1 Model-based ordination of pin-point cover data: effect of management on

2 dry heathland

- 3 Christian Damgaard¹, Rikke Reisner Hansen¹, and Francis K. C. Hui²
- 4 1 Department of Bioscience, Aarhus University, Vejlsøvej 25, 8600 Silkeborg, Denmark
- 5 2 Research School of Finance, Actuarial Studies and Statistics, Australian National University,
- 6 Acton ACT 2601, Australia
- 7 Corresponding author: Christian Damgaard, cfd@bios.au.dk, +45-30183153

9 Abstract

10 Recently, there has been an increasing interest in model-based approaches for the statistical 11 modelling of the joint distribution of multi-species abundances. The Dirichlet-multinomial 12 distribution has been proposed as a suitable candidate distribution for the joint species distribution of pin-point plant cover data and is here applied in a model-based ordination framework. Unlike 13 14 most model-based ordination methods, both fixed and random effects are in our proposed model 15 structured as *p*-dimensional vectors and added to the latent variables before the inner product with 16 the species-specific coefficients. This changes the interpretation of the parameters, so that the fixed 17 and random effects now measure the relative displacement of the vegetation by the fixed and 18 random factors in the *p*-dimensional latent variable space. This parameterization allows statistical 19 inference of the effect of fixed and random factors in vector space, and makes it easier for 20 practitioners to perform inferences on species composition in a multivariate setting. The method 21 was applied on plant pin-point cover data from dry heathlands that had received different 22 management treatments (burned, grazed, harvested, unmanaged), and it was found that treatment 23 have a significant effect on heathland vegetation both when considering plant functional groups or 24 when the taxonomic resolution was at the species level.

Keywords: Model-based ordination; plant pin-point cover data; statistical inferences on fixed and
 random effects; management of dry heathlands.

28 Introduction

The statistical treatment of whole communities with multiple species has traditionally relied on 29 30 ordination methods that use distance measures to reduce the plot by species abundance matrix into a low dimensional vector of distances among plots. Traditional methods for ordination, such as Non 31 32 Metric Multidimensional scaling (NMDS) or Principal Component Analysis (PCoA), rely on 33 strictly algorithmic approaches with no underlying statistical model of species abundance (McCune 34 and Grace 2002). To accommodate potential properties inherent in the data (e.g. a strong mean-35 variance relationship with count data), these methods may apply a variety of distance measures or 36 data transformations. However, commonly applied distance metrics, such as Euclidean, Manhattan, 37 or Bray–Curtis distance, often make implicit assumptions on the mean-variance relationship that 38 may not be entirely compatible with the data at hand. This can result in potentially incorrect 39 conclusions, e.g. location effects may be confounded within the dispersion effects (Warton et al. 40 2012, Warton and Hui 2017).

Recently, a number of ordination techniques, where the distribution of species abundance is 41 42 explicitly taken into account, have been developed. Collectively, these new techniques are referred 43 to as model-based ordination. In univariate statistics, for instance, these issues of mean-variance 44 relations in the species response have long been addressed with generalized linear models (GLM's) 45 and their mixed model counterparts, where the mean-variance relationship is modelled for each 46 response variable (Bolker et al. 2009). Model-based ordination can thus be regarded as multivariate 47 extensions of GLM's. This method offers the possibility of adjusting the distribution family to e.g. 48 negative binomial distribution for overdispersed count data and the Bernoulli distribution for 49 presence-absence data to better account for the inherent mean-variance relationship in the species 50 observations (Hui et al. 2015, Warton et al. 2015). Like GLM's, model- based ordination methods 51 further allow for a check of the validity of such model assumptions. Through row/column 52 standardizations, as incorporated in model-based approaches, each taxon or species contribution to 53 the sample location in ordination space is accounted for (Hawinkel et al. 2019) as well as potential 54 clustering of data across different levels of study structure. For instance, hierarchical structures 55 caused by host-parasite dynamics and reflecting nested design experiments (Björk et al. 2018). 56 Hence, additions to the model-based ordination literature are continually emerging (e.g. Sohn and 57 Li 2018, Hawinkel et al. 2019, Niku et al. 2019).

