupancy of the 6 adjacent grid cells that form its neighborhood. We used uninformative priors ($\sim N(\mu = 0, \sigma = 10^6)$) for all model parameters, and fit the model using 50,000 samples from 3 chains under Hamiltonian MCMC sampling with a No-U-Turn Sampler.

We used a metaheuristic approach, simulated annealing[5], to explore whether observed patterns of nesting are sub-optimal relative to the apparent preference for both abiotic habitat and the presence of adjacent nests. At each iteration, we proposed a stochastic movement of a nesting penguin. We then calculated the energy potential, defined as the suitability of the arrangement under our model, and accepted it with probability $min\left(1, e^{\frac{\gamma_o - \gamma_p}{T}}\right)$, where γ_o and γ_p are the energy of the original and proposed configurations, respectively, and T (systematically reduced in each iteration) is a temperature parameter controlling the probability of accepting a transition

[5]Python implementation of Simulated Annealing available at https://github.com/phil-mcdowall

to a less optimal state. This approach is well suited to finding local optima in a large configuration space where finding the global optimum is neither feasible nor necessary.

Results

The penguin colony at Beagle Island is in a sub-optimal arrangement, with greater patchiness than configurations with a lower 'energy' configuration. Not surprisingly, allowing individuals to make only ‘optimal’ moves (i.e., runs in which the ‘annealing’ part of the algorithm is disabled) yields sub-optimal arrangements compared to those achieved with annealing. While the best configuration achieved through simulation is likely to be far from globally optimal, our simulation easily finds lower energy (more optimal) configurations than are found empirically at Beagle Island (Fig 3-2c). Accordingly, our model and the existing map of penguins at Beagle Island strongly suggest that penguins persist in a highly sub-optimal arrangement.

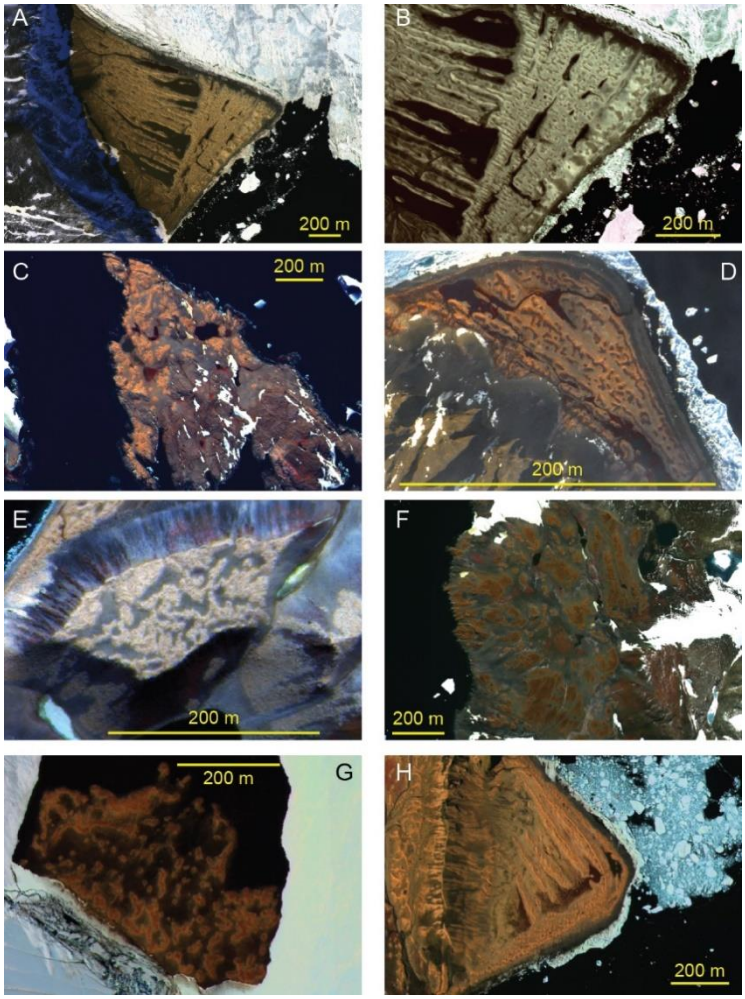

Figure 3-1. Satellite image (Imagery Copyright 2015 DigitalGlobe, Inc.) of *Pygoscelis* colonies illustrating the small scale spatial structuring of colonies. 1A&B: Cape Adare - 11 Dec 2015, 1C: Avian Island - 26 Jan 2010, 1D: Beaufort Island - 25 Dec 2011, 1E: Paulet Island - 4 Feb 2016, 1F: Hope Bay - 17 Jan 2016, 1G: Cape Crozier - 15 Dec 2015, 1H: Possession Island - 14 Jan 2016

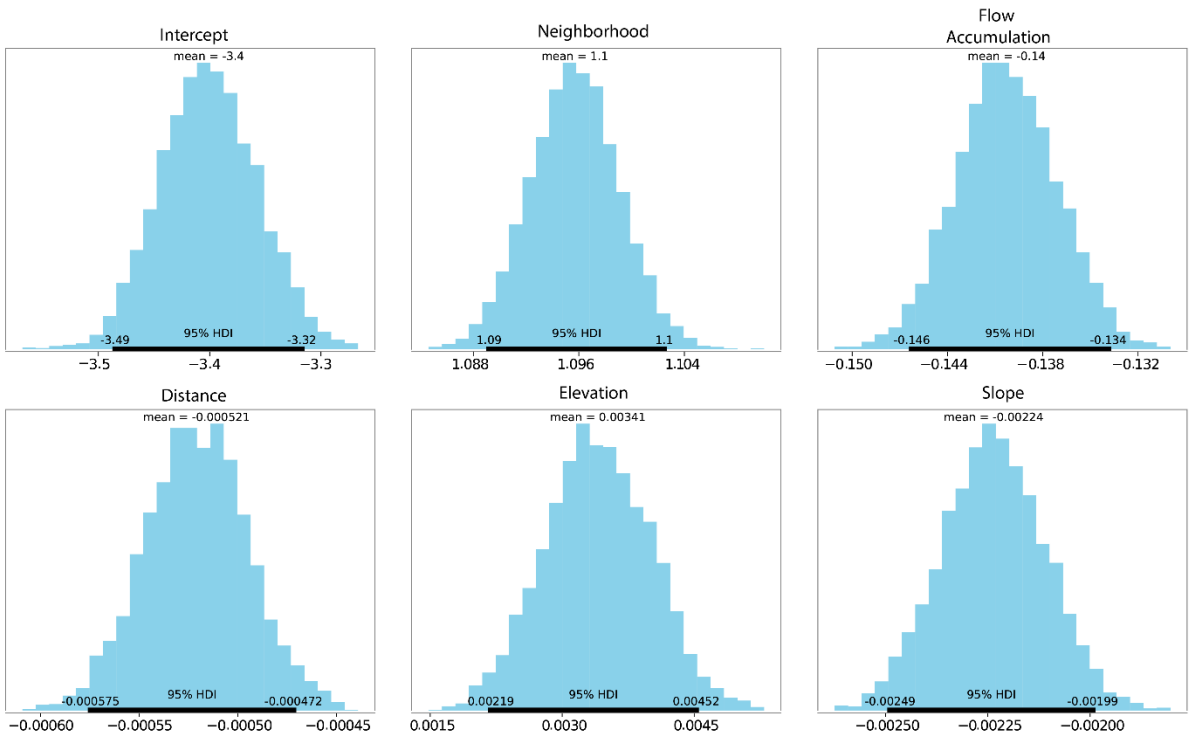

Figure 3-2. Posterior distribution of parameters of the auto logistic model of nest occupancy at Beagle Island. Intercept = -3.40 (95%CI: -3.49, -3.32), Neighborhood = 1.10 (95%CI: 1.09, 1.10), Flow Accumulation = -0.140 (95%CI: -0.146, -0.134), Cost Weighted Distance = -0.001 (95%CI -0.001, -0.001), Elevation = 0.003 (95%CI: 0.002, 0.005), Slope = -0.002 (95%CI: -0.002,-0.002).

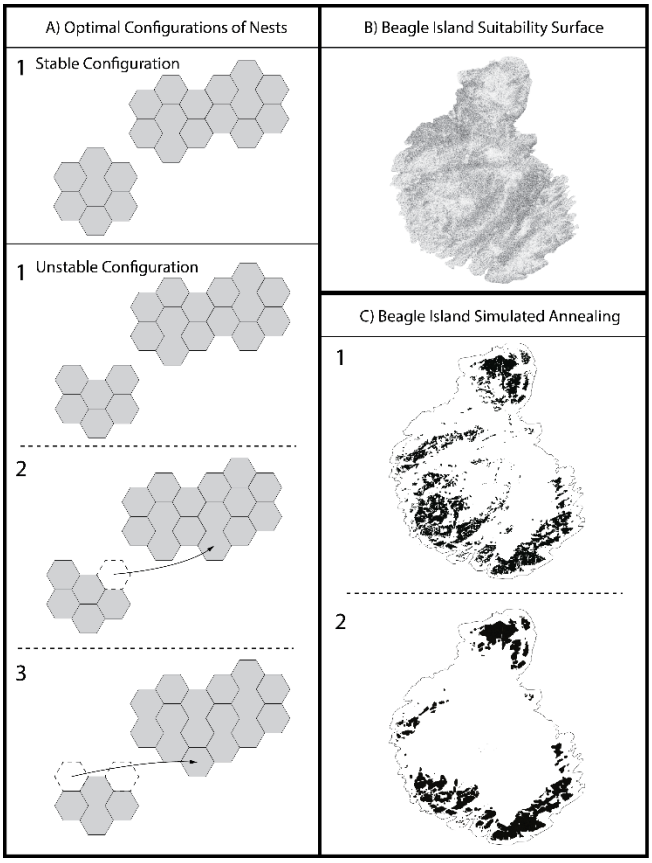

Figure 3-3. a) Optimal configurations of *Pygoscelis* nests in a hexagonal packing scheme on a homogeneous landscape. Some configurations of nests are 'stable' and no individual can improve their position by moving alone, while in other configurations sequential movement of individuals to optimal positions can lead to new configurations in which first order neighborhood occupancy is maximized. b) Suitability surface of Beagle Island estimated through high-resolution photogrammetry and auto-logistic modelling. c) Original configuration of nests observed at Beagle Island (1) and a more optimal arrangement reached through simulated annealing (2).

Discussion

When colony size decreases, geometry dictates that the proportion of sites available on the edges of the colony increases. This forces more birds to assume one of these lower quality nesting locations and potentially makes more birds vulnerable to predation, lowering the average reproductive success of even an ‘optimally’ arranged colony. However, as these colonies contract, they frequently appear to simultaneously fragment into smaller sub-colonies with greater exposed perimeter than would occur if the colony was not fragmented (Jackson et al. 2005). We demonstrate that penguin colonies are sub-optimal relative to inferred preference for landscape characteristics and are patchier than would be predicted based on the estimated strength of nearest neighbor interactions. This leads us to the question: Why are penguin colonies sub-optimally arranged?

We suggest several drivers, none mutually exclusive, that might explain the apparent sub-optimality of spatial dynamics: (1) nest site inertia, (2) short-range pair interactions, (3) individual level decision making, and (4) incomplete information due to limited perceptual range. The net result of these mechanisms leads to a configuration in which the 'selfish herd' becomes frozen in a sub-optimal configuration, with potential consequences for the temporal evolution of pattern, and therefore abundance, in a declining population. This scenario has received little attention in biology but is analogous to well-studied systems in statistical mechanics and condensed matter physics in which non-equilibrium states or incomplete phase separation becomes kinetically frozen.

Adélie penguins are faithful to an individual nest site, and by consequence, colonies of penguins display far greater spatiotemporal inertia than most other animal systems in which selfish herd dynamics have been examined. In a newly established colony, sequential choice of

sites might conceivably lead to an acceptably optimal arrangement as the selection of an optimal nest site is relatively trivial; each individual chooses a location adjacent to the founders' subcolony unit that maximizes their neighborhood and habitat quality. However, when an extant colony undergoes stochastic changes in size (through mortality, emigration or skipped breeding), high nest site fidelity and philopatry ensures that the new configuration must be a rearrangement of the existing configuration, which is more difficult to optimize. This rearrangement, in which each individual makes at most a single move in a breeding season, is slow relative to the life span of a breeding bird and population level dynamics. Additionally, in contrast to more traditional animal aggregations, in *Pygoscelis* penguins the benefits of nesting near conspecifics only occur at very short ranges, acting at the nearest neighbor level. This short-range interaction prevents distinct sub-colony units from experiencing attraction between each other and ensures that only individual movement decisions can be effective.

When the number of nesting individuals decreases, some proportion of individuals will find their nest site quality drops as a function of their neighborhood. If these established individuals decide to move, they may have difficulty finding a nest site that improves on their current location and may only be guaranteed a site of equivalent or lesser suitability at some other location on the periphery of the colony. Our simulated annealing approach demonstrates that arrangements of nests can be found that reduce the number of individuals on the exterior of colonies. Critically, however, local optima that cannot be escaped by rational sequential movement of individuals is common. As individuals act independently, each penguin must decide to move based on incomplete information; their final neighborhood depends upon the subsequent moves made by all other individuals that might complete its neighborhood. Additionally, only those penguins on the exterior of a sub-colony stand to make significant gains

from moving, while those occupying interior sites might prefer the colony to remain in their current configuration. Without coordinated movement these factors cause local optima in the configuration space in which no individual can guarantee it will improve its condition by moving without knowledge of the moves that might subsequently be made by other individuals. This lack of complete information, both of the landscape and the moves that might be made by other individuals, prevents the colony from reaching a globally optimal configuration.

A penguin's dilemma can be understood by analogy to the Nash equilibrium in game theory. Nest site fidelity results in an ‘iterative’ game where each year’s strategic decision must build off the previous year’s, and payoffs (reproductive success) are awarded at each iteration. Each individual in the colony has a strategy for reproduction as reflected in their chosen nest location. If all other individuals in the colony keep their strategies unchanged, no individual has an incentive to move as they cannot do better than their current strategy. While Nash equilbria are used in the context of the evolution of aggregative behavior (Wood and Ackland 2007), models of the 'selfish herd' tend to consider time periods over which closure can be assumed, often with the benefits of proximity to conspecifics as a continuous function of distance. However, when aggregations are considered over longer time periods and when conspecific interactions only occur at short distances, these Nash equilibria may arise within the aggregation. Though sub-optimal nesting arrangements may arise through a number of mechanisms, the existence of a sub-optimal state implies transient dynamics may be 'frozen' by nest site fidelity in stochastically changing populations. Without nest site fidelity the spatial configuration of a colony might be fluidly rearranged in each breeding season, avoiding the potential for these ‘Nash equilibrium’-driven frozen states. The existence of such 'frozen states' may allow hysteresis in the spatial patterning of penguin colonies under growth and declining abundance.

Colonies in decline may be susceptible to fragmentation, increased edge effects, and accelerated deterioration of a colony. Nest site fidelity may be a useful optimization strategy in increasing populations, but it also reduces the ability of penguins to optimize their spatial configuration and, in doing so, may make them susceptible to critical collapse.

Chapter 4 - Self-aggregated pattern formation in seabird colonies drives hysteresis in fluctuating populations[6]

Abstract

Aggregations are common in ecological systems, and may be broadly classified as externally driven, through constraints such as resource availability, or self-organized, in which interactions between individuals drive aggregation. Self-organization has been studied in a variety of highly temporally dynamic systems, however its role in more slowly evolving systems has not been well considered. We demonstrate that the aggregations observed in colonially nesting seabirds are a convolution of both external forcing arising from spatially structured terrain, and slowly evolving self-organized dynamics. Using an individual-based modelling approach we demonstrate that the spatial configuration of seabird colonies is a function of stochastic events at the individual level, and that declining abundance can lead to fragmentation even in a homogeneous environment. Strong edge effects from heterogeneous predation at colony edges creates a positive feedback cycle involving declining abundance, spatial fragmentation, and accelerating population decline. This model provides a mechanistic understanding of complex spatial structuring in penguin colonies and suggests the possibility of critical collapse in seabird populations.

[6] Author Contributions: (PM – Philip McDowall, HJL – Heather Lynch) Conceptualization: PM. Data curation: PM. Formal analysis: PM. Funding acquisition: HJL PM. Investigation: PM HJL. Methodology: PM HJL. Project administration: HJL. Supervision: HJL. Validation: PM. Visualization: PM HJL. Writing – original draft: PM HJL. Writing – review & editing: PM HJL.

Thanks to Thomas Sayre-McCord for his work in collecting the imagery used as a basis for this study.
Funding provided by the National Science Foundation Office of Polar Programs (NSF/OPP-1255058) and the Institute of Advanced Computational Science at Stony Brook University.

Introduction

Aggregations are observed across ecology, at macroscopic scales in the form of schooling fish (青木一郎 1982, Niwa 1994, Parrish and Edelstein-Keshet 1999), flocking birds (Emlen 1952, Hildenbrandt et al. 2010) or herds of ungulates (Couzin and Krause 2003, Conradt et al. 2009), through to microscopic scales such as bacterial aggregations (Ben-Jacob et al. 1994, Chowdhury et al. 2004) and, even down to pigmentation patterns within individuals (Murray 1981, Manukyan et al. 2017). These aggregations arise through individual responses to external stimulus, such as the availability of resources, or through the interactions of individuals. Many of the most complex and dynamic aggregative systems, however, occur through the interaction of individuals independent of any global information about the aggregation as a whole (Couzin and Krause 2003). Systems in which aggregation arises from (often short-range, or 'local') interactions between individuals, independent of external stimuli, are characterized as 'self-aggregating' (Rohani et al. 1997). Ecologists have long been fascinated with self-aggregating dynamics because even simple interactions can result in novel emergent properties, occurring at the level of the system rather than being specified at the individual level (Parrish and Edelstein-Keshet 1999, Camazine 2003). For instance, spatial patterning that exists at the level of the system but is not imposed by the environment (Couzin and Krause 2003) would be considered an emergent property of a system.

While much work has been done to model dynamic self-organizing systems, such as schooling fish, the role of self-organization may be missed when a system's spatial dynamics are slower than the period of observation. Colonial seabirds represent an extreme example, with spatial re-arrangement occurring on an annual basis and dynamics playing out over decades and even centuries. While the role of interactions in these systems has been considered from an

evolutionary point of view, the costs and benefits are usually calculated based on a static configuration of nests with little consideration for self-organization or other processes playing out over long periods of time.

In seabirds, colonial nesting is almost universal, with over 95% of species demonstrating some degree of colonial nesting (Wittenberger 1985, Rolland et al. 1998). Previous definitions of colonial breeding have focused on nesting space as the limiting resource that spatially defines the 'colony' as a biological unit (Wittenberger 1985, Kildaw et al. 2005), but this conception of the colony ignores the role of intraspecific interactions that may also drive spatiotemporal aggregation. Such aggregations are likely to result from a convolution of preference for terrain, with individuals selecting nesting locations based on the availability and suitability of nesting space, as well as self-aggregating dynamics driven by conspecific interactions. For many seabird species, colonization of new terrain is rare (Kildaw et al. 2005), and individuals may prefer to join large extant colonies rather than establishing new colonies. Despite decades of interest in coloniality as a biological phenomenon (Lack 1967, Wittenberger 1985), the role of conspecific interactions and their relationship to nesting terrain suitability, remains unclear. While interactions between individuals are functional, affecting reproductive success in the breeding season, interactions within these aggregations may become particularly important if the emergent properties of the system feed back into reproductive fitness of the individual and, by extension, of the larger breeding population. We use an individual-based model in which interactions play out on a real (empirically derived) high-resolution three-dimensional landscape to examine the factors leading to aggregation within seabird colonies, to explore the emergent properties of such spatial aggregations, and to understand its impacts on population dynamics.

Our analysis is based on observations of Adélie penguins (*Pygoscelis adeliae*), which nest in densely aggregated colonies comprising as many as several hundred thousand nests. At the smallest spatial scale, that of individual nests, colonies consist of densely packed but regularly-spaced nests within a contiguous nesting area. Groups of these tightly-packed aggregations – each of which we call a 'sub-colony' – may occupy a contiguous area of terrain, such as an island or cliffside, to form a single 'colony' of seabirds. Individuals within a colony compete for a range of resources, from distributed resources such as prey, to local resources such as suitable nest sites and nesting material (Carrascal et al. 1995, Moreno et al. 1995, Hunter and Davis 1998). In addition, colonies suffer density-related costs to fitness such as increased chances of disease and parasite outbreaks (Tella 2002, Brown and Brown 2004, Lynch et al. 2010, Rifkin et al. 2012) and, because large colonies attract more predators, increased predation (Emslie et al. 1995). There have been many attempts to explain the causative mechanisms of coloniality and the benefits that being in close proximity to conspecifics must confer; current theories involve information transfer about the quality and location of food resources (Bayer 1982) or high quality nesting habitat (Doligez et al. 2002), the availability of extra-pair copulations (Wagner 1993), predator swamping (Hamilton 1971, Wittenberger 1985, Ainley et al. 2005), and collaborative predator defense (Anderson and Hodum 1993, Brunton 1997), all of which would imply self-organized aggregative dynamics with varying levels of environmental forcing. While these factors may be the ultimate drivers of self-organization within a breeding colony, the proximate causes of self- aggregations must be mediated by the process of nest site selection. Initial nest site selection is likely to be based upon natal philopatry or, in the case of migrants, information acquired from the nest site selections of other individuals within the colony (Zador et al. 2009). In subsequent years, however, the process involves only two choices,

the choice to change one's nesting site from one breeding season to the next and, conditional on changing location, the choice of destination. Importantly, fidelity to a selected nest site from one breeding season to the next creates an 'inertia' that maintains the spatial arrangement of nests even as suitability changes.

