PHYTOSOCIOLOGICAL SURVEYS: TOOLS FOR WEED SCIENCE?[1]

Levantamentos Fitossociológicos: Ferramentas para a Ciência das Plantas Daninhas?

CONCENÇO, G.[2], TOMAZI, M.[3], CORREIA, I.V.T.[4], SANTOS, S.A.[5], and GALON, L.[6]

ABSTRACT - In simple terms, a phytosociological survey is a group of ecological evaluation methods whose aim is to provide a comprehensive overview of both the composition and distribution of plant species in a given plant community. To understand the applicability of phytosociological surveys for weed science, as well as their validity, their ecological basis should be understood and the most suitable ones need to be chosen, because cultivated fields present a relatively distinct group of selecting factors when compared to natural plant communities. For weed science, the following sequence of steps is proposed as the most suitable: (1) overall infestation; (2) phytosociological tables/graphs; (3) intra-characterization by diversity; (4) inter-characterization and grouping by cluster analysis. A summary of methods is established in order to assist Weed Science researchers through their steps into the realm of phytosociology.

Keywords: weed community; density; frequency; dominance; diversity; similarity.

RESUMO - Levantamento fitossociológico, em termos simples, é um grupo de métodos de avaliação ecológica com o objetivo de fornecer uma visão compreensiva tanto da composição como da distribuição de espécies vegetais em uma certa comunidade. Para compreender a aplicabilidade desses levantamentos para a ciência das plantas daninhas, bem como sua validade, devem-se escolher os métodos mais adequados e com base ecológica, uma vez que áreas cultivadas apresentam um grupo relativamente distinto de fatores de seleção, em comparação com os ambientes naturais. Para estudos fitossociológicos de plantas daninhas, a seguinte sequência de passos é proposta como a mais adequada: (1) infestação geral; (2) tabelas ou gráficos fitossociológicos; (3) intracaracterização por diversidade; e (4) intercaracterização e agrupamento por similaridade. Um apanhado dos métodos é apresentado, visando apoiar pesquisadores e estudantes da área de Plantas Daninhas em seus passos no reino da fitossociologia.

Palavras-chave: comunidade infestante; densidade; frequência; dominância; diversidade; similaridade.

INTRODUCTION

A phytosociological survey, in simple terms, is a group of ecological evaluation methods whose aim is to provide a comprehensive overview of both the composition and distribution of plant species in a given plant community. These methods were originally developed for describing relatively stable and solid plant communities, such as forests and prairies, with little to no human intervention (Pandeya et al., 1968), but they are widely used in other areas of knowledge.

In recent years, this group of methods has been vastly applied in studies of agricultural systems and arable fields (Adegas et al., 2010;

[1] Recebido para publicação em 22.11.2012 e aprovado em 9.3.2013.

[2] Agronomist, D.Sc., Weed Science researcher at Embrapa Western Agriculture, Dourados, MS, Brazil, <germani@cpao.embrapa.br>; [3] Agronomist, D.Sc., Climate Change researcher at Embrapa Western Agriculture, Dourados, MS, Brazil, <michely@cpao.embrapa.br>; [4] Undergraduate student in Agronomy, University Anhanguera, trainee in Weed Science at Embrapa Western Agriculture, Dourados, MS, Brazil, <igor.vinicius@aedu.com>; [5] Undergraduate student in Biology, University Unigran, trainee in Weed Science at Embrapa Western Agriculture, Dourados, MS, Brazil, <sabrinak3001@gmail.com>; [6] Agronomist, D.Sc., Weed Science professor at the Universidade Federal da Fronteira Sul, Erechim, RS, Brazil, <galonleandro@ig.com.br>.

Guglieri-Caporal et al., 2010; Fialho et al., 2011), actually assuming an important role for weed science. The term "phytosociology", however, is directly associated with the "structure of an association of plant species". Associations among plant species, although true in nature, are controversial in some aspects because they depend greatly on the effect of biotic and abiotic factors which act on the community (Greig-Smith, 1980). Thus, a given association may be valid only under certain conditions.

To understand the applicability of phytosociological surveys for weed science, their ecological basis need to be understood and the most suitable ones have to be chosen, because it is considered that arable fields have a relatively distinct group of selecting factors when compared to natural plant communities. A plowing operation or a herbicide application is a more powerful and instantaneous selection factor than most of the factors found in a natural, undisturbed forest (Frenedoso-Soave, 2003; Malik et al., 2007). This review will also address ecological concepts, in as much detail as required, to justify the aims and methods suggested for use in weed science. For more specific data, please refer to Pandeya et al. (1968), Barbour et al. (1998), Gurevitch et al. (2009) and Stohlgren (2007).

Synecology and autecology

Ecology may be roughly divided in two sub-sections: synecology and autecology (Barbour et al., 1998). These areas have distinct aims and one should be aware that methods suitable for one of these groups may not be fully applicable to the other because there is the risk of obtaining inaccurate data, as the only common point between them is Evolutionary Ecology (Figure 1).

Autecology deals with the adaptation and behavior of individual species or populations in relation to their environment (Barbour et al., 1998) and it encompasses seed germination (including soil seed banks), reproductive capacity of the species, behavior under distinct light intensities, tolerance to water deficit, plant identification (herbariums) and several other aspects (Pandeya et al., 1968).

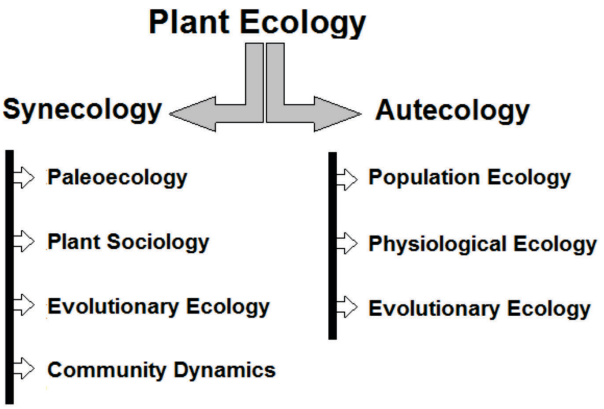


Figure 1 - Theoretical sub-divisions of plant ecology, which guides the nature of the methods adopted for ecological analyses. Embrapa Western Agriculture, Dourados-MS, 2012.

