

UNIVERSIDADE FEDERAL DO PARANÁ

SONIA MARIA KURCHAIT

**USO DE MÉTODOS ESTATÍSTICOS MULTIVARIADOS
EM FITOSSOCIOLOGIA FLORESTAL**

CURITIBA

2014

UNIVERSIDADE FEDERAL DO PARANÁ

SONIA MARIA KURCHAITD

**USO DE MÉTODOS ESTATÍSTICOS MULTIVARIADOS
EM FITOSSOCIOLOGIA FLORESTAL**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Florestal, Área de Concentração em Manejo Florestal, do Setor de Ciências Agrárias da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Ciências Florestais.

Orientador: Prof. Dr. Henrique Soares
Koehler

Co-Orientador: Prof. Dr. Luciano F.
Watzlawick

CURITIBA

2014

Ficha catalográfica elaborada por Denis Uezu – CRB 1720/PR
Biblioteca de Ciências Florestais e da Madeira - UFPR

Kurchaidt, Sonia Maria

Uso de métodos estatísticos multivariados em fitossociologia florestal / Sonia Maria Kurchaidt . – 2014

106 f. : il.

Orientador: Prof. Dr. Henrique Soares Koehler

Coorientador: Prof. Dr. Luciano F. Watzlawick

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Agrárias, Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 06/06/2014.

Área de concentração: Manejo Florestal

1. Comunidades vegetais. 2. Análise multivariada. 3. Mapeamento florestal. 4. Florestas - Paraná. 5. Teses. I. Koehler, Henrique Soares. II. Watzlawick, Luciano Farinha. III. Universidade Federal do Paraná, Setor de Ciências Agrárias. IV. Título.

CDD – 634.9

CDU – 634.0.182



Universidade Federal do Paraná
Setor de Ciências Agrárias - Centro de Ciências Florestais e da Madeira
Programa de Pós-Graduação em Engenharia Florestal

PARECER
Defesa n.º 1043

A banca examinadora, instituída pelo colegiado do Programa de Pós-Graduação em Engenharia Florestal, do Setor de Ciências Agrárias, da Universidade Federal do Paraná, após arguir o(a) doutorando(a) *Sonia Maria Kurchaidt* em relação ao seu trabalho de tese intitulado "**USO DE MÉTODOS ESTATÍSTICOS MULTIVARIADOS EM FITOSSOCIOLOGIA FLORESTAL**", é de parecer favorável à **APROVAÇÃO** do(a) acadêmico(a), habilitando-o(a) ao título de *Doutor(a)* em Engenharia Florestal, área de concentração em MANEJO FLORESTAL.

Dr. Solon Jonas Longhi
Universidade Federal de Santa Maria
Primeiro examinador

Dr. Luciano Farinha Watzlawick
Universidade Estadual do Centro Oeste - UNICENTRO
Segundo examinador

Dr. Mario Umberto Menon
Universidade Estadual do Centro Oeste - UNICENTRO
Terceiro examinador



Dr. Carlos Roberto Sanquetta
Universidade Federal do Paraná
Quarto examinador

Dr. Henrique Soares Koehler
Universidade Federal do Paraná
Orientador e presidente da banca examinadora

Curitiba, 06 de junho de 2014.

Antônio Carlos Batista

Coordenador do Curso de Pós-Graduação em Engenharia Florestal

AGRADECIMENTOS

Agradecimento é o sentimento de principal importância dentro da realização deste trabalho. Acredito que seria impossível a evolução do ser sem que houvesse, direta ou indiretamente a participação de outros. Essa interação influenciou significativamente a minha vida, permitindo-me crescer no sentido mais amplo da palavra. Por isso tentarei agradecer a todos os envolvidos na elaboração deste trabalho.

A Deus por estar ao meu lado me iluminando e me proporcionando a dádiva de poder desfrutar do meu maior presente: a Vida.

A minha mãe, exemplo de mulher, que sempre está ao meu lado com sua força, determinação e muito amor, incentivando-me a estudar e valorizando sempre as vitórias por mim alcançadas.

Ao meu querido pai que tão cedo nos deixou, mas continua sempre presente em meu coração, com certeza, onde está de alguma forma me ajudou a realizar este trabalho.

A minha filha Letícia que com seu carinho e amor, muito me incentivou para vencer mais esta etapa. Ao Cristhiano pelo amor e cuidado que sempre teve comigo. A Izadora, esta benção que Deus pôs em nossas vidas, pela certeza de que a vida sempre vale a pena.

Ao meu orientador Prof. Dr. Henrique Soares Koehler, agradeço pelos grandes ensinamentos, a dedicação e atenção oferecida a mim e principalmente pela amizade construída ao longo destes quatro anos.

Ao meu co-orientador Prof. Dr. Luciano Farinha Watzlawick, pelas sugestões e ensinamentos na realização deste trabalho.

Ao Prof. Dr. Solon Jonas Longhi, pela ajuda, parceria e incentivo na realização deste projeto.

Ao meu amigo e irmão Vitor Hugo Zanette grande companheiro de jornada, pelo incentivo e ajuda incansável. E também a seu filho João Vitor Zanette pela disponibilidade em me socorrer nas dificuldades encontradas sempre com atenção e carinho.

E a todos aqueles que de uma forma ou outra participaram da minha caminhada.

MUITO OBRIGADA!!!!

RESUMO

Em estudos de ecologia, inúmeras são as variáveis que afetam o estabelecimento, crescimento e permanência de espécies arbóreas. A análise multivariada de dados, sendo um conjunto de ferramentas de redução, classificação ou simplesmente utilizada como técnica exploratória torna-se útil e pertinente nestas situações. 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. Teve como objetivo principal aplicar técnicas multivariadas neste fragmento de Floresta, nos anos de 1995 a 2010, com o intuito de verificar o comportamento da floresta no período. Foi ajustada para cada uma das quatro parcelas, em cada um dos anos estudados, uma curva de acumulação espécie-área segundo um polinômio de 2º grau e testadas quanto a sua identidade. Os resultados revelaram que as equações diferem estatisticamente entre si e que não é possível utilizar uma única equação para representar a curva espécie/área no fragmento estudado. Foram verificadas se as intensidades amostrais empregadas (n=350) foram suficientes nos anos de 1995 e 2010, e todos os resultados obtidos mostraram que a intensidade efetuada foi maior do que a recomendada. Para os dados de ambos os anos foram obtidas as estimativas dos parâmetros fitossociológicos Densidade, Frequência, Dominância, Valor de Importância e Valor de Cobertura. Por meio da análise fatorial verificou-se que nenhum dos fatores, isolados ou em conjunto, destacou-se, uma vez que todos apresentaram altas cargas nos fatores. Foram identificadas as dez espécies que obtiveram maior valor de importância, nas quatro parcelas, com o objetivo de verificar se existem diferenças entre elas, por meio da utilização da MANOVA. Detectada a diferença entre as parcelas, a análise discriminante realizada selecionou três combinações lineares, para ambos os anos. Foram selecionadas as 20 espécies com maior Valor de Importância para cada uma das parcelas e foi realizada uma análise de agrupamentos, visando encontrar grupos similares, obtendo-se em cada um dos anos três grupos. Com os grupos formados foi realizada uma análise discriminante, que forneceu, para os anos em estudo, duas funções discriminantes, verificando-se que no ano de 1995, tanto na primeira como na segunda função discriminante, o grupo dois discrimina os demais, enquanto que para o ano de 2010, em ambas as funções o grupo um é que discrimina os demais. O grupo dois no ano de 1995 é formado predominantemente por subunidades pertencentes à parcela Rio, e para o ano de 2010, no grupo um predominam subunidades da parcela Araucária. A análise discriminante apontou as espécies *Araucaria angustifolia*, *Nectandra grandiflora*, *Ilex paraguariensis* e *Casearia decandra* como espécies predominantes na área estudada. Pode-se afirmar, com base nos resultados obtidos, que as parcelas Araucária, Rio, Fogo e Imbuia são pertencentes a uma única tipologia florestal, caracterizada como Floresta Ombrófila Mista.

Palavras chave: Floresta Ombrófila Mista, componentes principais, Manova, análise de agrupamentos; análise discriminante.

ABSTRACT

Ecological studies usually involve several variables which affect establishment, growth and permanence of three species. Multivariate analysis, used as a tool to reduce, classify or explore the data is very useful in these situations. The study was conducted using four permanent plots collected annually from the forest inventory of the "Long Term Ecological Program" (PELD), in São João do Triunfo Experimental Station of the Federal University of Paraná. The main objective was to apply multivariate techniques in a fragment of Araucaria forest, in 1995 and 2010, in order to verify the behavior and dynamics of the forest in the period. Initially, a second degree polynomial function was fitted to the species-area accumulation data, to each of the four plots, in each of the years studied. The analysis showed that the equations were statistically different and could not be replaced by a single equation to represent the type/area curve in the whole fragment. It was found that the sample intensities employed ($n = 350$) were sufficient in both years, with the results showing a sampling size higher than the recommended. For the data from both years, it was obtained estimates of the phytosociological parameters Density, Frequency, Dominance, Importance Value and Coverage Value. Factor Analysis showed that none of the factors, alone or combined, stood out, since all presented high loading factors. The ten species with the greatest amount of importance, in all four plots, were identified with the objective of determining if there were differences among the plots, through the use of multivariate analysis of variance. Once differences among the plots were detected, a discriminant analysis selected three linear combinations of species, for both years. Three discrimination functions were found. For both years, the 20 species with the highest importance value were selected from each of the plots and a cluster analysis was performed, aiming to find similar groups among them. Three groups were founded, in each of the years. Using the groups found by the discriminant analysis, that provided two discriminant functions, for the both years under study, it was found that in 1995, both the first and the second groups of subunits is discriminating the others, while for the year 2010, in both functions the group one is discriminating against others. Group two in 1995, is predominantly formed by subunits belonging to the Rio plot, and for the year 2010, in group one predominates the subunits from the Araucaria plot. The discriminant analysis showed the species *Araucaria angustifolia*, *Nectandra grandiflora*, *Ilex paraguariensis* and *Casearia decandra* as the predominant species in the studied area. It can be stated, based on the results obtained, that the Araucaria, Rio, Fogo and Imbuia plots belongs to a single forest type, characterized as Araucaria Forest .

Key words: Araucaria Forest , principal components, Manova, cluster analysis, discriminant analysis.

SUMÁRIO

	LISTA DE FIGURAS	9
	LISTA DE TABELAS	10
1	INTRODUÇÃO	13
1.1	OBJETIVO GERAL.....	15
1.2	OBJETIVOS ESPECÍFICOS	15
2	REVISÃO BIBLIOGRÁFICA	17
2.1	A VEGETAÇÃO DO ESTADO DO PARANÁ.....	17
2.2	FLORESTA OMBRÓFILA MISTA (FLORESTA COM ARAUCÁRIA)	18
2.3	FITOSSOCIOLOGIA.....	19
2.4	IDENTIDADE DE MODELOS	23
2.5	SUFICIÊNCIA AMOSTRAL	26
2.6	CURVA ESPÉCIE-ÁREA E CURVA DO COLETOR.....	29
2.7	ANÁLISE FATORIAL	33
2.8	ANÁLISE DOS COMPONENTES PRINCIPAIS	37
2.9	ANÁLISE DE VARIÂNCIA MULTIVARIADA - MANOVA	39
2.10	ANÁLISE DE AGRUPAMENTOS - CLUSTER	43
2.10.1	Métodos Hierárquicos.....	45
2.10.2	Métodos não-hierárquicos ou por particionamento.....	48
2.11	ANÁLISE DISCRIMINANTE	49
2.12	ANÁLISE MULTIVARIADA NA ÁREA FLORESTAL.....	51
3	MATERIAL E MÉTODOS	53
3.1	CARACTERIZAÇÃO DA ÁREA.....	53
3.2	ANÁLISE DE DADOS.....	54
3.2.1	Amostragem	55
3.3	IDENTIDADE DE MODELOS	56
3.4	SUFICIÊNCIA AMOSTRAL	57
3.5	ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS	58
3.6	ANÁLISE DE VARIÂNCIA MULTIVARIADA - MANOVA	58
3.7	ANÁLISE DE AGRUPAMENTOS - CLUSTER	59
3.8	ANÁLISE DISCRIMINANTE	60
4	RESULTADOS E DISCUSSÃO	61
4.1	IDENTIDADE DE MODELOS	61

4.2	SUFICIÊNCIA AMOSTRAL	70
4.3	ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS	74
4.4	ANÁLISE DE VARIÂNCIA MULTIVARIADA – MANOVA	78
4.5	ANÁLISE DE AGRUPAMENTOS - CLUSTER	85
4.6	ANÁLISE DISCRIMINANTE	89
5	CONCLUSÕES	96
6	REFERÊNCIAS	98

LISTA DE FIGURAS

FIGURA 01	PERFIL ESQUEMÁTICO DESTACANDO A ESTRUTURA DE UM SEGMENTO DE FLORESTA OMBRÓFILA MISTA NO MUNICÍPIO DE IRATI - PR, COM PREDOMINÂNCIA DE <i>Araucaria</i> , <i>Ocotea</i> , <i>Cedrela</i> , <i>Casearia</i> , <i>Sloanea</i> , <i>Podocarpus</i> , <i>Campomanesia</i> , <i>Ilex</i> e <i>Capsicodendron</i>	19
FIGURA 02	FLUXOGRAMA DOS PROCEDIMENTOS UTILIZADOS PARA A OBTENÇÃO DOS RESULTADOS.....	54
FIGURA 03	CROQUI DE LOCALIZAÇÃO DA ESTAÇÃO EXPERIMENTAL DE SÃO JOÃO DO TRIUNFO-PR.....	55
FIGURA 04	CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 1995.....	62
FIGURA 05	CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 2000.....	64
FIGURA 06	CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 2005.....	66
FIGURA 07	CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 2010.....	69
FIGURA 08	FUNÇÃO DISCRIMINANTE E CENTRÓIDES DAS PARCELAS PARA O ANO DE 1995.....	84
FIGURA 09	FUNÇÃO DISCRIMINANTE E CENTRÓIDES DAS PARCELAS PARA O ANO DE 2010.....	85
FIGURA 10	DENDROGRAMA RESULTANTE DA ANÁLISE DE AGRUPAMENTOS RELATIVO AO ANO DE 1995.....	87
FIGURA 11	DENDROGRAMA RESULTANTE DA ANÁLISE DE AGRUPAMENTOS RELATIVO AO ANO DE 2010.....	88
FIGURA 12	REPRESENTAÇÃO GRÁFICA DOS GRUPOS E SEUS CENTRÓIDES NO ANO DE 1995..	93
FIGURA 13	REPRESENTAÇÃO GRÁFICA DOS GRUPOS E SEUS CENTRÓIDES NO ANO DE 2010.	93

LISTA DE TABELAS

TABELA 1	RESUMO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE HIPÓTESE $H_0 : \beta_1 = \beta_2 = \dots = \beta_H$	26
TABELA 2	MANOVA PARA TESTAR A HIPÓTESE DE IGUALDADE DO VETOR DE EFEITO DOS TRATAMENTOS	41
TABELA 3	CARACTERIZAÇÃO DAS PARCELAS ESTUDADAS, NOS DOIS ANOS CONSIDERADOS	56
TABELA 4	EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE-ÁREA PARA CADA PARCELA, PARA O ANO DE 1995	61
TABELA 5	QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 1995	63
TABELA 6	EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE-ÁREA PARA CADA PARCELA, PARA O ANO DE 2000	63
TABELA 7	QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 2000	65
TABELA 8	EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE-ÁREA PARA CADA PARCELA, PARA O ANO DE 2005	65
TABELA 9	QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 2005	67
TABELA 10	EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE-ÁREA PARA CADA PARCELA, PARA O ANO DE 2010	67
TABELA 11	QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 2010	68
TABELA 12	FAMÍLIAS, NOME CIENTÍFICO E NOME COMUM DAS ESPÉCIES ENCONTRADAS NAS QUATRO PARCELAS TOMADAS NA ÁREA DE SÃO JOÃO DO TRIUNFO-PR.	70
TABELA 13	RESULTADOS DAS ESTATÍSTICAS E INTENSIDADES AMOSTRAIS ENCONTRADAS PARA A VARIÁVEL NÚMERO DE ESPÉCIES POR SUBUNIDADE, PARA AS MEDIÇÕES EFETUADAS NOS ANOS DE 1995 E 2010	72

TABELA 14	ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS PARA AS VARIÁVEIS D, F, Do, VI, VC, NOS ANOS DE 1995 E 2010. ..	76
TABELA 15	ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS PARA AS VARIÁVEIS D, F, Do, NOS ANOS DE 1995 E 2010.....	77
TABELA 16	ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS PARA AS VARIÁVEIS VI , VC, NOS ANOS DE 1995 E 2010.....	77
TABELA 17	ESPÉCIES SELECIONADAS PELO VALOR DE IMPORTÂNCIA, PARA OS ANOS DE 1995 E 2010.....	78
TABELA 18	RESULTADOS TESTE DE BOX E TESTE F PARA 1995 E 2010..	79
TABELA 19	RESULTADOS DOS TESTES PARA O INTERCEPTO E PARA OS GRUPOS NOS ANOS 1995 E 2010.....	79
TABELA 20	RESULTADOS DA ANÁLISE DE VARIÂNCIA MULTIVARIADA RELATIVA AOS DADOS DE 1995	80
TABELA 21	RESULTADOS DA ANÁLISE DE VARIÂNCIA MULTIVARIADA RELATIVA AOS DADOS DE 2010	81
TABELA 22	RESULTADOS DOS AUTOVALORES OBTIDO E PERCENTUAL DA VARIÂNCIA EXPLICADA, PARA OS ANOS DE 1995 E 2010	81
TABELA 23	RESULTADOS DO TESTE LAMBDA DE WILKS PARA OS ANOS DE 1995 E 2010	81
TABELA 24	COEFICIENTES PADRONIZADOS DA FUNÇÃO DISCRIMINANTE PARA O ANO DE 1995	82
TABELA 25	COEFICIENTES PADRONIZADOS DA FUNÇÃO DISCRIMINANTE PARA O ANO DE 2010	83
TABELA 26	RESULTADOS DOS CENTRÓIDES DAS FUNÇÕES DISCRIMINANTES PARA OS ANOS 1995 E 2010.....	84
TABELA 27	RESULTADOS DOS TESTES DE LAMBDA DE WILKS E F NA SELEÇÃO DE VARIÁVEIS DISCRIMINANTES, POR ORDEM DE ENTRADA, PARA O ANO DE 1995.....	89
TABELA 28	RESULTADOS DOS TESTES DE LAMBDA DE WILKS E F NA SELEÇÃO DE VARIÁVEIS DISCRIMINANTES, POR ORDEM DE ENTRADA, PARA O ANO DE 2010.....	90
TABELA 29	RESULTADOS DAS FUNÇÕES DISCRIMINANTES LINEARES OBTIDAS PARA AS ESPÉCIES SELECIONADAS NO ANO DE 1995.....	90

TABELA 30	RESULTADOS DAS FUNÇÕES DISCRIMINANTES LINEARES OBTIDAS PARA AS ESPÉCIES SELECIONADAS NO ANO DE 2010.....	91
TABELA 31	RESULTADOS DOS TESTES ESTATÍSTICOS DA SELEÇÃO DAS FUNÇÕES DISCRIMINANTES NO ANO DE 1995	91
TABELA 32	RESULTADOS DOS TESTES ESTATÍSTICOS DA SELEÇÃO DAS FUNÇÕES DISCRIMINANTES NO ANO DE 2010	92
TABELA 33	RESULTADOS DOS CENTRÓIDES PARA AS FUNÇÕES DISCRIMINANTES PARA OS TRÊS GRUPOS DE SUBUNIDADES, NOS ANOS DE 1995 E 2010.....	92
TABELA 34	RESULTADOS DA CLASSIFICAÇÃO, EM NÚMERO E PORCENTAGEM, DOS GRUPOS DE SUBUNIDADES, NO ANO DE 1995.....	94
TABELA 35	RESULTADOS DA CLASSIFICAÇÃO, EM NÚMERO E PORCENTAGEM, DOS GRUPOS DE SUBUNIDADES, NO ANO DE 2010.....	94

1 INTRODUÇÃO

O bioma Mata Atlântica ocupa 15% do território brasileiro e apresenta uma variedade de formações, dentre elas, a Floresta Ombrófila Mista (FOM). A designação desta tipologia advém da classificação fitogeográfica empregada pelo IBGE (2012), que denomina como Ombrófila as formações características de ambientes úmidos (SANQUETTA; MATTEI, 2006). O termo “Mista” refere-se à origem geológica da formação (IBGE, 2012), na qual contempla-se a coexistência de representantes das floras tropical (afro-brasileira) e temperada (austro-brasileira), em marcada relevância fisionômica de elementos Coniferales e Laurales (LEITE, 1994), com predomínio em regiões de clima tipicamente pluvial, principalmente em latitudes maiores a 23° S, altitudes superiores de 500 m, e em situações afastadas das influências marítimas (LEITE; KLEIN, 1990). As florestas de Araucária, de ocorrência predominante na região Sul do país, são de alta relevância econômica, oferecem diversos recursos madeireiros e outros produtos da floresta, além de ser considerada símbolo para a região e forma de vínculo afetivo para o seu povo (PIRES, *et al.*, 2012).

O ecossistema como um todo pode ser mantido ou alterado de acordo com a forma e os métodos de manejo adotados. Suas características podem ser modificadas até o ponto em que se torne impossível o reparo nos danos causados ao meio e à dinâmica ecológica para a manutenção da estrutura e dos componentes envolvidos na floresta. Dessa forma, o estudo da vegetação é importante não só para a Taxonomia Vegetal e Fitogeografia, mas também no âmbito de pesquisa aplicada e de gestão, principalmente, como subsídios à silvicultura, manejo de bacias hidrográficas, manejo de fauna, preservação de táxons, conservação do ambiente e interpretação do potencial da terra para uso agropecuário (LONGHI, 1997).

A demanda crescente destas informações implica no conhecimento de técnicas avançadas de organização e interpretação de dados, área onde a estatística multivariada apresenta novas técnicas e procedimentos advindos da evolução da informática no final do século passado.

Os métodos estatísticos disponíveis, para analisar variáveis, estão dispostos em dois grupos: um trata da estatística que analisa as variáveis de maneira isolada – a estatística univariada, e outro que trata as variáveis de forma conjunta – a estatística multivariada (VICINI, 2005).

Dentro da estatística univariada destaca-se a análise de regressão, onde com frequência deseja-se saber se um conjunto de g equações ajustadas são idênticas, ou seja, se o fenômeno em estudo pode ser representado por uma única equação. Vários autores apresentaram métodos para testar hipóteses relativas à identidade de modelos lineares, como por exemplo, Graybill (1976), Steel e Torrie (1980), Neter *et al.* (1996), Regazzi (1993, 1996), entre outros. Análises referentes a estas situações são comuns e de fundamental importância nas áreas de experimentação agropecuária, econometria e biometria florestal (MAGALHÃES e MAGINI, 2004).

Segundo Schilling e Batista (2008) um dos principais problemas enfrentados em estudos direcionados ao conhecimento da diversidade dos ecossistemas florestais nativos é a dificuldade de obtenção das informações necessárias à sua caracterização de forma confiável.

Uma representação fiel da diversidade de espécies de uma comunidade só seria atingida por meio de um levantamento com enumeração completa dos indivíduos, chamado censo florestal, que se torna inviável do ponto de vista econômico para a grande maioria dos estudos. Desta forma, o emprego de técnicas de amostragem, outro processo classificado como univariado, mostra-se como a forma mais exequível para realização de estudos de descrição e análise da vegetação.

A análise estatística de mensurações múltiplas efetuadas sobre uma amostra fornece um melhor entendimento na razão direta do número de variáveis utilizadas e permite considerar simultaneamente a variabilidade existente entre elas. Nesse contexto, a Análise Multivariada dispõe de uma diversidade de técnicas que favorecem o entendimento de muitos fenômenos (ALVES, 2007).

A análise multivariada de dados, sendo um conjunto de ferramentas de redução, classificação ou simplesmente utilizada como técnica exploratória torna-se útil nos estudos da ecologia, pois inúmeras são as variáveis que afetam o estabelecimento, crescimento e permanência de espécies arbóreas em diferentes ambientes (RODE *et al.*, 2011).

Verifica-se assim, a necessidade da realização de estudos que retratem a realidade da floresta, no que tange a florística, sua estrutura e dinâmica 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 ao máximo retratar a sua diversidade. É necessário também avaliar a adequação dos métodos e procedimentos amostrais capazes de contemplar a maior parte da variabilidade da comunidade estudada e, desta forma, realizar inferências mais exatas sobre a diversidade do local.

1.1 OBJETIVO GERAL

O objetivo deste trabalho foi aplicar técnicas de análise multivariada em dados biométricos de espécies arbóreas num fragmento de Floresta Ombrófila Mista, no período de 1995 a 2010, com o intuito de verificar o comportamento das variáveis fitossociológicas da floresta no período.

1.2 OBJETIVOS ESPECÍFICOS

- Ajustar modelos de regressão polinomial de grau dois em curvas espécie/área ou curva do coletor, visando avaliar a identidade destes modelos;
- Apresentar método para testar se as equações são idênticas e podem ser representadas por uma única equação de regressão;
- Verificar a intensidade amostral em função da variância do número de indivíduos por unidade amostral, para diferentes anos;
- Analisar a suficiência amostral por meio da curva de espécie/área ou curva do coletor com o objetivo de verificar o tamanho ideal da amostra;
- Caracterizar a composição florística e a estrutura arbórea da formação vegetal no fragmento estudado;

- Verificar se há diferenças entre grupos para as espécies selecionadas na análise e quais espécies contribuem para a diferença;
- Determinar e caracterizar diferentes grupos florísticos na área estudada.

2 REVISÃO BIBLIOGRÁFICA

2.1 A VEGETAÇÃO DO ESTADO DO PARANÁ

Levando em consideração a afirmação de Maack (1968), de que vegetação é “a expressão do clima em relação à altitude e latitude”, a vegetação original atual do Estado do Paraná é resultante do condicionamento ocorrido após o último máximo glacial (entre 26.000 e 17.000 anos atrás). Nesta época, a teoria de Damuth e Fairbridge (AB’SÁBER, 1977), estabelece um clima frio e seco para o Sul do Brasil. Este clima obrigou as comunidades vegetais que não suportavam estas condições a se dispersarem rumo ao norte, à procura de regiões menos frias (Brown; Lomolino, 2006 e Lavina; Fauth, 2010), dando espaço para as tipologias adaptadas aos climas frios (estepe e cerrado) ocuparem grande parte do território paranaense.

Com o reaquecimento do planeta, as florestas recobriram novamente grande parte do Estado, porém recombinaando-se em uma nova estrutura florística que é a que conhecemos hoje e chamamos de cobertura original do estado (BROWN; LOMOLINO, 2006).

O resultado desta readaptação foi descrito por Maack (1981) como sendo, originalmente, 83% da superfície paranaense cobertos por florestas e 17% ocupados por formações não florestais, como campos e cerrados, completados por vegetações pioneiras de influência marinha (restingas), fluviomarina (mangues) e flúvio-lacustre (várzeas) e também pela vegetação herbácea nas altitudes. Segundo SOCIEDADE DE PESQUISA EM VIDA SELVAGEM E EDUCAÇÃO AMBIENTAL - SPVS (1996), esta diversidade de ambientes é resultante de fatores ambientais, tais como tipo de solo, clima, efeitos de marés, rios, planícies e serras.

De acordo com o Sistema Brasileiro de Classificação da Vegetação (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE, 2012), baseado no sistema de classificação fisionômico-ecológico desenvolvido pelo projeto RADAMBRASIL, as formações vegetais encontradas no Paraná foram denominadas de Estepe (Campo Gerais), Savana (Cerrados), Floresta Ombrófila Mista (Floresta com Araucária), Floresta Ombrófila Densa (Floresta Atlântica) e a Floresta Estacional Semidecidual (Floresta Pluvial Tropical-subtropical) (IBGE, 2012).

2.2 FLORESTA OMBRÓFILA MISTA (FLORESTA COM ARAUCÁRIA)

Ao ultrapassar a serra, na direção oeste, no planalto do Estado (altitudes variando entre 500 e 1.200 m), situa-se a região de ocorrência da floresta com araucária. Nessa região, as chuvas também são bem distribuídas ao longo do ano, mas com médias de temperatura mais baixas e ocorrência regular de geadas, o que permite uma série de modificações tanto na composição das florestas como no seu funcionamento. No Estado do Paraná, distribui-se no primeiro, segundo e terceiro planaltos, ocorrendo também na região dos campos na forma de capões ou no vale dos rios (FERRETTI *et al.*, 2006).