58 Arguably, the most popular method of model-based ordination that has emerged is latent variable 59 modelling, which involves modelling community composition through a small set of underlying latent variables that reduce the dimension of the species abundance data from the number of species 60 61 to the dimension of the latent variables (Hui et al. 2015, Warton et al. 2015, Warton et al. 2016, 62 Niku et al. 2017). In doing so, the latent variables also express the residual variation owing to 63 factors beyond those of the measured predictors included in the model, e.g. possible biotic 64 interactions or phylogenetic relatedness between species. More specifically, in latent variable modelling the species composition of each plot is described by a vector with dimension p that is 65 much less than the number of species. These latent variables are then fitted to the species abundance 66 67 data using the relevant distribution of species abundance. Since the latent variables are not 68 observed, they are treated as random effects, meaning that they need to be predicted at the same 69 time that coefficients associated with them are estimated (often called loadings) along with other 70 potential site, treatment, and environmental effects.

71 Ecological multivariate data analyzed using these methods often presents as count data or presence-72 absence data. Because plants are sessile, they facilitate registration of multiple characteristics 73 related to vegetation structure and microclimate. The most common way to quantify plant species 74 abundance in light-open plant communities is to measure the cover, which is the relative area of the 75 species when projected onto the surface. Unlike plant counts or density, plant cover takes the size of 76 individuals into account. A relatively objective method for measuring plant cover is the pin-point 77 (or point-intercept) method in a frame or along a line, which has been widely employed in the field. 78 This involves vertically inserting a thin pin into the vegetation a number of times in a fixed design, 79 and the cover of a species is measured by the proportion of the inserted pins that touches the species 80 (Lindquist 1931, Levy and Madden 1933, Damgaard and Irvine 2019). The pin-point method is not 81 relevant for measuring the abundance of rare species and has been shown to underestimate species 82 richness (Bråkenhielm and Qinghong 1995). Importantly, the pin-point method allows unbiased aggregation of single species cover measures at the pin level into cover data for higher plant taxa, 83 84 e.g. the cover of grasses or herbal plants.

Many plant species are spatially aggregated, and it is important to take this spatial aggregation into
account when modelling multi-species pin-point plant cover data. Recently, a reparametrized
Dirichlet-multinomial distribution has been proposed for modelling such cover data (Damgaard
2015, 2018). The Dirichlet distribution is the conjugate prior of the multinomial distribution and a

89 multivariate generalization of the beta-binomial distribution, which has previously been used with

- 90 success to model pin-point cover data of spatially aggregated plant species (Damgaard 2009, 2013,
- 91 Damgaard and Irvine 2019). Moreover, the Dirichlet-multinomial distributed model has been
- 92 applied for modelling multi-species vegetation dynamics in e.g. heathland ecosystems (Damgaard
- 93 2015, Damgaard et al. 2017, Damgaard 2019).

94 The aim of this study is two-fold; i) to modify the model-based ordination method previously

- 95 suggested by among others Hui et al. (2015) so that it is applicable to multi-species pin-point cover
- 96 data, and ii) to reparametrize the underlying latent variable model so that the effect of fixed and
- 97 random factors on the species composition may be investigated and tested in the same underlying
- 98 multidimensional space as the latent variables characterizing the ordination.

99 The proposed method will be illustrated on multi-species pin-point cover data from six Danish dry

- 100 heathlands, where different nature management practices have been applied and, more specifically,
- 101 it will be used to assess whether the form of management has had an effect on heathland vegetation.

102 **Model**

103 The objective of the proposed model is to reduce a plant species abundance matrix with k plots and 104 *n* species to a k * p matrix of latent variables, where p < n is the number of latent variables, and the 105 plant abundance of each species is measured by its cover using the pin-point method. The effects of 106 observed covariates, including e.g. the experimental unit and design effects on species composition, 107 may be entered as fixed and/or random effects into the mean structure as well. The inclusion of such 108 effects means that the latent variables then model the so-called residual correlation between species, 109 i.e. any covariation that cannot be explained by the observed predictors and experimental design 110 effects (Warton et al. 2015, Niku et al. 2017).

