



ANÁLISE MULTIVARIADA: IMPORTÂNCIA DA TÉCNICA DE COMPONENTES PRINCIPAIS NA ANÁLISE FITOSSOCIOLÓGICA

Vitor Hugo Zanette¹, Sonia Maria Kurchaidt², João Vitor Zanette³, Henrique Soares Koehler⁴.

1. Doutorando em Engenharia Florestal, Universidade Federal do Paraná, Curitiba, PR; zanette@unicentro.br
2. Doutora em Engenharia Florestal, Professora Adjunta da Universidade Estadual do Centro-Oeste, Guarapuava, PR
3. Mestrando em Engenharia Aeronáutica, Instituto Tecnológico de Aeronáutica – ITA, São José dos Campos, SP
4. Engenheiro Florestal, Doutor em Engenharia Florestal, Professor Adjunto da Universidade Federal do Paraná, Curitiba, PR

Recebido em: 30/09/2014 – Aprovado em: 15/11/2014 – Publicado em: 01/12/2014

RESUMO

Este trabalho foi realizado com dados de quatro parcelas permanentes, coletadas anualmente pelo “Programa Ecológico de Longa Duração” (PELD), localizado em um fragmento de Floresta Ombrófila Mista em São João do Triunfo, PR, nos anos de 1995, 2000, 2005 e 2010, com o intuito de verificar se as estimativas fitossociológicas de densidade, frequência, dominância, valor de importância e valor de cobertura, apresentam diferenças quanto às cargas na análise fatorial. Também objetivou verificar quantos componentes principais são selecionados nos anos estudados. Os resultados mostram que o uso das variáveis fitossociológicas de densidade, frequência, dominância, valor de importância e valor de cobertura não apresentam diferenças nas cargas quando realizada a análise fatorial, indicando semelhança das parcelas quanto a sua composição. Observa-se, também, que foi selecionado somente um componente principal nas três situações em análise, sendo que o componente selecionado explica muito bem a variância em todos os anos estudados.

PALAVRAS-CHAVE: estimativas fitossociológicas, floresta de Araucária, fatores.

MULTIVARIATE ANALYSIS : IMPORTANCE OF TECHNIQUE OF PRINCIPAL COMPONENTS ANALYSIS IN PHYTOSOCIOLOGY

ABSTRACT

This study was conducted with data from four permanent plots, collected annually by the "Long Term Ecological Program" (PELD), located in a fragment of Araucaria Forest in São João do Triunfo, PR, in 1995, 2000, 2005 and 2010, in order to verify that the phytosociological estimates density, frequency, dominance, importance value and coverage value differ as to the charges in the factor analysis. Also aims to check how many principal components are selected in the years studied. The results show that the use of the phytosociological variables density, frequency, dominance, importance value and coverage value do differ in charges when performed factor analysis, indicating similarity of the plots on its composition. It is observed also that

was selected only a major component in the three situations under analysis, and the selected component explains very well the variance in all years studied.

KEYWORDS: phytosociological estimates, Araucaria forest, factors.

INTRODUÇÃO

O Brasil possui a flora arbórea mais diversificada do mundo. A falta de direcionamento técnico e conscientização ecológica na exploração dos recursos florestais têm acarretado prejuízos irreparáveis (LORENZI, 2000).

Pode-se manter ou alterar o ecossistema levando-se em consideração os métodos de manejo, podendo alterá-lo de tal forma que os danos causados sejam irreversíveis. Neste sentido, há necessidade do estudo da vegetação com o intuito de obter subsídios para a pesquisa aplicada e de gestão, possibilitando a preservação, conservação do ambiente e a interpretação do potencial do solo para o uso na pecuária, agricultura e florestas (KURCHAIDT, 2014).

Verifica-se assim, a necessidade da realização de estudos que retratem a realidade da floresta, no que tange a florística e sua estrutura para que essas informações possam ser úteis na elaboração e planejamento de ações que objetivem a conservação ou mesmo a recuperação dessa formação florestal, procurando no máximo retratar a sua diversidade (KURCHAIDT, 2014).

