

CHRISTOVÃO PEREIRA ABRAHÃO

**ESTIMATIVA DE ALGUMAS PROPRIEDADES DA MADEIRA DE
Eucalyptus urophylla POR ESPECTROMETRIA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2005

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

A159e
2005

Abrahão, Christovão Pereira, 1966-
Estimativa de algumas propriedades da madeira de
Eucalyptus urophylla por espectrometria / Christovão
Pereira Abrahão. – Viçosa : UFV, 2005.
x, 182f. : il. ; 29cm.

Inclui apêndice.

Orientador: Ricardo Marius Della Lucia.
Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas. F. 132-137.

1. Madeira - Testes. 2. Espectroscopia de reflectância.
3. Eucalipto. I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 634.9812

CHRISTOVÃO PEREIRA ABRAHÃO

**ESTIMATIVA DE ALGUMAS PROPRIEDADES DA MADEIRA DE
Eucalyptus urophylla POR ESPECTROMETRIA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

APROVADA: 19 de dezembro de 2005

Prof. Benedito Rocha Vital

(Conselheiro)

Prof. Francisco de Assis de Carvalho Pinto

(Conselheiro)

Prof. José Tarcisio da Silva Oliveira

Prof. José de Castro Silva

Prof. Ricardo Marius Della Lucia
(Orientador)

AGRADECIMENTOS

A Universidade Federal de Viçosa pela oportunidade de desenvolver meus estudos e me aperfeiçoar como ser humano. A Fundação de Amparo a Pesquisa de Minas Gerais – FAPEMIG, pelo período de bolsa a mim concedido.

Ao Prof. Ricardo Marius Della Lucia pelos longos anos de orientação e amizade. Ao Prof. José Tarcísio da Silva Oliveira, quem sempre esteve disposto a ajudar em meus trabalhos desde o amadurecimento de idéias, passando pela coleta de dados até a revisão de meus relatórios. Ao Prof. José de Castro e Silva, meu outro grande incentivador. Obrigado, amigos!

A meu pai Epiphaneo Walter Abrahão pelo exemplo de paixão pelo trabalho e pela família. A minha querida mãezinha (*in memoriam*), pelo exemplo de amor pelas pessoas. A meus irmãos Luciano e Eustáquio e Walter que sempre estiveram do meu lado pra qualquer parada. A minha esposa Léa e a meu filho Gabriel, pela constante presença e pelo carinho a mim dedicado durante os anos de desenvolvimento deste trabalho.

A meu querido amigo Antônio de Pádua Gomide, quem sempre manteve abertos para mim a sua casa e o seu atelier, onde passei momentos de deleite e reflexão sobre a arte de cultivar a amizade.

Aos funcionários do Laboratório de Propriedades Físicas e Mecânicas da Madeira, Osvaldo Ladeira e José Magalhães (*in memoriam*), pela amizade e auxílio em meus experimentos. A todos os meus professores e colegas de graduação e

pós-graduação, que tornaram essa jornada um agradável passeio pelo mundo da Ciência. Agradeço especialmente a Carlos Alberto Varela, Joseph Kalil Khoury Junior e Daniel Câmara Barcellos, Wagner de Souza Junior. Parceirões!

Dedico

A minha amada esposa, Léa,
mulher que é pura fibra
e sentimento.

A meu adorado filho, Gabriel,
anjo de ternura,
amigo inseparável.

BIOGRAFIA

CHRISTOVÃO PEREIRA ABRAHÃO, filho de Epiphaneo Walter Abrahão e Eny Pereira, nasceu no dia 13 de dezembro de 1966, em Andrelândia, Sul do Estado de Minas Gerais.

Concluiu seu curso de graduação em Engenharia Florestal na Universidade Federal de Viçosa em 1996. No ano de 2000, obteve pela mesma universidade o título *Magister Scientiae* em Ciência Florestal. Neste mesmo ano, deu início ao seu curso de doutorado em Ciência Florestal.

Desde a sua graduação, procurou desenvolver trabalhos em vários campos da tecnologia da madeira, como propriedades físicas e mecânicas, anatomia e identificação, equipamentos e métodos de preservação, estabilização dimensional, construções, secagem e energia da madeira.

ÍNDICE

	Página
RESUMO	viii
ABSTRACT	x
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1. Aplicações dos métodos não-destrutivos	3
2.2. Os métodos para a avaliação não-destrutiva da madeira	5
2.2.1. Avaliação visual	5
2.2.2. Propagação de ondas de tensão – Testes acusto-ultrasônicos.....	6
2.2.3. A atenuação de raios-X.....	9
2.2.4. Propriedades elétricas da madeira.....	12
2.2.5. Classificação mecânica	13
2.2.6. Propriedades ópticas da madeira.....	15
2.3. Princípios da espectroscopia	23
2.3.1. A Lei de Bouguer – Lambert – Beer.....	23
2.3.2. Interações entre matéria e energia.....	26
2.3.3. Respostas da madeira à radiação no UV-VIS.....	29
2.3.4. A variabilidade da madeira e a amostragem em espectroscopia de reflectância.....	34
3. OBJETIVOS	38
3.1. Objetivo geral.....	38
3.2. Objetivos específicos	38
4. MATERIAL E MÉTODOS	39
4.1. A madeira e a amostragem.....	39
4.2. A determinação das propriedades da madeira.....	41

4.2.1. Os espectros de reflectância.....	41
4.2.2. Propriedades físicas e mecânicas da madeira	42
4.3. Pré-processamento dos dados	43
4.3.1. Filtragem dos espectros	43
4.3.2. Transformações sobre os espectros.....	44
4.3.3. Compilação dos dados	45
4.4. Ajustes dos modelos de calibração e avaliação dos resultados.....	45
4.4.1. Regressões lineares simples e múltiplas por MQO	47
4.4.2. Regressão múltipla por componentes principais	47
4.4.3. Esquema geral de processamento dos dados	48
5. RESULTADOS E DISCUSSÃO	50
5.1. Colorimetria quantitativa	50
5.2. Densidade básica	53
5.2.1. Os espectros e suas correlações com a densidade básica.....	55
5.2.2. Modelos múltiplos ajustados por PCR.....	62
5.2.3. Otimização dos modelos	75
5.2.3.1. Remoção de pontos discrepantes	75
5.2.3.2. Resolução espectral.....	81
5.2.3.3. Suavização das curvas espectrais.....	84
5.2.3.4. Transformações sobre os espectros.....	87
5.2.3.4.1. Derivação das curvas espectrais.....	87
5.2.3.4.2. A absorbância.....	98
5.2.3.4.3. A função de Kubelka-Munk.....	102
5.2.3.4.4. Transformações para normalização dos dados espectros- cópicos	104
5.2.3.5. A regressão por PCR <i>forward</i>	113
5.3. A rugosidade da madeira.....	117
5.4. A resistência à compressão e a dureza	120
6. CONCLUSÕES E SUGESTÕES	129
REFERÊNCIAS BIBLIOGRÁFICAS.....	132
APÊNDICE A	138
A.1. A regressão por componentes principais.....	138
A.1.1. As direções dos eixos PC's	140
A.1.2. Procedimento matricial para o cálculo dos PC's	146
A.2. O equipamento e sua operação.....	150
A.2.1. O sistema óptico.....	150
A.2.2. Os espectrômetros	151
A.2.3. A esfera integradora e a sonda de reflectância.....	153
A.2.4. A aquisição dos espectros	154
A.3. O algoritmo computacional.....	157

RESUMO

ABRAHÃO, Christovão Pereira, D.S. Universidade Federal de Viçosa, dezembro de 2005. **Estimativa de algumas propriedades da madeira de *Eucalyptus urophylla* por espectrometria.** Professor Orientador: Ricardo Marius Della Lucia. Conselheiros: Benedito Rocha Vital e Francisco de Assis de Carvalho Pinto.

São escassas as referências na literatura mundial acerca da utilização da espectroscopia no UV-VIS-tNIR (200 a 1.000 nm) para predição das propriedades da madeira, apesar de os equipamentos que operam nesta faixa do espectro eletromagnético serem de custo muito inferior àqueles que operam no NIR (700-2.500 nm). O objetivo deste trabalho foi verificar a possibilidade da utilização de espectros no UV-VIS-tNIR para o desenvolvimento de modelos de calibração a serem utilizados na determinação de propriedades da madeira. Foi comprovada a existência de fortes correlações entre a reflectância da superfície da madeira sólida e sua densidade básica, sendo o mais alto coeficiente de correlação encontrado por volta de 375 nm, exatamente na região de maior absorção da energia radiante pela lignina. Foram ajustados modelos de calibração multivariados utilizando a técnica de regressão linear por componentes principais. Com o objetivo de se otimizar a capacidade de predição dos modelos de calibração foram empregadas técnicas de pré-processamento dos dados espectroscópicos, como variações na resolução espectral, filtragem polinomial e transformações matemáticas sobre a reflectância, como o cálculo da absorbância, a aplicação da função de Kubelka-Munk, a divisão pela variância, a extração da raiz quadrada e a função de Box-Cox, sucedidas ou não pelo cálculo das derivadas primeiras e segundas. Concluiu-se que é possível

obter-se modelos que apresentaram R^2 na validação igual a 0,89 e RMSEP igual a 5% ($40,75 \text{ kg/m}^3$) utilizando 5 componentes principais da reflectância como variáveis independentes. O método aqui empregado não forneceu bons resultados no desenvolvimento de modelos de calibração para a rugosidade, a dureza e a resistência à compressão.

ABSTRACT

ABRAHÃO, CHRISTOVÃO PEREIRA. DS. Universidade Federal de Viçosa. Dezembro de 2005. **Estimation for some properties of the wood of *Eucalyptus urophylla* by spectrometry**. Adviser: Ricardo Marius Della Lucia. Committee Members: Benedito Rocha Vital and Francisco de Assis de Carvalho Pinto.

References to UV-VIS-tNIR (200-1,000 nm) wood reflectance that studies is very scarce in world literature, in spite of the fact that equipments that operate in this range of electromagnetic spectrum have much lower costs than those that operate in the NIR (700-2,500 nm). The objective of the present work was to verify the possibility of using UV-VIS-tNIR spectra for calibration purpose in wood properties determination. It was proven that there exists of strong correlation between solid wood surface reflectance and its basic specific gravity, with the highest correlation coefficient found near 375 nm, exactly the region of major radiant energy absorption by lignin. Multivariate calibration models by principal components regression was constructed. To optimize the models performances, pre-processing techniques were applied on spectroscopic data, like reduction in spectral resolution, polynomial filtering and mathematics transformations on reflectance like absorbance computing, Kubelka-Munk function, division by variance, square root and Box-Cox function. It was concluded that is possible to obtain reasonably good estimates of basic specific gravity of wood ($R^2=0.89$ and $RMSEP=5\%$ $\Leftrightarrow 40,75 \text{ kg/m}^3$) with UV-VIS-tNIR reflectance using 5 components in a model based on spectroscopic data. However, no good results were obtained for roughness, hardness and crushing strength calibrations with the present method.

1. INTRODUÇÃO

Os métodos para a avaliação das propriedades da madeira se encontram em uma constante evolução que é impulsionada pela busca de maneiras mais rápidas, mais acuradas e de mais baixo custo para se conhecer o material. Se, além dessas características, um método alternativo é ainda não-destrutivo, são grandes as chances da sua utilização na indústria e em laboratórios de pesquisa. Na indústria, como ferramenta de controle de qualidade das matérias-primas, dos processos e dos produtos acabados; e como método laboratorial, para viabilização de pesquisas que dependem do conhecimento de propriedades da madeira a partir de um grande número de amostras, como no estudo de efeitos de sutis mudanças genéticas e ambientais, por exemplo.

São chamados métodos de avaliação não-destrutivos aqueles que não requerem, para a preparação de amostras e execução dos testes propriamente ditos, transformações na madeira que a tornem imprestável para a sua utilização prática ou para a aplicação de outros testes subsequentes (ROSS, 1999; ZOMBORI, 2000). Na maioria das vezes os métodos não-destrutivos se baseiam na investigação de correlações e ajuste de modelos de calibração entre as propriedades de interesse e outras de mais fácil mensuração, o que os caracteriza, também, como métodos indiretos.

Vários métodos não-destrutivos têm sido investigados por pesquisadores ao redor do mundo; alguns, com desdobramentos aplicáveis à detecção de defeitos e à classificação de peças de madeira na indústria, outros se tornam importantes auxiliares nos laboratórios de pesquisa, reduzindo custos e prazos necessários para o

conhecimento das propriedades da madeira. As principais técnicas utilizadas em tais métodos baseiam-se nas correlações entre as propriedades de interesse da madeira e suas propriedades mecânicas, ópticas, acústicas, térmicas, elétricas, dielétricas e radiográficas - atenuação de raios X, γ e nêutrons (SZYMANI; MCDONALD, 1981; FUNT; BRIAN, 1987; TROUGHTON; CLARKE, 1987; TREMBLAY, 1995; KLINE et al., 1998; STEELE et al., 1991; SAMSON et al., 1993, BINDZI; SAMSON, 1995; SAMSON, 1988; OJA et al., 2000).