A large proportion of chick mortality events in *Pygoscelis* colonies are the result of predation by the Brown and South Polar Skuas (*Stercorarius antarcticus and Stercorarius maccormicki*) and Giant Petrels (*Macronectes giganteus and M. halli*) (Emslie et al. 1995). Predation occurs at the edges of sub-colonies (Eklund 1964, Spurr 1975) both due to the greater accessibility of nests at the edge (predators usually land and attack from a standing position) and the reduced ability of smaller numbers of penguins to collaboratively defend edge sites. In fact, predation on the exterior of penguin colonies may be almost eight times higher than that experienced in interior nests (Emslie et al. 1995), with similar edge effects reported for other colonially nesting bird species (Coulson 1968). It is often assumed that the benefits of coloniality, in this case protection from predation, outweigh the costs and that being part of a colony confers greater fitness. However, the benefits accrued for large populations imply positive density dependence at low abundance (an Allee effect). Reluctance to settle at the peripheries of a colony and reduced breeding success at edge sites may constrain colony establishment or hasten local extirpation. Importantly, these edge effects operate at the sub-colony level, which suggests that Allee effects are not homogenous across a colony and that the optimal (predation-minimizing) arrangement in a homogeneous environment should be a single, circular sub-colony. Satellite observation of real colonies, however, reveal that colonies are often highly fragmented with many small, non-circular, sub-colonies in a complex spatial arrangement that does not solely reflect the underlying terrain. Understanding the mechanism creating these

complex patterns, recently revealed through aerial and satellite imagery, motivated the current study.

Model description

Here we use an individual-based model (IBM) to understand the linkages between individual-level behaviors and spatial patterning within an Adélie penguin colony[7]. This modelling framework easily incorporates interactions between individuals, and between individuals and their environment, allowing population level properties to emerge from individual level behaviors (DeAngelis and Grimm 2014). The individual-based approach is particularly suited to exploring questions with complex localized interactions between individuals and is therefore a tractable method to study spatial configurations of individuals within a population.

Adélie nests are typically tightly packed at small scales, with inter-nest distances of 0.68 m being highly conserved (mean = 0.68 m, sd = 0.09). We can therefore approximate continuous space as a discrete hexagonal grid, in which each cell represents a potential nest site. This approximation reduces computational complexity, incorporates the approximately hard-core repulsion between nest locations at short distance, a form of territoriality common in self-aggregating systems (Camazine 2003), and allows us to model interactions among nest sites as being isotropic.

We hypothesize that the spatial patterning observed in Adélie colonies results from both conspecific interactions and individual preference for suitable habitat and use a Bayesian auto-logistic model of nest site occupancy to separate the effects of conspecific interactions and

[7] Python code implementation available at https://github.com/phil-mcdowall

habitat preference. The auto-logistic model extends the traditional logistic model by incorporating an additional term which, for any given cell, is the proportion of the 6 neighboring cells which are occupied. Fitting this model requires information on both the location of penguin nests and the abiotic characteristics of available nest sites. We use a high-resolution digital elevation model (DEM), created through a photogrammetric process applied to aerial imagery captured at Beagle Island in the Danger Islands (McDowall and Lynch 2017, Borowicz et al. 2018), as the basis for this model. In addition to the derived topographic data, the imagery provides the locations of nesting penguins, yielding very high spatial resolution data on both the position of individuals and the landscape they occupy (McDowall and Lynch 2017).

Abiotic conditions at the nest site play an important role in reproductive success. While the Antarctic as a continent is, on average, very dry, the coastal regions where penguin colonies are found experience high levels of precipitation during the austral summer. In addition, temperatures are frequently above freezing, causing extensive snow melt, a high level of hydrological flow across the site, and frequent flooding of many potential nest sites. Nests that become flooded during the incubation phase are likely to be unsuccessful due to suffocation of the developing chick, while flooding occurring post hatching may result in the loss of the chick to hypothermia. Flow accumulation, a metric derived from the DEM which estimates the flow of water across the landscape, is used as a proxy for the likelihood of flooding for each nest site. Covariates used to model the probability of a nest sites being occupied include the DEM itself, which captures the role of nest elevation, the flow accumulation metric, and two other DEM-derived products: slope, cost-weighted distance to shore. The cost-weighted distance input uses slope as a friction surface for weighting and provides a metric of the difficulty of access of a particular location on an island. Using this auto-logistic model, we find that both conspecific

interactions and habitat preference are involved in nest site selection, with individuals preferentially selecting locations that are surrounded by conspecifics and likely to remain dry through the breeding season (Fig. 4-2).

We use the parameters of the best-fitting model and the DEM for Beagle Island to estimate the suitability of 1.8 million hexagonal cells across the landscape, each one of which represents a potential nest site location. Each of these cells is characterized by three state variables: abiotic quality derived from the auto-logistic model, current occupancy, and the proportion of the six adjacent locations that are also currently occupied. At each time step in the model these potential nest sites may be occupied by a single breeding individual. The model is initialized by seeding the landscape with a small sub-colony of individuals at a location that is selected with probability proportional to its quality. With the Allee effects introduced by Skua predation, new colonizations of *Pygoscelis* colonies are exceptionally rare. In fact, there has been only one known Adélie colonizations since 1900, and very few for the other *Pygoscelis* spp. penguins (Chinstrap and Gentoo) (Lynch et al. 2010, Lynch and LaRue 2014, Humphries et al. 2017).

Adélie penguins form strong bonds with a partner and have relatively low levels of extra-pair copulations. Additionally, there is a high degree of synchrony between the arrival at the colony of male and female birds (Davis and Darby 2012). In this model, we track only males in the population, who are reported to have the highest nest site fidelity, and assume that the actions of the other breeding partner are the same as those taken by the individual tracked in the IBM. Individuals are characterized by their age, reproductive history, and nest site location. The IBM uses an annual timestep, capturing four key processes within each year. Three of these, reproduction, nest-site fidelity, and nest site selection, occur during the 3-month breeding season,

while survival operates in the remaining 9 months of the year outside the breeding season (Fig 4-1).

Survival

Age-specific probabilities for overwinter survival from year $\boldsymbol{t}$ to $\boldsymbol{t+1}$ are modeled separately for immature (≤3 year old) and mature (>3 year old) individuals. We model survival as the result of a draw from a Bernoulli distribution, where the probability of survival ($p_{a,t}$) for a given year (t) is shared between individuals of each of the two age groups (a).

$$s_{i,t} \sim Bernoulli(p_{a,t})$$

In each time step of the model the probability ($p_{a,t}$) is drawn from a Gamma distribution for the appropriate age class. The hierarchical nature of this sub-model reflects the fact that survival is likely to be correlated between individuals in any given year in response to the conditions encountered during the non-breeding seasons. These distributions are parameterized using data from Ainley (2002), Hinke (2012) and Hinke et al. (2017).

$$p_{a,t} \sim \begin{cases} Gamma(\alpha = 10.82, \beta = 2.64) & a < 3 \\ Gamma(\alpha = 23.72, \beta = 9.87) & a \geq 3 \end{cases}$$

Adult overwinter survival is independent of the nest site selected; individuals who have died in the overwinter period are removed from the population at the start of each model year and their nest sites marked as unoccupied.

Nest Site Fidelity

Pygoscelis penguins are believed to have a high degree of nest site fidelity, often returning to the same nest location year after year. We apply three different models of nest site fidelity: Total – Individuals select a nest site and return to the exact same site each year; Partial – Individuals return to their previous nesting location unless the previous two years have resulted in total failure to successfully fledge chicks, in which case they will relocate their nest to an unoccupied nest site within the colony; None – Individuals demonstrate no nest site fidelity and select from any nest site across the landscape in each time step. While this final parameterization (no nest site fidelity) is biologically unrealistic for *Pygoscelis* penguins given known rates of natal philopatry and nest site fidelity, it allows us to demonstrate the impact of nest site fidelity on the spatial distribution of occupied nest sites. We compare these models on the basis of the similarity between the observed and simulated spatial patterns (see 'Exploring emergent spatial patterning') and select the partial nest site fidelity model as most producing spatial pattern most similar to that observed at Beagle Island. In contrast to the selected sub-model, total nest site fidelity resulted in populations that were under-aggregated relative to observed patterns. Additionally, this parameterization lead to unstable populations in which individuals remained in unsuitable nest locations and repeatedly failed to reproduce. The sub-model with no nest site fidelity lead to populations that were over-aggregated compared to observed patterns, with individuals nesting in large contiguous units.

Nest Site Selection

Nest site selection is based on the observed preferences of individuals in the Beagle Island colony, using the habitat suitability and strength of conspecific interaction estimated by

our auto-logistic model (Fig 3-2). In versions of the model parameterized with partial nest site fidelity, each individual that has decided to move is selects the best available nest site within a set search radius. If there are no unoccupied nests within this radius, the search radius is extended to double the original radius, which corresponds to an equivalent search operating from the exterior of the currently occupied sub-colony. For individuals entering the breeding population for the first time, we use the natal site as the origin for their search. The model assuming no site fidelity, by contrast, assumes individuals sequentially select nest sites from all available sites, with no preference for a previously occupied location.

Recruitment

While *Pygoscelis* penguins are capable of fledging two chicks per nest in a given year, as we model only the half the population, considering nests rather than both parents, recruitment is modelled as the result of a Bernoulli trial in which the probability of a nest being successful is a function of the age of the parent, and the suitability of the nest location. We fit a logistic regression to available data on recruitment for known aged individuals to generate the expected probability of success for individuals of a given age (from age 1 to 8+). The probability (F_a) of an individual of age 'a' successfully fledging a chick is calculated as

$$F_a = \frac{1}{1 + e^{-5.239+0.928a}}$$

The probability provided by this logistic regression are further modified by a factor that represent the biotic and abiotic influences resulting from nest site selection. We apply nest-site quality and neighborhood occupancy as multiplicative modifiers to the maximum reproductive success by age that are described by the logistic regression based on observed data. Additionally, we added a logistic density dependent term to recruitment that regulates population size. Several

authors (Ashmole 1963, Coulson 1985, Ainley et al. 1995) suggests a relationship between foraging range and maximum colony size in a range of seabird species, suggesting that prey-limitation for these central place foragers is an important aspect in the regulation of colony size. We apply density dependence to reproductive success rather than survival, as outside of the breeding season *Pygoscelis* penguins cease to operate as central-place foragers and may be widely dispersed, preventing local prey depletion.

The probability of the i_{th} individual successfully contributing a new individual to the population is therefore

$$p_i = \frac{k - N_t}{k} F_a q_i$$

Where k is a parameter controlling density dependence, N_t is the size of the population at timestep t, F_a is the age-specific reproductive probability for the i_{th} individual, and q_i is the suitability of the nesting location of the i_{th} individual. We estimated a value for k that provides reasonably sized populations, while limiting the population to sizes that could be simulated on a reasonable timescale.

While many of the parameters in the model are based either on values reported in the literature or on the parameters estimated in the auto-logistic model, we have no direct measure of the behavior used in searching for a suitable nest site. We use a grid search to find values for these parameters that best match the empirical spatial pattern, resulting in an initial search kernel of 2 nest units. Changing these parameter combinations in the model alter the results quantitatively, but not qualitatively.

Exploring emergent spatial patterning

To assess the ability of our model to replicate the spatial population dynamics of real penguin colonies, we use the pair correlation function (pcf), which measures the average number of nests *at* a distance *r* (in contrast to the commonly used Ripley's K function, which counts points *within* a range *r*) from a focal nest. The pcf is typically calculated for a range of r values, capturing spatial pattern operating at different spatial scales. The function pfc(r) can be compared against its equivalent under any given null model, allowing the properties of spatial distributions to be compared. In our discretized hexagonal lattice, nest sites adjacent to a focal nest can be visualized as a series of expanding 'rings' and r indexed in units of inter-nest distances. We compare the observed pcf(r) from the IBM to the observed spatial pattern of the colony at Beagle Island (Fig. 4-2) using root mean squared error (RMSE) where RMSE is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_{ir}^2}$$

where e_{ir} is the error between the observed and simulated values of pcf(r) in the i^{th} simulation, across all n simulations of a given model parameterization.

Our IBMs produce spatial patterns that are, on average, more densely aggregated than the observed colony. However, the degree of clustering is highly variable over time, and the pcf(r) values calculated for the real colony are within the 95th percentile envelopes of the simulated results. By exploring the correlations between the pcf(1) and growth rate integrated over a range of timespans, we observe that the spatial pattern in this model is best predicted by the growth rate over the previous 12 timesteps (Fig 4-2-C). Linear regression of the error in PCF at each time step of the simulation against the growth rate over the last 12 timesteps indicates that our

simulations are most similar to the observed pattern in periods following population decline. Regression of the bias in the error confirms that following increases in population size, the simulated patterns are more highly clustered than in the observed patterns. Regression of the pair correlation function at radius 1 (Fig 4-2-D) against the proportional change in population size over the previous 12 timesteps indicates that declines lead to a fragmented colony, with individuals having fewer immediate neighbors on average. Across simulations, declines in abundance lead to fragmentation exposing individuals to increased edge effects, while increases fill space and result in larger, more contiguous sub-colony units with fewer individuals on the peripheries of colonies.

Landscape Structure

The landscape on which we simulate populations is itself highly spatially structured. To understand how the landscape affects rates of fragmentation of sub-colonies, we ran additional simulations in which the landscape is spatially homogenous with suitability set to the median value of the observed landscape suitability's at Beagle Island. Spatially heterogeneous landscape features provide a template for the fragmentation of colonies, which typically fracture along contours of low habitat suitability. As a result, high quality sites are occupied for a larger proportion of the time (Fig 4-3). However, simulations in which the landscape is homogeneous (no terrain) illustrate that the template provided by landscape terrain is not required for fragmentation to occur, and populations under homogeneous conditions also fragment under decline due to strong conspecific interactions coupled with nest-site fidelity.

Perturbation

To understand the effects of spatial pattern on the stability of the population, we examined the effects of spatial configuration independent of population size. We selected multiple time points in the best fitting simulation at which the population size was identical (horizontal line in Fig. 4-4), but the spatial configuration was different. We use the state of the simulation at these points as the initial condition for an additional simulation. For each of these simulations we apply negative perturbations to the population by reducing the growth rate for five consecutive time steps, achieving an approximately 25% reduction in population, and record the time taken for the population to recover to its initial level. We found that the time taken for a population to recover to its initial size is related to the spatial configuration of the population (Fig. 4-4), consistent with the hypothesis that more fragmented populations are less resilient to perturbation. Slowed responses to perturbation in systems may also be an indicator that these systems are closer to a critical tipping point (Scheffer et al. 2012).

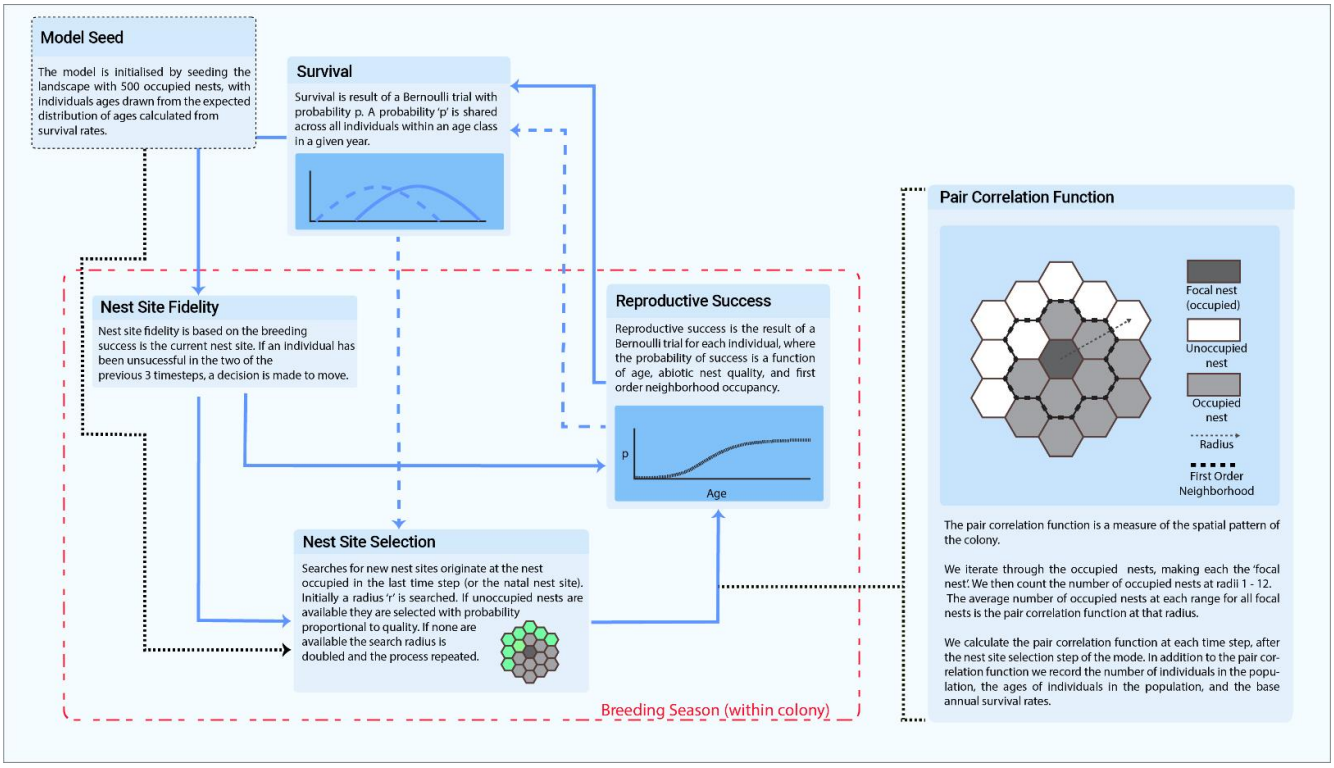

Figure 4-1. Schematic representation of the individual-based model of Adélie coloniality. Decisions about nesting location, and subsequent reproductive success are spatially explicit and occur within the breeding season (bounded by red dashed line), while survival occurs outside the breeding colony in the period between annual breeding.

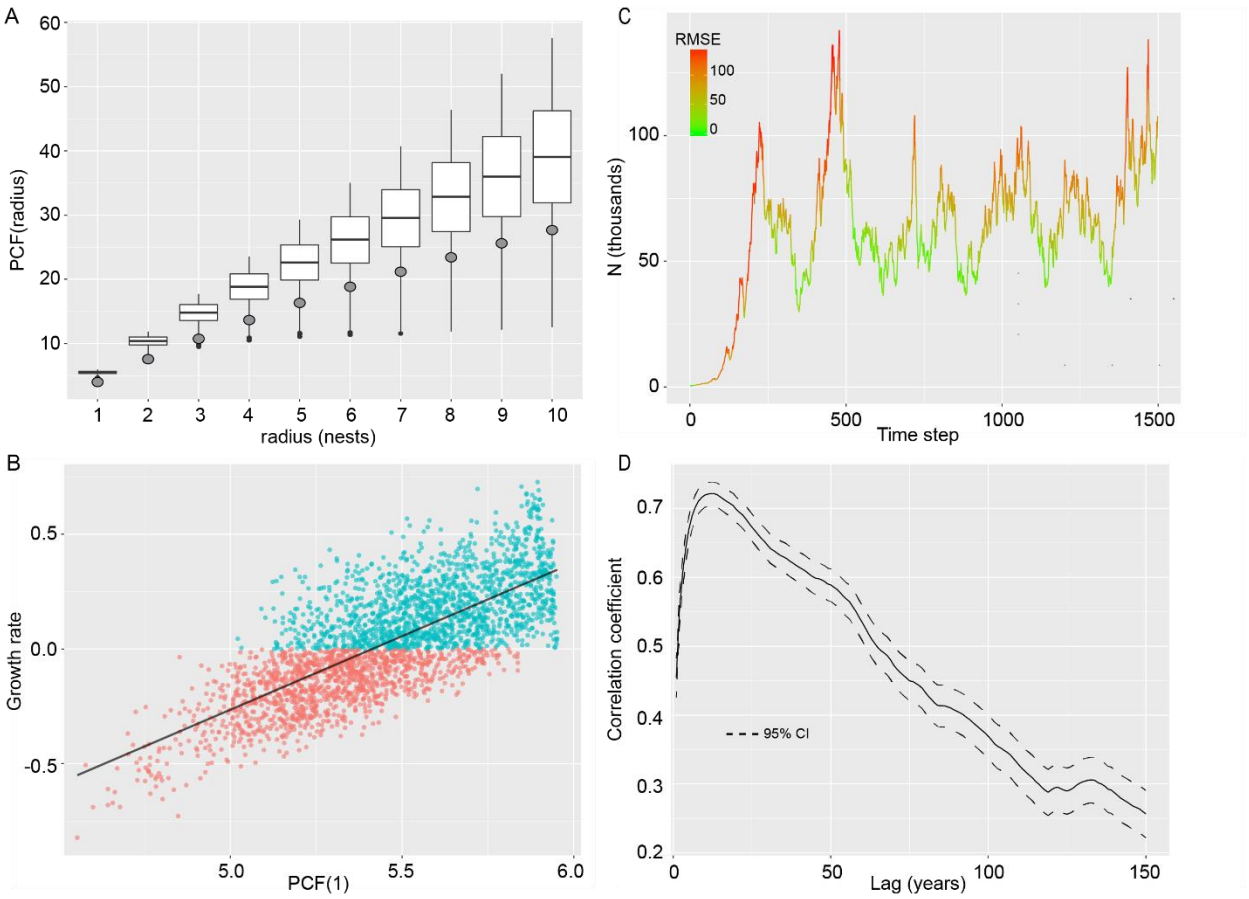

Figure 4-2. Population level trends and spatial pattern captured by our individual-based model. A) Pair correlation function (pcf) at radii 1-10 for the simulations (boxplots) and the pcf observed at the Beagle Island colony (gray points). B) Effects of growth rate (integrated over 12 timesteps) on the pcf for simulated results. Declines in population size result in a reduced value of pcf(radius =1), indicating reduced aggregation following decline. C) An example trajectory from a single run of the model. Color indicates the root mean squared error between the simulated pcf and the observed at Beagle Island, indicating that our model produces results most like Beagle Island following periods of decline. D) Correlation between population growth rate over lagged periods (x-axis) indicating that spatial pattern is best predicted by the change in population size over 12 timesteps.