LAMPRECHT (1990) cita que além dos sistemas de classificação das florestas terem validade universal, pelo menos, para todos os tipos de florestas tropicais, devem permitir que todos os resultados de levantamentos possam ser comparáveis entre si, de preferência com métodos estatísticos. Segundo SOUZA (2003), os estudos fitossociológicos tratam de levantamento de dados florísticos com diagnóstico qualitativo e quantitativo de uma área de floresta com determinado tipo de vegetação característica.

Os principais parâmetros a serem obtidos a partir dos dados são: densidade, frequência e dominância, posição sociológica, regeneração natural, valor de importância, valor de cobertura, índices fitossociológicos de comparação entre as parcelas e índices como o de homogeneidade das espécies. Outros pontos relevantes que influenciam a análise e que diferem os resultados fisiológicos em áreas do mesmo bioma são a dominância e a presença ou não de algumas espécies, e até mesmo as diferenças características individuais das áreas (KURCHAIDT, 2014).

Segundo REIS (1997) a Análise Fatorial (AF) é definida como um conjunto de técnicas estatísticas cujo objetivo é representar ou descrever um número de variáveis iniciais a partir de um menor número de variáveis. Na Análise Fatorial (AF) são importantes dois conceitos: o de ortogonalidade e o de carga fatorial. O primeiro está relacionado com independência, no qual deve haver dissociação entre as variáveis. Isto se consegue quando se realiza a Análise de Componentes Principais - (ACP), na qual cada componente é independente do outro e por isso a ACP é geralmente utilizada como uma técnica para se extrair fatores. O segundo conceito é de carga fatorial. A matriz de carga fatorial é um dos passos finais da análise fatorial. A carga fatorial é um coeficiente, um número decimal, positivo ou negativo, geralmente menor do que 1, que expressa o quanto um teste, ou variável observada está carregada, ou saturada em um fator (MANLY, 2008).

De acordo com MINGOTI (2005), a técnica estatística ACP foi introduzida por Karl Pearson em 1901, sendo um método fatorial, pois a redução do número dos

caracteres não se faz por uma simples seleção de alguns deles, mas pela construção de novos caracteres sintéticos obtidos pela combinação dos caracteres iniciais por meio dos “fatores”. É um método linear, pois se trata de combinações lineares (LIRIO, 2004).

A escolha do número de fatores é uma tarefa das mais importantes de uma AF. Segundo HAIR et al. (2005) se o pesquisador opta por um número muito reduzido, ele pode não identificar estruturas importantes existentes nos dados e, por outro lado, se o número é excessivo, ele pode vir a ter problemas de interpretação dos fatores. Como regra deve-se procurar um compromisso entre o número de fatores (a princípio o menor possível) e a sua interpretabilidade.

Desta forma, este trabalho teve como objetivo analisar se as estimativas fitossociológicas da formação vegetal na região de São João do Triunfo-PR, apresentam diferenças quanto às cargas ao serem submetidas à análise fatorial. Bem como, identificar o número de componentes responsáveis pela maior parte da variação dos dados originais por meio da análise de componentes principais e ainda identificar quais das estimativas fitossociológicas analisadas são as mais importantes.

MATERIAL E MÉTODOS

Os dados utilizados para o desenvolvimento deste trabalho são oriundos de parcelas permanentes de inventário florestal integrantes do “Programa Ecológico de Longa Duração” (PELD) localizadas na Estação Experimental de São João do Triunfo, da Universidade Federal do Paraná. A área de 32 hectares da estação está localizada no município de São João do Triunfo, região Centro Sul do Estado do Paraná, a 780 m de altitude, 25°34’18” de latitude S e 50°05’56” de longitude W. Os dados são coletados anualmente desde 1995.