A Floresta Ombrófila Mista é caracteristicamente dominada pela espécie *Araucaria angustifolia* (Bertol.) Kuntze, que forma um estrato dominante e contínuo acima de 30 metros de altura, podendo ocorrer indivíduos emergentes acima de 40 metros (RODERJAN *et al.*, 2002). A presença da araucária está associada às baixas temperaturas, sendo um fator fundamental para sua reprodução. Outros autores equiparam a ocorrência da família Araucariaceae com a família Lauraceae, baseados em Bolós *et al.* (1991), que definiu esta formação como “Floresta de Araucária e *Ocotea puberula*”.

Na época do descobrimento do Brasil a Floresta com Araucária se estendia numa faixa contínua desde o sul do estado de São Paulo até o norte do Rio Grande do Sul, chegando até a Argentina, além de manchas esparsas nos estados de Minas Gerais, Rio de Janeiro e Espírito Santo (IBGE, 1992).

A Floresta Ombrófila Mista é uma unidade fitoecológica onde se contempla a coexistência de representantes das floras tropical (afro-brasileira) e temperada (austro-brasileira), em marcada relevância fisionômica de elementos Coniferales e Laurales, com sua flora arbórea superior a 350 espécies com endemismo em torno de 40% (LEITE, 1994; IBGE, 1992) e originalmente com predomínio de araucária no dossel (RODERJAN; KUNIYOSHI, 1988; LEITE; KLEIN, 1990; IBGE, 1992, 2012).

A composição florística deste tipo de vegetação, dominada por gêneros primitivos como *Drimys* e *Araucaria* (australásicos) e *Podocarpus* (afro-asiático), sugere, em face da altitude e da latitude do Planalto Meridional, uma ocupação recente a partir de Refúgios Altomontanos (IBGE, 2012). A Figura 1 mostra um perfil esquemático de um segmento de Floresta Ombrófila Mista.

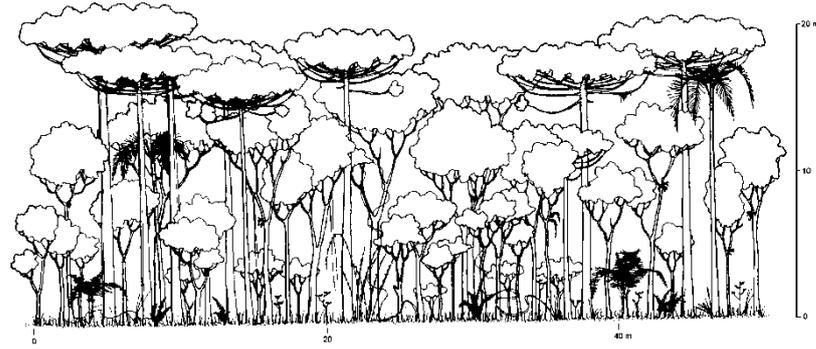


FIGURA 1 - PERFIL ESQUEMÁTICO DESTACANDO A ESTRUTURA DE UM SEGMENTO DE FLORESTA OMBRÓFILA MISTA NO MUNICÍPIO DE IRATI - PR, COM PREDOMINÂNCIA DE *Araucária*, *Ocotea*, *Cedrela*, *Casearia*, *Sloanea*, *Podocarpus*, *Campomanesia*, *Ilex* e *Capsicodendron*.
FONTE: Roderjan *et al.* (2002).

Segundo MAACK (1981), no Estado do Paraná o desenvolvimento das florestas com araucária constitui uma parte especial da floresta pluvial-subtropical, cujo desenvolvimento é intimamente relacionado à altitude. O limite inferior de crescimento é de 500 m. Em altitudes inferiores a araucária somente ocorre nas linhas de escoamento do ar frio, associada a *Syagrus romanzoffianum* (jerivá).

De acordo com IBGE (2012), são identificadas quatro formações da Floresta Ombrófila Mista: aluvial (em terraços antigos associados à rede hidrográfica), submontana (constituindo disjunções em altitudes inferiores a 400 m), Montana (situada aproximadamente entre 400 e 1000 m de altitude), e alto-montana (compreendendo as altitudes superiores a 1000 m). Porém, no Estado do Paraná, não há ocorrência de formações submontanas e, segundo Roderjan *et al.* (2002), as formações altomontanas ocupam uma superfície pouco expressiva, sendo fisionomicamente similares às formações montanas, mas com menor diversidade, devido ao rigor climático.

2.3 FITOSSOCIOLOGIA

Fitossociologia é o estudo das comunidades vegetais, incluindo composição de espécies, organização, interdependência, desenvolvimento, distribuição geográfica e classificação (Glossário de Ecologia, 1987). Galvão (1994) relata que fitossociologia é um ramo da geobotânica que se ocupa do estudo quantitativo da

composição florística, estrutura, funcionamento, dinâmica, distribuição e relações ambientais das comunidades vegetais.

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 sejam comparáveis entre si, de preferência com métodos estatísticos. 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 (SOUZA, 2003). No Brasil, os primeiros estudos desse ramo foram iniciados na década de 1940, relacionados principalmente a Floresta Ombrófila Densa (SCHORN, 2013).

Hanson e Churchill (1961) citam que as características de uma comunidade podem ser classificadas em dois principais grupos: analítico e sintético. O grupo analítico inclui características qualitativas, as quais são usualmente descritivas devido à grande dificuldade em mensurá-las, e quantitativas, as quais podem ser realmente medidas. As qualitativas implicam na composição florística, estratificação, periodicidade, vitalidade, forma de vida, sociabilidade e associação das espécies, enquanto que as quantitativas medem densidade populacional, cobertura, altura das plantas, peso das plantas, volume ocupado pelas plantas e frequência. Segundo os mesmos autores, os principais parâmetros a serem estimados a partir dos dados são estrutura horizontal (como densidade, frequência e dominância), estrutura diamétrica, estrutura vertical (posição sociológica, regeneração natural), valor de importância, variáveis fitossociológicas 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.

Densidade absoluta: refere-se ao número de indivíduos de cada espécie por unidade de área considerada. Densidade relativa é a proporção entre o número de indivíduos de uma espécie, em relação ao total de indivíduos amostrados (DAUBENMIRE, 1968).

$$DA = n/ha$$

$$DR = [(n/ha)/(N/ha)] \times 100$$

onde:

DA = densidade absoluta;

DR = densidade relativa;

n/ha = número de árvores de cada espécie por unidade de área;

N/ha = número total de árvores por unidade de área.

Frequência: é um conceito estatístico relacionado com a uniformidade da distribuição das espécies dentro de uma comunidade, em um número de áreas de igual tamanho (SOUZA, 1973). Frequência absoluta é determinada como a proporção entre o número de unidades amostrais onde a espécie ocorre e o número total de unidades amostrais. É a porcentagem de parcelas em que ocorre uma determinada espécie.

Frequência relativa é a proporção, expressa em porcentagem, entre a frequência absoluta de cada espécie e a frequência absoluta total por unidade de área (MULLER-DOMBOIS; ELLENBERG, 1974).

$$FR = [FA/\sum FA]* 100$$

onde:

FA= frequência absoluta;

FR = frequência relativa.

Dominância: Cain; Castro (1956) definem dominância como o termo usado para caracterizar a área coberta, espaço ocupado ou o grau de controle de uma comunidade por uma ou mais espécies. É a prevalência ou predominância relativa de indivíduos de uma espécie que resulta de seu número e volume ocupado.

Dominância é o grau de cobertura das espécies como expressão do espaço por elas requerido. Definem-se pelo somatório de todas as projeções horizontais dos vegetais sobre o solo. No caso de árvores, a dominância de uma espécie exprime-se, pela soma, em m², das projeções das copas de todos os seus respectivos indivíduos. Devido ser extremamente difícil e dispendioso determinar as projeções das copas, calculam-se em geral as áreas transversais individuais dos diâmetros das árvores por existir correlação entre o diâmetro de copa e o DAP (LAMPRECHT, 1990).

Dominância absoluta de uma espécie consiste na soma da área basal de todos os indivíduos da espécie, presentes na amostra. Dominância relativa é a

percentagem entre área basal total da espécie e a área basal total por unidade de área (MULLER-DOMBOIS; ELLENBERG, 1974).

$$\text{DoA} = g/\text{há}$$

$$\text{DoR} = [(g/\text{ha})/(G/\text{ha})] \times 100$$

DoA = dominância absoluta;

DoR = dominância relativa;

onde:

g/ha = área basal de cada espécie por unidade de área;

G/ha = área basal total por unidade de área.

Valor de Importância: A correlação entre os resultados de análises é feita por meio do valor de importância (LAMPRECHT, 1990). Obtém-se este índice somando para cada espécie os valores relativos de densidade, frequência e dominância, obtendo um valor máximo de 300,00%. O valor de importância pode ser convertido em percentagem de importância ao ser dividido por três.

Lamprecht (1990) menciona que por este meio, torna-se possível a comparação entre os "pesos ecológicos" das espécies dentro do respectivo tipo florestal. Valores mais ou menos iguais obtidos para os valores de importância das espécies mais características podem ser uma indicação da igualdade ou, pelo menos, semelhança das comunidades quanto à composição, estrutura, sítio e dinâmica.

$$\text{VI} = \text{DR} + \text{FR} + \text{DoR}$$

$$\text{PVI} = (\text{DR} + \text{FR} + \text{DoR})/3$$

onde

VI = valor de importância;

PVI = percentagem do valor de importância;

Valor de Cobertura: Cobertura é definida como a projeção vertical da copa ou das raízes de uma espécie sobre o solo, expressa sob a forma de porcentagem em relação a uma área definida (MULLER-DOMBOIS; ELLENBERG, 1974).

A importância que uma espécie adquire na floresta é caracterizada pelo número de árvores e suas dimensões (densidade e dominância), que determinam o espaço dentro da biocenose, não importando muito se as árvores aparecem isoladas ou em grupos (frequência) (LONGHI, 1997).

O uso do valor de cobertura, que agrega os totais, por espécie, obtidos para densidade e dominância relativas, tem um valor máximo de 200,00%. A porcentagem de cobertura consiste na média entre estes dois valores (MULLER-DOMBOIS; ELLENBERG, 1974).

$$VC = DR + DoR$$

$$PVC = (DR + DoR)/2$$

onde:

VC = valor de cobertura;

PVC = porcentagem do valor de cobertura.

2.4 IDENTIDADE DE MODELOS

Segundo REGAZZI (1993), a identidade de modelos trata-se de um teste geral que verifica a igualdade de um conjunto de equações de regressão, sob a hipótese de nulidade:

Ho: $\beta_1 = \beta_2 = \dots = \beta_H$ (as H equações são idênticas) e os modelos reduzem-se a forma:

$$Y_{hi} = bP_{1hi} + cP_{2hi} + e_i$$

onde Y_{hi} , P_{khi} e e_i têm as mesmas especificações dos modelo descrito inicialmente, b e c são os coeficientes comuns. Utilizando a notação matricial os modelos reduzidos podem ser escritos como:

$$\underline{Y} = \underline{Z} \underline{\theta} + \underline{\varepsilon}$$

em que:

\underline{Y} é o vetor dos valores observados da variável dependente, de dimensões $N \times 1$;

$$Z = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_H \end{bmatrix} \quad \text{onde } X_h \text{ com } H = 1, 2, \dots, H, \text{ de dimensões } N \times P;$$

$$\underline{\hat{\theta}} = \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{é o vetor dos coeficientes comuns, de dimensões } P \times 1;$$

ε é o vetor de erros aleatórios, de dimensões $N \times 1$.

O sistema de equações normais relativo ao modelo reduzido, obtido pelo método dos mínimos quadrados, é:

$$Z'Z \theta = Z' \underline{Y},$$

ou seja;

$$[\sum_{h=1}^H X'_h X_h] \cdot \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = [\sum_{j=1}^H X'_j Y_j] \quad (1)$$

Sendo $Z'Z$, matriz de dimensões $p \times p$ e não singular, o estimador do vetor de coeficientes comuns tem a seguinte expressão:

$$\underline{\hat{\theta}} = (Z'Z)^{-1} Z' \underline{Y} \quad (2)$$

onde $(Z'Z)$ é a matriz composta pela soma das matrizes $(X'_h X_h)$ de cada modelo, bem como a matriz $(Z' \underline{Y})$, o estimador do vetor dos coeficientes comuns, pode-se escrever a expressão como:

$$\underline{\hat{\theta}} = [\sum_{h=1}^H X'_h X_h]^{-1} \sum_{j=1}^H X'_j Y_j \quad (3)$$

A soma de quadrados da regressão, relativa ao modelo reduzido (1), é dada por

$$SQPar(r_1) = \underline{\hat{\theta}}' Z' \underline{Y} = (\sum_{j=1}^H X'_j Y_j) [\sum_{h=1}^H X'_h X_h]^{-1} [\sum_{t=1}^H X'_t Y_t] \quad (4)$$

com p graus de liberdade.

A soma de quadrados total não corrigida é dada por:

$$SQTot(r_1) = \underline{Y}'\underline{Y} = \left[\sum_{h=1}^H \underline{Y}'_h \underline{Y}_h \right] \quad (5)$$

com N graus de liberdade.

A soma de quadrado de resíduo relativa ao modelo reduzido é obtida pela diferença entre (5) e (4), ou seja,

$$SQRes(r_1) = \underline{Y}' [I - Z(Z'Z)^{-1} Z'] \underline{Y} \quad (6)$$

com (N – p) graus de liberdade.

O teste para verificar se as equações são idênticas é baseado na diferença entre a soma de quadrados de regressão do modelo completo e do modelo reduzido, ou seja, na redução que H_0 provoca na soma de quadrados de regressão do modelo completo. Assim a redução devida à hipótese H_0 , denotada por Redução (H_0), é obtida pela diferença entre

$$SQPar(c) = \underline{\hat{\beta}}' X' \underline{Y} \quad \text{e} \quad SQPar(r_1) = \underline{\hat{\theta}}' Z' \underline{Y}$$

ou seja:

$$\text{Redução } (H_0) = SQPar(c) - SQPar(r_1) \quad (7)$$

com (H-1) graus de liberdade.

Segundo GRAYBILL (1976), para testar a hipótese de nulidade:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_H,$$

a estatística usada é a F central, com (H-1)p e (N-Hp) graus de liberdade e o F_o é dado por:

$$F_o = \frac{[SQPar(c) - SQPar(r_1)] / (H-1)p}{SQRes(c) / (N-Hp)} \quad (8)$$

A Tabela 1 sumariza os procedimentos utilizados.

TABELA 1 - RESUMO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE HIPÓTESE $H_0 : \beta_1 = \beta_2 = \dots = \beta_H$

Fonte de variação	G.L	S.Q	Q. M	F_o
Parâmetros ($\underline{\beta}$)	(Hp)	$Q_1 = \widehat{\underline{\beta}}' X' \underline{Y}$		
Parâmetros ($\underline{\theta}$)	P	$Q_2 = \underline{\theta}' \sum_{t=1}^H X'_t \underline{Y}_t$		
Redução (H_0)	(H-1)p	$Q_3 = Q_1 - Q_2$	$V_1 = \frac{Q_3}{(H-1)p}$	V_1 / V_2
Resíduo	N-Hp	$Q_4 = Q_5 - Q_1$	$V_2 = \frac{Q_4}{N-Hp}$	
Total	N	$Q_5 = \underline{Y}' \underline{Y}$		

Fonte: GRAYBILL, 1976

Assim rejeita-se H_0 se e somente se $F_o \geq F_t$. A não rejeição da hipótese de nulidade permite concluir que, para um nível de significância α , as H equações não diferem entre si. Logo a equação ajustada com as estimativas dos coeficientes comuns pode ser usada como uma estimativa das H equações envolvidas (REGAZZI, 1993).

2.5 SUFICIÊNCIA AMOSTRAL

A intensidade de amostragem indica a porcentagem de área total da população que é incluída na amostra (HUSCH; MULLER, 1982). A primeira etapa para realizar uma amostragem consiste em determinar qual o tamanho necessário da amostra e uma resposta precisa pode não ser facilmente encontrada. A intensidade de amostragem ou fração de amostragem é a razão entre o número de unidades da amostra e o número total de unidades da população. A intensidade de amostragem pode ser determinada, basicamente por meio de dois procedimentos principais: em função da variabilidade da população, do erro de amostragem

admitido e da probabilidade de confiança fixada; ou em função do tempo e recursos disponíveis para a realização do inventário (PÉLLICO NETTO; BRENA, 1997). É importante sempre aferir a intensidade amostral com base na precisão, que é estabelecida via limite de erro especificado e probabilidade definida, geralmente 10% e 95%, respectivamente.

Para fazer o cálculo do número ideal de unidades amostrais é necessário saber se a população é finita ou infinita (PÉLLICO NETTO; BRENA, 1997). A diferença estatística de população finita e infinita é feita pelo valor do fator de correção (1-f), que é uma função de amostragem (f), aplicado às estimativas. Assim, se:

$$(1 - f) \geq 0,98 \rightarrow \text{a população é considerada infinita;}$$

$$(1 - f) \leq 0,98 \rightarrow \text{a população é considerada finita,}$$

onde f é a fração de amostragem que é calculado a partir do quociente entre o número de parcelas amostradas (n) e o número total de parcelas da população (N) ou número potencial de unidades amostrais:

$$f = \frac{n}{N} \quad \text{sendo que} \quad N = \frac{A}{a}$$

onde:

A - é a área total da população (m²);

a- é a área da unidade amostral ou parcela (m²).

Quando a população for considerada infinita, o fator de correção pode ser desprezado, tanto no cálculo da intensidade amostral, como no cálculo de estimativas que levam em consideração o fator de correção (PÉLLICO NETTO; BRENA, 1997). O cálculo da intensidade amostral (n) muda conforme o processo de amostragem escolhido e é efetuado com base no valor (t) de Student, do limite do erro (LE) pré-estabelecido e na variância ou no coeficiente de variação, que são medidas de heterogeneidade da população florestal.

Para se obter o número de unidades amostrais necessário para estimar os parâmetros da população com a precisão e confiabilidade fixadas, recomenda-se o recálculo da intensidade amostral (n₁), porém utilizando o valor (t) com base no novo número de graus de liberdade encontrado (n-1), até encontrar um valor constante

para a intensidade amostral. Esta estabilidade garante que o valor de (n) calculado será o ideal para a população inventariada (PÉLLICO NETTO; BRENA, 1997).

O cálculo da intensidade amostral para populações finitas, calculada em função da variância é dado por:

$$n = \frac{Nt^2 s^2}{NE^2 + t^2 s^2}$$

onde:

N = número total de amostras de dado tamanho existentes na população;

t = valor tabelado segundo a distribuição de Student;

S² = variância

E = (LE * \bar{X})

LE é o limite do erro fixado em função da média.

O procedimento de cálculo para populações infinitas é dado pela fórmula:

$$n = \frac{t^2 s^2}{E^2}$$

Se a intensidade amostral for calculada em função do coeficiente de variação e para uma população finita, seria calculada pela fórmula:

$$n = \frac{Nt^2 CV^2}{N(LE)^2 + t^2 CV^2}$$

onde:

CV = coeficiente de variação.

Para uma população infinita a fórmula se reduz a:

$$n = \frac{t^2 CV^2}{(LE)^2}$$

2.6 CURVA ESPÉCIE-ÁREA E CURVA DO COLETOR

A riqueza e a diversidade de espécies são diretamente influenciadas pela natureza da comunidade e pelo esforço amostral despendido, uma vez que o número de espécies aumenta com o aumento do número de indivíduos amostrados (BARROS, 2007; COLWELL et al., 2004). Portanto, é necessário selecionar amostras que representem o padrão florístico do local. Para uma amostragem eficiente, devem-se conhecer os princípios básicos dos métodos e procedimentos a serem adotados, e ainda, ter uma noção da estrutura da comunidade estudada. A maneira mais adequada de compreender-se a composição florística, a estrutura e a dinâmica de uma floresta de forma completa, é monitorá-la sistematicamente e por longos períodos. A metodologia conhecida mais indicada para tal finalidade é a instalação e medição de parcelas permanentes. Goldsmith e Harrison (1976) afirmam que mudanças na vegetação em longos períodos são melhor estudadas por meio de parcelas permanentes ou de transectos permanentes. Para Alder; Synnott (1992), uma amostragem da dinâmica, utilizando-se parcelas permanentes, concentra-se inicialmente na medição de taxas de mudança, incluindo o incremento em diâmetro, a taxa de mortalidade e de ingresso.

A distinção entre os termos “curva espécie/área” e “curva do coletor” torna-se importante devido às diferentes questões ecológicas que são abordadas em cada caso. Colwell; Coddington (1994) utilizam o termo “curva de acumulação de espécies” para as curvas construídas com dados de habitats aproximadamente homogêneos no tempo e no espaço, sendo a expressão “curva espécie/área” reservada para padrões biogeográficos, o que torna implícita a heterogeneidade da área.

Em 1835 Watson, de acordo com Hill *et al.* (1994) foi o primeiro a citar que o número de espécies tende a aumentar contínua e monotonicamente com a área. Porém, a ênfase na expressão matemática da relação ocorreu somente a partir dos trabalhos de Arrhenius (1921) e Gleason (1922,1925). Esses autores alimentaram uma controvérsia sobre a validade da fórmula empírica proposta por Arrhenius (1921), em que a soma das probabilidades de encontrar cada espécie em particular em uma área dava o número provável de espécies dessa área e que teve sua correção demonstrada para diversos tamanhos de amostra. Numa crítica a esse trabalho, Gleason (1922) sustentou que a equação proposta por Arrhenius (1921)

estava totalmente errada, pois, caso estivesse correta, o número de espécies deveria aumentar continuamente com o aumento da área (SCHILLING; BATISTA, 2008).

O estudo da curva espécie/área levou Braun-Blanquet (1932) a propor o uso dessa curva para indicar o “tamanho mínimo” de uma comunidade. Essa “área mínima” foi definida como a menor área que pode conter uma representação adequada da comunidade e o ponto em que a curva espécie/área torna-se aproximadamente horizontal determina o chamado “número mínimo” ou “combinação característica” de espécies. Porém, segundo Cain (1943), esse tamanho mínimo não garante por si só a representatividade da amostra, pois Braun-Blanquet (1932) utilizara povoamentos menores que a área mínima sob o argumento de que, de qualquer forma seria necessário observar a presença da combinação característica de espécies na área escolhida e que mesmo uma área maior do que a área mínima poderia não conter todas as espécies características daquela associação.

A curva construída pela agregação, ligada ao conceito de área mínima da associação, pressupõe o uso de uma parcela única formada a partir da lista de espécies em uma parcela relativamente pequena, que é gradativamente aumentada (normalmente dobrando seu tamanho) e tem as espécies adicionais contadas. Esse procedimento continua até que a adição de novas áreas não acrescente mais novas espécies ou, ao menos, um número significativo delas (BRAUN-BLANQUET, 1932; MUELLER-DOMBOIS; ELLEMBERG, 1974). Por esse sistema, a maior parcela sempre contém todas as anteriores e é então definida uma ordem não arbitrária para a construção da curva espécie-área. Porém, quando essa curva é construída de acordo com o conceito de representação quantitativa da comunidade vegetal, são utilizadas várias parcelas, geralmente pequenas e disjuntas e a ordem de entrada das parcelas torna-se arbitrária (SCHILLING; BATISTA, 2008).

A curva de acumulação de espécies ou curva do coletor é uma expressão da diversidade de espécies numa comunidade. Comparando-se comunidades de maior e menor diversidade, a curva das primeiras terão uma inclinação mais acentuada e uma assíntota mais elevada. Uma variação da curva proposta como representação da diversidade é a curva do número de espécie por número de indivíduos amostrados (COLWELL; CODDINGTON, 1994; GOTELLI; COLWELL, 2001).

Contudo, tradicionalmente o procedimento mais usado para definir a suficiência amostral a partir da curva do coletor, apesar de equivocado, é a

“inspeção visual” da curva para identificar o “ponto onde se inicia o patamar”. Cain; Castro (1956) já argumentavam que a visualização do patamar é fundamentalmente um artifício gráfico, pois dependendo da relação entre o comprimento dos eixos x e y (abscissas e ordenadas) o patamar pode ou não ser visualizado. A própria ideia da existência de um patamar, entendido como o ponto em que a inclusão de unidades amostrais não resultaria na inclusão de novas espécies na amostra, é outro aspecto limitante da curva do coletor (SCHILLING; BATISTA, 2008). Dessa forma, admitindo-se a inexistência do patamar na curva do coletor, torna-se impossível definir objetivamente, sem arbitrariedade, um ponto na curva onde o tamanho da amostra é considerado adequado ou suficiente para representar a comunidade.

Nos trabalhos de fitossociologia, a curva do coletor é geralmente construída utilizando-se a ordem em que as parcelas foram medidas no campo como ordem “natural” para acumular o número de espécies e o número de parcelas ou a área amostrada. Nas curvas assim geradas, é comum que o pesquisador interprete rápidos crescimentos ou patamares temporários como atributos característicos da comunidade sendo estudada. Numa perspectiva de métodos de amostragem, no entanto, a comunidade sendo estudada é a população alvo e qualquer ordem das unidades amostrais é absolutamente arbitrária, inclusive a ordem em que as unidades foram medidas em campo (COLWELL; CODDINGTON, 1994).

Cain; Castro (1956) propuseram um método para definir o tamanho ótimo da amostra definido pelo ponto em que a curva atinge determinada taxa de crescimento, que seria de fácil localização em qualquer curva do coletor, independente da razão entre os eixos. Ele sugeriu que a amostragem estaria adequada quando a um aumento de 10% na área amostrada correspondesse um aumento de 10% no número de espécies. Conforme explicou em um trabalho Cain, (1943) esse é o ponto na curva em que o incremento em novas espécies é igual à média desse incremento (número total de espécies encontrado dividido pela área total amostrada). Porém, a localização desse ponto depende da área total amostrada e em estudo realizado por Rice; Kelting (1955) foram comparadas curvas baseadas em amostras de tamanho crescente, obtendo-se valores diferentes de área mínima com o aumento do tamanho da amostra – quanto maior a área total amostrada, maior a área mínima obtida.

Segundo Schilling; Batista (2008), suficiência amostral é definida como uma área mínima que garanta a “composição característica” da associação ou comunidade vegetal. Essa área mínima pode ser composta de uma única área

referente a um “stand” observado no campo ou pode ser composta de várias parcelas de mesma área, que juntos compõem uma amostra da associação ou comunidade.

Porém, apesar do conceito teórico de suficiência amostral estar ligado à composição característica da comunidade, o uso da curva do coletor como método para determinação da área mínima não implica numa composição característica, mas sim num número mínimo de espécies presentes na amostra SCHILLING; BATISTA (2008).

Segundo os autores a pressuposição de uma composição florística característica, fixa e definida, para a comunidade está estreitamente ligada ao conceito de comunidade como associação vegetal. A impossibilidade de operacionalizar o conceito de suficiência amostral por meio de uma técnica quantitativa vai além da questão da curva do coletor e remete ao próprio conceito de comunidade vegetal.

O uso da amostragem como método científico se fundamenta na amostragem estatística e, nesse contexto, não é possível avaliar a qualidade da amostra sem uma definição objetiva e empírica da população a ser amostrada. Outro aspecto fundamental na avaliação da qualidade de uma amostra estatística e que é ignorado nas definições de suficiência amostral é o custo da informação gerada pela amostra SCHILLING; BATISTA (2008).

O conceito de suficiência amostral remete à qualidade com que a composição florística é representada, o que pode ser interpretado estatisticamente como a precisão da informação gerada pela amostra. O tamanho ótimo da amostra será aquele que maximize a precisão, mantendo o custo constante ou, de forma equivalente, minimize o custo mantendo a precisão constante (SHIVER; BORDERS, 1996).

Estatisticamente esse é o maior problema com o conceito de suficiência amostral e com a própria curva do coletor, quando aplicada a florestas tropicais. Na curva do coletor, à medida que se observam mais unidades amostrais (parcelas ou pontos), a representação florística sempre melhora, isto é, aumenta a exatidão da informação obtida.

2.7 ANÁLISE FATORIAL

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, onde 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).

Os fatores explicam parte da variabilidade total dos dados por meio da soma das variâncias das variáveis originais. Um método bastante conhecido para a obtenção dos fatores é o da máxima verossimilhança (Johnson; Wichern, 1992), que consiste na suposição de que as variáveis envolvidas sigam distribuição normal, fato este que limita sua utilização.