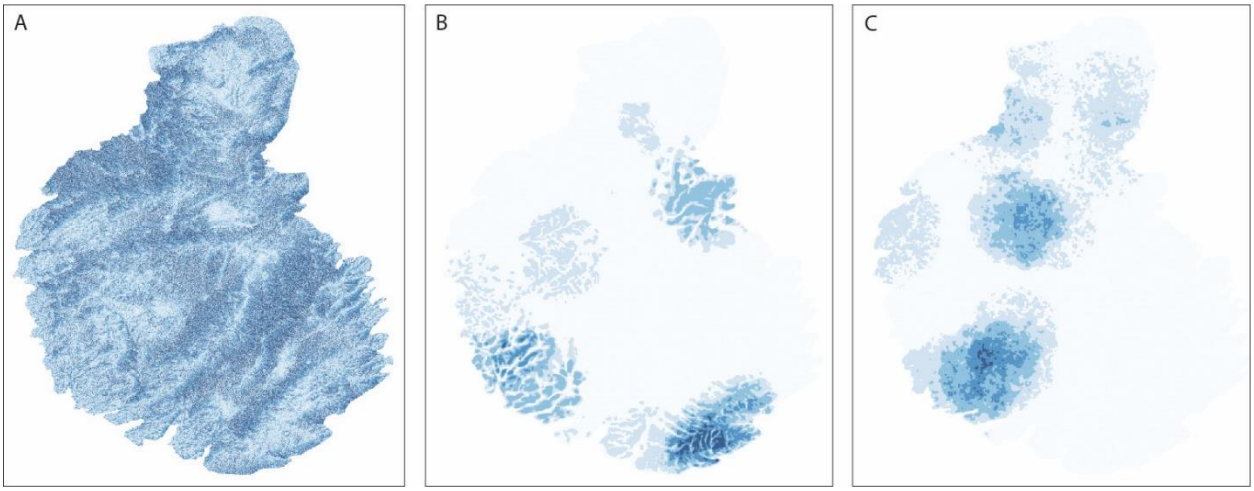

Figure 4-3. Effects of topography on the fragmentation of colonies A) Habitat suitability for nesting Adélie penguins at Beagle Island in the Danger Islands group. B) Heatmap of occupancy of nest locations (darker blue nests occupied more frequently), demonstrate the role terrain plays as a template for fragmentation of colonies. C) Heatmap of occupancy of nest locations with preference for terrain removed from the model. While colonies still fragment, the spatial patterning is independent of location.

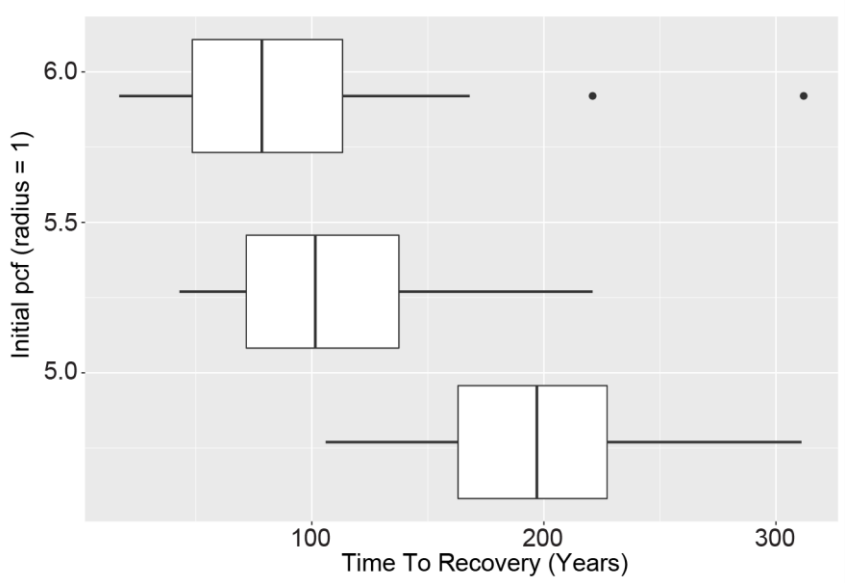

Figure 4-4. Distribution of time taken for populations to recover to initial starting size after perturbation, given differing initial spatial configurations. Vertical lines represent the median time to recovery, while boxes represent the interquartile range and horizontal lines represent the upper and lower quartiles. Compact, highly clustered configurations characterized by higher initial pair correlation functions (top) recover twice as quickly as those that are in a fragmented state before perturbation (lower pair correlation function -bottom).

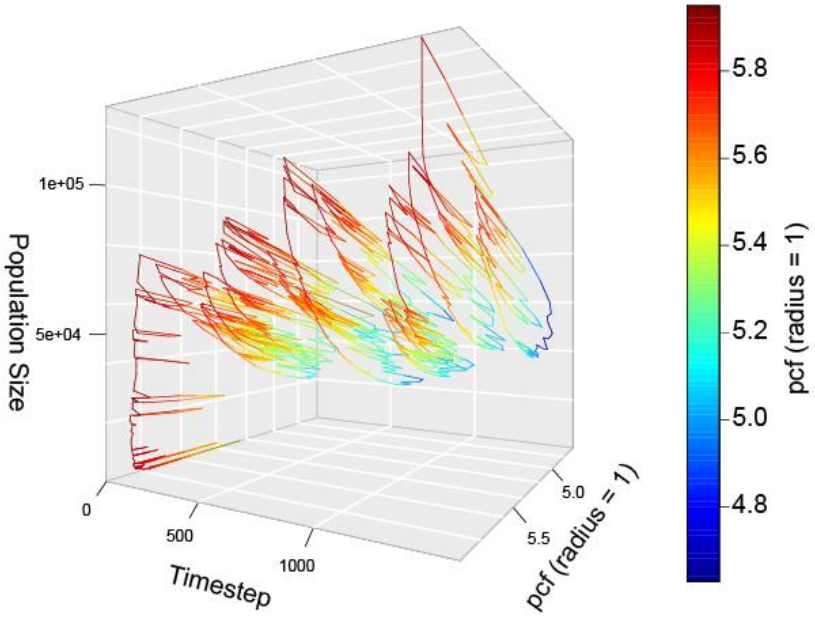

Figure 4-5. Trajectory of a simulated population (Population size axis) over time (Timestep axis) illustrating the additional dimension (pcf-axis) introduced by spatial configuration (pcf(1)) in a population in which spatial configuration affects population vital rates.

Discussion

Our simulated populations demonstrate that with only a simple set of assumptions about the processes occurring at the individual level, we observe both complex pattern formation, and strong effects of spatial pattern on population level dynamics. Nest site fidelity and conspecific attraction appear to drive fragmentation in sub-colonies as populations decrease (Fig 4-2). This fragmentation exposes more individuals to exterior sites and this in turn reduces the average reproductive success of the population. The result is a feedback loop that may drive populations to decline when fragmentation (likely to accumulate over the history of the colony) is high enough. Experiments with perturbations to this system illustrate that recovery may be impeded by the spatial configuration of the colony, and that sub-optimal arrangements of individuals in the colony may limit the system's ability to respond to conditions causing reduced reproductive success or high mortality. Additionally, when the system has degraded to a sub-optimal arrangement, there is no clear mechanism for sub-colonies to recombine into contiguous units without the population undergoing a growth phase. As such, abrupt declines followed by periods of stability may result in the persistence of sub-optimal arrangements, resulting in a colony that is more prone to decline. While these results indicate that the colony observed at Beagle Island is likely to have undergone periods of decline in the past, review of historic records indicates that this population may have been relatively stable for the past decades. Where we observe fragmented patterns, we may be observing the result of declines at any point in the colonies history. Due to the scarcity of new colonization events, observation of a colonies initial phases of growth may be impossible.

While fragmentation, or splitting, has been demonstrated in self-aggregating systems in which repulsion between individuals is balanced by attraction at other spatial scales (Chen et al.

2012), our IBM suggests that in a dynamic system such repulsive forces are not necessary for fragmentation. While interactions in real colonies may include intermediate or long range repulsive forces and are likely more complex than in our simulation, the simple mechanisms we explore are well supported by biological studies on colonial seabirds and sufficient for pattern formation. Conspecific attraction, often assumed to drive aggregative feedback loops, may also be instrumental in fragmentation when the ability to reorganize is limited by nest site fidelity.

Models of population dynamics often treat a population as a single well-mixed group whose dynamics are memoryless Markov processes and thus do not display long-range 'memory' of past dynamics. Our results demonstrate that when interactions between individuals or between individuals and their environment are spatially structured, the configuration of individuals within the environment represents an additional system variable. Routes through this higher dimensional space may be equally deterministic, but when projected to a lower dimensional space in a non-spatial setting, we may observe hysteretic effects. In systems with a strong coupling between individual level spatial configuration and population level dynamics, it may be possible to make inferences about the state of one level of the organization from the knowledge of state of the other level (DeAngelis and Grimm 2014). Understanding the temporal dynamics of populations requires long-term datasets that can be expensive and time consuming to collect. If the data are intended to be used to make management decisions based on the trends of some population, by the time required to collect a time series of sufficient length is often too long to be of practical benefit. The feedback loops between population size and spatial configuration maybe particularly important when environmental variability is high, as the stochastic degradation of spatial optimality may result in unstable populations which, when assessed via population size, appear to be stable.

If spatial configuration can be used as an indicator of population trend, we may be able to use it to rapidly identify populations that may be undergoing change rather than investing in futile datasets. While the drone based imagery used in this model to capture spatial pattern that was used in this model requires visiting the colony, it is possible to capture the shape of the colony using high resolution satellite imagery in which the outline of the colony is visible due to guano deposition (LaRue et al. 2014). This may allow us to calculate additional metrics of stability for the system to improve our predictions of past and projected future population states. While our IBMs illustrate that simple, individual level mechanisms may be responsible for macro-scale pattern formation, they also suggest that our current predictive ability in these systems may be low. The sheer number of possible spatial configurations and the number of accumulating individual level decisions makes predicting spatial pattern difficult. The role of spatial configuration may complicate our ability to forecast future abundance.

An interesting aspect of our results is the distribution of birds by age across the colony. Even with no explicit competitive advantage for older breeders, we see that older birds are more likely to be found within the interior of the colony with the younger birds on the outside. This 'central-periphery' distribution, either in terms of quality or age (which is often a proxy for quality) has been reported in several studies (Minias 2014) and interpreted as the result of competitive advantage. Our model suggests that this age structure within the colony might be an emergent property of individual level choices about nest choice in the face of nest site fidelity and the desire for complete neighborhoods.

Finally, it is worth noting that while terrain clearly influences the probability of occupancy for each potential nesting site, long-term occupancy changes habitat suitability. *Pygoscelis* penguins modify their environment through the accumulation of nesting materials,

and through the stabilization of soil through excretion of guano. This biogeomorphological feedback loop drives auto-correlation in occupancy as individuals improve the quality of their nest site, encouraging return to the same nest site, as well as encouraging others to occupy the same location in future generations. This feedback loop amplifies the importance of the history of a colony on its spatial configuration and potential future growth rates.

In our model, the formation of new sub-colonies is driven by declines in population, however this mechanism does not explain how brand-new colonies (e.g., on a previously unoccupied island) arise. The process of colonization, especially of sites not occupied by other penguin species, is unknown because it is virtually never observed, though we know that over geological time, Adélie penguins have shifted their ranges and thus must be capable of de novo colonization of new habitat. Regardless, the barriers to colonization are clearly related to the Allee effects associated with predation, and our model may shed some new light on the processes underlying these important but very rare events.

We have demonstrated that relatively simple assumptions about individual level behavioral choices may lead to emergent properties at the population level, leading to complex population trajectories that may not be adequately captured by point models of populations. Due to linkages between spatially structured habitat and individual behavior, stochastic processes that may impact the trajectory of a population over many subsequent timesteps, and the difficulty in capturing all the behavioral choices made and the distribution of behaviors amongst individuals, quantitative prediction from this type of model is likely to be difficult. However, we have demonstrated that such simple assumptions are able to reproduce the patterns seen in real populations (Grimm and Railsback 2005), and that spatial configuration is indicative of temporal dynamics and may provide an early warning signal for the health of the population. Given these

likely linkages between population level process and spatial pattern, attention should be paid to the spatial configuration of colonies under existing monitoring programs. The adoption of UAV census methodologies which provide the spatial resolution to resolve individual birds (Borowicz et al. 2018) should be a priority as, in addition to improving the efficiency and potential accuracy of abundance estimates, the information provided on the locations of nests may be invaluable for testing hypotheses about decline driven fragmentation, and for identifying those colonies that are at greatest risk of decline. While the application of this technology is in its infancy, and the survey this work is based on (Borowicz et al. 2018) represents a single snapshot of colony spatial patterning, with concerted effort to produce annual time-series of such high-spatial resolution our potential ability to understand the links between process and pattern should be greatly enhanced.

While the data produced through UAV based monitoring is exceptional, it's adoption as a continent wide monitoring tool is impractical, however understanding the links between process and spatial pattern provides two major benefits. The encoding of process into spatial pattern should allow us to use opportunistic and temporally patchy surveys to identify those colonies undergoing change in the absence of long term time-series. Even in the absence of the highest spatial resolution data, it may be informative to examine the shapes of colonies at lower spatial resolution, substituting the pair correlation function for courser scale spatial metrics. LaRue et al. (2014) demonstrated that satellite based remote sensing provided sufficient resolution to identify the areas occupied by *Pygoscelis* nests through the spectral signature of the guano stain that accumulates during the breeding season. While this approach does not allow us to resolve individual nests, this type of data set may provide informative metrics such as the ratio of area to perimeter of the guano stain, or the connectivity of sub-colony units, both of which are likely to decline in response to declining population, and subsequently, may be indicators that the

populations is at risk of further decline. While these metrics may respond on a slower time scale than individual level features, remotely sensed data may allow for better temporal resolution and coverage of a wider range of colonies, allowing us to augment continent wide monitoring of nesting populations with additional information on the spatial configuration of colonies, and to potentially fit statistical models linking these augmented datasets with population level trends.

Chapter 5 - Combining UAS based Spectrometry with 3D model reconstruction[8]

Abstract

The rapidly decreasing cost and increasing capabilities of unmanned aerial systems (UASs), in combination with the latest generations of miniaturized sensor packages, provides the potential for a new scale of remote sensing in ecology. Through the application of modern computer vision techniques, UASs can be used to produce high resolution topographic data from set of overlapping aerial imagery. When combined with additional sensors, this presents a novel opportunity for exploring the intersection of structure and function. There has been considerable interest in UAS based spectral readings in precision agriculture, but the methods for spatial registration of this data, particularly in the case of non-imaging sensors, are not well developed. We present a method for integrating high spectral resolution data with high spatial resolution three-dimensional topographic data, that allows for variable geometries of sensor layouts, correction for instability in the UAS platform, and for collection and processing of spectral data in off-nadir orientations.

[8] Author Contributions: (PM – Philip McDowall, HJL – Heather Lynch): Conceptualization: PM. Data curation: PM. Formal analysis: PM. Investigation: PM. Methodology: PM. Supervision: HJL. Validation: PM. Visualization: PM. Writing – original draft: PM. Writing – review & editing: PM HJL.

Thanks to Shawn Serbin, Meng Ran and Andrew McMahon at Brookhaven National Lab for providing the data used in this analysis.

Funding provided by the National Science Foundation Office of Polar Programs (NSF/OPP-1255058) and the Institute of Advanced Computational Science at Stony Brook University.

Introduction

Recent advances in the capabilities of Unmanned Aerial Systems (UASs) have resulted in systems with increased capacity to carry weight, increased ease of use, and rapidly lowering costs, making them an increasing useful tool for ecologists. One of the primary functions of these systems has been in census work, utilizing the high-resolution cameras that can be carried to capture aerial imagery in a rapid, and cost-effective manner (Borowicz et al. 2018). In addition to utilizing the imagery collected by these systems directly, a computer vision algorithm, Structure-from-Motion (SfM) can be used to reconstruct the 3D structure of the surveyed scene in very high (centimeter scale) resolution (McDowall and Lynch 2017). SfM algorithms are based on triangulating matched points between sets of overlapping images, allowing the estimation of three-dimensional structure from two-dimensional image sequences. This method differs from stereo-matching algorithms traditionally used to create digital elevation models (DEMs), and from more common photogrammetric methods, in that the camera position and scene geometry are solved simultaneously, negating the need for data on camera position, or three-dimensional control points. This is achieved using algorithms which track multiple correspondences between multiple overlapping, offset images, and solving a series of equation which link the position of points in images, through a model of the cameras optics, to the locations of those pixel in a three-dimensional coordinate system (Torr and Zisserman 1999). While the three-dimensional model that is created is unscaled and arbitrarily orientated, the addition of supplementary information, such as ground control points of known positions within the scene, or the locations of the cameras in a real-world coordinate system, allows us to reconstruct a scaled, georectified three-dimensional model of the scene. The resulting model is similar in nature and resolution to point

clouds created by LiDAR systems, however the cost of the cameras needed for SfM are often significantly lower (Fonstad et al. 2013). The information that can be extracted from such high resolution topographic surveys is only beginning to be explored, but this approach has already been applied to modelling the spatial distribution of a species within the landscape, and to quantifying features such as canopy cover and structure in vineyards and forests (Mathews and Jensen 2013, Wallace et al. 2016).

With the increasing payload capabilities of modern UASs, and the constantly decreasing size and weight of various sensor packages, the potential for creating high resolution general surveying tools is increasing. The combination of structural information provided by SfM and additional information may allow us to gain a better understanding of how structure influences function and may allow us to refine the resolution of data collected while greatly expanding spatial coverage. One of the sensor packages with the most potential in UAS based ecology is small form-factor spectrometers which are now available at weights less than 250g, far within the capabilities of many commercially available UAS systems (Burkart et al. 2014). There are a range of commercial available packages, including light-weight multi- and hyper-spectral cameras, as well as miniature non-imaging spectrometers, providing options that can cover a wide variety of wavelengths, from ultraviolet, through the visible spectrum, to near-infrared.

Spectral information can greatly enhance our understanding of the function of ecosystems through correlations between derived indices calculated from spectral data with known functions. Spectral data from satellite and manned aerial platforms has been applied extensively in ecology, with applications including estimating vegetation biomass, thematic classification of landcover (Cohen and Goward 2004), mapping the locations of invasive species (Ustin et al. 2002, Hestir et al. 2008, He et al. 2011) and determining species compositions (Martin et al. 1998).

Additionally, ground based spectrometers have been applied to problems such as measuring the levels of photosynthetic pigments in the intertidal zone (Carrère et al. 2004) A common use of spectral data is in the calculation of indices such as the Normalized Difference Vegetation Index (NDVI). NDVI uses the ratio of the difference of the recorded red and infrared radiances, over their sum, to give a single metric of spectral characteristic which normalizes the effect of solar zenith angle. This derived metric correlates strongly with net primary productivity and absorbed photosynthetically active radiation (Tucker 1979). There are many such vegetation indices, formed by ratios, differences, or linear combinations of bands in spectral data (Ye et al. 2008), all of which may correlate with some aspect of a vegetation's physical properties. In addition to information on the functional nature of the ecosystem, spectral information can be applied to thematic mapping, allowing the easier identifications of different types of ground cover, for instance through the detection of distinctive spectral patterns of vegetation or water.

While the spatial resolution of satellite-based sensors has been increasing, the readings available from these platforms are relatively course, with resolutions typically in the range of 10s of meters. Additionally, these platforms suffer from low temporal resolution, with the availability of data dependent on the orbit of the satellite. In the best case, data are available on a daily time scale with low spatial resolution. While manned aerial platforms can both increase the spatial resolution of data available, and allow data to be collected when required, these systems may be prohibitively expensive. As such, interest in UAS-based spectrometry is rapidly increasing, particularly in the field of precision agriculture (Link et al. 2013), and solutions exist to combine 3D scenes with readings created using multi- and hyper-spectral camera systems (Aasen et al. 2015). However, the use of imaging sensors often limits the available spectral resolution, with most systems offering 10’s or low hundreds of spectral bands, in comparison to

1000+ for a non-imaging system. This lack of spectral information may reduce the inferences that can be drawn from the resulting dataset.

