Methods

Grassland vegetation sampling — a practical guide for sampling and data analysis

Bianca Ott Andrade[1, 5*] (iD), Ilsi Iob Boldrini[1, 2] (iD), Monica Cadenazzi[3] (iD), Valério D. Pillar[4] (iD) and Gerhard Ernst Overbeck[1, 2] (iD)

ABSTRACT

Grassland and savanna ecosystems are the original vegetation types of more than 30% of the Brazilian territory, but conservation and training of future professionals has largely focused on forests. In fact, no standard protocols exist for sampling grassland vegetation for environmental planning or licensing. Neglecting non-forest ecosystems may have deleterious consequences for the maintenance of biodiversity and the provisioning of ecosystem services. Herein, we provide practical guidelines on how to conceive and develop ecological studies of grassland vegetation for scientific, monitoring or technical purposes. Using examples mostly from southern Brazil, we explain and discuss the various components of the research process, from the question to be investigated to defining and implementing the sampling design, performing data analyses and presenting the results. Our guidelines should prove useful for training technicians and researchers working on grassland ecosystems and other non-forest ecosystems in Brazil and surrounding regions.

Keywords: non-forest ecosystems, plant community, phytosociological studies, quantitative analysis, relevé, vegetation survey

Introduction

Conservation of biodiversity and ecosystem services depends on knowledge of the ecosystems to be preserved and of the species within them. The Brazilian law for the protection of native vegetation (Brasil 2012) specifies that any suppression of natural vegetation needs to be authorized by the responsible authorities. However, the criteria based on which authorization can be granted, or not, still need to be established for grasslands in Brazil. The characterization of the vegetation at the site under question provides crucial information on the conservation state of the ecosystems; especially quantitative data allows for an in-depth analysis of plant community structure. In addition, information regarding the occurrence of rare, threatened or exotic plant species most likely will be important for any evaluation of conservation value and thus for decisions on management

1 Programa de Pós-Graduação em Botânica, Universidade Federal do Rio Grande do Sul, 91501-970, Porto Alegre, RS, Brazil
2 Departamento de Botânica, Universidade Federal do Rio Grande do Sul, 91501-970, Porto Alegre, RS, Brazil
3 Departamento de Biometria, Estadística y Computacion, Universidad de la República, 60000, Paysandú, Uruguay
4 Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, 91501-970, Porto Alegre, RS, Brazil
5 Department of Agronomy and Horticulture, University of Nebraska-Lincoln, 68583-0915, Lincoln, NE, USA

* Corresponding author: andradebo@gmail.com

and land use change. If put into a broader context, this data also allows to infer about uniqueness of the vegetation at a given site, which is important for licensing processes, as in many regions large parts of natural vegetation have already been lost.

The state of knowledge of vegetation and vegetation sampling in different Brazilian biomes and vegetation types has been presented to a wider academic public (Felfili *et al.* 2011; Eisenlohr *et al.* 2015). This has included a general outline of methods for characterizing grassland vegetation (Munhoz & Araújo 2011). However, even in regions where grasslands form the characteristic vegetation, focus both in vegetation sampling for licensing purposes and in training of future professionals in the past has largely been on forest, or on the tree component, as non-forest system have been less valued for conservation, despite high biodiversity (Overbeck *et al.* 2007; 2015). This contrasts with high levels of degradation by land use change (Andrade *et al.* 2015; Oliveira *et al.* 2017). Increased efforts to assess grassland vegetation in south Brazil for management and planning of land use change, and especially stronger efforts for conservation, are necessary and recently under way (Andrade *et al.* 2019), but need to be spread further, including into procedures of environmental licensing.

Several publications and guidelines on methods of vegetation sampling exist, starting with those in classical books of vegetation ecology (*e.g.* Mueller-Dombois & Ellenberg 1974), and some are also available in Portuguese or Spanish (*e.g.* Braun-Blanquet 1979; Munhoz & Araújo 2011). However, the objective of these works usually is not to give guidance for a specific kind of study, but to present general information on sampling procedures and calculation of **phytosociological parameters**, usually in a theoretic approach. More accessible guidance of what kind of information should be collected, and how, in specific regions is mostly missing. Such a practical guidance, however, is essential to implement legal requirements and to comply with environmental legislation. While sampling of vegetation for purposes foreseen in legislation does not need to become a scientific study, nonetheless the quality of both field work and analyses/presentation of the data must be guaranteed.

Here, we wish to fill this gap and provide a practical guide for studies on grassland vegetation in Brazil. Our paper synthesizes key issues regarding the **sampling design**, choice of field measurement methods and data analysis for grassland vegetation studies. We emphasize the necessity to link sampling design and the planned analyses of the collected data, *i.e.* the necessity to define the study design based on the objective of the study. Our aim explicitly is not to give a full review of methods (this has been done elsewhere, *e.g.* Mueller-Dombois & Ellenberg 1974; Kent 2012) but to be selective based on what we – from our working experience – consider to be the most important aspects. We focus on the Campos Sulinos region where we

work and where grasslands have for long been neglected in conservation; however, the guidelines we present should be, likely with adaptations, transferable to other regions with grassland or non-forest vegetation, *e.g.* in the grasslands adjacent to southern Brazil, the Río de la Plata grasslands (Andrade *et al.* 2018) or to the Cerrado.

A practical guide for grassland vegetation sampling

Objectives of sampling

The objectives of the study always have implications for the sampling design. Broadly, we can distinguish studies aiming to test causal hypotheses or to search for patterns in space and time and their relations to environmental factors or ecosystem effects and studies aiming to estimate vegetation composition for practical purposes such as in environmental impact assessments or licensing vegetation suppression (Tab. 1). By 'grassland vegetation sampling' we imply that the grassland plant community as a whole is the subject of interest. Studies may of course also focus on certain species populations of special interest (*e.g.* monitoring of endangered species or of invasive species, for example as part of a management plan) or other issues, but this kind of study is out of the scope of our guidelines.