Outro método bastante utilizado é baseado na Análise de Componentes Principais ACP, cuja vantagem é que não há pressuposição da normalidade das variáveis envolvidas. Os fatores são obtidos a partir da decomposição da matriz de correlação ou de covariância. Na decomposição da origem cargas fatoriais indicam quanto cada variável está associada a cada fator e os autovalores associados a cada um dos fatores envolvidos (MARRIOTT, 1974).

Os autovalores são números que refletem a importância do fator. Quando o número de fatores é igual ao número de variáveis, a soma dos autovalores corresponde à soma das variâncias dessas variáveis. Logo a razão entre um autovalor e a soma das variâncias indica a proporção da variabilidade total dos dados que é explicada pelo fator. A soma das proporções relativas aos fatores considerados na análise reflete o quanto da variabilidade dos dados é explicado pelo conjunto de fatores (MARRIOTT, 1974).

Segundo Johnson e Wichern (1992) seja o vetor aleatório observável \underline{X} , com p componentes, $\underline{X} \sim (\underline{\mu}, \Sigma)$ o modelo fatorial postula que \underline{X} é linearmente dependente

sobre algumas variáveis aleatórias não observáveis, F_1, F_2, \dots, F_m , para $m \leq p$, chamados fatores comuns e p fontes de variações aditivas: $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamados erros ou fatores específicos.

O modelo fatorial é escrito por:

$$X_i - \mu_i = l_j F_j + \varepsilon_i$$

onde:

μ_i média da i -ésima variável

ε_i erro ou fator específico da i -ésima variável

l_j peso ou carregamento na i -ésima variável X_i do i -ésimo fator F_j , $i = 1, 2, \dots, p$ e

$j = 1, 2, \dots, m$

Em notação matricial:

$$\underline{X} - \underline{\mu} = L * \underline{F} + \underline{\varepsilon}$$

onde:

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad L = \begin{bmatrix} e_{11} & e_{12} & e_{1m} \\ e_{21} & e_{22} & e_{2m} \\ e_{1p} & e_{2p} & e_{pm} \end{bmatrix} \quad \underline{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

Os vetores aleatórios não observáveis \underline{F} e $\underline{\varepsilon}$ satisfazem as seguintes hipóteses:

- \underline{F} e $\underline{\varepsilon}$ são independentes;
- $E(\underline{F}) = 0$, $cov(\underline{F}) = I$
- $E(\underline{\varepsilon}) = 0$, $cov(\underline{\varepsilon}) = \varphi$, onde φ é a matriz diagonal.

O modelo fatorial ortogonal implica uma estrutura de covariância de X , tal que:

$$\Sigma = LL' + \varphi$$

Ainda segundo Johnson e Wichern (1992) a porção da variância de i -ésima variável aleatória X_i , advinda como contribuição dos m fatores comuns é chamada comunalidade e a porção dessa variância, $V(X_i) = \sigma_{ii} = \sigma_i^2$, advinda do fator específico, ou seja, do erro específico é a variância específica. Designado a i -ésima comunalidade por h_i^2 , tem-se

$$\sigma_i^2 = h_i^2 + \varphi_i$$

As comunalidades são índices atribuídos às variáveis originais que expressam em termos percentuais o quanto da variação de cada variável é explicada pelo modelo de AF estimado. Quanto mais perto de 1 estiverem as comunalidades, melhor será o ajuste da AF.

$$h_i^2 = e_{1i}^2 + e_{2i}^2 + \dots + e_{mi}^2$$

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. Existem na literatura vários critérios que auxiliam na determinação do número de fatores que, invariavelmente, quando empregados em um mesmo conjunto de dados, conduzem a resultados diferentes. Como regra deve-se procurar um compromisso entre o número de fatores (a princípio o menor possível) e a sua interpretabilidade.

Para Kachigan (1991), existem vários métodos de escolha do número de fatores. O critério de Kaiser, também conhecido como critério da raiz latente, determina que o número de fatores deve ser igual ao número de autovalores maiores ou igual à média das variâncias das variáveis analisadas. No caso em que a AF é feita sobre a matriz de correlação, esse critério corresponde à exclusão de fatores com autovalores inferiores a 1. No critério da porcentagem da variância explicada o número é determinado de modo que o conjunto de fatores explique uma porcentagem pré-definida (em geral 70%) da variabilidade global.

Rotação dos fatores: Segundo Johnson; Wichern (1992), nem sempre a matriz de pesos originais obtida permite uma interpretação clara dos fatores e para se obter uma estrutura mais simples, utiliza-se a rotação dos fatores, que pode ser feita sem afetar a validade do modelo e fornece uma matriz onde os grupos de variáveis correlacionadas aparecem com maior nitidez. Novos fatores comuns podem ser encontrados: F_1^* , F_2^* , ..., F_m^* que são combinações lineares dos fatores comuns F_1, F_2, \dots, F_m e são não correlacionados com variâncias unitárias.

Para o autor uma transformação ortogonal corresponde a uma rotação rígida dos eixos coordenados. O ideal é obter uma estrutura de pesos tal que cada variável tenha alto peso em um único fator e baixo ou moderado nos demais e isso nem sempre é possível embora a rotação forneça um resultado bem próximo disso.

Na rotação denominada Varimax (maximiza a variância) o processo tem como objetivo encontrar eixos com o menor número possível de pesos altos e tantos pesos próximos de zero quanto possível e é definido como sendo os coeficientes rotacionados escalonados pela raiz quadrada das comunalidades.

Escore fatorial são os valores estimados dos fatores comuns. São estimativas dos valores não observáveis dos vetores aleatórios F_j $j = 1, 2, \dots, n$, isto é, os escores fatoriais \hat{f}_j são a estimativa do valor f_j atingido por F_j (j -ésimo caso).

Pode-se usar os componentes principais para estimar os escores fatoriais e para tanto usa-se o método dos mínimos quadrados ordinários que consiste em considerar todas as variâncias iguais. Desta forma:

$$\hat{f}_j = (L'L)^{-1} L' (x_j - \bar{X}) \quad \text{para matriz de covariância}$$

$$\hat{f}_j = (L'L)^{-1} L'_z Z_j \quad \text{para matriz de correlação}$$

Escore fatorial final é o resultado da transformação dos diversos escores em um escore único, representado por:

$$(\text{proporção explicada} / \lambda_1) * \hat{f}_1 + (\text{proporção explicada} / \lambda_2) * \hat{f}_2$$

2.8 ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais - ACP pode ser resumida em três passos, quais sejam a derivação da matriz de correlação ou de variância e covariância, o cálculo dos autovetores e autovalores, e a transformação linear do conjunto de dados (RICHARDS, 1993).

Esta transformação tem como base a rotação do espaço de atributos na direção dos autovetores com o intuito de tornar ortogonal o conjunto de dados, promovendo a união das informações de maior correlação nas primeiras componentes (MATHER, 1999). A análise de componentes principais reduz a dimensionalidade dos dados, concentrando as informações significativas nas primeiras componentes geradas pela transformação (SCHOWENGERDT, 1997).

O objetivo da Análise de Componentes Principais é abordar aspectos como a geração, a seleção e a interpretação das componentes investigadas. Para a determinação das componentes principais, é necessário calcular a matriz de variância e covariância (Σ), ou a matriz de correlação (R), encontrar os autovalores e os autovetores e, por fim, escrever as combinações lineares, que serão as novas variáveis, denominadas de componentes principais, sendo que cada componente principal é uma combinação linear de todas as variáveis originais, independentes entre si e estimadas com o propósito de reter, em ordem de estimação e em termos da variação total, contida nos dados iniciais. A interpretação de uma componente principal é feita mediante o grau de importância, ou, ainda, a influência que cada variável tem sobre cada componente, sendo que esta importância é dada pela correlação entre cada variável X_{ij} e o componente Y_i que estiver sendo interpretado (REGAZZI, 2001).

Geração dos componentes principais: algebricamente componentes principais são combinações lineares particulares das p variáveis aleatórias originais x_1, x_2, \dots, x_p . Geometricamente estas combinações lineares representam a seleção de um novo sistema de coordenadas obtido por rotação do sistema original com x_1, x_2, \dots, x_p como eixos. Os novos eixos y_1, y_2, \dots, y_p , representam as direções com variabilidade máxima, e fornecem uma descrição mais simples e mais parcimoniosa da estrutura de covariância. As componentes principais dependem da matriz de covariância (ou da matriz de correlação) de x_1, x_2, \dots, x_p . Seu desenvolvimento não necessita da suposição de gaussianidade (distribuição normal). Por outro lado,

Componentes Principais derivados de populações normais multivariadas tem sua interpretação em termos de elipsoide de densidade constante (JOHNSON; WICHERN, 1992).

Seja $\underline{X} = [X_1, X_2, \dots, X_p]$, o vetor aleatório p-dimensional com um vetor de médias $\underline{\mu}$ e matriz de covariância Σ , com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, considere as combinações lineares:

$$Y_1 = \underline{c}'_1 \underline{X} = c_{11} X_1 + c_{21} X_2 + \dots + c_{p1} X_p$$

$$Y_2 = \underline{c}'_2 \underline{X} = c_{12} X_1 + c_{22} X_2 + \dots + c_{p2} X_p$$

.....

$$Y_i = \underline{c}'_i \underline{X} = c_{1i} X_1 + c_{2i} X_2 + \dots + c_{pi} X_p$$

Resulta em:

$$\text{Var}(Y_i) = \underline{c}'_i \Sigma \underline{c}_i \quad i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i, Y_k) = \underline{c}'_i \Sigma \underline{c}_k \quad i, k = 1, 2, \dots, p$$

Os Componentes Principais são essas combinações lineares não correlacionadas y_1, y_2, \dots, y_i cujas variâncias são as maiores possíveis. O Primeiro Componente Principal é a combinação linear $\underline{c}'_1 \underline{X}$, com máxima variância, que maximiza a $\text{Var}(\underline{c}'_1 \underline{X})$ sujeito a $\underline{c}'_1 \underline{c}_1 = 1$. O Segundo Componente Principal é a combinação linear $\underline{c}'_2 \underline{X}$, que maximiza a $\text{Var}(\underline{c}'_2 \underline{X})$ sujeito a $\underline{c}'_2 \underline{c}_2 = 1$ e com $\text{Cov}(\underline{c}'_1 \underline{X}, \underline{c}'_2 \underline{X}) = 0$. Assim, o i-ésimo Componente Principal é a combinação linear $\underline{c}'_i \underline{X}$, que maximiza $\text{Var}(\underline{c}'_i \underline{X})$ sujeito a $\underline{c}'_i \underline{c}_i = 1$ e $\text{Cov}(\underline{c}'_i \underline{X}, \underline{c}'_k \underline{X}) = 0$, para $k < i$.

Da aplicação de alguns resultados, resulta que (JOHNSON; WICHERN, 1992) :

- Os componentes principais não são correlatos e possuem variâncias iguais aos autovalores de Σ ;
- Se algum λ_i (autovalor) é igual a outro, a escolha dos correspondentes (autovetores) e, portanto Y_i , não são únicos.

- A variação total é igual a soma das variâncias dos componentes principais e, conseqüentemente, a proporção da variância total explicada pela k-ésima componente principal é:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

- Se a maior parte da variância pode ser atribuída a uma, duas ou três componentes principais, então estas podem substituir as originais p variáveis, sem muita perda de informação.
- Se $Y_1 = \underline{e}'_1 \underline{X}$; $Y_2 = \underline{e}'_2 \underline{X}$; ... ; $Y_p = \underline{e}'_p \underline{X}$, são componentes principais da matriz Σ , então:

$$\rho = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

são os coeficientes de correlação entre as y_i componentes e as variáveis x_k .

Do ponto de vista geométrico, tem-se um diagrama de dispersão dos dados de n pontos em um espaço p-dimensional:

- a análise de componentes principais é uma rotação de eixos tal que a variância total das projeções dos pontos sobre o primeiro é máximo (primeiro componente principal);
- o segundo eixo (segundo componente principal) é ortogonal ao primeiro e absorve tanto quanto possível da variância remanescente.

2.9 ANÁLISE DE VARIÂNCIA MULTIVARIADA - MANOVA

Os dados multivariados se parecem muito com os univariados: consistem de uma ou mais variáveis independentes (preditoras) e duas ou mais variáveis dependentes (respostas). A distinção entre os dados univariados e multivariados recai em grande parte sobre como os dados são organizados e analisados, e não em como são coletados (GOTELLI; AARON, 2011)

Para Johnson; Wichern (1982), a análise de variância multivariada permite comparar diversas amostras multivariadas, a partir de vetores médios. A proposta desta técnica é verificar se as diversas amostras de unidades multivariadas são oriundas de populações centradas no mesmo ponto médio. Algumas pressuposições da estrutura dos dados devem ser obedecidas para validade da inferência estatística:

- (a) $X_{i1}, X_{i2}, \dots, X_{ini}$ deve ser uma amostra aleatória de tamanho n_i do tratamento i , com média μ_i , $i=1, 2, \dots, g$. As amostras dos tratamentos devem ser independentes;
- (b) todos os tratamentos possuem covariância comum Σ ; e
- (c) cada tratamento tem distribuição normal multivariada.

O modelo de análise de variância multivariada é apresentado a seguir. Neste modelo cada componente é um vetor de p componentes.

$$X_{ij} = \underline{\mu} + \underline{\tau}_i + \underline{e}_{ij} \quad i = 1, 2, \dots, g \text{ e } j = 1, 2, \dots, n$$

em que,

- \underline{e}_{ij} é independentemente e identicamente distribuído e $N_p(0, \Sigma)$ para todo i e j ;
- $\underline{\mu}$ é o vetor média geral;
- $\underline{\tau}_i$ representa o vetor de efeitos do i -ésimo tratamento. Pode-se adotar a restrição paramétrica $\sum_{i=1}^g n_i \underline{\tau}_i = \underline{0}$

Os erros do vetor X_{ij} são correlacionados, no entanto a matriz de covariância Σ é a mesma para todos os tratamentos. O vetor de observações pode ser decomposto em:

$$\begin{array}{ccccccc} \underline{X}_{ij} & = & \underline{\bar{X}} & + & (\underline{\bar{X}}_i - \underline{\bar{X}}) & + & (\underline{X}_{ij} - \underline{\bar{X}}_i) \\ \text{Observação} & & \text{Estimativa da} & & \text{Estimativa do} & & \text{resíduo} \\ & & \text{média geral} & & \text{efeito do tratamento} & & \end{array}$$

Analogamente, demonstra-se que a soma de quadrados e produtos totais possui a seguinte decomposição:

Soma de quadrados e produtos (SQP) = SQP tratamentos + SQP resíduo total corrigido

$$\sum_{i=1}^g \sum_{j=1}^n (\underline{X}_{ij} - \bar{X})(\underline{X}_{ij} - \bar{X})^t = \sum_{i=1}^g n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^t + \sum_{i=1}^g \sum_{j=1}^n (\underline{X}_{ij} - \bar{X}_i)(\underline{X}_{ij} - \bar{X}_i)^t$$

A soma de quadrados e produtos do resíduo pode ser expressa por:

$$\sum_{i=1}^g \sum_{j=1}^n (\underline{X}_{ij} - \bar{X}_i)(\underline{X}_{ij} - \bar{X}_i)^t = (n_1-1)s_1 + (n_2-1)s_2 + \dots + (n_g-1)s_g$$

onde:

S_i é a matriz de covariância amostral do i-ésimo tratamento.

O teste da hipótese de inexistência de efeitos de tratamentos é realizado considerando as magnitudes das somas de quadrados e produtos de tratamento e resíduo pela variância generalizada. Desta forma:

$$H_0 = T_1 = T_2 = \dots = T_g = 0$$

H_1 = pelo menos um dos tratamentos T_i é diferente dos demais.

O esquema de análise de variância multivariada (MANOVA) é apresentado na Tabela 2. A fonte de variação total é particionada em causas de variação devido ao tratamento e ao erro experimental ou resíduo.

TABELA 2 - MANOVA PARA TESTAR A HIPÓTESE DE IGUALDADE DO VETOR DE EFEITO DOS TRATAMENTOS.

FONTE DE VARIAÇÃO	GL	MATRIZ DAS SQPC
TRATAMENTO	$g - 1$	$B = \sum_{i=1}^g n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^t$
RESÍDUO	$\sum_{i=1}^g n_i - g$	$W = \sum_{i=1}^g \sum_{j=1}^n (\underline{X}_{ij} - \bar{X}_i)(\underline{X}_{ij} - \bar{X}_i)^t$
TOTAL	$\sum_{i=1}^g n_i - 1$	$B+W = \sum_{i=1}^g \sum_{j=1}^n (\underline{X}_{ij} - \bar{X})(\underline{X}_{ij} - \bar{X})^t$

Fonte: Johnson e Wichern (1992)

Se H_0 é verdadeira, $F \sim F_{g-1, n-g}$. A hipótese de nulidade é rejeitada ao nível de significância α se $F > F_{g-1; n-g, (1-\alpha)}$.

Segundo Field (2009) quatro critérios existem para o teste desta hipótese: O traço de Pillai-Bartlett (V) representa os autovalores (λ) para cada uma das funções discriminantes e o número dessas funções. O traço de Pillai é a soma das proporções da variância explicada nas funções discriminantes; é semelhante à razão SS_M/SS_T , ou seja, é a razão entre a soma dos quadrados devido à variável de agrupamento ou variação sistemática (SS_M) e a soma de quadrados total SS_T :

$$V = \sum_{i=1}^S \frac{\lambda_i}{1 + \lambda_i}$$

O valor T^2 de Hotelling (T) é a soma dos autovalores para cada combinação linear, e desse modo se compara à razão F da ANOVA. Esta estatística é a razão SS_M/SS_T para cada combinação linear:

$$T = \sum_{i=1}^S \lambda_i$$

O valor de Lambda de Wilks (Λ) representa a razão da variância do erro pela variância total para cada combinação linear, ou seja, é a razão SS_R/SS_T onde SS_R é a soma de quadrado do resíduo. Muitos autores recomendam utilizar o critério lambda de Wilks como referência, por se tratar de um teste baseado na razão de verossimilhança.

$$\Lambda^* = \frac{|W|}{|B+W|}$$

A maior raiz de Roy representa a proporção da variância explicada pela não explicada, isto é SS_M/SS_R para a primeira função discriminante. Pode-se dizer que é o autovalor para a primeira combinação linear.

$$\text{Maior raiz} = \lambda_{\text{máximo}}$$

Olson (1974, 1976 e 1979) observou que para tamanhos amostrais pequenos e moderados, as quatro estatísticas diferem pouco em termos de poder. Se as

diferenças estiverem concentradas na primeira combinação linear, a estatística de Roy deverá ser mais poderosa, seguido pelo traço de Hotelling, λ de Wilks e o traço de Pillai. Contudo, quando os grupos diferem ao longo de mais de uma combinação linear, o poder maior é do traço de Pillai e a raiz de Roy é o menor.

Outra questão relativa ao poder dos testes é quanto ao tamanho da amostra e o número de variáveis dependentes. Stevens (1980) recomenda a utilização de um número pequeno de variáveis dependentes (menos de 10), exceto se os tamanhos amostrais forem grandes.

Em termos de robustez, os quatro testes estatísticos são relativamente robustos a violação da normalidade multivariada, embora a raiz de Roy seja afetada por distribuições platicúrticas. A raiz de Roy também não é robusta quando a homogeneidade das matrizes de covariâncias não se verifica.

Bray e Maxwell (1985) concluíram que quando tamanhos amostrais são iguais o traço de Pillai é o mais robusto a violação das hipóteses, contudo, quando os tamanhos amostrais são diferentes, essa estatística é afetada pela violação da hipótese de igualdade das matrizes de covariâncias.

2.10 ANÁLISE DE AGRUPAMENTOS - CLUSTER

A análise de agrupamentos ou “cluster” é um processo de partição de uma população heterogênea em vários subgrupos mais homogêneos. No agrupamento, não há classes pré-definidas, os elementos são agrupados de acordo com a semelhança, o que a diferencia da tarefa de classificação (DONI, 2004).

Dentre os procedimentos para a realização de uma análise de agrupamentos destacam-se os métodos hierárquicos e os não-hierárquicos. Para descrevê-los apropriadamente é necessário definir inicialmente as medidas de similaridade e o uso da matriz de similaridade, para, em seguida, descrever os métodos, seus algoritmos, funções distância e algumas características.

Segundo Zaine (2003), dada a dificuldade de se examinar todas as combinações de grupos possíveis em um grande volume de dados, foram desenvolvidas diversas técnicas capazes de auxiliar na formação dos agrupamentos. Uma análise de cluster criteriosa exige métodos que apresentem as seguintes características:

- ser capaz de lidar com dados com alta dimensionalidade;
- ser “escalável” com o número de dimensões e com a quantidade de elementos a serem agrupados;
- habilidade para lidar com diferentes tipos de dados;
- capacidade de definir agrupamentos de diferentes tamanhos e formas;
- exigir o mínimo de conhecimento para determinação dos parâmetros de entrada;
- apresentar resultado consistente independente da ordem em que os dados são apresentados;

Em geral, nenhum algoritmo atende a todos esses requisitos e, por isso, é importante entender as características de cada algoritmo para a escolha de um método adequado a cada tipo de dado ou problema (HALDIKI, *et al.*, 2001).

A maioria dos métodos de análise de agrupamentos requer uma medida de similaridade entre os elementos a serem agrupados, normalmente expressas como uma função distância ou métrica.

Distância euclidiana: a distância euclidiana é a distância geométrica no espaço multidimensional. A distância euclidiana entre dois elementos $X = [X_1, X_2, \dots, X_p]$ e $Y = [Y_1, Y_2, \dots, Y_p]$, é definida por:

$$d_{xy}^2 = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2}$$

Distância euclidiana quadrática : A distância euclidiana quadrática é definida pela expressão:

$$d_{XY} = (X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2$$

Distância de Manhattan: a distância de Manhattan é definida pela expressão:

$$d_{xy} = |X_1 - Y_1| + |X_2 - Y_2| + \dots + |X_p - Y_p|$$

Em muitos casos a distância de Manhattan apresenta resultados similares ao da distância Euclidiana. Entretanto, nessa medida, o efeito de uma grande diferença entre uma das dimensões de um elemento é minimizado, já que a mesma não é elevada ao quadrado.

Distância de Chebychev: a distância de Chebychev é apropriada no caso em que se deseja definir dois elementos como diferentes, se apenas umas das dimensões é diferente. Ela é definida por:

$$d_{xy} = \text{máximo} (|X_1 - Y_1|); (|X_2 - Y_2|); \dots\dots\dots; (|X_p - Y_p|)$$

As medidas de similaridade são utilizadas na análise de agrupamentos de forma a determinar a distância entre elementos. Essa distância é normalmente representada na forma de matriz, ou seja, em uma matriz de similaridade.

2.10.1 Métodos Hierárquicos

O método hierárquico de cluster consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, onde os mesmos são agregados ou desagregados. Os métodos hierárquicos são subdivididos em: aglomerativos e divisivos. Os grupos, nos métodos hierárquicos, são geralmente representados por um diagrama bi-dimensional, chamado de dendograma ou diagrama de árvore (DONI, 2004).

Segundo Anderberg (1973), no método aglomerativo, cada elemento representa um grupo, e a cada passo, um grupo ou elemento é ligado a outro de acordo com sua similaridade, até o último passo, onde é formado um grupo único com todos os elementos. Como o objetivo da análise de agrupamentos ou de Cluster é reunir objetos semelhantes, torna-se necessária alguma medida para avaliar o quão semelhantes, ou diferentes, são os objetos (REGAZZI, 2001).

Existe uma variedade de métodos aglomerativos, que são caracterizados de acordo com o critério utilizado para definir as distâncias entre grupos. Entretanto, a maioria dos métodos parecem ser formulações alternativas de três grandes conceitos de agrupamento aglomerativo, quais sejam os métodos de ligação, centroide e o de Ward. Os métodos de ligação se dividem em ligação por vizinho mais próximo ou “single linkage”, ligação por vizinho mais distante ou “complete linkage” e ligação média ou “average linkage”.

O método de ligação por vizinho mais próximo emprega a distância de valor mínimo:

$$d(AB)W = \min (dAW, dBW)$$

Segundo Anderberg (1973) são características desse método que em geral os grupos muito próximos podem não ser identificados, e permite detectar grupos de formas não-elípticas, que apresenta pouca tolerância a ruído, pois tem tendência a incorporar os ruídos em um grupo já existente, que apresenta bons resultados tanto para distâncias euclidianas quanto para outras distâncias e que tem tendência a formar longas cadeias (encadeamento).

No método de ligação por vizinho mais distante é empregada a distância máxima, dada por:

$$d(AB)W = \max (dAW, dBW)$$

Kaufmann; Rousseeuw (1990) cita que esse método apresenta bons resultados tanto para distâncias euclidianas quanto para outras distâncias, tem tendência a formar grupos compactos e os ruídos demoram a serem incorporados ao grupo.

Os métodos de ligação pelo vizinho mais próximo e pelo vizinho mais distante trabalham em direções opostas. Se eles apresentam resultados semelhantes, significa que o grupo está bem definido no espaço, ou seja, o grupo é real. Mas se ocorre o contrário, os grupos provavelmente não existem (ROMESBURG, 1984).

No método de ligação por média a função distância é definida por:

$$\frac{(N_X d_{XW} + N_Y d_{YW})}{N_X + N_Y}$$

onde N_x e N_y são os números dos elementos no grupo X e Y respectivamente e d_{xw} e d_{yw} são as distâncias entre os elementos XW e YW, respectivamente.

Kaufmann; Rousseeuw (1990) citam como algumas características desse método a menor sensibilidade a ruídos que os métodos de ligação por vizinho mais próximo e por vizinho mais distante, que apresenta bons resultados tanto para distâncias euclidianas quanto para outras distâncias e que tem tendência a formar grupos com número de elementos similares.

Para o método de ligação por centroide, a distância entre dois grupos é definida como a distância euclidiana ao quadrado entre os vetores de médias centróides, dos dois grupos. A função distância é definida por:

$$\frac{(N_X d_{XW} + N_Y d_{YW})}{N_X + N_Y} - \frac{N_X N_Y d_{XY}}{(N_X + N_Y)^2}$$

onde N_x e N_y são os números dos elementos no grupo X e Y respectivamente e d_{xw} e d_{yw} são as distâncias entre os elementos XW e YW, respectivamente.

Kaufmann; Rousseeuw (1990) citam como características desse método a robustez a presença de ruídos e que devido ao fenômeno de reversão, o método não é muito usado. O fenômeno da reversão consiste na junção dos grupos em que as distâncias não são estritamente crescentes. Nesse caso, ocorre um cruzamento, interseção no dendograma resultante (FERREIRA, 2011).

A função distância para o método de ligação por mediana é dada por:

$$d_{(XY)W} = \frac{d_{XW} + d_{YW}}{2} + \frac{d_{XY}}{4}$$

onde d_{xw} e d_{yw} são as distâncias entre os elementos XW e YW, respectivamente.

Dentre as características referentes a esse método, citadas por Kaufman; Rousseeuw (1990) destaca-se que apresenta resultado satisfatório quando os grupos possuem tamanhos diferentes, que pode apresentar resultado diferente quando permutado os elementos na matriz de similaridade, que apresenta robustez à presença de outliers e que apresenta o fenômeno da reversão.

Ward (1963) propõe um processo geral de classificação em que n elementos são progressivamente reunidos dentro de um grupo por meio da minimização de uma função objetiva para cada (n-2) passos de fusão.

Nesse método, a função distância é dada por:

$$d_{(XY)W} = \frac{((N_W - N_X)d_{XY} + (N_W - N_Y)d_{YW} - N_W d_{XW})}{N_W + N_X + N_Y}$$

onde N_x , N_w e N_y são os números dos elementos no grupo X, Y e W respectivamente e d_{xw} , d_{yw} e d_{xy} são as distâncias entre os elementos XW, XY e YW, respectivamente.

As características desse método, citadas por Romesburg (1984), são que apresenta bons resultados tanto para distâncias euclidianas quanto para outras distâncias, que pode apresentar resultados insatisfatórios quando o número de elementos em cada grupo é praticamente igual, que tem tendência a combinar grupos com poucos elementos e que é sensível à presença de outliers.

Os métodos divisivos trabalham na direção oposta dos métodos aglomerativos, ou seja, um grupo inicial contendo todos os elementos é dividido em dois subgrupos, de tal forma que os elementos em um subgrupo estejam distantes dos elementos do outro subgrupo. Esses subgrupos são então divididos em subgrupos dissimilares e o processo continua até cada elemento formar um grupo.