Recent work has demonstrated the potential of mounting the latest generation of miniature, non-imaging spectrometers on UASs, providing increased spectral resolution at the cost of spatial resolution. The data recorded from these systems have been shown to have a high level of correlation with calibration readings taken on the ground. However, the systems currently employed lack the ability to combine 3D structure with this higher spectral resolution data. Additionally, previously presented methods have assumed that the readings collected by the spectral sensor are captured on-nadir, relying on the stabilization of the aerial platform by its internal navigational hardware and software, or the addition of a stabilizing gimbal. However, in forward flight the UAS platform is rarely level, and making assumptions about the orientation of the sensor may result in large errors in registration between the reading and the ground plane. Here we present a method to project the data collected from non-imaging spectrometers to a 3D model created through SfM, utilizing the refined estimates of location and orientation derived during processing in an attempt to mitigate against projection errors.

Methods

We used data collected by the Terrestrial Ecosystem Science & Technology group at Brookhaven National Laboratory. These data were collected using a CarbonCore Cortex UAS in an octocopter configuration, carrying a high-resolution digital camera and a light-weight OceanOptics USB2000+ spectrometer, both of which were controlled and synchronized by a Raspberry Pi connected to the UAS's flight controller. In addition to the spectrometer mounted

on the UAS, this system employs a second, upward facing, spectrometer which collects concurrent spectral readings allowing readings to be adjusted for solar irradiance. The downward facing spectrometer is mounted on the same gimbal system used to orient the camera, allowing us to use the camera position estimated through SfM to orient the spectrometer readings relative to the 3D reconstructed surface. We used an adjustable collimating lens, which allows the field of view to be set between 0° (total collimation) and 45°, providing adjustments for the optimal tradeoff between collection efficiency and spatial resolution. In the readings presented here, the collimating lens was set to a field of view of 14°.

While there are a range of software products available for SfM processing, both open source and proprietary, we used the commercial package, Agisoft Photoscan for the reconstruction step. Our workflow would be equally applicable to other reconstruction packages that allow the export of the three-dimensional model and the estimated camera locations. The initial result of processing using SfM algorithms is a three-dimensional point cloud. These point clouds may be very dense, with a point for each pixel that can be matched between at least images. When flight profiles are designed such that there is overlap between multiple images, this may result in a point cloud density that is greater than the pixel resolution of a single image. The potential density of this point cloud makes it computationally difficult to work with, however methods exist to create a 3D triangulated mesh from these point clouds, aiming to reconstruct the surface from measured points on the surface (Fabio 2001). We exported the results of surface reconstruction as a mesh, represented as a collection of point vertices in x,y,z space, which are grouped together to form the faces of a polygonal mesh, significantly reducing the computational burden for projection of spectrometer readings to the 3D scene.

While the UASs navigational systems include an inertial measurement unit (IMU) and GPS which give an estimate of the location and orientation of the platform throughout the flight, our main concern when combining results from the spectrometer and scene reconstruction is potential registration errors between the two data streams. To minimize these potential registration errors, we use a control system that synchronizes the capture of images for SfM, and the collection of spectrometer readings, and then use the location and orientations estimated as part of the SfM reconstruction process for registration of the spectrometer data. The imagery, 3D models, and projected spectral data are scaled and georectified to known ground control points, providing a final product in a real-world coordinate system.

We modeled the spectrometer reading as integrating surfaces which intersect a conical beam from the spectrometers fiber optic (Fig. 5-1), with angle equal to half the field of view of the fiber optic or lens. To project this beam onto the 3D model, we find the spatial intersection of this conical beam and the vertices of the surface model created through SfM[9]. This allows us to identify a set of vertices of the model which may have contributed to the recorded spectral reading. By testing for an intersection between a vector projected back from a vertex to the sensor location, and the 3D surface, we can remove vertices from this set that are occluded from the reading. We attached the corresponding spectra to each of the selected vertices, and additionally attached the distance between the vertex and the sensor.

The radius of an on-nadir reading over a planar surface is $h\tan(x/2)$, while the area covered is given by $\pi\left(h\tan\left(\frac{x}{2}\right)\right)^2$, where x is the field of view of the spectrometer, and h is the height above the surface. For a typical flight altitude of 30m and field of view of 14°, this results

[9] Implementation available at https://github.com/phil-mcdowall

in a circular reading with radius 3.6 m, with an area of 42 m^2. The strength of our method of projecting is that it allows us to deal with readings taken off nadir and over non-planar surfaces, allowing us to take readings of different aspects of the three-dimensional scene. While we utilize a setup with the spectrometer mounted in the same orientation as the camera, our method allows for an offset between the position of the camera and the spectrometer reading, permitting the system to be calibrated for different configurations. In addition to allowing different configurations of mountings on the UAS, we can modify the beam radius to specifications of other systems by altering the radius of the conical beam.

The resulting dataset is a three-dimensional mesh in which those vertices which were within the projected cone from the spectrometer reading contain corresponding spectral information. The vertex data may be projected to a planar form using an orthographic projection, providing data in a similar form to other applications of UAS based spectrometry, and allowing the data to be integrated in to more traditional remote sensing pipelines. We can then apply any existing corrections or spectral processing algorithms on this dataset, providing any of the derived metrics and indices available from other spectral datasets.

Results

In some cases, the assumption that readings are collected on nadir is reasonable. However, in the majority of readings the footprint calculated using the orientations estimated through SfM results in a significant difference between the two estimation methods (Fig. 5-2). For a collection of 346 readings taken over two flights, we found that the median difference between the centroids of the footprint estimated on-nadir and the footprint corrected for

orientation was 5.2 m, with a standard deviation of 2.8 m. The largest difference between two corresponding centroids was 10.86 m, or 124% of the radius of the on-nadir footprint. At a flight height of ~30 m, this resulted in no area of overlap between the two estimated footprints. This level of error can occur with relatively small changes in orientation of the UASs gimbal, with a difference of 7° being sufficient to result in no overlap. The 346 estimated footprints have an average area of 34 m^2 (standard deviation 4.5 m^2), which is slightly larger than the 32m^2 estimated using the assumption of on-nadir readings. The geometry of the projection means that off-nadir readings will always have a larger footprint than those collected on-nadir (except in cases where there is occlusion), and as such readings should be taken orthogonal to the surface of interest to maximize spatial resolution.

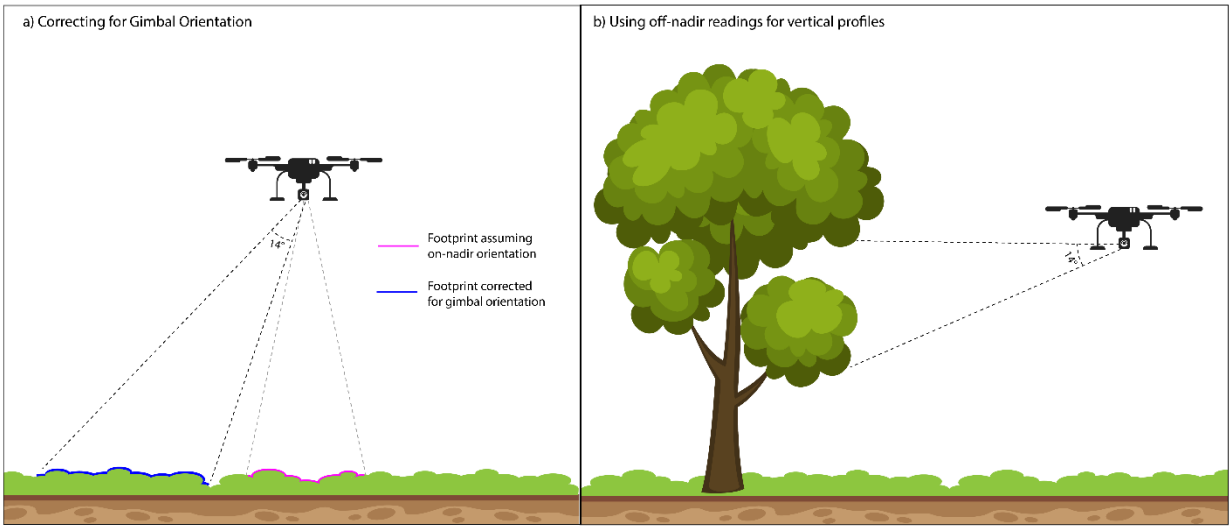

Figure 5-1. Schematic of UAS as a personal scale remote sensing platform. a) Incorrect assumptions about the orientation of sensors may result in significant spatial registration errors, and attributing readings to surfaces which do not overlap the surface being read. b) The ability to use sensors in off-nadir orientations allows for novel survey plans, such as vertical profiles.

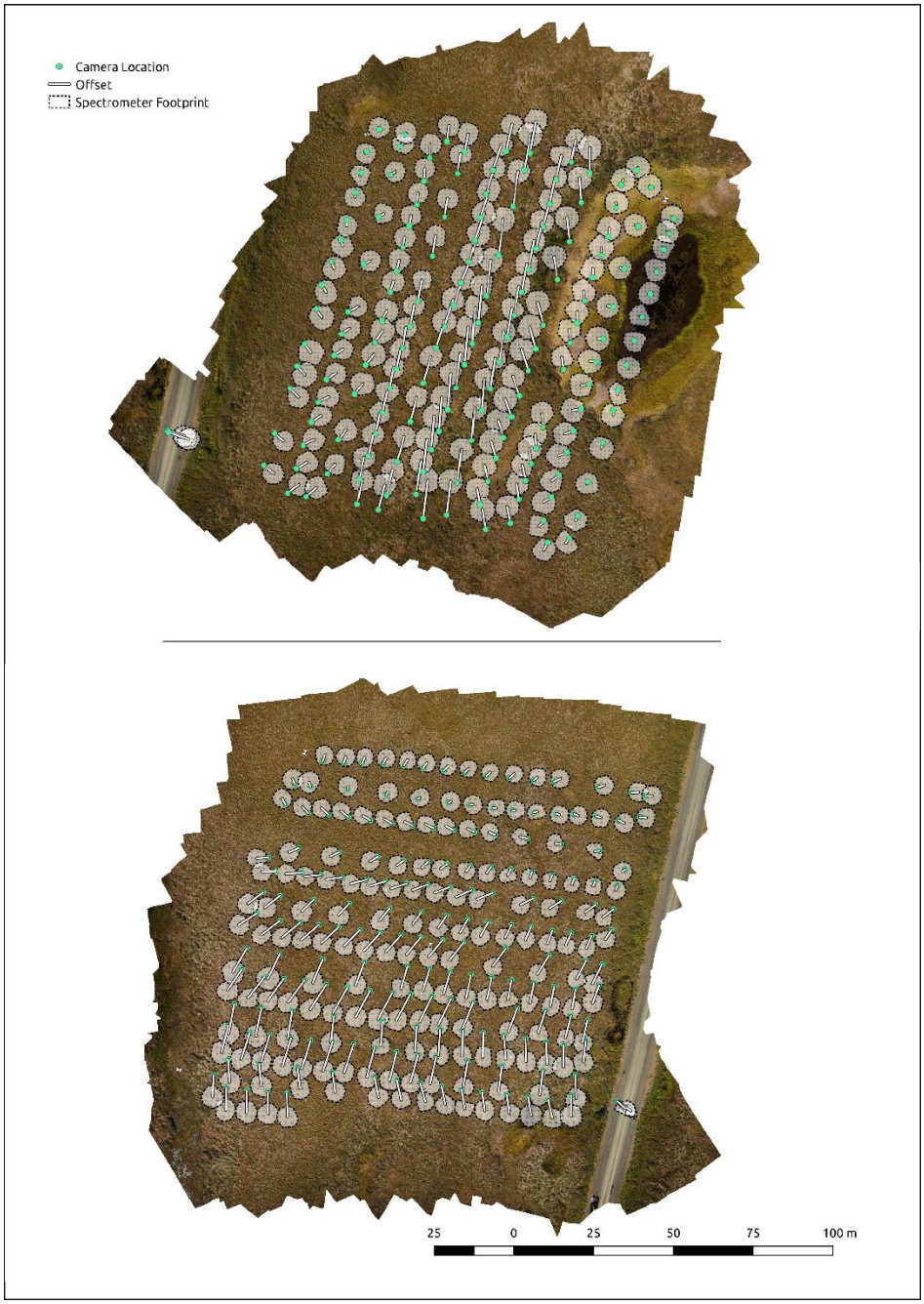

Figure 5-2. Orthophotos created from 346 UAS images collected during two flights by the Terrestrial Ecosystem Science & Technology group at Brookhaven National Laboratory. The location of the UAS platform at the time of each measurement is shown as a green dot. Under the assumption that spectrometer measurements are made on-nadir, the footprint of these readings would be centered on the UAS location. White circles represent the footprints calculated through the orientations estimated during 3D reconstruction, while the white lines indicate the difference between the centroids of the two footprints.

Conclusions

One of the primary issues with combining remotely sensed spectral data and measurements of ecological function collected in the field is that there tends to be a mismatch in the scale of these datasets, with remote sensed data integrating readings across pixels covering >100 m^2, while measurements of ecological function taken in the field may be point measurements. When there is a high degree of homogeneity in the spatial distribution of ecological function, this scale mismatch may be significant and individual pixels in the remotely sensed dataset may represent the contribution of multiple sub-pixel scale objects. While the effects of spectral-mixing can be modelled, and attempts made to separate out the contributing spectra, increasing the spatial resolution of data may be preferable. In addition to increasing resolution, UASs allow for much greater temporal resolution and fine control over the surveying period. This is important both for survey design, and for calibration of the remotely sensed data, as the use of UASs allows for survey data and calibration data to be collected in the same temporal reference, reducing potential mismatches.

By bringing this technology to a small scale that provides high spatial resolution we can begin to better understand the distribution of function within the ecosystem. However, as we have demonstrated, operating at these small scales brings additional technical challenges. While traditional aerial and satellite platforms carry high quality IMUs and have higher inertia and stability due to their mass, UAS platforms are less stable and the sensors that position the gimbal used to stabilize gimbals may drift during operation, leading to readings being unintentionally collected off-nadir. By utilizing the gimbal orientations estimated during SfM we can attempt to correct for this drift and can also plan data collection missions in which readings are intentionally

taken off-nadir, for instance allowing vertical profiles of vegetation to be collected. It is important to note that the orientation of the gimbal between readings is non-independent, with a high degree of temporal (and therefore spatial) auto-correlation in the angle off-nadir. These errors tend to accumulate over the course of a flight (gimbal drift), and as such, under the assumption that readings are taken on-nadir, the resultant errors will be spatially heterogeneous.

Our method for projecting spectral data represents an improvement over the assumptions previously made with regards to the orientation of the sensor, however further work is required in modelling the contribution of different vertices to the spectral reading. Our method should provide sufficient information for the contribution of vertices to be modelled, as it provides the distance to the sensor, the angle between the sensor and the normal of the recorded vertex, and the density of vertices in each reading. One of the benefits of this approach to projecting spectral data is that we can calculate the slope and aspect of the points that contribute to the integrated reading. This information could then be used, in combination with supplementary information on the orientation of illumination at the time of data collection, to apply topographic corrections to the spectral information. Topography influences the local illumination of a surface, and the radiance of the surface may vary if oriented towards or away from the incident sunlight.

In addition to 3D structure, the visual information present in the original images, and the spectrometer data, modern UASs may be equipped with additional sensor packages such as thermal imaging cameras. While some of these datasets may be integrated through existing methods such as directly through the SfM pipeline, or through registration in a post-processing step such as feature matching (Jakob et al. 2017), the method we have presented here could be generalized to alternative sensors in the future, potentially allowing for novel classification of ecosystem function at very small spatial scales.

Bibliography

Aasen, H., A. Burkart, A. Bolten, and G. Bareth. 2015. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. ISPRS Journal of Photogrammetry and Remote Sensing **108**:245-259.

Ainley, D. 2002. The Adélie penguin: bellwether of climate change. Columbia University Press.

Ainley, D. G., G. Ballard, B. J. Karl, and K. M. Dugger. 2005. Leopard seal predation rates at penguin colonies of different size. Antarctic Science **17**:335-340.

Ainley, D. G., N. Nur, and E. J. Woehler. 1995. Factors affecting the distribution and size of pygoscelid penguin colonies in the Antarctic. The Auk:171-182.

Anderson, D. J., and P. J. Hodum. 1993. Predator behavior favors clumped nesting in an oceanic seabird. Ecology **74**:2462-2464.

Ashmole, N. P. 1963. The regulation of numbers of tropical oceanic birds. Ibis **103**:458-473.

Baddeley, A., P. J. Diggle, A. Hardegen, T. Lawrence, R. K. Milne, and G. Nair. 2014. On tests of spatial pattern based on simulation envelopes. Ecological Monographs **84**:477-489.

Baddeley, A., E. Rubak, and R. Turner. 2015. Spatial point patterns: methodology and applications with R. CRC Press.

Baddeley, A., and R. Turner. 2005. Spatstat: an R package for analyzing spatial point patterns. Journal of statistical software **12**:1-42.

Baddeley, A. J., J. Møller, and R. Waagepetersen. 2000. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica **54**:329-350.

Barbosa, A., J. Moreno, J. Potti, and S. Merino. 1997. Breeding group size, nest position and breeding success in the chinstrap penguin. Polar Biology **18**:410-414.

Barry, S., and J. Elith. 2006. Error and uncertainty in habitat models. Journal of Applied Ecology **43**:413-423.

Bayer, R. D. 1982. How important are bird colonies as information centers? The Auk:31-40.

Ben-Jacob, E., O. Schochet, A. Tenenbaum, I. Cohen, A. Czirok, and T. Vicsek. 1994. Generic modelling of cooperative growth patterns in bacterial colonies. Nature **368**:46.

Bennie, J., B. Huntley, A. Wiltshire, M. O. Hill, and R. Baxter. 2008. Slope, aspect and climate: spatially explicit and implicit models of topographic microclimate in chalk grassland. Ecological modelling **216**:47-59.

Besag, J. E. 1972. Nearest-neighbour systems and the auto-logistic model for binary data. Journal of the Royal Statistical Society. Series B (Methodological):75-83.

Bivand, R. S., E. J. Pebesma, V. Gómez-Rubio, and E. J. Pebesma. 2008. Applied spatial data analysis with R. Springer.

Borowicz, A., P. McDowall, C. Youngflesh, T. Sayre-McCord, G. Clucas, R. Herman, S. Forrest, M. Rider, M. Schwaller, and T. Hart. 2018. Multi-modal survey of Adélie penguin mega-colonies reveals the Danger Islands as a seabird hotspot. Scientific reports **8**:3926.

Brown, C. R., and M. B. Brown. 2004. Group size and ectoparasitism affect daily survival probability in a colonial bird. Behavioral Ecology and Sociobiology **56**:498-511.

Brunton, D. H. 1997. Impacts of predators: center nests are less successful than edge nests in a large nesting colony of Least Terns. Condor:372-380.

Burkart, A., S. Cogliati, A. Schickling, and U. Rascher. 2014. A novel UAV-based ultra-light weight spectrometer for field spectroscopy. IEEE sensors journal **14**:62-67.

Camazine, S. 2003. Self-organization in biological systems. Princeton University Press.

Campomizzi, A. J., J. A. Butcher, S. L. Farrell, A. G. Snelgrove, B. A. Collier, K. J. Gutzwiller, M. L. Morrison, and R. N. Wilkins. 2008. Conspecific attraction is a missing component in wildlife habitat modeling. Journal of Wildlife Management **72**:331-336.

Carrascal, L., J. Moreno, and J. Amat. 1995. Nest maintenance and stone theft in the Chinstrap penguin (Pygoscelis antarctica). Polar Biology **15**:541-545.

Carrère, V., N. Spilmont, and D. Davoult. 2004. Comparison of simple techniques for estimating chlorophyll a concentration in the intertidal zone using high spectral-resolution field-spectrometer data. Marine Ecology Progress Series **274**:31-40.

Chen, Z., H. Liao, and T. Chu. 2012. Aggregation and splitting in self-driven swarms. Physica A: Statistical Mechanics and its Applications **391**:3988-3994.

Chowdhury, D., K. Nishinari, and A. Schadschneider. 2004. Self-organized patterns and traffic flow in Colonies of organisms: from bacteria and social insects to vertebrates. Phase Transitions **77**:601-624.

Cimino, M. A., W. R. Fraser, A. J. Irwin, and M. J. Oliver. 2013. Satellite data identify decadal trends in the quality of Pygoscelis penguin chick-rearing habitat. Global change biology **19**:136-148.

Cohen, W. B., and S. N. Goward. 2004. Landsat's Role in Ecological Applications of Remote Sensing. BioScience **54**:535-545.

Conradt, L., J. Krause, I. D. Couzin, and T. J. Roper. 2009. “Leading according to need” in self-organizing groups. The American Naturalist **173**:304-312.

Coulson, J. 1968. Differences in the quality of birds nesting in the centre and on the edges of a colony. Nature **217**:478-479.

Coulson, J. 1985. Density regulation in colonial seabird colonies. Proceedings of the XVIII International Ornithological Congress, Moscow.:783-791.

Couzin, I. D., and J. Krause. 2003. Self-organization and collective behavior in vertebrates. Advances in the Study of Behavior **32**:1-75.

Cressie, N. A. 1993. Statistics for spatial data. Wiley Online Library.

Dakos, V., S. Kéfi, M. Rietkerk, E. H. Van Nes, and M. Scheffer. 2011. Slowing down in spatially patterned ecosystems at the brink of collapse. The American Naturalist **177**:E153-E166.

Dandois, J. P., and E. C. Ellis. 2010. Remote sensing of vegetation structure using computer vision. Remote Sensing **2**:1157-1176.

Dandois, J. P., and E. C. Ellis. 2013. High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. Remote sensing of Environment **136**:259-276.

Davis, L. S., and J. T. Darby. 2012. Penguin biology. Elsevier.

DeAngelis, D. L., and V. Grimm. 2014. Individual-based models in ecology after four decades. F1000prime reports **6**.

