

Using the beta distribution to analyse plant cover data

Christian F. Damgaard¹  | Kathryn M. Irvine² 

¹Bioscience, Aarhus University, Silkeborg, Denmark

²Northern Rocky Mountain Science Center, U.S. Geological Survey, Bozeman, Montana

Correspondence

Christian F. Damgaard
Email: cfd@bios.au.dk

Funding information

US Geological Survey; National Park Service, Grant/Award Number: IAA P12PG70586; US Fish and Wildlife Service, Grant/Award Number: IAA 450063656

Handling Editor: Iain Stott

Abstract

1. Most plant species are spatially aggregated. Local demographic and ecological processes (e.g. vegetative growth and limited seed dispersal) result in a clustered spatial pattern within an environmentally homogenous area. Spatial aggregation should be considered when modelling plant abundance data.
2. Commonly, plant abundance is quantified by measuring cover within multiple areal plots or along multiple lines randomly placed within a study area. A common practice for analysing plant cover is to use statistical methods that rely on the normal distribution for quantifying uncertainty. This is problematic because plant cover data tend to be left-skewed (J-shaped), right skewed (L-shaped) or U-shaped and, therefore, commonly violate classic statistical assumptions, such as normality.
3. We outline statistical analyses that explicitly account for spatial aggregation by assuming that plant cover is beta-distributed. The beta distribution is a flexible choice because within the open unit interval it can take on a wide range of shapes (L, J, U, or a bell-shaped). We discuss and introduce extensions to the beta distribution that address common analysis issues encountered in plant cover datasets, such as (a) the treatment of zero and one cover values, (b) hierarchical data structures and (c) observations errors. For heuristic purposes, we focus on single species analyses, but we demonstrate how the outlined methods can be generalized to more species.
4. The assumption that plant cover is beta-distributed allows us to estimate the degree of spatial aggregation, and the ecological significance of this new knowledge is discussed. We provide a summary of available software for analyses (emphasizing standard R packages) and include worked examples and a simulation study comparing analysis options as supplemental information.
5. *Synthesis.* Previously, the state of the statistical software made it practically difficult for empirical plant ecologists to analyze their cover data correctly, but new theory and R-packages have been developed, and this difficulty no longer exists. We recommend that empirical plant ecologists embrace the new statistical possibilities for exploring the exciting ecological features in spatial variation of plant cover.

KEYWORDS

beta distribution, hierarchical models, ordinal cover data, pinpoint cover data, plant cover, spatial aggregation, statistical models

1 | INTRODUCTION

Most plant species are spatially aggregated. Vegetative growth and limited seed dispersal result in a clustered spatial pattern within a homogenous environment (Pacala & Levin, 1997; Watt, 1947). Commonly, plant abundance is measured using a pre-defined spatial unit, a line or area for plot-based sampling. Then, multiple plots or lines are randomly placed within an environmentally homogeneous area. In this context, plant abundance data typically displays a U-shaped distribution for dominating spatially aggregated species. That is, a bulk of the plots or lines land within a patch and the measured abundance is larger compared to the average abundance at the site, or they randomly fall outside a patch and the measured abundance is lower than the site average. On the other hand, following the same rationale, less dominating spatially aggregated plant species display empirical distributions that are right-skewed (L-shaped). Most plots have zero or small abundance recorded, but occasionally a plot will intersect a patch and the recorded values are large. The common practice for analyzing plant abundance is to use statistical methods that rely on the normal distribution for quantifying uncertainty (e.g. Elzinga, Salzer, & Willoughby, 1998). Here, we review statistical methods that rely on the beta distribution as an alternative for analyzing plant abundance data.

The most commonly used measure of plant abundance is cover, which is the relative area of the species when projected onto the surface (Figure 1). Different monitoring programmes have generated large collections of cover datasets that span many years and/or large spatial extents (e.g. Nielsen et al., 2012). In a survey of 66 published studies where plant cover was measured (search criteria

used provided in Appendix S1), we found that most studies used multi-species ordination methods for analyzing their cover data, and normally or log-normally distributed data was implicitly assumed in six out of seven studies where single-species analyses were of interest. Additionally, outside of the peer-reviewed literature, gray literature and accessible web-resources are prompting the use of incorrect statistical approaches for analyzing plant data. For example, in an often cited guidance manual for measuring and monitoring plant populations (Elzinga et al., 1998), it is recommended to analyze plant cover data by ANOVA or *t*-tests. This common practice is problematic due to the spatial aggregation of plant species. Even transforming the cover data is not sufficient in all instances. The standard statistical model assumptions of linearity, no outliers, constant variance and residuals that are consistent with a bell-shaped normal distribution are often violated for these data. In this account, we focus on plant cover, but the statistical issues discussed here will be relevant to other types of plant abundance measures as well. The spatial patterns exhibited in plant cover datasets can be accommodated within a statistical model by assuming a beta distribution. The beta distribution is a flexible option because within the open unit interval ($0 < \text{plant cover} < 1$), it can take on a wide range of shapes (L, J, U, or a bell-shaped), and there are extensions to the beta distribution that accommodate datasets that contain zero and one values (Ospina & Ferrari, 2010). In fact, the use of the beta distribution to model plant cover data dates back to Robinson in 1954. Since then, a number of empirical and theoretical studies have shown that the underlying distribution of cover at a site is consistent with a beta probability distribution (Chen, Shiyomi, Bonham, et al., 2008; Chen, Shiyomi, Hori, Shiyomi, Hori, & Yamamura, 2008; Chen, Shiyomi, Yamamura, & Hori, 2006; Damgaard, 2008, 2009, 2012, 2013, 2014; Irvine, Rodhouse, & Keren, 2016; Kemp & Kemp, 1956; Robinson, 1954; Wright, Irvine, Warren, & Barnett, 2017). Further, we show the utility of the beta distribution for modelling plant abundance data gathered using different field methods; plot-based, point intercept, line intercept, and pinpoint methods (Box).

We outline beta distribution-based analysis options for the most-often used field methods to quantify areal cover of a species: point-based methods or visually estimated cover (Box). This review will for heuristic purposes focus on the analysis of single species cover data from randomly placed plots at a single environmentally homogeneous site, but it will be demonstrated how the outlined methods may be generalized to account for environmental gradients, hierarchical sampling designs, and multiple species. Further, we demonstrate how the beta distribution can be used to analyse ordinal cover data avoiding incorrectly using the mid-point value of the ordinal intervals, and an extension to beta regression that accounts for observation errors, an emergent issue for field-based plant studies. Previously, statistical methods available for empirical plant ecologists were limited, but now suitable R-packages exist for implementing our proposed beta distribution-based methods; in our review, we provide a summary of these software solutions (using standard R packages). We include worked examples and accompanying R code (Irvine, 2019).

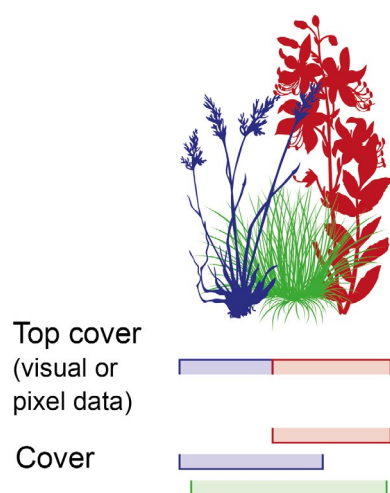


FIGURE 1 Plant cover is defined as the relative area of the species when projected onto the surface and should not be confused with top cover (Wilson, 2011). Note that the green plant species has a sizeable cover, but has a top cover of zero. Often, it is the top cover that is measured when cover is assessed visually in relatively large plots, or when plant cover is estimated from pixel data. The sum of the cover values of all species may be larger than one, whereas the sum of the top cover values of all species is less than or equal to one [Colour figure can be viewed at wileyonlinelibrary.com]

BOX 1 Different methods of measuring plant cover

Plant cover is defined as the relative area of the species when projected onto the surface (Figure 1). Unlike plant counts or density, plant cover takes the size of individuals into account and is an important and often measured characteristic of the composition of herbal plant communities (Bonham, 1989; Kent & Coker, 1992). Different monitoring programmes have generated large collections of temporal and spatial plant cover datasets (e.g. Nielsen et al., 2012).

The most common way to measure plant cover in herbal plant communities is to make visual assessments of the area covered by the different species in a small circle or quadrat. Often, the visual estimates of cover percentages are categorized using different ordinal classification schemes (Table 1). For example, a common ordinal scale is from Daubenmire (1959) in which cover class 0: small value, cover class 1: >0%–5%, cover class 2: >5%–25%, cover class 3: >25%–50%, cover class 4: >50%–75%, cover class 5: >75%–95%, and cover class 6: >95%–100%.