Os métodos divisivos são pouco mencionados na literatura, pois exigem uma maior capacidade computacional que os métodos aglomerativos, é possível notar que o método divisivo exige um maior número de iterações. No primeiro passo do método aglomerativo são consideradas todas as uniões possíveis de dois elementos. Este número cresce quadraticamente à medida que N aumenta, ou seja, seu crescimento é acelerado. Mesmo assim a implementação do algoritmo é computacionalmente viável.

Já o método divisivo, baseado no mesmo princípio, começará considerando todas as divisões dos elementos em dois grupos, com pelo menos um elemento em cada grupo (KAUFMAN; ROUSSEUW, 1990). Comparando apenas o primeiro passo cresce exponencialmente à medida que N aumenta, e dessa forma, para um grande número de elementos, torna-se inviável sua implementação computacional.

2.10.2. Métodos não-hierárquicos ou por particionamento

Os métodos não-hierárquicos, ou por particionamento, foram desenvolvidos para agrupar elementos em K grupos, onde K é a quantidade de grupos definida previamente. Nem todos os valores de K apresentam grupos satisfatórios, sendo assim, aplica-se o método várias vezes para diferentes valores de K , escolhendo os

resultados que apresentem melhor interpretação dos grupos ou uma melhor representação gráfica (BUSSAB, 1990).

A ideia central da maioria dos métodos por particionamento é escolher uma partição inicial dos elementos e, em seguida, alterar os membros dos grupos para obter-se a melhor partição (ANDERBERG, 1973).

Quando comparado com o método hierárquico, o método por particionamento é mais rápido porque não é necessário calcular e armazenar, durante o processamento, a matriz de similaridade. Em geral, os métodos por particionamento diferem entre si pela maneira que constituem a melhor partição. Os métodos por particionamento mais conhecidos são o método de k-médias ou “k-means” e o método k-medóides ou “k-medoid”.

O dendrograma ou fenograma, também chamado de gráfico em árvore, representa uma síntese gráfica do trabalho desenvolvido, resumindo a informação, que ocasiona uma pequena perda da mesma, pelo fato de ser uma síntese. Apesar disto o gráfico é de grande utilidade para a classificação, comparação e discussão de agrupamentos.

Segundo Landim (2006), no dendrograma estão dispostas linhas ligadas segundo os níveis de similaridade que agrupam pares de espécies ou de variáveis. Como este gráfico é uma simplificação de duas dimensões de uma relação unidimensional é inevitável que algumas distorções quanto à similaridade apareçam. A medida de tal distorção pode ser obtida por um coeficiente de correlação, dito, cofenético, entre os valores da matriz inicial de similaridade e aqueles derivados do dendrograma.

2.11 ANÁLISE DISCRIMINANTE

A análise Discriminante é uma ferramenta estatística utilizada para classificar um determinado elemento num determinado grupo de variáveis, entre os diversos grupos existentes. Para tal é necessário que o elemento a ser classificado pertença realmente a um dos grupos, e que sejam conhecidas as características dos elementos dos diversos grupos (KASZNAR, 1986).

Conhecer os grupos significa que o número de grupos e os objetos de cada grupo são perfeitamente identificados antes de a análise ser aplicada. O objetivo é

determinar a qual dos grupos ou populações, cujas densidades de probabilidades são conhecidas, irá pertencer um novo objeto ou um novo conjunto de objetos (FERREIRA, 2011)

Segundo Johnson; Wichern (1982), a ideia consiste em separar duas classes de objetos ou fixar um novo objeto, em uma das duas classe (populações) de π_1 e π_2 . Os objetos são ordinariamente separados ou classificados com base nas medidas das p-variáveis aleatórias associadas

$$\underline{X}' = [X_1, X_2, \dots, X_p]$$

A ideia de Fischer foi transformar as observações multivariadas X , nas observações univariadas Y , tal que os Y 's derivados das populações π_1 e π_2 sejam separados tanto quanto possível. Fischer sugeriu o uso da combinação linear das variáveis aleatórias de X para a obtenção dos Y 's, por que são funções simples de se obter matematicamente. Não há suposição de normalidade para este método.

Sejam:

- $\mu_1 y$ - as médias dos Y 's obtidos dos X 's pertencentes à população ou grupo π_1 ;
- $\mu_2 y$ - as médias dos Y 's obtido dos X 's pertencentes à população ou grupo π_2 .

O método de Fischer sugere que se selecione a combinação linear que maximiza a distância quadrática entre $(\mu_1 Y)$ e $(\mu_2 Y)$ relativa a variabilidade dos Y 's.

A análise discriminante reduz o número de variáveis para um número menor de parâmetros, que são funções discriminantes linearmente dependentes das variáveis originais. Desta forma, os grupos poderão ser visualizados num espaço multidimensional, menor que o anterior e os coeficientes das funções discriminantes indicarão a contribuição das variáveis originais para cada função discriminante (LONGHI, 1997).

O método discriminante mais usado é o "stepwise", que seleciona as variáveis que vão participar da análise, baseada na capacidade de discriminação. O processo tem início com a seleção da variável que tem maior valor de discriminação, esta variável é pareada com as demais, uma de cada vez e o critério usado para a seleção é novamente comparado. A variável que, em conjunto com a primeira selecionada, produzir o melhor valor para o critério é a segunda a ser escolhida para entrar na análise. Essas duas variáveis selecionadas são combinadas com as demais e a combinação que apresentar maior valor para o critério de seleção

determinará a terceira a entrar na análise, e assim sucessivamente, até que todas sejam ordenadas pela sua capacidade de discriminação, ou eliminadas por não apresentarem potencial mínimo tolerado de discriminação (FIELD, 2009).

Muitas vezes as técnicas de análise discriminante e agrupamento são utilizadas juntas. Após a realização do agrupamento, aplica-se a análise discriminante e eventualmente realocam-se os objetos, que estão classificados incorretamente, aos grupos cuja probabilidade a posteriori de classificação for maior. Nova análise discriminante é então aplicada, até que não haja mais modificações e os grupos formados nesse último estágio são os grupos que se deve utilizar (KACHIGAN, 1991).

2.12 ANÁLISE MULTIVARIADA NA ÁREA FLORESTAL

O uso de técnicas multivariadas sofreu grande incremento com o advento dos computadores. Seu uso facilitou sobremaneira a execução deste tipo de análise, dada a complexidade de seu conteúdo computacional.

Na área florestal, a partir de 1980, o uso de técnicas multivariadas começou a ser utilizado. Longhi (1980) utilizou técnicas de agrupamento, análise discriminante e análise de variância multivariada em estudos de vegetação. Kanieski (2010) utilizou análise de correspondência canônica para correlacionar diversidade e características ambientais na Floresta Nacional de São Francisco de Paula, RS.

Floss (2011) destaca que existem muitos trabalhos em fitossociologia, mas poucos utilizam análises multivariadas, com correlação dos fatores ambientais e abundância de espécies. Em meio aos poucos trabalhos fitossociológicos, com o uso de Análise de Correspondência Canônica (CCA) entre abundância das espécies arbóreo-arbustivas, com fatores ambientais de solo e topografia na região sul, pode-se citar Narvaes (2004), Gomes (2005) e Budke et al. (2007). Os estudos ecológicos entre estrutura florestal relacionada com fatores topográficos e solos, utilizando análises multivariadas, são mais difundidos na região sudeste do país (BOTREL et al., 2002; FAGUNDES, 2003; MARTINS et al., 2003; SOUZA et al., 2003; OLIVEIRA-FILHO et al., 2004; CARVALHO et al., 2005) e centro-oeste (PINTO et al., 2005).

Além da CCA, existem outros métodos para a caracterização de comunidades

vegetais, os quais levam apenas em consideração a composição florística e os parâmetros fitossociológicos. Dentre os métodos, pode-se citar Análise de Agrupamento, utilizada por Longhi (1997) em comunidades florestais na sub-bacia hidrográfica do rio Passo Fundo (RS), e análise de espécies indicadoras – TWINSpan (TWo-way INdicator SPecies ANalysis) utilizada por Araujo (2004), na Floresta Estacional Decidual Aluvial, em Cachoeira do Sul (RS). Ambos os métodos de agrupamento são classificados como hierárquicos. O primeiro faz a separação pelo método aglomerativo, e o segundo, pelo método divisivo. Ambos os métodos utilizam como variáveis o número de indivíduos de cada espécie nas parcelas amostrais encontrada entre as matrizes originais.

Ribeiro (2004) utilizou técnicas de Análise de Agrupamentos, utilizando o programa TWINSpan para determinar os agrupamentos florísticos da vegetação arbórea e a Análise de Correspondência Canônica para a determinação dos fatores ecológicos limitantes para a ocorrência das espécies indicadoras dentro dos grupos. Araujo (2004) classificou espécies em grupos ecológicos e, com base nos dados de densidade, realizou uma análise de agrupamento utilizando o método TWINSpan. Os agrupamentos foram caracterizados em relação a sua estrutura horizontal.

3 MATERIAL E MÉTODOS

3.1 CARACTERIZAÇÃO DA ÁREA

Os dados utilizados para o desenvolvimento deste trabalho são oriundos de quatro parcelas permanentes de inventário florestal do “Programa Ecológico de Longa Duração” (PELD), localizado na Estação Experimental de São João do Triunfo, da Universidade Federal do Paraná. A área da estação experimental, que totaliza 32 hectares, 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 (Figura 2).

Histórico: a área foi adquirida em 1969 de colonos da região, tendo sofrido corte seletivo em algumas parcelas (LONGHI, 1980). A floresta, apesar de não ter sofrido corte raso, foi alterada pela exploração seletiva de araucárias, exploração de erva-mate, limpeza e um incêndio, em 1982, que afetou os talhões próximos à estrada, incluindo o talhão 1 e parte do talhão 6.

Clima: De acordo com o sistema de classificação climática de Köppen, a região enquadra-se no tipo Cfb, clima temperado propriamente dito, com temperatura média no mês mais frio abaixo de 18°C (mesotérmico), com verões frescos, temperatura média no mês mais quente abaixo de 22°C e sem estação seca definida (IAPAR, 1994).

Solo: Segundo o mapa esquemático de solos da Região Sul, a região de estudo pode ser classificada no grande grupo de solos dominantes em Latossolo Vermelho-Escuro, e no grupo de solos com menor extensão em Podzólico Vermelho-Escuro, Podzólico Vermelho-Amarelo, Latossolo Vermelho-Escuro Húmico e Cambissolo (IBGE, 1990). Com base no levantamento de reconhecimento dos solos do Estado do Paraná, a área de estudo em São João do Triunfo encontra-se solo Podzólico Vermelho-Amarelo Álico Tb A proeminente, textura média/argilosa, fase floresta subtropical perenifólia, relevo suave ondulado de vertentes curtas (EMBRAPA, 1984).

Tipologia: a vegetação da área é classificada como Floresta Ombrófila Mista Montana, de 400 a 1000m de altitude.

3.2 ANÁLISE DE DADOS

Os resultados obtidos obedeceram ao fluxograma apresentado na Figura 3, que tem por objetivo facilitar o entendimento de todos os procedimentos utilizados.

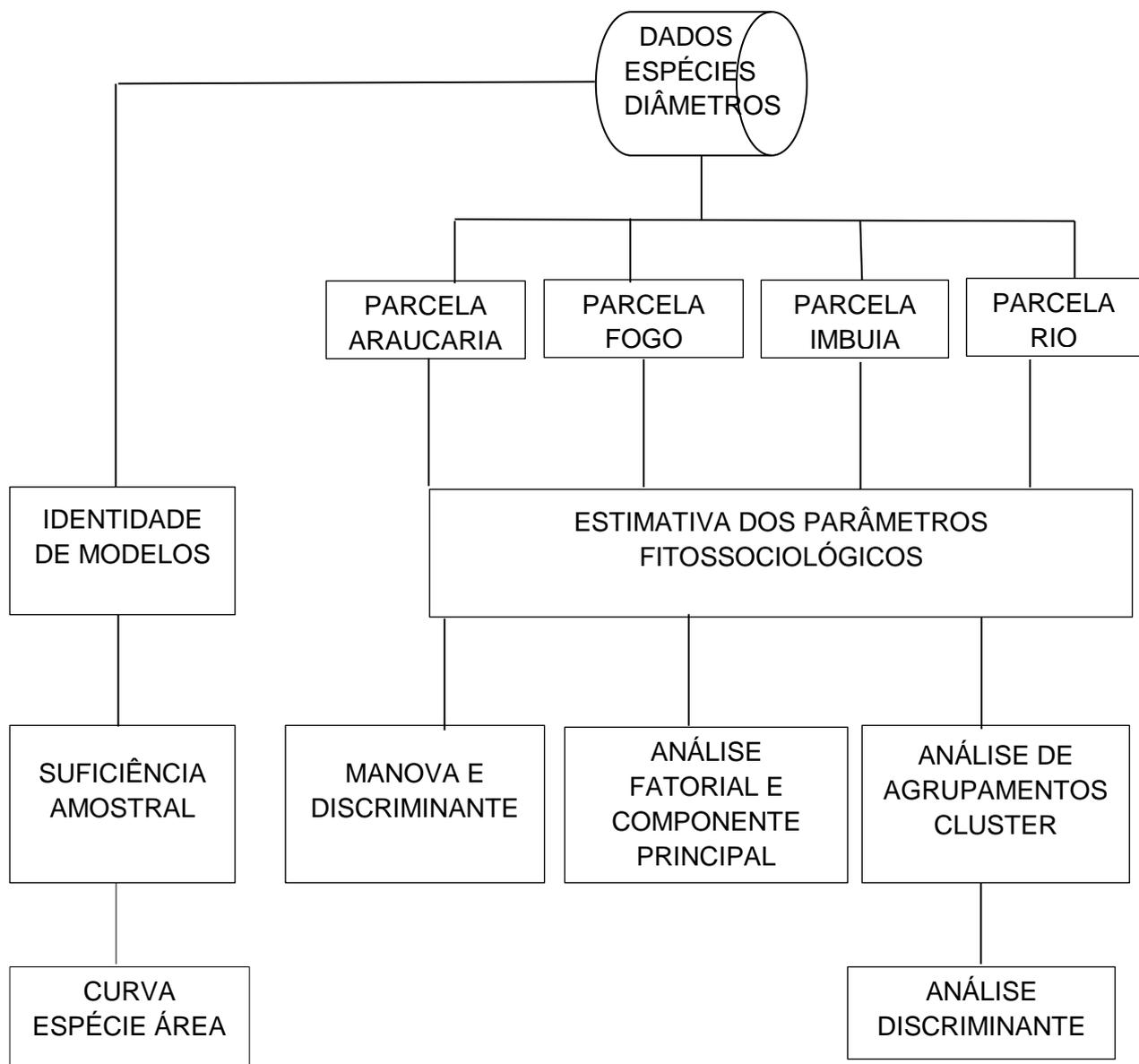


FIGURA 2 FLUXOGRAMA DOS PROCEDIMENTOS UTILIZADOS PARA A OBTENÇÃO DOS RESULTADOS.

3.2.1 AMOSTRAGEM

A área florestal foi dividida em parcelas de 100 x 100m (um ha). Os dados são coletados anualmente em quatro parcelas denominadas, somente com propósito de identificação das mesmas, como Araucária, Imbuia, Rio e Fogo (Figura 2).

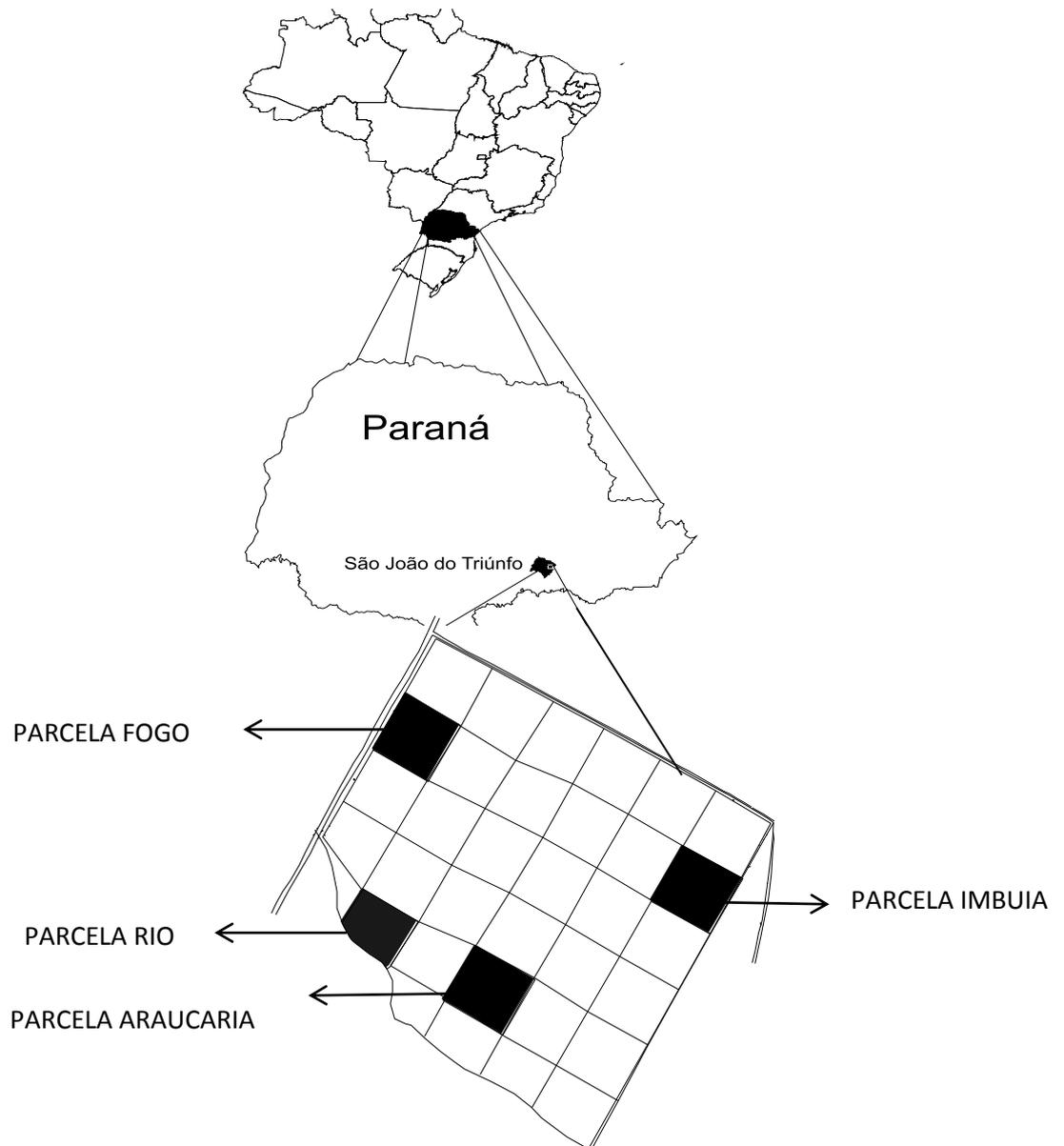


FIGURA 3. CROQUI DE LOCALIZAÇÃO DA FAZENDA EXPERIMENTAL DE SÃO JOÃO DO TRIUNFO, PR.

Cada parcela é dividida em sub-parcelas de 10 x 10 m, onde todas as árvores com diâmetros à altura do peito (DAP) igual ou superior a 10 cm são medidas com fita métrica ou trena. Cada árvore recebeu um número, sendo que as árvores que ingressam no ano de medição recebem um número acrescido a 1000 unidades. As medições usadas neste trabalho são as referentes aos anos de 1995, 2000, 2005 e 2010, nas quatro parcelas consideradas. Um sumário da caracterização das parcelas é apresentado na Tabela 3.

TABELA 3 CARACTERIZAÇÃO DAS PARCELAS ESTUDADAS, NOS DOIS ANOS CONSIDERADOS.

ANOS	PARCELAS	N ^a DE INDIVÍDUOS	N ^o DE ESPÉCIES	DAP MÉDIO (cm)
1995	ARAUCARIA	733	41	23,90
	RIO	472	38	10,30
	FOGO	616	46	22,20
	IMBUIA	609	56	26,60
2010	ARAUCARIA	734	42	23,40
	RIO	613	41	9,78
	FOGO	760	54	22,10
	IMBUIA	654	59	25,60

3.3 IDENTIDADE DE MODELOS

De posse dos dados das medições realizadas foi elaborada para cada uma das quatro parcelas, em cada um dos anos estudados, uma curva de acumulação espécie-área. Para cada uma das subunidades de cada parcela foram contadas quantas espécies novas ocorreram em cada uma delas, somando-se cumulativamente as espécies ocorrentes. Estes dados compõe o vetor “Y”, enquanto a matriz “X” é formada por colunas que representam as áreas da sub parcelas e as

mesmas ao quadrado, uma vez que o modelo é de 2ª ordem, e a matriz β é formada pelos coeficientes β_1 e β_2 .

Para cada uma das curvas geradas foi ajustado um modelo de regressão polinomial de 2º grau:

$$\hat{Y} = b_1 X + b_2 X^2$$

onde:

Y = Número de espécies acumulado;

X = Área amostrada acumulada.

A seguir as equações obtidas foram testadas quanto a sua identidade para verificar a possibilidade de se representar as curvas de acumulação de cada parcela por meio de uma única equação, dentro de cada um dos anos estudados.

Os dados foram analisados utilizando os programas MS EXCEL e MatLab (2009a). Para a realização dos testes de identidade foi desenvolvida uma rotina específica dentro do programa MatLab.

3.3 SUFICIÊNCIA AMOSTRAL

Tendo em vista que a amostragem da área em estudo já havia sido feita, quando foram medidas quatro parcelas, sendo três de um hectare cada e uma de meio hectare, divididas em sub-parcelas de 10 x 10 m, foi determinado se a intensidade amostral empregada, considerando as subunidades como unidade de amostra (n=350), foi suficiente nos quatro anos de medição. Para determinar a suficiência amostral foi considerada a variância do número de indivíduos por sub-parcela, para um erro de no máximo 10% do número médio, para os níveis de probabilidade de 5% e 10%.

A suficiência amostral também foi avaliada por meio das curvas do coletor ou curvas Espécie/Área ajustadas para cada uma das parcelas, onde por meio da derivada primeira destas equações, foram obtidos o número de espécies em cada sub-parcela correspondentes ao esforço amostral utilizando o número de sub-parcelas.

3.4 ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS

Com base nos dados coletados nos anos de 1995 e 2010, foram obtidas as estimativas dos parâmetros fitossociológicos Densidade (D), Frequência (F), Dominância (Do), Valor de Importância (VI) e Valor de Cobertura (VC), com o auxílio do programa FLOREXCEL (v.3.12.01). As estimativas dos parâmetros fitossociológicos foram feitas, em cada um dos anos, para cada parcela separadamente e para o conjunto delas.

De posse destas estimativas foi realizada, para cada conjunto de dados resultantes, uma análise fatorial e de componentes principais objetivando reduzir o número de variáveis e definir o número de fatores. Essas análises foram feitas para um grupo contendo todas as cinco variáveis (D, F, Do, VI e VC), para um grupo contendo as variáveis D, F e Do e para um grupo contendo somente as variáveis VI e VC.

A análise fatorial englobou o cálculo da matriz de correlação e respectivo nível de significância, o determinante desta matriz, que testa a multicolinearidade entre as variáveis, e os testes KMO, que testa a correlação das variáveis e o teste Bartlett, que verifica a esfericidade. A extração dos fatores foi feita pelo método de componentes principais, utilizando a matriz de correlação para os autovalores maiores que 1. A rotação dos fatores obedeceu o critério de máxima variância (VARIMAX). Os escores foram obtidos pelo método Anderson-Ruben, que assegura escores não correlacionados. O pacote estatístico utilizado para obtenção das análises foi o SPSS (*Statistical Package for the Social Sciences*) versão 19, para o Windows.

3.5 ANÁLISE DE VARIÂNCIA MULTIVARIADA - MANOVA

Considerando os anos de 1995 e 2010, as árvores foram agrupadas por espécies e verificado quantas vezes cada espécie ocorreu dentro de cada sub-parcela. A seguir foi realizada a análise fitossociológica, onde foram identificadas as 10 espécies que obtiveram maior valor de importância (VI), nas quatro parcelas, com

o objetivo de verificar se existe diferença entre elas (Araucária, Rio, Fogo e Imbuia). Uma vez selecionadas as 10 espécies mais importantes foi realizada a análise de variância multivariada (MANOVA) utilizando o método “stepwise” e considerando as espécies como variáveis dependentes no modelo.

Detectada a diferença entre as parcelas (Araucária, Rio, Fogo e Imbuia) foi realizada análise discriminante com o objetivo de verificar quais espécies contribuem e quanto contribuem na diferenciação entre as parcelas e se alguma parcela ou combinação linear delas discrimina as demais.

Foi realizado o teste de BOX, para verificar a igualdade das matrizes de covariâncias, e os testes Traço de Pillai, Lambda de Wilks, Traço de Hotelling e Maior raiz de Roy, para determinar se existe diferença entre os grupos. Na análise discriminante foi usada a matriz de covariância dentro dos grupos. O pacote estatístico utilizado para obtenção das análises foi o SPSS (*Statistical Package for the Social Sciences*) versão 19, para o Windows.

3.6 ANÁLISE DE AGRUPAMENTOS - CLUSTER

Com as matrizes de dados relativas aos anos de 1995 e 2010 foram selecionadas as 20 sub-parcelas que possuíam no mínimo 10 indivíduos cada, para cada uma das quatro parcelas estudadas, totalizando 80 sub-parcelas. Este procedimento foi feito uma vez que a variável considerada no agrupamento foi o número de indivíduos por espécie. Como existe um grande número de espécies presentes nas quatro parcelas estudadas, se fossem consideradas todas as 350 subunidades, o limite de variáveis do programa estatístico seria ultrapassado, inviabilizando a análise.

A partir destes dados foi realizada uma análise de agrupamentos hierárquica. A união entre as subunidades foi feita pelo método de Ward, que calcula a média da variável número de indivíduos por espécie em cada subunidade e para cada subunidade calcula a distância Euclidiana ao quadrado entre as médias encontradas. A definição dos grupos foi feita pelo traçado de uma linha paralela ao eixo horizontal no ponto médio da distância Euclidiana, conforme recomendam alguns autores. A análise foi realizada com o pacote estatístico SPSS (*Statistical Package for the Social Sciences*) versão 19, para o Windows.

3.7 ANÁLISE DISCRIMINANTE

Com base na análise de agrupamentos, que define os grupos em ambos os anos, foi realizada uma análise discriminante visando verificar as diferenças entre os grupos. Inicialmente foi realizada uma análise fitossociológica para cada um dos grupos definidos pela análise de agrupamentos (“cluster”), sendo selecionadas as dez espécies com maior Valor de Importância (VI). A escolha do VI como variável utilizada na análise discriminante foi devido ao fato que este valor leva em conta não só o número de árvores existentes, mas também a área coberta, a proporção entre o número de unidades amostrais onde a espécie ocorre e o número total de unidades amostrais.

Desta forma as variáveis dependentes foram as dez espécies com maior VI e a variável independente os grupos definidos. Foi utilizado o procedimento “stepwise”, que permite a entrada das variáveis dependentes passo a passo, ao invés de todas ao mesmo tempo. Como a entrada das variáveis independentes foi “stepwise”, o método de seleção foi o lambda de Wilks, que considera como critério o valor da estatística F para o teste de diferença entre os centroides dos grupos.

Os coeficientes da função discriminante foram obtidos de forma padronizada. A classificação foi feita considerando as probabilidades para grupos de mesmo tamanho, utilizando a matriz variância-covariância dentro dos grupos.

Foi utilizado o programa estatístico SPSS (*Statistical Package for the Social Sciences*) versão 19 para o Windows, para realização das análises.

4 RESULTADOS

4.1 IDENTIDADE DE MODELOS

Para cada ano avaliado, consideraram-se quatro modelos de regressão ($H = 4$), com quatro submodelos polinomiais do segundo grau ($r = 4$) de dois parâmetros ($p = 2$), assumindo-se que os modelos passam pela origem, ou seja, $\beta_0 = 0$, partindo do pressuposto que ao se trabalhar com coordenadas cartesianas, a variável X é a área e no ponto $(0,0)$ tem-se área nula e conseqüentemente nenhuma espécie.

Na Tabela 4 são apresentadas as equações de regressão ajustadas para o ano de 1995, para cada uma das quatro áreas em estudo, bem como o coeficiente de determinação (R^2), o erro padrão (s_{XY}) que expressa a dispersão entre os valores observados e estimados pela regressão e o erro padrão da estimativa em percentual ($s_{XY\%}$), que expressa o quanto em termos médios os valores observados variam em relação aos valores estimados.

TABELA 4 EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE ÁREA PARA CADA PARCELA, PARA O ANO DE 1995.