Bailleres et al. (2002) afirmaram serem os altos custos e a morosidade dos métodos tradicionais de determinação de propriedades tecnológicas da madeira as maiores restrições na triagem de indivíduos em extensos programas de melhoramento genético florestal. Os mesmos autores apontam a espectroscopia no infravermelho próximo (NIRS, do inglês: *Near Infrared Spectroscopy*) como uma alternativa viável aos métodos tradicionais. De fato, muitos trabalhos têm sido desenvolvidos ao redor do mundo com o emprego da NIRS para a determinação de propriedades da madeira, principalmente no setor de celulose e papel.

Algumas dezenas de trabalhos enfocando o uso da NIRS para a predição de propriedades da madeira podem ser encontradas na literatura. Menos comuns, porém, são os relatos de aplicações da espectroscopia na faixa do visível e do ultravioleta (UV-VIS, do inglês: *ultraviolet and visible spectroscopy*). Acredita-se que esta seja uma lacuna que deve ser preenchida, pois espectrômetros que operam na faixa do visível (400 a 700 nm) e outros que incluem as transições para o ultravioleta e para o infravermelho (300 a 1.000 nm), são equipamentos de custo muito inferior aos daqueles que operam somente no infravermelho próximo (1.000 a 2.500 nm). Os fabricantes atribuem as diferenças nos preços às tecnologias de produção dos sensores dos aparelhos. Sensores para operação no infravermelho exigem maior controle na dopagem dos semicondutores. Além disso, no processo de dopagem são utilizados elementos químicos mais raros como índio e gálio.

Assim, o presente trabalho se propôs a investigar a possibilidade da utilização da espectroscopia de reflectância nas regiões do ultravioleta, do visível e transição visível-infravermelho como método não destrutivo para a determinação de propriedades da madeira.

2. REVISÃO DE LITERATURA

2.1. Aplicações dos métodos não-destrutivos

Bodig (2000) sugeriu o uso do termo “avaliação não destrutiva – AND” (em inglês, *non-destructive evaluation – NDE*), em preferência ao termo “ensaios não destrutivos” (em inglês, *non-destructive testing - NDT*), já que o primeiro apresenta um caráter mais geral para a descrição das atividades que podem ser desenvolvidas por um perito na determinação das propriedades e do estado em que se encontra uma peça de madeira ou de seus derivados. Assim, no decorrer do presente texto será utilizada a sigla AND sempre que se referir ao termo “avaliação não destrutiva”.

Ainda Bodig (2000), comentou que as AND's, tanto para a madeira, quanto para outros materiais, encontram-se ainda na infância de seu desenvolvimento. O autor justifica esta opinião lembrando a fragilidade dos modelos físicos diante da complexidade gerada pelas interações entre as diversas variáveis do mundo real que não podem ser ponderadas *a priori*. Tal observação mostra-se especialmente apropriada no que diz respeito à madeira, que é produto de uma atividade biológica e que, mesmo após a sua formação, ainda tem as suas propriedades afetadas por fatores ambientais, inclusive pela sua manipulação e beneficiamento (ROSS, 1999).

Apesar das limitações existentes, pode-se listar alguns sucessos na pesquisa e desenvolvimento de equipamentos industriais para a realização de AND's sobre a madeira. Os métodos e princípios aplicados estão relacionados à natureza, geometria e localização na cadeia produtiva em que se encontram os materiais a serem testados. Por

exemplo, a avaliação de vigas de madeira já em uso numa construção não poderia ser realizada com um método que se baseie em testes de flexão, o que ainda seria possível antes da sua instalação.

Pode-se separar as AND's em avaliações qualitativas - das quais fazem parte a detecção de defeitos e identificação de espécies; e avaliações quantitativas - que visam fornecer valores para propriedades e dimensões dos defeitos da madeira. Os métodos podem se aplicar à árvore ainda viva, ou em algum ponto da cadeia produtiva, incluindo o material já em uso. Agrupando as classificações propostas por Bodig (2000), Zombori (2000) e Szymani e McDonald (1981), as principais aplicações dos métodos para AND na indústria de produtos florestais podem ser assim separadas:

Detecção de defeitos em madeira sólida, laminados e painéis.

Detecção e quantificação de defeitos como nós, inclinação de grã, bolsas de resina, ataques de fungos e insetos, presença da medula, esmoado, rachaduras, desvios dos padrões de dimensões etc. Uma acurada detecção de defeitos pode auxiliar na tomada de decisão durante o desdobro de toras, resserra e laminação, com conseqüentes incrementos no rendimento, grau de padronização e valor dos produtos. Um grande número de técnicas se aplica a estes casos, desde a inspeção visual até a geração de imagens tridimensionais, através de tomógrafos de raios-X.

Segregação ou classificação de produtos para uso estrutural.

As propriedades da madeira e da maioria dos produtos de madeira apresentam variabilidade muito maior do que a que se obtém nos materiais sintéticos como os metais, o concreto e os plásticos. Por isto a resistência e a rigidez dos produtos de madeira devem ser avaliadas individualmente com o objetivo de se classificar cada uma das peças segundo seu potencial de utilização. Tradicionalmente tal classificação é realizada através da inspeção visual. O desenvolvimento de técnicas menos subjetivas veio conferir maior confiabilidade à avaliação. As máquinas de classificação mecânica para a madeira sólida e de propagação de ultra-som para laminados são bons exemplos de evoluções da inspeção visual.

Avaliação *in situ* de elementos estruturais

Níveis mínimos de resistência e rigidez são exigidos para o desempenho seguro de elementos estruturais. Com o avanço do tempo de uso, as estruturas de madeira estão

sujeitas a perder parte de sua capacidade de suportar cargas devido ao ataque de organismos xilófagos ou por suas próprias características reológicas. A inspeção visual de um perito pode não ser suficiente para a avaliação das condições mecânicas de um grande elemento estrutural, o que, muitas vezes, leva à substituição desnecessária de peças que poderiam ainda desempenhar com segurança sua função na estrutura. Métodos baseados na propagação de ondas de tensão são comumente aplicados nestes casos.

Avaliação da sanidade e propriedades da madeira de árvores vivas.

O estado de sanidade em que se encontra a madeira dos troncos e galhos das árvores vivas é especialmente importante para a segurança contra quedas em áreas urbanas, que podem causar danos a construções, redes elétricas e que representam riscos a vidas humanas. A sanidade dos troncos e o monitoramento de propriedades da madeira como resultado de tratamentos silviculturais também fornecem informação úteis para o manejo adequado de florestas, uma vez que delas depende a rentabilidade de um empreendimento madeireiro. O uso de métodos rústicos como a introdução de barras metálicas ou de instrumentos sofisticados como penetrômetros e sistemas de análise de propagação de ondas de tensão são de grande auxílio no diagnóstico de vazios, zonas de baixa densidade, apodrecimento etc. (WANG et al., 2000; YAMAMOTO, 2000).

2.2. Os métodos para a avaliação não-destrutiva da madeira

As várias aplicações para os métodos de AND da madeira descritas anteriormente requerem técnicas adequadas a cada situação. A seguir serão listados os principais métodos AND com uma breve descrição de seus princípios de funcionamento.

2.2.1. Avaliação visual

O mais antigo e talvez aquele que ainda é o mais utilizado método de AND baseia-se na observação a olho nu das condições em que se encontram a madeira ou seus produtos. Naturalmente, este tipo de avaliação apresenta um alto grau de subjetividade e somente possui validade se realizada por peritos extremamente cautelosos e bem treinados. Os principais fatos que auxiliam o perito em uma tomada de decisão são

a presença de rachaduras, deformações excessivas, outros tipos de danos mecânicos e a extensão do ataque organismos xilófagos como podridão e infestações por insetos.

A avaliação visual muitas vezes é auxiliada pelos chamados métodos *pseudo-não-destrutivos* que, segundo BODIG (2000), podem ser:

- Percutir a peça supostamente afetada com algum objeto rígido e pesado para observar o som produzido. Uma sonoridade retumbante, por exemplo, pode indicar a presença de cavidades internas.
- O uso de instrumentos cortantes para remover as camadas superficiais das peças com o objetivo de se verificar a extensão dos danos.
- A remoção de baguetas com trados para examinar danos internos.
- O uso de resistógrafos para localizar as seções que apresentam danos internos e avaliar sua profundidade.

2.2.2. Propagação de ondas de tensão – Testes acusto-ultrasônicos

Os métodos de AND que empregam a análise da propagação de ondas de tensão são os mais empregados e estudados na atualidade. As ondas de tensão na madeira são geradas por impacto ou por vibração forçada. Devido à maior complexidade dos equipamentos utilizados para a indução e análise de vibrações forçadas, o método de produção das ondas por impacto é o mais difundido.

A Figura 1 apresenta um esquema de montagem de um teste de propagação de ondas de tensão em uma amostra de madeira. O elemento indutor pode ser um impacto ou um gerador de vibrações forçadas do tipo eletroímã, por exemplo. O sentido de propagação das ondas pode se paralelo ou transversal às fibras.



Figura 1 - Esquema de montagem de um teste de propagação de ondas de tensão.

Os parâmetros obtidos através dos testes de propagação de ondas de tensão são a frequência de ressonância e a velocidade das ondas. Somente a velocidade de

propagação já fornece informações importantes acerca das propriedades elásticas do material através da expressão (ROSS et al. 2000; KOLLMANN, 1959):

$$E = v^2 \cdot \rho \quad (\text{Eq. 1})$$

em que:

E = módulo de elasticidade;

v = velocidade de propagação

ρ = densidade do material.

A frequência de ressonância, por sua vez, também pode ser utilizada no cálculo do módulo de elasticidade através da expressão:

$$E = \frac{f_r W L^3}{2,46 I g} \quad (\text{Eq. 2})$$

em que:

E = módulo de elasticidade;

f_r = frequência de ressonância (modo fundamental);

W = peso da peça;

I = momento de inércia da peça;

g = aceleração da gravidade (9,8 m/s²).

A Figura 2 apresenta a montagem de um teste para a determinação da frequência de ressonância de uma peça de madeira. A peça é posta em vibração aplicando e removendo-se rapidamente uma pequena carga sobre o centro do vão. A determinação da frequência de ressonância é possibilitada pela análise das oscilações na carga registrada pelo sensor posicionado em uma das extremidades do vão.

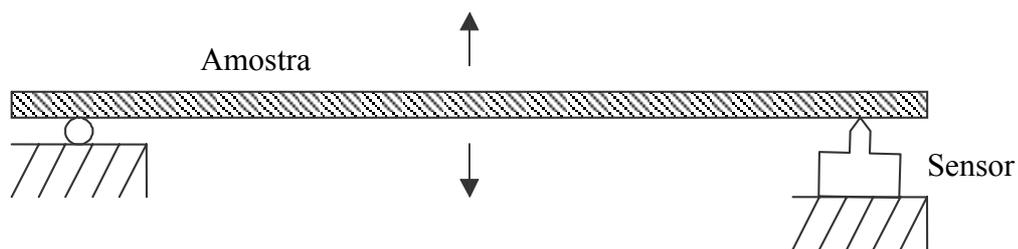


Figura 2 - Esquema de montagem de um teste para a determinação da frequência de ressonância de uma peça de madeira.

Ross et al. (2000) e Burdzik e Nkwera (2002) apresentaram uma série de resultados altamente promissores ao comparar os valores para o módulo de elasticidade obtidos com os métodos de propagação de ondas de tensão e aqueles obtidos pelos ensaios de flexão estática convencional. Os autores, em adição, relataram que o método

de vibrações forçadas (Figura 2) fornece resultados ainda melhores que aqueles obtidos pela análise da velocidade de propagação das ondas geradas por impacto. A capacidade de predição do módulo de ruptura (MOR) que, muitas vezes é o principal objetivo dos testes continua, no entanto, limitada às baixas correlações encontradas entre esta propriedade e o módulo de elasticidade. A aplicação dos testes de propagação de ondas no sentido transversal às fibras ao longo de uma peça pode auxiliar na localização e quantificação de defeitos como nós e rachaduras, informações que podem ser adicionadas aos modelos para melhorar a sua capacidade de predição do MOR (FULLER et al., 1994; SCHAD et al., 1996).

Bekhta (2000) apresentou bons resultados na determinação do MOE em chapas de partículas e MDF empregando métodos de propagação de ondas de tensão, neste caso, ultra-sons. O autor comentou que o método testado pode ser prontamente aplicado na indústria. Uma das vantagens da aplicação do método nos painéis é que estes, diferentemente da madeira sólida, apresentam altas correlações entre MOE e MOR.

O uso de ultra-sons (acima de 20.000Hz) nos testes de propagação de ondas de tensão segue a princípios similares àqueles anteriormente descritos sendo aplicados às vibrações de mais baixa frequência (20Hz a 20.000Hz - som audível). Entretanto, os ultra-sons apresentam a vantagem de poderem ser concentrados em pequenas áreas, conferindo maior resolução espacial às medições. Além disso, os ultra-sons podem ser utilizados em sistemas de eco, ou seja, um sinal é emitido e captado no mesmo ponto, o que simplifica a construção e utilização dos equipamentos (BODIG, 2000; BUCUR, 2000).

Lisette e Ferenc (2000) e Grimberg et al. (2000) apresentaram resultados promissores no uso de ultra-sons para a detecção e delimitação de áreas contendo falhas de colagem em compensados e madeira laminada colada.