Synecology deals with the community as a whole, including the full set of species present, with all the interactions surrounding it (Barbour et al., 1998). This set is called a "basic unit of vegetation" (Pandeya et al., 1968). Within synecology, plant sociology holds the description and mapping of vegetation types and communities (Barbour et al., 1998).

Two contrasting theories

Due to the nature of the Basic Science of ecology, there is a range of controversies regarding current concepts (Gurevitch et al., 2009). The main discussion surrounding synecological methods is focused on the nature of community. "Community" is defined as a group of populations which co-exist in space and time, interacting with one another directly or indirectly (Gurevitch et al., 2009).

The concept of community is based on the principle of "Associations", which are different clusters of plant species, found generally together in sites with similar environmental conditions. The nature of the relationship among species inside a cluster is the basis for the most controversial point of phytosociological studies.

The discrete view

The first theory regarding plant association and the interdependence of species within the

same community was proposed by Clements (1916). Such theory states that plant communities are very organized entities composed by mutually inter-dependent species – the so called "super-organisms". Thus, the emergence and disappearance of a given plant community could be easily and precisely estimated, because it was considered a sole organism (Clements, 1936).

Two of Clements's most remarkable affirmations were the occurrence of several narrow connections between two or more species, and the cooperation among species for survival (Ludwig & Reynolds, 1988). Clements's theories were widely accepted and rarely challenged while he was alive, mainly because of his energetic and dominant personality (Gurevitch et al., 2009). Clements's theory predicted that the optimal and the amplitude of species were expected to present distinct clusters; hence, changes in vegetation were expected to be abrupt.

The continuum view

In contrast to Clements's ideas, Gleason (1926) believed that communities were a result of interactions both among species and between species and the environment, combined to casual historically extreme climatic events. Gleason defended the idea that each species had its own tolerance to given selection factors; thus, they answered to environmental stresses in particular ways.

According to Gleason, within the range of stress which a species is able to tolerate, casual events determine when a species is actually found in a given place (Gurevitch, 2009). Gleason's theories were confirmed by Curtis & McIntosh (1951). Gleason's theory predicted that the optimal and the amplitude of species were expected to be independent, creating a gradient of occurrence as the environment (e.g. temperature, rain, soil fertility, altitude) changed.

Unifying points and limitations

Currently, both theories contribute with a share to the concept most widely accepted among modern plant ecologists: plant association exists to a certain degree; the gradient of plant composition of clusters is defined by the environment (or management in arable areas); and abrupt changes are observed in the composition of species inside clusters when abrupt selection factors are applied (Pandeya et al., 1968; Barbour et al., 1998; Stohlgren, 2007; Gurevitch et al., 2009). For example, in frequently plowed and harrowed areas, plant species are expected to differ greatly from those in a nearby area grown for several years under no-tillage system, as found by Concenço et al. (2011).

Aims and methods

The aim of phytosociological studies for weed science is similar to that of ecological studies. Weed science researchers should, however, take into account that the nature of agricultural experiments usually implies (1) plots with much smaller size than the one expected for phytosociological samplings; and (2) much stronger selection factors than those acting in the natural environment. Moreover, selection factors are usually momentaneous as the treatments are applied, e.g. distinct crop planting densities, row spacings or crop canopy structure; previous residual or frequent post-emergence herbicides applications, and sometimes the unknown use history of the area.

In this context, the use of phytosociological methods for weed science should be directly associated with the nature of the treatments applied, considering as mandatory a common history for all the area where the whole experiment will be installed, with no differential selection factors other than those comprised by the treatments.

In long-term field trials, phytosociological surveys may be more comprehensively interpreted because of the larger size of the plots and the consolidation of a "system" in each one of the treatments. In addition, the soil seed bank of plant species will tend to be equalized and to reflect more reliably the effects of management. In other words, plant communities in long-term, consolidated trials are usually more closely in conformity with Gleason's theory of gradient occurrence of species as the selection factors are changed.

The methods used in plant sociology rely on two key points: (1) sampling the areas accurately and (2) describing the plant community as clearly as possible so that the data can reflect the real plant community.

Methods for sampling the community

Synecologists seek to understand the degree of species interdependence within communities, how the distribution of communities depends upon past and present environmental factors (in long-term trials), and the role played by communities in such ecosystem or agricultural system (Barbour et al., 1998).

Pandeya et al. (1968) and Barbour et al. (1998) point out several sampling methods, but considering the limitations imposed by experiments in agriculture, only two of them will be addressed in this article: the relevé and the random quadrats methods.

Relevé

The relevé method was improved, if not developed, by Josias Braun-Blanquet, a Swiss ecologist who helped classify much of Europe's vegetation. His sampling method is also referred to as SIGMA, Braun-Blanquet, and Zurich-Montpellier (Barbour et al., 1998). Both this method and the theories involving its application are controversial, because its theoretical background is mainly associated with Clements's theory; thus, it may lead to inaccurate samplings in constantly disturbed environments, e.g., arable fields.

The first step of the relevé method is to draw an species-area curve (Figure 2). For a reliable sampling with the Relevé method, researchers should essentially consider a minimum size of sampled area from the total area, where all species present in the field are represented. This means that the area sampled needs to be calibrated to the environment it will represent. The researcher should start sampling a small area (0.25 m^2 in Figure 2) and count the number of species found in the single quadrat. The size of the quadrat should be progressively increased until the number of species has been stabilized (about 8 m^2 in Figure 2). When the number of species is stabilized, the size of the single quadrat which should be evaluated is defined for that given area.