Towards an adequate sampling design for grassland vegetation studies

"Sampling" implies that we evaluate vegetation in a limited number of sampling units, the elements of the study, that build our sample – and this sample should be defined in a way ("sampling design") that it adequately represents the sampling universe (see Tab. 2 for terms and concepts).

Vegetation varies in time: when to sample?

Subtropical grasslands, such as those found in south Brazil, are strongly influenced by seasonality of climatic conditions, with implications for plant phenology (Freitas & Sazima 2006). Considering south Brazil grasslands, at the end of the winter, *i.e.* from August on, grasses with C3 photosynthetic pathway will start flowering, together with several herbs with similar phenological patterns. From late spring on, *i.e.* late November, summer species, *e.g.* grasses with a C4 metabolism, will start their development, coming to a flowering peak in summer. Species from many other families also show clear seasonal patterns during the year. Flowering of summer species will continue until March and April, with *e.g.* many species from the Asteraceae flowering late in the season (*e.g.* Oleques *et al.* 2017). The implications for vegetation sampling in the region are clear: depending on the timing of sampling, some species are not easily

detectable, which may produce differences in the description of vegetation composition. An ideal period for sampling of grasslands in south Brazil are October and November in the Pampa grasslands, and between the end of November and end of December in the highland grasslands, when both summer and winter species are present. In January, dry spells may occur that complicate sampling, especially in the southern part of the region, and winter species will usually not be detectable. In the highland grasslands in Rio Grande do Sul, Santa Catarina and Paraná States, as well as in the coastal region, we usually have sufficient humidity for sampling later in the year, until March, but then early season species may not be detectable. Sampling should be avoided from April to August, as the number of species that will be missed is high, and the lack of flowers will make species identification more difficult. Differences in species composition due to the sampling period is especially important in comparative studies based on published data that may stem from surveys conducted at different time

periods. Studies in other regions will likely require the definition of other periods, usually avoiding sampling in the dry season.

In savanna ecosystems, the flowering and fruiting peak of herbaceous vegetation are also associated to climate seasonality, and mostly occur in the middle and at end of rainy season, in contrast to the woody component that generally flowers during the late dry and early wet season (*e.g.* Batalha & Martins 2004; Munhoz & Felfili 2005). Thus, sampling of the herbaceous stratum in the Cerrado should be avoided from May to September due to the lack of flowers or fruits which make it difficult to identify the plant species and may increase the risk of misidentification. However, besides precipitation seasonality, fire has also been reported as an important factor influencing flowering and fruiting, both in Cerrado (*e.g.* Munhoz & Felfili 2005; Massi *et al.* 2017) and in the *Campos Sulinos* (Fidelis & Blanco 2014), and sampling shortly after fire may thus be interesting to answer specific questions related to responses of the plant community to fire.

Table 1. Main types of plant community studies, their objectives and resulting implications for study design.

Objective	Study approach	Implications	Examples
Testing causal hypotheses.	Randomized experiments with manipulated factors.	Replication and adequate controls for treatments are essential. May involve monitoring in time.	Responses of plant communities to grassland management (Baggio *et al.* 2018), restoration procedures (Thomas *et al.* 2019), or other types of treatments, *e.g.* species removal (Guido *et al.* 2017).
Searching for patterns in space and for relations to environmental factors or ecosystem effects.	Surveys (observational studies).	Controlled, randomized experimental manipulation is not involved. The sample should adequately represent the sampling universe, thus proper stratification may be needed. Sampling sufficiency guided by stability of detected patterns. Causal inference may be possible with path analysis and structural equation modelling.	Effects of land use change or management history on plant communities (Staude *et al.* 2018; Koch *et al.* 2016); classification of grasslands over a larger region (Andrade *et al.* 2019).
Searching for patterns in time.	Surveys (observational, monitoring studies).	Controlled, randomized experimental manipulation may not be involved. Sampling based on replicated permanent plots representative of the monitored area. Monitoring of reference systems is desirable for restoration studies.	Long-term ecological research (Andrade *et al.* 2016).
Estimation of vegetation composition for practical purposes such as environment impact assessment (EIA/RIMA), authorization of suppression of native vegetation.	Surveys (observational study).	Sampling sufficiency guided by precision requirements for species composition and diversity estimation.	Description of plant communities at a given site (Setubal & Boldrini 2010).

Table 2. Terms used in sampling design.

Term	Meaning	Example
Sampling universe	Target of investigation: All measurable sampling units.	All possible vegetation plots in the delineated study area.
Stratum	A subset of the sampling universe.	All possible plots between 600m and 800 m a.s.l.
Sampling unit	One element of the sampling universe.	One plot.
Sample	Set of selected sampling units taken from the sampling universe.	A vegetation table based on a certain number of plots.
Attribute	Descriptors of the sampling units.	Plant species presence or abundance, environmental factors.
Sampling design	Location of sampling units, their size, shape, and number.	A sampling grid systematically covering the sampling universe.

Vegetation varies in space: where to sample?

Vegetation studies should take place in clearly defined areas and thus the selection of sampling sites is the first step of any study. In classical **phytosociological studies**, plots are preferentially placed in vegetation stands previously defined (Chytrý 2001); this kind of data is often used to build vegetation classification systems over larger areas (*e.g.* Willner *et al.* 2017; for discussion see *e.g.* Ewald 2003) but is not usually the objective of single studies that aim at evaluating a specific area or situation and where the question of representability of study points is important. As vegetation patterns in space are influenced by bedrock, soil conditions, topography, management type and frequency, among others, it is important to consider if it is important to take this variation into account or not – depending on the objectives of the study.

Points or plots: which sampling method is the most adequate?

Two principal types of sampling methods exist, the plot method, and the point (or point intercept) method.