111 Partly following Hui et al. (2015) and adopting a Bayesian framework for estimation and inference,

112 we propose that the mean cover of species j in plot i, denoted here as q_{ij} , be modelled as follows

113
$$logit(q_{ij}) = (\boldsymbol{\alpha}_{fixed[i]} + \boldsymbol{\gamma}_{random[i]} + \boldsymbol{z}_i)' \cdot \boldsymbol{\theta}_j + \beta_j$$
(1),

114 where all parameters in bold denote vectors of dimension *p*.

115 In this model, the vector $\boldsymbol{\alpha}_{fixed[i]}$ denotes a fixed effect applied to plot *i*, for instance a treatment

- 116 effect, and are assigned weak prior distributions, e.g. $N_p(\mathbf{0}, 100 \text{ I})$, which is a p-dimensional
- 117 multivariate normal distribution with zero mean vector and a covariance matrix with a diagonal
- 118 matrix with all diagonal elements set to 100. Note that for reasons of parameter identifiability, it is
- assumed that these fixed effects are not unique to each plot (which is almost always the case).
- 120 Next, the vector $\boldsymbol{\gamma}_{random[i]}$ model denotes a random effect applied to plot *i*, e.g. the location in a 121 hierarchical experimental design, and are assumed to be drawn from a common multivariate normal 122 distribution with a zero mean vector and an unstructured p^*p covariance matrix, $N_n(\mathbf{0}, \boldsymbol{\Sigma})$. For
- 123 example, when p = 2, as is commonly considered for the purposes of ordination, $\Sigma =$
- $\begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$. In this case, the standard deviations σ_i are assigned a uniform positive prior, 124 125 while the correlation coefficient ρ is assigned a uniform prior between -1 and 1. As with the fixed 126 effects, the random effects are assumed not to be at plot level for reasons of parameter identifiability, e.g. in our case study the random effect is at the level of the site, in which multiple 127 plots are nested. In the above notation (eq. 1), the model is fitted with only one fixed factor and one 128 129 random factor, however, the model may be extended to include either more fixed or random effects, 130 as well as including a spatial correlated random component. Conversely, the model may be 131 simplified by either omitting the fixed and random effect.
- Finally, the vector \mathbf{z}_i denotes a vector of p latent variables for plot i and, as is standard, is assumed to come from a standard normal distribution $N_p(\mathbf{0}, \mathbf{I})$, where the zero mean vector and identity covariance matrix are used to avoid location and scale invariance and ensure that the parameters in the model are identifiable (Hui et al. 2015, Niku et al. 2017).
- 136 The vector $\boldsymbol{\theta}_j$ denotes a *p*-dimensional vector of coefficients or loading for species *j* (e.g. Hui et al.
- 137 2015) and is, similarly to the fixed effects above, assigned a weak prior $N_p(0, 100 \text{ I})$. Note that in
- 138 order to ensure that the parameters are identifiable, we further apply a standard constraint of
- 139 assuming that the upper triangular portion of the n * p matrix of species-specific coefficients is
- 140 constrained to be zero, while the diagonal elements are constrained to be positive (see Hui et al.
- 141 2015, Niku et al. 2017). The loadings can be interpreted as quantifying the species responses to the
- 142 unmeasured latent variables and can also be plotted in conjunction with the ordinations as a model-

based biplot (e.g. Warton et al. 2015). Finally, the quantities β_j denote a vector of constants chosen *a priori*, which are required to ensure that $logit(q_{ij})$ is a real number.