Rust (2000) apresentou um inovador sistema de detecção de defeitos internos em toras e árvores vivas utilizando ondas de tensão. Tal sistema pode fornecer imagens de secções transversais ao longo da peça sob análise, à semelhança das tomografias computadorizadas. O sistema baseia-se no cruzamento de informações sobre a velocidade de propagação das ondas obtidas simultaneamente em vários pontos, o que permite a criação de imagens bidimensionais das secções. Os sensores e emissores dos sinais acústicos são presos em torno da peça por uma cinta métrica que possibilita o conhecimento das distâncias entre eles. O autor, apesar de atestar a eficiência do “tomógrafo sônico” na detecção de rachaduras, vazios e podridão pela comparação de

seus tomogramas com resultados obtidos com trados e penetrômetros, fez ainda algumas ressalvas acerca da utilização prática de seu invento. Tais ressalvas incluem a realização de um número maior de testes envolvendo diferentes espécies em diferentes estados de sanidade. O autor sugeriu também que seja investigada a capacidade de detecção da presença de fungos específicos através dos padrões dos tomogramas.

2.2.3. A atenuação de raios-X

Os *raios-X*, também denominados *raios de Roetgen* em homenagem ao seu descobridor Wilhelm Conrad Roetgen (1845-1923), são produzidos quando elétrons muito acelerados atingem um alvo metálico. Os raios-X, assim como a luz visível, o ultravioleta e o infravermelho, são radiações eletromagnéticas e, como tal, suas interações com a matéria obedecem às leis da Mecânica Quântica. Os raios-X apresentam comprimentos de onda entre 0,001 e 10 nm. A quantidade de energia de um fóton de raios-X é muito superior àquela carregada por um fóton de luz visível. Assim, ao atingir um obstáculo, os fótons de raios-X com sua alta energia não são absorvidos com as transições de elétrons para níveis de mais alta energia como ocorre com a luz visível. No impacto, os elétrons dos orbitais externos são arrancados das moléculas do corpo que se tornam carregadas positivamente - ionizadas. Parte da energia dos fótons é absorvida no deslocamento dos elétrons e parte é reemitida na forma de radiação de comprimentos de onda maiores (de menor energia) que, por sua vez, atingirá moléculas de camadas mais interiores do corpo, deslocando elétrons e sendo re-emitida até a extinção total da energia dos fótons que atingiram a superfície. Se a quantidade de energia que incidiu sobre a superfície é suficientemente alta, parte dela poderá atravessar o corpo e ser medida através da sensibilização de filmes fotográficos ou da excitação de foto-sensores eletrônicos. A intensidade com que a radiação original é atenuada (absorvida) é diretamente proporcional à densidade do corpo e à distância percorrida. Este é o princípio da radiografia. O fenômeno de ionização das moléculas atingidas pelos raios-X é o que os inclui entre as *radiações ionizantes* e o que representa o maior risco à saúde humana na utilização de equipamentos que empregam este tipo de radiação (SEARS; ZEMANSK, 1977).

Desde o início do Séc. XX têm-se relatos da utilização da atenuação de raios-X como auxiliar na detecção de danos causados por fungos e insetos, nós e rachaduras

internas em peças de madeira (BROWN et al., 1949). As diferenças de densidade entre regiões de uma peça de madeira são prontamente detectáveis em uma radiografia.

Szymani e McDonalds (1981) comentaram que as radiografias convencionais, isto é, quando são empregados filmes fotográficos, requerem muito tempo para sua preparação e análise, o que torna inviável a sua aplicação na indústria. Sistemas que utilizam monitores de televisão já oferecem informações em tempo real para que um operador de máquinas tome decisões sobre a melhor maneira de se desdobrar uma tora segundo a presença de defeitos e até metais incrustados no lenho. O desenvolvimento de equipamentos de raios-X digitais ampliou a possibilidade da aplicação desta tecnologia em serrarias para tomadas de decisão automáticas em tempo real, via computadores, para otimização do rendimento da operação, atentando-se para padrões de qualidade.

Funt e Bryant (1987), enfocando a ineficiência de radiografias transversais na localização de defeitos no plano da seção transversal de uma tora, investigaram o uso de um tomógrafo hospitalar para a obtenção de imagens radiográficas tridimensionais de toras de *Pseudotsuga menziesii*. A tomografia computadorizada (CT) aliada a *softwares* especializados permite localizar e discriminar com grande exatidão, defeitos como nós, podridões e rachaduras, além das variações de densidade devidas aos anéis de crescimento. Porém, a técnica encontra limitações no seu emprego sobre a madeira verde ou com teores de umidade muito elevados, pois a presença da água eleva a densidade aparente da madeira, reduzindo os contrastes. Os autores relataram uma grande absorção da radiação no alburno de toras verdes devido ao maior teor de umidade tipicamente encontrado nessa porção do xilema, o que cria dificuldade na implementação de algoritmos computacionais para tomadas de decisão automáticas. Wiberg e Lindgren (1996) valeram-se da possibilidade da TC detectar a água na madeira para o monitoramento do gradiente de umidade no interior de toras durante o processo de secagem em estufa.

Schad et al. (1996), empregando um tomógrafo industrial na detecção de defeitos em toras, também se depararam com os problemas gerados pela influência da umidade sobre a densidade da madeira. Os autores concluem que muitos desenvolvimentos, principalmente das técnicas de interpretação dos dados devem ser atingidos antes do emprego comercial da TC em serrarias. Os altos custos dos equipamentos industriais disponíveis no mercado constituem também fortes restrições à disseminação do uso desta tecnologia. Somente indústrias cujos produtos possuem alto valor

agregado, como a aeroespacial, automotiva e eletrônica, são capazes de absorver tais custos.

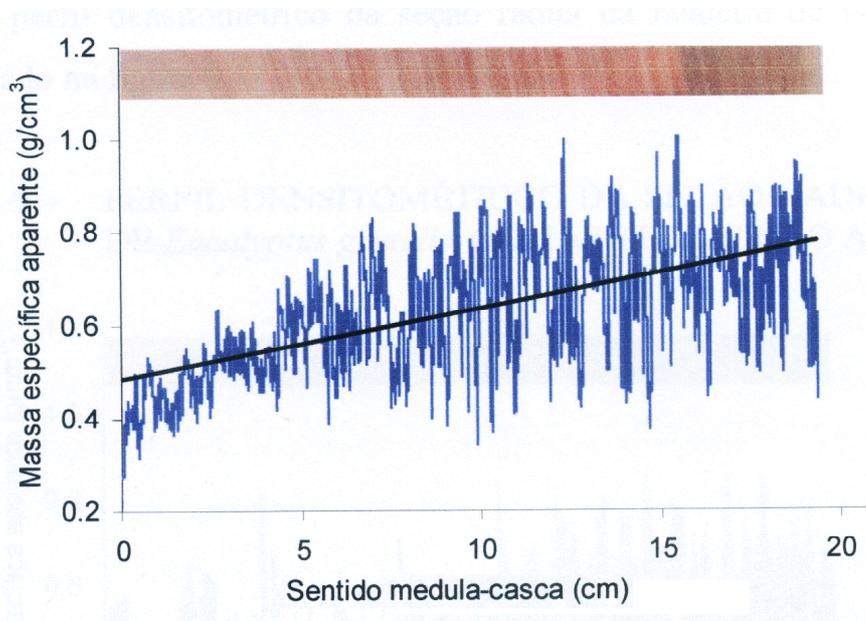
Schajer (2001), utilizando informações acerca da distribuição da densidade ao longo do comprimento e da largura de uma peça de madeira serrada, desenvolveu um método de classificação com base na estimação do módulo de ruptura das peças. A distribuição da densidade ao longo do comprimento das peças no plano da sua largura, é fornecida por um equipamento de raios-X simples, composto por uma fonte de radiação e um detector digital. O autor desenvolveu modelos para a predição do módulo de ruptura de peças de dimensões comerciais que levam em consideração o módulo de ruptura de amostras livres de defeitos e a distribuição da densidade dentro das peças. Tais modelos apresentaram coeficientes de determinação de até 0,73, enquanto que modelos baseados apenas na rigidez das peças serradas apresentaram coeficientes de determinação de 0,55. Os erros padrões da estimativa obtidos com os modelos propostos também foram inferiores àqueles obtidos com os modelos baseados somente na rigidez. O autor atribui a maior eficácia do método baseado na distribuição da densidade às considerações mecânicas embutidas nos modelos – distribuição das tensões nas peças quando em flexão. Equipamentos que utilizam o método desenvolvido pelo autor encontram-se em operação no mercado norte-americano.

Por outro lado, em laboratórios de pesquisa, a densitometria de raios-X tem sido uma ferramenta de grande auxílio no estudo da variação radial da densidade da madeira. Os chamados *perfis densitométricos* são obtidos com grande resolução espacial (Figura 3). Tais estudos podem servir a propósitos de caracterização da densidade da madeira de um povoamento florestal para a obtenção de indicações de processamento e uso, ou para investigações dos efeitos de fatores ambientais e práticas silviculturais sobre a taxa de crescimento da árvore e a densidade da madeira.

Moschler e Winistorfer (2000) apresentaram um relato sobre o histórico do desenvolvimento de densitômetros baseados na atenuação de raios-X e radiação gama. Nos primórdios do seu desenvolvimento, a calibração dos aparelhos era dificultada pela instabilidade das fontes de radiação disponíveis em termos da manutenção do comprimento de onda e da intensidade do fluxo radiante. Equipamentos antigos utilizavam filmes fotográficos na coleta da radiação.

A disponibilização dos dados obtidos com os instrumentos antigos dependia de um processo demorado que tinha início na revelação dos filmes que eram analisados em um outro equipamento – o fotodensitômetro – que avaliava atenuação de um feixe de

luz visível pelo filme. A densidade era estimada comparando-se a atenuação da luz das radiografias da madeira com radiografias de materiais padrão de densidades conhecidas. Equipamentos modernos utilizam foto-sensores que convertem a radiação que atravessa a amostra em impulsos elétricos que, por sua vez, são convertidos em sinais digitais e assim armazenados ou apresentados graficamente em tempo real. Illman e Dowd (1999), relataram sucessos na aplicação de técnicas de microtomografia para a investigação de danos causados por fungos na estrutura celular da madeira. As imagens produzidas com a resolução de 3 micrômetros permitiram quantificar o volume de parede celular destruído e a visualização tridimensional dos danos causados pelos fungos no interior das amostras. A aplicação da técnica em campo ainda é limitada pelos altos custos e pelas grandes dimensões dos equipamentos.



Fonte: Silva (2002).

Figura 3 - Densitograma de raios-X de uma seção radial da madeira de *E. grandis* com 20 anos de idade.

2.2.4. Propriedades elétricas da madeira

As relações entre as propriedades elétricas da madeira e seu teor de umidade são conhecidas desde o início do Séc. XX. A relação direta entre a condutividade elétrica da madeira e o teor de umidade se dá de forma exponencial e cresce rapidamente até a umidade de saturação das fibras, a partir do qual passa a crescer muito lentamente até atingir, na saturação completa, valores próximos à condutividade da

água. O comportamento da constante dielétrica da madeira é bastante similar, ou seja, também apresenta uma relação direta com o teor de umidade (KOLLMANN, 1959; PANSHIN; DE ZEEUW, 1970).

O emprego destas relações na construção de instrumentos para a medição do teor de umidade da madeira talvez tenha sido um dos primeiros aprimoramentos nos métodos AND's para a madeira. Os medidores de umidade que se baseiam na variação da resistividade são chamados xilohigrômetros resistivos, enquanto que aqueles baseados na constante dielétrica são chamados capacitivos. Os medidores resistivos fazem passar uma corrente contínua pela madeira através da diferença de potencial existente entre as pontas de eletrodos que precisam ser introduzidos na madeira. Os capacitivos, por sua vez, operam submetendo a madeira a um campo eletromagnético variável por intermédio de uma corrente alternada que lhe é transmitida por eletrodos dispostos em sua superfície. Os medidores resistivos ordinários disponíveis no mercado são capazes de determinar a umidade de peças de madeira que apresentam entre 5 e 25% de umidade. Os capacitivos operaram na faixa entre 0 e 25%.

Outra aplicação bastante difundida para a constante dielétrica da madeira em avaliações não destrutivas de propriedades baseia-se na sua relação com a direção das fibras. Sabe-se que a constante dielétrica é fortemente afetada pela direção das fibras, sendo 50 vezes maior no sentido longitudinal que no transversal. Assim, a avaliação desta constante pode levar ao conhecimento do ângulo de inclinação da grã, informação muito útil na segregação de peças que serão utilizadas para propósitos estruturais, posto que as propriedades de resistência e rigidez da madeira também estão correlacionadas ao ângulo da grã em relação à direção do esforço. Equipamentos industriais para determinação do ângulo de grã auxiliados por programas de computador encontram-se em plena operação ao redor do mundo. Entretanto, as medições ainda são limitadas a peças que apresentam teor de umidade abaixo do ponto de saturação das fibras (SAMSON et al. 1993; SAMSON, 1988; SAMSON, 1984).