Identificação da parcela	Equação ajustada	R^2	S_{YX}	$S_{YX\%}$
Araucária	$\hat{Y}_1 = 0,8972 X - 0,0053X^2$	83,87%	1,92	6,75
Rio	$\hat{Y}_2 = 1,5584 X - 0,0170X^2$	94,14%	0,86	3,33
Fogo	$\hat{Y}_3 = 1,2387 X - 0,0082X^2$	94,33%	1,47	4,12
Imbuia	$\hat{Y}_4 = 1,3746X - 0,0087 X^2$	91,38%	2,17	5,31

A Figura 4 mostra as equações ajustadas para todas as parcelas consideradas, no ano de 1995, e respectivos coeficientes de determinação (R^2).

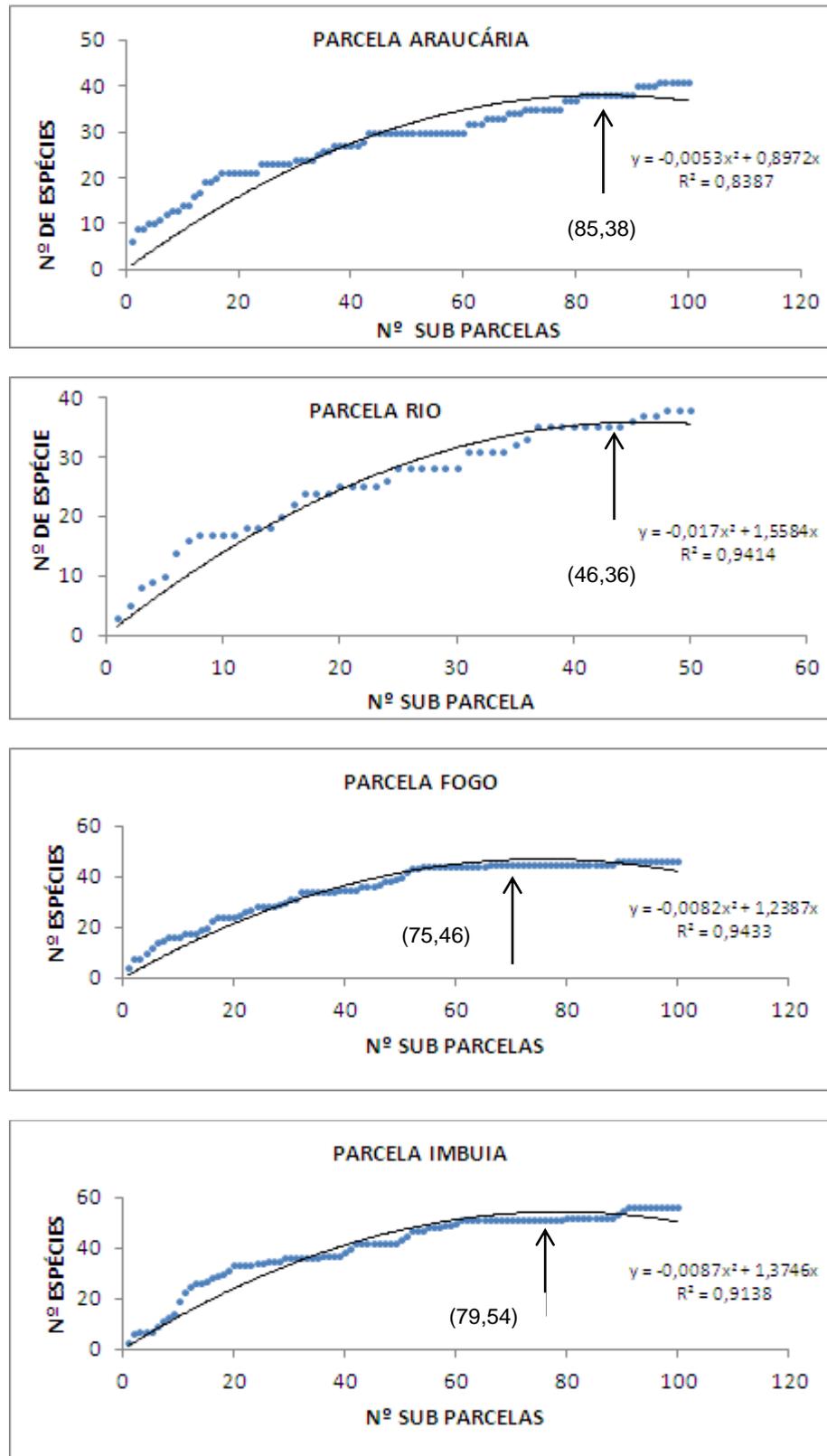


FIGURA 4 CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 1995.

A Tabela 5 contém os resultados do teste para verificar se as quatro equações de regressão podem ser representadas por uma única equação. Observa-se que as equações diferem estatisticamente entre si, ou seja, rejeita-se a hipótese de nulidade para o ano de 1995, tendo em vista que o valor de $F_t = 2,12$ é menor do que o valor de F_o , logo as equações não podem ser representadas por uma única equação.

TABELA 5 QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 1995.

Fonte de variação	G.L.	Soma de	Quadrado Médio	F_o
Coeficiente modelo completo (β)	8	447810	55976,25	
Coeficiente modelo reduzido (θ)	2	438350	219175	
Redução (H_0)	6	9460,6	1576,77	140,22*
Resíduo	342	3845,6	11,245	
Total	350	451652		

*significativo ao nível de 5% de probabilidade pelo teste F.

Na Tabela 6 estão apresentadas as equações de regressão ajustadas para o ano de 2000, para cada uma das quatro parcelas em estudo, com os respectivos coeficientes de determinação, o erro padrão da estimativa e erro padrão da estimativa em percentagem.

TABELA 6 EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE-ÁREA PARA CADA PARCELA, PARA O ANO DE 2000.

Identificação da	Equação ajustada	R^2	S_{YX}	$S_{YX\%}$
Araucária	$\hat{Y}_1 = 0,8147X - 0,0043X^2$	87,60	1,77	6,43
Rio	$\hat{Y}_2 = 1,5271 X - 0,0151X^2$	94,03	0,93	3,49
Fogo	$\hat{Y}_3 = 1,2270 X - 0,0077X^2$	94,21	1,58	4,29
Imbuia	$\hat{Y}_4 = 1,4129 X - 0,0996 X^2$	87,44	2,34	5,87

A Figura 5 mostra as equações ajustadas para o ano de 2000, para cada uma das parcelas consideradas e respectivos coeficientes de determinação (R^2).

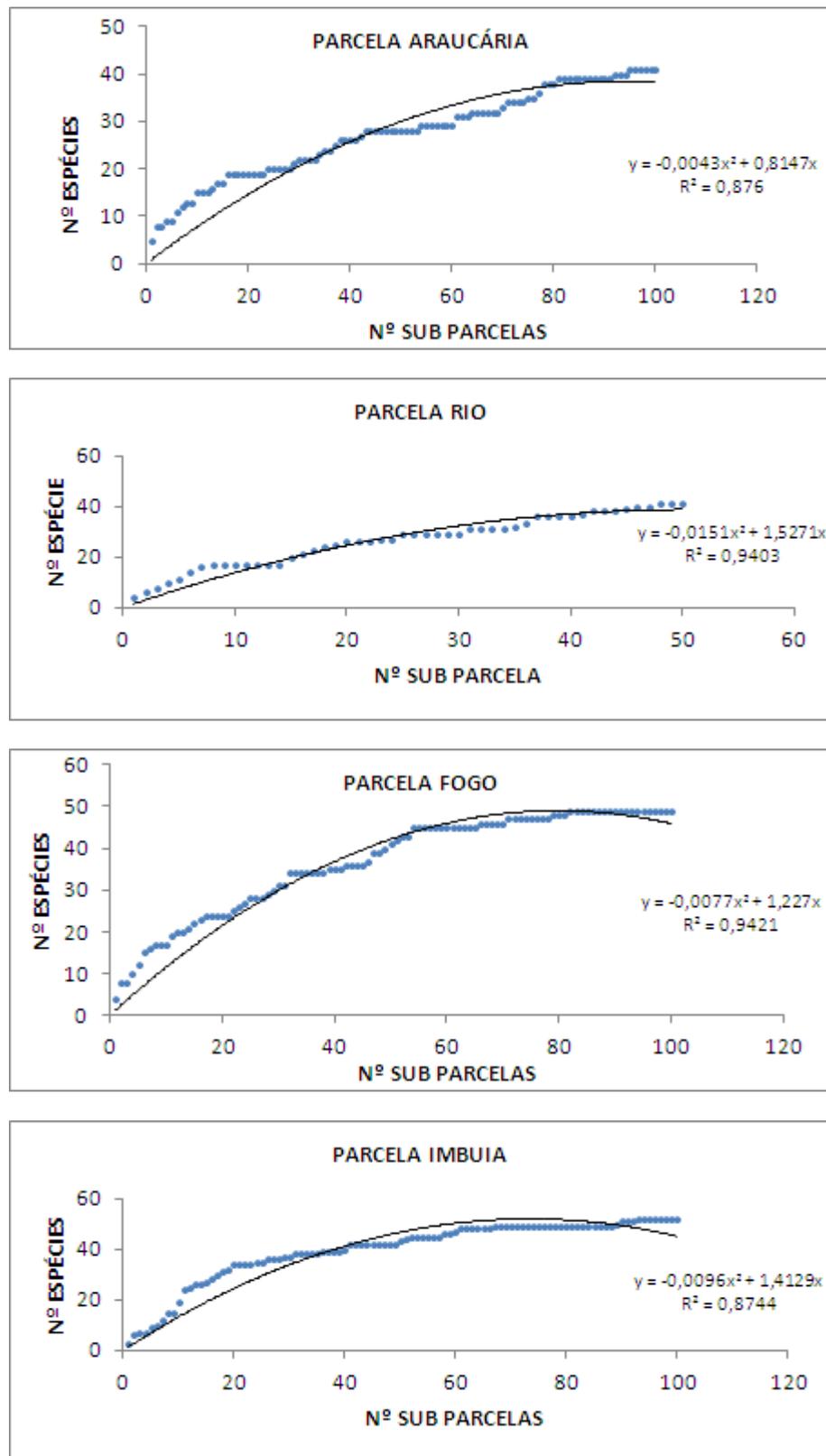


FIGURA 5 CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 2000.

A Tabela 7 apresenta os resultados do teste de identidade para verificar se as quatro equações de regressão podem ser representadas por uma única equação para os dados do ano de 2000. Pelos resultados obtidos as equações ajustadas diferem estatisticamente, ou seja, rejeita-se a hipótese de nulidade de que as quatro equações ajustadas podem ser substituídas por uma única equação, para o referido ano.

TABELA 7 QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 2000.

Fonte de variação	G.L.	Soma de Quadrado	Quadrado Médio	F _o
Coeficiente modelo completo (β)	8	444930	55616,25	
Coeficiente modelo reduzido (θ)	2	435160	217580	
Redução (H ₀)	6	9767,4	1627,9	135,89
Resíduo	342	4097,1	11,979	
Total	350	449029		

* significativo ao nível de 5% de probabilidade pelo teste F.

A Tabela 8 contém as equações de regressão ajustadas para a curva espécie/área para o ano de 2005, para cada uma das quatro parcelas em estudo, bem como os respectivos coeficientes de determinação, erros padrão e erros padrão das estimativas em percentagem.

TABELA 8 EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE-ÁREA PARA CADA PARCELA, PARA O ANO DE 2005.

Identificação da parcela	Equação ajustada	R ²	S _{YX}	S _{YX%}
Araucária	$\hat{Y}_1 = 0,8409 X - 0,0043 X^2$	88,95	1,79	6,20
Rio	$\hat{Y}_2 = 1,5586 X - 0,0152 X^2$	95,03	0,88	3,18
Fogo	$\hat{Y}_3 = 1,2237 X - 0,0075 X^2$	94,59	1,57	4,21
Imbuia	$\hat{Y}_4 = 1,4381 X - 0,0096 X^2$	91,21	2,08	5,04

A Figura 6 apresenta os resultados do ajuste da equação para a curva espécie-área referentes as parcelas analisadas no ano de 2005, mostrando a equação ajustada e o respectivo coeficiente de determinação.

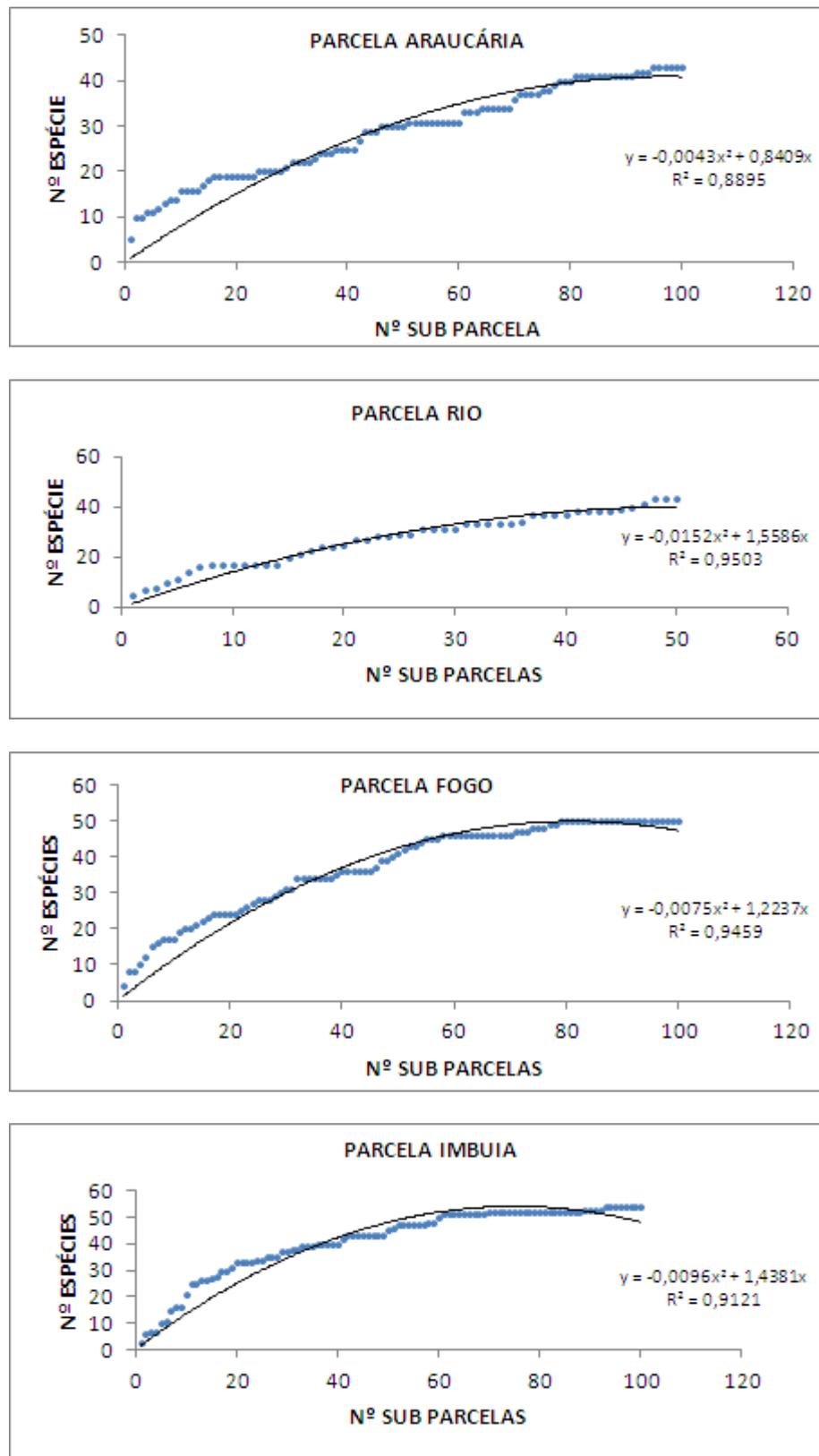


FIGURA 6 CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 2005.

Na Tabela 9 são apresentados os resultados do teste de identidade para as equações ajustadas para cada uma das parcelas, para o ano de 2005. A análise de variância revelou que para o ano em questão, as quatro equações de regressão referentes a cada uma das parcelas não podem ser representadas por uma única equação, uma vez que houve diferença estatística entre elas.

TABELA 9 QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 2005.

Fonte de variação	G.L.	Soma de Quadrado	Quadrado Médio	F _o
Coefficiente modelo completo (β)	8	474030	59253,75	
Coefficiente modelo reduzido (θ)	2	464420	232,21	
Redução (H ₀)	6	9609,4	1601,57	147,88*
Resíduo	342	3704	10,83	
Total	350	477730		

* significativo ao nível de 5% de probabilidade pelo teste F.

Na Tabela 10 são apresentadas as equações de regressão ajustadas para o ano de 2010, para cada uma das quatro áreas em estudo, bem com os respectivos coeficientes de determinação, erros padrão da estimativa e erros padrão das estimativas em porcentagem.

TABELA 10 EQUAÇÕES DE REGRESSÃO AJUSTADAS PARA A CURVA ESPÉCIE-ÁREA PARA CADA PARCELA, PARA O ANO DE 2010.

Identificação da parcela	Equação ajustada	R ²	S _{YX}	S _{YX%}
Araucária	$\hat{Y}_1 = 0,9088X - 0,0057 X^2$	81,21	2,10	7,23
Rio	$\hat{Y}_2 = 1,6338 X - 0,0170 X^2$	96,47	0,74	2,68
Fogo	$\hat{Y}_3 = 1,2422 X - 0,0074 X^2$	92,21	1,96	5,05
Imbuia	$\hat{Y}_4 = 1,4189 X - 0,0087 X^2$	94,02	1,96	4,54

A Tabela 11 apresenta os resultados do teste de identidade para verificar se as quatro equações de regressão podem ser representadas por uma única equação. A análise de variância revelou que as equações diferem estatisticamente, ou seja, rejeita-se a hipótese de nulidade, para o ano de 2010, e não é possível utilizar somente uma equação para representar a curva espécie-área para a parcela em questão.

TABELA 11 QUADRO DA ANÁLISE DE VARIÂNCIA RELATIVA AO TESTE DE IDENTIDADE DAS EQUAÇÕES AJUSTADAS PARA AS QUATRO PARCELAS, PARA O ANO DE 2010.

Fonte de variação	G.L.	Soma de Quadrado	Quadrado Médio	F _o
Coeficiente modelo completo (β)	8	509310	63663,75	
Coeficiente modelo reduzido (θ)	2	496750	248375	
Redução (H ₀)	6	12569	2094,83	164,98*
Resíduo	342	4342,4	12,697	
Total	350	513657		

* significativo ao nível de 5% de probabilidade pelo teste F.

Por meio dos gráficos apresentados é possível observar que o polinômio de 2º grau apresentou estatísticas de ajuste boas, contrariando Schilling e Batista (2008), Gomide et al. (2005), Camargo (1997), Lima (1997), Vasconcelos (1992) e Volpato (1994). Esses autores afirmam que seria mais indicado ajustar modelos não lineares, utilizando a técnica de regressão linear de platô, auxiliada pelo teorema do limite central.

A Figura 7 apresenta as equações de regressão ajustadas e respectivos coeficientes de determinação para os dados referentes às quatro parcelas avaliadas, para o ano de 2010.

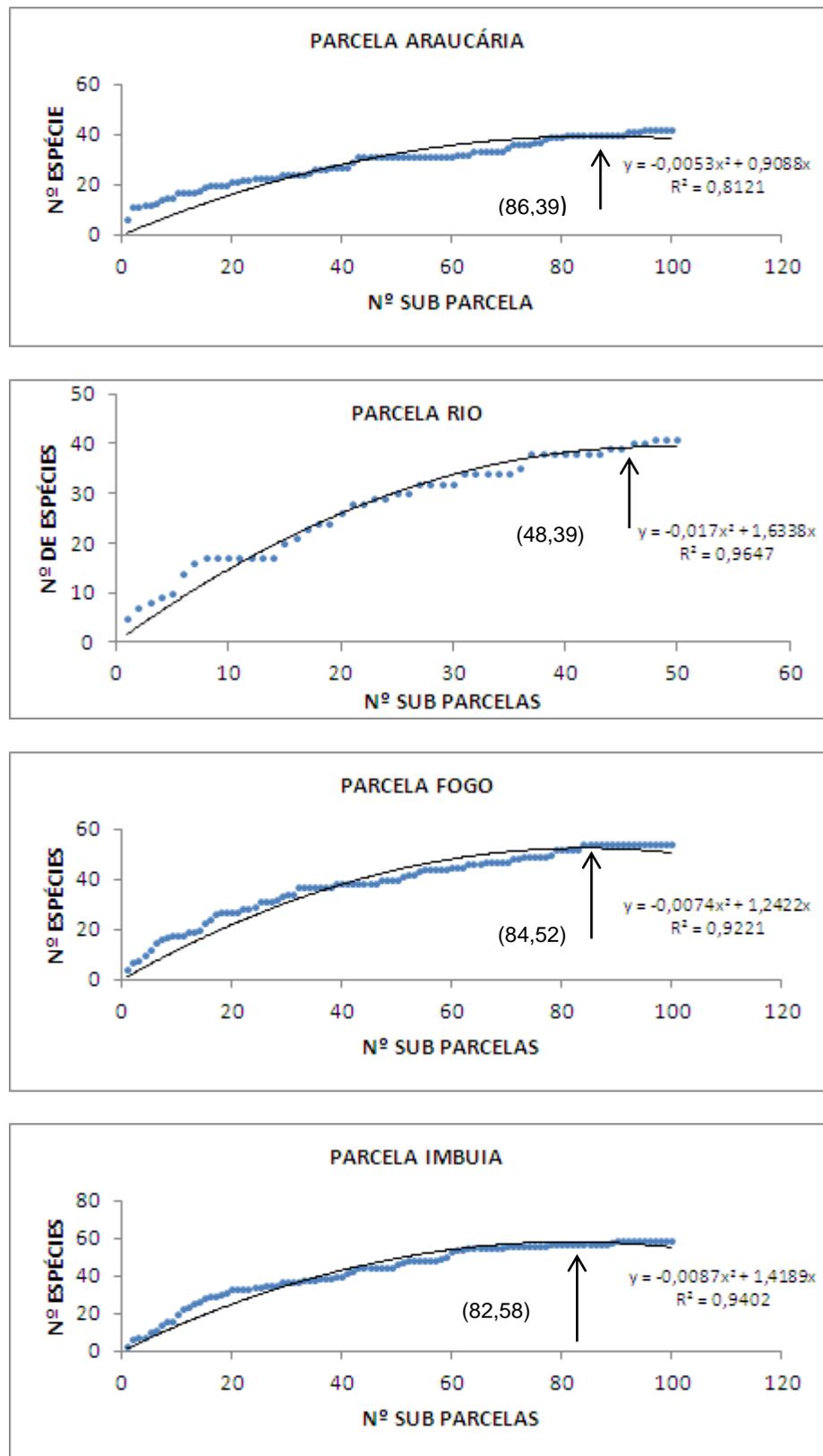


FIGURA 7 CURVAS ESPÉCIE-ÁREA, EQUAÇÕES AJUSTADAS E COEFICIENTES DE DETERMINAÇÃO PARA AS PARCELAS AVALIADAS, PARA ANO DE 2010.

4.2 SUFICIÊNCIA AMOSTRAL

Com base nos levantamentos efetuados pelo Programa Ecológico de Longa Duração (PELD) realizado na área em estudo, foram verificadas aproximadamente a existência de 77 espécies pertencentes a 30 famílias botânicas, sendo que alguns exemplares não foram identificados. A dificuldade de identificação dessas espécies ocorreu pela impossibilidade da obtenção de material botânico fértil, dificultando uma análise mais detalhada das mesmas. Durigan (1999) trabalhando na mesma área em 1997 encontrou 64 espécies, o que indica ter havido um ingresso de aproximadamente 20% no número de espécies.

Na Tabela 12, são apresentadas as famílias, o nome científico e o nome comum de cada espécie encontrada nas quatro parcelas medidas.

TABELA 12 FAMÍLIAS, NOME CIENTÍFICO E NOME COMUM DAS ESPÉCIES ENCONTRADAS NAS QUATRO PARCELAS TOMADAS NA ÁREA DE SÃO JOÃO DO TRIUNFO-PR.

Familia	Nome Científico	Nome comum
ANACARDIACEAE	<i>Lithrea brasiliensis</i> Marchand	Bugreiro
ANACARDIACEAE	<i>Schinus terebinthifolius</i> Raddi	Aroeira
ANNONACEAE	<i>Annona</i> sp. Linn	Ariticum
ANNONACEAE	NI	NI
AQUIFOLIACEAE	<i>Ilex brevicuspis</i> Reissek	Caúna
AQUIFOLIACEAE	<i>Ilex dumosa</i> Reissek	Caúna-miúda
AQUIFOLIACEAE	<i>Ilex paraguariensis</i> A. St-Hil.	Erva-mate
AQUIFOLIACEAE	<i>Ilex theezans</i> Martex Reissek	Orelha-de-mico
ARAUCARIACEAE	<i>Araucaria angustifolia</i> (Bertol.) Kuntze	Pinheiro-do-Paraná
ARECACEAE	<i>Syagrus romanzoffiana</i> (Cham.) Glassman	Jerivá
ASTERACEAE	<i>Gochnatia polymorpha</i> (Less.) Cabrera	Cambará
ASTERACEAE	<i>Piptocarpha axillaris</i> (Less.) Baker	Vassourão-graúdo
ASTERACEAE	<i>Piptocarpha angustifolia</i> Dusén ex. Malme	Vassourão-branco
ASTERACEAE	<i>Vernonanthura discolor</i> (Spreng.) H. Rob.	Vassourão-preto
BIGNONIACEAE	<i>Jacaranda puberula</i> Cham.	Caroba
CANELACEAE	<i>Cinnamodendron dinisii</i> Schwacke	Pimenteira
CELASTRACEAE	<i>Maytenus ilicifolia</i> Mart. ex. Reissek	Espinheira-santa
CLETHRACEAE	<i>Clethra scabra</i> ex. Pers.	Carne-de-vaca
ELAEOCARPACEAE	<i>Sloanea lasiocoma</i> K. Schum.	Sapopema
ERYTHROXYLACEAE	<i>Erythroxylum deciduum</i> A. St-HILL.A. St.-Hil.	Cocão
EUPHORBIACEAE	<i>Sapium glandulosum</i> (L) Morong	Leiteiro
EUPHORBIACEAE	<i>Sebastiania commersoniana</i> (Bail.) L. B.Sm &Down	Branquilha
FABACEAE	<i>Fabaceae</i>	NI
FABACEAE	<i>Inga</i> sp.	Ingá
FABACEAE	<i>Lonchocarpus</i> sp.	NI
FABACEAE	<i>Machaerium</i> sp.	NI
FABACEAE	<i>Machaerium stipitatum</i> Vogel	Sapuvinha
FABACEAE	<i>Mimosa scabrella</i> Benth.	Bracatinga
LAMIACEAE	<i>Vitex megapotamica</i> (Spreng.) Moldenke	Tarumã
LAURACEAE	<i>Cinnamomum sellowianum</i> (Nees &Mart.) Kosterm.	Canela-raposa

TABELA 12 FAMÍLIAS, NOME CIENTÍFICO E NOME COMUM DAS ESPÉCIES ENCONTRADAS NAS QUATRO PARCELAS TOMADAS NA ÁREA DE SÃO JOÃO DO TRIUNFO-PR (Continuação).