The visually assessed cover data has been criticized for being rather subjective, i.e. dependent on the person who makes the observation, and can be quite variable (Crocker & Tiver, 1948; Floyd & Anderson, 1987; Kennedy & Addison, 1987; Klimeš, 2003; Milberg, Bergstedt, Fridman, Odell, & Westerberg, 2008; Vittoz et al., 2010).

A more objective method for measuring plant cover is the pinpoint (or point-intercept) method in a frame or along a line, which has been widely employed. In the pinpoint method, a thin pin is inserted vertically into the vegetation a number of times in a fixed design, and the cover of a species is measured by the proportion of the inserted pins that touches the species (Damgaard, 2009; Levy & Madden, 1933; Lindquist, 1931). The pinpoint method is not relevant to measuring the abundance of rare species and has been shown to underestimate species richness (Bråkenhielm & Qinghong, 1995).

One of the advantages of the pinpoint method is that it is possible to aggregate taxa at the pin level. For example, if we want to aggregate the cover of two species, then we simply count the number of pins that hit either one or both of the species and that measure the estimate of the cover of the two species. A related method to the point-intercept is the line intercept in which the proportion of a fixed length line that intercepts a species is recorded. Choices regarding the line width and length should be considered, as they influence how well the empirical value represents true plant cover (Elzinga et al., 1998, pp. 112–113).

In the near future, it is expected that many plant cover measurements will be automated using drone image analysis. This will enable plant top cover measurements in more and larger plots, but also with increased taxonomic uncertainty.

TABLE B1 Different plant cover ordinal classification schemes. The column ordinal classes show the cutoff points for the intervals that define the different ordinal cover classes

Scheme	Ordinal cover classes	Reference
Braun-Blanquet	0.01, 0.05, 0.25, 0.5, 0.75	Braun-Blanquet, 1964
Daubenmire	0.05, 0.25, 0.5, 0.75, 0.95	Daubenmire, 1959
Hult-Sernander	0.0625, 0.125, 0.25, 0.5	Du Rietz, 1930
Domin	0.01, 0.05, 0.1, 0.25, 0.33, 0.5, 0.75, 0.95	Currall, 1987

1.1 | Plant cover process model

We assume that the basic analysis unit for plant cover is based on a pre-defined spatial unit i (e.g. areal plot or line, Box) and spatial units are randomly selected and surveyed within a larger pre-defined domain (or site) j that is ecologically meaningful, such as a wetland, management unit, allotment, etc. ($i = 1, \dots, n_j$ for each site j). To justify the statistical assumption that plant cover measured on a set of plots or lines within a site consists of observations or replicates from the same probability distribution with a common mean and variance parameter, we assume that the underlying environmental conditions are homogeneous and the plant species has the same spatial aggregation within a site j . This assumption can easily be relaxed, as we show later.

When plant cover is visually estimated within a plot i or by the line-intercept method along a line i within a site j (Box and Table 2), the response denoted as Y_{ij} is a continuous proportion that is bounded between zero and one. This type of response is appropriately modelled using a beta distribution with mean cover parameter μ_j and shape parameter δ ,

$$[Y_{ij} | \mu_j, \delta] \sim \text{Beta}(\mu_j/\delta - \mu_j, (1 - \mu_j)(1 - \delta)/\delta), \quad (1)$$

where $\text{Beta}(\cdot, \cdot)$ refers to the classical parameterization of the beta-distribution with the two shape parameters α and β (Figure 2). Both the mean cover parameter μ_j and shape parameter δ are defined in the open interval $0 < (\mu_j, \delta) < 1$. An important benefit to using the beta distribution is that the shape parameter δ can be interpreted as a measure of spatial aggregation for a given site j . The size of δ is

TABLE 2 Statistical analysis options for different field measurement methods for plant cover

Field measurement method(s)	Response variable type	Statistical model options	Software options	References
Pinpoint Point-intercept	Binomial proportion	Beta-binomial	R-INLA R package VGAM package	Rue, Martino, & Chopin, 2009; Yee, 2015)
		Logit-trans- formed empirical proportions	stats R package (lm function)	Ramsey & Schafer, 2002
		Logistic regres- sion binomial counts	stats R package (glm function)	Ramsey & Schafer, 2002
		Multiple binomial logistic regres- sions for each species	mvabund, boral, or gllvm R packages	Hui, 2016; Niku, Warton, Hui, & Taskinen, 2017; Wang, Naumann, Wright, & Warton, 2012
Visually estimated Cover Classes in areal plot	Ordered categorical	Cumulative logit models Multiple cumula- tive logit models for each species	MASS package (polr function) ordinal pack- age (clm function) mvabund, boral or gllvm R packages	Agresti, 2010; Hui, 2016; Irvine & Rodhouse, 2010; Niku et al., 2017; Wang et al., 2012
		Ordinal beta model	JAGS, Stan, WinBUGS	Irvine et al., 2016
Visually estimated cover in areal plot Line-intercept	Continuous proportion	Beta regression	betareg R package	Cribari-Neto & Zeileis, 2010; Ferrari & Cribari-Neto, 2004
		Multiple beta regressions for each species	boral R package	Hui, 2016

not a species-specific constant, but depends on the size and shape of the spatial unit i used to record data in relation to the true species spatial pattern within a site, as is true for any measure of spatial aggregation. For example, when the size of a plot increases, then the probability that the plot will cover both patch and non-patch areas of a spatially aggregated plant increases as well, and δ is expected to decrease (Figure 2c: $\hat{\delta}=0.68$, Figure 2d: $\hat{\delta}=0.36$). The fit of the beta distribution for two examples of empirical plant cover data is shown in Figure 2e and f.

The relationship between the parameterization (Equation 1) and another parameterization of the beta distribution, $Beta(\mu_j\phi, \phi - \mu_j\phi)$, where ϕ is a precision parameter (Ferrari & Cribari-Neto, 2004), is $\delta=1/(1+\phi)$. As spatial aggregation decreases ($\delta \rightarrow 0$), precision increases ($\phi \rightarrow \infty$) and the among-plot variation in plant cover decreases, meaning the measured cover values within a site approach a common value.

The parameter of interest of the underlying beta-distribution of plant cover (Equation 1) will, in most empirical studies, be the mean cover, μ_j . Typical research questions are whether the mean cover of a specific plant species (a) differs among sites as related to management practices, (b) is related to environmental or disturbance gradients, and (c) changes over time. In the latter section, *comparison of beta models to naïve linear models with normally distributed errors* and Appendix S2, we show how these questions can be addressed using a beta distribution for visual estimates and ordinal plant cover datasets (Irvine, 2019).

1.2 | Extension to process model that accounts for zero and one cover observations

Commonly, zero cover values are recorded in field-based plant studies. In fact, it is expected that some species will be absent from a site due to large-scale extinction and colonization dynamics (MacArthur & Wilson, 1967). In some cases, for example, in studying the invasion of a nonnative species and depending on the spatial extent of a site, there may be parts of the site or the entire site where the species is not found, manifesting in an overabundance of zeros (Irvine et al., 2016). Sometimes, the zero and one cover values are omitted from the dataset, but this practice will clearly lead to a biased mean cover estimate and should be avoided. Instead, a zero-augmented beta distribution can be used (Ospina & Ferrari, 2010). Basically, the probability of observing zero cover is modelled using logistic regression. More formally, the zero values are modelled by an independent Bernoulli process $U_i \sim Bernoulli(\psi)$, where ψ is the probability of a species occurring in a plot and abundance or percent cover is modelled as a beta random variable conditional on the species occurring in a plot, $[Y_i|U_i=1] \sim Beta(\mu, \delta)$. If this model describes the observed patterns in the empirical data, the interpretation is that there is an underlying ecological process that leads to a zero value, and a different process that describes the values between 0 and 1 (non-zeros), as in a two-part or hurdle model (Appendix S2 provides code and examples). A similar approach can be used to model one values, however, in most situations it is unlikely that a species will cover a plot