2.2.5. Classificação mecânica

Vários trabalhos têm sido realizados nos últimos quarenta anos com o objetivo de se conhecer a correlação entre o módulo de ruptura (MOR) e o módulo de elasticidade (MOE) da madeira, que é uma propriedade que pode ser conhecida sem a destruição das peças. Os resultados de tais trabalhos levaram ao desenvolvimento de

equipamentos (Figura 5) industriais para a classificação de peças de madeira serrada segundo sua resistência com base na rigidez à flexão (SENFT; DELLA LUCIA, 1977). A classificação de peças de madeira roliça como postes, também se baseia em testes de flexão, mas, neste caso, dada a complexidade das formas que podem assumir, não se busca a determinação do MOE para predição do MOR, mas sim a verificação se uma peça suporta determinada carga.

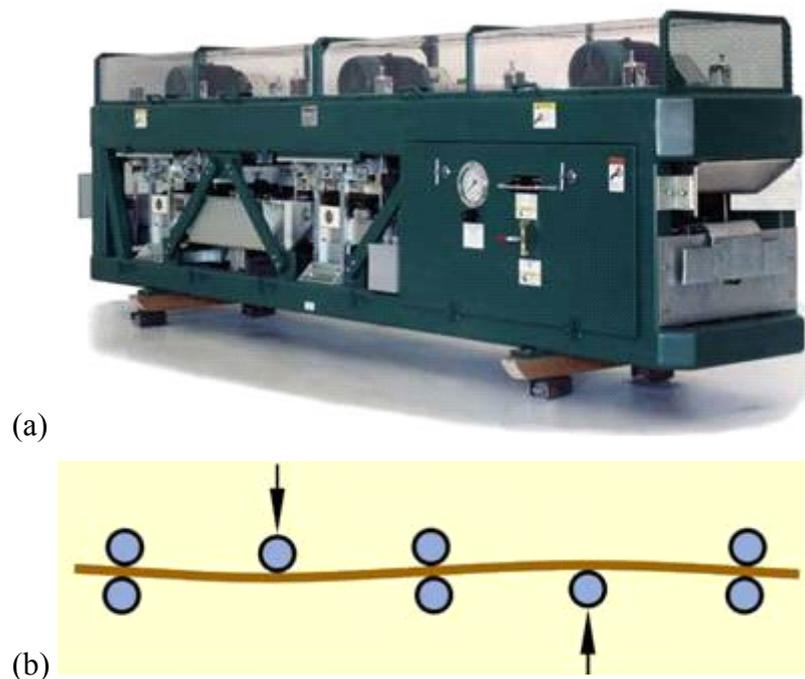


Figura 5 - Máquina para classificação mecânica por flexão de madeira serrada (a). Modelo Metriguard 7200HCLT de fabricação norte-americana. (b) Rolos tratores e aplicadores de carga (setas).

Samson e Hailey (1989) apresentaram uma análise histórica e perspectivas para o mercado das máquinas de classificação mecânica, que, àquela época, completavam 25 anos de existência. Entre 1982 e 1987 foram colocadas 200 novas máquinas no mercado mundial, uma taxa de 50 máquinas por ano. Nos cinco anos anteriores esta taxa encontrava-se em torno de 30 máquinas por ano. Ou seja, até o final da década de 1980 o mercado destas máquinas encontrava-se em plena ascensão. Os autores comentam que a manutenção deste crescimento dependia, àquela época, da inclusão da Alemanha entre os países que adotam o sistema em sua indústria de construção civil, já que este país é o maior importador de madeira da Europa. Ou seja, a inclusão da Alemanha teria reflexos

em todos os países produtores de madeira. A entrada do Japão no mercado também abriria ainda mais o espaço mundial para o uso das máquinas.

A resistência apresentada por alguns países na adoção do sistema mecânico de classificação baseia-se no fato de que a correlação entre MOE e MOR e peças de dimensões estruturais é muito baixa ($R=0,55$, em média), já que o MOR é fortemente influenciado pelos defeitos da madeira. Chui e Smith (1991) ainda verificaram que a inclusão do módulo de rigidez (elasticidade no cisalhamento) aos modelos que contêm apenas o módulo na flexão como variável independente para a previsão do módulo de ruptura, não levam a ganhos significativos em termos de acurácia. FPL (1999) sugere a inspeção visual das peças como auxiliar na classificação mecânica. Porém, este tipo de avaliação não é capaz de fornecer informações objetivas que possam ser incluídas nos modelos para predição do MOR como posição e dimensões dos nós. Sistemas de visão artificial integrados às máquinas de classificação mecânica se apresentam como uma possível solução ao problema, porém implicam em mais custos a um sistema que já é bastante caro. Os autores ainda comentam que a popularização do sistema depende também do desenvolvimento de equipamentos mais simples e de baixo custo.

2.2.6. Propriedades ópticas da madeira

Dentre os métodos baseados em propriedades ópticas da madeira incluem-se os estudos das interações do material com a radiação eletromagnética compreendida na faixa de comprimentos de onda de 300 a 15.000 nm, ou seja, desde o ultravioleta até o infravermelho distante, passando pela janela do visível (400 a 700 nm). Essa radiação pode ser refletida e refratada por materiais sólidos na forma de espelhos, prismas e lentes e compõe a gama das *radiações ópticas* (Swain e Davis, 1978). Os métodos baseados em propriedades ópticas podem ser agrupados em: espectroscópicos, de imageamento, de sensores térmicos e de sensores monocromáticos.

As técnicas de imageamento utilizam informações fornecidas por câmeras fotográficas ou de vídeo para a avaliação de defeitos e planejamento de planos de cortes, o que poderia ser aplicado na otimização da produção em serrarias e fábricas de laminados e compensados e outros segmentos da indústria madeireira. Tais imagens podem ser obtidas a partir de equipamentos que produzem imagens digitais diretamente, ou a partir de equipamentos analógico que terão suas imagens convertidas em imagens digitais as quais alimentam softwares especialistas. Existem vários padrões de imagens,

porém um dos mais utilizados é o padrão RGB. As imagens em RGB (do inglês: *red, green and blue*) são representadas digitalmente como matrizes tridimensionais, nas quais os índices das linhas e colunas representam as posições dos pontos no plano da imagem e as camadas (R, G e B), de maneira simplificada, representam as refletâncias nas bandas entre 400 e 500 nm (B), 500 e 600 nm (G) e 600 e 700 nm (R). Os elementos desta matriz são chamados *pixels* (corruptela do inglês: *picture elements*). Aos pixels são atribuídos valores padronizados variando de 0 a 255. Estes valores são chamados *valores de brilho* dos pixels (GONZALEZ; WOODS, 1993; CAVALCANTI, 2000).

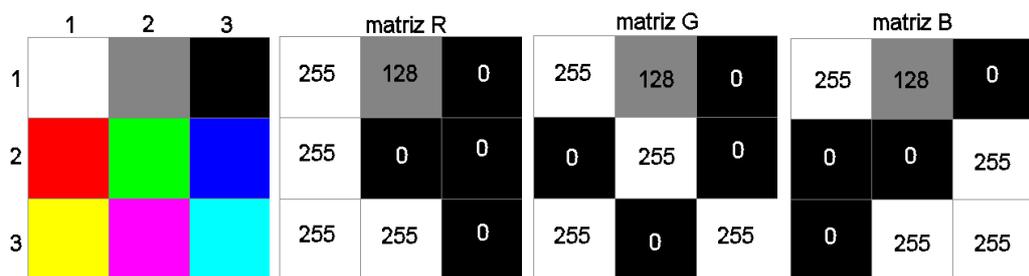


Figura 6 - Imagem colorida e suas componentes R, G e B representadas em tons de cinza com os respectivos valores dos pixels em cada banda.

A Figura 6 ilustra como é realizado o armazenamento de dados em uma imagem digital RGB. As imagens das bandas individuais são mostradas em tons de cinza com os valores de brilho variando de 0 a 255. Os valores de brilho em cada banda indicam a intensidade em que diminutos elementos fosforescentes dispostos na tela do monitor devem se acender para criar a sensação de cor para o observador. Note-se na Figura 6 que o primeiro pixel da primeira linha da imagem colorida apresenta a cor branca que é obtida sobrepondo-se pixels com valores de brilho de 255 nas bandas R, G e B. O segundo elemento da primeira linha apresenta uma tonalidade de 50% de cinza obtida combinando-se os valores de brilho de 128 em cada banda. Finalmente, o terceiro pixel da primeira linha, de cor preta, é resultado da combinação do valor zero em todas as bandas. A segunda linha da imagem colorida da Figura 6 contém as cores puras vermelho, verde e azul, as quais são geradas tomando-se o valor de 255 para cada banda de mesmo nome da cor e zero para as demais. A terceira linha já apresenta cores que são combinações das cores puras duas a duas. Considerando-se o número de 256 valores permitidos em cada uma das três bandas (0 a 255), um total de 16.777.216 ($=256^3$) de cores podem ser obtidas com as combinações.

Assim, tomando-se os valores de brilho em cada banda de uma imagem colorida como uma referência válida para a reflectância naquelas faixas do espectro eletromagnético, podemos assumir que um *scanner* de mesa, câmeras fotográficas e de vídeo podem funcionar como rústico espectrômetros operando na faixa do visível. Nishino e Shinagawa (1997) demonstraram a eficácia de um *scanner* de mesa na obtenção de coordenadas cromáticas XYZ comparando sua performance com a de um colorímetro. Os autores desenvolveram modelos lineares multivariados em que as coordenadas XYZ fornecidas por um colorímetro eram previstas com base nas coordenadas RGB fornecidas por um scanner de mesa ordinário.

A Figura 7 apresenta as imagens coloridas e em tons de cinza para cada banda RGB resultado do escaneamento de uma amostra de madeira de *Eucalyptus grandis*. Na imagem colorida da amostra (Figura 7, no alto, à esquerda) é possível perceber-se na face longitudinal, com uma certa clareza, a presença de faixas escuras correspondentes ao lenho tardio observado em três anéis de crescimento da madeira (melhor vistos no corte transversal). Na imagem correspondente à banda R, entretanto, o contraste entre as regiões de lenho tardio e inicial é muito reduzido, o que significa que a amostra de madeira reflete de forma quase que indiscriminada a luz de comprimentos de onda mais elevados. Na imagem da banda G o contraste volta a ser notado. A banda B, por sua vez, evidencia com clareza ainda maior a diferença na refletância das regiões de diferentes densidades. O gráfico apresentado na Figura 7 revela a variação dos valores de brilho dos pixels de uma linha ao longo da largura na região central do corte longitudinal. Percebe-se que na banda R estes valores permanecem praticamente constantes, enquanto que nas bandas G e B, depressões no perfil (mais notáveis em B) revelam as menores reflectâncias apresentadas pelo lenho tardio nestas faixas do espectro.

Enfim, nas bandas G e B pode-se observar que a resposta espectral varia quando varia a densidade (e outras propriedades) da madeira, o que é explicado pela maior concentração de lignina no lenho tardio, pois, como será mais bem exposto adiante, este componente da madeira é o principal responsável pela absorção da radiação nessa região do espectro (aproximadamente de 400 a 600 nm). Note-se que a maior quantidade de lignina não é implicação de seu mais alto teor e sim da maior concentração da “substância madeira” como um todo por unidade de área irradiada, pois é de conhecimento comum que o lenho tardio possui, em geral, menor teor de lignina que o lenho inicial em relação ao peso seco de madeira (Lin e Dence, 1992).

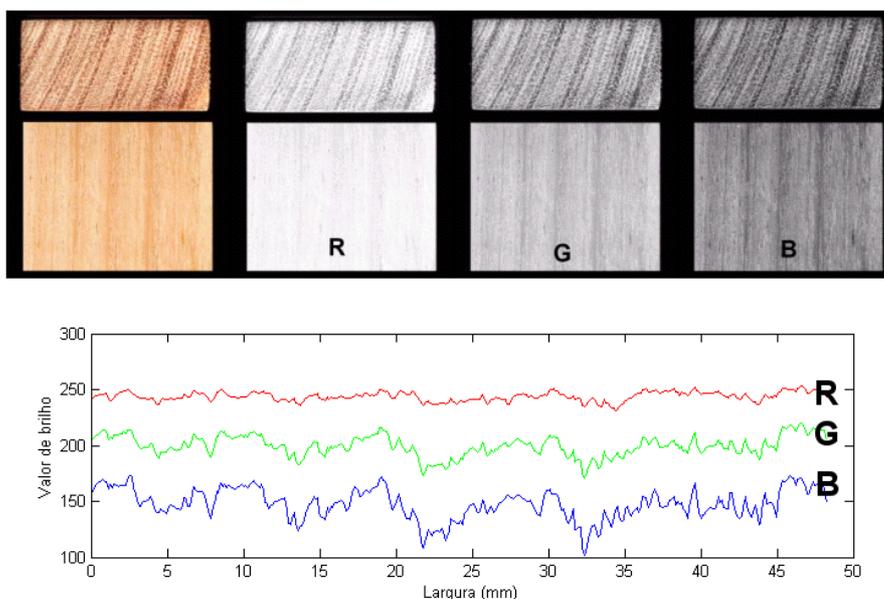


Figura 7 - Imagens coloridas e em cada banda RGB de uma amostra de *E. grandis* em corte transversal e longitudinal aproximadamente radial. O gráfico evidencia os valores de brilho dos pixels de uma linha central ao longo da largura da face longitudinal em cada banda.