The main limitation of the relevé (Braun-Blanquet) method is that it assumes almost no variation all along the sampled ecosystem ("superorganism" of Clements). As a consequence, in Braun-Blanquet's view, there is no need to sample more than one point inside the community, since the minimum size of the quadrat is properly calibrated. Some authors, however, adopt the relevé while "splitting" its size in several pieces after

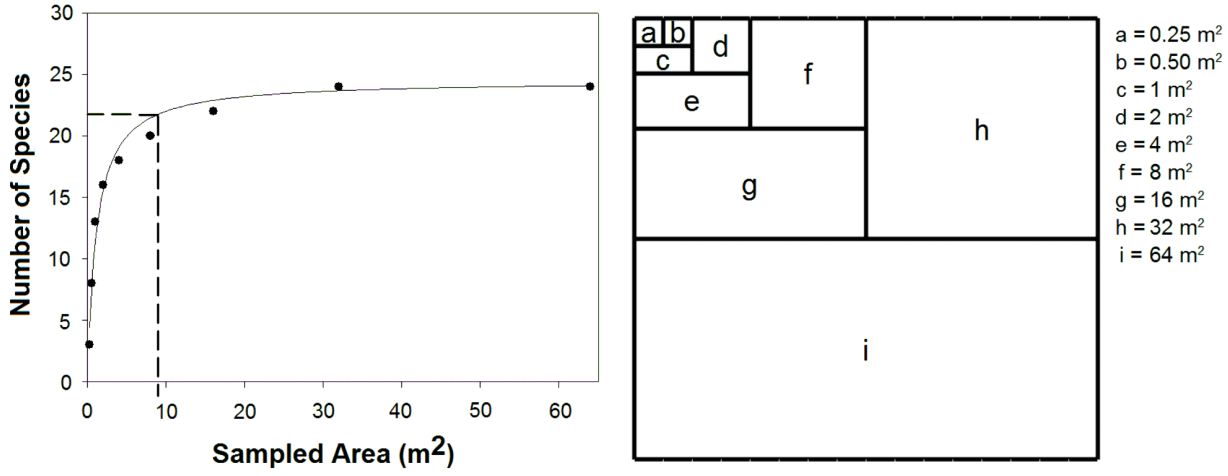


Figure 2 - Calibration of the relevé method – determination of the minimal area of the single quadrat to be sampled for fidelity in terms of number of plant species. Embrapa Western Agriculture, Dourados-MS, 2012.

calibration. This will not, on the one hand, overcome the limitations of this method and may, on the other hand, even eliminate the representative community as well, by positioning each piece of sampling in locations which differ greatly from the original theoretical whole quadrat.

Although Mueller-Dombois & Ellenberg (1974) defend the use of this sampling method, the relevé method should be avoided in agroecosystems because of differences regarding soil type, fertility and natural random dispersion of weeds in the area. In fact, one of the main criticisms to this method is that there will always be a differential bias between two different sampled quadrats with the same area, and thus the calibration based on number of species would not be enough to provide precision to community description (Barbour et al., 1998). This method also makes it difficult to obtain data of frequency for the species (which will be addressed later). In addition, it is difficult to obtain statistical data such as standard error of the samples.

For each type of plant community, the minimum average size of areas to be sampled with the revelé method can be found in Mueller-Dombois & Ellenberg (1974), and it ranges from 0.1 m^2 for lichen communities to 50,000 m^2 for tropical rain forests. Thus, this method can be very human labor-intensive depending on the type and size of the community to be represented by the sample.

Random quadrats

Sampling by random quadrats is widely adopted by North-American ecologists, who are not satisfied with the European vision of simply understanding the structure of a community (Barbour et al., 1998). This method consists in subjectively finding patterns inside the community to be sampled, and sampling in such a way not to favor a particular pattern (Pandeya et al., 1968; Barbour et al., 1998). It means that for data to be as reliable as possible, sampling should be accomplished as randomly as possible. For arable fields, these "patterns" may consist of regions of the field with distinct traits (wet soil opposite to dry soil) or weeds distribution, e.g. a region with predominance of a given species because it is the point from where that species started to spread in the field, or a region with a predominance of species which survived to the last application of herbicides.

In addition to the correct identification of the patterns, the empirical accuracy of this method also relies on the number and size of the individual quadrats (Pandeya et al., 1968). Several considerations are made by Barbour et al. (1998) regarding these aspects, but only the most significant ones for arable fields are highlighted: (1) quadrat shape should preferably be square, with equal side lengths. This will make the sampling less likely to follow a particular pattern (e.g. crop interrows). Rectangular forms result in higher perimeter of the quadrat thus increasing the Edge Effect; (2) the square quadrat size should be as large as possible to dilute the Edge Effect. The Edge Effect is associated with the mistake of the evaluator while deciding if plants close to the border of the sampled area are inside or outside the quadrat.

In addition, for agricultural ecosystems, there are some additional observations: (1) there is no need for statistical replications (experimental design) to allow the use of phytosociological methods for evaluating weed occurrence, once variation comes from the differences among quadrats (descriptive statistics). A minimum representative area of each treatment/community, however, should be sampled, and the community should be as large as possible to dilute external influences; and (2) statistical blocking of experiments in areas with low uniformity (e.g. half area constantly plowed and half area with no tillage) will not give higher reliability for the comparison. The areas to be chosen should be as homogeneous as possible, when experimental design trials are installed to evaluate weed occurrence by phytosociological methods. The only source of variation for statistical data should be the quadrats.

Barbour et al. (1998) cite three rules of thumb, from distinct authors, to be adopted when decisions are made about the size of the quadrat. The most appropriate rule for arable fields is the one proposed by Greig-Smith (1964): the size of the quadrat should be at least twice as large as the average canopy spread of

the largest species. Considering traditional arable fields, it is hypothesized that a quadrat with 0.5m of side length will be suitable for most of the situations.

Although the Accuracy of the sampling cannot be calculated – because it requires knowledge of the true mean of the community, Precision (Pr) is a good indicator of the efficiency of the sampling procedure (Barbour et al., 1998). In the Random Quadrats method, Pr can be calculated with the formula:

$$Pr = \frac{1}{variance\ of\ sample\ means}$$

Pandeya et al. (1968) describes two methods for sampling the area when using the random quadrats method: "random" and "transects". As transects are unlikely to be suitable for arable fields, only "random" will be discussed. As regards random samplings, there are three sub-types: "even spaced", "chance distribution", and "zoned random" (Figure 3).

The even spaced method requires previous knowledge of the area and previous planning of the locations to be sampled, which is usually not a problem for arable fields, and the quadrats are equally distributed in the area (Figure 3A). Chance distribution comprises a completely randomized choice of the locations to be sampled, thus increasing the possibility of not detecting abundant species which are not frequent (this issue will be addressed later), as well as increasing the chance of leaving big gaps of areas of unknown evenness (gray zones in Figure 3B) with no sampling.