Diggle, P. 1983. Statistical Analysis of Spatial Point Patterns. Academic Press.

Doligez, B., E. Danchin, and J. Clobert. 2002. Public information and breeding habitat selection in a wild bird population. Science **297**:1168-1170.

Eklund, C. 1964. Population studies of Antarctic seals and birds. Biologie Antarctique,(R. Carrick, M. Holdgate and J. Prevost, eds.) Hermann, Paris:4l5-4l9.

Emlen, J. T. 1952. Flocking behavior in birds. The Auk **69**:160-170.

Emslie, S. D., N. Karnovsky, and W. Trivelpiece. 1995. Avian predation at penguin colonies on King George Island, Antarctica. The Wilson Bulletin:317-327.

Fabio, R. 2001. From point cloud to surface: the modeling and visualisation problem– International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol XXXIV-5/W10-2001.

Fargallo, J. A., A. De León, and J. Potti. 2001. Nest-maintenance effort and health status in chinstrap penguins, Pygoscelis antarctica: the functional significance of stone-provisioning behaviour. Behavioral Ecology and Sociobiology **50**:141-150.

Fonstad, M. A., J. T. Dietrich, B. C. Courville, J. L. Jensen, and P. E. Carbonneau. 2013. Topographic structure from motion: a new development in photogrammetric measurement. Earth Surface Processes and Landforms **38**:421-430.

Funwi-Gabga, N., and J. Mateu. 2012. Understanding the nesting spatial behaviour of gorillas in the Kagwene Sanctuary, Cameroon. Stochastic Environmental Research and Risk Assessment **26**:793-811.

Garabedian, J. E., R. J. McGaughey, S. E. Reutebuch, B. R. Parresol, J. C. Kilgo, C. E. Moorman, and M. N. Peterson. 2014. Quantitative analysis of woodpecker habitat using high-resolution airborne LiDAR estimates of forest structure and composition. Remote sensing of Environment **145**:68-80.

Green, S., A. Bevan, and M. Shapland. 2014. A comparative assessment of structure from motion methods for archaeological research. Journal of Archaeological Science **46**:173-181.

Grimm, V., and S. F. Railsback. 2005. Individual-based Modeling and Ecology:(Princeton Series in Theoretical and Computational Biology).

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology letters **8**:993-1009.

Haase, P. 1995. Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. Journal of vegetation science **6**:575-582.

Hahn, S., and H.-U. Peter. 2003. Feeding territoriality and the reproductive consequences in brown skuas Catharacta antarctica lonnbergi. Polar Biology **26**:552-559.

Hamilton, W. D. 1971. Geometry for the selfish herd. Journal of Theoretical Biology **31**:295-311.

He, K. S., D. Rocchini, M. Neteler, and H. Nagendra. 2011. Benefits of hyperspectral remote sensing for tracking plant invasions. Diversity and Distributions **17**:381-392.

Hestir, E. L., S. Khanna, M. E. Andrew, M. J. Santos, J. H. Viers, J. A. Greenberg, S. S. Rajapakse, and S. L. Ustin. 2008. Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. Remote sensing of Environment **112**:4034-4047.

Hildenbrandt, H., C. Carere, and C. K. Hemelrijk. 2010. Self-organized aerial displays of thousands of starlings: a model. Behavioral Ecology **21**:1349-1359.

Hughes, J., M. Haran, and P. C. Caragea. 2011. Autologistic models for binary data on a lattice. Environmetrics **22**:857-871.

Humphries, G., R. Naveen, M. Schwaller, C. Che-Castaldo, P. McDowall, M. Schrimpf, and H. Lynch. 2017. Mapping application for penguin populations and projected dynamics (MAPPPD): data and tools for dynamic management and decision support. Polar Record **53**:160-166.

Hunter, F., and L. Davis. 1998. Female Adelie penguins acquire nest material from extrapair males after engaging in extrapair copulations. The Auk:526-528.

Illian, J. B., J. Møller, and R. P. Waagepetersen. 2009. Hierarchical spatial point process analysis for a plant community with high biodiversity. Environmental and Ecological Statistics **16**:389-405.

Jackson, A. L., S. Bearhop, and D. R. Thompson. 2005. Shape can influence the rate of colony fragmentation in ground nesting seabirds. Oikos **111**:473-478.

Jakob, S., R. Zimmermann, and R. Gloaguen. 2017. The Need for Accurate Geometric and Radiometric Corrections of Drone-Borne Hyperspectral Data for Mineral Exploration: MEPHySTo—A Toolbox for Pre-Processing Drone-Borne Hyperspectral Data. Remote Sensing **9**:88.

Jenness, J. S. 2004. Calculating landscape surface area from digital elevation models. Wildlife Society Bulletin **32**:829-839.

John, E. A. 1990. Fine scale patterning of species distributions in a saxicolous lichen community at Jonas Rockslide, Canadian Rocky Mountains. Ecography **13**:187-194.

Jones, J. 2001. Habitat selection studies in avian ecology: a critical review. The Auk **118**:557-562.

Kersten, T. P., and M. Lindstaedt. 2012. Image-based low-cost systems for automatic 3D recording and modelling of archaeological finds and objects. Pages 1-10 *in* Euro-Mediterranean Conference. Springer.

Kildaw, S. D., D. B. Irons, D. R. Nysewander, and C. L. Buck. 2005. Formation and growth of new seabird colonies: the significance of habitat quality. Marine Ornithology **33**:49-58.

King, A. J., A. M. Wilson, S. D. Wilshin, J. Lowe, H. Haddadi, S. Hailes, and A. J. Morton. 2012. Selfish-herd behaviour of sheep under threat. Current Biology **22**:R561-R562.

Lack, D. 1967. Interrelationships in breeding adaptations as shown by marine birds. Pages 3-42 *in* Proceedings of the international Ornithological Congress.

LaRue, M. A., H. Lynch, P. Lyver, K. Barton, D. Ainley, A. Pollard, W. Fraser, and G. Ballard. 2014. A method for estimating colony sizes of Adélie penguins using remote sensing imagery. Polar Biology **37**:507-517.

Lavy, A., G. Eyal, B. Neal, R. Keren, Y. Loya, and M. Ilan. 2015. A quick, easy and non-intrusive method for underwater volume and surface area evaluation of benthic organisms by 3D computer modelling. Methods in Ecology and Evolution **6**:521-531.

Lawson, C. R., J. Bennie, J. A. Hodgson, C. D. Thomas, and R. J. Wilson. 2014. Topographic microclimates drive microhabitat associations at the range margin of a butterfly. Ecography **37**:732-740.

Lieshout, M. v., and A. Baddeley. 1996. A nonparametric measure of spatial interaction in point patterns. Statistica Neerlandica **50**:344-361.

Link, J., D. Senner, and W. Claupein. 2013. Developing and evaluating an aerial sensor platform (ASP) to collect multispectral data for deriving management decisions in precision farming. Computers and electronics in agriculture **94**:20-28.

Lookingbill, T. R., and D. L. Urban. 2005. Gradient analysis, the next generation: towards more plant-relevant explanatory variables. Canadian Journal of Forest Research **35**:1744-1753.

Loosmore, N. B., and E. D. Ford. 2006. Statistical inference using the G or K point pattern spatial statistics. Ecology **87**:1925-1931.

Lynch, H., and M. LaRue. 2014. First global census of the Adélie Penguin. The Auk **131**:457-466.

Lynch, H. J., W. F. Fagan, and R. Naveen. 2010. Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. Polar Biology **33**:493-503.

Manukyan, L., S. A. Montandon, A. Fofonjka, S. Smirnov, and M. C. Milinkovitch. 2017. A living mesoscopic cellular automaton made of skin scales. Nature **544**:173.

Martin, M., S. Newman, J. Aber, and R. Congalton. 1998. Determining forest species composition using high spectral resolution remote sensing data. Remote sensing of Environment **65**:249-254.

Mathews, A. J., and J. L. Jensen. 2013. Visualizing and quantifying vineyard canopy LAI using an unmanned aerial vehicle (UAV) collected high density structure from motion point cloud. Remote Sensing **5**:2164-2183.

Mcclure, C. J., B. W. Rolek, and G. E. Hill. 2012. Predicting occupancy of wintering migratory birds: Is microhabitat information necessary? The Condor **3**:482-490.

McDowall, P., and H. J. Lynch. 2017. Ultra-Fine Scale Spatially-Integrated Mapping of Habitat and Occupancy Using Structure-From-Motion. PloS one **12**:e0166773.

McKaye, K. R., D. E. Mughogho, and T. Lovullo. 1992. Formation of the selfish school. Environmental biology of fishes **35**:213-218.

Melfi, G., and G. Schoier. 2004. Simulation of random distributions on surfaces. Societa Italiana di Statistica:173-176.

Minias, P. 2014. Evolution of within-colony distribution patterns of birds in response to habitat structure. Behavioral Ecology and Sociobiology **68**:851-859.

Møller, J., and R. P. Waagepetersen. 2007. Modern statistics for spatial point processes. Scandinavian Journal of Statistics **34**:643-684.

Moreno, J., J. Bustamante, and J. Viñuela. 1995. Nest maintenance and stone theft in the chinstrap penguin (Pygoscelis antarctica). Polar Biology **15**:533-540.

Murray, J. D. 1981. A pre-pattern formation mechanism for animal coat markings. Journal of Theoretical Biology **88**:161-199.

Nash, J. 1951. Non-cooperative games. Annals of mathematics:286-295.

Niwa, H.-S. 1994. Self-organizing Dynamic Model of Fish Schooling. Journal of theoretical Biology **171**:123-136.

Opedal, Ø. H., W. S. Armbruster, and B. J. Graae. 2015. Linking small-scale topography with microclimate, plant species diversity and intra-specific trait variation in an alpine landscape. Plant Ecology & Diversity **8**:305-315.

Parrish, J. K., and L. Edelstein-Keshet. 1999. Complexity, pattern, and evolutionary trade-offs in animal aggregation. Science **284**:99-101.

Photoscan, A. 2016. URL: http://www. agisoft. com, Agisoft LLC, 11 Degtyarniy per., St. Petersburg. Russia.

Pickard, J. 2002. Assessing vegetation change over a century using repeat photography. Australian Journal of Botany **50**:409-414.

Portugal, S. J., G. Maurer, G. H. Thomas, M. E. Hauber, T. Grim, and P. Cassey. 2014. Nesting behaviour influences species-specific gas exchange across avian eggshells. Journal of Experimental Biology **217**:3326-3332.

Questad, E. J., J. R. Kellner, K. Kinney, S. Cordell, G. P. Asner, J. Thaxton, J. Diep, A. Uowolo, S. Brooks, and N. Inman-Narahari. 2014. Mapping habitat suitability for at-risk plant species and its implications for restoration and reintroduction. Ecological Applications **24**:385-395.

Rand, A. S. 1964. Ecological distribution in anoline lizards of Puerto Rico. Ecology **45**:745-752.

Renner, I. W., J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic, and D. I. Warton. 2015. Point process models for presence-only analysis. Methods in Ecology and Evolution **6**:366-379.

Renner, I. W., and D. I. Warton. 2013. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. Biometrics **69**:274-281.

Rifkin, J. L., C. L. Nunn, and L. Z. Garamszegi. 2012. Do animals living in larger groups experience greater parasitism? A meta-analysis. The American Naturalist **180**:70-82.

Ripley, B. D. 1976. The second-order analysis of stationary point processes. Journal of applied probability **13**:255-266.

Ripley, B. D. 1977. Modelling spatial patterns. Journal of the Royal Statistical Society. Series B (Methodological):172-212.

Rohani, P., T. J. Lewis, D. Grünbaum, and G. D. Ruxton. 1997. Spatial self-organisation in ecology: pretty patterns or robust reality? Trends in Ecology & Evolution **12**:70-74.

Rolland, C., E. Danchin, and M. d. Fraipont. 1998. The evolution of coloniality in birds in relation to food, habitat, predation, and life-history traits: a comparative analysis. The American Naturalist **151**:514-529.

Scheffer, M., S. R. Carpenter, T. M. Lenton, J. Bascompte, W. Brock, V. Dakos, J. Van de Koppel, I. A. Van de Leemput, S. A. Levin, and E. H. Van Nes. 2012. Anticipating critical transitions. Science **338**:344-348.

Schreiber, E. A., and J. Burger. 2001. Biology of marine birds. CRC Press.

Seabra, R., D. S. Wethey, A. M. Santos, and F. P. Lima. 2011. Side matters: microhabitat influence on intertidal heat stress over a large geographical scale. Journal of Experimental Marine Biology and Ecology **400**:200-208.

Spurr, E. 1975. Breeding of the Adélie penguin Pygoscelis adeliae at Cape Bird. Ibis **117**:324-338.

Stonehouse, B. 1975. The Biology of penguins. Macmillan.

Stoyan, D., and A. Penttinen. 2000. Recent applications of point process methods in forestry statistics. Statistical Science:61-78.

Suggitt, A. J., P. K. Gillingham, J. K. Hill, B. Huntley, W. E. Kunin, D. B. Roy, and C. D. Thomas. 2011. Habitat microclimates drive fine-scale variation in extreme temperatures. Oikos **120**:1-8.

Tattoni, C., F. Rizzolli, and P. Pedrini. 2012. Can LiDAR data improve bird habitat suitability models? Ecological modelling **245**:103-110.

Tella, J. 2002. The evolutionary transition to coloniality promotes higher blood parasitism in birds. Journal of Evolutionary Biology **15**:32-41.

Tenaza, R. 1971. Behavior and nesting success relative to nest location in Adelie penguins (Pygoscelis adeliae). The Condor **73**:81-92.

Torr, P. H., and A. Zisserman. 1999. Feature based methods for structure and motion estimation. Pages 278-294 *in* International workshop on vision algorithms. Springer.

Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote sensing of Environment **8**:127-150.

Ustin, S. L., D. DiPietro, K. Olmstead, E. Underwood, and G. Scheer. 2002. Hyperspectral remote sensing for invasive species detection and mapping. Pages 1658-1660 *in* Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International. IEEE.

Van Laarhoven, P. J., and E. H. Aarts. 1987. Simulated annealing. Pages 7-15 Simulated annealing: Theory and applications. Springer.

Velázquez, E., I. Martínez, S. Getzin, K. A. Moloney, and T. Wiegand. 2016. An evaluation of the state of spatial point pattern analysis in ecology. Ecography **39**:1042-1055.

Vierling, K. T., L. A. Vierling, W. A. Gould, S. Martinuzzi, and R. M. Clawges. 2008. Lidar: shedding new light on habitat characterization and modeling. Frontiers in Ecology and the Environment **6**:90-98.

Viscido, S. V., and D. S. Wethey. 2002. Quantitative analysis of fiddler crab flock movement: evidence for 'selfish herd'behaviour. Animal behaviour **63**:735-741.

Wagner, R. H. 1993. The pursuit of extra-pair copulations by female birds: a new hypothesis of colony formation. Journal of Theoretical Biology **163**:333-346.

Wallace, L., A. Lucieer, Z. Malenovský, D. Turner, and P. Vopěnka. 2016. Assessment of forest structure using two UAV techniques: A comparison of airborne laser scanning and structure from motion (SfM) point clouds. Forests **7**:62.

Weidinger, K. 1998. Effect of predation by skuas on breeding success of the Cape petrel Daption capense at Nelson Island, Antarctica. Polar Biology **20**:170-177.

Wiegand, T., P. Grabarnik, and D. Stoyan. 2016. Envelope tests for spatial point patterns with and without simulation. Ecosphere **7**.

Wittenberger, J. F. 1985. The adaptive significance of coloniality in birds. Avian biology:1-78.

Wood, A. J., and G. J. Ackland. 2007. Evolving the selfish herd: emergence of distinct aggregating strategies in an individual-based model. Proceedings of the Royal Society of London B: Biological Sciences **274**:1637-1642.

Wu, C. 2007. SiftGPU: A GPU implementation of sift. http://cs. unc. edu/~ ccwu/siftgpu.

Wu, C. 2013. Towards linear-time incremental structure from motion. Pages 127-134 *in* 3D Vision-3DV 2013, 2013 International conference on. IEEE.

Wu, C., S. Agarwal, B. Curless, and S. M. Seitz. 2011. Multicore bundle adjustment. Pages 3057-3064 *in* Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE.

Ye, X., K. Sakai, H. Okamoto, and L. O. Garciano. 2008. A ground-based hyperspectral imaging system for characterizing vegetation spectral features. Computers and electronics in agriculture **63**:13-21.

Young, E. 2005. Skua and penguin: predator and prey. Cambridge University Press.

Zador, S. G., J. K. Parrish, and A. E. Punt. 2009. Factors influencing subcolony colonization and persistence in a colonial seabird, the common murre Uria aalge. Marine Ecology Progress Series **376**:283-293.

Zahawi, R. A., J. P. Dandois, K. D. Holl, D. Nadwodny, J. L. Reid, and E. C. Ellis. 2015. Using lightweight unmanned aerial vehicles to monitor tropical forest recovery. Biological conservation **186**:287-295.

青木一郎. 1982. A simulation study on the schooling mechanism in fish. 日本水産学会誌 **48**:1081-1088.

Appendix

Orthorectification of Point Process Models

Implementing simulations of point processes on surfaces

The following sections provide R code to simulate Poisson Point Processes over parametric and rasterized surfaces.

For a parametric surface of the form $z = f(x, y)$:

1. Define a function $g(x, y)$ such that $g(x, y) = \lambda\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$, where λ is the intensity of the point process on the surface.
2. Generate a set of points uniformly randomly over the domain W with intensity equal to the maximum of $g(x, y)$ over the region
3. Calculate $g(x_i, y_i)$ for each point
4. Accept points with probability equal to $\frac{g(x_i, y_i)}{max_{i \in W} g(x_i, y_i)}$

Create a function `f(x,y)` which returns the elevation value `z`.

```
SINFun = function(x,y){
  return(sin(x)*2)
}
```

Define additional functions which return the derivative of `f(x,y)` w.r.t `x` and `y`.

```
dzdx = function(x,y){
  return(cos(x))
}

dzdy = function(x,y){
  return(0)
}
```

The intensity on the surface at any point x_i, y_i is given by $g(x, y) = \lambda\sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$

```
m1 = function(x,y){
  return(sqrt(1+dzdx(x,y)^2+dzdy(x,y)^2))
  }
```

The rejection sampling algorithm requires the maximum of `g(x,y)` over the window `W`. For many functions this can be found using the `stats::optim` function.

```
m1optim = function(XY){
  return(-sqrt(1+dzdx(XY[1],XY[2])^2+dzdy(XY[1],XY[2])^2))
}
```

```
maxM1 = -optim(par = c(1,1),lower=c(minX,minY),upper=c(maxX,maxY),fn =
m1optim,method="L-BFGS-B")$value
```

We then generate a homogeneous Poisson point pattern over the window W with intensity equal to $\lambda * \max M1$, and from that pattern we accept points with probability equal to $\frac{g(x_i,y_i)}{max_{i\in W} g(x_i,y_i)}$.

```
area = (maxX-minX)*(maxY-minY)
x=runif(lambda*maxM1*area,minX,maxX)
y=runif(lambda*maxM1*area,minY,maxY)
w=runif(lambda*maxM1*area,0,1)
x=x[w<(m1(x,y)/maxM1)]
y=y[w<(m1(x,y)/maxM1)]
```

Examples

Sine Function

```
# Generate a Poisson point process on a surface described by a sine function

generateSinPoints <- function(lambda,xrange,yrange,plot=FALSE){

SINFun = function(x,y){
  return(sin(x))
}

minX <- xrange[1]
maxX <- xrange[2]
minY <- yrange[1]
maxY <- yrange[2]

x_ <- seq(minX,maxX,by=0.5)
y_ <- seq(minY,maxY,by=0.5)
z_ <- outer(x_, y_, SINFun)

dzdx = function(x,y){
  return(cos(x))
}

dzdy = function(x,y){
  return(0)
}

m1 = function(x,y){
  return(sqrt(1+dzdx(x,y)^2+dzdy(x,y)^2))
  }

m1optim = function(XY){
  return(-sqrt(1+dzdx(XY[1],XY[2])^2+dzdy(XY[1],XY[2])^2))
  }

maxM1 = -optim(par = c(1,1),lower=c(minX,minY),upper=c(maxX,maxY)
               ,fn = m1optim,method="L-BFGS-B")$value

area = (maxX-minX)*(maxY-minY)
x=runif(lambda*maxM1*area,minX,maxX)
y=runif(lambda*maxM1*area,minY,maxY)
w=runif(lambda*maxM1*area,0,1)
keep = w<(m1(x,y)/maxM1)
x=x[keep]
y=y[keep]

if(plot == TRUE){
rot<-persp(x_, y_, z_,main="Sine Function",xlab="x",ylab="y",zlab="z")
points(trans3d(x = x,y=y,z = SINFun(x,y),pmat=rot),col="blue",pch=3)
}
return(ppp(x,y,window=owin(xrange=xrange,yrange=yrange)))

}
```

```
# Fig. S1a - CSR points on a synthetic sinusoidal surface
generateSinPoints(10,c(0,8),c(0,2),plot=TRUE)
```

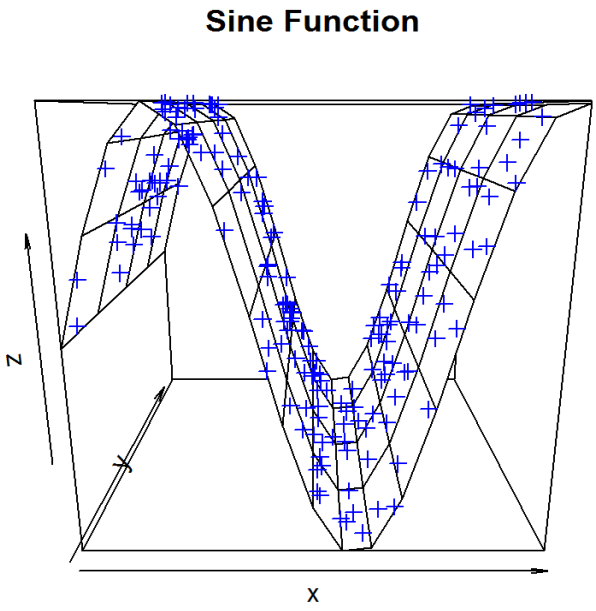

Fig. S1a CSR points on a synthetic sinusoidal surface.

```
# Fig. S1b - KDE of projection of CSR points on a synthetic sinusoidal surface
plot(density(generateSinPoints(200,c(0,8),c(0,2))),main="")
```

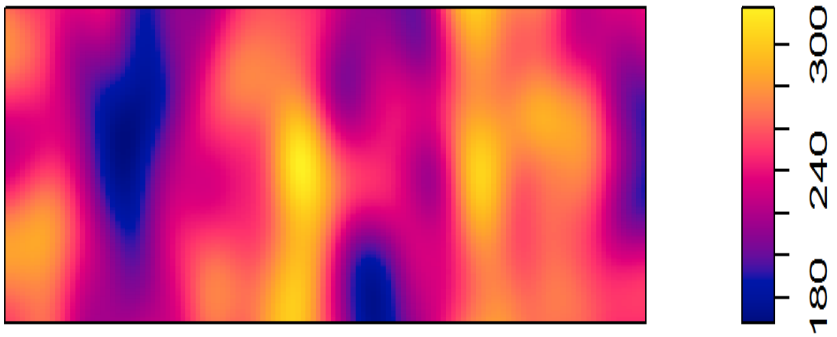

Fig. S1b Kernel Density Estimate of planar projection of points simulated on a synthetic sinusoidal surface.