Khoury Junior et al. (2005) demonstraram ser possível a identificação de defeitos em madeira serrada de várias espécies de eucalipto a partir de imagens de vídeo coloridas.

Apesar das recentes evoluções nos equipamentos e técnicas de processamento de imagens digitais, não existem ainda no mercado sistemas que, operando em tempo real, apresentem performances satisfatórias em ambientes industriais (PHAM; ALCOCK, 1998; KLINE et al., 1998; BUTLER et al., 2002).

Sensores térmicos são aparatos que utilizam a detecção da radiação no infravermelho para a determinação da temperatura. Esses termosensores possuem ampla aplicação na indústria. A determinação do teor de umidade da madeira acima da umidade de saturação das fibras (USF) por métodos não destrutivos tem sido um problema tecnológico desde o início do desenvolvimento dos medidores de umidade elétricos. Equipamentos empregando microondas, radiação nuclear (nêutrons e raios gama) e raios-X têm sido testados e alguns se encontram em operação na indústria. Porém, tais métodos apresentam sérias limitações devidas, em alguns, à necessidade do conhecimento *a priori* da densidade da madeira, em outros à grande sensibilidade de sua acurácia a variações na temperatura e entre espécies. Ainda, outros requerem grandes investimentos e apresentam problemas de segurança no trabalho dos operários devido à exposição à radiação.

Troughton e Clarke (1987) apresentaram um inovador sistema para a determinação do teor de umidade acima do (USF) com base em propriedades térmicas da madeira. Sabe-se que o calor específico da madeira é afetado positivamente e de maneira quadrática pelo seu teor de umidade mesmo acima da USF (KOLLMANN, 1959). O sistema proposto por Troughton e Clarke (1987), consiste na medição da elevação da temperatura em laminados provocada pela exposição a uma fonte de calor por um tempo fixo. A temperatura do material é determinada antes e depois de seu aquecimento através de sensores de radiação infravermelha. De posse de modelos de calibração que relacionam o calor específico com o teor de umidade e conhecendo-se a quantidade de energia térmica fornecida ao material é possível calcular-se o teor de umidade com grande exatidão.

Dentre os sensores monocromáticos que possuem aplicação na indústria de madeira encontram-se os sensores de presença e medição sensíveis à luz *laser*. Tais sensores são amplamente utilizados em processos industriais para o acionamento de máquinas (sensores de presença da peça) e determinação de conformidade nas dimensões dos produtos (sensores de medição).

Os métodos espectroscópicos são aqueles de uso mais incipiente e também se apresentam entre os mais promissores no desenvolvimento de sistemas de AND para produtos florestais. Independente da faixa do espectro utilizada, a espectroscopia aplicada a propósitos quantitativos baseia-se no ajuste de modelos que têm como variáveis dependentes as propriedades da madeira de interesse e, como independentes, a reflectância ou absorbância em vários comprimentos de onda ao longo da faixa de trabalho. No método AND com base nas propriedades mecânicas da madeira (item 2.2.5.), por exemplo, o módulo de ruptura era previsto com um modelo que possuía apenas o módulo de elasticidade como variável independente. Em espectroscopia, porém, podem figurar centenas ou até milhares de variáveis independentes, de acordo com a necessidade apresentada pelo problema em estudo, até o limite da resolução do espectrômetro utilizado. Ou seja, a reflectância em cada comprimento de onda do espectro representa uma variável independente. Para o ajuste de modelos de tal porte se faz necessário o uso de técnicas estatísticas de análise multivariada e é imprescindível o auxílio de computadores para a realização dos cálculos. A regressão por componentes principais tem sido uma técnica muito utilizada em espectroscopia porque, ao mesmo tempo em que resolve o problema da colinearidade, típico em dados espectrais de alta resolução, realiza uma drástica compressão no número de variáveis independentes a

serem computadas no modelo final. Note-se que a compressão obtida na regressão por componentes principais não significa obrigatoriamente a eliminação de variáveis originais de pouca relevância. Na verdade, são criadas novas variáveis, que são combinações lineares daquelas originais. Essas novas variáveis, que contêm de forma implícita os valores das originais, são hierarquizadas segundo o montante da variância global das originais que carregam e adicionadas uma a uma no modelo até que a inclusão de mais uma delas não tenha efeitos significativos sobre a acurácia (ADAMS, 1995; MARTEN; NÆS, 1996; NÆS et al., 2002; JACKSON, 1991).

A obtenção de um modelo de calibração acurado em espectroscopia é consequência de uma série de procedimentos bem estabelecidos, que tem início na escolha das populações e amostras e é seguida pela inspeção, filtragem e aplicação de transformações matemáticas sobre os dados até a escolha dos modelos e métodos de ajuste e validação. As obras de MARTEN e NÆS (1996) e NÆS et al. (2002) têm sido referidas como excelentes guias para o desenvolvimento de modelos de calibração em espectroscopia. Considerando que nenhum modelo físico-fenomenológico é perfeito, posto que todos lançam mão de premissas simplificadoras, os modelos estatísticos são de grande utilidade quando são desconhecidas as relações causais entre os fenômenos.

Especialmente no que diz respeito às propriedades espectrais da madeira e suas correlações com suas propriedades físicas, químicas e anatômicas, seria um esforço hercúleo e possivelmente infrutífero a tentativa de desenvolvimento de modelos físicos que rejam as relações entre as variáveis, dada a extrema complexidade química e estrutural do material.

Dentro da larga faixa do espectro denominada *radiação óptica*, tem se destacado nos estudos relacionados à madeira e seus derivados, a espectroscopia de reflectância na região do infravermelho próximo, de 1000 a 2.500 nm (NIRS, do inglês: *Near Infrared Spectroscopy*). De fato, a NIRS tem sido intensivamente utilizada para a determinação de propriedades químicas e tecnológicas de interesse para a indústria de papel e celulose ao redor do mundo (MICHELL; HIGGINS, 2002). Empresas do ramo de celulose e papel sediadas no Brasil já possuem seus próprios espectrômetros e têm utilizado a NIRS como auxiliar em seus programas de melhoramento genético e seleção de clones.

Alguns trabalhos encontrados na literatura relatam a realização da NIRS irradiando-se diretamente a madeira sólida na forma de cavacos, de baguetas extraídas com trados ou de corpos-de-prova usinados; em outros, a madeira é reduzida a um pó de

granulometria padronizada (aqui referido apenas como “serragem”). Em todos os casos são obtidos bons resultados nas calibrações, como será visto a seguir.

Schimleck et al. (1999) relataram sucessos no desenvolvimento de um modelo de calibração usando NIRS sobre a madeira na forma de serragem para a predição da densidade básica da madeira de *Eucalyptus globulus* em árvores com 8 anos de idade. Nesse trabalho os autores constataram que a NIRS apresentou uma performance equivalente àquela observada para o densitômetro “Pilodyn”, de uso tradicional em trabalhos de campo.

Raymond e Schimleck (2002) demonstraram que o teor de celulose da madeira de *Eucalyptus globulus* também pode ser determinado através da NIRS sobre serragem com grande acurácia.

Schimleck e Evans (2002), utilizando amostras na forma de baguetas, verificaram que os valores para o ângulo microfibrilar da madeira de *Pinus radiata* determinado através da NIRS são equivalentes àqueles determinados por difração de raios X no SILVISCAN-2. O SILVISCAN-2 é a segunda geração de um equipamento multifuncional desenvolvido especialmente para a investigação de propriedades da madeira pela Organização para Investigações Industriais e Científicas da Comunidade Australiana (CSIRO). O aparelho congrega um densitômetro, um difratômetro de raios X e um sistema de análise de imagens.

Fardim et al. (2002) obtiveram bons resultados em calibrações com a NIRS para predição de propriedades físico-químicas de polpas celulósicas Kraft não branqueadas de *Eucalyptus grandis*. Nesse trabalho, a irradiação foi feita diretamente sobre a polpa na forma de folhas de papel. Marklund et al. (1999) demonstraram que os espectros obtidos irradiando-se os cavacos de madeira geram modelos com a mesma capacidade de predição que outros ajustados sobre os espectros de reflectância obtidos diretamente sobre as polpas de vinte espécies de coníferas.

O emprego da espectroscopia no infravermelho para a predição de propriedades da madeira de interesse para o seu uso na forma de produtos sólidos ou em painéis de madeira reconstituída também tem obtido sucessos. Thumm e Meder (2001), irradiando diretamente a madeira sólida, obtiveram bons resultados em seus modelos de calibração com NIRS para a predição do módulo de elasticidade na flexão de amostras da madeira de *Pinus radiata*. Gindl et al. (2001), também irradiando a madeira sólida, demonstraram ser a NIRS uma excelente maneira para estimar os módulos de ruptura e de elasticidade na flexão e a resistência à compressão da madeira de *Larix occidentalis*.

Lebow et al. (1996) relataram ser possível a identificação de características macroscópicas como nós, esmoados, manchas de fungos, apodrecimento, madeira de compressão, anéis de crescimento e outros em laminados de *Pseudotsuga menziesii* através da VIS. Note-se que esse trabalho apresenta um caráter qualitativo, ou seja, através dos espectros pode-se discriminar um ou outro tipo de característica. Gindl e Teischinger (2002), em um enfoque quantitativo, demonstraram ser possível a determinação do ângulo de grã da madeira de *Picea abies* na forma de peças serradas, com modelos de calibração desenvolvidos a partir da VIS.

Muitos dos trabalhos com a NIR e a VIS acima citados foram desenvolvidos utilizando-se amostras de madeira na forma de serragem, o que impossibilita a extrapolação de seus resultados para aplicações sobre a madeira serrada ou na forma de laminados. E, ainda, aqueles trabalhos em que os espectros foram tomados sobre a madeira sólida ou laminada, apesar de terem sido obtidas boas calibrações, apresentam a limitação de terem utilizado amostras locais. Ou seja, os espectros e as propriedades da madeira foram determinados em pequenas amostras de madeira devido a limitações dos espectrômetros utilizados (pequenas áreas irradiadas). Assim, com uma única irradiação não seria possível obter-se informações acerca da distribuição das propriedades em uma peça de grandes dimensões como no trabalho de Schajer (2001), nem tampouco, seria possível obter-se perfis com alta resolução como aqueles fornecidos por densitômetros de raios-X. Tal fato não reduz o mérito dos trabalhos citados, pois o emprego industrial dos métodos espectroscópicos depende apenas do desenvolvimento de equipamentos adequados, uma vez que a existência de correlações entre as propriedades da madeira e o espectro de reflectância vem sendo, há tempos, comprovada.

Um recente avanço na espectroscopia de reflectância e nos dispositivos de controle de qualidade de produtos e processos tem sido proporcionado pelo surgimento de equipamentos de baixo custo capazes de produzir imagens hiperespectrais (HEIJDEN et al. 2000; HYVARINEN, 1998). De maneira semelhante aos radiômetros utilizados em sensoriamento remoto, porém com resolução espectral muito superior, as câmeras hiperespectrais (também denominadas *espectrômetros imageadores*) fornecem os espectros de reflectância de cada ponto em suas imagens. Os dados são armazenados em arquivos digitais na forma de matrizes multidimensionais. Enquanto os dados gerados por câmeras e *scanners* RGB são dispostos em três matrizes que se sobrepõem para a formação de uma imagem digital colorida, as câmeras hiperespectrais geram arquivos com um número de matrizes definido pela resolução espectral utilizada. Tal

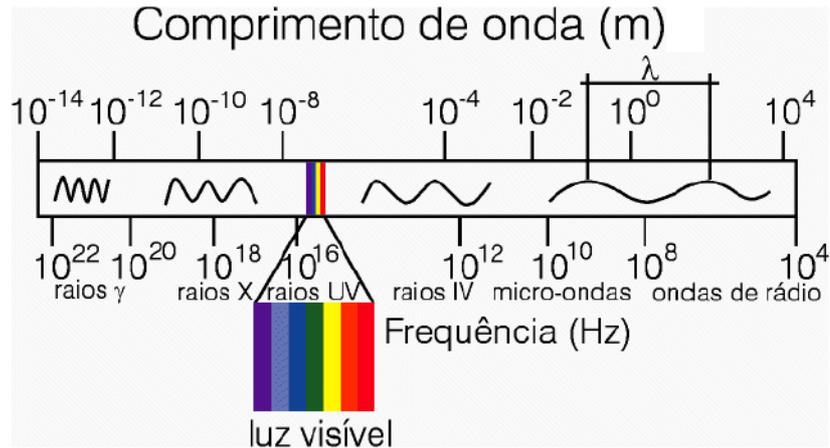
número pode chegar a 200 matrizes quando se utiliza a máxima resolução de alguns desses equipamentos (2 nm). A resolução espacial das câmeras hiperespectrais depende da distância que as separa do alvo imageado. No posicionamento mais próximo da amostra permitido por um desses aparelhos, espectros de pontos adjacentes separados apenas por 10 μm – a mesma resolução geralmente fornecida por densitômetros de raios-X – podem ser obtidos. Tais equipamentos podem preencher as lacunas apresentadas pelas performances dos espectrômetros de irradiação local.