145 The main difference between the present method and the original latent variable ordination 146 approach proposed by, among others, Hui et al. (2015) and Warton et al. (2015) is that both the fixed and random effects, $\boldsymbol{\alpha}_{fixed[i]}$ and $\boldsymbol{\gamma}_{random[i]}$ are vectors of dimension p and are added to the 147 latent variables prior to their inner product with the species-specific coefficients. This changes the 148 149 interpretation of the parameters so that the fixed and random effects now measure the relative 150 displacement of the vegetation by the fixed and random factors in the *p*-dimensional latent space. 151 Consequently, the latent variable z_i may now more properly be referred to as modelling the residual 152 correlation between species, i.e. the covariation after the fixed and random effects are taken into 153 account. The scale of the displacement by the fixed and random effects is measured relatively to the 154 residual latent variables, which are fixed to a unit standard variation for ease of interpretation (and parameter identifiability). This parameterization of the underlying latent variable model uses the 155 156 concepts of fixed and random effects in an analogous way as to how the terms are used 157 in generalized linear mixed effects models (Bolker et al. 2009). It is anticipated that this analogy 158 will make it easier for practitioners to test e.g. treatment effects on species composition in a 159 multivariate setting.

160 The underlying motivation for using model-based ordination instead of traditional ordination is to take the distributional properties of the species abundance sampling into account such that e.g. the 161 162 correct mean-variance relationship is used in the modelling (e.g. Warton et al. 2012, Warton and 163 Hui 2017). In this study, we consider the multi-species pin-point cover data, and it has previously 164 been demonstrated that a reparametrized Dirichlet-multinomial distribution is a suitable candidate 165 distribution to model multi-species pin-point cover data (Damgaard 2015, Damgaard et al. 2017, 166 Damgaard 2019). The advantage of using the reparametrized Dirichlet-multinomial distribution is 167 that the degree of spatial aggregation in plant communities is taken into account and explicitly 168 modelled by a parameter δ (Damgaard 2018). More specifically, the observed hierarchical multispecies pin-point cover data, Y, is modelled by a mean cover vector of the n-1 first species, q_i , and a 169 parameter δ , which measures the degree of intra-specific spatial aggregation, by, 170

171
$$Y \sim Mn(\sum_{n} y_i, (p_1, \dots, p_{n-1}, 1 - p_1 - \dots - p_{n-1}))$$
 (2),

172
$$\Lambda(p_1, \dots, p_{n-1}) \sim Dir\left(\frac{q_1 - q_1\delta}{\delta}, \dots, \frac{q_{n-1} - q_{n-1}\delta}{\delta}, \frac{1 - \delta}{\delta} - \frac{q_1 - q_1\delta}{\delta} - \dots - \frac{q_{n-1} - q_{n-1}\delta}{\delta}\right)$$

At the limit when $\delta \rightarrow 0$, the reparametrized Dirichlet-multinomial distribution degenerates into the multinomial distribution. Here, the prior distributions of q_j and δ are assumed to be uniformly distributed between (0, 1) and (0.01, 0.95), respectively. For more details on the properties of the reparametrized Dirichlet-multinomial distribution, see Damgaard (2018).

177 Estimation

178 The proposed latent variable model was estimated using Bayesian Markov Chain Monte Carlo

179 (MCMC) methods using the Metropolis-Hastings algorithm with normally distributed candidate

180 distributions. Specifically, we considered a MCMC chain with a burn-in period of 70,000 iterations

181 followed by 30,000 additional iterations.

182 Trace plots of the sampling chains of all parameters and latent variables were examined to assess

183 their mixing properties and convergence of the MCMC chain. Additionally, the overall fitting

184 properties of the model were checked by examining the regularity and shape of the marginal

185 distribution of parameters as well as the distribution of the deviance $(= -2 \log L(Y|\theta))$. The

186 efficiency of the MCMC procedure was assessed by inspecting the evolution in the deviance.