Note: An alternative method to generate topographically correct point patterns is to use the `spatstat` package to generate an inhomogeneous point pattern by creating a `spatstat im` object containing a rasterized version of the correction factor. While this method is approximate, it is more efficient and simpler to implement, and the approximation is likely to be sufficient for most cases.

The function `spatstat::as.im` converts an input, `x`, into an object of class `im`, representing a 2D pixel image which is compatible with other functions in the `spatstat` package. When the input argument `x` is a function, this function is evaluated at all cells within the window `W`. As "the function X will be evaluated in the form X(x, y, …) where x and y are vectors containing the x and y coordinates of all the pixels in the image mask", the function used to generate the image must be vectorized. The `Vectorize` function is used to generate a rasterized version of the function $z = f(x, y)$. The function `sp::surfaceArea` can then be used on this pixel image to determine the ratio between the true and projected areas of each cell in the window.

```
#Not Run
step=as.im(SINFun,W=owin(xrange=c(0,8),yrange=c(0,2)),eps=0.01)
stepSA =
as.im(surfaceArea(as.matrix(step),cellx=step$xstep,celly=step$ystep,byCell=TRUE),W
=owin(xrange=c(0,8),yrange=c(0,2)))
stepCF = stepSA/(step$xstep*step$ystep)
stepCF = as.im(stepCF,W=as.owin(step))
```

Step Function

```
# Generate a Poisson point process on a surface described by a step function

generateStepPoints <- function(lambda,xrange,yrange,plot=FALSE){

STEPFun = function(x,y){
  if(x < 2){return(4)}
  else if(x>=2 & x<4){return(4 + (x-2)*1)}
  else if(x>=4 & x<6){return(6)}
  else if(x>=6){return(6 + (x-6)*2)}
}

STEPFun <- Vectorize(STEPFun)

minX <- xrange[1]
maxX <- xrange[2]
minY <- yrange[1]
maxY <- yrange[2]

x_ <- seq(minX,maxX,by=0.5)
y_ <- seq(minY,maxY,by=0.5)
z_ <- outer(x_, y_, STEPFun)

dzdx = function(x,y){
  if(x < 2){return(0)}
  else if(x>=2 & x<4){return(1)}
  else if(x>=4 & x<6){return(0)}
  else if(x>=6){return(2)}
}

dzdx = Vectorize(dzdx)

dzdy = function(x,y){
  return(0)
}

m1 = function(x,y){
  return(sqrt(1+dzdx(x,y)^2+dzdy(x,y)^2))
  }

maxM1 = 2.236068
area = (maxX-minX)*(maxY-minY)
x=runif(lambda*maxM1*area,minX,maxX)
y=runif(lambda*maxM1*area,minY,maxY)
w=runif(lambda*maxM1*area,0,1)
keep=w<(m1(x,y)/maxM1)
x=x[keep]
y=y[keep]

if(plot == TRUE){
rot<-persp(x_, y_, z_,main="Step Function",xlab="x",ylab="y",zlab="z")
points(trans3d(x = x,y=y,z = STEPFun(x,y),pmat=rot),col="blue",pch=3)
}

return(ppp(x,y,window=owin(xrange=xrange,yrange=yrange)))

}
```

```
# Fig S2a - CSR points on a synthetic stepped surface
generateStepPoints(10,c(0,8),c(0,2),plot=TRUE)
```

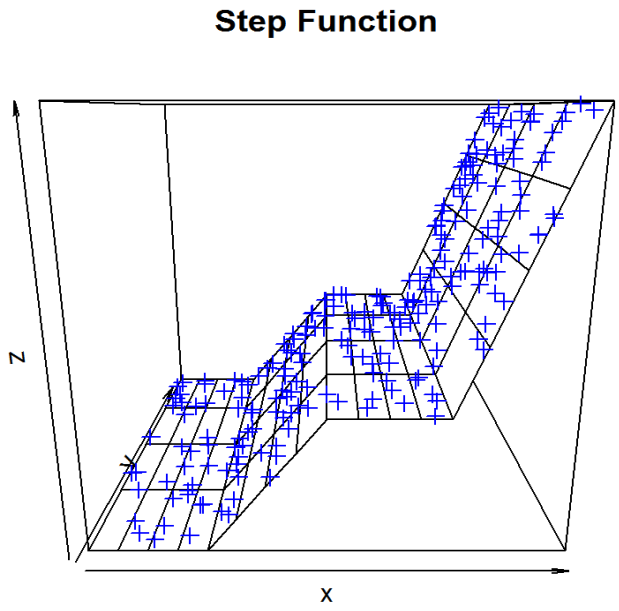

Fig. S2a CSR points on a synthetic stepped surface.

```
# Fig S2b - KDE of projection of CSR points on a synthetic stepped surface
plot(density(generateStepPoints(200,c(0,8),c(0,2))),main="")
```

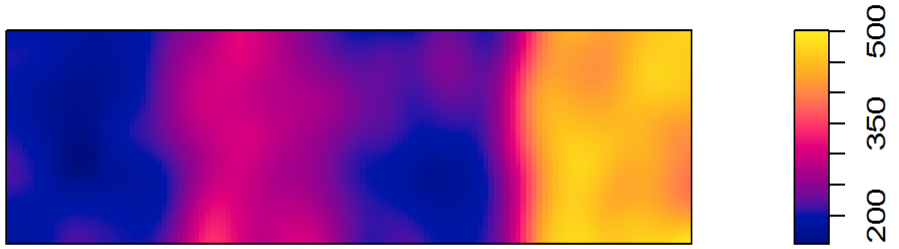

Fig. S2b Kernel Density Estimate of planar projection of points simulated on a synthetic stepped surface.

Exponential Function

```
# Generate a Poisson point process on a surface described by an exponential
# function

generateExpPoints <- function(lambda,xrange,yrange,plot=FALSE){

EXPFun = function(x,y){
  return(exp(x/4))
}

minX <- xrange[1]
maxX <- xrange[2]
minY <- yrange[1]
maxY <- yrange[2]

x_ <- seq(minX,maxX,by=0.5)
y_ <- seq(minY,maxY,by=0.5)
z_ <- outer(x_, y_, EXPFun)

dzdx = function(x,y){
  return(exp(x/4)/4)
}

dzdy = function(x,y){
  return(0)
}

m1 = function(x,y){
  return(sqrt(1+dzdx(x,y)^2+dzdy(x,y)^2))
  }

m1optim = function(XY){
  return(-sqrt(1+dzdx(XY[1],XY[2])^2+dzdy(XY[1],XY[2])^2))
  }

maxM1 = -optim(par = c(1,1),lower=c(minX,minY),upper=c(maxX,maxY),fn =
m1optim,method="L-BFGS-B")$value


area = (maxX-minX)*(maxY-minY)
x=runif(lambda*maxM1*area,minX,maxX)
y=runif(lambda*maxM1*area,minY,maxY)
w=runif(lambda*maxM1*area,0,1)
keep=w<(m1(x,y)/maxM1)
x=x[keep]
y=y[keep]

if(plot == TRUE){
rot<-persp(x_, y_, z_,main="Exponential Function",xlab="x",ylab="y",zlab="z")
points(trans3d(x = x,y=y,z = EXPFun(x,y),pmat=rot),col="blue",pch=3)
}
return(ppp(x,y,window=owin(xrange=xrange,yrange=yrange)))

}
```

```
# Fig. S3a - CSR points on a synthetic exponential surface
generateExpPoints(12,c(0,8),c(0,2),plot=TRUE)
```

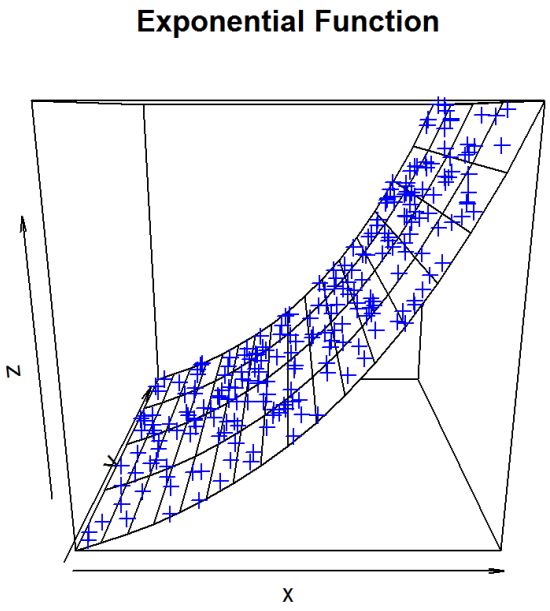

Fig. S3a CSR points on a synthetic exponential surface.

```
# Fig. S3b - KDE of projection of CSR points on a synthetic exponential surface
plot(density(generateExpPoints(200,c(0,8),c(0,2))),main="")
```

Fig. S3b Kernel Density Estimate of planar projection of points simulated on a synthetic stepped surface.

Hypothesis Testing

We use two Monte Carlo based approaches to hypothesis testing; The Diggle (1986) / Cressie (1991) / Loosmore and Ford (2006) (DCLF) test and the Maximum Absolute Deviation test. Additionally, we use one analytic test, Analytic Global Envelopes (AGE). For the implementation of the AGE, see the supplemental file "AGEpcf.r". Due to the stochastic nature of these tests, results may vary between runs. Additionally, the number of iterations (set by the variables `total` and `nSim`) have been reduced in this document due to long run times.

```
source("AGEpcf.r")

dclf_fail_plane<-0
mad_fail_plane<-0
dclf_fail_plane_pcf<-0
mad_fail_plane_pcf<-0
age_fail_plane<-0
total <- 500
nSim <- 99
for(i in 1:total){
  ppp_ <- rpoispp(lambda = 12, win = as.owin(c(0,8,0,2)))
  if(dclf.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_plane=dclf_fail_plane+1}
  if(mad.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_plane=mad_fail_plane+1}
  if(dclf.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_plane_pcf=dclf_fail_plane_pcf+1}
  if(mad.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_plane_pcf=mad_fail_plane_pcf+1}
  if(AGEpcfPp(ppp_,0.05)$p<=0.05){age_fail_exp = age_fail_plane+1}
}



dclf_fail_step<-0
mad_fail_step<-0
dclf_fail_step_pcf<-0
mad_fail_step_pcf<-0
age_fail_step<-0
total <- 500
nSim <- 99

for(i in 1:total){
  ppp_ <- generateStepPoints(10,c(0,8),c(0,2))
  if(dclf.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_step=dclf_fail_step+1}
  if(mad.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_step=mad_fail_step+1}
  if(dclf.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_step_pcf=dclf_fail_step_pcf+1}
  if(mad.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_step_pcf=mad_fail_step_pcf+1}
  if(AGEpcfPp(ppp_,0.05)$p<=0.05){age_fail_step = age_fail_step+1}
}


dclf_fail_sin<-0
mad_fail_sin<-0
```

```
dclf_fail_sin_pcf<-0
mad_fail_sin_pcf<-0
age_fail_sin<-0
total <- 500
nSim <- 99

for(i in 1:total){
  ppp_ <- generateSinPoints(10,c(0,8),c(0,2))
  if(dclf.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_sin=dclf_fail_sin+1}
  if(mad.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_sin=mad_fail_sin+1}
  if(dclf.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_sin_pcf=dclf_fail_sin_pcf+1}
  if(mad.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_sin_pcf=mad_fail_sin_pcf+1}
  if(AGEpcfPp(ppp_,0.05)$p<=0.05){age_fail_sin = age_fail_sin+1}
}


dclf_fail_exp<-0
mad_fail_exp<-0
dclf_fail_exp_pcf<-0
mad_fail_exp_pcf<-0
age_fail_exp<-0
total <- 500
nSim <- 99

for(i in 1:total){
  ppp_ <- generateExpPoints(12,c(0,8),c(0,2))
  if(dclf.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_exp=dclf_fail_exp+1}
  if(mad.test(ppp_, Kest,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_exp=mad_fail_exp+1}
  if(dclf.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){dclf_fail_exp_pcf=dclf_fail_exp_pcf+1}
  if(mad.test(ppp_, pcf,
nsim=nSim,verbose=FALSE)$p<=0.05){mad_fail_exp_pcf=mad_fail_exp_pcf+1}
  if(AGEpcfPp(ppp_,0.05)$p<=0.05){age_fail_exp = age_fail_exp+1}
}
```

Results

For each surface we record the number of times that we would reject the null hypothesis of complete spatial randomness. Using an alpha level of 0.05 we would expect the null to be rejected at a rate of 5% due to chance.

Point Pattern	**M.A.D. (K(r))**	**dclf (K(r))**	**M.A.D. (g(r))**	**dclf (g(r))**	**AGE (g(r))**
CSR on Plane	3.2%	5.4 %	3.8%	4.4%	0%
CSR Step Function	81%	77.4 %	8.2%	11.4%	13%
CSR Sine Function	7.8%	8%	6.4%	6%	7.2%
CSR Exponential Function	35.6%	35.8%	5.6%	5.6%	14%

Table S1. Realized rates of rejection of the null for hypothesis tests of CSR, when CSR is assumed to operate in the x-y plane, but the generating process operates on a 2D manifold in 3D space.

Simulation Envelopes - Ripleys K

```
# Generate simulation envelopes for CSR on plane and on parametric surface

iters = 100
tmp=Kest(rpoispp(12,win=owin(c(0,8),c(0,2))),rmax=1)
RANupper = rep(0,length(tmp$r))
RANlower = rep(10,length(tmp$r))
for(i in 1:iters){
  points_= rpoispp(12,win=owin(c(0,8),c(0,2)))
  tmp=Kest(points_,rmax=1)
  RANupper=pmax(RANupper,tmp$iso)
  RANlower=pmin(RANlower,tmp$iso)
}
poisEnvelope = list(upper=RANupper,lower=RANlower,r=tmp$r)

generateEnvelopes = function(iters,genFunc,lambda){
  tmp = Kest(genFunc(lambda,c(0,8),c(0,2)),rmax=1)
  upper = rep(0,length(tmp$r))
  lower = rep(10,length(tmp$r))
  for(i in 1:iters){
    points_=genFunc(lambda,c(0,8),c(0,2))
    tmp=Kest(points_,rmax=1)
    upper=pmax(upper,tmp$iso)
    lower=pmin(lower,tmp$iso)
  }
  return(list(upper=upper,lower=lower,r=tmp$r))
}

plotEnvelopes <- function(testEnvelope,poissonEnvelope,title){

plot(testEnvelope$r,testEnvelope$upper,type="l",col="red",ylab="K(r)",xlab="r",mai
n=title)
  lines(testEnvelope$r,testEnvelope$lower,col="red")
  lines(testEnvelope$r,poissonEnvelope$lower,col="blue")
  lines(testEnvelope$r,poissonEnvelope$upper,col="blue")
}

stepEnvelopes <- generateEnvelopes(iters = 100,genFunc = generateStepPoints,lambda
= 10)
sinEnvelopes <- generateEnvelopes(iters = 100,genFunc = generateSinPoints,lambda =
10)
expEnvelopes <- generateEnvelopes(iters = 100,genFunc = generateExpPoints,lambda =
12)
```

```
# Fig. S4a - Simulation envelopes of Ripleys K for sine surface
plotEnvelopes(sinEnvelopes,poisEnvelope,"Ripleys K, Sine Function")
```

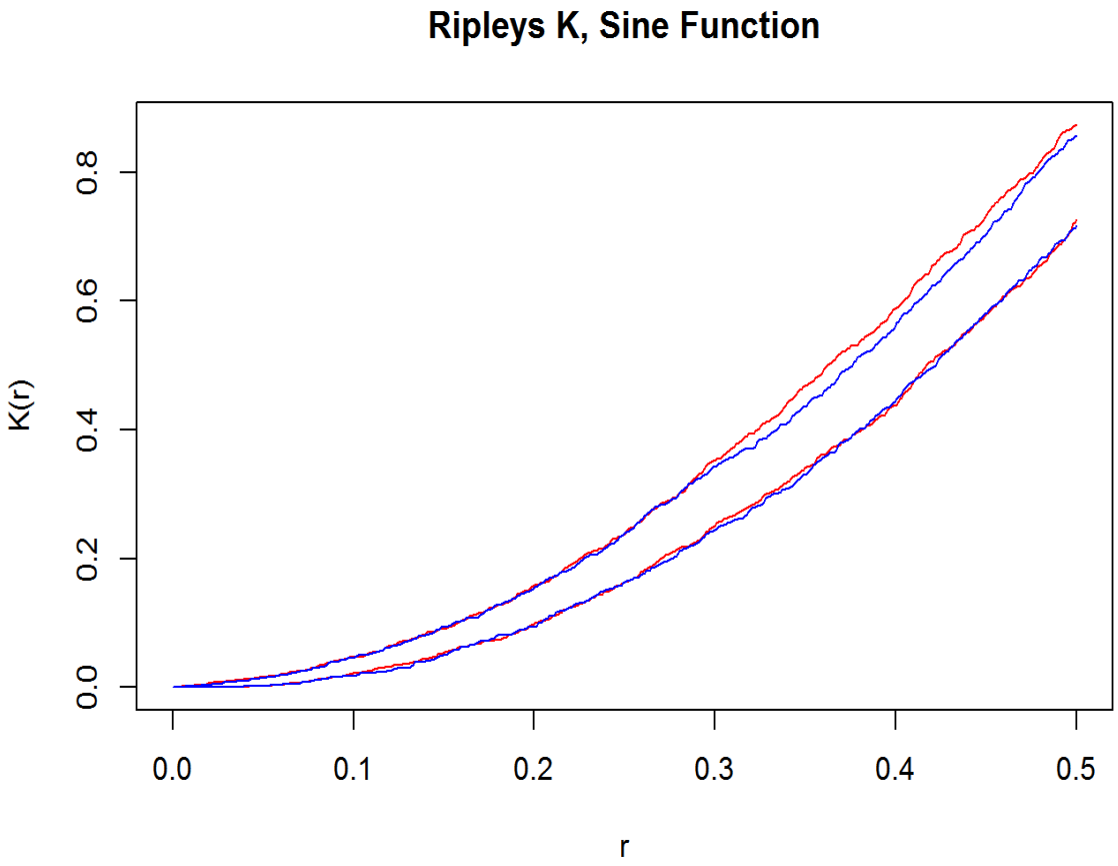

Fig S4a Ripleys K calculated for the x-y plane projection of CSR points generated on synthetic sinusoidal surface.

```
# Fig. S4b - Simulation envelopes of Ripleys K for step surface
plotEnvelopes(stepEnvelopes,poisEnvelope,"Ripleys K, Step Function")
```