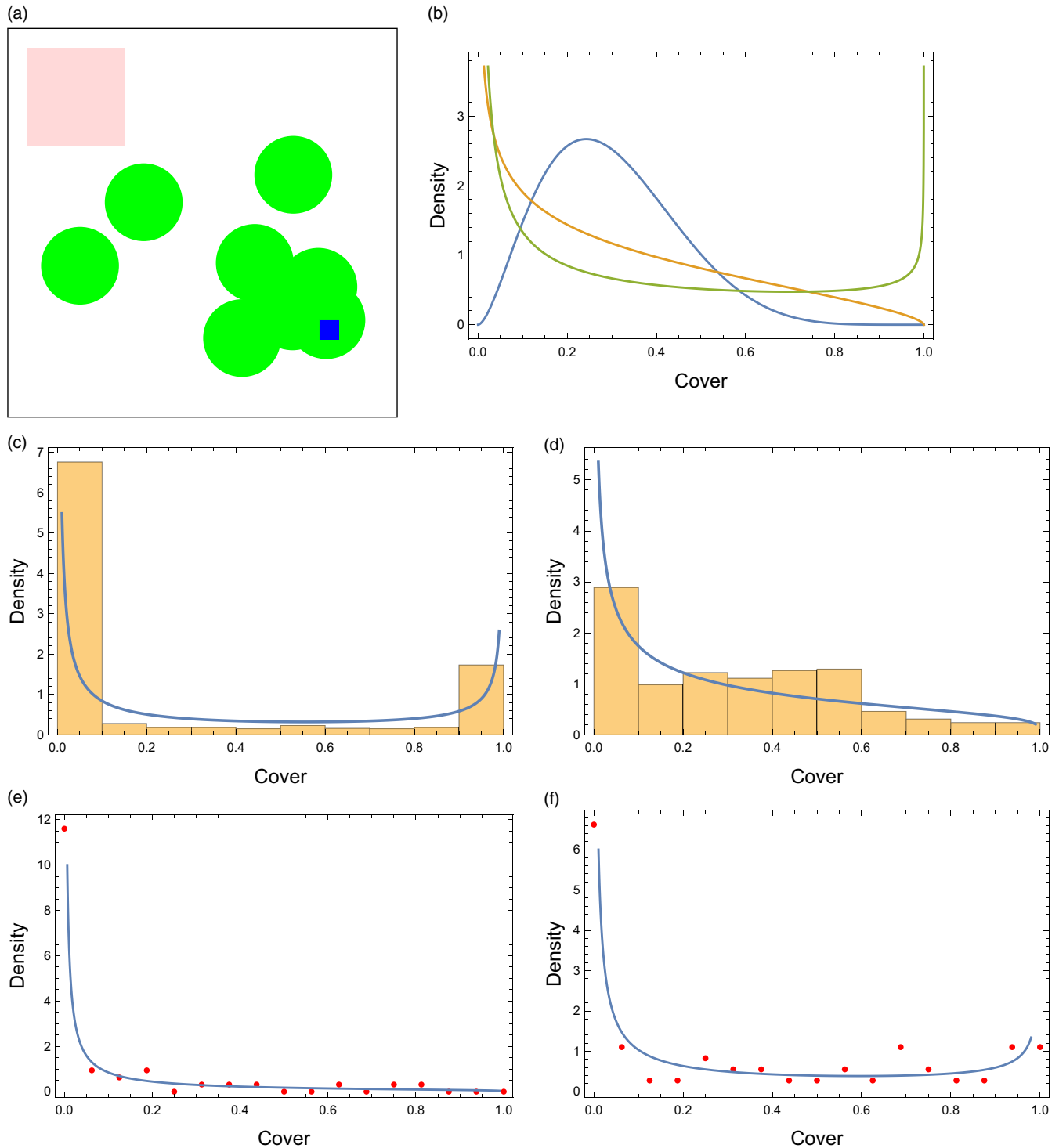


FIGURE 2 (a) Plant abundance of a patchy plant species (green) is measured by its cover using randomly placed relatively small plots (blue square). (b) The beta distribution (Equation 1) is flexible and may be fitted to both unimodal and U-shaped plant cover data. For all lines, mean cover, μ , is 0.3 and the parameter that measures the spatial aggregation, δ , is 0.1 (blue line), 0.3 (orange line) and 0.5 (green line), respectively. (c) The resulting distribution of plant cover from a 1,000 randomly placed small plots (sampling plots area = 1/400: the blue square in (a), blue line: the fitted beta-binomial distribution with $\hat{\mu}=0.33$, $\hat{\delta}=0.68$). (d) The resulting distribution of plant cover from 1,000 randomly placed plots (sampling plots area = 1/16: the light red square in (a), blue line: the fitted beta-binomial distribution with $\hat{\mu}=0.30$, $\hat{\delta}=0.36$). (e) Pinpoint cover samples of *Erica tetralix* from a wet heathland site (red points: pinpoint cover data shown on the cover scale, blue line: the fitted beta-binomial distribution with $\hat{\mu}=0.081$, $\hat{\delta}=0.43$). (f) Pinpoint cover samples of *Calluna vulgaris* from a dry heathland site (red points: pinpoint cover data shown on the cover scale, blue line: the fitted beta-binomial distribution with $\hat{\mu}=0.31$, $\hat{\delta}=0.60$). (The sampling method of the two pinpoint cover examples is explained in Damgaard, 2019b) [Colour figure can be viewed at wileyonlinelibrary.com]

completely and a dataset dominated by one values could indicate an issue with the study design and the chosen plot size. A dataset that includes plots with zero cover and plots completely covered by a species requires a three-part model as described in Ospina and Ferrari (2010). We provide an empirical example of fitting the zero-and-one inflated beta model (see section Sago Pondweed (*Stuckenia pectinata*) Example: Percent Cover and Appendix S2 for R code).

Alternatively, the zero cover values may arise from the same ecological process as the non-zero values. For example, sampling a highly aggregated species with low average cover within a homogenous site could result in a high probability of observing very small cover values (Figure 2a). In this case, zero values may be adequately modelled by assuming a common U-shaped beta distribution (Equation 1). In practice, this may be done by adding a small trace value to the observed zero cover values. If the zero and non-zero cover values are sufficiently (i.e. no major evidence of assumption violations) modelled using the two-parameter beta distribution suggested in Equation 1, there may be interpretational and computational advantages using a simpler model.

For example, in a study of *Erica tetralix* at 84 sites that were classified as wet heathland, where plant cover was measured by the pin-point method in a total of 1,148 plots, it was found that the variation in site mean cover was adequately modelled by the beta distribution and that the data did not support an independent zero-generating process (Damgaard, 2013). Since *E. tetralix* is a characteristic species of wet heathlands in this case, it was expected that there would be no indication of an independent zero generating process. It is always possible to evaluate whether there is a need for explicitly modelling the zeros separately from the non-zero values by using a formal maximum likelihood ratio test or residual-based model assessments (see Appendix S2 for worked examples).

In some settings, the more complex model (zero-augmented beta distribution) provides insights into the two ecological processes of distribution and abundance (Irvine et al., 2016; Wright et al., 2017). For example, cheatgrass (*Bromus tectorum*) is a notorious invader in the Western USA. In an analysis of cheatgrass cover in John Day Fossil Beds National Monument, a relatively intact protected area; there was evidence of an overabundance of zeros, likely because certain areas were yet to be invaded. By explicitly modelling the zero process and the abundance process separately, there was evidence that cheatgrass was less likely to occur in steep north facing areas, whereas abundance was higher in flatter areas of the monument. These statistical associations suggest potential ways for managers to prioritize areas in the monument for restoration activities to contain the current invasion or early detection efforts to prevent invasion (Rodhouse et al., 2014).

1.3 | Ordinal cover data

In many plant studies, visually estimated percent cover of a bounded quadrat is binned into a priori defined intervals or cover classes (see Box). A common practice is to convert the observed cover classes to the corresponding mid-point value. For example,

a plot with a recorded cover class or interval between 0% and 5% is replaced with 2.5%. These mid-point values are then used as the response and analysed with multiple linear regression or analysis of variance, this approach is inadequate for many reasons that are a consequence of the bounded nature of percent cover (see Agresti, 2010, 5–7 for a thorough treatment). For example, the assumption of constant variance is violated because the response variable is naturally heteroscedastic, with less variation near the lowest and highest cover class relative to moderate cover values (Agresti, 2010, p. 5). Further, some researchers will assume that averaging mid-point values from many plots is reasonable, likely loosely invoking the central limit theorem. This approach makes the tenuous assumption that the true cover values, if observed, are close to the cover class mid-point value. Potentially, averaging the mid-points can severely over- or underestimate the true average percent cover if the true cover values are close to the lower or upper boundary of the interval class.