A vegetação da área é classificada como Floresta Ombrófila Mista, de 400 a 1000m de altitude. A área florestal foi dividida em três parcelas de 100 x 100m (1ha) e uma parcela de 50 X 100 (0,5 ha). Os dados são coletados anualmente em quatro parcelas, denominadas, somente com propósito de identificação das mesmas, como Araucária, Imbuia, Rio e Fogo. Cada parcela é dividida em sub parcelas de 10 x 10 m, nas quais todas as árvores com diâmetros à altura do peito (DAP) igual ou superior a 10 cm são medidas com fita métrica ou trena. As medições usadas neste trabalho são referentes aos anos de 1995, 2000, 2005 e 2010, tendo em vista que estudos preliminares mostram que não há muita variação de crescimento da floresta no período menor do que cinco anos.

Foram encontradas e quantificadas nesta área as espécies ocorrentes. A quantificação foi realizada pelo uso de três parcelas de 1ha e uma parcela de 0,5 ha, tomadas no todo, denominadas de Araucária, Imbuia, Rio e Fogo, nos anos de 1995, 2000, 2005 e 2010, após, por meio, do software FLOREXCEL foram calculadas as estimativas fitossociológicas para os anos em estudo, de posse das estimativas foi realizada a análise fatorial e de componentes principais, fazendo uso do software SPSS versão 17.

RESULTADOS E DISCUSSÃO

As árvores foram agrupadas e verificado quantas vezes cada espécie ocorreu dentro de cada sub parcela. O Quadro 1 mostra o número de espécies encontradas nos quatro anos estudados nas quatro parcelas definidas e também considerando essas quatro parcelas em conjunto, que foi denominada Geral. Observa-se no

Quadro 1 o número de indivíduos que não foram identificados representando a dificuldade de identificação dessas espécies, que ocorreu pela impossibilidade de obtenção de material botânico que permitisse a identificação.

QUADRO 1 - Número de espécies por área e por ano estudado

ÁREA	ANOS	Nº ESPÉCIES	Nº IND. NÃO IDENTIFICADOS
GERAL	1995	79	14
	2000	79	12
	2005	81	14
	2010	84	16
ARAUCARIA	1995	41	6
	2000	41	6
	2005	43	6
	2010	42	7
IMBUIA	1995	57	25
	2000	53	20
	2005	54	23
	2010	59	28
RIO	1995	39	11
	2000	41	18
	2005	43	20
	2010	41	20
FOGO	1995	47	10
	2000	49	8
	2005	50	9
	2010	54	11

Os dados foram agrupados por espécies, sendo que inicialmente considerou-se a área como um todo para cada um dos anos estudados. A seguir realizou-se o cálculo das estimativas dos parâmetros fitossociológicos, para cada um dos anos em cada uma das parcelas.

Com as estimativas dos parâmetros, densidade (D), frequência (F), dominância (Do), valor de importância (VI) e valor de cobertura (VC), foi realizada a Análise de Componentes Principais, para a área conjunta das quatro parcelas e para cada uma das quatro parcelas separadamente.

A Tabela 1 mostra os resultados da análise de componente principais, nas quais foram consideradas as cinco estimativas dos parâmetros fitossociológicos. Observa-se que no ano de 1995, as áreas Araucária, Imbuia, Rio e Fogo, não mostram o valor do teste de KMO. Isto se deve ao fato de que o determinante da matriz de correlação R, foi igual a zero, indicando que há multicolinearidade entre os dados. Outro fato que se pode avaliar é o Teste de Bartlett, que verifica a hipótese nula de que a matriz de correlação original é uma matriz identidade. Rejeita-se esta hipótese quando há significância ($p < 0,05$) para todos os dados analisados,

indicando que a matriz R não é matriz identidade. Essa técnica também foi utilizada por SALOMÃO et al. (2012) em estudo na Floresta Ombrófila Densa, em que os resultados respaldam o emprego da análise fatorial para a extração de fatores e a estimação dos escores fatoriais. CARVALHO (2009) também reporta a utilização de técnicas estatísticas multivariadas na análise de desenvolvimento da comunidade vegetal da Caatinga e que divisão de grupos ecológicos mostrou-se eficiente.

A Tabela 1 mostra também que foi selecionado somente um componente, retirado pelo critério de Kaiser, com autovalores maiores que um. Os autovalores associados com cada componente representam a variância explicada por aquele componente. Assim, para todas as áreas no ano de 1995, o componente um explica 77,29% da variância total, e os demais componentes explicam pequenas variâncias e por isso não são considerados.