The point method (Levy & Madden 1933) uses a wooden frame with pins fixed every ten centimeters. The pins are lowered into the ground and all species touched by the pins are recorded. In case a pin touches more than one species, they all are recorded. This technique is commonly used in short-term experiments, and requires very good expertise in plant species identification. A more practical option concerning mobility is the use of a string with knots every ten centimeters, or a measuring tape. The point method may detect a smaller number of species (Vittoz & Guisan 2007) and generate higher cover estimates (Kercher *et al.* 2003) than the plot method. For illustrations and more detailed explanation of the method see Herrick *et al.* (2005).

The plot method consists in the description of standardized sampling areas (plots). All species found in a plot are registered, and usually their abundance is recorded (for example, percent cover which is more applicable to herbaceous vegetation than number of individuals, the parameter used in studies on woody species). Additionally, general features of the vegetation may be recorded, *e.g.* vegetation height, percentage of bare soil, etc. As plot-based sampling today is much more common and as it offers more possibilities in data analysis, we will focus on this approach in the following (see also Boldrini & Overbeck 2015).

Practical issues of plot sampling: size, distribution and number

Plot size: The smaller the plot, the higher the accuracy of its description, as it is easier to observe smaller areas in detail. Also, in general, vegetation within smaller plots is more homogeneous than within larger plots, which may

allow more robust conclusions about relations between vegetation and micro-environmental conditions. In any case, since plot size influences the number of recorded species and thus estimates of species richness (Chytrý 2001; Dengler *et al.* 2008), the analysis of spatial patterns of plant populations (Smith 1964), correlations between species occurrences or performances (Smith 1964) and vegetation-environment correlations (Davies *et al.* 2005), it is useful to use a standard plot size when working in a given type of vegetation; different plots sizes will always complicate comparison between studies. For adequate sampling, a plot must be large enough to represent the pattern or community to be investigated in a meaningful way. Thus, the plot size depends on the physical **vegetation structure**, which is influenced by the size of plants. In the Campos Sulinos region, most studies use plots of 0,25 m^2 or of 1 m^2. Pragmatically, we recommend 1 m^2, as this allows for sampling in different grassland types, *e.g.* those dominated by small species, but also those dominated by tall-growing tussock grasses; thus, this plot size should also be useful *e.g.* in Cerrado vegetation. If the objective is to additionally sample upper level vegetation, *e.g.* the shrub or tree component, when present, larger plots should be applied for these species. In this case, a nested design can be used, with 1 m^2 plots for the grassland layer inserted into 25-1000 m^2 plots for the shrub or tree component (*e.g.* Felfili *et al.* 2005; Koch *et al.* 2016), depending on the grassland ecosystem under consideration. Plots of 1 m^2 are also of a suitable size to be sampled by two observers, which may increase efficiency and accuracy, as two observers together may detect a higher number of species (Vittoz & Guisan 2007). For purposes of vegetation classification over larger areas, sometimes larger plots, denominated stands, are used (see subsection *The vegetation stand*).

Location of plots in space: Usually, plots should be located in a way that their totality is representative of the entire plant communities. A single plot, *e.g.* in the center of a given grassland area, will not be sufficient for vegetation characterization (unless it is of a very large, impractical size), as it represents only one possible condition of the vegetation. Plots should be placed randomly or systematically in the area under investigation. Further, previous stratification can be applied when the objective is to evaluate the effects of gradients or contrasts (*e.g.* differences on topography, soil water content, etc.) and the different strata are to be represented equally; this design will then allow for formal testing of hypothesis involving factors that might influence vegetation composition (see section on *Data analysis* below). The strata need not necessarily be of equal area, if the number of sampling units taken from each stratum is proportional to its area (Smith 1964). Within each stratum, plots can be placed randomly or systematically. This type of sampling is called stratified random sampling and it ensures a representative sample of the different types of environment in the sampling universe. In systematic

sampling, the plots are located at fixed distances along transects or at the nodes of an imaginary grid over the area to be sampled (*e.g.* Andrade *et al.* 2019).

Plot number (sample size): The criterion to be considered to assess **sampling sufficiency** in the number of plots depends on the objectives of the study (Tab. 1, see also Kenkel *et al.* 1989). If the aim is the estimation of vegetation composition, such as for environment impact assessment (EIA/RIMA), or to obtain an authorization of suppression of native vegetation, sampling sufficiency should be guided by precision requirements for species composition and diversity estimation. For mean plot richness, diversity, species composition or any other variable evaluated in the plots, bootstrap resampling can be used for the estimation of precision based on confidence limits (Pillar 1998). For estimating richness in the study area, species-area relationships are perceived as saturation curves, indicating a 'minimum area' for sampling, which can also be achieved by other indicators such as Jackknife or Chao indices (Magurran 2004). On the other hand, when the aim is to search for patterns of vegetation in space or time and its relations to environmental factors or ecosystem effects, sampling sufficiency should be guided by the stability of conclusions regarding these patterns (Pillar 1998; 1999a; b). That is, a sample maybe too small, *e.g.* for estimating species richness, but sufficient for detecting a stable pattern, *e.g.* the classification of plant communities.

Importantly, if the aim is to evaluate conservation value of an area or to get complete species lists (or at least as complete as possible), plot-based sampling may never be able to sample all species. Rare species – that may be especially interesting from a conservation perspective – will often be missed. Thus, additional **floristic sampling,** where species are simply listed when present, may be necessary. This kind of sampling should also be repeated at different times of the year to get a full list, due to seasonal differences (see above) that may also affect rare or endangered species.

We must be aware of the trade-offs between plot size and plot number and decide – based on the purpose of the study – what is the best design (Kenkel *et al.* 1989; Pillar 1998). This is also relevant as plot size and number will influence the variance of the measured variables. In general terms, for estimation purposes, fewer plots that are large and internally heterogeneous may provide estimates with better precision than many smaller, more homogeneous plots. However, the opposite will be needed for detecting patterns, since otherwise the patterns will be blurred by sampling plots that are too heterogenous internally.