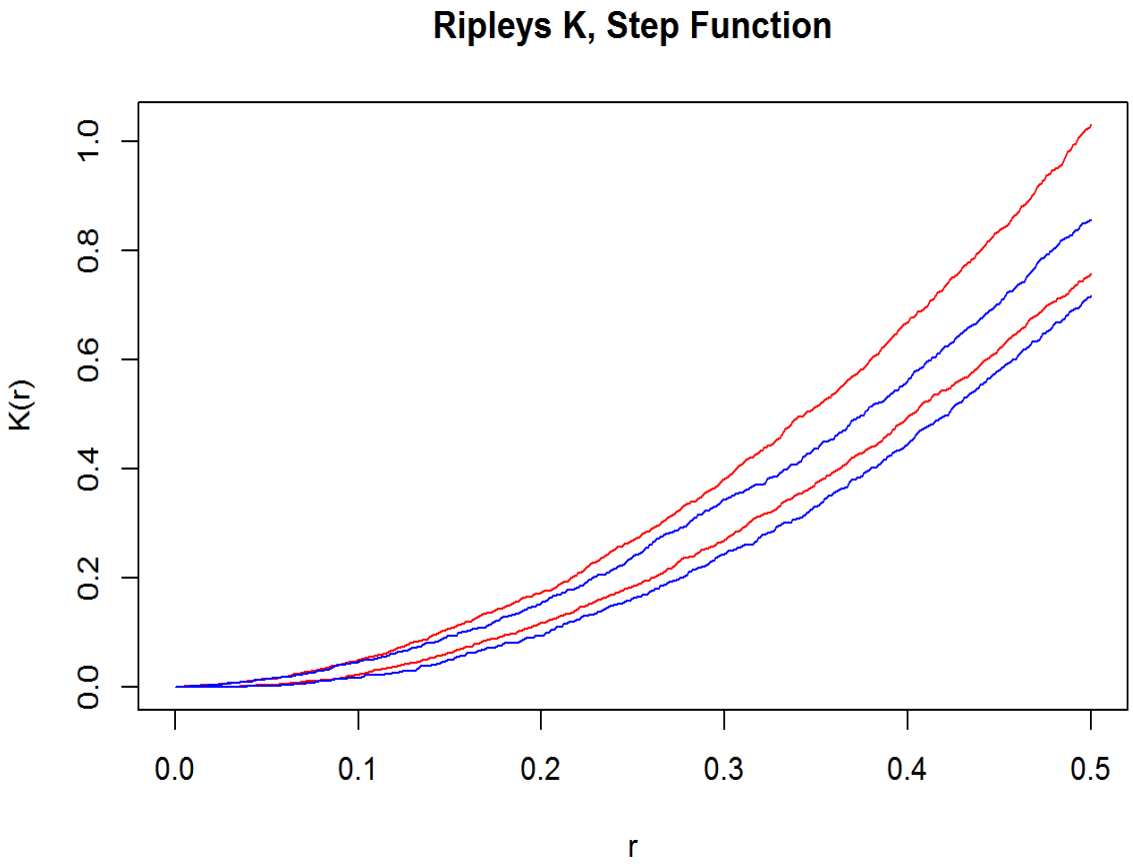

Fig. S4b Ripleys K calculated for the x-y plane projection of CSR points generated on synthetic stepped surface.

```
# Fig. S4c - Simulation envelopes of Ripleys K for exponential surface
plotEnvelopes(expEnvelopes,poisEnvelope,"Ripleys K, Exponential Function")
```

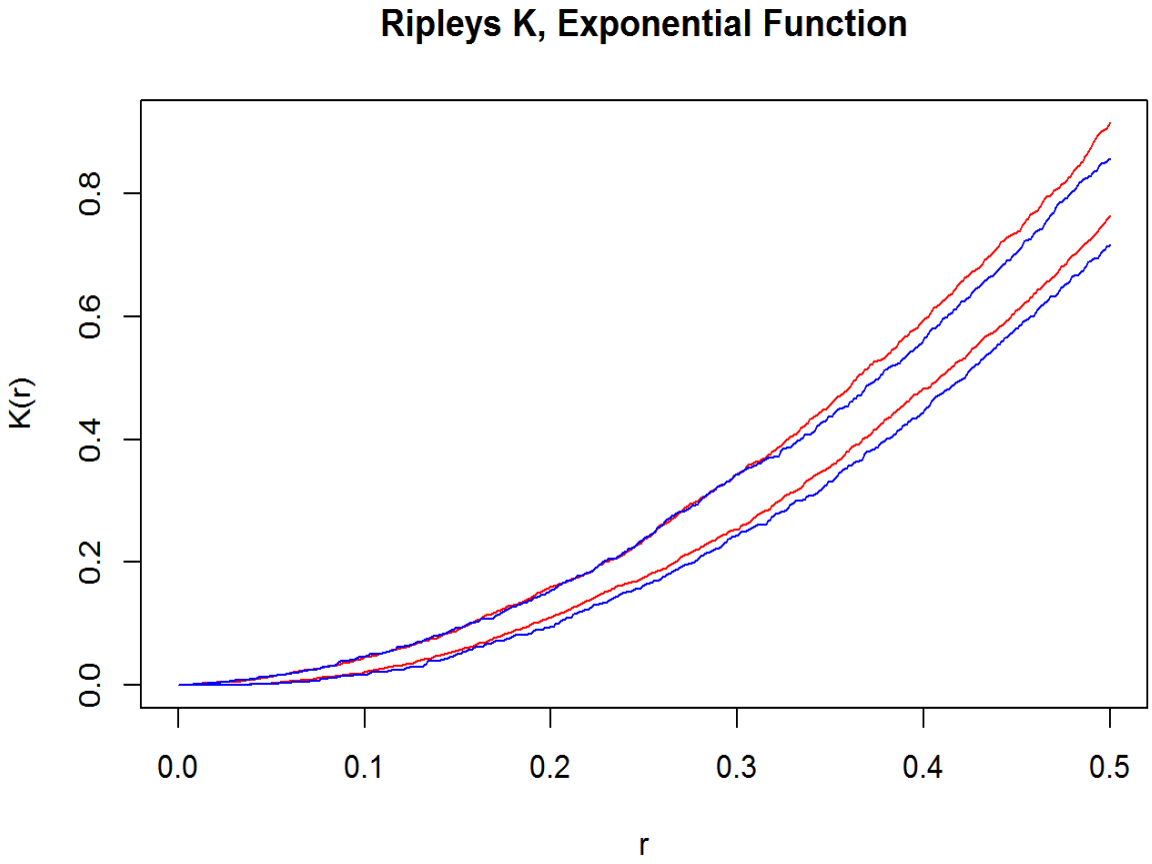

Fig. S4c Ripleys K calculated for the x-y plane projection of CSR points generated on synthetic exponential surface.

Topographically Corrected Model Fitting

Bei dataset

"The dataset bei gives the positions of 3605 trees of the species *Beilschmiedia pendula* (Lauraceae) in a 1000 by 500 metre rectangular sampling region in the tropical rainforest of Barro Colorado Island. The accompanying dataset bei.extra gives information about the altitude (elevation) in the study region. It is a list containing two pixel images, elev (elevation in metres) and grad (norm of elevation gradient). These data are part of a much larger dataset containing the positions of hundreds of thousands of trees belong to thousands of species; see Hubbell and Foster (1983), Condit, Hubbell and Foster (1996) and Condit (1998)." –Reproduced from spatstat package documentation, version 1.45-0

```
# Fig. S5 - Bei dataset overlaid on included terrain model

M <- persp(bei.extra$elev,
           colmap=terrain.colors(128),
           apron=TRUE, theta=-30,
           phi=20,zlab="Elevation",
           main="Bei Dataset",
           expand=6,
           visible=TRUE,
           shade=0.3,
           xlab="1000m",
           ylab="500m")

perspPoints(bei,
           Z=bei.extra$elev,
           M=M,
           pch=16,
           cex=0.3)

image.plot(legend.only=T,
          zlim=range(bei.extra$elev),
          col=terrain.colors(128),
          horizontal=TRUE)
```

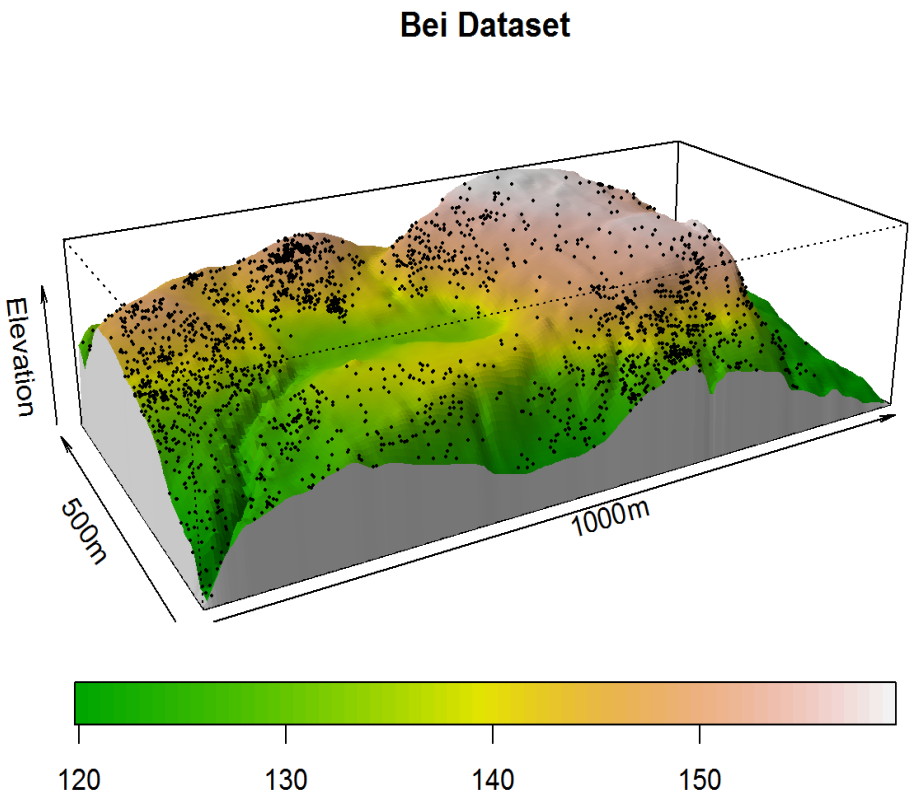

Fig. S5 Bei dataset from spatstat package. Points indicate the position of *Beilschmiedia pendula* in a 1000 by 500 meter region on Barro Colorado Island.

Bei model without topographic correction

```
# Fit model of bei intensity against slope with no topographic correction

beiNoCorrection <- ppm(bei ~  slope,covariates = list(slope = bei.extra$grad))

summary(beiNoCorrection)
## Point process model
## Fitting method: maximum likelihood (Berman-Turner approximation)
## Model was fitted using glm()
## Algorithm converged
## Call:
## ppm.formula(Q = bei ~ slope, covariates = list(slope = bei.extra$grad))
## Edge correction: "border"
##  [border correction distance r = 0 ]
## --------------------------------------------------------------------------------
## Quadrature scheme = data + dummy + weights
## Data pattern:
## Planar point pattern:  3604 points
## Average intensity 0.00721 points per square metre
## Window: rectangle = [0, 1000] x [0, 500] metres
## Window area = 5e+05 square metres
## Unit of length: 1 metre
##
## Dummy quadrature points:
## (130 x 130 grid, plus 4 corner points)
## Planar point pattern:  16904 points
## Average intensity 0.0338 points per square metre
## Window: rectangle = [0, 1000] x [0, 500] metres
## Window area = 5e+05 square metres
## Unit of length: 1 metre
##
## Quadrature weights:
## (counting weights based on 130 x 130 array of rectangular tiles)
## All weights:
##  range: [1.64, 29.6] total: 5e+05
## Weights on data points:
##  range: [1.64, 14.8] total: 41000
## Weights on dummy points:
##  range: [1.64, 29.6] total: 459000
## --------------------------------------------------------------------------------
## FITTED MODEL:
##
## Nonstationary Poisson process
##
## ---- Intensity: ----
##
## Log intensity: ~slope
## Model depends on external covariate 'slope'
## Covariates provided:
##  slope: im
##
## Fitted trend coefficients:
## (Intercept)       slope
##   -5.390553    5.022021
##
##               Estimate       S.E.   CI95.lo   CI95.hi Ztest       Zval
## (Intercept) -5.390553 0.03001716 -5.449385 -5.331720   *** -179.58236
## slope        5.022021 0.24540264  4.541041  5.503002   ***   20.46441
```

```
##
## ------------ gory details -----
##
## Fitted regular parameters (theta):
## (Intercept)       slope
##   -5.390553    5.022021
##
## Fitted exp(theta):
##  (Intercept)        slope
## 4.559453e-03 1.517176e+02
```

```
#Fig S6a - 2D plot of fitted intensity
plot(predict(beiNoCorrection),main = '')
```

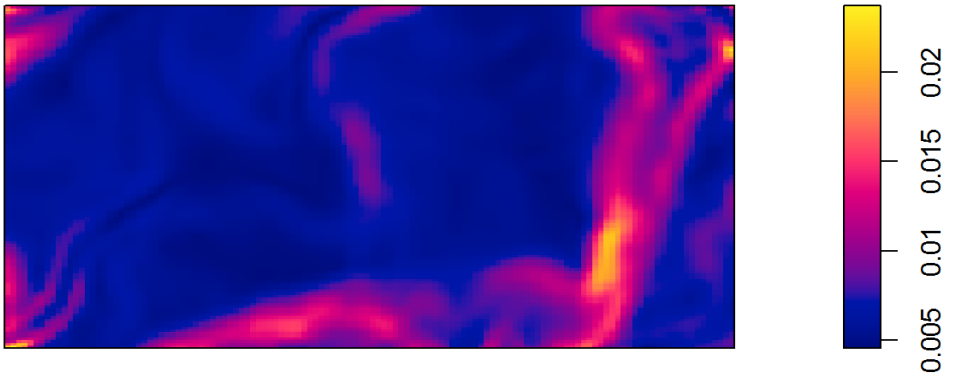

Fig. S6a Expected number of points (locations of *Beilschmiedia pendula*) predicted by model fitted without correction for projection to the x-y plane

```
#Fig S6b - 3D plot of fitted intensity
 M <- persp(bei.extra$elev,
            colmap=heat.colors(128),
            colin = predict(beiNoCorrection),
            apron=TRUE,
            theta=-30,
            phi=20,
            zlab="Elevation",
            expand=6,
            visible=TRUE,
            shade=0.3,
            xlab="1000m",
            ylab="500m",
            main = '')

 image.plot(legend.only=T,
            zlim=range(predict(beiNoCorrection)),
            col=heat.colors(128),
            horizontal=TRUE)
```

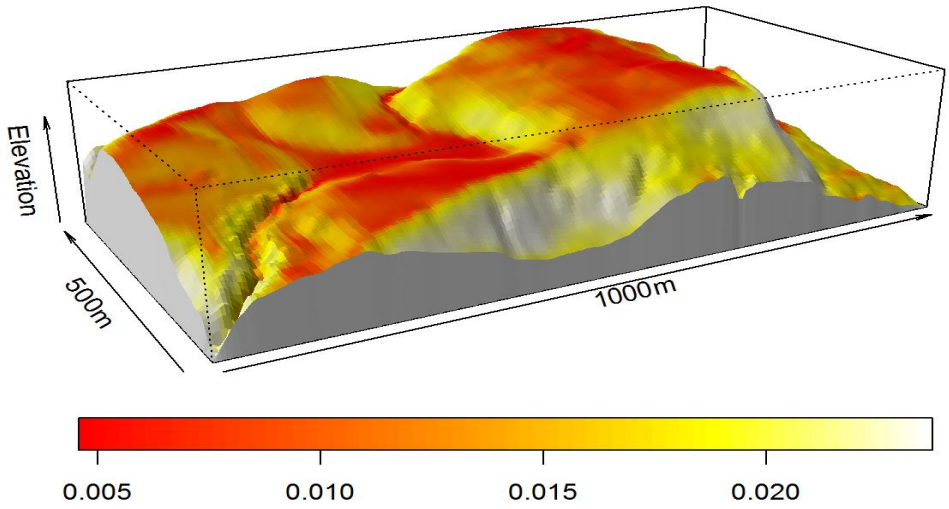

Fig. S6b Expected number of points (locations of *Beilschmiedia pendula*) predicted by model fitted without correction for projection to the x-y plane

Bei model with correction

The topographically corrected surface area for each cell in a raster DEM can be calculated using the `surfaceArea` function from the `sp` package. This method implements the 8 point algorithm from Jenness (2004). The resulting raster can be converted back to a `spatstat` compatible dataset using the `as.im` function from the `spatstat` package. This raster must then be divided by the area of each cell in the raster to produce the correction factor between S and W.

```
SurfaceArea=as.im(surfaceArea(as.matrix(bei.extra$elev),
                              cellx=5,
                              celly=5,
                              byCell=TRUE),
                  W=as.owin(bei.extra$elev))

TopoCorrection = SurfaceArea/(bei.extra$elev$xstep*bei.extra$elev$ystep)
```

Models of inhomogeneous point processes can be topographically corrected by including $\log(S_i/W_i)$ as an offset term when fitting models using the spatstat function 'ppm'.

```
# Fit model of bei intensity against slope with topographic correction

beiWithCorrection <- ppm(bei ~ offset(log(TopoCorrection)) + slope,
                         covariates = list(TopoCorrection = TopoCorrection, slope
= bei.extra$grad))

summary(beiWithCorrection)
## Point process model
## Fitting method: maximum likelihood (Berman-Turner approximation)
## Model was fitted using glm()
## Algorithm converged
## Call:
## ppm.formula(Q = bei ~ offset(log(TopoCorrection)) + slope, covariates =
list(TopoCorrection = TopoCorrection,
##     slope = bei.extra$grad))
## Edge correction: "border"
##  [border correction distance r = 0 ]
## --------------------------------------------------------------------------
## Quadrature scheme = data + dummy + weights
## Data pattern:
## Planar point pattern:  3604 points
## Average intensity 0.00721 points per square metre
## Window: rectangle = [0, 1000] x [0, 500] metres
## Window area = 5e+05 square metres
## Unit of length: 1 metre
##
## Dummy quadrature points:
## (130 x 130 grid, plus 4 corner points)
## Planar point pattern:  16904 points
## Average intensity 0.0338 points per square metre
## Window: rectangle = [0, 1000] x [0, 500] metres
## Window area = 5e+05 square metres
## Unit of length: 1 metre
##
## Quadrature weights:
```

```
## (counting weights based on 130 x 130 array of rectangular tiles)
## All weights:
##  range: [1.64, 29.6] total: 5e+05
## Weights on data points:
##  range: [1.64, 14.8] total: 41000
## Weights on dummy points:
##  range: [1.64, 29.6] total: 459000
## ---------------------------------------------------------------------------
## FITTED MODEL:
##
## Nonstationary Poisson process
##
## ---- Intensity: ----
##
## Log intensity: ~offset(log(TopoCorrection)) + slope
## Model depends on external covariates 'TopoCorrection' and 'slope'
## Covariates provided:
##  TopoCorrection: im
##  slope: im
##
## Fitted trend coefficients:
## (Intercept)       slope
##   -5.384777    4.860740
##
##             Estimate       S.E.   CI95.lo   CI95.hi Ztest       Zval
## (Intercept) -5.384777 0.02999295 -5.443562 -5.325992   *** -179.53476
## slope        4.860740 0.24511659  4.380321  5.341160   ***   19.83032
##
## ----------- gory details -----
##
## Fitted regular parameters (theta):
## (Intercept)       slope
##   -5.384777    4.860740
##
## Fitted exp(theta):
##  (Intercept)        slope
## 4.585863e-03 1.291198e+02
```

```
# Fig S7a - 2D plot of fitted intensity
plot(predict(beiWithCorrection),main = '')
```

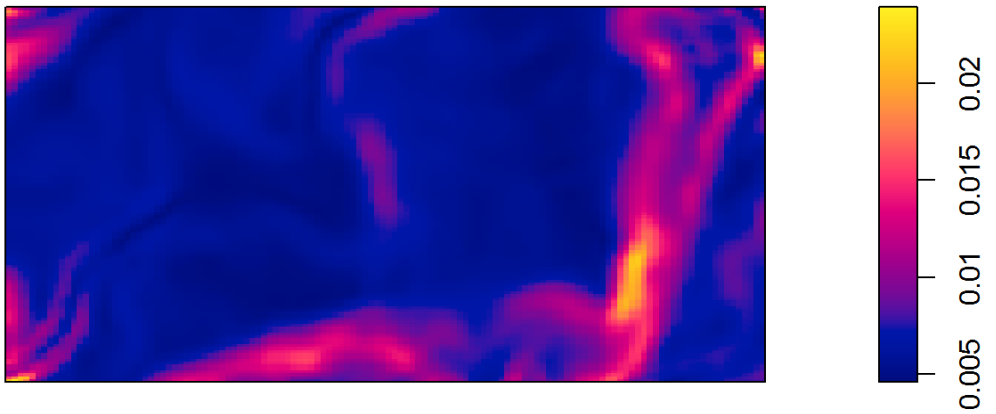

Fig. S7a Expected number of points (locations of *Beilschmiedia pendula*) predicted by model fitted with correction for projection to the x-y plane

```
# Fig S7b - 3D plot of fitted intensity
M <- persp(bei.extra$elev,
           colmap=heat.colors(128),
           colin = predict(beiWithCorrection),
           apron=TRUE,
           theta=-30,
           phi=20,
           zlab="Elevation",
           expand=6,
           visible=TRUE,
           shade=0.3,
           xlab="1000m",
           ylab="500m",
           main = '')

image.plot(legend.only=T,
           zlim=range(predict(beiWithCorrection)),
           col=heat.colors(128),
           horizontal=TRUE)
```