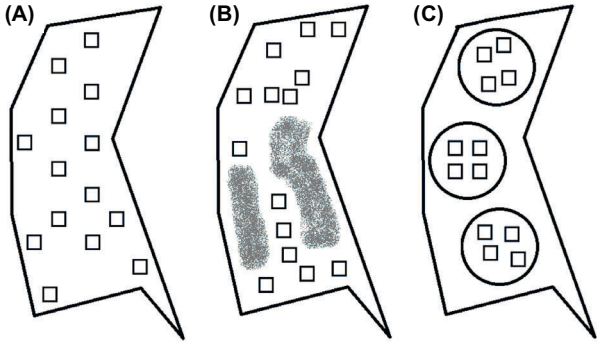


Figure 3 - Distribution of samplings for the random quadrats method. (A) even spaced distribution, (B) chance distribution, and (C) zoned random distribution of quadrats. Embrapa Western Agriculture, Dourados-MS, 2012.

The zoned random consists in previously defined sub-areas with distinct traits, and randomly choosing the locations to be sampled inside each zone. For this method, the number of quadrats to be sampled in each zone will depend on the proportion of the total area it represents (Figure 3C). For example, in pastures under grazing, animals tend to concentrate for overnight in specific locations where most of the feces (and seeds of some weeds) tend to concentrate. If the overnight area represents about 15% of the total, only 15% of the quadrats should be sampled inside that area.

Methods for describing the community

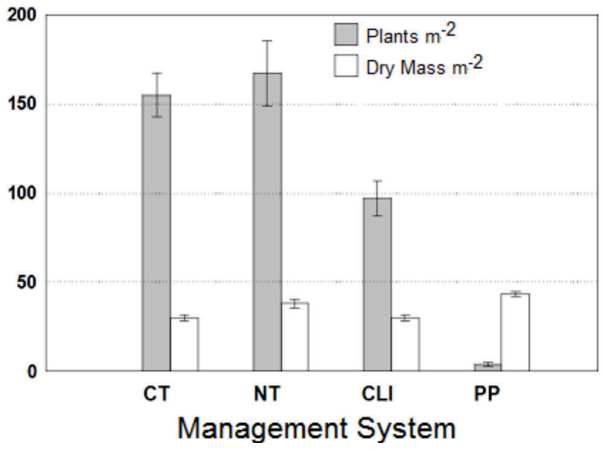
After data are collected in the field, they need to be translated into easily understandable tables and graphs. The most diverse methods are reported in the literature, and researchers are actually encouraged to find their own ways to show their data to the scientific community. Most of the authors, however, agree that both tables and graphics should be used in the same manuscript (with no repeated data), avoiding excessively long tables or excessively summarized graphs.

It is suggested that researchers should use the following sequence for data presentation: (1) explorative graphs, with % of area covered (if available), number of plants and dry mass; (2) importance components, in tables or graphs; (3) intra-population inferences; and (4) inter-population inferences.

A simple exploratory graph showing the number of individuals and dry mass accumulated per area (no need for species distinction) in each treatment, will provide readers with a comprehensive overview of the data to be further explored. Figure 4 shows the results from a long-term trial with four treatments, provided as an example (adapted from Concenço et al., 2011).

Importance components

Importance components are associated with plant traits which turn a given species



CT = conventional tillage; NT = no-tillage; CLI = crop-livestock integration; PP = permanent pasture. Error bars are presented above each column. Adapted from Concenço et al. (2011).

Figure 4 - Illustration of the overall weed infestation of a long-term trial, by treatment.

into a weed inside the community. Several synecological parameters may be considered for the importance of each species in the system (Pandeya et al., 1968; Barbour et al., 1998), namely: abundance, density, cover, frequency, homogeneity, dominance, sociability, vitality, periodicity, constance and fidelity. Differences in application of these terms are observed if the researcher chooses either the relevé or the random quadrats method (Barbour et al., 1998). A summarized survey about these parameters is found in Pandeya et al. (1968). These parameters, however, were developed for ecological studies of natural, undisturbed environments.

Although authors are relatively free to decide which parameters they are going to consider in a particular analysis, the parameters chosen should be as independent as possible. Three of these parameters are suggested as the most significant ones for describing weeds dynamics in arable fields: density, frequency and dominance. Abundance is a rather nebulous term, but it is often used as a synonym for density (Barbour et al., 1998). The clearest definitions for these parameters are found in Barbour et al. (1998), summarized below.

Density (or for instance abundance) is the number of plants rooted within each quadrat. The average density per quadrat of each species can be extrapolated to any convenient unit area. Frequency is the proportion of total quadrats which contains at least one rooted individual of a given species. A Dominant species of a community is the overstory species which contributes the most cover or basal area (in case of large trees) to the community, compared to other overstory species. These parameters are graphically shown in Figure 5. According to Barbour et al. (1998), *frequency* in the random quadrats method is roughly equivalent to *sociability* in the relevé method, because frequency itself cannot be obtained in the latter because of its unique sampling point.

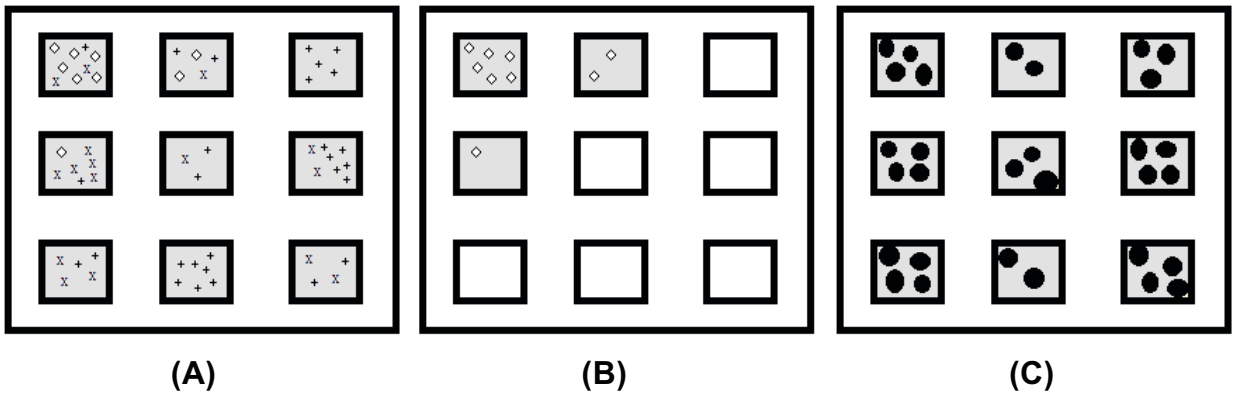


Figure 5 - (A) Density or Abundance (DE), associated with the number of plants of a given species found in all quadrats; (B) Frequency (FR), associated with the number of quadrats where a given species was found, regardless of the number of individuals; (C) dominance (DO), associated with the amount of space in the canopy attributed to a given species, in arable fields measured usually by dry mass accumulation. Embrapa Western Agriculture, Dourados-MS, 2012.