2.3. Princípios da espectroscopia

2.3.1. A Lei de Bouguer – Lambert – Beer

A espectroscopia tem como objetivo revelar o efeito da interação entre a radiação eletromagnética e a matéria, estando esta em estado sólido, líquido ou gasoso, sobre a quantidade e qualidade da radiação que é refletida ou transmitida por um corpo irradiado. É chamada radiação eletromagnética à transmissão da energia na forma de ondas que possuem um componente elétrico e outro magnético – ondas eletromagnéticas. Tais ondas são produzidas pela movimentação de uma carga em um campo magnético (THOMAS, 1996; PERKAMPUS, 1992; SEARS; ZEMANSK, 1977).

O espectro completo da radiação eletromagnética engloba, em ordem crescente de comprimento de onda, os raios cósmicos, os raios gama, os raios X, o ultravioleta, a luz visível, o infravermelho, as microondas e as ondas de rádio cujos comprimentos podem ser da ordem de quilômetros. (Figura 8). Esses nomes indicam gamas de comprimentos de onda separadas por motivos didáticos e práticos, pois o espectro eletromagnético é contínuo, não há diferenças abruptas entre essas formas do mesmo fenômeno físico. Por exemplo, todas se irradiam pelo espaço com a mesma velocidade $c = 2,998 \times 10^8$ m/s (velocidade da luz). As diferenças estão no comprimento das ondas, na sua frequência e na energia carregada pela radiação, o que fazem com que tenham diferentes comportamentos, como o poder de penetração dos raios X ou o aquecimento provocado pelo infravermelho. Uma fonte de radiação, como o Sol, pode emitir luz dentro de uma ampla faixa do espectro. Por exemplo, decompondo-se a luz solar com um prisma é possível ver um espectro de cores, como as do arco-íris. Outras faixas são invisíveis ao olho humano, mas detectáveis por instrumentos.



Fonte: Modificado de Almeida e Santos (2001).

Figura 8 - As várias regiões do espectro eletromagnético.

Espectros de absorção ou reflexão e as razões pelas quais pode ocorrer o fenômeno da absorção da radiação no ultravioleta, visível e no infravermelho próximo (UV-Vis-Nir) são objetos de estudo da Espectroscopia Óptica. Os primeiros estudos espectrométricos foram realizados ainda no Séc. XVII por Pierre Bouguer (1729) e Johann Heinrich Lambert (1760). Estes dois cientistas efetuaram observações independentes sobre as relações entre a intensidade da luz incidente (I_o) e a intensidade da luz transmitida (I_t) por um meio transparente e homogêneo. Eles verificaram que as propriedades associadas ao processo de absorção de luz pelo meio podiam ser enunciadas em termos de duas leis fundamentais:

A intensidade de luz monocromática transmitida (I_t) por um corpo homogêneo é proporcional à intensidade de luz incidente (I_o). Assumindo k como uma constante de proporcionalidade característica do meio tem-se:

$$I_t = kI_o$$

A intensidade de luz (monocromática) transmitida (I_t) decresce exponencialmente com o aumento da espessura (b) do corpo homogêneo. Ou, em termos matemáticos:

$$I_t = I_o e^{-kb}$$

August Beer, somente no Séc. XIX, com base nos estudos de Bouguer e Lambert, demonstrou que a concentração (c) de um soluto pode ser determinada com base na relação entre a intensidade da luz incidente e transmitida por uma solução

contida em um recipiente com um comprimento de caminho óptico a ser percorrido pela radiação igual a “b” (Figura 9).

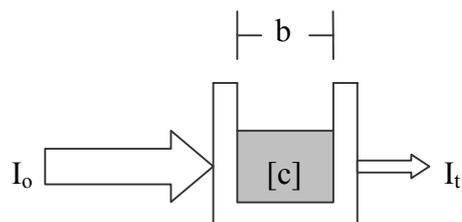


Figura 9 - Radiação de intensidade incidente I_o atravessando uma cubeta com caminho óptico igual a b , contendo uma solução de concentração c e resultando em uma radiação transmitida de intensidade I_t .

Desta maneira, estabeleceu-se o princípio fundamental da espectroscopia óptica que é chamado hoje Lei de Beer e expresso matematicamente por:

$$A = \log\left(\frac{I_o}{I_t}\right) = kcb$$

em que A é a absorvância do sistema. Assim, uma vez conhecida a constante k e o comprimento b , é possível determinar-se a concentração do soluto na solução através da relação entre a quantidade de radiação que incide e aquela que escapa do meio. Assim, conhecendo-se a absorção A através da relação entre a intensidade de luz incidente (I_o) sobre uma solução pura e a intensidade da luz que é por ela transmitida (I_t), para um dado comprimento de onda λ , a concentração do soluto c está associada à um coeficiente de absorção k a ser determinado por regressão linear simples, mantendo-se constante a distância b – determinada na construção do espectrômetro. A presença de outros solutos ou variações nas condições físicas da solução impossibilitam a aplicação direta da lei de Beer-Lambert. No estudo espectrométrico de materiais biológicos que, na maioria das vezes, se apresentam como misturas de diversos compostos em diversas proporções e dispostos em arranjos estruturais variados, técnicas de calibração capazes de compensar as interferências químicas e físicas devem ser utilizadas para a determinação dos modelos para estimação das propriedades de interesse (CIENFUEGOS; VAITSMAN, 2000; ADAMS, 1995; MARTEN; NÆS, 1996; NÆS et al., 2002).

O princípio da espectroscopia óptica pode também ser aplicado à reflectância da radiação incidente sobre um corpo opaco ou translúcido. Ou seja, a maneira com que radiação de variados comprimentos de onda é refletida por um corpo está intimamente relacionada à sua composição química e estrutura física.

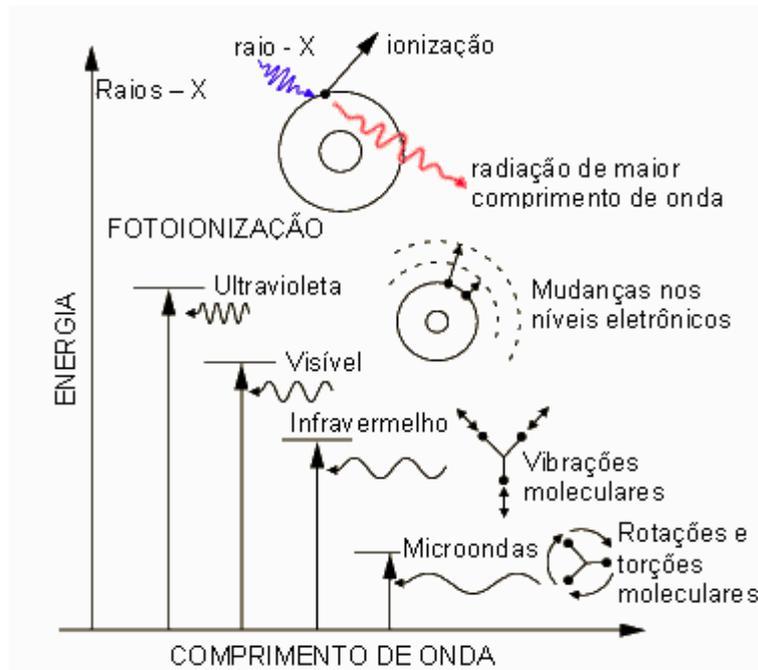
2.3.2. Interações entre matéria e energia

A absorção da energia eletromagnética pela matéria ocorre segundo Mecanismos específicos para cada faixa do espectro (Figura 10). Segundo a equação de Einstein, a energia (E) carregada por um fóton é igual ao produto da frequência da radiação (f) pela constante de Plank ($h=6,63 \times 10^{-34} \text{Js}$):

$$E = fh \text{ (Equação de Einstein)} \quad (\text{Eq. 1})$$

Assim, para uma mesma velocidade de propagação, o comprimento de onda é inversamente proporcional à frequência e, conseqüentemente, à energia carregada pelo fóton.

Radiações de alta energia, como os raios cósmicos, raios-X e raios gama, são absorvidas com o deslocamento de elétrons para fora das moléculas, o que justifica a sua denominação “radiações ionizantes”, pois perdendo elétrons as moléculas tornam-se carregadas eletricamente - ionizadas. Com a saída do elétron ocorre a emissão de uma radiação de menor energia. Tal fenômeno é chamado *dispersão de Compton*. Se a energia que chega ao corpo é grande o suficiente para que a radiação de Compton ainda possua energia na faixa dos raios-X e se o corpo possui densidade baixa o bastante para que nem toda a energia residual seja atenuada, esta radiação reemitida poderá atravessar a matéria e ser captada por sensores ou filmes radiográficos.



Fonte: Modificado de <www.hyperphysics.com>.

Figura 10 - Interações entre matéria e energia ao longo do espectro eletromagnético.

Entre os raios-X e o ultravioleta próximo, a energia da radiação ainda é alta o suficiente para arrancar os elétrons de seus orbitais com a reemissão de fótons de menor energia. Nesta região do espectro ocorre a chamada fotoionização. O potencial mutagênico desta radiação, assim como daquelas de comprimentos de onda menores (raios gama, X e cósmicos), apresenta perigos à saúde humana por seus efeitos carcinogênicos. Tal potencial é utilizado, porém, para a indução de mutações em embriões de sementes em pesquisas de melhoramento genético. Outra utilidade encontrada pelo homem para o poder letal das radiações ionizantes está na esterilização de embalagens, ambientes etc.

A absorção das radiações cujos comprimentos de onda situam-se no ultravioleta próximo (UVA: 320-400 nm), no visível (VIS: 400-700 nm) e na transição entre o visível e infravermelho (tNIR: 700-1.100 nm) dá-se através da transferência da energia dos fótons para os elétrons do alvo que se encontram em orbitais de menor energia, os quais serão deslocados para orbitais mais externos (de maior energia), provocando o chamado *estado excitado da matéria*. Tal absorção mostra-se bastante seletiva e dependente da composição química e do arranjo físico das moléculas do alvo irradiado.

A determinação dos comprimentos de onda no UV-VIS-tNIR a serem absorvidos por um certo composto pode ser realizada através do cômputo das diferenças entre os níveis de energia dos seus orbitais eletrônicos. Para qualquer átomo ou molécula, a diferença entre a energia de seu orbital mais interno ocupado por elétrons e de outro orbital mais externo desocupado indica a quantidade de energia necessária para o deslocamento de um elétron do orbital mais interno para o outro mais externo. A partir da equação de Einstein pode-se então, calcular qual é o comprimento de onda (λ) da radiação eletromagnética capaz de fornecer aquela quantidade de energia (ΔE) por cada fóton. Da Física Ondulatória tem-se:

$$v = f \cdot \lambda \quad (\text{Eq. 2})$$

em que v é a velocidade de propagação, f é a frequência e λ o comprimento da onda. Fazendo-se $v = c$ (velocidade da luz), tem-se:

$$f = c/\lambda .$$

Ainda, fazendo-se $E = \Delta E$ e resolvendo a equação de Einstein (Eq. 1) para λ , tem-se:

$$\lambda = ch/\Delta E \quad (\text{Eq. 3})$$

Assim, a radiação de comprimento de onda λ será absorvida em maior intensidade pela molécula em questão. O cálculo teórico dos níveis de energia dos orbitais moleculares constitui uma das mais complexas tarefas da Mecânica Quântica e baseia-se na *equação de Schrödinger*:

$$\hat{H}\Psi = E\Psi \quad (\text{Eq. 4})$$

em que H é o operador diferencial Hamiltoniano; Ψ são as funções de onda (autofunções) que descrevem a probabilidade de se encontrar os elétrons em certas posições dentro dos orbitais; e E são os autovalores resultantes da aplicação do operador sobre as autofunções e representam as energias dos elétrons. Em realidade, a Eq. 3 é uma simplificação da equação de Schrödinger, pois cada um de seus termos se expande em expressões altamente complexas quando a análise é feita sobre um sistema real.

À medida que aumentam o tamanho da molécula, o número de elementos químicos envolvidos e os tipos de ligações entre os átomos, a complexidade dos cálculos cresce enormemente. Isto porque as equações de onda para os elétrons dos orbitais moleculares (ψ_{molec}) são obtidas a partir de combinações lineares das equações de onda dos elétrons em seus orbitais atômicos (ψ_{atom}) considerando-se as configurações eletrônicas originais dos átomos isolados. Em face da grande especificidade do conhecimento necessário ao seu desenvolvimento, tal enfoque teórico não será aqui apresentado, pois, apesar da sua utilidade para a compreensão das interações entre matéria e energia, é bastante comum que no estudo de moléculas complexas como aquelas encontradas em compostos orgânicos, tal análise seja simplesmente inviabilizada pela enorme quantidade de considerações a serem feitas sobre as condições de contorno do problema. (SOLOMONS, 1992; CAREY; SUNDBERG, 1990).

As absorções no infravermelho próximo, médio e distante (1100 a 15.000 nm) não envolvem a ionização nem tampouco a transferência de elétrons de seus orbitais. Ao atingir um corpo a radiação de baixa energia do IR é capaz de provocar apenas vibrações nas moléculas. Por esta razão, a espectroscopia no infravermelho é também chamada *vibracional*. As ligações em uma molécula podem ser comparadas a cordas com pesos que são os átomos ou radicais presos em suas extremidades. Assim o sistema todo pode vibrar em uma frequência que depende da tensão das cordas – energia da ligação - e das massas dos pesos – massas dos átomos ou radicais. Entretanto, uma absorção vibracional só será observada no espectro infravermelho se a molécula irradiada apresenta um momento dipolo. Quando isso ocorre a substância é dita

infravermelho-ativa. Moléculas simétricas, como o N₂, por exemplo, normalmente não são IR-ativas (COATES, 2000; CLARK; ROUSH, 1984). A radiação de microondas, por sua vez, induz rotações e torções moleculares.