The vegetation stand

The stand is often considered the basic unit for the study of plant community. In phytosociological studies, stands are to be representative of a given vegetation type and are used for descriptive purposes, providing the basic data to build up named units for classification of the vegetation into distinct and stable communities, and to form primary hypotheses of causal factors differentiating them (Poore 1955). The unit stand (also known as "relevé") should be (i) large enough to contain all species belonging to the plant community, (ii) uniform within the stand area, and (iii) its plant coverage should be as homogeneous as possible (*i.e.* should not contain large open areas, or be dominated by different species at distinct patches within the stand) (Mueller-Dombois & Ellenberg 1974). Size of a stand may vary from 0.25 to 1 ha, depending on the specific plant community, and needs to be evaluate a priori, *e.g.* by species-area curves. For grassland studies, 25 m² (5 x 5 m) is a common plot size (*e.g.* Lezama *et al.* 2019). While phytosociological studies have a long tradition in Europe (European phytosociology school) they are actually not much used in South America, even though phytosociological parameters are being calculated. See Ewald (2003) for a critical discussion concerning European relevé analysis and American continuum analysis

The sampling process: estimating species abundances

When studying a tree community, it is rather easy to evaluate species **abundances**: we can simply count the number of individuals present in a plot. In grasslands, it is difficult to identify individuals, as many plants present clonal growth by rhizomes or stolons or may sprout from underground organs. Apparently separated shoots can belong to the same individual. Thus, we usually visually estimate plant species abundance based on percent cover in the plot. Visual estimates of plant cover are widely used both in descriptive and experimental studies. The Braun-Blanquet (1979) scale probably is the most well-known cover scale. Alternatively, the Londo (1976) 13-classes scale may be adopted (see Tab. 3). However, depending on the aims of the study and the range of environmental conditions in the sampled area, it may be adequate to record only species presence-absence; this will be faster and may allow conclusions that are not different than using species abundance or cover (Wilson 2012).

The sampling process: dealing with unidentified species

More often than not, it is impossible to identify all plant species in the field, especially when they do not present flowers or fruits. Unidentified plant material needs to be collected, dried and organized for future identification in the laboratory using taxonomic literature and help of specialists. Note that some journals require a list of voucher numbers for the species; in this case, researchers should collect well developed individuals that will then be deposited in a herbarium.

Table 3. Cover classes of the Londo (1976) and Braun-Blanquet (1979) scales and mean values of each cover class used for analytical purpose.

Londo scale			Braun-Blanquet scale		
Cover (%)	Class	Value for analysis*	Cover (%)	Class	Value for analysis*§
<1	0.1	0.5	<5 (solitary to few individuals)	r	0.02
1-<3	0.2	2		+	0.1
3-<5	0.4	4	<5 (numerous individuals)	1	2.5
5-<15	1	10	5-<25	2	15
15-<25	2	20			
25-<35	3	30	25-<50	3	37.5
35-<45	4	40			
45-<55	5	50			
55-<65	6	60	50-<75	4	62.5
65-<75	7	70			
75-<85	8	80	75-<100	5	87.5
85-<95	9	90			
95-100%	10	100			

*corresponds to the mean cover value; § see Maarel (2007) for discussion.

The sampling process: recording additional information

When conducting vegetation sampling, it is advisable to collect additional information that may be important in interpretation of the vegetation data. This refers, for example, to surface inclination and aspect of the plot, both variables that may directly influence plant composition. This also refers to structural characteristics of the vegetation: vegetation height may be a consequence of disturbance regimes or resource availability, and often there is a relationship between plant species richness and vegetation height, as higher (and larger) plants tend to outcompete smaller species. Species richness as well as plant species composition often will also be influenced by substrate availability and conditions. Recording percent cover of open soil, litter and rocks is neither complicated nor time consuming, and can present very meaningful information. Depending on the objectives of sampling, it may be interesting to record additional parameters. For example, number or cover of cow dung patches may be a proxy for grazing pressure, an important predictor variable in grazed grassland.

Of course, environmental variation can be recorded by many other ways. For instance, soil sampling and subsequent analysis can give us detailed information on chemical and physical features of the habitat the plant species live in. However, this information is usually not sampled at the plot scale, due to costs, but for larger areas. For more details concerning soil sampling methods for ecological studies see Perkins *et al.* (2013).

Permanent plots

Permanent plots allow for the repeated recording of vegetation descriptors such as cover of plant species in a fixed area. The main objective of permanent plots is to detect vegetation change over time (for example, succession or long-term effects of management), which may be achieved in two ways: (i) fine-scale monitoring of permanent plots, or (ii) re-sampling of old vegetation plots without permanent marking. The former will allow precise information on trends in species composition, including increases and decreases of the cover of each species, as well as changes in the structural attributes of the community, among other descriptors. The second approach means that the plot location cannot be established exactly, and thus analyses necessarily will be a bit coarser, usually conducted not at the individual plot level but considering a set of plots in the same stratum. In both cases standardized sampling is important (see subsection *Long-term biodiversity monitoring: the need of standardized sampling design*).

For permanent plot installation in grassland, we recommend use of metal specs at the ground level. If not found by visual inspection, a metal detector can be used. We do not advise on using sticks aboveground in grasslands with grazing animals, as they attract domestic grazers which may lead to destruction or to heavy trampling right at the plot. Also, sticks may serve as perches for birds and thus may lead to woody species recruitment in the plots or close by. For easier location of plots in the field, sticks – possibly painted – may be placed at fixed distances from the actual plots.