Mueller-Dombois & Ellenberg (1974) consider Density and Abundance as different parameters, but in fact they are both based on the same raw data: number of individuals. Thus, authors who adopt this reference are advised not to use the two parameters in the same study, under the risk of giving more importance to number of individuals compared to species distribution and dry mass accumulation, which would imbalance the Importance Value (IV).

As for the application of management practices to control weed species, although not widely accepted and still not well consolidated, we propose to plan the control of abundant species preferably in pre-emergence; the less frequent species by localized applications or management practices, and the most dominant in post-emergence, preventing them from accumulating mass and dominating the field.

Abundant species are widely distributed in the area; hence the application of pre-emergence herbicides will play an important role in reducing their occurrence. As less frequent species occur in specific locations of the field, in many cases there should be no need to apply the control all over the area in order to eliminate these species. Dominant species, which are not frequent, may present just a few individuals randomly distributed in the field; thus, it will be difficult to locate them before emergence. As a result, locating them in the area in early post-emergence may be the correct time to apply control practices.

Based on the three parameters (density, frequency, dominance), the Importance Value of each species in the community can be easily estimated. The most important weed species will be those with a higher number of individuals (density), widely distributed in the area (frequency), and capable of suppressing the other species as a result of faster growth and mass accumulation (dominance). Thus, the Importance Value for each weed species can be obtained with the formula:

$$\mathrm{IV}(\%) = \frac{(density(\%) + frequency(\%) + dominance(\%))}{3}$$

In addition, none of these parameters need to be presented in the absolute form, and it is advisable to present only the relative scores. In Table 1, provided as an example, grey columns can be suppressed in the final presentation. Authors are also free to decide whether or not they should present the entire list of species found in a given treatment, or only the main weed species, grouping the remaining ones as "others". In the example in Table 1, only the four main species are presented.

For the IV, authors are advised to obtain the mean of the three parameters instead of the simple sum, because in this way the importance value will be associated with a "100%" of importance, e.g., if the IV of *Bidens pilosa* (Table 1) is 32.9, this means that 32.9% of the overall importance for infestation is attributed to that species. This option, however, is more viable when data are presented in

Table 1 - Phytosociological parameters which comprise the Importance Value of infestation of weed species. Embrapa Western Agriculture, Dourados-MS, 2012

Species	DE	DE (%)	FR	FR (%)	DO	DO (%)	IV (%)
Bidens pilosa	15	20.8	8	26.7	325.8	47.2	32.9
Commelina benghalensis	25	34.7	3	10.0	125.7	18.2	21.5
Richardia brasiliensis	12	16.7	5	16.6	36.4	5.3	13.7
Conyza bonariensis	8	11.1	2	6.7	82.7	12.0	10.3
Other species	12	16.7	12	40.0	120.1	17.4	21.6
Total	72	100	30	100	690.7	100	100

NOTE: DE = density; FR = frequency; DO = dominance. This example comprises a theoretical sampling of 8 quadrats per treatment. Thus, the maximum "frequency" to be observed for each species is 8, except for "other species". When simplifying the table from the "complete" to the "main species" model, authors should sum the frequencies of the species involved, as should be done with DE and DO.

tables (Table 1).The graphical representation of the same data (in stacked bar graphs) is shown in Figure 6, with the limitation that in the graphical form, the IV will be by default presented in the basis of 300 instead of 100, otherwise DE (or AB), FR and DO would have to be each divided by three. Barbour et al. (1998) present other optional graphs (pictograms) which are suitable for representing data graphically, mainly when a few areas and a few weed species are being evaluated.

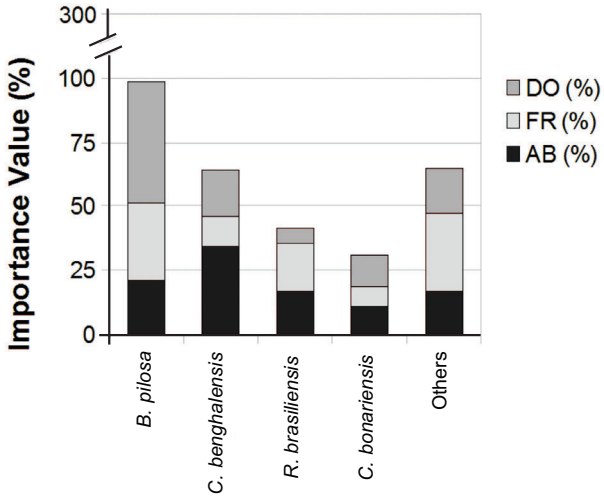


Figure 6 - Graphical representation of survey data of AB (or DE), FR, DO and IV in relative terms. Embrapa Western Agriculture, Dourados-MS, 2012.

After sampling is finished, the precision of sampling (Pr) can be calculated as previously stated. For this purpose, researchers are free to decide whether or not they should use abundance (number of individuals per quadrat) or dominance (total dry mass per quadrat) for that. It is not defined in the literature which of them is the most appropriate, but authors are advised to use abundance to represent the precision of sampling.

Diversity indexes

There are three types of diversity: (1) diversity of differentiation, (2) diversity of standard and (3) diversity of inventory (Gurevitch et al., 2009). We will focus only on diversity of inventory. A diversity index is a statistic which is intended to understand the variety of individuals of a given population, thus allowing inferences about a particular plant community in terms of both the number of species found and the balancing in the number of individuals per species (Barbour et al., 1998). These indexes allow the intra-characterization of each area, supplying additional support for researchers' inferences about a given community. In general, intra-population inferences do not properly receive the deserved attention, and authors are encouraged to explore this aspect.