2.3.3. Respostas da madeira à radiação no UV-VIS

A madeira reflete seletivamente a radiação eletromagnética nas regiões espectrais do ultravioleta, visível e infravermelho. Dentro da faixa do espectro (UV-VIS) proposta no presente trabalho, maiores variações na refletância do tecido lenhoso, quando variam também a sua estrutura anatômica e composição química são encontradas no ultravioleta e na região do azul até parte do verde, o que sugere a existência de correlações entre estes fenômenos. Apesar de se tratar de um fenômeno ainda não completamente elucidado, vários autores concordam que a absorção no ultravioleta e no azul-verde deve-se às características estruturais da lignina (DENCE; REEVE, 1996).

Nassau (1985), discorrendo longamente acerca dos fenômenos físicos e químicos responsáveis pela absorção seletiva da radiação observada em vários tipos de substâncias e sistemas ópticos, afirmou que os grupos funcionais que conferem cor às moléculas orgânicas são chamados *chromóforos* (do grego: *chromo*-‘cor’ + *phoros*-‘portador’) e a maioria deles possui em suas estruturas ligações insaturadas, como C=O, C=C ou N=N, em sistemas conjugados. A alternância entre ligações simples e duplas existente entre os carbonos do anel benzênico, presente em grande quantidade na lignina, o caracteriza como um sistema conjugado. Em um sistema deste tipo, os elétrons das duplas ligações estão, na verdade, envolvendo toda a molécula em orbitais moleculares híbridos do tipo π . A Figura 11 ilustra como podem estar dispostas as insaturações no anel aromático. Note-se na figura que as estruturas 1-1 e 1-2 (formas de Kekule) e 1-3, 1-4 e 1-5 (formas de Dewar) são todas plausíveis e contribuem, cada uma, com um certo percentual para a estrutura real do benzeno que é, de fato, uma resultante delas. Por esta razão é preferível a representação 1-6 que simboliza o *híbrido de ressonância* de todas as possíveis formas.

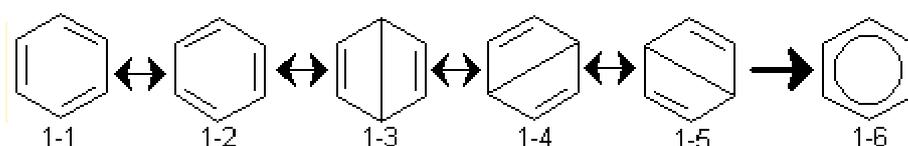
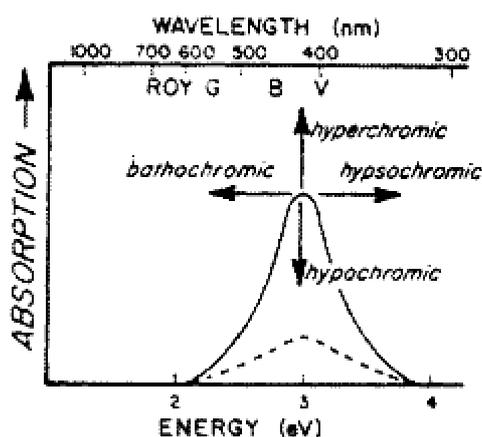


Figura 11 - Representações das possíveis formas para o benzeno.

Sabe-se, porém, que um anel aromático isolado (benzeno) apresenta picos de absorção da radiação eletromagnética situados ainda na região do ultravioleta. Ou seja, é ainda muito alta a energia necessária para a promoção dos elétrons de seus orbitais mais externos para orbitais de maior energia. Porém, se a um cromóforo como o benzeno, estão ligados grupos funcionais como hidroxilas (-OH), carbonilas (>C=O), aminas (-NH₂), halogênios (Cl, Br etc) entre outros, ocorre uma redução na quantidade de energia necessária para a promoção dos elétrons. Neste caso, a molécula poderá absorver a radiação de comprimentos de onda mais altos (de menor energia), como na região do visível, o que conferirá um aspecto colorido à substância. Tais grupos funcionais são chamados *auxocromos* (do grego: *auxién*-‘estimular’) e o seu efeito de redução na demanda por energia para a promoção de elétrons é chamado *batocromismo* [do grego: *bathos*-‘aprofundar’ (neste caso, baixar a frequência ou a energia da onda)]. Por outro lado, existem outros grupos que aumentam tal demanda e são chamados *hipsocrômicos* (do grego: *hypso*-‘altura, elevar’). O conjunto cromóforo + auxocromo é denominado *cromógeno* (do grego: *génos*-‘nascer, originar’). Existem ainda os auxocromos *hipercrômicos* (do grego: *hypér*-‘superior’) que são os grupos que promovem a elevação na intensidade da absorção de um cromóforo e também os *hipocrômicos* (do grego: *hypér*-‘inferior’) que reduzem a intensidade da absorção. Os efeitos dos auxocromos mais comuns no estudo da fotoquímica da cor são apresentados na Figura 12.

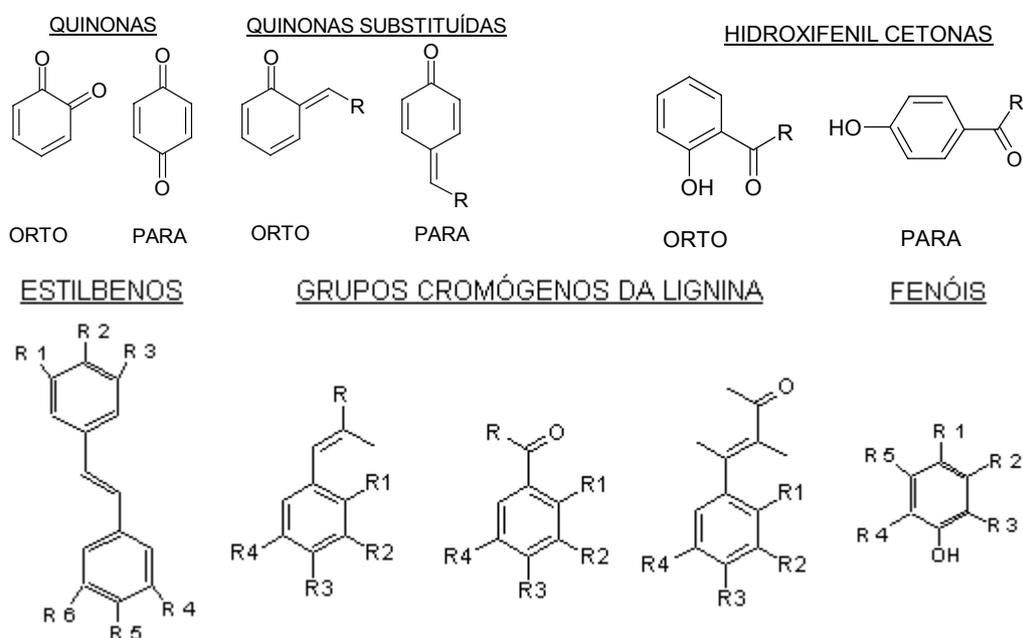


Fonte: Nassau (1985).

Figura 12 - Efeitos de diferentes grupos auxocrômicos sobre a absorção de um cromóforo. R=vermelho; O=laranja; G=verde; B= azul e V=violeta.

A presença dos grupos guaiacil propano ou seringil propano na estrutura da lignina é suficiente para explicar a absorção pela madeira da radiação ultravioleta em

torno de 200 nm. Porém, a absorção no ultravioleta próximo (UVA: 320-400 nm) e no visível (400 a 700 nm), é atribuída a cromóforos como as quinonas substituídas, as hidroxifenil cetonas e cromógenos derivados das unidades básicas guaiacil propano ou seringil propano (Figura 13).



Fonte: Modificado de SENAI-IPT (1984) e Sundqvist (2004).

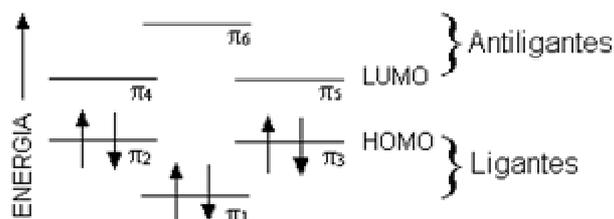
Figura 13 - Grupos cromóforos presentes na madeira associados à lignina. Onde: R, R1, R2, R3, R4 = H, OH, ou substituintes orgânicos.

Note-se que todos os grupos apresentados na Figura 13 possuem em suas estruturas anéis aromáticos ou são deles derivados, como as quinonas. Em acréscimo a tal observação é importante ressaltar que os carboidratos apresentam baixíssima absorção no UV-VIS quando isolados (GRACE et al. 1989; SUNDQVIST, 2004), o que conduz à conclusão de que o caráter aromático da lignina é determinante para a capacidade de absorção da radiação no ultravioleta e no visível, pois são os anéis aromáticos as principais fontes de sistemas conjugados na madeira (STENIUS, 2000; SJÖSTRÖM, 1993; EKLUND; LINDSTRÖM, 1992; SENAI-IPT, 1988). Segundo Müller et al. (2003), 80 a 95% da absorção da luz pela madeira são atribuídos aos cromóforos da lignina. Os extrativos fenólicos como os taninos são também importantes para a absorção no visível e a sua polimerização decorrente do processo de cernificação contribui para o escurecimento do lenho na região do cerne de espécies que apresentam cerne e alburno distintos. Outros extrativos que possuem grupos cromóforos específicos

conferem as colorações típicas apresentadas pelas madeiras de certas essências florestais, por apresentarem absorção em regiões específicas do visível (DENCE; REEVE, 1996; PANSHIN; DEZEEW, 1970; BROWN et al., 1949).

Dentre as substâncias minerais que ocorrem na madeira, os óxidos férricos e ferrosos são os que mais interferem nas propriedades ópticas do lenho (DENCE; REEVE, 1996), pois formam complexos fortemente absorvedores da radiação nas faixas do azul-verde. Entretanto, qualquer madeira, ainda que totalmente livre de extrativos, continuará a apresentar alguma absorção na região do UV-azul-verde, o que corrobora a afirmação de que os cromóforos da lignina são os maiores responsáveis pela absorção da radiação nessa faixa do espectro eletromagnético. Mais especificamente, a cor amarelada apresentada por muitas espécies de madeiras e pelas madeiras em geral quando livres de extrativos, deve-se aos grupamentos terminais aldeídicos da lignina. A absorção desse tipo de estrutura se estende desde o UV até dentro da faixa da luz visível correspondente à sensação do azul, em torno de 400 nm. Como exemplos dessas estruturas podem ser citadas a vanilina e o p-hidroxibenzaldeído (DENCE; REEVE, 1996).

Assim, o anel aromático possui seis orbitais moleculares híbridos do tipo π cujo estado fundamental dos níveis de energia é apresentado no esquema da Figura 14.



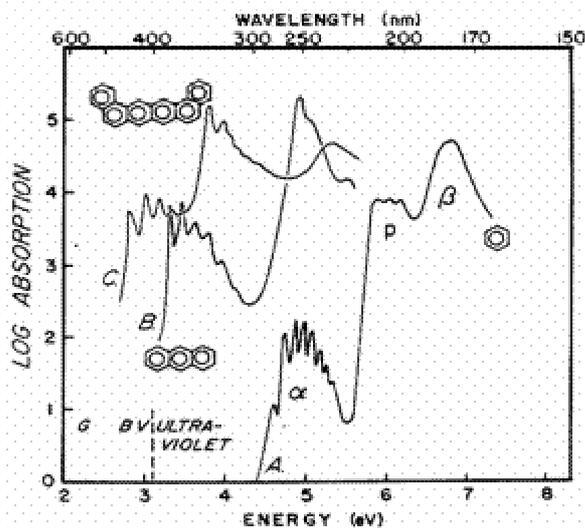
Fonte: Modificado de Nassau (1985).

Figura 14 - Estado fundamental dos níveis de energia dos orbitais moleculares π na molécula do benzeno. As setas nos orbitais indicam os elétrons e os sinais de seus *spins*.

Os orbitais π_1 , π_2 e π_3 apresentam os menores níveis de energia e são chamados ligantes. O termo HOMO (do inglês: *highest occupied molecular orbitals*) é usado para designar os orbitais de maior energia entre aqueles ocupados por elétrons (π_2 e π_3). Os orbitais π_4 , π_5 e π_6 , que na verdade estão vazios, são chamados antiligantes e apresentam os maiores níveis de energia. Os níveis não ocupados de menor energia (π_4 e π_5) são chamados LUMO (do inglês: *lowest unoccupied molecular orbitals*). O espectro de

absorção típico do benzeno apresenta os três picos mostrados na Figura 15 (curva A). Às transições entre π_1 e π_2 ou π_3 , corresponde o pico de absorção de menor energia chamado α e situado entre 225 e 300 nm. O pico entre 190 e 220 nm é chamado p , apresenta nível energético intermediário e corresponde às transições entre π_2 ou π_3 e π_4 ou π_5 . O pico β , bem marcado por volta de 180nm, representa as transições entre os níveis de maior energia π_4 ou π_5 e π_6 .