187 For ordination and when p = 2, we constructed plots of the posterior mean values of $(\alpha_{fixed[i]} + \gamma_{random[i]} + z_i)$ for each plot along with 95% credibility regions of $\alpha_{fixed[i]}$, noting 188 189 that these are p-dimensional regions rather than the unusual one-dimensional credibility intervals 190 seen in mixed models and the standard latent variable ordination approach. These credibility regions 191 enabled us to make statistical inferences on the effect of the fixed factors on species composition, e.g. if the 95% credibility regions of two fixed factors did not overlap. In such cases, we concluded 192 193 that there is clear statistical evidence that the effects of the two factors on the species composition 194 differ substantially. Consequently, the suggested parameterization in equation (1) will allow us to 195 utilize the model-based ordination framework for testing purposes on the latent space. Furthermore, 196 posterior means of the species coefficients, θ_i for each cover class were also extracted and 197 superimposed on top of the ordination diagram to visualize the contribution of each cover class 198 (species) to the location of plots in the diagram, although note that the mean of the last species (q_n) 199 is not estimated in eqn. 2.

200 All calculations were done using *Mathematica* (Wolfram 2016).

201 Case study

202 To assess the effect of heathland management on plant communities, plant species were registered 203 using the pin-point method. During the fall of 2018, the cover of heathland plant species was 204 registered at 6 heathland sites encompassing four different heathland management regimes: 205 harvesting, grazing, burning and unmanaged, i.e. abandonment of management. At each site several 206 of the four management regimes were applied. Within each site and management combination, the 207 pin-point measurements were repeated in each of ten randomly positioned plots (Electronic 208 Supplement, Appendix A). The pin-point frames had 25 pins within an area of 0.5m x 0.5m. 209 Pinpoint data were then aggregated at the pin-level into five cover classes: Dwarf shrubs, 210 graminoids, forbs, mosses and lichens. That is, if e.g. two graminoid species where hit by the same 211 pin, they were aggregated to a single hit (Fig. 1).

212 We modelled both the aggregated pin-point data and the underlying single species pin-point data

213 (number of species: 34) using the proposed latent variable ordination approach formulated above,

214 where the four different heathland management regimes were assumed to be *p*-dimensional fixed

effects, i.e. $\alpha_{treatment[i]}$ in equation (1), and the site was assumed to be a *p*-dimensional random

effect, i.e. $\gamma_{site[i]}$ in equation (1). For the purpose of visualization, the dimension of the latent

217 variable model (*p*) was set to 2.

218 Results

219 The MCMC iterations demonstrated fair mixing properties (Electronic supplement, Appendix B),

and the marginal posterior distribution of selected parameters and compound parameters are

summarized in Table 1 and Fig. 2.

222 The main results are shown in Fig. 2, where the mean of the posterior distribution of $(\alpha_i + \gamma_i + z_i)$

vectors are displayed for each plot for the aggregated species classes and all species, respectively.

The distance between plots indicates the difference in the cover of the species, where the scale is

relative to the unit standard deviation of the residual latent variables, i.e. the variation in species

cover among plots that could not be explained by the fixed or random factors.

227 The 95% credibility areas of the effect of fixed factors, i.e. the different management regimes, are 228 shown by colored ovals in Fig. 2. Since no ovals overlap, we concluded that the four different 229 management regimes lead to characteristic heathland plant communities, which on average were 230 distinct from each other. In particular, grazing has a large effect on the composition of the heathland 231 vegetation, whereas the effects of the other three management regimes were comparably more 232 similar (Fig. 2). This is further demonstrated in Fig. 1, where grazing leads to a shift in vegetation 233 cover from a high dwarf-shrub dominance towards forb and graminoid covered heathland. Inclusion of all species in the model still yielded a distinct plant species composition with non-overlapping 234 235 credibility intervals. However, the relative positions of burned and harvested plots were shifted, 236 indicating that inference of management affects species composition depending on the taxonomic 237 resolution (functional groups vs. all species). Furthermore, increasing the taxonomic resolution 238 leads to higher within treatment variation in the unmanaged plots compared to the aggregated 239 dataset (Fig 2b, larger credibility area).

The scale of the random effects, i.e. the effect of site, is shown in Table 1. Given that the scale of the random effect is relative to the unit standard deviation of the residual latent variables, then we conclude that the displacement due to sites tended to be significantly larger than the variation among the residual latent variables (Table 1, σ_2 is significantly larger than one for both cases). This indicated a relatively large influence of site-specific species pools. There was also a significant positive correlation between the two dimensions of the random effects when all species were analyzed (Table 1).