The most widely used diversity indexes are Margalef (α), Menhinick (Dm), Simpson (D) and modified Shannon-Weiner (H'), in addition to density of species itself (Gurevitch et al., 2009). Simpson's D relates the likelihood that two randomly selected individuals from an infinitely large community will belong to different species (Simpson, 1949). The Simpson index considers the abundance of species in the sample while being less sensitive to species richness (Simpson, 1949; Barbour et al., 1998). Giavelli et al. (1986) state that D is less prone to errors because of factors associated with sampling problems, and it should be chosen instead of H', and α was reported for its notable statistical imprecision. In addition, Simpson's D gives very little weight to rare species, and it is most sensitive to the numbers of abundant species; Shannon-Weiner's H', in contrast, is more sensitive to rare species; this is where sampling errors may be most pronounced (Barbour et al., 1998).

As D and H' are theoretically distinct and affected in different ways by rare or abundant species, authors are advised to use both indexes in order to make inferences about the diversity of a given plant community. There are optional formulas with different parameters to calculate both D and H', but the easiest ones to use and, hence, less prone to errors, are the ones supplied by Barbour et al. (1998):

$$D = 1 - \Sigma (pi)^2$$

$$H' = - \Sigma (pi)(\log_2 pi)$$

where p_i = proportion of all individuals in the sample which belong to species i. Thus, by using the formulas supplied by Barbour et al. (1998), only the relative abundance (divided by

100) is necessary for obtaining D and H'. More recently, the Natural Log (ln) is being used instead of Log base 2 for obtaining H'. Although it makes less sense, it also makes no real difference for the final value. Authors are suggested to use log base 2 for H', but ln is also valid. Data of relative abundance from Table 1 will be used to illustrate both, adopting $\log_2$ for H':

$$D = 1 - [(0.208)^2+(0.347)^2 +(0.167)^2+ (0.111)^2+(0.167)^2] = 0.77$$

$$H' = - [(0.208*-2.26)+(0.347*-1.53)+(0.167* -2.58)+(0.111*-3.17)+(0.167*-2.58)=2.22$$

As both coefficients are affected differently by abundant or rare species, H' will usually be more appropriate for areas recently submitted to big changes in management (e.g. shifted from conventional tillage to sod seeding), where different species start to appear as a consequence of the new environment. In this situation, higher H' would usually tend to represent higher environmental sustainability of the cropping system.

Simpson's D, in contrast, is more appropriate for well consolidated areas where no abrupt changes were recently implemented. For example, long-term RoundupReady® soybean fields tend to present high infestation of *Commelina benghalensis*, *Conyza* spp. and *Ipomoea spp.*, which are also the most important weeds; thus, D will more accurately reflect the diversity in this area as it is mostly weighted by abundant species.

As an example of differences in the application of such coefficients, Table 2 shows both D and H' from areas submitted to distinct managements for 16 years (Concenço et al., 2011).

For the situation shown in Table 2, Simpson's D showed a higher diversity for the CT area compared to NT while H' indicated higher diversity for NT compared to CT. This means that the most important weeds in CT are the most abundant – probably the ones largely selected by management. In contrast, H' indicated that diversity may be on the increase at the NT area because of the emergence of some new plant species not present at CT.

Table 2 - Diversity coefficients for weed occurrence in areas submitted to distinct types of management for 16 years. Embrapa Western Agriculture, Dourados-MS, Brazil, 2011

Treatment[1/]	D[2/]	H' [2/]
CT	0.64	0.43
NT	0.59	0.47
CLI	0.02	0.02
PP	0.63	0.50

[1/] CT = conventional tillage; NT = no-tillage; CLI = crop-livestock integration; PP = permanent pasture. Adapted from Concenço et al. (2011). [2/] D = Simpson; H' = Shannon-Weiner (based on density).

Stohlgren (2007) reports that low productivity (high stress) areas usually present low diversity, but this is also true for very productive sites, as a result of competitive exclusion (a link with autecological studies); high diversity is usually observed in sites with intermediate productivity. As a consequence, long-term fertilization tends to decrease plant diversity because it will select those species with higher ability to use a given fertilizer. This helps to explain why stressed areas usually increase their diversity as they are recovered from stress, thus highlighting the importance of diversity indexes for inferences in long-term field trials.

Another widely used coefficient is the Evenness (E') one, based on Shannon-Weiner's diversity. This index reflects the degree of dominance of species in a given community (Magurran, 2003). McManus & Pauly (1990), however, propose the use of a derivative coefficient: Shannon-Weiner Evenness Proportion, which is able to evaluate trends of stress in a given environment over time. This coefficient seems to be applicable for phytosociological studies, with three advantages: (1) it considers both Density and Dominance, creating a new link of synecology with competition studies (autecology); (2) it allows inferences about ecological stress from static data; and (3) it allows inferences about stressing factors in long-term trials over time – in fact, it was developed with this aim.

$$SEP = \frac{H' \, dominance}{H' \, density}$$

where SEP = Shannon-Weiner Evenness Proportion; *H' dominance* = Shannon-Weiner based on biomass; *H' density* = Shannon-Weiner based on number of individuals. Authors are encouraged to read the original study (McManus & Pauly, 1990) for further information about SEP.

Multivariate analysis

Because the diversity indexes allow inferences about each given studied area with no comparison across areas, there is the need to adopt statistical methods which allow researchers to infer which areas are similar in terms of weed infestation. This can be accomplished in two ways: univariate or multivariate analysis (Gurevitch et al., 2009).

Univariate analysis consists in studying individually each one of the traits evaluated for the group of communities, e.g. if the dry mass of weeds or the number of weed individuals per area is equal among communities. For this purpose, usual tests such as ANOVA and the subsequent multiple mean comparison can be used (Thornley, 1976), and each quadrat is considered as a "replication" if the area under study is a trial with no experimental design (like observation units).

Multivariate analysis, however, focuses on a group of traits evaluated for all communities which, when put together, allow the estimation of the differences among communities through a complex way which yields a "distance" between each pair of communities (Barbour et al., 1998). For this purpose, the Euclidean distance is usually used (Danielsson, 1980). Authors can refer to Podani (2000) for further information about multivariate analysis in biological systems.