LAURACEAE	<i>Cinnamomum cimoenum</i> (Nees & Mart.) Kosterm.	Pau-alho
LAURACEAE	<i>Nectandra grandiflora</i> Nees	Canela-amarela
LAURACEAE	<i>Nectandra</i> sp.	Canela-imbuia
LAURACEAE	<i>Nectandra megapotamica</i> (Spreng.) Mez	Canela-preta
LAURACEAE	<i>Ocotea corymbosa</i> Mez	Canela-de-porco
LAURACEAE	<i>Ocotea puberula</i> (Rich.) Nees	Canela-guaicá
LAURACEAE	<i>Ocotea pulchella</i> (Nees & Mart.) Mez	Canela-lageana
LAURACEAE	<i>Ocotea</i> sp	Canela
LAURACEAE	NI	NI
LAURACEAE	<i>Persea</i> sp.	Pau-andrade
MALVACEAE	<i>Luehea divaricata</i> Mart. & Zucc.	Açoita-cavalo
MELIACEAE	<i>Cedrela fissilis</i> Vell.	Cedro
MELIACEAE	<i>Trichilia elegans</i> A. Juss.	Pau-de-ervilha
MYRTACEAE	<i>Blepharocalyx salicifolius</i> (Kunth) O. Berg	Murteira
MYRTACEAE	<i>Myrcia hartwegiana</i> O. Berg.	Guamirim
MYRTACEAE	<i>Campomanesia xanthocarpa</i> (Mart.) O. Berg.	Guabiroma
MYRTACEAE	<i>Campomanesia guazumifolia</i> (Cambess.) O. Berg	Sete-capotes
MYRTACEAE	<i>Eugenia involucrata</i> DC.	Cerejeira
MYRTACEAE	<i>Eugenia</i> sp.	NI
MYRTACEAE	<i>Eugenia uniflora</i> L.	Pitanga
MYRTACEAE	<i>Myrceugenia miersiana</i>	Cambrum-piloso
MYRTACEAE	<i>Myrceugenia myrcioides</i> (Cambess.) O. Berg.	Guamirim-branco
MYRTACEAE	<i>Myrcia Splendens</i> (SW.) DC.	Guamirim-chorão
MYRTACEAE	<i>Eugenia hyemalis</i> Cambess. NR(EFC):9506	Guamirim-vermelho
MYRTACEAE	<i>Myrcia</i> sp. 1	Guamirim II
MYRTACEAE	Myrtaceae 2	Guamirim-graúdo
MYRTACEAE	NI	NI
MYRTACEAE	<i>Eugenia speciosa</i> Cambess	Araça
NI	NI	Embira-do-mato
NI	NI	Liana-lenhosa
NI	NI	NI
PRIMULACEAE	<i>Myrsine coriacea</i> (Sw.) R. Br.ex. Roem. & Schult.	Capororoquinha
PRIMULACEAE	<i>Myrsine umbellata</i> Mart.	Capororocão
RHAMNACEAE	<i>Hovenia dulcis</i> Thunb.	Uva-do-japão
ROSACEAE	<i>Eriobotrya japonica</i> (Thunb.) Lindl.	Ameixeira
ROSACEAE	<i>Prunus brasiliensis</i> (Charm. & Schltld.) D. Dietr.	Pessegueiro-bravo
RUBIACEAE	NI	NI
RUTACEAE	<i>Zanthoxylum kleinii</i> (R.S. Cowan) Waterman	Juvevê-branco
RUTACEAE	<i>Zanthoxylum rhoifolium</i> Lain.	Mamica de cadela
SALICACEAE	<i>Banara tomentosa</i> Clos	NI
SALICACEAE	<i>Casearia decandra</i> Jacq.	Guaçatunga
SALICACEAE	<i>Casearia obliqua</i> Spreng.	Guaçatunga-vermelha
SALICACEAE	<i>Xylosma pseudosalzmanii</i> Sleumer	Sucará
SAPINDACEAE	<i>Allophylus edulis</i> (A. St-Hil, et al.) Hieron. Ex Nierderl.	Chal-chal
SAPINDACEAE	<i>Allophylus guaraniticus</i> (A. St-Hil.) Radlk.	Vacum
SAPINDACEAE	<i>Cupania vernalis</i> Cambess.	Cuvatã
SAPINDACEAE	<i>Matayba elaeagnoides</i> Radlk.	Miguel-pintado
SOLANACEAE	<i>Solanum mauritianum</i> Scop.	Tabaco
SOLANACEAE	<i>Solanum sanctaecatharinae</i> Dunal	Fumeiro
SOLANACEAE	<i>Solanum</i> sp.	NI
STYRACACEAE	<i>Styrax leprosus</i> Hook. & Arn.	Pau-de-remo
SYMPLOCACEAE	<i>Symplocos celastria</i> Mart.	Pau-de-anta
WINTERACEAE	<i>Drimys brasiliensis</i> Miers	Cataia

Tendo em vista que a amostragem da área em estudo já havia sido feita, quando foram medidas quatro parcelas, sendo que três delas com um hectare cada e uma com meio hectare, divididas em sub-parcelas de 10 x 10 m, foi determinado se a intensidade amostral empregada, considerando as sub-parcelas como unidade de amostra (n=350), foi suficiente nos quatro anos de medição.

Dessa forma, tendo como base $N = 3200$ sub-parcelas possíveis e 350 sub-amostras efetuadas calculou-se o fator de correção que resultou na definição da população como sendo finita.

A intensidade de amostragem da população finita foi feita considerando-se a variância do número de indivíduos por sub-parcela, para um erro de no máximo 10% do número médio, para os níveis de probabilidade de 5% e 10%.

Para obter o número de unidades amostrais necessários para estimar os parâmetros da população com a precisão e confiabilidade fixadas, foi feito o recálculo da intensidade amostral, utilizando o valor de (t) com base no novo número de graus de liberdade encontrado, até que se encontrasse um valor constante para a intensidade amostral. A Tabela 13 mostra os resultados obtidos na determinação da intensidade amostral, relativas aos anos de 1995 e 2010.

TABELA 13 RESULTADOS DAS ESTATÍSTICAS E INTENSIDADES AMOSTRAIS ENCONTRADAS PARA A VARIÁVEL NÚMERO DE ESPÉCIES POR SUBUNIDADE, PARA AS MEDIÇÕES EFETUADAS NOS ANOS DE 1995 E 2010.

ESTATÍSTICAS		
ÁREA TOTAL	32 ha	
ÁREA DA AMOSTRA	0,01 ha	
Nº DE PARCELAS	350	
N	3200	
1-f	0,8906	
ESTATÍSTICAS	1995	2010
MEDIA	33,66	37,74
DESVIO PADRÃO	12,58	13,79
VARIÂNCIA	157,69	190,38
COEFICIENTE DE VARIAÇÃO	37,30	38,59
$E(5\%) = LE \cdot \bar{X}$	1,68	1,79
$E(10\%) = LE \cdot \bar{X}$	3,37	3,57
VARIÂNCIA DA MÉDIA	0,40	0,48
ERRO PADRÃO	0,63	0,69
ERRO DE AMOSTRAGEM RELATIVO	3,70	3,83
DIMENSIONAMENTO DA AMOSTRA (5%)	201,75	215,05
DIMENSIONAMENTO DA AMOSTRA (10%)	144,57	154,29

A suficiência amostral também foi avaliada por meio da curva do coletor ou curva Espécie-Área, onde foram obtidos o número de espécies em cada sub-parcela (grafado no eixo das ordenadas) contra o esforço amostral utilizando o número de sub-parcelas (grafado no eixo das abscissas).

Nas Figuras 4 e 7 são apresentadas as curvas de acumulação de espécies para cada uma das quatro amostras tomadas na área, para os anos de 1995 e 2010.

A área estudada é um fragmento de Floresta Ombrófila Mista onde foram encontradas 30 famílias e 77 espécies, nas quatro parcelas medidas, para um total de 2.761 indivíduos com diâmetro à altura do peito (DAP) maior ou igual a 10 cm. As famílias mais representativas da área são Mirtácea seguida de Laurácea e Fabácea. A composição florística apresenta grande heterogeneidade na família Mirtácea, com um total de 16 espécies, enquanto que a família Laurácea apresentou 14 espécies e a família Fabácea 6 espécies.

A intensidade amostral, considerando o número de espécies por sub-parcelas amostral como variável de estudo, indicou que a população era finita uma vez que o fator de correção para populações foi 0,89, menor do que o limite (0,98). Considerando um nível de probabilidade de 5%, para um limite de erro de 10% da média do número de indivíduos, foram encontradas como intensidade amostral 202 e 215 sub-parcelas para os anos de 1995 e 2010 respectivamente. Para um nível de probabilidade de 10%, e o mesmo limite de erro, o tamanho de amostra recomendado foi de 145 e 154 para cada um dos anos considerados.

Na comparação entre as intensidades calculadas e as efetuadas em ambos os anos, pode-se concluir que a amostragem efetuada foi maior do que a sugerida, para ambos os níveis de probabilidade. Neste sentido, a intensidade de 350 sub-parcelas, para cada um dos anos estudados (1995 e 2010) estima os parâmetros da população com a precisão e a confiabilidade fixadas e garante a determinação da “composição característica” da associação ou comunidade vegetal (SCHILLING e BATISTA, 2008).

Ao se analisar as curvas espécie/área, para cada uma das amostras, para o ano de 1995 (Figura 4) observa-se que todas as curvas apresentaram uma inclinação suave, atingindo um ponto de máxima, obtido pela derivada da equação ajustada, de 85 sub-parcelas, o que corresponde a 38 espécies, para a parcela denominada Araucária. Para a parcela Rio o ponto de máxima foi 46 sub-parcelas e 36 espécies. Para as parcelas Fogo e Imbuia, os pontos de máxima foram 75 e 79 sub-parcelas, para 46 e 54 espécies, respectivamente.

No que diz respeito ao ano de 2010 (Figura 7), foram encontrados pontos de máximo de 86, 48, 84 e 82 sub-parcelas, com 39, 39, 52 e 58 espécies para cada uma das quatro parcelas (Araucária, Rio, Fogo e Imbuia) avaliadas. Em todas as situações e para os anos de 1995 e 2010, as intensidades amostrais efetuadas foram superiores as intensidades calculadas. A medição de sub-parcelas adicionais acrescentou um número pequeno de espécies ao número existente, significando que a assíntota foi atingida. Dada a natureza do modelo ajustado, os pontos de máximo de todas curvas coincide com o ponto de assíntota, permitindo que se verifique matematicamente onde se iniciam, como pode ser visto pela indicação desses pontos nas Figuras 4 e 7.

A construção da curva do coletor para determinar o tamanho ótimo da amostra apresenta limitações segundo SCHILLING e BATISTA (2008). Os resultados encontrados no cálculo da intensidade mostraram que existe similaridade entre os valores recomendados pelos dois processos, ou seja, o ponto onde se inicia o patamar de estabilidade e o valor do dimensionamento da amostra, o que contraria o exposto pelos autores citados.

Para a área de estudo, parcelas de um hectare, subdivididas em sub-parcelas de 10 x 10m mostraram-se mais do que suficientes para a realização de estudos fitossociológicos. Deve-se salientar que, segundo Schilling e Batista (2008), o tamanho ótimo de uma amostra estatística é encontrado comparando a precisão desejada com o custo de se obter a amostra. O tamanho ótimo será aquele que maximize a precisão e minimize o custo, fator esse que não foi considerado no presente uma vez que as amostras avaliadas fazem parte do Programa Ecológico de Longa Duração (PELD) cujo objetivo principal é avaliar a dinâmica do fragmento de floresta ao longo dos anos, e para tanto utilizou quatro parcelas de um hectare, divididas em 100 sub-parcelas de 0,1 ha cada.

4.3 ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS

Com base nas estimativas obtidas para parâmetros Densidade (D), Frequência (F), Dominância (Do), Valor de Importância (VI) e Valor de Cobertura (VC), foram realizadas as análises Fatorial e de Componentes Principais.

A Tabela 14 apresenta os resultados dos testes estatísticos referentes à análise fatorial e de componentes principais para o grupo composto pelas cinco estimativas dos parâmetros fitossociológicos. Observa-se que no ano de 1995, as parcelas Araucária, Imbuia, Rio e Fogo, não apresentaram resultados para os testes KMO e Bartlett, pois o determinante da matriz de correlação R foi igual à zero, indicando que há multicolinearidade entre os dados. Quando considerado as quatro parcelas juntas o valor do teste KMO é maior do que 0,5, o que indica que o tamanho da amostra é razoável e que não há necessidade de se incluir mais variáveis, enquanto que o teste de Bartlett mostra que a matriz de correlação não é uma matriz identidade, indicando que existem relacionamentos entre as variáveis sendo analisadas.

Quanto ao número de componentes (Tabela 14) foi selecionado somente um componente, retirado pelo critério de Kaiser, onde os autovalores devem ser maiores do que um, para o conjunto de amostras, explicando 77,3% da variância total. Quando se consideram as parcelas de modo isolado, as variâncias explicadas, também por somente um componente, foram superiores a 89,6%.

Outro item a ser considerado na Tabela 14, para o ano de 1995, são as cargas de cada variável em cada componente. As cargas foram obtidas pelo método Varimax, cuja função foi maximizar a dispersão dentro dos componentes. Os resultados obtidos foram muito próximos de um ou 100%, com exceção do conjunto de amostras que apresentou valores entre 0,85 e 0,90.

Com relação ao ano de 2010, tanto para as parcelas isoladas como para as parcelas agrupadas, o valor do teste de KMO foi maior que 0,5 significando que o tamanho da amostra é razoável e ainda, que não há necessidade de incluir mais variáveis. Quanto à multicolinearidade entre os dados, o teste indica que ela não existe, pois os determinantes das matrizes de correlação são diferentes de zero, indicando que há correlação entre os dados, para ambos os grupos. O teste de Bartlett indica que em todas as parcelas e no conjunto delas, a matriz de correlação não é uma matriz identidade. Os resultados do ano de 2010, também mostram que foi selecionado somente um componente, com autovalores maiores que um. Assim, para todas as parcelas e para o conjunto delas, o componente explica mais do que 87,2% da variância total.

Quanto às cargas de cada variável, para o ano de 2010, em cada componente pode-se observar que estão todas muito próximas de um ou 100%.

TABELA 14 ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS PARA AS VARIÁVEIS D, F, Do, VI e VC, NOS ANOS DE 1995 E 2010.

ANO	ÁREA	KMO	BART.	Nº DE COMP.	VARIÂN 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
	IMBUA	-	-	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
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
	IMBUA	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 15 apresenta os resultados dos testes estatísticos referentes à análise fatorial e de componentes principais para o grupo composto pelas estimativas dos parâmetros fitossociológicos D, F e Do.

Os valores do teste de KMO revelaram valores todos maiores do que 0,5 para as parcelas isoladas e para as parcelas agrupadas, da mesma forma que ocorreu quando foram utilizadas cinco variáveis na análise. O teste de Bartlett mostrou valores altamente significativos ($p < 0,05$) para todas as parcelas nos dois anos. Somente um componente foi extraído e as variâncias explicadas maiores do que 81,7%, com as cargas apresentando valores muito próximos.

Na Tabela 16 encontram-se os resultados das análises levando em consideração somente as variáveis, Valor de Importância e Valor de Cobertura. Pode-se verificar que os valores obtidos para o teste de KMO estão no limite do recomendado 0,50 para as cinco parcelas nos dois anos analisados. O teste de Bartlett é significativo em todas as situações estudadas ($p < 0,05$). Foi selecionado somente um componente principal para cada uma das parcelas isoladas e agrupadas, com variâncias explicadas acima de 99%, e com valores para as cargas acima de 0,99 em todas as situações.

TABELA 15 ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS PARA AS VARIÁVEIS D, F e Do, NOS ANOS DE 1995 E 2010.

ANO	ÁREA	KMO	BARTLETT	Nº COMP.	VARIA. EXPL. %	CARGAS		
						D	Do	F
1995	TODAS	0,686	675,94*	1	91,12	0,981	0,921	0,961
	ARAUC	0,514	216,73*	1	93,02	0,994	0,965	0,933
	IMBUA	0,657	184,44*	1	84,72	0,958	0,830	0,967
	RIO	0,526	185,09*	1	93,31	0,996	0,951	0,950
	FOGO	0,740	205,97*	1	94,50	0,984	0,950	0,982
2010	TODAS	0,625	423,45*	1	92,63	0,988	0,925	0,973
	ARAUC	0,544	217,13*	1	91,98	0,990	0,917	0,969
	IMBUA	0,613	192,48*	1	81,68	0,953	0,779	0,967
	RIO	0,484	188,96*	1	99,39	0,994	0,910	0,946
	FOGO	0,675	258,42*	1	92,11	0,976	0,918	0,984

* Teste de Bartlett significativo a 5% de probabilidade.

TABELA 16 - ANÁLISE FATORIAL E DE COMPONENTES PRINCIPAIS PARA AS VARIÁVEIS VI e VC, NOS ANOS DE 1995 E 2010.

ANO	ÁREA	KMO	BARTLETT	Nº de COMP.	VARIA. EXPL %	CARGAS	
						VI	VC
1995	TODAS	0,50	669,99*	1	99,36	0,997	0,997
	ARAUC	0,50	176,99*	1	99,75	0,999	0,999
	IMBUA	0,50	214,86*	1	99,51	0,998	0,998
	RIO	0,50	173,51*	1	99,78	0,999	0,999
	FOGO	0,50	220,57*	1	99,82	0,999	0,999
2010	TODAS	0,50	387,40*	1	99,78	0,999	0,999
	ARAUC	0,50	180,04*	1	99,74	0,999	0,999
	IMBUA	0,50	212,75*	1	99,42	0,997	0,997
	RIO	0,50	168,43*	1	99,68	0,998	0,998
	FOGO	0,50	254,28*	1	99,82	0,999	0,999

* Teste de Bartlett significativo a 5% de probabilidade.

Quando se trabalha com as cinco variáveis fitossociológicas, observa-se que a carga de cada variável é bastante alta e com valores próximos. Considerando que a carga deve indicar qual variável é a mais importante, no presente, mesmo com uma pequena diferença, destaca-se o Valor de Importância, no entanto as demais variáveis não podem ser descartadas, pois apresentam valores significativos.

Quando consideradas as variáveis D, F e Do, também se observa uma proximidade dos valores das cargas, e neste caso destaca-se a variável Densidade. Porém não se recomenda descartar as demais devido à significância das cargas. Para as variáveis VI e VC, observa-se que não há diferença entre as cargas, e não se recomenda descartar nenhuma delas.

Como o Valor de Importância é composto pela soma da densidade, da frequência e da dominância, pode-se fazer uso dele, pois ele reúne três variáveis, como utilizado por (LONGHI 1997). Este cálculo permite comparar os pesos ecológicos das espécies dentro do respectivo tipo florestal. Valores muito próximos obtidos para o valor de importância das espécies mais características podem ser uma indicação da igualdade ou, pelo menos, semelhança das comunidades quanto à composição, estrutura, sítio e dinâmica (LAMPRECHT, 1990).

4.4 ANÁLISE DE VARIÂNCIA MULTIVARIADA - MANOVA

As espécies selecionadas com maior valor de importância, em ordem alfabética, para os anos de 1995 e 2010, são apresentadas na Tabela 17. Estas espécies foram utilizadas na composição dos quatro grupos para testar se há diferença entre eles. As espécies mais importantes encontradas no presente coincidem com os resultados obtidos por Longhi (1997), Durigan (1999) e Pizzato (1999), trabalhando no mesmo fragmento florestal.

TABELA 17 ESPÉCIES SELECIONADAS PELO VALOR DE IMPORTÂNCIA, PARA OS ANOS DE 1995 E 2010.

ESPÉCIE	1995 – VI	2010 – VI
<i>Araucaria angustifolia</i>	84,90	75,63
<i>Blepharocalyx salicifolius</i>	-	9,08
<i>Cinnamodendron dinisii</i>	11,10	7,58
<i>Casearia decandra</i>	9,88	-
<i>Cinamomum amoenum</i>	-	8,73
<i>Ilex brevicuspis</i>	8,05	-
<i>Ilex paraguariensis</i>	18,57	13,01
<i>Llithraea brasiliensis</i>	6,98	-
<i>Matayba elaeagnoides</i>	11,31	14,61
<i>Nectandra grandiflora</i>	20,54	26,33
<i>Ocotea porosa</i>	10,65	9,94
<i>Ocotea puberula</i>	11,16	7,57
<i>Prunus brasiliensis</i>	-	7,55

Na Tabela 18 encontram-se os resultados do teste de Box, que verifica a igualdade entre as matrizes de variâncias e covariâncias. Este teste deve ser considerado sempre que os grupos tiverem tamanhos diferentes, como no presente,

onde três dos grupos têm 100 sub-parcelas e um deles têm 50. O resultado do teste de Box, para os dois anos, foi significativo ao nível de 5% de probabilidade, rejeitando a hipótese de nulidade, ou seja, as matrizes não são iguais, o que é requerido pela análise. Segundo Field (2009) o teste de Box precisa de fato ser examinado quando os tamanhos amostrais diferem e não se deve acreditar nele quando a normalidade multivariada não puder ser assumida.

TABELA 18 RESULTADOS TESTE DE BOX E TESTE F PARA 1995 E 2010

TESTES	1995	2010
BOX	2793,88	2075,86
F	15,98	11,87
SIGNIFICÂNCIA	0,00	0,00

* significativo a 5% de probabilidade.

Na Tabela 19 encontram-se os resultados do teste de poder e robustez para os dados relativos aos anos de 1995 e 2010. São citadas estatísticas para o intercepto do modelo, porque a MANOVA pode ser caracterizada como um modelo de regressão, e para a variável que representa o agrupamento. São apresentadas quatro estatísticas multivariadas e a transformação dos valores para a razão F, com sua significância. Essas estatísticas determinam se há ou não diferença entre os grupos. Pode-se observar pelos resultados apresentados que tanto para o intercepto como para os grupos, os quatro testes realizados são significativos ao nível de 5% de probabilidade.

TABELA 19 RESULTADOS DOS TESTES PARA O INTERCEPTO E PARA OS GRUPOS NOS ANOS 1995 E 2010.

TESTE	1995			2010		
	VALOR	F	SIG	VALOR	F	SIG
INTERCEPTO						
TRAÇO DE PILLAI	0,71	80,45	0,000*	0,78	116,67	0,000*
LAMBDA DE WILKS	0,29	80,45	0,000*	0,22	116,67	0,000*
TRAÇO DE HOTELLING	2,39	80,45	0,000*	3,46	116,67	0,000*
MAIOR RAIZ DE ROY	3,39	80,45	0,000*	3,46	116,67	0,000*
GRUPO						
TRAÇO DE PILLAI	0,74	11,12	0,000*	0,79	11,87	0,000*
LAMBDA DE WILKS	0,40	12,02	0,000*	0,38	12,89	0,000*
TRAÇO DE HOTELLING	1,16	12,93	0,000*	1,25	13,94	0,000*
MAIOR RAIZ DE ROY	0,81	27,30	0,000*	0,88	29,75	0,000*

* significativo a 5% de probabilidade.

Os resultados obtidos na análise de variância multivariada dos dados, referentes ao ano de 1995, são apresentados na Tabela 20. O teste multivariado leva em consideração as correlações entre as variáveis dependentes, podendo assim detectar com maior precisão se há diferença entre os grupos. A Tabela 20 mostra que todas as espécies são significativas ao nível de 5% de probabilidade e que existe diferença entre os grupos.

TABELA 20 RESULTADOS DA ANÁLISE DE VARIÂNCIA MULTIVARIADA RELATIVA AOS DADOS DE 1995.

ESPÉCIES SELECIONADAS	SOMA DE QUADRADO	QUADRADO MÉDIO	F	Sig.
<i>Araucaria angustifolia</i>	248,819	82,940	38,244	0,000*
<i>Nectandra grandiflora</i>	46,070	15,357	10,787	0,000*
<i>Ilex paraguariensis</i>	67,190	22,397	36,075	0,000*
<i>Matayba elaeagnoides</i>	13,077	4,359	6,290	0,000*
<i>Ocotea puberula</i>	14,614	4,871	10,433	0,000*
<i>Cinnamodendron dinisii</i>	9,627	3,209	4,824	0,003*
<i>Ocotea porosa</i>	3,007	1,002	5,860	0,001*
<i>Casearia decandra</i>	11,754	3,918	13,830	0,000*
<i>Ilex brevicuspis</i>	6,667	2,222	10,019	0,000*
<i>Lithrea brasiliensis</i>	8,127	2,709	5,311	0,001*

* significativo a 5% de probabilidade.

Na Tabela 21 encontram-se os resultados da análise de variância multivariada dos dados referentes ao ano de 2010, onde observa-se que todas as espécies são significativas, ao nível 5% de probabilidade, com exceção da espécie *Cinnamodendron dinisii*, que não difere das demais.

A análise de variância multivariada mostrou que existe relacionamento entre as variáveis, mas há necessidade de se investigar a natureza do relacionamento. A análise discriminante é a melhor maneira de se obter este resultado.

Assim foi realizada uma análise discriminante para os anos de 1995 e 2010. A Tabela 22 mostra os autovalores encontrados. Esses autovalores foram convertidos em porcentagem da variância das combinações lineares sendo que, para 1995 a primeira combinação linear foi responsável por 69,7% da variância. A segunda combinação linear foi responsável por 19,4% e a terceira por 10,9%. Com relação ao ano de 2010, observa-se que a primeira combinação linear foi responsável por 70,4% da variância, a segunda por 16,9%, e a terceira por 12,7% da variância.

TABELA 21 RESULTADOS DA ANÁLISE DE VARIÂNCIA MULTIVARIADA RELATIVA AOS DADOS DE 2010.

ESPÉCIES SELECIONADAS	SOMA DE QUADRADO	QUADRADO MÉDIO	F	Sig.
<i>Araucaria angustifolia</i>	252,590	84,197	41,625	0,000*
<i>Blepharocalyx salicifolius</i>	26,150	8,717	18,022	0,000*
<i>Cinnamodendron dinisii</i>	2,120	,707	1,748	0,157 ^{NS}
<i>Cinnamomum amoenum</i>	22,077	7,359	13,686	0,000*
<i>Ilex paraguariensis</i>	85,077	28,359	30,255	0,000*
<i>Matayba eleagnoides</i>	15,057	5,019	7,026	0,000*
<i>Nectandra grandiflora</i>	113,980	37,993	18,902	0,000*
<i>Ocotea porosa</i>	3,227	1,076	6,153	0,000*
<i>Ocotea puberula</i>	9,144	3,048	9,099	0,000*
<i>Prunus brasiliensis</i>	6,449	2,150	5,449	0,001*

* significativo a 5% de probabilidade; ns-não significativo.

TABELA 22 RESULTADOS DOS AUTOVALORES OBTIDOS E PORCENTUAL DA VARIÂNCIA EXPLICADA, PARA OS ANOS DE 1995 E 2010.

FUNÇÃO	1995		2010	
	AUTOVALOR	% DA VARIÂNCIA	AUTOVALOR	% DA VARIÂNCIA
1ª	0,805	69,7	0,878	70,4
2ª	0,225	19,4	0,211	16,9
3ª	0,126	10,9	0,158	12,7

A Tabela 23 apresenta os resultados do teste Lambda de Wilks, que gerou o mesmo valor obtido na MANOVA (0,40) para o ano de 1995 e (0,38) para o ano de 2010, ambos ao nível 5% de probabilidade. As três combinações lineares são significativas, logo a diferença entre os grupos detectada pela MANOVA pode ser explicada em três dimensões.

TABELA 23 RESULTADOS DO TESTE LAMBDA DE WILKS PARA OS ANOS DE 1995 E 2010

TESTE DE FUNÇÃO	1995			2010		
	LAMBDA	DF	SIG.	LAMBDA	DF	SIG.
1 ATÉ 2	0,40	30	0,00*	0,38	30	0,000*
2 ATÉ 3	0,73	18	0,00*	0,71	18	0,000*
3	0,89	8	0,00*	0,86	8	0,000*

*significativos a 5% probabilidade.

Os coeficientes padronizados da função discriminante são apresentados na Tabela 24. Observa-se para o ano de 1995, que na primeira combinação linear a

maior contribuição é da espécie *Ilex paraguariensis* seguida da *Araucaria angustifolia*, da *Matayba elaeagnoides* e *Nectandra grandiflora* (0,569; 0,567, -0,278 e 0,139 respectivamente). Na segunda combinação linear as maiores contribuições são da *Casearia decandra*, *Lithrea brasiliensis* e *Ocotea porosa* (0,607; -0,335 e 0,311, respectivamente) e na terceira combinação linear as espécies que mais contribuem são *Ocotea puberula*, *Ilex brevicuspis* e *Cinnamodendron dinisii* (0,748; -0,399; -0,289 e 0,245).

TABELA 24 COEFICIENTES PADRONIZADOS DA FUNÇÃO DISCRIMINANTE PARA O ANO DE 1995.

ESPÉCIES	FUNÇÃO		
	1ª CL	2ª CL	3ªCL
<i>Araucaria angustifolia</i>	0,567	-0,178	0,404
<i>Necatandra grandiflora</i>	0,139	-0,049	-0,399
<i>Ilex paraguariensis</i>	0,569	0,462	-0,042
<i>Matayba elaeagnoides</i>	-0,278	-0,204	-0,115
<i>Ocotea puberula</i>	-0,077	0,179	0,748
<i>Cinnamodendron dinisii</i>	0,241	0,046	0,245
<i>Ocotea porosa</i>	-0,166	0,311	0,072
<i>Casearia decandra</i>	-0,271	0,607	0,039
<i>Ilex brevicuspis</i>	0,264	0,168	-0,289
<i>Lithraea brasiliensis</i>	0,084	-0,335	0,254

A Tabela 25 apresenta os coeficientes padronizados da função discriminante, para o ano de 2010. Na primeira combinação linear a maior contribuição é da *Araucaria augustifolia*, seguida de *Cinnamomum amoenum* e *Ocotea porosa* (0,623; 0,400; -0,212 respectivamente). Na segunda combinação linear as maiores contribuições são da *Ilex paraguariensis*, *Blepharocalyx salicifolius* e *Matayba elaeagnoides* (0,628; 0,527 e -0,337, respectivamente) e na terceira combinação linear as espécies que mais contribuem são *Ocotea puberula*, *Nectandra grandiflora*, e *Prunus brasiliensis* (0,678; 0,386; -0,346).