In both cases, the estimated posterior distribution of δ was dominated by the chosen lower limit of the prior distribution of the parameter (0.01), and the degree of spatial aggregation could not be distinguished from zero in a statistical sense, i.e. random expectations.

250 **Discussion**

Over the past five years, there has been an increasing interest in model-based approaches for the statistical modelling of the joint distribution of multi-species abundances (e.g. Clark et al. 2014, Warton et al. 2015, Warton et al. 2016, Ovaskainen et al. 2017). The current study is an example reflecting this important trend in community ecology, as the importance of how species abundance data are distributed, sampled, and driven by a complex interplay of different ecological processes is 256 given increasing recognition. Plant species are sedentary, and in many terrestrial ecosystems plants 257 dominate the ground cover, leaving few bare patches. These characteristic features of plant growth 258 have important consequences for the joint distribution of plant species cover, such as large spatial 259 aggregation among single species and relatively strong negative correlation among them. To model 260 these features, the Dirichlet-multinomial distribution has previously been suggested as a suitable 261 candidate distribution for the joint species distribution of pin-point plant cover data (Damgaard 262 2015, Damgaard et al. 2017, Damgaard 2019). The Dirichlet-multinomial distribution has markedly 263 different distributional properties than both the Bernoulli distribution and the negative binomial 264 distribution, which till now have typically dominated the development of model-based ordination 265 methods and software (e.g. Niku et al. 2019).

In the suggested reparametrized Dirichlet-multinomial distribution, the spatial aggregation among plant species is modelled by the parameter δ , which increases with the within-site spatial variation in cover, i.e. the degree of spatial aggregation. In the present case, the degree of spatial aggregation could not be distinguished from random expectations, and it was concluded that the observed covariation pattern among the species was adequately modelled by the residual latent variables. Consequently, we suggest to test whether the more simple multinomial distribution is suitable in future model-based ordination of pin-point cover data.

273 Unlike in Hui et al. (2015) and Warton et al. (2015), among others, both the fixed and random 274 factors are here modelled as p-dimensional vectors and added to the latent variables before the inner 275 product with the species-specific coefficients. This means that the estimated displacement effects of 276 management and sites are estimated and statistical inference performed in the same *p*-dimensional 277 latent ordination space. When p = 2 in particular, the estimated average displacements of e.g. the different management regimes may be shown in the plane as ovals of 95% credibility areas. Such 278 279 plots, we find, are an illustrative way to show the effects of different treatments on the species 280 composition. This is in contrast to most latent variable models, which consider the sample scores as 281 random effects and make prior distributional assumptions on them without incorporating both 282 random and fixed effects in the same model framework. Naturally, the suggested parameterization 283 with fixed and random effects may also be applicable for count data or absence-presence data.

Not surprisingly, in our case study the interpretation of the fixed effects is sensitive to taxonomic
resolution of the dataset. This was evident when comparing Fig 2a and b. If data was aggregated

286 into functional groups (Fig. 2a), it could be concluded that management abandonment leads to 287 relatively homogenous plant communities. However, the model based on species level resolution 288 (Fig. 2b) showed that the variation within the unmanaged plots is high. Regardless of taxonomic 289 resolution, both models exhibited clear evidence that a varied management regime with different 290 management methods produces an overall heterogeneous plant species composition. Hence, 291 interpretation of fixed effects in this model-based framework may provide important information 292 for heathland conservation at both the level of plant functional groups and at the species level 293 resolution.

If p > 2, then the 95% credibility regions, which illustrate the mean effect of the fixed and random factors, must be imagined as a *p*-dimensional ellipsoidal shape. We expect that the statistical power to separate the effects of the different fixed factors on species composition will increase with the dimension of the latent variable used in the ordination. However, this preliminary notion needs to be explored in a more systematic way. Finally, the chosen estimation method is relatively slow, and in cases with many plots and many species faster estimation procedures may be needed and should be developed in the future.