Clustering by similarity

For phytosociological studies, the distance based on a set of community characters is usually not the most suitable technique for verifying the level of resemblance of a given pair of areas or communities. Gurevitch et al. (2009) reports that the abundance of species, the main trait used for comparing plant communities, is not usually a simple relation as assumed for a usual multivariate cluster analysis. Based on this, communities in phytosociological studies should be grouped based on binary data, e.g. presence or absence of each weed species in each community. The most frequently used binary similarity coefficients (all based on number of individuals) are Jaccard, Sørensen, Sørensen-Dice, simple combination, Ochioi, and asymmetric similarity (Barbour et al., 1998; Gurevitch et al., 2009). Among them, the most accurate for most situations is Jaccard (J):

$$J = \frac{c}{a + b - c}$$

where J = Jaccard similarity index; a = total number of plant species in area "*a*"; b = total number of plant species in area "*b*"; c = total number of plant species common to areas "*a*" and "*b*". J may also be presented in a slightly different way, by adding "*c*" instead of subtracting it. In fact, as this index aims to attribute higher similarity to areas with the highest numbers of plant species in common, it is nonsense to add "*c*" to the denominator of J, because it will decrease the similarity for areas which are actually more similar; thus, authors are strongly advised to adopt the formula presented by Barbour et al. (1998). For more detailed studies, the same author also presents an additional formula: Jaccard weighted by cover (Jc).

These data will be used for generating the similarity matrix. For this purpose, first the researcher needs to analyze the list of species for each area. The following data were extracted from the raw tables of Concenço et al. (2012), supplied as an example (Table 3).

Table 3 - Number of plant species required for generating the similarity matrix. Embrapa Western Agriculture, Dourados-MS, 2012

Area	# of species	Area	# of species
1	13	1x3	7
2	6	1x4	7
3	11	2x3	5
4	9	2x4	5
1x2	5	3x4	7

Example from raw data of Concenço et al. (2012). NOTE: area crossings indicate number of species common to both areas.

Table 4 shows the similarity matrix, obtained from the data in Table 3 by using the Jaccard similarity coefficient (J). Some authors prefer to present only the data in Table 4 and not proceed to the cluster analysis. This is correct and can be done, and for these authors, it is advised to consider two areas as "similar" if the Jaccard coefficient between areas is equal to or higher than 0.25 (Mueller-Dombois & Ellenberg, 1974), or when Sørensen's coefficient is equal to or higher than 0.50 (Felfili & Venturoli, 2000). Authors, however, are advised to proceed and present the cluster analysis by similarity (Figure 7). In this case, Tables 3, 4 and 5 should not be presented as they are intermediary steps for cluster analysis.

Most software products, however, are not able to generate a cluster analysis from similarity data, and stating that areas are either "equal" or "not different" actually means distinct things in statistical terms (Thornley, 1976; Ludwig & Reynolds, 1988). The dissimilarity matrix should be generated from the similarity matrix by "1-J" (Table 5).

The dissimilarity matrix should be supplied to an appropriate statistical software product for cluster analysis – the software must be informed it is a dissimilarity matrix. The statistical environment R (R-Development, 2012) is highly recommended for this task, but several other software products will also be suitable. Areas should be grouped by cluster analysis considering the qualitative trait only (presence or absence of the species), according to the dissimilarities obtained from the inverse of Jaccard's similarity matrix. Hierarchical grouping should be obtained from the distance matrix (dissimilarities) (Barbour et al., 1998)

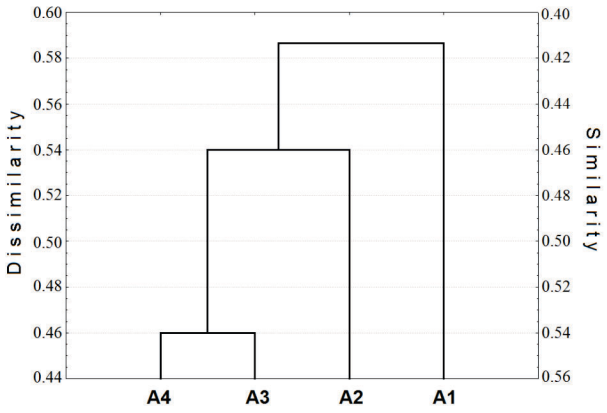


Figure 7 - Areas and/or treatments grouped by cluster analysis for the raw data from Concenço et al. (2012), based on the UPGMA method.

by using the *Unweighted Pair Group Method with Arithmethic Mean* (UPGMA) (Sneath & Sokal, 1973). The final expected result for the cluster analysis is shown in Figure 7.

After cluster analysis, grouping validation should be accomplished by the cophenetic correlation coefficient, obtained by the Pearson linear correlation between the cophenetic matrix and the original matrix of distances (Sokal & Rohlf, 1962). The cophenetic matrix is easily obtained under statistical environments like R (R-Development, 2012). The cophenetic coefficient should be equal or above 0.85, which indicates that the grouping properly reflects the original data (Sokal & Rohlf, 1962). In the example given, the cophenetic coefficient was equal to 0.98 (areas and/or treatments were reliably grouped by the cluster analysis).

Additional data are needed for a complete cluster analysis in order to determine the

Table 4 - Similarity matrix based on Jaccard's similarity coefficient. Embrapa Western Agriculture, Dourados-MS, 2012

Areas	A1	A2	A3	A4
A1	1	0.36	0.41	0.47
A2	0.36	1	0.42	0.50
A3	0.41	0.42	1	0.54
A4	0.47	0.50	0.54	1

Table 5 - Dissimilarity matrix based on Jaccard's similarity coefficient. Embrapa Western Agriculture, Dourados-MS, 2012

Areas	A1	A2	A3	A4
A1	0	0.64	0.59	0.53
A2	0.64	0	0.58	0.50
A3	0.59	0.58	0	0.46
A4	0.53	0.50	0.46	0

threshold level (either by similarity or dissimilarity) to be used for determining the number of homogeneous groups. This is usually an empirical task to be chosen among distinct criteria available in specialized bibliography, but authors are encouraged to define the threshold level by the simple mean of the matrix (either similarity or dissimilarity, according to the scale at the graph). This mean, however, should not consider matching areas ("1s" at the similarity and "0s" at the dissimilarity matrix). Thus, the proposed threshold level for Figure 7 would be 0.45 (for the similarity scale), or 0.55 (for the dissimilarity scale). As a consequence, only area 1 is considered as distinct from the others in terms of composition of infestation, at 45% similarity (Figure 7).