TABELA 25 COEFICIENTES PADRONIZADOS DA FUNÇÃO DISCRIMINANTE PARA O ANO DE 2010.

ESPÉCIES	FUNÇÃO		
	1ª CL	2ª CL	3ª CL
<i>Araucaria angustifolia</i>	0,623	-0,011	0,118
<i>Blepharocalyx salicifolius</i>	-0,235	0,527	0,241
<i>Cinnamodendron dinisii</i>	0,142	0,079	0,115
<i>Cinnamomum amoenum</i>	0,400	-0,261	0,315
<i>Ilex paraquariensis</i>	0,410	0,628	-0,041
<i>Matayba eleagnoides</i>	-0,102	-0,337	0,281
<i>Nectandra grandiflora</i>	0,311	-0,168	-0,346
<i>Ocotea porosa</i>	-0,212	0,198	0,046
<i>Ocotea puberula</i>	0,107	0,070	0,678
<i>Prunus brasiliensis</i>	0,202	0,105	0,386

Os centróides são as médias dos escores das combinações lineares de cada grupo e o sinal do centróide (positivo ou negativo) permite verificar onde as diferenças entre eles são maiores dentro da função e desta forma indica quem discrimina.

A Tabela 26, referente ao ano de 1995, mostra que na primeira combinação linear as parcelas Imbuia e Araucária discriminam as outras duas parcelas, tendo em vista que as diferenças entre os centróides são as maiores para essas parcelas. Na segunda combinação linear a parcela Fogo discrimina as demais, e conseqüentemente, na terceira combinação linear a parcela Rio é a que têm a maior diferença entre os centróides. Com relação ao ano de 2010, as parcelas Rio e Imbuia discriminam as outras parcelas na primeira combinação linear. Para a segunda combinação linear a parcela que apresenta a maior diferença entre os centróides é a parcela Fogo e na terceira combinação linear a que melhor discrimina é a parcela a Araucária.

Nas Figuras 8 e 9 observa-se o relacionamento entre as combinações lineares e as parcelas. Esses gráficos apresentam os escores das combinações lineares para cada espécie, agrupados de acordo com a sub-parcela a qual a espécie pertence. Além disso, os centróides das parcelas são ressaltados como o escore médio da combinação linear para cada parcela.

TABELA 26 RESULTADOS DOS CENTRÓIDES DAS FUNÇÕES DISCRIMINANTES PARA OS ANOS 1995 E 2010.

GRUPO	1995			2010		
	1ª CL	2ª CL	3ª CL	1ª CL	2ª CL	3ª CL
ARAUCARIA	1,083	0,374	-0,221	0,708	0,447	-0,387
RIO	0,669	-0,296	0,791	1,354	-0,084	0,776
FOGO	-0,270	-0,646	-0,257	-0,119	-0,673	-0,219
IMBUIA	-1,147	0,419	0,082	-1,267	0,268	0,218

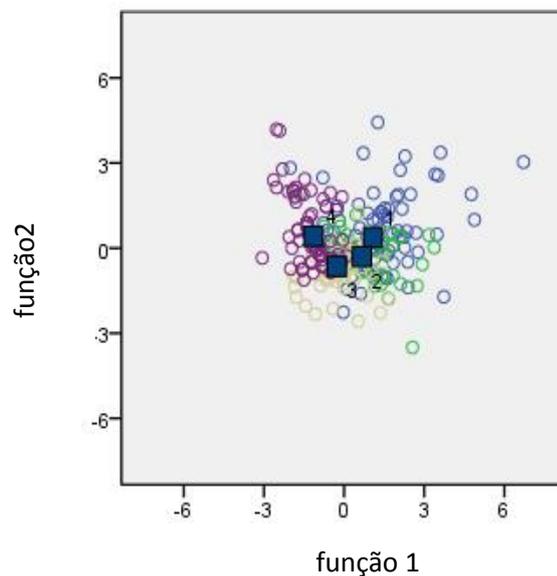


FIGURA 8 FUNÇÃO DISCRIMINANTE E CENTRÓIDES DAS PARCELAS PARA O ANO DE 1995.

A análise de variância multivariada efetuada mostrou que há diferença significativa entre as parcelas denominadas Araucária, Rio, Fogo e Imbuia, para as 10 espécies selecionadas. Considerando as espécies que mais contribuíram para a formação dos grupos nos anos de 1995 e 2010, pode-se inferir que a área em estudo trata-se de um fragmento de Floresta Ombrófila Mista (floresta com araucária), conforme descrito pelo Sistema Brasileiro de Classificação da Vegetação (IBGE 2012), baseado no sistema de classificação fisionômico-ecológico desenvolvido pelo projeto RADAMBRASIL.

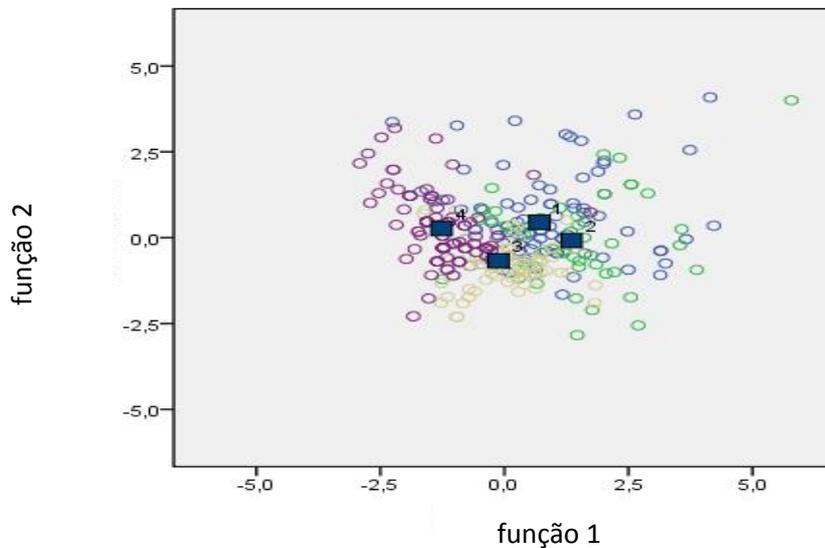


FIGURA 9 FUNÇÃO DISCRIMINANTE E CENTRÓIDES DAS PARCELAS PARA O ANO DE 2010.

4.5 ANÁLISE DE AGRUPAMENTOS – CLUSTER

A partir dos dados das 80 sub-parcelas selecionadas com base no número de indivíduos encontrados por espécie, nos anos de 1995 e 2010, foi realizada uma análise de agrupamentos, visando encontrar grupos formados pela similaridade entre as sub-parcelas. Como método de agrupamento hierárquico optou-se pelo de Ward. A sequência dos grupos obtidos foi apresentada por meio de um Dendograma.

A definição dos grupos no Dendograma foi feita pelo traçado de uma linha paralela ao eixo horizontal, no ponto médio da distância Euclidiana. Essa linha é denominada Linha de Fenon (JOHNSON, WICHERN, 1982). Essa separação também poderia ser sido feita segundo diferentes critérios, subjetivos e pessoais.

Pode-se observar nas Figuras 10 e 11, referentes aos anos de 1995 e 2010, que a linha de Fenon interceptou três ramos de ligações nos agrupamentos formados, classificando o conjunto de 80 sub-parcelas e 83 espécies, em três grupos distintos.

O primeiro grupo, no ano de 1995, foi formado agrupando 37 sub-parcelas, sendo 18 da parcela Araucária, 12 da parcela Rio, seis da parcela Fogo e uma da parcela Imbuia. O segundo grupo foi formado por cinco sub-parcelas, sendo quatro da parcela Rio e uma da parcela Fogo. No terceiro grupo, formado por 38 sub-parcelas, sendo 19 da parcela Imbuia, 13 da parcela Fogo, quatro da parcela Rio e duas da parcela Araucária.

No ano de 2010, o primeiro grupo foi formado por 8 sub-parcelas, sendo seis da parcela Araucária e duas da parcela Rio. O segundo grupo, formado por 29 sub-parcelas, tem 11 da parcela Araucária, dez da parcela Rio, seis da parcela Fogo e duas da parcela Imbuia. O terceiro grupo, formado por 43 sub-parcelas, sendo 18 da parcela Imbuia, 14 da parcela Fogo, oito da parcela Rio e três da parcela Araucária.

Analisando o número de sub-parcelas que compõem cada grupo, em ambos os anos, é possível afirmar que o primeiro grupo é formado pelas sub-parcelas das parcelas Araucária e Rio, o que comprova a proximidade física destas parcelas (Figura 2). No segundo grupo a presença de sub-parcelas da parcela Rio é marcante no ano de 1995, enquanto que no ano de 2010, as parcelas Rio e Araucária são as mais predominantes. O terceiro grupo, em ambos os anos é formado por sub-parcelas originárias das parcelas Fogo e Imbuia, que são fisicamente distantes das duas outras parcelas. Cumpre salientar que a influência da parcela Fogo é mais marcante no ano de 2010 do que no ano de 1995, devido ao ingresso de 144 indivíduos de 8 espécies em 2010 (Tabela 3). A parcela Fogo foi a que apresentou maior incremento no número de indivíduos e espécies dentre as quatro parcelas estudadas, no período avaliado, sendo esta provavelmente a principal resposta da parcela ao incêndio ocorrido (Tabela 3).

Cabe ressaltar que foram testadas outras medidas de similaridade, além da distância euclidiana, tais como a distância de Manhattan e a distância de Chebychev. Os resultados obtidos não foram satisfatórios, tendo em vista que resultaram em muitos agrupamentos.

Quanto aos métodos aglomerativos, foram testados o método do vizinho mais próximo, do vizinho mais distante, da ligação por média, pelo centróide e pelo método de Ward. O método de Ward apresentou melhores resultados conforme preconiza Romesburg (1984), pois o número de elementos em cada grupo resultante não é igual e por esta razão foi o método selecionado.

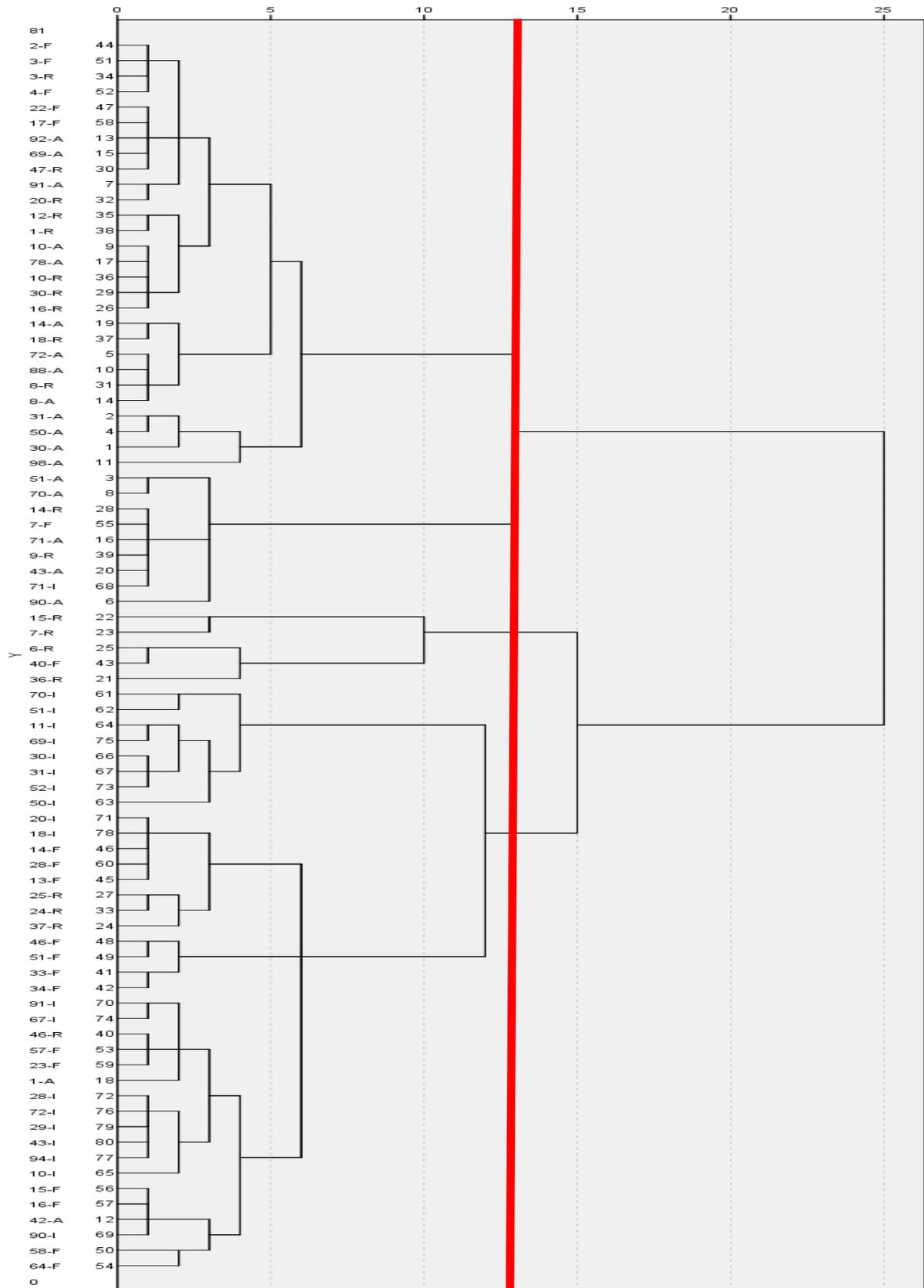


FIGURA 10 DENDROGRAMA RESULTANTE DA ANÁLISE DE AGRUPAMENTOS RELATIVO AO ANO DE 1995.

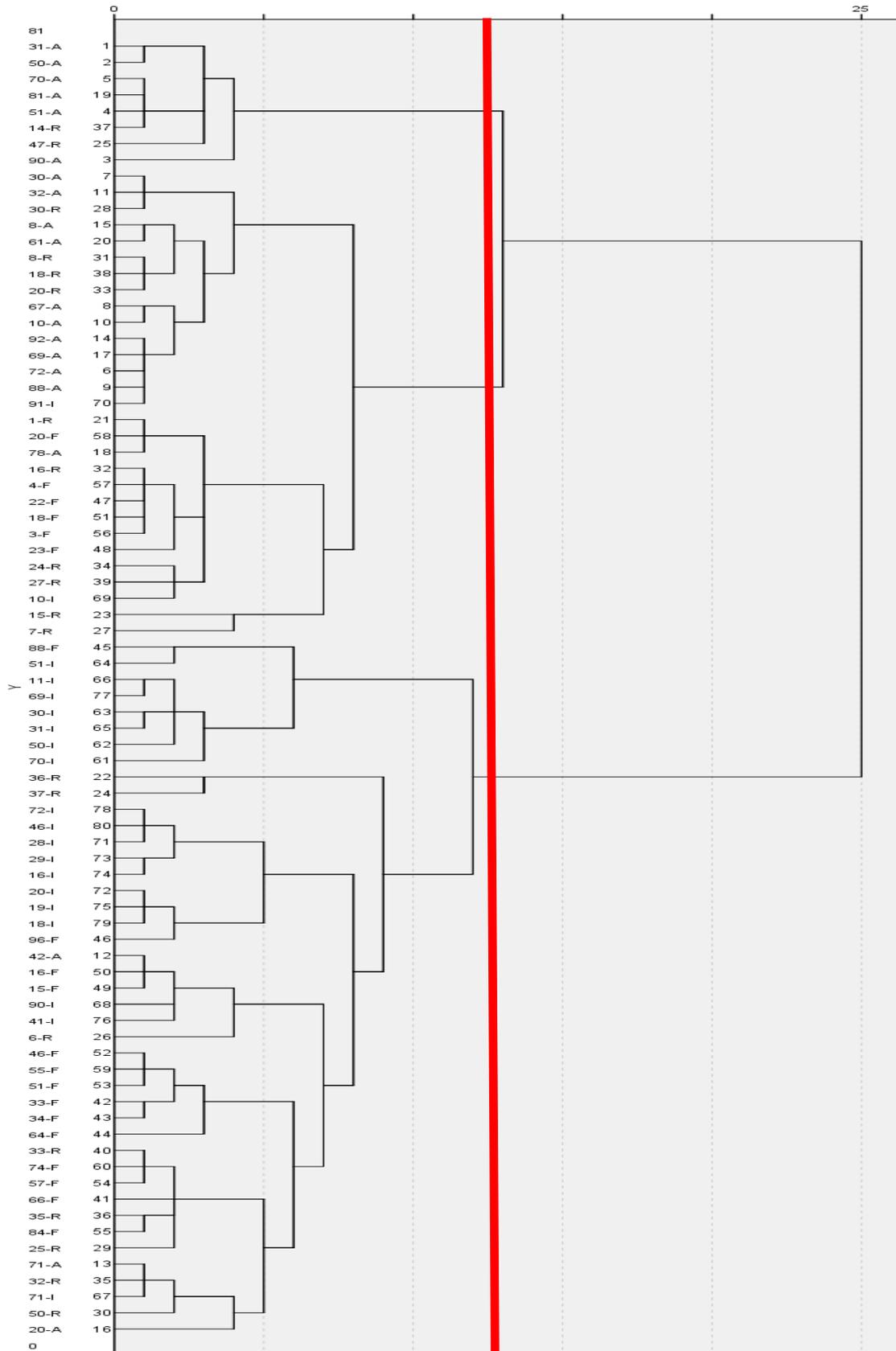


FIGURA 11 DENDROGRAMA RESULTANTE DA ANÁLISE DE AGRUPAMENTOS RELATIVO AO ANO DE 2010.

4.6 ANÁLISE DISCRIMINANTE

Os resultados do procedimento “stepwise”, obtidos para o ano de 1995 são apresentados na Tabela 27. Todas as dez espécies utilizadas como variáveis dependentes foram selecionadas. Os resultados do teste de lambda de Wilks e da estatística F para cada espécie, permitem inferir que as dez espécies contribuem significativamente e tem alto poder discriminante dos grupos. Importante salientar que as dez espécies estão entre as mais características do tipo de floresta estudado, segundo IBGE (2012).

TABELA 27 RESULTADOS DOS TESTES DE LAMBDA DE WILKS E F NA SELEÇÃO DE VARIÁVEIS DISCRIMINANTES, POR ORDEM DE ENTRADA, PARA O ANO DE 1995.

PASSO	ESPÉCIE	LAMBDA	F	SIGNIF.
1	<i>Cinnamodendron dinisii</i>	0,414	54,46	0,000*
2	<i>Araucaria angustifolia</i>	0,225	42,02	0,000*
3	<i>Lithraea brasiliensis</i>	0,160	37,42	0,000*
4	<i>Nectandra grandiflora</i>	0,121	34,71	0,000*
5	<i>Vitex megapotamica</i>	0,099	31,75	0,000*
6	<i>Ilex paraguariensis</i>	0,084	29,49	0,000*
7	<i>Myrsine coriacea</i>	0,072	27,73	0,000*
8	<i>Xylosma pseudosalzmanii</i>	0,061	26,66	0,000*
9	<i>Myrcia sp.</i>	0,053	25,59	0,000*
10	<i>Campomanesia xanthocarpa</i>	0,047	24,70	0,000*

*significativo a 95% de probabilidade.

Na Tabela 28, observam-se os resultados dos testes de lambda de Wilks e F que selecionaram quatro espécies, dentre as dez avaliadas pelo procedimento “stepwise”, para o ano de 2010. As quatro espécies selecionadas também constam da relação das obtidas em 1995. A ordem de entrada destas quatro espécies é a igual a ordem de entrada em 1995, sendo que a importância de cada espécie aumentou no ano de 2010.

TABELA 28 RESULTADOS DOS TESTES DE LAMBDA DE WILKS E F NA SELEÇÃO DE VARIÁVEIS DISCRIMINANTES, POR ORDEM DE ENTRADA, PARA O ANO DE 2010.

STEP	ESPÉCIE	LAMBDA	F	SIGNIF.
1	<i>Araucaria angustifolia</i>	0,396	58,73	0,000*
2	<i>Nectandra grandiflora</i>	0,237	39,99	0,000*
3	<i>Ilex paraguariensis</i>	0,203	30,43	0,000*
4	<i>Myrcia sp.</i>	0,179	25,24	0,000*

*significativo a 95% de probabilidade.

Após a definição das variáveis discriminantes, procedeu-se a determinação das funções discriminantes, cujos coeficientes foram computados de forma padronizada. Foram encontradas duas funções discriminantes para o ano de 1995 (Tabela 29), sendo que as espécies mais importantes na primeira função discriminante foram *Xylosma pseudosalzmanii*, *Vitex megapotamica*, *Cinnamodendrom dinisii*, e *Lithrea brasiliensis*. Na segunda função destacaram-se as espécies *Myrsine coriacea*, *Myrcia sp.*, *Araucaria angustifolia*, *Nectandra grandiflora* e *Ilex paraguariensis*. Na primeira função discriminante as espécies *Xylosma pseudosalzmanii* e *Vitex megapotamica* apresentaram pesos bem maiores que as demais.

TABELA 29 RESULTADOS DAS FUNÇÕES DISCRIMINANTES LINEARES OBTIDAS PARA AS ESPÉCIES SELECIONADAS NO ANO DE 1995.

ESPÉCIES	1ª FUNÇÃO	2ª FUNÇÃO
<i>Araucaria angustifolia</i>	0,077	0,506
<i>Campomanesia xanthocarpa</i>	0,352	0,104
<i>Cinnamodendrom dinisii</i>	0,983	-0,526
<i>Ilex paraguariensis</i>	0,112	0,425
<i>Lithraea brasiliensis</i>	0,714	0,168
<i>Myrcia sp.</i>	-0,088	-0,541
<i>Myrsine ferruginea</i>	0,846	0,908
<i>Nectandra grandiflora</i>	0,022	0,493
<i>Xylosma pseudosalzmanii</i>	4,742	0,927
<i>Vitex megapotamica</i>	-4,303	0,807
CONSTANTE	-1,682	-2,396

Na Tabela 30, são apresentados os coeficientes padronizados das duas funções discriminantes encontradas para o ano de 2010. As espécies mais importantes na primeira função discriminante foram *Araucaria angustifolia* e *Myrcia sp.*, enquanto na segunda função destacam-se as espécies *Ilex paraguariensis* e

Nectandra grandiflora. Na primeira função discriminante todas as espécies tiveram pesos baixos, não havendo destaque para nenhuma delas.

TABELA 30 RESULTADOS DAS FUNÇÕES DISCRIMINANTES LINEARES OBTIDAS PARA AS ESPÉCIES SELECIONADAS NO ANO DE 2010.

ESPÉCIES	1ª FUNÇÃO	2ª FUNÇÃO
<i>Araucaria augustifolia</i>	0,677	0,090
<i>Nectandra grandiflora</i>	0,192	0,444
<i>Ilex paraguariensis</i>	0,518	1,020
<i>Myrcia sp.</i>	0,309	-0,481
CONSTANTE	-2,604	0,317

Os resultados dos testes estatísticos relativos ao procedimento de seleção do número de discriminantes, para o ano de 1995, encontram-se na Tabela 31. Analisando os autovalores da função obtida, nota-se que não houve destaque de uma função sobre a outra, tendo em vista que a primeira função explica 59% da proporção da variância total. Esse fato deve-se ao elevado valor dos coeficientes de correlação canônica, que indicam um alto grau de relacionamento entre as funções discriminantes e o grupo de espécies.

Os testes, Qui-quadrado e Lambda de Wilks, mostram como as informações são retiradas das funções discriminantes. Antes de se obter a função, o valor de lambda era muito baixo (0,047), indicando que as séries de espécies sendo analisadas possuem um considerável poder de discriminação. Depois da alocação dessa capacidade de discriminação na primeira função, o valor de Lambda aumenta pouco e o alto valor de X^2 (222,2*) assegura que as variáveis continuam contendo poder de discriminação. Assim sendo, nenhuma das funções discriminantes deve ser excluída da análise.

TABELA 31 RESULTADOS DOS TESTES ESTATÍSTICOS DA SELEÇÃO DAS FUNÇÕES DISCRIMINANTES NO ANO DE 1995.

FUNÇÃO	AUTO-VALOR	% VARIÂNCIA EXPLICADA	CORRELAÇÃO CANÔNICA	LAMBDA	X^2
1	4,343	59,0	0,902	0,047*	222,2*
2	3,016	41,0	0,867	0,249*	100,8*

* significativo a 5% de probabilidade

Os resultados dos testes estatísticos relativos ao procedimento de seleção do número de discriminantes no ano de 2010 encontram-se na Tabela 32. Observando-se o autovalor de cada função discriminante, conclui-se que houve destaque de uma função discriminante sobre a outra, tendo em vista que a primeira função explica 77,4% da proporção da variância total.

Os testes, Qui-quadrado e Lambda de Wilks no ano de 2010 resultaram, assim como no ano de 1995, que nenhuma função discriminante deve ser excluída da análise (Tabela 32).

TABELA 32 RESULTADOS DOS TESTES ESTATÍSTICOS DA SELEÇÃO DAS FUNÇÕES DISCRIMINANTES NO ANO DE 2010.

FUNÇÃO	AUTO-VALOR	% VARIÂNCIA EXPLICADA	CORRELAÇÃO CANÔNICA	LAMBDA	X ²
1	2,327	77,4	0,836	0,179*	129,9*
2	0,680	22,6	0,636	0,595*	39,2*

* significativo a 5% de probabilidade

A Tabela 33 apresenta os resultados encontrados para os centroides, ou pontos médios de cada grupo, nas duas funções discriminantes, para os dois anos em estudo. Verifica-se que em 1995, tanto na primeira como na segunda função discriminante, o grupo dois discrimina os demais, pois apresenta a maior diferença entre os centroides. Para o ano de 2010, em ambas as funções discriminantes o grupo um é o que discrimina os demais.

O grupo dois, em 1995, é formado predominantemente por sub-parcelas pertencentes à parcela Rio, como pode ser observado na Figura 10, enquanto que para o ano de 2010, no grupo um predominam sub-parcelas da parcela Araucária (Figura 11). Essas duas parcelas encontram-se fisicamente próximas, como pode ser verificado na Figura 2, e talvez este fato explique o poder discriminante delas.

TABELA 33 RESULTADOS DOS CENTRÓIDES PARA AS FUNÇÕES DISCRIMINANTES PARA OS TRÊS GRUPOS DE SUB-PARCELAS, NOS ANOS DE 1995 E 2010.

GRUPO	1995		2010	
	FUNÇÃO 1	FUNÇÃO 2	FUNÇÃO 1	FUNÇÃO 2
1	0,265	2,083	3,071	-1,771
2	7,566	-2,528	1,060	0,907
3	-1,215	-1,598	-1,286	-0,282

As Figuras 12 e 13 apresentam as representações gráficas das sub-parcelas em seus respectivos grupos, bem como os centróides que os representam. Os grupos encontram-se bem separados, e isto se deve ao alto índice de acerto na classificação dos grupos.

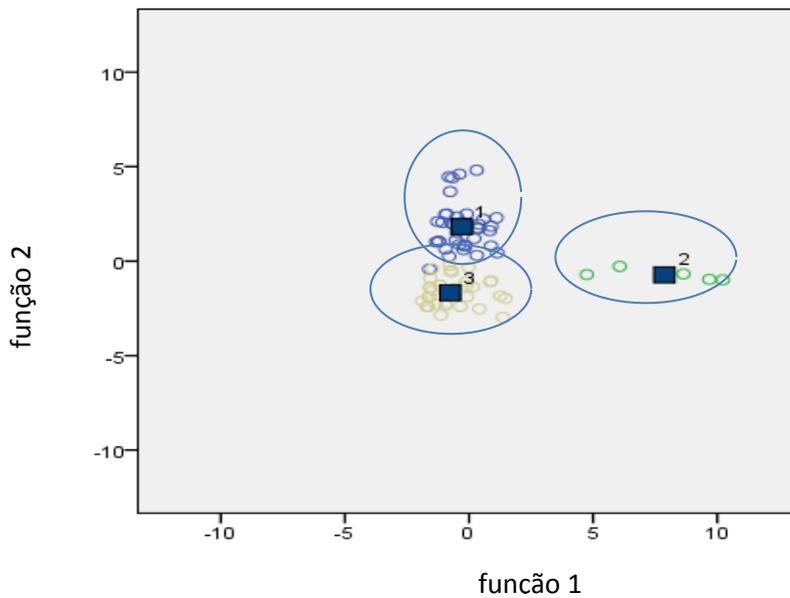


FIGURA 12 REPRESENTAÇÃO GRÁFICA DOS GRUPOS E SEUS CENTRÓIDES NO ANO DE 1995.