Long-term biodiversity monitoring: the need of standardized sampling design

Long-term vegetation monitoring is important to detect biodiversity changes over time, *e.g.* due to land-use change. One example for South Brazil is the LTER/PELD Campos Sulinos project (Fig. 1), today with seven sites distributed in different grassland regions in Rio Grande do Sul State. Aim of the project is to collect long-term data on grassland plant and arthropod communities under different types of grassland management. In contrast to many other LTER projects, it thus contains an experimental approach testing for the effects of different types of grassland management (see *e.g.* Andrade *et al.* 2016), and thus is designed to deliver management-relevant data on a regional scale.

While the LTER/PELD aims to respond specific questions on grassland management, the need for broader biodiversity monitoring has led to the development of standardized sampling designs such as the RAPELD protocol, where subsampling points are situated systematically on transects located along isoclines. Plots on an isocline tend to be more similar each other, which allows enlarging the subsampled area within each sampling unit (transect) without increasing its heterogeneity. As the aim is monitoring of different groups of organisms, plot size and design are defined specifically for each group, but this protocol allows the integration of data (*e.g.* different trophic levels) collected on the same transects for which also environmental data

(*e.g.* soil features, land use data) are available. For details on the RAPELD method, see Magnusson *et al.* 2005.

Figure 1. Contrasting vegetation structures in response to the application of different management treatments in the context of a long-term ecological research located in Lavras do Sul, Rio Grande do Sul, Brazil. The experimental plot in the left side of the fence was submitted to grazing management while at the right side it was excluded from management.

Data is collected – now what?

Description of vegetation patterns

Usually, vegetation data will be organized in a contingency table where plots (sampling units) are described by species (variables) or by other variables such as total biomass or vegetation height. First, basic information to be compiled from the plant community table – and already providing a first vegetation description – include (i) the number of species per plot, (ii) the frequency of each species, (iii) the mean abundance of each species over the total number of plots, (iv) information on total vegetation cover and cover of bare soil/rock or rocks. Thus, the calculation of the classical phytosociological parameters (*sensu* Müller-Dombois & Ellenberg 1974) usually is the first step of data analysis. Second, the species in the table can be ordered according to mean abundance values, thus providing a good characterization of principal patterns of the community. If the study followed a stratified sampling approach (see above), it is useful to present data for the different strata. In most types of study, it will be important to give the full species list and respective abundance values, but sometimes it may be sufficient to indicate only the most abundant species in the main document. However, species abundances should always be informed for all species in an appendix or in supplementary material. For any studies that are important in a context of environmental licensing, information on presence of threatened species (on state or national lists) is important here; it needs to be critically evaluated if sampling was appropriate (*e.g.* in terms of season) to include these species in sampling.

Depending on the aim of the study, simple, exploratory analyses, may be useful, *e.g.* by correlating vegetation height and species richness per plot to reveal possible processes shaping the plant community. Similarly, it may be interesting to correlate environmental data (*e.g.* soil features) with features of the plant community, *e.g.* by correlation or linear regression; environmental data should also be prepared in a table. Note that in correlation and regression analyses, no statistical inference on causal relations can be made, and in the latter values of one parameter are used to predict another.

Basic issues for statistical analysis

Alongside with the pure description of a plant community and its features, the aim of the study may be testing causal hypotheses, or searching for spatial or temporal patterns and its relations to environmental factors or ecosystem effects (Tab. 1), in which case the analysis may require multivariate methods to compare sampling units regarding vegetation composition and/or environmental variables. Testing for multivariate differences between groups (or experimental treatments) can be achieved by means of Multivariate Analysis of Variance (MANOVA), for which permutation (or randomization) methods are available (Pillar & Orlóci 1996; Anderson 2001; Pillar 2013). The same methods (Manly 2007), as well as bootstrap resampling (Kirby & Gerlanc 2013), can be used whenever statistical testing for the effects of individual variables are needed. Using classical Analysis of Variance (ANOVA) or comparison of means (*e.g.* Tukey test) requires that assumptions of normal distribution and homogeneity of variances of data are met. Normality can be assessed by the visual interpretation of a Quantile-Quantile plot (QQ plot) or by an appropriate test, such as Shapiro-Wilks and homogeneity of variances by help of the Levene-test (*e.g.* Montgomery 2017; Das & Imon 2016). However, when working with field data, it is common that the data do not possess a normal distribution (especially cover data measured in percent, where distributions usually are skewed). If we cannot accept the premises mentioned above, data transformations, such as logarithmic, square root and angular (or arc sine), are often recommended. One problem can be that ecological data often contains many zeros; in this case, it is difficult to achieve normality by transformations. If it is not possible to reach normality and homogeneity of variance by transformation, parametric tests like ANOVA or t-Test are not possible, as they will require our data to follow the normal distribution and/or large sample size. In this case, we can still apply non-parametric tests that are based on the rank order of species, such as Kruskal-Wallis (for three or more groups of sampling units) or the Friedman test (for two groups of sampling units), but they may have lower power. This is why more and more commonly randomization and bootstrap methods are used, as these do not imply in assumption about data distribution, though the tests may be affected by unequal group variances (Pillar 2013). Of course, adequate replication is necessary for all

analysis of differences between groups of sampling units (see subsection *Beware of pseudo-replication!*).

Today, researchers often wish to compare groups of plots regarding a larger number of variables taken individually, *e.g.* different environmental factors or cover of different species groups. However, the more tests are done based on the same data set, the higher also the chance that one of them will reveal a statistically significant result by pure chance (Pike 2010). Often, this problem of multiple comparisons in neglected, but adjusting for multiple tests is required in confirmatory studies whenever results from multiple tests have to be combined in one final conclusion and decision (Bland & Altman 1995). The correction prosed by Bonferroni is one of the most widely used for this problem, but leads to loss in test power, so that other methods may be preferred (see discussion in Pike 2010).