301 Acknowledgements

We thank K. E. Nielsen for field help providing pin-point data. CFD and RRH both acknowledge
the funding provided by Aage V. Jensen foundation. FKCH acknowledges the funding provided by
two grants from the Australian Research Council

306 Tables

307 Table 1. Marginal credible intervals for the posterior distribution of the parameters in $\Sigma =$

$$308 \quad \begin{pmatrix} \sigma_1^2 & \rho \, \sigma_1 \, \sigma_2 \\ \rho \, \sigma_1 \, \sigma_2 & \sigma_2^2 \end{pmatrix}$$

Parameter	2.5%	50%	97.5%					
Aggregated cover classes:								
σ_1	0.828	1.355	1.820					
σ_2	1.044	1.811	2.651					
ρ	-0.710	-0.046	0.563					
All species:								
σ_1	0.356	0.753	1.346					
σ_2	1.270	1.965	3.005					
ρ	0.377	0.870	0.993					

309

311 Figures

- Fig. 1. The cover of the aggregated species at the different management treatments B: burned, G:
- 313 grazed, H: harvested, U: unmanaged.



Fig. 2. Map of latent variables, posterior means of $(\alpha_i + \gamma_i + z_i)$ and 95% credibility areas of the effect of fixed factors (management), brown: burned, light green: harvested, dark brown:

unmanaged, dark green: grazed. Mean species coefficients, θ_i for each cover class (species) are

319 shown. A: four aggregated species classes. B: All species, 1: *Eriophorum angustifolium*, 2:

320 Calamagrostis epigejos, 3: Trichophorum cespitosum, 4: Danthonia decumbens, 5: Agrostis

321 capillaris, 6: Nardus stricta, 7: Festuca ovina, 8: Deschampsia flexuosa, 9: Festuca rubra, 10:

322 Molinium caerulea, 11: Poa compressa, 12: Carex pilulifera, 13: Carex arenaria, 14: Carex

323 panacea, 15: Carex nigra, 16: Luzula campestris, 17: Galium saxatile, 18: Hieracium pilosella, 19:

324 Potentilla erecta, 20: Rumex acetosella, 21: Hypochoeris radicata, 22: Vaccinium vitis-idaea, 23:

325 Calluna vulgaris, 24: Empetrum nigrum, 25: Genista anglica, 26: Erica tetralix, 27: Dicranum

326 scoparium, 28: Pseudoscleropodium purum, 29: Pleurozium schreberi, 30: Bryum subsp., 31:

327 Sphagnum compactum, 32: Campylopus introflexus, 33: Hypnum cupressiforme, 34: Cladonia

328 rangiferina.