Phytosociological surveys are useful as tools to shed light on the dynamics of weed species and their interactions in arable fields. The methods, however, are the most diverse as several indexes and coefficients are available, depending on the literature used as a reference by a given author. Basic care should be taken, however, when sampling and describing the plant community as well.

For weed science, the following sequence of steps is proposed as the most suitable for a phytosociological survey: (1) overall infestation; (2) phytosociological tables/ graphs; (3) intra-characterization by diversity; (4) inter-characterization and grouping by cluster analysis. Any other set of data or way of presentation, however, may still be adequate depending on the nature of the environment that is being studied.

The literature is definitely not clear about methods for phytosociological surveys, and the authors were not able to find all the set of information in the same source. Even classical references miss some important aspects of phytosociological studies. In this review, a summary of methods was made in order to to assist Weed Science researchers through their steps into the realm of phytosociology.

LITERATURE CITED

ADEGAS, F. S. et al. Levantamento fitossociológico de plantas daninhas na cultura do girassol. **Planta Daninha**, v. 28, n. 4, p. 705-716, 2010.

BARBOUR, M. G. et al. **Terrestrial plant ecology**. Menlo Park: Benjamin/Cummings, 1998. 688 p.

CLEMENTS, F. E. Nature and structure of the climax. **J. Ecol.**, v. 24, n. 2, p. 252-284, 1936.

CLEMENTS, F. E. **Plant succession**: an analysis of the development of vegetation. Washington, DC: Carnegie, 1916. (Publication, 242).

CONCENÇO, G. et al. Phytosociology in agricultural areas submitted to distinct wintercropping management. **Planta Daninha**, v. 30, n. 2, p. 297-304, 2012.

CONCENÇO, G.; SALTON, J. C.; CECCON, G. **Dinâmica de plantas infestantes em sistemas de cultivo**. Dourados: Embrapa Agropecuária Oeste, 2011. 49 p. (Embrapa Agropecuária Oeste. Documentos, 114).

CURTIS, J. T.; McINTOSH, R. P. An upland forest continuum in the prairie-forest border region of Wisconsin. **Ecology**, v. 32, n. 3, p. 476-498, 1951.

DANIELSSON, P. Euclidean distance mapping. **Comput. Graphics Image Process.**, v. 14, n. 1, p. 227-248, 1980.

FELFILI, J. M.; VENTUROLI, F. Tópicos em análise de vegetação. **Comum. Técn. Flor.**, v. 2, n. 2, p. 34-44, 2000.

FIALHO, C. M. T. et al. Fitossociologia da comunidade de plantas daninhas na cultura da soja transgênica sob dois sistemas de preparo do solo. **Sci. Agr.**, v. 12, n. 1, p. 9-17, 2011.

FRENEDOSO-SOAVE, R. C. Phytosociological studies on natural establishment of vegetation in an unreclaimed limestone mining. **Braz. Arch. Biol. Technol.**, v. 46, n. 2, p. 259-269, 2003.

GIAVELLI, G.; ROSSI, O.; SARTORE, F. Comparative evaluation of four species diversity indices related to two specific ecological situations. **Field Stud.**, v. 6, n. 3, p. 429-438, 1986.

GLEASON, H. A. The individualistic concept of the plant association. **B. Torrey Bot. Club**, v. 53, n. 1, p. 7-26, 1926.

GREIG-SMITH, P. The development of numerical classification and ordination. **Vegetatio**, v. 42, n. 1, p. 1-9, 1980.

GREIG-SMITH, P. **Quantitative plant ecology**. London: Butterworths, 1964. 256 p.

GUGLIERI-CAPORAL, A.; CAPORAL, F. J. M.; POTT, A. Phytosociology of sown pasture weeds under two levels of degradation in Brazilian savanna areas, Mato Grosso do Sul State, Brazil. **Pesq. Agropec. Trop.**, v. 40, n. 3, p. 312-321, 2010.

GUREVITCH, J.; SCHEINER, S. M.; FOX, G. A. **Ecologia vegetal**. Porto Alegre: Artmed, 2009. 574 p.

LUDWIG, J. A.; REYNOLDS, J. F. **Statistical ecology**. New York: J. Wiley, 1988. 337 p.

McMANUS, J. W.; PAULY, D. Measuring ecological stress: variations on a theme by R. M. Warwick. **Marine Biol.**, v. 106, n. 2, p. 305-308, 1990.

MAGURRAN, A. E. **Measuring biological diversity**. London: Blackwell Publishing, 2003. 260 p.

MALIK, N. Z.; ARSHAD, M.; MIRZA, S. N. Phytosociological attributes of different plant communities of Pir Chinasi hills of Azad Jammu and Kashmir. **Intern. J. Agric. Biol.**, v. 9, n. 4, p. 569-574, 2007.

MUELLER-DOMBOIS, D.; ELLENBERG, H. **Aims and methods of vegetation ecology**. New York: Wiley, 1974. 547 p.

PANDEYA, S. C.; PURI, G. S.; SINGH, J. S. **Research methods in plant ecology**. New York: Asia Publishing House, 1968. 272 p.

PODANI, J. **Introduction to the exploration of multivariate biological data**. Leiden: Backhuys Publishers, 2000. 407 p.

R DEVELOPMENT. **R:** a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2012. Available at: <http://www.R-project.org/>. Access in: 10, Aug, 2012.

SIMPSON, E. H. Measurement of diversity. **Nature**, v. 163, n. 4148, p. 688, 1949.

SNEATH, P. H.; SOKAL, R. R. **Numerical taxonomy**: the principles and practice of numerical classification. San Francisco: W. H. Freeman, 1973. 573 p.

SOKAL, R. R.; ROHLF, F. J. The comparison of dendrograms by objective methods. **Taxon**, v. 11, n. 1, p. 33-40, 1962.

STOHLGREN, T. J. **Measuring plant diversity:** lessons from the field. New York: Oxford University Press, 2007. 390 p.

THORNLEY, J. H. M. **Mathematical models in plant physiology**. London: Academic Press, 1976. 318 p.