Diversity indices

Species richness is one way to quantify biodiversity but ignores completely that different species may have different abundances. Indices of diversity, in contrast, rely on two dimensions of the plant community: species richness and the evenness of the distribution of individuals within species. Well-known indices are, for example, the Shannon index and the Simpson index. Possibly, the Shannon index is the most widely used, however, it is not easy to interpret and, Shannon values of different studies cannot be compared, as they depend on species richness which is influenced by the sampling design. The Simpson index, scaled between 0 and 1, is more useful in this regard. At any rate, based on a single index, it still is impossible to say which community (or plot) is the most diverse, as different indices vary in the weight they give to species richness and the evenness. Melo (2008) discusses the issue and suggests the use of diversity profiles that present a more detailed picture of diversity patterns in the plant community.

Beware of pseudo-replication!

Despite a wealth of highly cited literature (*e.g.* Hurlbert 1984; 2004; Oksanen 2001; 2004), pseudoreplication continues to be a problematic issue in vegetation studies. It arises when we do not have true replicates of the experimental situation. For instance, if we sample vegetation in 20 plots of one burned grassland site and in 20 plots of another unburned grassland site, we cannot make interference on the differences between burned and unburnt grasslands, because we only studied one site for each, and thus our sampling units cannot be considered independent replicates for each class – a pre-requisite for allowing conclusions about statistical effects. We would need replicates of the different types of site to be able to do so, at minimum three or (better) more. Unfortunately, many studies without true replicates (pseudoreplicates

instead) still are being conducted – and published. Often low number of sites is a consequence of the desire to reach sampling sufficiency in terms of species composition at each site – which, however, is not usually needed if the aim is to compare general ecological patterns among groups of sampling units.

Ordination and classification/cluster analysis

Ordination (Goodall 1954) consists of the arrangement of vegetation units in relation to each other in terms of their similarity of species composition and/or their environmental drivers, it thus is a very useful way to synthetize – and display graphically – similarities among sampling units. The main goals of an ordination are primarily the reduction of dimensionality and the description of patterns in species composition. Depending on how successful we are at this process, the ordination space allow us to: (i) summarize the plant community data and identify the nature of vegetation variation in the area under study, (ii) examine and compare the distribution of a given species within communities, (iii) summarize the variation within sets of vegetation units which can be correlated to biotic and abiotic controls defining a gradient (Kent 2012).

There is a large variety of available methods of ordination analysis, such as CCA (Canonical Correspondence Analysis), PCA (Principal Component Analysis), PCoA (Principal Coordinate Analysis), RDA (Redundancy Analysis), NMDS (Non-Metric Multidimensional Scaling), among others (Podani 2005; Legendre & Legendre 2012; Wildi 2017). The choice of ordination method will depend on the dataset (*e.g.* high number of zeroes; the category of data [*e.g.* quantitative, qualitative, mixed]) and the goal of the study. While PCoA, NMDS, PCA consist of indirect gradient analysis (*i.e.* have no external data), CCA and RDA assess plant-environment relationships in a direct gradient analysis (*i.e.* utilizes external environmental data in addition to the species data). The PCoA and NMDS are distance-based approaches

Ordination and clustering (or classification) are complementary approaches. The clustering is the placement of vegetation sampling units into groups. It is often useful when one wishes to differentiate and map ecological communities (*e.g.* Andrade *et al*. 2019).

Conclusions

For all kinds of vegetation studies, sampling needs to be planned beforehand based on clearly defined objectives and clear ways to reach them. It does not matter if the main goal of a vegetation survey is to develop basic ecological research or if the work is to be an applied study in conservation and environmental monitoring: the methods adopted in both cases must be defined carefully to give rise to reproducible and robust data. If the resulting study is of

Bianca Ott Andrade, Ilsi Iob Boldrini, Monica Cadenazzi, Valério D. Pillar and Gerhard Ernst Overbeck

quality depends on three major points: First, and discussed in this contribution, a sampling design that is adequate for the posed situation/question. Second, and also discussed, adequate and correct analysis of data and interpretation of results. Additionally, there is one third important issue that we did not treat here: knowledge of plant species and experience in field sampling is a basic requirement for all studies (see Menezes *et al.* 2018 for discussion), and training of professionals capable to do and to interpret vegetation studies of grassland plant communities an important issue.

At present efficient strategies and mechanisms for conservation and restoration of the Campos Sulinos are lacking, and the same goes for other regions in Brazil, including those with non-forest vegetation (Overbeck *et al.* 2015). Without doubt, a better knowledge of grassland vegetation is an important step to effectively include grasslands in conservation measures, even though this alone is not important: the understanding that ultimately human society depends on biodiversity and the ecosystem services it provides (IPBES 2018). The political will to manage our natural resources accordingly is, obviously, the basis for any conservation action. At any rate, the more well-conducted studies are available, land-management and conservation decisions will have a more solid grounding.

Glossary

Abundance: estimation of local density of a given plant species; in grasslands, usually percent cover over a given area.

Analysis of sampling sufficiency: evaluation if the sampling effort in an ecological study is sufficient for its objective. In general terms, sampling sufficiency indicates stability in the conclusions of the study.

Floristic sampling: sampling that consists in listing all plant species found at a given locality (as opposed to plot-based sampling).

Sampling design: selection and arrangement of the sampling units in the field.

Phytosociological parameters: quantitative attributes which describe the vegetation patterns in a given study site, commonly used in descriptive vegetation studies (*e.g.* absolute and relative abundance, absolute and relative frequency, importance value index).

Phytosociological studies: any study that describes the co-occurrence of plant species. In the Central European school of Phytosociology, higher level classifications of plant communities are developed.

Vegetation structure: usually used for the description of the plant community by help of phytosociological parameters, but sometimes also used to describe the physical vegetation structure (*e.g.* height of different strata, total cover).

References

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26: 32-46.