O próximo item a ser considerado na Tabela 1 são as cargas de cada variável em cada componente. As cargas são obtidas pelo método Varimax, e sua função foi maximizar a dispersão das cargas dentro dos componentes, pode-se observar que estão muito próximos de 1 ou 100%.

TABELA 1 – Análise de componentes principais para as estimativas dos parâmetros fitossociológicos .

ANO	ÁREA	KMO	BARTLET	Nº DE COMP.	VAR. EXPL %	CARGAS				
						VI	VC	D	Do	F
1995	TODAS	0,630	1515,44*	1	77,29	0,850	0,855	0,897	0,887	0,906
	ARAUC	-	-	1	95,25	0,999	0,991	0,987	0,953	0,947
	IMBUIA	-	-	1	89,58	0,999	0,983	0,926	0,886	0,935
	RIO	-	-	1	95,27	0,999	0,990	0,994	0,971	0,925
	FOGO	-	-	1	92,28	0,999	0,994	0,976	0,965	0,971
2000	TODAS	0,696	2394,99*	1	96,29	0,999	0,993	0,987	0,966	0,961
	ARAUC	0,619	1539,29*	1	95,21	0,999	0,991	0,985	0,952	0,950
	IMBUIA	0,578	1725,23*	1	88,77	0,998	0,981	0,915	0,880	0,932
	RIO	0,582	1467,65*	1	94,65	0,998	0,989	0,993	0,966	0,917
	FOGO	0,647	1741,66*	1	95,91	0,999	0,993	0,972	0,962	0,971
2005	TODAS	0,693	2452,95*	1	95,53	0,999	0,991	0,982	0,957	0,957
	ARAUC	0,629	1604,65*	1	94,78	0,999	0,991	0,979	0,946	0,951
	IMBUIA	0,560	1765,17*	1	87,75	0,998	0,978	0,908	0,864	0,928
	RIO	0,587	1560,54*	1	93,62	0,998	0,986	0,989	0,954	0,908
	FOGO	0,648	3793,38*	1	95,66	0,999	0,992	0,973	0,956	0,970
2010	TODAS	0,687	2516,82*	1	94,89	0,999	0,990	0,976	0,950	0,955
	ARAUC	0,612	1563,76*	1	94,52	0,999	0,990	0,978	0,943	0,949
	IMBUIA	0,577	1919,26*	1	87,22	0,998	0,977	0,905	0,860	0,923
	RIO	0,624	1412,12*	1	93,13	0,998	0,985	0,983	0,945	0,911
	FOGO	0,651	1937,40*	1	94,61	0,999	0,991	0,959	0,944	0,970

* Teste de Bartlett significativo a 5% de probabilidade

A Tabela 2 mostra os resultados da ACP considerando apenas densidade, frequência e dominância. Observa-se que não há multicolinearidade entre os dados, pois os determinantes da matriz de correlação são diferentes de zero. Os valores de

KMO são todos maiores que 0,5 conforme recomendado pela literatura (FIELD, 2009). Já o teste Bartlett se mostra altamente significativo ($p < 0,05$) para todas as áreas em todos os anos. Somente um componente foi extraído, as variâncias explicadas são bastante altas, e as cargas têm valores muito próximos.

TABELA 2- Análise de componentes principais para densidade, freqüência e dominância.