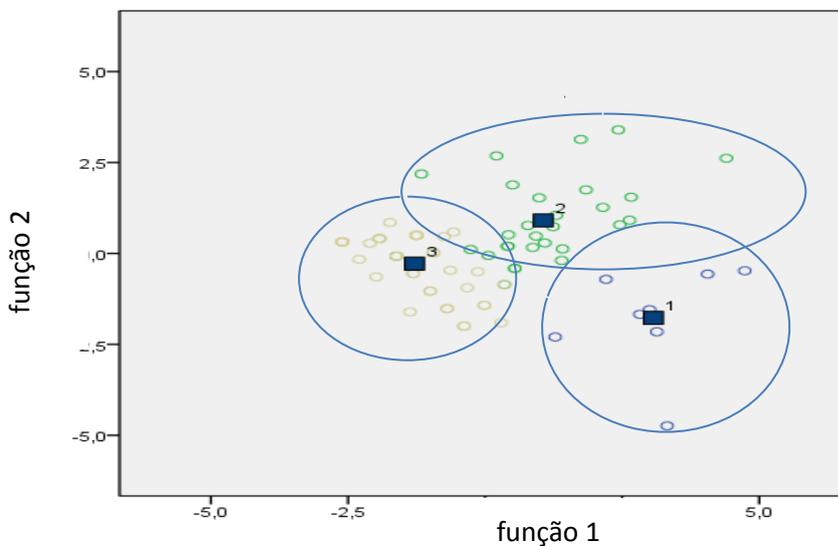


FIGURA 13 REPRESENTAÇÃO GRÁFICA DOS GRUPOS E SEUS CENTRÓIDES NO ANO DE 2010.

Os percentuais de classificações corretas e incorretas nos três grupos são apresentados na Tabela 34. Para o ano de 1995, no grupo um somente um indivíduo foi classificado erroneamente, enquanto que nos grupos dois e três todos os indivíduos foram classificados corretamente. No geral, houve 98,8% de classificações corretas, indicando precisão da técnica de agrupamento.

TABELA 34 RESULTADOS DA CLASSIFICAÇÃO, EM NÚMERO E PORCENTAGEM, DOS GRUPOS DE SUB-PARCELAS, NO ANO DE 1995.

GRUPOS	CLASSIFICAÇÃO DOS GRUPOS			Total
	1	2	3	
1	36	0	1	37
2	0	5	0	5
3	0	0	38	38
% 1	97,3	0,0	2,7	100,0
% 2	0,0	100,0	0,0	100,0
% 3	0,0	0,0	100,0	100,0

Para o ano de 2010, os percentuais de classificações obtidos nos três grupos, são apresentados na Tabela 35. Nos grupos de sub-parcelas um e três houve 100% de acerto, enquanto que para o grupo dois, 17,2% das classificações foram errôneas.

TABELA 35 RESULTADOS DA CLASSIFICAÇÃO, EM NÚMERO E PORCENTAGEM, DOS GRUPOS DE SUB-PARCELAS, NO ANO DE 2010.

GRUPOS	CLASSIFICAÇÃO DOS GRUPOS			Total
	1	2	3	
1	8	0	0	8
2	0	24	5	29
3	0	0	43	43
% 1	100,0	0,0	0,0	100,0
% 2	0,0	82,8	17,2	100,0
% 3	0,0	0,0	100,0	100,0

A análise discriminante forneceu duas funções discriminantes que separam os grupos formados pela análise de agrupamentos de forma efetiva. Decorrente dessas

funções discriminantes, não houve necessidade de reclassificar as espécies, tendo em vista os altos percentuais de acerto nas classificações efetuadas. Foram classificadas de forma correta 93,8%.

Os resultados obtidos na análise discriminante indicam que das espécies selecionadas com maior valor de importância, no ano de 2010, as que se destacam para os três grupos são *Araucaria angustifolia*, *Nectandra grandiflora*, *Ilex paraguariensis* e *Casearia decandra*, como já havia sido indicado na análise realizada para o ano de 1995. Os resultados indicam que ao longo de 15 anos a floresta manteve suas características florísticas, com exceção da parcela Fogo que apresentou um aumento de 8 novas espécies.

5. CONCLUSÕES

Com base nas análises realizadas no presente é possível concluir que:

- A representação da curva espécie/área pode ser feita pelo ajuste de um modelo de equação de regressão polinomial de 2º grau;
- As curvas obtidas para cada uma das parcelas não podem ser substituídas por uma única equação, em todos os anos estudados;
- O modelo ajustado para a curva espécie/área, dada sua natureza, permite verificar matematicamente onde se inicia o patamar assintótico da curva, por meio da determinação dos pontos de máxima das equações obtidas;
- Para a população finita que representa a área em estudo, o número de amostras efetuadas, para os anos de 1995 e 2010, foram maiores do que os recomendados para os níveis de erro e probabilidade considerados;
- O uso das variáveis fitossociológicas Densidade, Frequência, Dominância, Valor de Importância e Valor de Cobertura não apresentaram diferenças nas cargas quando foi realizada a análise fatorial, indicando semelhança das parcelas quanto sua a composição;
- A análise de variância multivariada (MANOVA) realizada, revelou diferenças entre os grupos de espécies selecionadas nos dois anos;
- A MANOVA indicou que existe relacionamento entre as espécies e a investigação da natureza desses relacionamentos, por meio da análise discriminante, apontou três combinações lineares de espécies para cada um dos anos, e suas respectivas contribuições;
- A análise de agrupamentos indicou a formação de três grupos distintos nos dois anos estudados;
- O uso da distância euclidiana ao quadrado e do método de Ward mostraram-se satisfatórios na determinação dos agrupamentos existentes, utilizando o número de indivíduos por sub-parcela;
- Houve diferença entre o número de espécies selecionadas nos anos de 1995 e 2010 pelo uso do método de “stepwise”, na análise discriminante;
- Para ambos os anos foram determinadas duas funções discriminantes distintas;
- As espécies que mais contribuíram na primeira função discriminante, para o ano de 1995, foram *Xylosma pseudosalzmanii*, *Vitex megapotamica*, *Cinnamodendron*

dinisii, e *Lithrea brasiliensis*. Na segunda função destacaram-se as espécies *Myrsine coriacea*, *Myrcia* sp., *Araucaria augustifolia*, *Nectandra grandiflora* e *Ilex paraguariensis*;

- As espécies mais importantes, no ano de 2010, na primeira função discriminante foram *Araucaria augustifolia* e *Myrcia* sp., enquanto na segunda função destacam-se as espécies *Ilex paraguariensis* e *Nectandra grandiflora*;
- As duas funções discriminantes obtidas para os anos de 1995 e 2010, não podem ser descartadas;
- Para o ano de 1995, tanto na primeira como na segunda função discriminante, o grupo dois discrimina os demais enquanto que para o ano de 2010 o grupo um é o que discrimina os outros grupos;
- Em ambos os anos estudados o nível de acerto nas classificações dos grupos foi efetivo, apresentando níveis de acerto maiores do que 90%;
- As várias técnicas de análise multivariada utilizadas mostraram-se eficientes na consecução dos objetivos visando verificar o comportamento das variáveis fitossociológicas em florestas nativas.

6 REFERÊNCIAS

AB'SÁBER, A. N. Os domínios morfoclimáticos da América do Sul. Primeira aproximação. **Geomorfologia**, São Paulo, v. 52, p. 1-22, 1977.

ALDER, D.; SYNNOTT, T. J. Permanent sample plot techniques for mixed tropical forest. **In: Tropical Forestry Papers**. Oxford Forestry Institute, Department of Plant Sciences, University of Oxford, n. 25, 1992.

ALVES, L. B. **Tratamento de dados multivariados por análise de correspondência e de agrupamento em dados de idosos de São José dos Campos**. 2007. 110p. Dissertação de Mestrado, Instituto Tecnológico de Aeronáutica. Curso de Engenharia Aeronáutica e Mecânica. São José dos Campos, 2007.

ANDERBERG, M. R. **Cluster analysis for applications**. New York: Academic Press, 1973.

ARAÚJO, M. M. et al. Análise de agrupamento da vegetação de um fragmento de floresta estacional decidual aluvial, Cachoeira do Sul, RS, Brasil. **Ciência Florestal**, Santa Maria, v. 14, n.1, p. 133-147, 2004.

ARRHENIUS, O. Species and área. **Journal of Ecology**, New York, NY, v.9, n.1, sept 1921.

BARROS, R. S. M. **Medidas de diversidade biológica**. Programa de Pós Graduação em Ecologia – Universidade Federal de Juiz de Fora, 2007. Disponível <http://www.ecologia.ufjf.br/admin/upload/File/Estagio_docencia_Ronald.pdf>. Acesso em: 6 de janeiro 2013.

BOLÓS, O. de; CERVI, A. C.; HATSCHBACH, G. Estudios sobre la vegetación del estado de Paraná (Brasil meridional). **Collectanea Botanica**, Barcelona, v. 20, p. 79-182, 1991.

BOTREL, R.; OLIVEIRA-FILHO, A. T; RODRIGUES, L. A.; CURI, N. Influência do solo e topografia sobre as variações da composição florística e estrutura da comunidade arbórea-arbustiva de uma floresta estacional semidecidual em Ingaí, MG. **Revista Brasileira de Botânica**, São Paulo, v.25, n.2, p. 195-213, jun. 2002.

BRAY, J. H., MAXWELL, S. E. Multivariate Analysis of variance. Sage university paper series on quantitative applications in the social sciences. Newbury Park(CA): Sage. P.07-54, 1985.

BRAUN-BLANQUET, J. **Plant Sociology: the study of plant communities**. New York/London: McGraw-Hill, 1932.

BROWN, J. H.; LAMOLINO, M. V. **Biogeografia**. 2 ed. Ribeirão Preto: FUNPEC, 2006. 691 p.

BUDKE, J. C.; JARENKOW, J. A.; OLIVEIRA-FILHO, A. T. Relationships between tree component structure, topography and soils of a riserside forest, Rio Botucaraí, Southern Brazil. **Plant Ecol**, Ames, n.189, p. 187 – 200, 2007.

BUSSAB, W. de O.; MIAZAKI, É. S.; ANDRADE, D. F. de. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 1990.

CAIN, S. A., Sample-plot technique applied to alpine vegetation in Wyoming. **American Journal of Botany** , New York, NY, v.30, p. 240-247, 1943.

CAIN, S.A.; CASTRO, G.M.de O. Application of some phytosociological techniques to Brazilian Rain Forest. **American Journal Botany**., Bronx, NY, v.43, n.3, p.205-217, 1956.

CAMARGO, F. M. **Caracterização da vegetação lenhosa e dos solos de um mosaico de serrado, floresta semidecídua, floresta decídua em Bocaiuva-MG**. 1997, 55p. Dissertação de Mestrado em Engenharia Florestal- Universidade Federal de Lavras, Lavras.

CARVALHO, D. A. et al. Distribuição de espécies arbóreo-arbustivas ao longo de um gradiente de solos e topografia em um trecho de floresta ripária do Rio São Francisco em Três Marias, MG, Brasil. **Rev. Brasileira de Botânica**, São Paulo, v. 28, n.2, p. 329-345, abr-jun., 2005.

COLWELL, R.; CODDINGTON, J. Estimating terrestrial biodiversity through extrapolation. **Philosophical Transactions of Royal Society of London**, London, v. 345, p.101-118, 1994.

COLWELL, R. K.; MAO, C. X.; CHANG J. Interpolating, extrapolating, and comparing incidencebased species accumulation curves. **Ecology**, New York, NY, v. 85, p. 2717-2727, 2004.

DAUBENMIRE, R. **Plant communities** - a textbook of plant synecology. New York: Harper & Row. Pub., 1968. 300p.

DONI, M. V. **Análise de cluster**: métodos hierárquicos e de particionamento. São Paulo, 2004.

DURIGAN, M.E. **Florística, dinâmica e análise proteica de uma Floresta Ombrófila Mista em São João do Triunfo – PR**, 1999, 125 p. Dissertação de Mestrado em Engenharia Florestal, Universidade Federal do Paraná, Curitiba.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária, Serviço Nacional de Levantamento e Conservação de Solos. **Levantamento de reconhecimento dos solos do estado do Paraná** – Tomo I. Curitiba: EMBRAPA – SNLCS/SUDESUL/IAPAR, 1984. 414 p

FAGUNDES, L. M. **Caracterização florística e estrutural do componente arbóreo de dois fragmentos de florestas ciliares na região do médio Rio Grande, em Alpinópolis e Passos, MG**, 2003, 87p. Dissertação de Mestrado em Engenharia Florestal, Universidade Federal de Lavras, Lavras.

FERREIRA, D. F. **Estatística multivariada**. 2. ed. Lavras: Ed. UFLA, 2011. 676 p.

FERRETTI, A. R.; BORGES, R. S.; BRITZ, R. M. Os estados da Mata Atlântica: Paraná. In.: CAMPANILI, M.; PROCHNOW, M. **Mata Atlântica – uma rede pela floresta**. Brasília: RMA, 2006. p. 58-76.

FIELD, A. **Descobrimos a estatística usando SPSS**. 2. ed. Porto Alegre: Artmed, 2009 688p.

FLOSS, P. A., **Aspectos ecológicos e fitossociológicos no entorno de nascentes em formações florestais do Oeste de Santa Catarina**, 2011. 154p. Tese de Doutorado em Engenharia Florestal, Universidade Federal de Santa Maria, Santa Maria, 2011.

GALVÃO, F. Métodos de levantamento Fitossociológico. In: **A Vegetação Natural do Estado do Paraná**. Curitiba: IPARDES-CTD, 1994. sp. Curso.

GLEASON, H. A. On the relation between species and area. **Ecology** New York, NY, v. 3, p158-162, 1922.

GLEASON, H. A., Species and area. **Ecology**, New York, NY, v. 6, p. 66-74, 1925.

GLOSSÁRIO DE ECOLOGIA – 1. Ed., São Paulo: Publicação ACIESP, 1987, n. 57.

GOLDSMITH, F. B.; HARRISON, C. M. Description and analysis of vegetation. In CHAPMAN, S.B. ed. **Methods in plant ecology**. London: Blackwell Scientific Publications, 1976. p. 85-155.

GOMES, J. F. **Classificação e crescimento de grupos ecológicos da Floresta Ombrófila Mista na FLONA de São Francisco de Paula, RS**. 2005, 75 p. Dissertação de Mestrado em Engenharia Florestal, Universidade Federal de Santa Maria, Santa Maria, 2005.

GOMIDE, L. R.; SCOLFORO, J. R. S.; OLIVEIRA, A. D.; THIERSCH, C. R. Suficiência amostral para fragmentos florestais do rio São Francisco. In: SCOLFORO, J. R. S.; OLIVEIRA, A. D. **Um modelo fitogeográfico para as áreas de preservação permanente da bacia do rio São Francisco, em Minas Gerais**. Lavras: UFLA, 2005. p. 128-136.

GOMIDE, L. R. *et al.* Uma nova abordagem para definição da suficiência amostral em Fragmentos florestais nativos. **Cerne**, Lavras, v. 11, n. 4, p. 376-388, out./dez. 2005.

GOTELLI, N. E., COLWELL, R. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. **Ecology Letters**, Malden, MA, v.4, p.379-391, 2001.

GOTELLI, N. J.; AARON, M. E. **Princípios de estatística em ecologia**, Porto Alegre: Editora Artmed, 2011. 528 p.

GRAYBILL, F. A. **Theory and application of the linear model**. Belmont: Duxbury Press, 1976. 704 p.

HAIR, J. F. ANDERSON, et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Editora Bookman, 2005.

HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. On clustering validation techniques. **Journal of Intelligent Information Systems**, New York, NY. v. 17, n. 2-3, p. 107-145, 2001.

HANSON, H. C.; CHURCHILL, E. D. **The plant community**. New York: Reinhold Pub. Co., 1961. 218 p.

HILL, T.; MARQUES, L.; O'CONNOR, M. REMUS, W. Artificial neural networks for forecasting and decision making. International **Journal of Forecasting**, Malden, MA. v.10, p. 5-15, 1994

HUSCH, B.; MULLER, C. I. **Forest mensuration** 2. ed. New York: John Wiley Sons, 1982. 202 p.

IAPAR. Instituto Agrônomo do Paraná. **Cartas climáticas do Paraná**. 1994. Disponível <<http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=863>>. Acesso em 11/11/2013.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Geografia do Brasil - Região Sul**. v.2. Rio de Janeiro: FIBGE, 1990. 420p.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Manual técnico da vegetação brasileira**. Rio de Janeiro: 1992, 92p. (Série Manuais Técnicos em Geociências, 1).

IBGE. Instituto Brasileiro de Geografia e Estatística. **Manual técnico da vegetação brasileira**. Rio de Janeiro: 2012, 275p., 2. ed. (Série Manuais Técnicos em Geociências, 1).

JOHNSON, R. A., WICHERN, D. W. **Applied Multivariate Statistical Analysis**. Madison: Prentice Hall International, 1982. 607p.

JOHNSON, R. A.; WICHERN, D. W., **Applied multivariate statistical analysis**. 4 ed. New Jersey: Prentice Hall, 1992.

KACHIGAN, S.K. **Multivariate statistical analysis: a conceptual introduction**. 2. ed. New York: Radius Press, 1991.

KANIESKI, M. R., **Caracterização Florística, Diversidade e Correlação Ambiental na Floresta Nacional de São Francisco de Paula, RS**. 94p. 2010
Dissertação (Mestrado em engenharia Florestal), Universidade Federal de Santa Maria, Santa Maria, 2010

KASZNAR, I. K. **Falências e concordatas de empresas: modelos teóricos e estudos empíricos – 1978- 1982**. 1986, Tese (Doutorado em Economia), Fundação Getúlio Vargas, Rio de Janeiro, 1986.

KAUFMAN, L.; ROUSSEEUW, P. J. **Finding groups in data: an introduction to cluster analysis**. New York: Wiley, 1990.

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

LANDIM, P. M. B., **Texto Didático 15**, São Paulo, UNESP, 2006.

LAVINA, E.; FAUTH, G. Evolução geológica da América do Sul nos últimos 250 milhões de anos. In: CARVALHO, C.; ALMEIDA, E. (Orgs.). **Biogeografia da América do Sul: padrões e processos**. São Paulo: Roca, v. 1, 2010. p. 3-13.

LEITE, P. F; KLEIN, R. M. Vegetação. In: IBGE. **Geografia do Brasil: Região Sul**. Rio de Janeiro, 1990, p. 115-150.

LEITE, P. F. **As diferentes unidades fitoecológicas da Região Sul do Brasil: proposta de classificação**, 1994, 160 p. Dissertação (Mestrado em Engenharia Florestal), Universidade Federal do Paraná, Curitiba. 1994.

LIMA, C. S. A. **Desenvolvimento de um modelo para manejo sustentado do serrado**. 1997, 157p. Dissertação (Mestrado em Engenharia Florestal), Universidade Federal de Lavras, Lavras.1997.

LONGHI, S. J. **A estrutura de uma floresta natural de *Araucaria angustifolia* (Bert.) O. Ktze., no sul do Brasil**. 1980. 198 f. Dissertação (Mestrado em Engenharia Florestal), Universidade Federal do Paraná, Curitiba, 1980.

LONGHI, S. J. **Agrupamento e análise fitossociológica de comunidades florestais na sub-bacia hidrográfica do Rio Passo Fundo - RS**. 1997. 193 f. Curitiba. Tese (Doutorado em Ciências Florestais), Universidade Federal do Paraná. Curitiba, 1997.

MAACK, R. **Geografia física do Estado do Paraná**. Curitiba: José Olympio, 1968.

MAACK, R. **Geografia física do Estado do Paraná**. 2. ed. Rio de Janeiro: J.Olympio; Curitiba: Secretaria da Cultura e do Esporte do Governo do Estado do Paraná, 1981. 450 p.

MAGALHÃES, S. R. S, MAGINI, M. Avaliação de Métodos para Comparação de Modelos de Regressão por Simulação de dados. **SEMINA: Ciências Exatas e Tecnológicas**, Londrina, v.25, p. 117- 122, 2004.

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

MARTINS, S. V; SILVA, N. R. S; SOUZA, A. L; NETO, J. A. A. M. Distribuição de espécies arbóreas em um gradiente topográfico de Floresta Estacional Semidecidual em Viçosa, MG. **Scientia Florestalis**, Piracicaba, n. 64, p. 172 – 181, Dez. 2003.

MARRIOTT, F. H. C. **The interpretation of multiple observations**. New York: Academic Press, 1974.

MATHER, P. M. **Computer processing of remotely sensed images**. New York: John Wiley e Sons, Inc., 1999. 292 p.

MATLAB r 2009a. Matrix Laboratory

MUELLER-DOMBOIS, E.; ELLENBERG, F. **Aims and methods of vegetation ecology**. New York: Willey & Sons, 1974. 547 p.

NARVAES, I. da S. **Classificação e caracterização da regeneração natural em Floresta Ombrófila Mista na Floresta Nacional de São Francisco de Paula, RS**. 2004. 143 f. Dissertação (Mestrado em Engenharia Florestal), Universidade Federal de Santa Maria, Santa Maria, 2004.

NETER, J.; KUTNER, M. NACHTSHEIM, C.; WASSERMANN, W. **Applied linear statistical models**, New York: McGraw-Hill Companies Inc, 1996.

OLIVEIRA FILHO, A. T. et al. Diversity and structure of the tree community of a fragment of tropical secondary forest of the Brazilian Atlantic Forest domain 15 and 40 years after logging. **Rev. bras. Bot.**, São Paulo, v.27, n.4, p. 685-701, out-dez. 2004.

OLSON, C. L. Comparative robustness of six tests in multivariate of variance. **Journal of the American Statistical Association**, Montreal, v. 69, p. 894-908, 1974.

OLSON, C. L. On choosing a test statistic in multivariate analysis of variance. **Psychological Bulletin, California**, v. 83, p. 579- 586, 1976.

OLSON, C. L. Practical considerations in choosing a MANOVA test statistic: a rejoinder to Stevens. **Psychological Bulletin, California**, v. 86, p. 1350-1352, 1979.

PÉLLICO NETTO, S.; BRENA, D. A. **Inventário florestal**. Curitiba. Editorados pelos autores, 1997, 316 p.

PINTO, J. R. R.; OLIVEIRA-FILHO A. T.; HAY, J. D. V. Influence of soil and topography on the composition of a tree community in central brazilian valley forest. **Edinburgh Journal of Botany**, Cambridge, v. 62, n. 1-2, p. 69-90, 2005.

PIRES, P. T. L. de; ZENI JUNIOR, D. M.; GAULKE, D. As unidades de conservação e a Floresta Ombrófila Mista no estado do Paraná, **Ciência Florestal**, Santa Maria, v.22, n.3, p. 589-603, 2012.

PIZATTO, W. **Avaliação biométrica da estrutura e da dinâmica de uma Floresta Ombrófila Mista em São João do Triunfo - PR: 1995 a 1998**. 1999, 172 f. Dissertação (Mestrado em Engenharia Florestal), Universidade Federal do Paraná, Curitiba, 1999.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. **Revista Ceres**, Viçosa, v. 40, n. 228, p. 176-195, 1993.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão. **Revista Agropecuária Brasileira**, Brasília, v.31, n. 1, p. 1-17, 1996.

REGAZZI, A. J. **Análise Multivariada**. Viçosa: Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas. Departamento de Informática, 2001, 166 p.

RIBEIRO, S. B. **Classificação e ordenação da comunidade arbórea da Floresta Ombrófila Mista da Flona de São Francisco de Paula, RS**. 2004. 181 f. Tese (Doutorado em Engenharia Florestal), Universidade Federal de Santa Maria, Santa Maria, 2004.

RICE, E.L.; KELTING, R.W. The species-area curve. **Ecology**, New York, v.36, p. 7-11. 1955.

RICHARDS, J. A. **Remote Sensing Digital Image Analysis - An Introduction**. 2nd ed. Springer. Berlin: Verlag, 1993, 281 p.

RODE, R.; FIGUEIREDO FILHO, A.; MACHADO, S.A.; GALVAO, F. Grupos florísticos e espécies dominantes em povoamento de Araucaria angustifolia e uma floresta ombrófila mista. **Revista Árvore**, Viçosa, v. 35, n. 2, p. 319-327, 2011.

RODERJAN, C. V.; KUNIYOSHI, Y. S. **Macrozoneamento Florístico da Área de Proteção Ambiental de Guaraqueçaba: APA – GUARAQUEÇABA.**, 15, Curitiba FUPEF, 1988. 53 p. (

RODERJAN, C. V.; GALVÃO, F.; KUNIYOSHI, Y. S.; HATSCHBACH, G. G. As unidades fitogeográficas do Estado do Paraná. **Ciência & Ambiente**, Santa Maria, v. 24, p. 75-92, 2002.

ROMESBURG, Charles H. **Cluster analysis for researchers**. Belmont: Lifetime Learning Publications, 1984.

SANQUETTA, C. R.; MATTEI, E. (Org.). **Perspectivas de recuperação e manejo sustentável das Florestas de Araucária**. Curitiba: Multi-Graphic, 2006, 297 p.

SCHILLING, A.C.; BATISTA, J.F. Curva de acumulação de espécies e suficiência amostral em florestas tropicais. **Revista Brasileira de Botânica**, São Paulo, v. 31, p. 179-187, 2008.

SCHORN, L. A., Composição do Banco de Sementes no Solo em Áreas de Preservação Permanente sob diferentes tipos de Cobertura, **Floresta**, Curitiba, v. 43, n. 1, p. 49-58, jan/ mar. 2013.

SCHOWENGERDT, R.A. **Remote Sensing: Models and Methods for Image Processing**. 3. ed. London: Academic Press.1997. 560 p.

SHIVER, B. D.; BORDERS, B. E. **Sampling techniques for forest resource inventory**, New York, John Wiley, 1996.

SPVS. Sociedade de Pesquisa em Vida Selvagem e Educação Ambiental – SPVS. **Nossas Árvores: manual para recuperação da Reserva Florestal Legal**. Curitiba: FNMA,1996. 84 p.

SOUZA, P. F. **Terminologia florestal** - glossário de termos e expressões florestais. Rio de Janeiro: Fundação IBGE, 1973. 302 p.

SOUZA, J. S. et al. Análise das variações florísticas e estruturais da comunidade arbórea de um fragmento de floresta semidecidual às margens do rio Capivari, Lavras – MG. **Árvore**, Viçosa, v. 27, n. 2, p. 185-206, 2003.

SOUZA, D. C. **Estudo Fitossociológico no Remanescente Florestal nas margens dos rios Córrego Água dos Papagaios e do Campo, Campo Mourão Paraná**. SIAUT, Simpósio Ambiental da Universidade Federal do Paraná, Junho 2003.

STELL, R. G. D.; TORRIE, J. H. **Principles and Procedures of Statistics**. 2. Ed. New York: MacGraw-Hill Book Co, 1980.

STEVENS, J. P. Power of the multivariate analysis of variance tests. **Psychological Bulletin**. California, v. 88, p.728-737. 1980.

VASCONCELOS, P. C. S. **Fitossociologia de uma vegetação em sucessão secundária, no Vale do Paraíba, São Paulo.** 1992, 116 f. Dissertação (Mestrado em Engenharia Florestal) Universidade Federal de Viçosa, Viçosa, 1992.

VICINI, L. **Análise multivariada da teoria à prática.** Santa Maria: UFSM, CCNE, 2005.

VOLPATO, M. M. L. **Regeneração Natural em uma floresta secundária no domínio de mata Atlântica:** uma análise fitossociológica. 1994, 123 f. Dissertação (Mestrado em Engenharia Florestal), Universidade Federal de Viçosa, Viçosa, 1994.

WARD, J. H. Hierarchical grouping to optimize an objective function. **Journal of. American Statistical Association**, Montreal, v. 58, p. 236- 244, 1963.

ZAIANE, O. R. et al. **On data clustering analysis: scalability, constraints and validation.** Edmonton Alberta, University of Alberta, 2003.