The visually assessed cover data may also be reported simply by the assessed percent cover, but, typically, the reported values are biased towards numbers that are divisible by five or 10. For example, values like 32% are less frequently recorded than 30%, and, consequently, the reported percent cover values also have some more or less well-defined ordinal classification intervals, and it would be more accurate to partition the assessed percent cover into ordinal cover classes that reflect the field sampling resolution.

Cover class data are inherently an ordered categorical variable and could be appropriately modelled using proportional odds models (Irvine & Rodhouse, 2010). Alternatively, more recently, a model was developed based on using the beta distribution as a latent variable that is “chopped up” into intervals or cover classes. This model allows for interpretation of the ecologically meaningful beta parameters (μ, δ) instead of odds ratios of cumulative category probabilities, as is done in the proportional odds model (Irvine et al., 2016). The desire to interpret the data on the percent cover scale can be accomplished by using the model proposed by Irvine et al. (2016). Further, Irvine et al. (2016) directly model the zero cover class as opposed to assuming all zeros are a small trace value. We compare the naïve mid-point approach to the ordinal beta model of Irvine et al. (2016) with an empirical dataset in the section *Comparison of beta models to naive linear models with normally distributed errors* and the code to recreate the analyses are available in Appendix S2.

1.4 | Pinpoint or point-intercept cover data

When plant cover of a spatially aggregated species at an environmentally homogenous site, j , is beta distributed, it follows from (Equation 1) that for pinpoint or point-intercept cover data (see Box) with observed response, Z_{ij} , the number of hits in a frame or along a line with n possible hits are beta-binomial distributed with the probability distribution function,

$$Betabin(Z_{ij}; n, \mu_j, \delta) = \binom{n}{Z_{ij}} \frac{\varphi\left(\mu_j \left(\frac{1}{\delta} - 1\right), Z_{ij}\right) \varphi\left(\frac{(1-\mu_j)(1-\delta)}{\delta}, n - Z_{ij}\right)}{\varphi\left(\frac{1}{\delta} - 1, n\right)} \quad (2)$$

where φ is the Pochhammer function, $\varphi(x, n) = \Gamma(x+n)/\Gamma(x) = x(x+1)\dots(x+n-1)$ (Damgaard, 2009). This can be written hierarchically to show the connection to the beta distribution as $[Z_{ij}|Y_{ij}] \sim \text{Binomial}(n, Y_{ij})$ and $[Y_{ij}|\mu_j, \delta]$ as in Equation 1. Notice that zero and one cover values are within the domain of the beta-binomial distribution (Equation 2).

The beta-binomial distribution (2) may be generalized to the Pólya-Eggenberger distribution that is derived from a Pólya urn model (Qu, Greene, & Piedmonte, 1993) and has the same probability distribution function as the beta-binomial distribution. The advantage of using the Pólya-Eggenberger distribution is that δ may take negative values, $\delta > -\text{Min}(\mu/(n-1-\mu), (1-\mu)/(n-1))$, which may be relevant to the rare cases when plant species are more regularly positioned than random (e.g. Hovanes, Harms, Gagnon, Myers, & Elder, 2018). When $\delta = 0$, the Pólya-Eggenberger distribution degenerates to the binomial distribution with probability parameter μ_j . The shape parameter δ that measures the degree of spatial aggregation has the additional property in the Pólya-Eggenberger distribution that it is a measure of the intra-plot correlation among pin hits (Qu et al., 1993).

We recommend using the beta-binomial distribution (Equation 2) for modelling pinpoint or point-intercept data because of its ability to account for and estimate spatial aggregation using the δ parameter. However, alternatives, such as using a logit-transformation of the empirical proportions or logistic regression for binomial counts (Table 2), may be sufficient in some cases. For completeness, we include worked examples for all of these approaches (Table 2, Appendix S2).

1.5 | Solutions for typical complications in the analysis of plant cover

1.5.1 | Observation errors

When plant top cover is visually estimated, this field method is subject to two types of observation errors: measurement and/or detection errors. A measurement error is when a recorded value differs from the true value. A detection error is when an observer does not see a plant species, even when it is present within a plot. To account for these errors, the survey design should be altered such that multiple observations are available per plot i arising from multiple observers recording independent measurements or multiple independent visits to the same plot. Then, a hierarchical formulation can be used to extend the beta model to account for both measurement error and imperfect detection (Wright et al., 2017). An interesting finding is that detection errors impact mean percent cover estimates more negatively than measurement errors because of the inherent connection between plant abundance and detectability. This suggests that for studies of rare or sparsely distributed species, the additional effort of gathering multiple independent observations from the same plot may be warranted to guard against potentially serious

underestimation of occurrence (distribution) and overestimation of abundance (Wright et al., 2017).

1.5.2 | Variable sampling methods

Plant cover is measured using different methods (Box) and with variable plot sizes. Consequently, if a study seeks to compare plant cover among sites that belong to different monitoring programmes, it is to be expected that the sampling method may differ between the sites. Furthermore, in long-term monitoring programmes, it is not unusual that the sampling method changes during the lifetime of the programme. Therefore, it is important to develop statistical methods that allow for an integrated analysis of different types of plant cover data, which is possible by the methodology outlined in Damgaard (2014) under certain simplifying assumptions.

As mentioned earlier, the size of spatial aggregation parameter, δ , is not a species-specific constant, but depends on the size and shape of the sampling plot. Thus, assuming that the degree of spatial aggregation does not vary among sites or with time and plant cover data are measured without systematic errors, then visually determined ordinal plant cover data as well as pinpoint cover data may be fitted into the same underlying beta distribution of plant cover (Equation 1) with mean cover parameter, μ_j , if δ is allowed to depend on sampling scheme (Damgaard, 2014).

1.5.3 | Hierarchical or gradient spatial designs.

Often, a two-stage spatial design is used to survey plants with multiple sites and randomly placed plots within each site. In such designs, mean cover may vary among sites because of historic legacy effects or other unmeasurable factors. The beta model can be extended to accommodate this multi-level structure by including

$$\text{logit}(\mu_j) = \gamma_j, \quad (3)$$

where $\gamma_j \sim \text{Normal}(\mu_\gamma, \sigma_\gamma^2)$ which accounts for potential differences in mean cover among sites j .

For example, the cover of *Erica tetralix* at 23 Danish wet heathland sites was found to decrease significantly in the period between 2004 and 2009 using a spatio-temporal hierarchical model, where the “true” plant cover at each site and year was modelled by a latent variable (Damgaard, 2012).

If species–environment relationships are of interest, plots and/or sites may be chosen to span the gradients of interest, such as elevation or precipitation. These complexities can easily be accommodated by allowing for heterogeneity in the mean parameter by using a logit-link function to include explanatory variables of interest, $\text{logit}(\mu_{ij}) = \mathbf{X}_{ij}\beta$. The mean structure parameterization should be consistent with the resolution of the explanatory variables. For example, if an explanatory variable is the same for each plot within a site, but varies among sites (\mathbf{X}_j), these variables should likely be included in the site-level portion of the model by extending μ_j in Equation 3 (see

Gelman and Hill (2007) for a more thorough review of multi-level or hierarchical models).

1.5.4 | Multi-species

The statistical treatment of whole plant communities with multiple species has for many years relied on different ordination methods that use ad hoc distance measures, for example, Euclidian or Bray–Curtis distances, to reduce the species by sample abundance matrix into a vector of distances between samples. However, there is an increasing interest in the statistical modelling of species abundance matrices (e.g. Clark, Gelfand, Woodall, & Zhu, 2014, Warton et al., 2015). Additionally, it has been suggested that in order to infer ecological processes from observed patterns of species abundance, we need to investigate the covariance in species abundance (e.g. Brown, Law, Illian, & Burslem, 2011) and, if possible, to compare the observed covariance to the expected covariance when spatially aggregated plants associate randomly (Damgaard, Ehlers, Ransijn, Schmidt, & Svenning, 2018).

For multi-species plant cover data, it seems promising to use the Dirichlet distribution as the multi-species generalization of the beta-distribution, where the “true” relative cover of each species is modelled by latent variables, and the multi-species spatial aggregation is modelled with a single parameter (Damgaard, 2015, 2019b; Damgaard, Nielsen, & Strandberg, 2017). Alternatively, multi-species model extensions can be constructed by augmenting the beta distribution (Equation 1) to allow for species-specific mean and spatial aggregation parameters (μ_s, δ_s , where s is for each unique plant species; see Table 2 for R packages available and references for modelling multivariate binomial count, ordinal, and continuous percent cover abundance data), analogous to modelling known species assemblages with detection/non-detection data (e.g. Royle & Dorazio, 2008). Clark, Nemergut, Seyednasrollah, Turner, and Zhang (2017) advocate the use of a generalized joint attribute model (Clark et al., 2017), but this model is not able to account for U-shaped data,

which makes it less applicable to spatially aggregated plant abundance data.