ANO	ÁREA	KMO	BARTLET	DETERM.	Nº COMP.	VAR. EXPL %	CARGAS		
							D	Do	F
1995	TODAS	0,686	675,94*	0,024	1	91,12	0,981	0,921	0,961
	ARAUC	0,514	216,73*	0,003	1	93,02	0,994	0,965	0,933
	IMBUIA	0,657	184,44*	0,033	1	84,72	0,958	0,830	0,967
	RIO	0,526	185,09*	0,006	1	93,31	0,996	0,951	0,950
	FOGO	0,740	205,97*	0,009	1	94,50	0,984	0,950	0,982
2000	TODAS	0,589	429,38*	0,004	1	94,63	0,993	0,949	0,975
	ARAUC	0,526	217,88*	0,003	1	92,98	0,993	0,968	0,931
	IMBUIA	0,638	163,41*	0,038	1	83,67	0,953	0,816	0,967
	RIO	0,506	193,78*	0,006	1	92,48	0,997	0,945	0,942
	FOGO	0,735	215,86*	0,009	1	94,01	0,982	0,944	0,982
2005	TODAS	0,572	442,19*	0,003	1	93,57	0,992	0,936	0,973
	ARAUC	0,563	216,65*	0,005	1	92,34	0,990	0,922	0,969
	IMBUIA	0,610	182,59*	0,028	1	82,35	0,954	0,789	0,969
	RIO	0,476	207,68*	0,006	1	91,08	0,997	0,923	0,942
	FOGO	0,718	537,06*	0,004	1	93,69	0,984	0,936	0,983
2010	TODAS	0,625	423,45*	0,005	1	92,63	0,988	0,925	0,973
	ARAUC	0,544	217,13*	0,004	1	91,98	0,990	0,917	0,969
	IMBUIA	0,613	192,48*	0,032	1	81,68	0,953	0,779	0,967
	RIO	0,484	188,96*	0,007	1	99,39	0,994	0,910	0,946
	FOGO	0,675	258,42*	0,006	1	92,11	0,976	0,918	0,984

* Teste de Bartlett significativo a 5% de probabilidade

Na Tabela 3 encontram-se os resultados da análise levando-se em consideração somente as variáveis: valor de importância e valor de cobertura. Pode-se verificar que os valores obtidos para KMO estão no limite recomendado de 0,50 (FIELD, 2009) para as cinco áreas em todos os anos analisados. Já o teste de Bartlett é significativo em todas as situações estudadas ($p < 0,05$).

Foi selecionado somente um componente principal para cada uma das áreas, com variância explicada bastante alta, acima de 99%, e as cargas têm valores iguais em todos os níveis testados.

TABELA 3 – Análise de fatorial para valor de importância e valor de cobertura

ANO	ÁREA	KMO	BARTLET	DETERM.	Nº	CARGAS		
						COMP.	EXPL%	VI
1995	TODAS	0,50	669,99*	0,025	1	99,36	0,997	0,997
	ARAUC	0,50	176,99*	0,010	1	99,75	0,999	0,999
	IMBUA	0,50	214,86*	0,019	1	99,51	0,998	0,998
	RIO	0,50	173,51*	0,009	1	99,78	0,999	0,999
	FOGO	0,50	220,57*	0,007	1	99,82	0,999	0,999
2000	TODAS	0,50	377,91*	0,007	1	99,82	0,999	0,999
	ARAUC	0,50	179,78*	0,009	1	93,76	0,999	0,999
	IMBUA	0,50	197,19*	0,020	1	99,49	0,997	0,997
	RIO	0,50	179,37*	0,009	1	99,76	0,999	0,999
	FOGO	0,50	232,77*	0,007	1	99,83	0,999	0,999
2005	TODAS	0,50	377,19*	0,008	1	99,79	0,999	0,999
	ARAUC	0,50	187,49*	0,010	1	99,75	0,999	0,999
	IMBUA	0,50	196,92*	0,022	1	99,45	0,997	0,997
	RIO	0,50	179,97*	0,012	1	99,71	0,999	0,999
	FOGO	0,50	481,88*	0,007	1	99,82	0,999	0,999
2010	TODAS	0,50	387,40*	0,009	1	99,78	0,999	0,999
	ARAUC	0,50	180,04*	0,010	1	99,74	0,999	0,999
	IMBUA	0,50	212,75*	0,023	1	99,42	0,997	0,997
	RIO	0,50	168,43*	0,013	1	99,68	0,998	0,998
	FOGO	0,50	254,28*	0,007	1	99,82	0,999	0,999

*Teste de Bartlett significativo a 5% de probabilidade

CONCLUSÃO

Ao realizar a análise fatorial e a análise de componentes principais com as estimativas dos parâmetros fitossociológicos pode-se observar que no ano de 1995 os dados referentes às áreas Araucária, Imbuia, Rio e Fogo, quando foram usadas as cinco estimativas fitossociológicas, apresentam multicolinearidades, e as correlações entre os parâmetros foram muito altas, acima de 0,8.