Andrade BO, Bonilha CL, Ferreira PMA, Boldrini II, Overbeck GE. 2016. Highland Grasslands at the Southern tip of the Atlantic Forest Biome: Management options and conservation challenges. Oecologia Australis 20: 37-61.

Andrade BO, Bonilha CL, Overbeck GE, *et al.* 2019. Classification of South Brazilian grasslands: Implications for conservation. Applied Vegetation Science 22: 168-184.

Andrade BO, Koch C, Boldrini II, *et al.* 2015. Grassland degradation and restoration: A conceptual framework of stages and thresholds illustrated by southern Brazilian grasslands. Natureza & Conservação 13: 95-104.

Andrade BO, Marchesi E, Burkart S, *et al.* 2018. Vascular plant species richness and distribution in the Río de la Plata grasslands. Botanical Journal of the Linnean Society 188: 250-256.

Baggio R, Medeiros RB, Focht T, Boavista LR, Pillar VD, Müller SC. 2018. Effects of initial disturbances and grazing regime on native grassland invasion by *Eragrostis plana* in southern Brazil. Perspectives in Ecology and Conservation 16: 158-165.

Batalha MA, Martins FR. 2004. Reproductive phenology of the Cerrado plant community in Emas National Park (central Brazil). Australian Journal of Botany. 52: 149-161.

Bland JM, Altman DG. 1995. Multiple significance tests: the Bonferroni method. BMJ 310: 170.

Boldrini II, Overbeck GE. 2015. Estudos fitossociológicos em vegetação campestre. In: Eisenlohr PV, Felfili JM, Melo MMRF, Andrade LA, Meira-Neto JAA. (eds.) Fitossociologia no Brasil: métodos e estudos de casos. Vol. II. Viçosa, Editora UFV. p. 228-249.

Brasil. 2012. Lei Federal nº 12.651, 25 de maio de 2012. Brasília, DOU.

Braun-Blanquet J. 1979. Fitosociología. Bases para el estudio de las comunidades vegetales. Madrid, Blume Ediciones.

Chytrý M. 2001. Phytosociological data give biased estimates of species richness. Journal of Vegetation Science 12: 439-444.

Das KR, Imon AHMR. 2016. A brief review of tests for normality. American Journal of Theoretical and Applied Statistics 5: 5-12.

Davies KF, Chesson P, Harrison S, Inouye BD, Melbourne BA, Rice KJ. 2005. Spatial heterogeneity explains the scale dependence of the native-exotic diversity relationship. Ecology 86: 1602-1610.

Dengler J, Chytrý M, Ewald J. 2008. Phytosociology. In: Jörgensen SE, Fath BD. (eds.) Encyclopedia of ecology. Amsterdam, Elsevier. p. 2767-2779.

Eisenlohr PV, Felfili JM, Melo MMRF, Andrade LA, Meira-Neto JAA. 2015. Fitossociologia no Brasil: métodos e estudos de casos. Vol. II. Viçosa, Editora UFV.

Ewald J. 2003. A critique for phytosociology. Journal of Vegetation Science 14: 291-296.

Felfili JM, Carvalho FA, Haidar RF. 2005. Manual para monitoramento de parcelas permanentes nos biomas Cerrado e Pantanal. Brasília, Universidade de Brasília.

Felfili JM, Eisenlohr PV, Melo MMRF, Andrade, LA, Meira-Neto JAA. 2011. Fitossociologia no Brasil: métodos e estudos de casos. Vol.1. Viçosa, Editora UFV.

Fidelis A, Blanco C. 2014. Does fire induce flowering in Brazilian subtropical grasslands?. Applied Vegetation Science 17: 690-699.

Freitas L, Sazima M. 2006. Pollination biology in a tropical high-altitude grassland in Brazil: Interactions at the community level. Annals of the Missouri Botanical Garden 93: 465-516.

Goodall DW. 1954. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. Australian Journal of Botany 2: 304-324.

Guido A, Pillar VD. 2017. Invasive plant removal: assessing community impact and recovery from invasion. Journal of Applied Ecology 54: 1230-1237.

Herrick JE, Zee JW, Havstad KM, Burkett LM, Whitford WG. 2005. Monitoring manual for grassland, shrubland and savanna ecosystems. Vol. I: Quick Start. Vol. II: Design, supplementary methods and interpretation. Las Cruces, USDA-ARS Jornada Experimental Range.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54: 187-211.

Hurlbert SH. 2004. On misinterpretations of pseudoreplication and related matters: a reply to Oksanen. Oikos 104: 591-597.

IPBES – Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. 2018. The IPBES regional assessment report on biodiversity and ecosystem services for the Americas. Bonn, Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

Kenkel NC, Juhász-Nagy P, Podani J. 1989. On sampling procedures in population and community ecology. Vegetatio 83: 195-207.

Kent M. 2012. Vegetation description and data analysis: a practical approach. Chichester, John Wiley & Sons.

Kercher SM, Frieswyk CB, Zedler JB. 2003. Effects of sampling teams and estimation methods on the assessment of plant cover. Journal of Vegetation Science 14: 899-906.

Kirby KN, Gerlanc D. 2013. BootES: An R package for bootstrap confidence intervals on effect sizes. Behavior Research Methods 45: 905-927.

Koch C, Conradi T, Gossner MM, et al. 2016. Management intensity and temporary conversion to other land-use types affect plant diversity and species composition of subtropical grasslands in southern Brazil. Applied Vegetation Science 19: 589-599.

Legendre P, Legendre L. 2012. Numerical ecology. Kidlington, Elsevier.

Levy EB, Madden EA. 1933. The point method of pasture analysis. New Zealand Journal of Agriculture 46: 179-267.

Lezama F, Pereira M, Altesor A, Paruelo JM. 2019. Grasslands of Uruguay: classification based on vegetation plots. Phytocoenologia (in press). doi: 10.1127/phyto/2019/0215

Londo G. 1976. The decimal scale for releves of permanent quadrats. Vegetatio 33: 61-64.