330

331 References

- Björk, J. R., F. K. C. Hui, R. B. O'Hara, and J. M. Montoya. 2018. Uncovering the drivers of host-associated
 microbiota with joint species distribution modelling. Molecular Ecology 27:2714-2724.
- Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White.
 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in
 Ecology & amp; Evolution 24:127-135.
- Bråkenhielm, S. and L. Qinghong. 1995. Comparison of field methods in vegetation monitoring. Water, Air,
 and Soil Pollution **79**:75-87.
- Clark, J. S., A. E. Gelfand, C. W. Woodall, and K. Zhu. 2014. More than the sum of the parts: forest climate
 response from joint species distribution models. Ecological Applications 24:990-999.
- 342 Damgaard, C. 2009. On the distribution of plant abundance data. Ecological Informatics **4**:76-82.
- 343 Damgaard, C. 2013. Hierarchical and spatially aggregated plant cover data. Ecological Informatics **18**:35-39.
- Damgaard, C. 2015. Modelling pin-point cover data of complementary vegetation classes. Ecological
 Informatics **30**:179-184.
- 346 Damgaard, C. 2018. The joint distribution of pin-point plant cover data: a reparametrized Dirichlet 347 multinomial distribution. arXiv e-prints.
- Damgaard, C. 2019. Spatio-Temporal Structural Equation Modeling in a Hierarchical Bayesian Framework:
 What Controls Wet Heathland Vegetation? Ecosystems 22:152-164.
- Damgaard, C. and K. M. Irvine. 2019. Using the beta distribution to analyze plant cover data. Journal of
 Ecology.
- Damgaard, C., K. E. Nielsen, and M. Strandberg. 2017. The effect of nitrogen deposition on the vegetation
 of wet heathlands. Plant Ecology 218:373-383.
- Hawinkel, S., F.-M. Kerckhof, L. Bijnens, and O. Thas. 2019. A unified framework for unconstrained and
 constrained ordination of microbiome read count data. PLOS ONE 14:e0205474.
- Hui, F. K. C., S. Taskinen, S. Pledger, S. D. Foster, and D. I. Warton. 2015. Model-based approaches to
 unconstrained ordination. Methods in Ecology and Evolution 6:399-411.
- Levy, E. B. and E. A. Madden. 1933. The point method of pasture analyses. New Zealand Journal of
 Agriculture 46:267-279.
- 360 Lindquist, B. 1931. Den skandinaviska bokskogens biologi. Svenska Skogsvårdsföeningens Tidskrift 3.
- McCune, B. and J. Grace. 2002. Analysis of ecological communities. MJM Software Design, Gleneden Beach,
 OR.
- Niku, J., F. K. C. Hui, S. Taskinen, and D. I. Warton. 2019. gllvm: Fast analysis of multivariate abundance data
 with generalized linear latent variable models in r. Methods in Ecology and Evolution 10:2173-2182.
- Niku, J., D. I. Warton, F. K. C. Hui, and S. Taskinen. 2017. Generalized Linear Latent Variable Models for
 Multivariate Count and Biomass Data in Ecology. Journal of Agricultural, Biological and
 Environmental Statistics 22:498-522.
- Ovaskainen, O., G. Tikhonov, A. Norberg, B. F. Guillaume, L. Duan, D. Dunson, T. Roslin, and N. Abrego.
 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters 20:561-576.
- Sohn, B. M. and H. Li. 2018. A GLM-based latent variable ordination method for microbiome samples.
 Biometrics **74**:448-457.
- Warton, D. I., F. G. Blanchet, R. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui. 2016.
 Extending Joint Models in Community Ecology: A Response to Beissinger et al. Trends in Ecology &
 Evolution **31**:737-738.

- Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui. 2015. So
 Many Variables: Joint Modeling in Community Ecology. Trends in Ecology & Evolution **30**:766-779.
- Warton, D. I. and F. K. C. Hui. 2017. The central role of mean-variance relationships in the analysis of
 multivariate abundance data: a response to Roberts (2017). Methods in Ecology and Evolution
 8:1408-1414.
- Warton, D. I., S. T. Wright, and Y. Wang. 2012. Distance-based multivariate analyses confound location and
 dispersion effects. Methods in Ecology and Evolution 3:89-101.
- 383 Wolfram, S. 2016. Mathematica. Wolfram Research, Inc., Champaign, USA.

385 Electronic supplements

386 Appendix A: Details of the case study

- 387 A map of Jutland showing the location and size of the heathlands investigated in the study. Inset
- 388 figure shows the location of the selected plot within the management parcel. Inset figure shows the
- 389 experimental design within each plot. The squares show an example of the randomised locations for
- 390 pin-point analyses.

391



392

An overview of the selected sites and the management carried out in the selected plots. The number
under burn, harvest, and unmanaged is the number of years since the management was last applied.
For grazing, the number is the number of years it has been grazed. Frequency denotes how often the
management is performed.

Site	Harvest	Grazing	Unmanaged	Frequency

	Burn	(Years)	(Years)	(Years)	
	(Years)				
Harrild	3	2		> 50	10 years
Kongenshus	1	1		> 50	5 years
Borris	1	1		> 30	1 year
Ovstrup		2	> 50	> 50	10 years
Randboel		2	> 30	> 100	Once
Noerholm			> 50	> 100	-