1.6 | Comparison of beta models to naïve linear models with normally distributed errors

As mentioned above, there is substantial empirical evidence to support the use of the beta distribution for plant cover. An important consideration is that the flexible shape of the beta distribution easily handles the typical L- or U-shaped distribution of plant cover data (Figures 2 and 3). Although applying a logit-transformation to percent cover may result in a unimodal distribution, this approach does not necessarily fix the inherent skewness and heavy tails common to empirical datasets (Figure 4, Panel (b)). Furthermore, transforming the response and applying a linear model produces an inherently biased estimator of mean cover compared to directly fitting the beta regression model due to Jensen's inequality (Cribari-Neto & Zeileis, 2010). Since plant cover data are bounded between zero and one, the fitting of cover data with, for example, the normal distribution will, generally, lead to statistical problems when making ecological predictions. However, the use of classic statistical models based on Gaussian errors continues as found in our above-mentioned literature search. Next, we present two real data analyses to show the potentially major interpretational differences between fitting a naïve linear model versus fitting a beta regression model to plant cover data. These worked examples and the accompanying R code to recreate the analyses are included in Appendix S2. Additionally, we compare the beta and more familiar “classic” analysis options in a simulation study (Appendix S3).

1.7 | Sago pondweed (*Stuckenia pectinata*) example: percent cover

First, consider a field study interested in whether there is evidence of a difference in mean plant cover after water was removed within

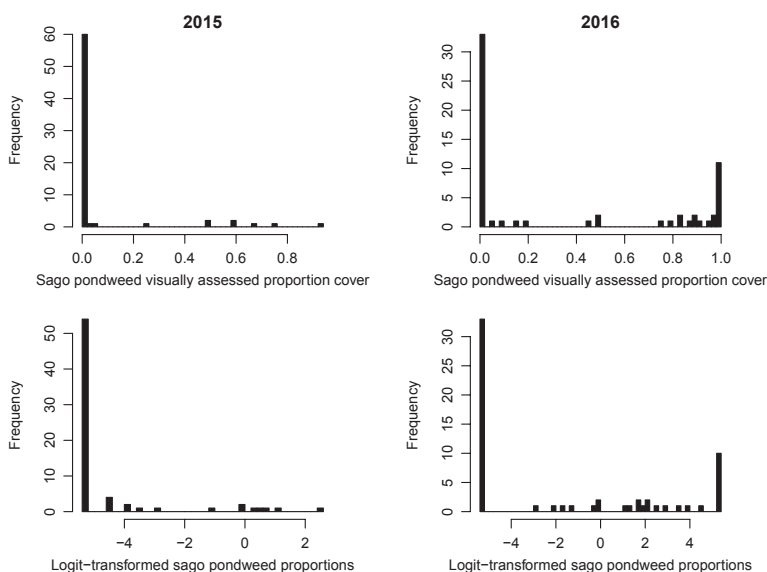


FIGURE 3 Empirical histograms of recorded percent cover values divided by 100, based on visually assessing sago pondweed (*Stuckenia pectinata*) within 1-m × 1-m areal plots in 2015 (left column) and 2016 (right column). Bottom two panels display logit-transformed proportion cover values and demonstrate that in the case of L- or U-shaped distributions transforming the data does not guarantee a unimodal distribution, as is assumed when analyzing the data with a linear model with normally distributed errors

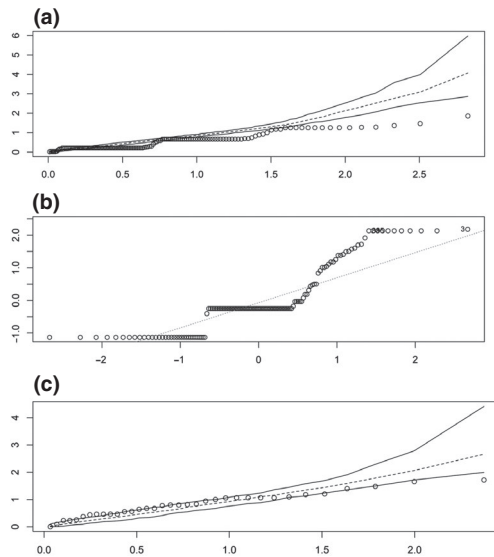


FIGURE 4 Comparison of default qq-plots for beta and linear model objects available using R and code and explanation available in Appendix S2. Beta models use standardized weighted residuals 2 recommended in (Espinheira, Ferrari, & Cribari-Neto, 2008; panel (a) and (c)) and linear model are the response residuals (panel b). Panel (a) is the beta model using a trace value for 0 values and 0.995 for 1 values and assuming a separate δ for each year. Panel (b) is using the logit-transformed response and applying a linear model. Panel (c) is the beta residuals for the zero-augmented beta model assuming a common δ

a wetland unit. Seventy 1-m \times 1-m plots were randomly placed throughout the wetland unit located at Bowdoin National Wildlife Refuge (Montana, USA) in the summer of 2015 and the following summer (2016), after the water was removed, 62 plots were surveyed. The top cover for all submerged aquatic plants was recorded in each plot, here, we focus on only one species, sago pondweed (*Stuckenia pectinata*), which is of management interest because it provides high quality forage for migratory waterfowl. These data were consistent with a L- and then U-shaped distribution, which is unchanged even when a logit-transformation is applied to the percent cover values (Figure 3). Note these data are typical of plant assemblage datasets in that 87 of the 132 plots had zero cover recorded for sago pondweed.

We analyzed the observed proportions with a beta and zero-and-one-augmented beta model and a naïve linear regression model. We also transformed the proportions using a logit function, $\log(p/(1-p))$, and fit a linear regression model. One difference among the different approaches is the interpretation of the model coefficient estimates, an additive shift for a linear model versus multiplicative change in the odds of percent cover (the odds is the ratio of the plot covered to uncovered or $p/(1-p)$) for beta models and the logit-transformed response variable. Also, the beta models allow for exploring whether the spatial aggregation (δ) changed after the wetland dried. Further, we investigated whether there was the need for a four parameter model (zero-and-one-augmented beta), which would

suggest the zero- and one-processes were distinct from the process leading to the continuous percent cover (values greater than 0 and less than 1).

We found that the odds of percent cover increased by 17 times after the wetland dried using the logit-transformed response (95% confidence interval: 6 to 54 times) versus only two times (95% confidence interval: 1.3 to 3 times) based on fitting a beta model with a common spatial aggregation parameter (δ). However, based on an AIC model comparison and the histograms, allowing for δ or spatial aggregation to vary among years seemed sensible (Figure 3). Under this more flexible model, the odds of percent cover were six times higher post-drawdown (95% confidence interval: 3.5 to 10 times) and the spatial aggregation of sago pondweed increased (L- to U-shaped distribution). Using the naïve approach would suggest the mean proportion cover was 0.29 units greater in 2016 after drying (95% confidence interval: 0.17, 0.41). The residual diagnostics suggested that neither the beta models nor the linear models fit the data very well (Figure 4, Panel (a) and Panel (b)). However, if the zero and one cover values were assumed to arise partly due to a zero-one-augmented beta distribution (Irvine et al., 2016; Ospina & Ferrari, 2010; Wright et al., 2017), then the fit markedly improved (Figure 4, Panel (c)).

In the wetland study, applying a zero-augmented beta distribution was insightful because the plots without any sago pondweed (zero values) were potentially dominated by emergent vegetation species. Wetlands are often managed such that the ratio of open water versus emergent habitat is 50:50 (the so-called hemi-marsh). Therefore, modelling sago pondweed presence versus absence provided information on the change in the ratio of open water and emergent habitat after the wetland dried (54 of 70 plots in 2015 and 33 of 62 plots in 2016 with recorded zeros for sago pondweed). We estimated that the odds of zero cover decreased by a factor of 0.33 in 2016 and the odds of sago proportion cover increased by a factor of 4 in 2016 compared to 2015 (see Appendix S2 for details).