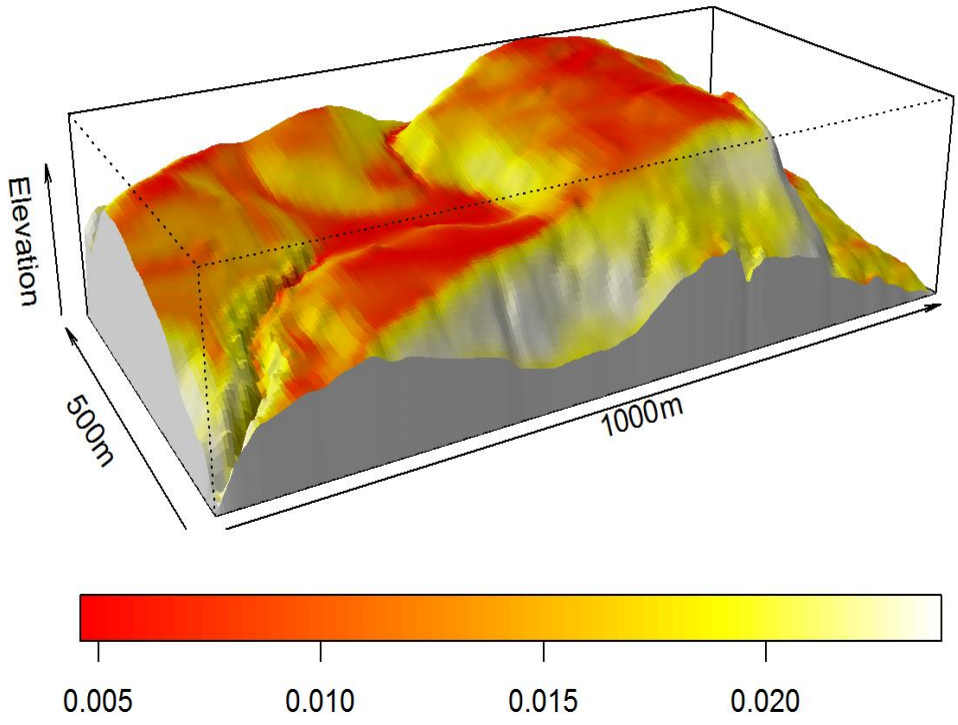

Fig. S7b Expected number of points (locations of *Beilschmiedia pendula*) predicted by model fitted with correction for projection to the x-y plane

```
# Fig S7c - Change in estimated intensity with and without topographic correction
plot((predict(beiWithCorrection)-
predict(beiNoCorrection))/predict(beiNoCorrection), main = '')
```

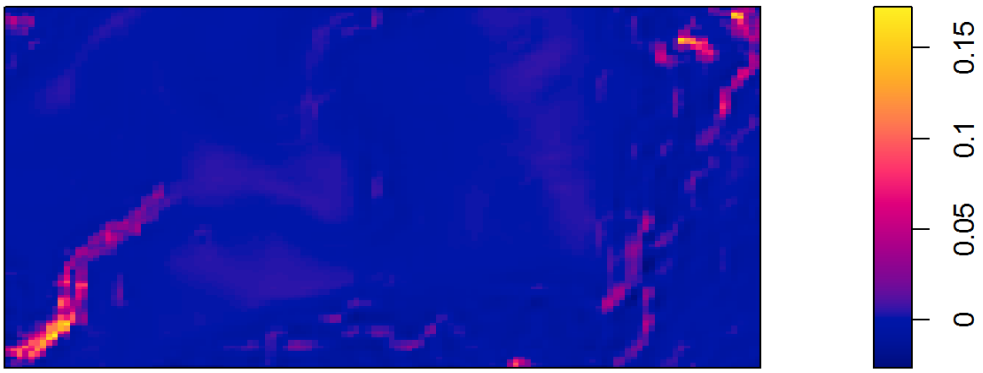

Fig. S7c Proportional change in expected intensity between model with correction for projection and model without.

```
# Fig S7d - Change in estimated intensity with and without topographic correction

M <- persp(bei.extra$elev,
           colmap=heat.colors(128),
           colin = (predict(beiWithCorrection)-
predict(beiNoCorrection))/predict(beiNoCorrection),
           apron=TRUE,
           theta=-30,
           phi=20,
           zlab="Elevation",
           expand=6,
           visible=TRUE,
           shade=0.3,
           xlab="1000m",
           ylab="500m",
           main = '')

image.plot(legend.only=T,
           zlim=range((predict(beiWithCorrection)-
predict(beiNoCorrection))/predict(beiNoCorrection)),
           col=heat.colors(128),
           horizontal=TRUE)
```

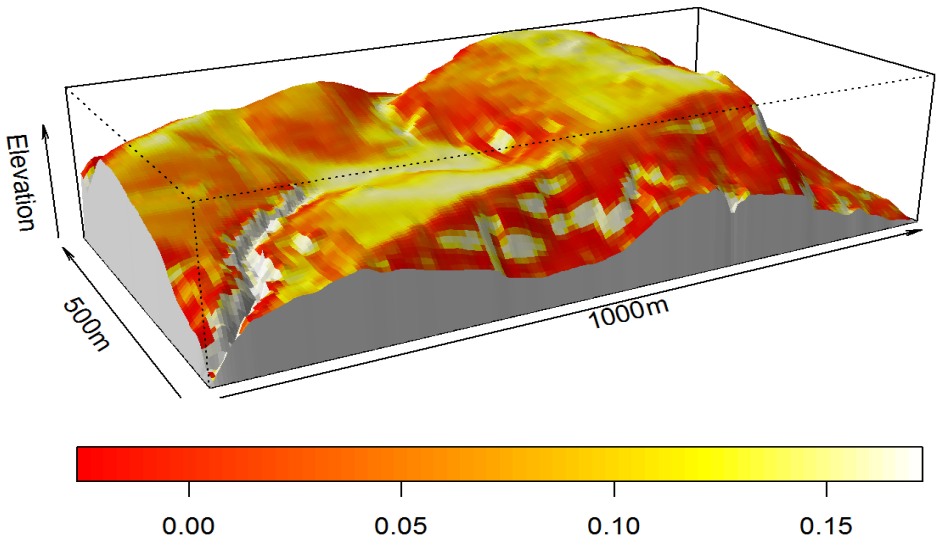

Fig. S7c 3D representation of proportional change in expected intensity between model with correction for projection and model without.

Distribution of correction factors required in real datasets

```
#Synthetic surfaces
STEPFun = function(x,y){
  if(x < 2){return(4)}
  else if(x>=2 & x<4){return(4 + (x-2)*1)}
  else if(x>=4 & x<6){return(6)}
  else if(x>=6){return(6 + (x-6)*2)}
}

STEPFun <- Vectorize(STEPFun)

SINFun = function(x,y){
  return(sin(x)*2)
}

EXPFun = function(x,y){
  return(exp(x/4))
}

step=as.im(STEPFun,W=owin(xrange=c(0,8),yrange=c(0,2)),eps=0.01)
stepSA = as.im(surfaceArea(as.matrix(step),
                           cellx=step$xstep,
                           celly=step$ystep,
                           byCell=TRUE),
               W=owin(xrange=c(0,8),yrange=c(0,2)))

stepCF = stepSA/(step$xstep*step$ystep)
stepCF = as.im(stepCF,W=as.owin(step))

SIN=as.im(SINFun,W=owin(xrange=c(0,8),yrange=c(0,2)),eps=0.01)
SINSA = as.im(surfaceArea(as.matrix(SIN),
                          cellx=SIN$xstep,
                          celly=SIN$ystep,
                          byCell=TRUE),
              W=owin(xrange=c(0,8),yrange=c(0,2)))

SINCF = SINSA/(SIN$xstep*SIN$ystep)
SINCF = as.im(SINCF,W=as.owin(SIN))

EXP=as.im(EXPFun,W=owin(xrange=c(0,8),yrange=c(0,2)),eps=0.01)
EXPSA = as.im(surfaceArea(as.matrix(EXP),
                          cellx=EXP$xstep,
                          celly=EXP$ystep,
                          byCell=TRUE),
              W=owin(xrange=c(0,8),yrange=c(0,2)))

EXPCF = EXPSA/(EXP$xstep*EXP$ystep)
EXPCF = as.im(EXPCF,W=as.owin(EXP))

histDFEXP = as.data.frame(EXPCF)
histDFEXP$dataset="EXP"
histDFSIN = as.data.frame(SINCF)
histDFSIN$dataset="SINE"
histDFSTEP = as.data.frame(stepCF)
```

```
histDFSTEP$dataset="STEP"

# Real datasets
histDFGOR = as.data.frame(1/cos((pi*gorillas.extra$slopeangle)/180))
histDFGOR$dataset="Gorilla"
histDFBEI = as.data.frame(1/cos(atan(bei.extra$grad)))
histDFBEI$dataset="Bei"
histDFFIRE = as.data.frame(1/cos((pi*clmfires.extra$clmcov100$slope)/180))
histDFFIRE$dataset="CLM Fires"
hhistDF = rbind(histDFEXP,histDFSIN,histDFSTEP,histDFGOR,histDFBEI,histDFFIRE)
hhistDF$dataset = factor(hhistDF$dataset,
                         levels = c("EXP","SINE","STEP","Bei","CLM
Fires","Gorilla"))
```

```
# Fig S8 - Distribution of correction factors for explored surfaces
ggplot(data=hhistDF, aes(value,fill=dataset))+
  geom_histogram(bins=100)+
  facet_grid(dataset~.,scales="free_y")+
  theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())+
  guides(fill=guide_legend(title="Dataset"))
```

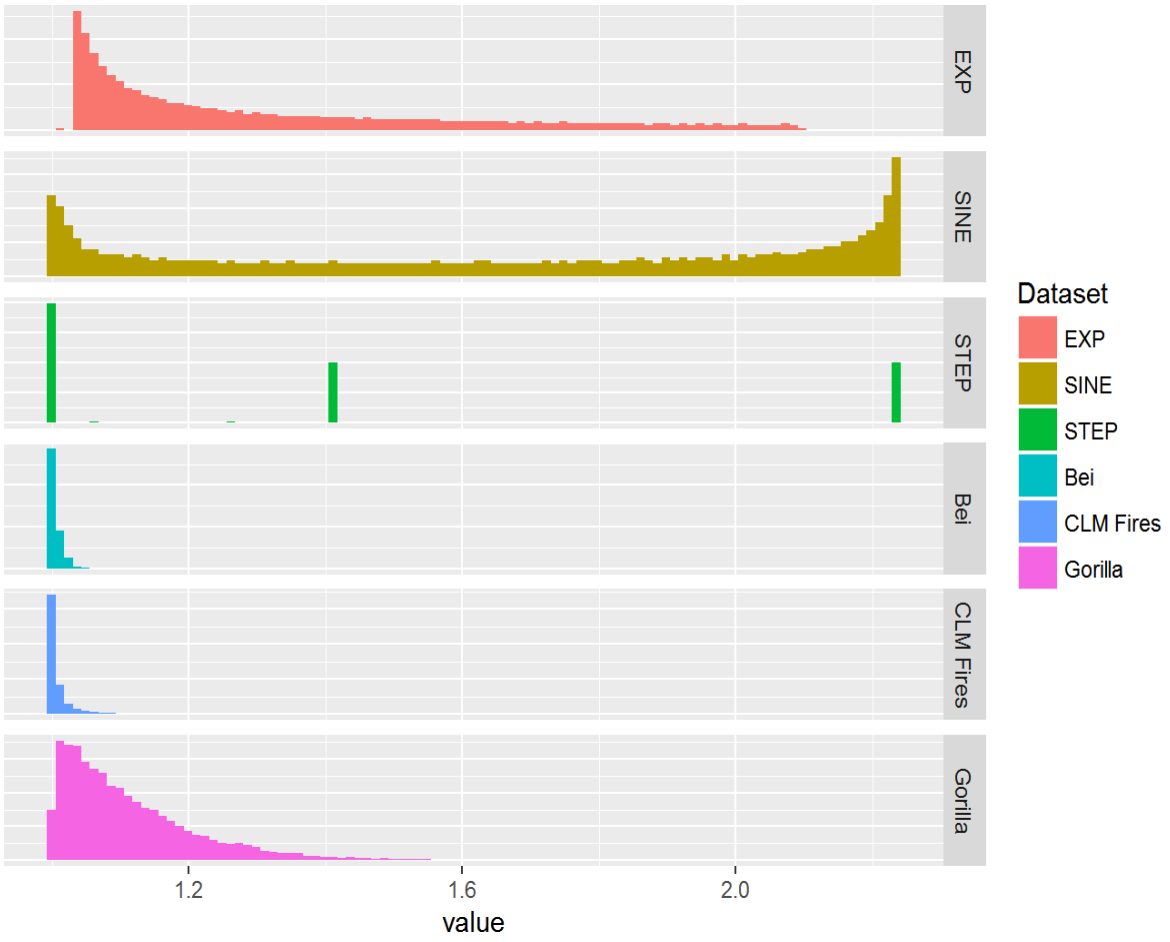

Fig. S8 Comparison of the distributions of correction factors in synthetic surfaces used in previous examples and the correction factors required for several real datasets from the spatstat package.

Effects of Slope

The following code generates a set of homogeneous Poisson point patterns over a surface in which half of the window is in the x-y plane, while the other half of the surface is a sloped plane set at a variable angle (theta) from the x-y plane. For each pattern we test whether we would reject a null of CSR at a level of 0.05, using the Maximum Absolute Deviance test.

```
#convert degrees to radians for use in trigonometric functions
toRadians = function(theta){
  return((theta*pi)/180)
}

#Return a vectorized function with arguments x,y which can be passed to rpoispp
lam_by_slope = function(theta){
  lam = function(x,y){
    #return ppp generator
    if(x<=4){
      return(10)
    }
    else{
      return(10*(1/cos(toRadians(theta))))
    }
  }
  return(Vectorize(lam))
}

iters=500
theta = 1:60
result = vector(length=60)
for(i in theta){
  for(j in 1:iters){
  result[i] = result[i] + (mad.test(X =
rpoispp(lam_by_slope(i),win=owin(c(0,8),c(0,8))),nsim=99,verbose=F)$p<=0.05)
  }
}
```

```
# Fig S9 - Effect of partial inclined plane on rates of reject CSR null
plot(x=theta,y=result/iters,xlab="Angle of Inclined Plane",ylab="Failure
Rate",type='b',main = '')
abline(h=0.05,col="green",lty=2)
```

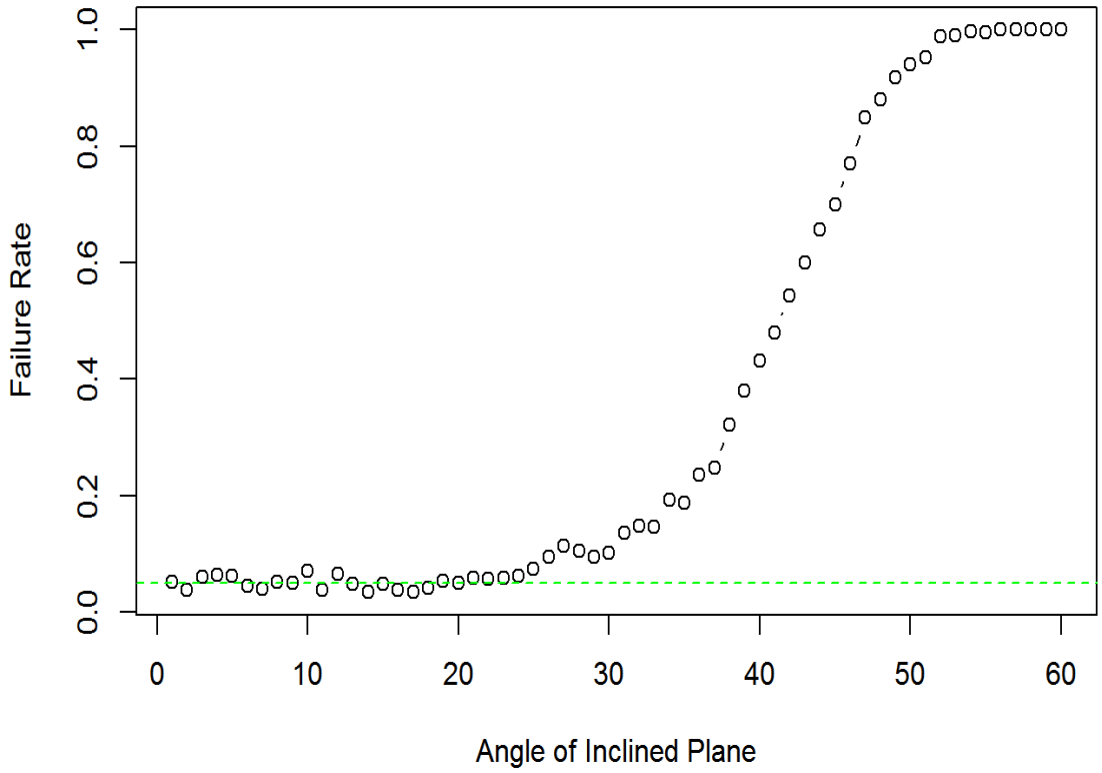

Fig. S9 Rate at which null of CSR (in x-y plane) would be rejected under the Maximum Absolute Deviance test if generating point process acts on a surface comprising a planar area and an inclined plane.

Algorithms for generating point processes on surfaces in stored as 3D Meshes

An alternate way of storing topographic information is in the form of a polygonal mesh, where a surface is represented by a set of point locations and relationships between these points together form the faces of the surface. One of the advantages of storing topographic data in this form is that the density of vertices within the mesh is variable and as such information storage is more efficient, and the mesh can more closely approximate the continuous surface it is intended to describe.

Poisson point processes may be simulated directly on the faces of the mesh using a barycentric coordinate system, providing correctly distributed points with x, y and z components.

1. For each face in the mesh, calculate the area A_F of the polygon
2. Generate a random variable N such that $N \sim Poisson(\lambda A_F)$
3. Define vectors $(\overrightarrow{ab})$ and $(\overrightarrow{ac})$ from one vertex of the triangle (a) to vertices b and c
4. Generate N pairs of random variables u and v such that $u, v \sim Uniform(0,1)$
5. If $u + v > 1$, then $u = 1 - u$, $v = 1 - v$
6. Calculate point positions as $\vec{a} + (\overrightarrow{ab})u + (\overrightarrow{ac})v$

These points can be converted to a format suitable for existing point pattern analysis packages by producing an orthographic projection to the x-y plane. Alternatively we can calculate the areas of each polygon in their original x,y,z form, then project these polygons to the x-y plane and form a correction factor between projected and real areas.

Wavefront .OBJ file format (.OBJ)

The .OBJ file format stores the structure of the mesh as a plain-text series of vertex positions and edges which group vertices into faces. The .OBJ format is commonly used in 3D modelling software and may be exported from a range of photogrammetry packages. With the increased availability and use of unmanned aerial vehicles equipped with cameras we foresee photogrammetrically derived topography datasets becoming more prevalent in ecology. The following python code will read a .OBJ format mesh and generate a Poisson point pattern across the faces of the mesh.

NOTE: While the .OBJ format allows the faces of meshes to be any arbitrary n-sided polygon, this code requires a mesh with triangular faces. Meshes can be triangulated using freely available 3D modelling software such as Blender.

```
import numpy as np

class Mesh():
    def __init__(self,filename):
        numFaces = 0
        verts = []
        polygons = []
        area = []
        for line in open(filename, "r"):
            vals = line.split()
            if vals:
                if vals[0] == "v":
                    v = map(float, vals[1:4])
                    verts.append(np.array(v))
                if vals[0] == "f":
                    vertices = map(Mesh.line_split, vals[1:4])
                    print [verts[i] for i in vertices]
                    polygon = {'vertices': vertices,
                               'area':Mesh.polygon_area([verts[i] for i in ver
tices])}
                    polygons.append(polygon)
                    numFaces += 1
            self.polygons=polygons
            self.verts = verts

    @staticmethod
    def line_split(line):
        return int(line.split("/")[0])-1

    @staticmethod
    def unit_normal(a, b, c):
        x = np.linalg.det([[1,a[1],a[2]],[1,b[1],b[2]],[1,c[1],c[2]]])
        y = np.linalg.det([[a[0],1,a[2]],[b[0],1,b[2]],[c[0],1,c[2]]])
        z = np.linalg.det([[a[0],a[1],1],[b[0],b[1],1],[c[0],c[1],1]])
        mag = (x**2 + y**2 + z**2)**0.5
        return (x/mag, y/mag, z/mag)

    @staticmethod
    def polygon_area(polygon):
        if len(polygon) < 3:
            return 0
        total = [0, 0, 0]
        for i in range(3):
            a = polygon[i]
            b = polygon[(i+1) % 3]
            crossProd = np.cross(a, b)
            total[0] += crossProd[0]
            total[1] += crossProd[1]
            total[2] += crossProd[2]
            area = abs(np.dot(total, Mesh.unit_normal(polygon[0], polygon[1],
```

```
polygon[2]))/2)
        return area

    def pointProcess(self, lam):
        points = []
        for face in self.polygons:
            area = face['area']
            nPoints = np.random.poisson(area * lam)
            a, b, c = tuple([self.verts[x] for x in face['vertices']])
            ab = b - a
            ac = c - a
            for point in range(0, nPoints):
                r1, r2 = np.random.uniform(0, 1, 2)
                if r1 + r2 >= 1:
                    r1 = 1 - r1
                    r2 = 1 - r2
                point = a + r1 * ab + r2 * ac
                points.append(point)
        return (points)
```

ESRI .TIN

The TIN format produced by ESRI applications such as ArcScene is similar to the .OBJ however the format is proprietary and ESRI tools must be used to work with data in this form.

1. Run the TIN Triangle tool from the `3D Analyst` toolbox to create a shapefile of the projections of the faces of the TIN.
2. Open the attribute table for the newly create feature layer. Add a field called `NRand` to the attribute table and open the field calculator. Select Python as the parser. In the Pre-Logic Script Code block add `import numpy as np`. In the `NRand =` block add `int(np.random.poisson(lambda,1))*!Shape_Area!`, replacing lambda with an integer representing the desired intensity.
3. Run the Create Random Points tool from the Data Management toolbox. Select the projection of the TIN as the Constraining Feature Class and set Number of Points to Field and select NRand.