As análises realizadas nos anos de 1995, 2000, 2005 e 2010 demonstram haver correlação entre os dados, tendo em vista que os resultados do teste KMO estão acima de 0,5 que é recomendado pela literatura. Por meio do teste de Bartlett observa-se que são todos significativos a 5% de probabilidade, mostrando que as matrizes de correlação originais não são matrizes identidades. Em todos os estudos realizados observa-se a extração de um único componente principal com autovalor acima de um e com a porcentagem da variância explicada bastante alta.

No entanto quando se trabalha com cinco estimativas dos parâmetros observou-se que a carga de cada parâmetro foi bastante alta e com valores próximos. Considerando que a carga deve indicar qual parâmetro é mais importante, neste caso com uma pequena diferença destaca-se o valor de importância, porém as demais estimativas dos parâmetros não podem ser descartadas, pois apresentam valores significativos.

No estudo onde foram consideradas as variáveis densidades, frequência e dominância, também se observa uma proximidade dos valores nas cargas, e neste

caso há uma pequena diferença na variável densidade, mas isso não significa que podemos descartar as demais devido à significância das cargas.

Quando foram consideradas as variáveis, valor de importância e valor de cobertura observa-se que não há diferença entre as cargas das duas estimativas, e não se recomenda descartar nenhuma delas.

REFERÊNCIAS

CARVALHO, E.C.D. **Estrutura e estágios de sucessão ecológica da vegetação de caatinga em ambiente serrano no Cariri paraibano**. Dissertação de mestrado (Mestrado em Ciência e Tecnologia Ambiental), Universidade Estadual da Paraíba, 2009.

FIELD, A. **Descobrimo a estatística usando o SPSS** 2. ed., 688p. Porto Alegre: Artemed, 2009.

HAIR, J. F.; BLACK, W.C; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R.L. **Análise multivariada de dados**. 5. ed. Porto Alegre: Editora Bookman, 2005.

KURCHAIDT, S. M., **Uso de Métodos Estatísticos Multivariados em Fitossociologia Florestal**, 2014, 106f.(Tese de Doutorado), Doutorado em Engenharia Florestal, Universidade Federal do Paraná, Curitiba, 2014.

LAMPRECHT, H. **Silvicultura nos trópicos: ecossistemas florestais e respectivas espécies arbóreas - possibilidades e métodos de aproveitamento sustentado**. Rossdorf: TZ-Verl.-Ges., 1990. 343 p.

LIRIO, G. S. W., **Métodos Multivariados: Uma metodologia para avaliar a satisfação dos clientes da RBS-TV na região noroeste do RS**. Dissertação de Mestrado (Mestrado em Engenharia Florestal), Universidade Federal de Santa Maria. Santa Maria, 2004.

LORENZI, H. **Árvores Brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil**. 3. ed. Nova Odessa, SP: Plantarum, 2000.

MANLY, B. F. J. **Métodos estatísticos multivariados: uma introdução**. 3. ed. Porto Alegre: Bookman, 2008.

MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada**, Belo Horizonte: UFMG, 2005.

REIS, E. **Estatística multivariada aplicada**. Lisboa: Silabo, 1997.

SALOMÃO, R.P; SANTANA, A.C; COSTA NETO, S.V. Construção de índices de valor de importância de espécies para análise fitossociológica de floresta ombrófila através de análise multivariada. **Floresta**, v.42, p. 115-128, 2012.

SOUZA, D. C. Estudo fitossociológico no remanescente florestal nas margens dos rios Córrego Água dos Papagaios e do Campo, Campo Mourão Paraná. In: III

Simpósio Ambiental da Universidade Tecnológica Federal do Paraná, 2003, Campo Mourão.