1.8 | Cheatgrass (*Bromus tectorum*) example: ordinal cover classes

Next, we consider a long-term monitoring effort of upland vegetation communities in the Clarno unit of the John Day Fossil beds national monument in Oregon USA (Yeo, Rodhouse, Dicus, Irvine, & Garrett, 2009). A random sample of 1-m \times 1-m plots was selected every year for 7 years (minimum sample size was 91 and maximum 220 per year for a total sample size of 1,390). We analyzed observed ordinal cover classes of an invasive annual grass, *Bromus tectorum*, to assess whether there was evidence of a directional change over time (Figure 5). There were only 4% zeros (57 observations) in this dataset; however, other visualization diagnostics suggested a zero-augmented ordinal model would be more appropriate (see Appendix S2 for analytical details). Using the ordinal beta model proposed in Irvine et al. (2016), we found evidence that the probability of detection or presence increased (odds increased 1.3 times per year: 95% credible interval 1.15 to

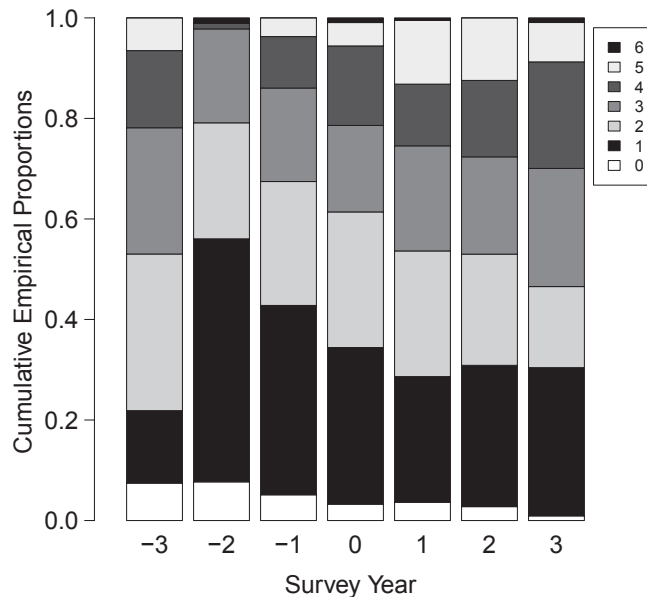


FIGURE 5 Cumulative proportions for observed cover classes of cheatgrass (*Bromus Tectorum*) surveyed in John Day Fossil Beds National Monument in Oregon, USA. The height of each colored bar represents the observed proportion of plots in the corresponding cover class (Y-axis) for a given survey year (X-axis). For example, the proportion of plots recorded as cover class 0 (salmon colored) decreases over time. A stacked bar plot best reflects the true ordinal nature of the data and can be used to show changes over time or an environmental gradient

1.5) and inconclusive evidence for a decrease in abundance over time (odds decreased by a factor of 0.98:95% credible interval 0.94 to 1.01). Naively using the mid-points as opposed to the ordinal classes suggests strong evidence that proportion cover increases by 0.015 units each year (95% confidence interval: 0.007, 0.02). However, the normal qq-plot highlights the severe left skewness and outliers indicating the linear model is not appropriate for these data (Appendix S2, Figure B8).

The linear regression model assumptions are violated to various degrees, depending on the underlying shape of the percent cover distribution (L-, J-, or U-shaped). Visualization tools (residual plots, qq-plots) should be used to diagnose whether there is a lack of constant variance, nonlinearity, and outliers. These assumption violations are not as much of an issue with the statistical approaches we propose because the beta distribution inherently accommodates skewness and heteroskedasticity (Table 2). Better data-model agreement and increased ecological interpretability are additional benefits to using the beta models.

2 | DISCUSSION

The development of statistical methods based on the beta distribution and associated software for implementation provides an accessible alternative for plant ecologists to consider when exploring spatial and temporal patterns in plant cover data. In fact, a number

of theoretical and empirical papers since 1954 have advocated for the beta distribution as the underlying distribution of plant cover and, in some cases, classic statistical tools that rely on the normal distribution could lead to erroneous conclusions, particularly in the case of using the mid-points for ordinal cover classes. Furthermore, in a comparison of statistical power, our proposed beta method consistently had equivalent or higher power compared to the “classic” approaches in all cases (Appendix S3). Beta regression was substantially higher powered in the case of simulating data from an L-shaped beta distribution (Appendix S3 Figure C2) and assuming that zero observations, if present, were analyzed as trace values. We suggest considering alternative statistical methods outlined in Table 2 for modelling plant cover data. To facilitate the adoption of these methods, we provide the details for our worked examples in Appendix S2.

In our own extensive empirical work with plant cover data, the beta distribution, including the above-mentioned generalizations, has typically provided an adequate fit to the studied plant cover data. However, it is possible that the beta distribution in some cases may not provide an acceptable fit to empirical cover data and then one option is to use non-parametric techniques, for example quantile regression (Schröder, Andersen, & Kiehl, 2005) for modelling such plant cover data. Thus, for some species, it may be adequate to model plant cover data without assuming an underlying beta distribution from a statistical perspective, but we strongly suggest considering the models that are linked to the beta distribution because of the ability to connect ecological processes with observed patterns more directly.

By exploiting the beta distribution, researchers can explore and test different hypotheses on the ecological role of plant spatial aggregation for single species or at the community level. For example, the measured degree of spatial aggregation of *Calluna vulgaris* on dry heathlands in Denmark has increased significantly with nitrogen deposition and decreased significantly with annual precipitation and through time since 2004 (Damgaard, 2019a). The ecological mechanisms underlying such spatial patterns need to be determined, but nitrogen deposition may possibly have positive effects on individual plant and patch sizes of *C. vulgaris*, which is expected to increase the degree of spatial aggregation. The parameterization of the beta distribution (Equation 1) and the outlined generalizations for ordinal data, zero-values, complex sampling designs and multi-species allow us to explore important ecological features of the spatial distribution of plant cover.

Spatial aggregation may play an important role in plant community dynamics. Generally, spatial aggregation is expected to reduce the importance of interspecific competitive interactions (Rees, Grubb, & Kelly, 1996). The hypothesis that increased intraspecific aggregation reduces the importance of interspecific competition has been supported by the few experiments that have been carried out (Lenssen, Herschok, Speek, During, & Kroon, 2005; Mokany, Ash, & Roxburgh, 2008; Monzeglio & Stoll, 2005; Schmidt, 1981; Stoll & Prati, 2001). Moreover, a number of theoretical studies suggest that intraspecific aggregation (or interspecific segregation) in an otherwise environmentally homogenous area more readily leads to

coexistence (e.g. Bolker & Pacala, 1999; Bolker, Pacala, & Neuhauser, 2003; Chesson, 2000; Neuhauser & Pacala, 1999). On the other hand, if interspecific competition is weak relative to intraspecific competition, then spatial aggregation may lead to demographic extinction of the slightly weaker competitor (Bolker et al., 2003; Neuhauser & Pacala, 1999). These hypotheses regarding spatial aggregation and community dynamics can be tested by exploiting our beta models for plant cover. As an example, multi-species pinpoint cover data has been fitted to the multivariate Pólya-Eggenberger distribution to measure the degree of spatial aggregation at the plant community level by a single parameter and use this parameter to test different ecological hypotheses. Using this approach, it has been tested (a) whether the observed spatial aggregation in grasslands may be explained by ecological processes that contradict the neutral theory (Damgaard & Ejrnæs, 2009), (b) whether species aggregation reduces the importance of competitive interactions (Damgaard, 2010), and (c) whether climate change or nitrogen deposition effect individual plant size, which again may affect the spatial aggregation at the plant community level (Damgaard, Ejrnæs, & Stevens, 2013). Similar hypotheses can be investigated for ordinal and line-intercept data using the beta models introduced in Table 2 and demonstrated in Appendix S2.

There are still remaining questions and possible advancements in the analysis of plant cover data that need to be looked at in the coming years. For example, more thorough guidance and examples of statistical models appropriate for multispecies ordinal cover data are needed. Dirichlet regression is a possibility for modelling percent cover for multi-species; however, it is unclear if the implied covariance structure among species is too restrictive and interpretations can be cumbersome. It is also important to further develop methods for inferring possible mechanistic ecological processes from observed plant cover changes in permanent plots. For example, whether observed plant cover changes may be explained by competitive interactions (e.g. Ransijn, Damgaard, & Schmidt, 2015). Additionally, it will be interesting to integrate the above-mentioned models of the spatial aggregation of plant abundance with geostatistical models of the local abiotic environment. This would allow a separation of the degree of spatial aggregation caused by plant ecological processes, such as vegetative growth and limited seed dispersal, from more deterministic effects caused by small-scale variation in the environment.