Magnusson WE, Lima AP, Luizão R, et al. 2005. RAPELD: a modification of the Gentry method for biodiversity surveys in long-term ecological research sites. Biota Neotropica 5:19-24.

Magurran AE. 2004. Measuring biological diversity. Malden, Blackwell.

Manly BFJ. 2007. Randomization, bootstrap, and Monte Carlo methods in biology. Boca Raton, Chapman & Hall/ CRC.

Maarel E. 2007. Transformation of cover-abundance values for appropriate numerical treatment:alternatives to the proposals by Podani. Journal of Vegetation Science 18: 767-770.

Massi KG, Eugênio CUO, Franco CA. 2017. Post-fire reproduction of herbs at a savanna-gallery forest boundary in Distrito Federal, Brazil. Brazilian Journal of Biology 77: 876-886.

Melo AS. 2008. O que ganhamos "confundindo" riqueza de espécies e equabilidade em um índice de diversidade? Biota Neotropica 8: 21-27

Menezes LS, Ely CV, Lucas DB, Silva GH, Boldrini II, Overbeck GE. 2018. Plant species richness record in Brazilian Pampa grasslands and implications. Brazilian Journal of Botany 41: 817-23.

Montgomery DC. 2017. Design and analysis of experiments. New York, John Wiley & Sons.

Mueller-Dombois D, Ellenberg H. 1974. Community sampling: The releve method. In: Mueller-Dombois D, Ellenberg H. (eds.) Aims and methods of vegetation ecology. New York, John Willey & Sons, Inc. p. 45-66.

Munhoz CBR, Araújo GM. 2011. Métodos de amostragem do estrato herbáceo-subarbustivo. In: Felfili JM, Eisenlohr PV, Melo MMRF, Andrade LA, Meira-Neto JAA. (eds.) Fitossociologia no Brasil: métodos e estudos de casos. Viçosa, Editora UFV.p. 213-230.

Munhoz CBR, Felfili JM. 2005. Fenologia do estrato herbáceo-subarbustivo de uma comunidade de campo sujo na Fazenda Água Limpa no Distrito Federal, Brasil. Acta Botanica Brasilica 19: 979-988.

Oksanen L. 2001. Logic of experiments in ecology: is pseudoreplication a pseudoissue? Oikos 94: 27-38.

Oksanen L. 2004. The devil lies in details: reply to Stuart Hurlbert. Oikos 104: 598-605.

Oleques SS, Overbeck GE, Avila RS. 2017. Flowering phenology and plant-pollinator interactions in a grassland community of Southern Brazil. Flora 229: 141-146.

Oliveira TE, de Freitas DS, Gianezini M, et al. 2017. Agricultural land use change in the Brazilian Pampa Biome: The reduction of natural grasslands. Land Use Policy 63: 394-400.

Overbeck GE, Müller SC, Fidelis A, et al. 2007. Brazil's neglected biome: the South Brazilian Campos. Perspectives in Plant Ecology, Evolution and Systematics 9: 101-116.

Overbeck GE, Vélez-Martin E, Scarano FR, et al. 2015. Conservation in Brazil needs to include non-forest ecosystems. Diversity and Distributions 21: 1455-1460.

Perkins LB, Blank RR, Ferguson SD, Johnson DW, Lindemann WC, Rau BM. 2013. Quick start guide to soil methods for ecologists. Perspectives in Plant Ecology, Evolution and Systematics 15: 237-244.

Pike N. 2010. Using false discovery rates for multiple comparisons in ecology and evolution. Methods in Ecology and Evolution 2: 278-282.

Pillar VD. 1998. Sampling sufficiency in ecological surveys. Abstracta Botanica 22: 37-48.

Pillar VD. 1999a. How sharp are classifications? Ecology 80: 2508-2516.

Pillar VD. 1999b. The bootstrapped ordination re-examined. Journal of Vegetation Science 10: 895-902.

Pillar VD. 2013. How accurate and powerful are randomization tests in multivariate analysis of variance? Community Ecology 14: 153-163.

Pillar VD, Orlóci L. 1996. On randomization testing in vegetation science: multifactor comparisons of relevé groups. Journal of Vegetation Science 7: 585-592.

Podani J. 2005. Multivariate exploratory analysis of ordinal data in ecology: Pitfalls, problems and solutions. Journal of Vegetation Science 16: 497-510.

Poore MED. 1955. The use of phytosociological methods in ecological investigations: II. Practical issues involved in an attempt to apply the Braun-Blanquet system. Journal of Ecology 43: 245-269.

Setubal RB, Boldrini II. 2010. Floristic and caracterization of grassland vegetation at a granitic hill in Southern Brazil. Revista Brasileira de Biociências 8: 85-111.

Smith PG. 1964. Quantitative plant ecology. London, Butterworths Scientific Publications.

Staude IR, Vélez-Martin E, Andrade BO, et al. 2018. Local biodiversity erosion in south Brazilian grasslands under moderate levels of landscape habitat loss. Journal of Applied Ecology 55: 1241-1251.

Thomas PA, Overbeck GE, Müller S. 2019. Restoration of abandoned subtropical Brazilian highland grasslands: fast effect of mowing, but not of hay transfer. Acta Botanica Brasilica 33: 405-411.

Vittoz P, Guisan A. 2007. How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. Journal of Vegetation Science 18: 413-422.

Wildi O. 2017. Data analysis in vegetation ecology. 3rd. edn. Boston, CABI Publishing.

Willner W, Kuzemko A, Dengler J, et al. 2017. A higher-level classification of the Pannonian and western Pontic steppe grasslands (Central and Eastern Europe). Applied Vegetation Science 20: 143-158.

Wilson JB 2012. Species presence/absence sometimes represents a plant community as well as species abundances do, or better. Journal of Vegetation Science 23: 1013-1023.