Previously, the state of the statistical software made it practically difficult for empirical plant ecologists to analyze their cover data correctly, but new theory and R-packages have been developed, and this difficulty no longer exists. However, software development that includes options for modelling zero-and-one augmented beta models directly and that account for hierarchical designs (i.e. random effects structures) still need to be provided. We highly recommend that empirical plant ecologists embrace the new statistical methods because of the desire to deepen ecological understanding by harnessing the beta distribution. Such an uptake will allow for exploring the exciting ecological features in spatial variation of plant cover.

ACKNOWLEDGEMENTS

Thanks to Wilson J. Wright for valuable discussions on the statistical analyses of plant cover data. We appreciate Stefanie Wacker's valuable comments on a previous version of the paper. Dr. K. Irvine's participation was supported by the US Geological Survey, National Park Service (IAA P12PG70586), and US Fish and Wildlife Service (IAA 450063656). The sago pondweed and cheatgrass data were made available from US Fish and Wildlife Service's Inventory and Monitoring Program (thanks to Jeff Warren, Jenny Barnett, and Jessica Larson) and the US National Park Service's Inventory and Monitoring Upper Columbia Basin Network (thanks to Tom Rodhouse). The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service nor National Park Service. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

AUTHORS' CONTRIBUTIONS

The writing was done in collaboration. C.F.D. made the first draft of the main text. K.M.I. made the appendices.

DATA AVAILABILITY STATEMENT

The datasets and corresponding code for reproducing the examples are available through a US Geological Survey public access catalog <https://doi.org/10.5066/P9OHMLL1> (Irvine, 2019).

ORCID

Christian F. Damgaard  <https://orcid.org/0000-0003-3932-4312>

Kathryn M. Irvine  <https://orcid.org/0000-0002-6426-940X>

REFERENCES

- Agresti, A. (2010). *Analysis of ordinal categorical data*. Hoboken, NJ: John Wiley and Sons.
- Bolker, B. M., & Pacala, S. W. (1999). Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. *The American Naturalist*, 153, 575–602. <https://doi.org/10.1086/303199>
- Bolker, B. M., Pacala, S. W., & Neuhauser, C. (2003). Spatial dynamics in model plant communities: What do we really know. *The American Naturalist*, 162, 135–148. <https://doi.org/10.1086/376575>
- Bonham, C. D. (1989). *Measurements for terrestrial vegetation*. New York, NY: John Wiley & Sons.
- Bräkenhielm, S., & Qinghong, L. (1995). Comparison of field methods in vegetation monitoring. *Water, Air, and Soil Pollution*, 79, 75–87. <https://doi.org/10.1007/BF01100431>
- Braun-Blanquet, J. (1964). *Pflansensoziologie*. Wien: Springer.
- Brown, C., Law, R., Illian, J. B., & Burslem, D. F. R. P. (2011). Linking ecological processes with spatial and non-spatial patterns in plant communities. *Journal of Ecology*, 99, 1402–1414. <https://doi.org/10.1111/j.1365-2745.2011.01877.x>

- Chen, J., Shiyomi, M., Bonham, C. D., Yasuda, T., Hori, Y., & Yamamura, Y. (2008). Plant cover estimation based on the beta distribution in grassland vegetation. *Ecological Research*, 23, 813–819. <https://doi.org/10.1007/s11284-007-0443-3>
- Chen, J., Shiyomi, M., Hori, Y., & Yamamura, Y. (2008). Frequency distribution models for spatial patterns of vegetation abundance. *Ecological Modelling*, 211, 403–410. <https://doi.org/10.1016/j.ecolmodel.2007.09.017>
- Chen, J., Shiyomi, M., Yamamura, Y., & Hori, Y. (2006). Distribution model and spatial variation of covering grassland vegetation. *Grassland Science*, 52, 167–173. <https://doi.org/10.1111/j.1744-697X.2006.00065.x>
- Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31, 343–366. <https://doi.org/10.1146/annurev.ecolsys.31.1.343>
- Clark, J. S., Gelfand, A. E., Woodall, C. W., & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, 24, 990–999. <https://doi.org/10.1890/13-1015.1>
- Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, 87, 34–56. <https://doi.org/10.1002/ecm.1241>
- Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R, 34, 24.
- Crocker, R. L., & Tiver, N. S. (1948). Survey methods in grassland ecology. *Grass and Forage Science*, 3, 1–26. <https://doi.org/10.1111/j.1365-2494.1948.tb00836.x>
- Currall, J. E. P. (1987). A transformation of the domin scale. *Vegetatio*, 72, 81–87.
- Damgaard, C. (2008). Modelling pin-point plant cover data along an environmental gradient. *Ecological Modelling*, 214, 404–410. <https://doi.org/10.1016/j.ecolmodel.2008.03.012>
- Damgaard, C. (2009). On the distribution of plant abundance data. *Ecological Informatics*, 4, 76–82. <https://doi.org/10.1016/j.ecoinf.2009.02.002>
- Damgaard, C. (2010). Intraspecific aggregation does not increase species richness in dune grasslands. *Journal of Ecology*, 98, 1141–1146. <https://doi.org/10.1111/j.1365-2745.2010.01692.x>
- Damgaard, C. (2012). Trend analyses of hierarchical pin-point cover data. *Ecology*, 93, 1269–1274. <https://doi.org/10.1890/11-1499.1>
- Damgaard, C. (2013). Hierarchical and spatially aggregated plant cover data. *Ecological Informatics*, 18, 35–39. <https://doi.org/10.1016/j.ecoinf.2013.06.001>
- Damgaard, C. (2014). Estimating mean plant cover from different types of cover data: A coherent statistical framework. *Ecosphere*, 5, art20. <https://doi.org/10.1890/ES13-00300.1>
- Damgaard, C. (2015). Modelling pin-point cover data of complementary vegetation classes. *Ecological Informatics*, 30, 179–184. <https://doi.org/10.1016/j.ecoinf.2015.10.006>
- Damgaard, C. (2019a). Is the degree of spatial aggregation of *Calluna vulgaris* regulated by nitrogen deposition or precipitation? *Acta Oecologica*, 97, 1–5. <https://doi.org/10.1016/j.actao.2019.04.002>
- Damgaard, C. (2019b). Spatio-temporal structural equation modeling in a hierarchical Bayesian framework: What controls wet heathland vegetation? *Ecosystems*, 22, 152–164. <https://doi.org/10.1007/s10021-018-0259-8>
- Damgaard, C., Ehlers, B. K., Ransijn, J., Schmidt, I. K., & Svenning, J.-C. (2018). Insights on plant interaction between dominating species from patterns of plant association: Expected covariance of pin-point cover measurements of two species. *Environmental and Ecological Statistics*, 25, 221–235. <https://doi.org/10.1007/s10651-018-0400-z>
- Damgaard, C., & Ejrnæs, R. (2009). Quantification of the intra-plot correlation in plant abundance data: A possible test of the neutral theory. *Ecological Complexity*, 6, 64–69. <https://doi.org/10.1016/j.ecocom.2008.10.008>
- Damgaard, C., Ejrnæs, R., & Stevens, C. J. (2013). Intra-specific spatial aggregation in acidic grasslands: Effects of acidification and nitrogen deposition on spatial patterns of plant communities. *Journal of Vegetation Science*, 24, 25–32. <https://doi.org/10.1111/j.1654-1103.2012.01438.x>
- Damgaard, C., Nielsen, K. E., & Strandberg, M. (2017). The effect of nitrogen deposition on the vegetation of wet heathlands. *Plant Ecology*, 218, 373–383. <https://doi.org/10.1007/s11258-016-0693-7>
- Daubenmire, R. F. (1959). Canopy coverage method of vegetation analysis. *Northwest Science*, 33, 43–64.
- Du Rietz, G. (1930). Vegetationsforschung auf soziationsanalytischer Handbuch der biologischen Arbeitsmethoden (pp. 293–480). Berlin and Wien.
- Elzinga, C. L., Salzer, D. W., & Willoughby, J. W. (1998). Measuring and monitoring plant populations. *BLM Technical Reference*, 1730–1731.
- Espinheira, P. L., Ferrari, S. L. P., & Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, 35, 407–419. <https://doi.org/10.1080/02664760701834931>
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31, 799–815. <https://doi.org/10.1080/0266476042000214501>
- Floyd, D. A., & Anderson, J. E. (1987). A comparison of 3 methods for estimating plant cover. *Journal of Ecology*, 75, 221–228.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Hovanes, K. A., Harms, K. E., Gagnon, P. R., Myers, J. A., & Elder, B. D. (2018). Overdispersed spatial patterning of dominant bunchgrasses in southeastern pine savannas. *The American Naturalist*, 191, 658–667. <https://doi.org/10.1086/696834>
- Hui, F. K. C. (2016). bral – Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution*, 7, 744–750.
- Irvine, K. M. (2019). Plant cover datasets for implementing beta distribution based models. (ed U. G. S. d. release). 5066/P9OHMLL1
- Irvine, K. M., & Rodhouse, T. J. (2010). Power analysis for trend in ordinal cover classes: Implications for long-term vegetation monitoring. *Journal of Vegetation Science*, 21, 1152–1161. <https://doi.org/10.1111/j.1654-1103.2010.01214.x>
- Irvine, K. M., Rodhouse, T. J., & Keren, I. N. (2016). Extending ordinal regression with a latent zero-augmented beta distribution. *Journal of Agricultural, Biological and Environmental Statistics*, 21, 619–640. <https://doi.org/10.1007/s13253-016-0265-2>
- Kemp, C., & Kemp, A. (1956). The analysis of point quadrat data. *Australian Journal of Botany*, 4, 167–174. <https://doi.org/10.1071/BT9560167>
- Kennedy, K. A., & Addison, P. A. (1987). Some considerations for the use of visual estimates of plant cover in biomonitoring. *Journal of Ecology*, 75, 151–157. <https://doi.org/10.2307/2260541>
- Kent, M., & Coker, P. (1992). *Vegetation description and analysis*. London, UK: Belhaven Press.
- Klimeš, L. (2003). Scale-dependent variation in visual estimates of grassland plant cover. *Journal of Vegetation Science*, 14, 815–821. <https://doi.org/10.1111/j.1654-1103.2003.tb02214.x>
- Lenzen, J. P. M., Hershock, C., Speek, T., During, H. J., & de Kroon, H. (2005). Experimental ramet aggregation in the clonal plant *Agrostis stolonifera* reduces its competitive ability. *Ecology*, 86, 1358–1365. <https://doi.org/10.1890/04-0939>
- Levy, E. B., & Madden, E. A. (1933). The point method of pasture analyses. *New Zealand Journal of Agriculture*, 46, 267–279.
- Lindquist, B. (1931). Den skandinaviska bokskogens biologi. *Svenska Skogsvårdsföreningens Tidskrift*, 3.
- MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton, NJ: Princeton University Press.
- Milberg, P., Bergstedt, J., Fridman, J., Odell, G., & Westerberg, L. (2008). Observer bias and random variation in vegetation

- monitoring data. *Journal of Vegetation Science*, 19, 633–644. <https://doi.org/10.3170/2008-8-18423>
- Mokany, K., Ash, J., & Roxburgh, S. (2008). Effects of spatial aggregation on competition, complementarity and resource use. *Austral Ecology*, 33, 261–270. <https://doi.org/10.1111/j.1442-9993.2007.01813.x>
- Monzeglio, U., & Stoll, P. (2005). Spatial patterns and species performances in experimental plant communities. *Oecologia*, 145, 619–628. <https://doi.org/10.1007/s00442-005-0168-3>
- Neuhauser, C., & Pacala, S. W. (1999). An explicitly spatial version of the Lotka-Volterra model with interspecific competition. *The Annals of Applied Probability*, 9, 1226–1259.
- Nielsen, K. E., Bak, J. L., Bruus, M., Damgaard, C., Ejrnæs, R., Fredshavn, J. R., ... Strandberg, M. (2012). NATURDATA.DK - Danish monitoring program of vegetation and chemical plant and soil data from non-forested terrestrial habitat types. *Biodiversity & Ecology*, 4, 375. <https://doi.org/10.7809/b-e.00165>
- Niku, J., Warton, D. I., Hui, F. K. C., & Taskinen, S. (2017). Generalized linear latent variable models for multivariate count and biomass data in ecology. *Journal of Agricultural, Biological and Environmental Statistics*, 22, 498–522. <https://doi.org/10.1007/s13253-017-0304-7>
- Ospina, R., & Ferrari, S. (2010). Inflated beta distributions. *Statistical Papers*, 51, 111–126. <https://doi.org/10.1007/s00362-008-0125-4>
- Pacala, S., & Levin, S. A. (1997). Biological generated spatial pattern and the coexistence of competing species. Spatial ecology. In D. Tilman, & P. Kareiva (Eds.), *The role of space in population dynamics and interspecific interactions* (pp. 204–232). Princeton, NJ: Princeton University Press.
- Qu, Y., Greene, T., & Piedmonte, M. R. (1993). Symmetric Bernoulli distributions and generalized binomial distributions. *Biometrical Journal*, 35, 523–533. <https://doi.org/10.1002/bimj.4710350503>
- Ramsey, F., & Schafer, D. (2002). *The statistical sleuth. A course in methods of data analysis*. Boston, MA: Duxbury Press.
- Ransijn, J., Damgaard, C., & Schmidt, I. (2015). Do competitive interactions in dry heathlands explain plant abundance patterns and species coexistence? *Plant Ecology*, 216, 199–211. <https://doi.org/10.1007/s11258-014-0428-6>
- Rees, M., Grubb, P. J., & Kelly, D. (1996). Quantifying the impact of competition and spatial heterogeneity on the structure and dynamics of a four-species guild of winter annuals. *The American Naturalist*, 147, 1–32. <https://doi.org/10.1086/285837>
- Robinson, P. (1954). The distribution of plant populations. *Annals of Botany*, 18, 35–45. <https://doi.org/10.1093/oxfordjournals.aob.a083380>
- Rodhouse, T. J., Irvine, K. M., Sheley, R. L., Smith, B. S., Hoh, S., Esposito, D. M., & Mata-Gonzalez, R. (2014). Predicting foundation bunchgrass species abundances: Model-assisted decision-making in protected-area sagebrush steppe. *Ecosphere*, 5, 1–16. <https://doi.org/10.1890/ES14-00169.1>
- Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. San Diego, CA: Elsevier Ltd.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71, 319–392.
- Schmidt, W. (1981). Über das Konkurrenzverhalten von *Solidago canadensis* und *Urtica dioica*. *Verhandlungen der Gesellschaft für Ökologie*, 9, 173–188.
- Schröder, H. K., Andersen, H. E., & Kiehl, K. (2005). Rejecting the mean: Estimating the response of fen plant species to environmental factors by non-linear quantile regression. *Journal of Vegetation Science*, 16, 373–382. <https://doi.org/10.1111/j.1654-1103.2005.tb02376.x>
- Stoll, P., & Prati, D. (2001). Intraspecific aggregation alters competitive interactions in experimental plant populations. *Ecology*, 82, 319–327.
- Vittoz, P., Bayfield, N., Brooker, R., Elston, D. A., Duff, E. I., Theurillat, J.-P., & Guisan, A. (2010). Reproducibility of species lists, visual cover estimates and frequency methods for recording high-mountain vegetation. *Journal of Vegetation Science*, 21, 1035–1047. <https://doi.org/10.1111/j.1654-1103.2010.01216.x>
- Wang, Y., Naumann, U., Wright, S. T., & Warton, D. I. (2012). mvabund—An R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*, 3, 471–474.
- Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution*, 30, 766–779. <https://doi.org/10.1016/j.tree.2015.09.007>
- Watt, A. S. (1947). Pattern and process in the plant community. *Journal of Ecology*, 35, 1–22.
- Wilson, J. B. (2011). Cover plus: Ways of measuring plant canopies and the terms used for them. *Journal of Vegetation Science*, 22, 197–206. <https://doi.org/10.1111/j.1654-1103.2010.01238.x>
- Wright, W. J., Irvine, K. M., Warren, J. M., & Barnett, J. K. (2017). Statistical design and analysis for plant cover studies with multiple sources of observation errors. *Methods in Ecology and Evolution*, 8, 1932–1941. <https://doi.org/10.1111/2041-210X.12825>
- Yee, T. W. (2015). *Vector generalized linear and additive models: With an implementation in R*. New York, NY: Springer.
- Yeo, J. J., Rodhouse, T. J., Dicus, G. H., Irvine, K. M., & Garrett, L. K. (2009). Upper Columbia Basin Network Sagebrush Steppe Vegetation Monitoring Protocol: Narrative and Standard Operating Procedures (Appendix to Narrative). *Natural Resource Report*. NPS/UCBN/NRR—2009/142, National Park Service, Fort Collins, Colorado.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Damgaard CF, Irvine KM. Using the beta distribution to analyse plant cover data. *J Ecol*. 2019;107:2747–2759. <https://doi.org/10.1111/1365-2745.13200>