Se outros anéis são agregados a um anel aromático isolado, ocorre o efeito batocrômico. Ou seja, a demanda por energia para promover um elétron do orbital mais externo é reduzida. Assim, os picos de absorção, antes situados no UV para a o anel isolado, se deslocam na direção das regiões espectrais de maiores comprimentos de onda. Na Figura 14 são também apresentados os espectros de absorção típicos do antraceno - $C_{14}H_{10}$ (curva B) e do 1,2,7,8 dibenzonaftaceno - $C_{26}H_{16}$ (curva C). Os picos α e p , de menor energia, antes distintos no espectro do benzeno, encontram-se fundidos para as moléculas maiores (B e C). Entre os três compostos apresentados na Figura 14, somente o 1,2,7,8 dibenzonaftaceno apresenta-se como uma substância colorida, já que parte de sua curva espectral apresenta absorção na região do visível (violeta e azul). A supressão da reflectância na faixa do azul confere ao composto uma coloração amarelada.



Fonte: Extraído de Nassau (1985).

Figura 15 - Espectros de absorção do benzeno - C_6H_6 (A), do antraceno - $C_{14}H_{10}$ (B) e do 1,2,7,8 dibenzonaftaceno - $C_{26}H_{16}$ (C). Note-se o deslocamento dos picos de absorção em direção ao visível com o aumento do tamanho das moléculas.

O batocromismo resultante do aumento no tamanho das moléculas pode ser explicado pela redução da diferença entre a energia do orbital mais externo e aquela do orbital imediatamente sobrejacente (ΔE). Moléculas maiores (polímeros) sempre absorverão a radiação em comprimentos de onda superiores àqueles em que suas unidades (monômeros) apresentam picos de absorção. Com exemplo, pode-se citar o comportamento do adesivo à base de fenol que, antes de se polimerizar apresenta um aspecto esbranquiçado ou incolor, enquanto que, após a adição do formaldeído em condições adequadas de pH, torna-se um polímero escuro devido a absorção no visível.

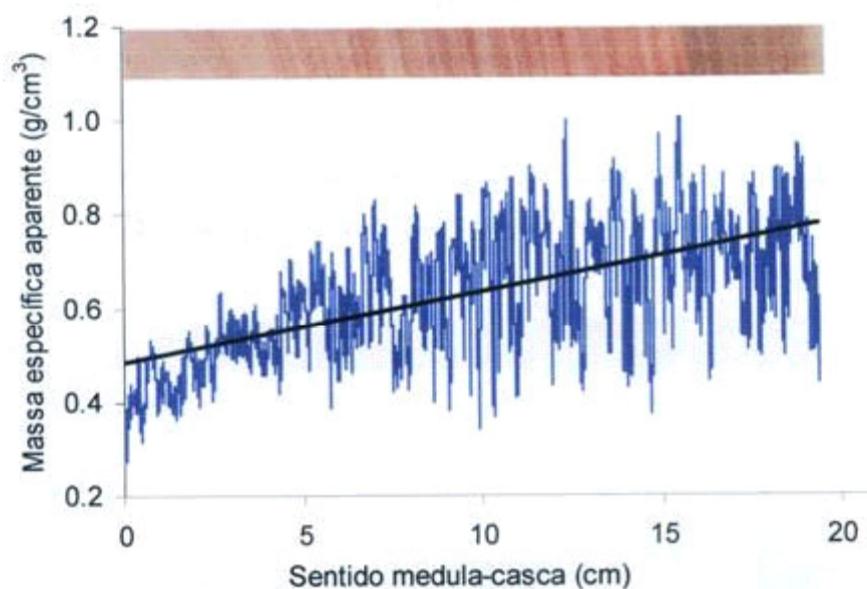
2.3.4. A variabilidade da madeira e a amostragem em espectroscopia de reflectância

As variações nas propriedades da madeira decorrentes da idade da árvore, transição cerne/alburno, transição madeira juvenil/adulta e, ainda, lenho tardio/inicial resultam em variabilidade na resposta espectral de sua superfície. Considerando-se que tais transições implicam em variações simultâneas na composição química e na estrutura anatômica da madeira, modelos de calibração univariados - quando apenas um comprimento de onda é utilizado - podem ter sua capacidade de predição prejudicada devido a interferências não consideradas no momento dos ajustes. No entanto, como a resposta espectral se apresenta como resultante dessas mudanças, os modelos multivariados são capazes de compensá-las intrinsecamente, contando que a amostragem seja suficientemente grande para conter combinações entre os diversos níveis das variáveis. Tal compensação é geralmente necessária e eficiente na análise espectrométrica de misturas de compostos em solução.

Variações no teor de umidade e na cor devido à foto-oxidação das camadas superficiais da madeira também podem gerar interferências na acurácia dos modelos. Mais uma vez, tais variações fazem parte daquelas interferências que são compensadas pela análise multivariada.

Zobel e Buijtenen (1989) apresentaram uma extensa revisão acerca das variações encontradas nas propriedades da madeira, suas causas e controle. Variações que ocorrem, seja dentro do tronco de uma mesma árvore, seja entre árvores de diferentes espécies, seja entre árvores da mesma espécie, são atribuídas à influência de fatores genéticos e ambientais. Assim, como resultado da ação desses fatores na fisiologia do câmbio vascular, responsável pela formação do tecido lenhoso, origina-se o material fibroso e heterogêneo denominado madeira.

Ainda que isolando-se os fatores genéticos responsáveis pelas variações entre árvores e entre espécies, limitando-se o problema à situação em que apenas uma única árvore deve ter determinadas as propriedades de sua madeira, restarão as questões: *Onde se darão as medições? Quais devem ser as dimensões das amostras?* Considerando-se que as variações ocorrem continuamente em qualquer direção do tronco, tais questões se revelam um tanto embaraçosas. Comitês de especialistas ao redor do mundo têm elaborado normas técnicas para a determinação das propriedades da madeira, padronizando-se dimensões de amostras e procedimentos. Entretanto, os resultados obtidos seguindo-se o que é preconizado por tais normas, invariavelmente, expressam médias ponderadas das propriedades que variam mesmo dentro de pequenas amostras. O perfil de densidade da madeira no sentido medula-casca obtido através da densitometria de raios-X mostrado na Figura 16, publicado por Silva (2000), ilustra bem a variabilidade das propriedades da madeira no sentido medula-casca.



Fonte: Extraído de Silva (2002).

Figura 16 - Densitograma de raios-X de uma seção radial da madeira de *E. grandis* com 20 anos de idade.

No perfil da Figura 16 encontra-se para a densidade o valor médio de $0,63\text{g/cm}^3$, enquanto que o valor mínimo é de $0,29$ e o máximo de $0,91\text{g/cm}^3$. Note-se que o valor médio foi calculado com base em todos os quase 20.000 pontos em que se realizaram as medições, já que a resolução espacial do aparelho é de $10\ \mu\text{m}$. Assim, é necessário se associar ao valor da densidade a posição de onde se realizou a medição.

Pode-se supor que, com a densidade, variam a composição química e a estrutura anatômica da madeira. Aliás, seria mais correto dizer que a densidade varia como consequência da variação da sua composição e estrutura. Concluindo, ao se medir a densidade em diferentes localizações do tronco, são obtidos valores que são expressões das combinações entre diversas características da madeira. Aos espectros de reflectância também pode ser aplicado o mesmo enfoque. Ou seja, os espectros são expressões da composição e estrutura da madeira e, sendo assim, devem possuir informações acerca de tais características.

Assim, ao tomar-se os espectros de toda a superfície de uma amostra de madeira, estão sendo ponderadas as variações químicas e estruturais ocorridas dentro daquela pequena porção de material. E, como os valores das propriedades para as quais serão ajustados os modelos de calibração, são também expressões de ponderações, pode-se afirmar que há coerência no estudo das correlações entre tais números. Uma ressalva deve ser feita para o caso dos espectros das faces tangenciais que não podem revelar o que há nas camadas mais profundas das amostras. A Figura 17 ilustra o problema com uma fotografia de um corte transversal de uma amostra de *E. grandis*. Note-se na figura que as faces tangenciais, Tg1 e Tg2, coincidem com camadas de lenho inicial (faixas verticais claras) e tardio (faixas verticais escuras), respectivamente, enquanto que nas faces radiais, os tecidos estão representados de forma ponderada. Ou seja, a média aritmética dos espectros das faces tangenciais não poderá expressar a proporção de lenho tardio/inicial presente na amostra.

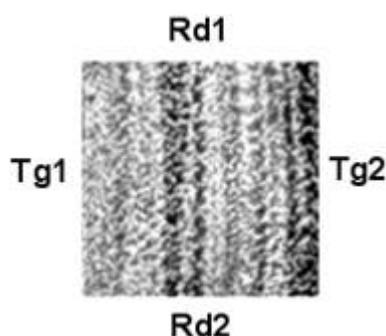


Figura 17 - Imagem do corte transversal de uma amostra da madeira de *E. grandis*. Tg1 e Tg2: faces tangenciais; Rd1 e Rd2: faces radiais.

Cientes de que as respostas espectrais representam expressões ponderadas de uma certa área de uma amostra, resta a importante consideração de que as variações

simultâneas em diversas propriedades da madeira podem resultar em interferências nos espectros de reflectância com sérias conseqüências para a habilidade de predição dos modelos de calibração univariados. A espectroscopia de reflectância da superfície da madeira encontra problemas análogos àqueles presentes na espectroscopia de absorção de uma solução na qual se deseja determinar a concentração de um certo soluto na presença de outros solutos contaminantes. Supondo-se que o soluto em estudo apresenta um pico de absorção em determinado comprimento de onda, porém variam também nas amostras da solução utilizadas para a calibração, a concentração de outro soluto e, ou, um parâmetro físico, como a temperatura, ocorrerão interferências na resposta espectral da solução. Como resultados obter-se-ão espectros que apresentam picos ligeiramente deslocados da posição esperada e suas intensidades não variarão linearmente com a concentração do soluto que se deseja quantificar. Em adição, novos picos podem surgir em diferentes regiões do espectro. Neste caso a lei de Beer-Lambert é subvertida, pois a presença de outros solutos, ou variações nas condições físicas da solução impossibilitam a sua aplicação direta. No estudo espectrométrico de materiais biológicos que, na maioria das vezes, se apresentam como misturas de diversos compostos em diversas proporções e dispostos em arranjos estruturais variados, técnicas de calibração capazes de compensar as interferências químicas e físicas devem ser utilizadas para a determinação dos modelos para estimação das propriedades de interesse. Tais técnicas, usualmente, empregam regressões multivariadas sobre novas variáveis independentes que são combinações lineares das originais, como a regressão por componentes principais (ADAMS, 1995; MARTEN; NÆS, 1996; NÆS et al., 2002).

3. OBJETIVOS

3.1. Objetivo geral

Verificar a existência de correlações entre propriedades físicas e mecânicas da madeira de *Eucalyptus urophylla* e os espectros de reflectância difusa da madeira sólida na região do visível e transições para o ultravioleta e infravermelho.

3.2. Objetivos específicos

a) Desenvolver e testar modelos de calibração multivariados com a técnica de regressão por componentes principais para a predição de propriedades físicas e mecânicas da madeira de *Eucalyptus urophylla* com base na reflectância de sua superfície. As propriedades da madeira para as quais foram ajustados os modelos foram:

- Resistência à compressão paralela às fibras;
- Dureza;
- Densidade básica;
- Rugosidade;

b) Avaliar a influência de filtragens e transformações matemáticas aplicadas nos espectros sobre a acurácia dos modelos.

c) Descrever o comportamento das coordenadas cromáticas da madeira no sentido medula-casca.

4. MATERIAL E MÉTODOS

4.1. A madeira e a amostragem

O trabalho foi desenvolvido com a madeira de *Eucalyptus urophylla* procedente de plantios comerciais nos municípios de Paraopeba e Turmalina, doravante referidos simplesmente como Local 1 e Local 2, respectivamente. Foram utilizadas seis árvores de cada localidade. As árvores foram abatidas e traçadas em toras de 3 m de comprimento que foram imediatamente conduzidas ao Laboratório de Propriedades Físicas e Mecânicas da Madeira do Departamento de Engenharia Florestal da Universidade Federal de Viçosa.

As toras correspondentes aos primeiros 3 m a partir da base do tronco e aquelas imediatamente sobrejacentes (3 a 6 m) foram serradas ainda verdes para a retirada de pranchas diametraais. As pranchas assim obtidas foram deixadas secar ao ar até atingirem o equilíbrio higroscópico quando, finalmente foram usinadas em um desengrosso até a espessura de 2 cm. Neste ponto é importante ressaltar que as superfícies radiais das pranchas não sofreram mais nenhum tipo de usinagem ou acabamento antes da tomada dos espectros.

Das porções intermediárias do comprimento de cada prancha (cerca de 1,5 m da extremidade) foram retiradas baguetas radiais isentas de defeitos, as quais forneceram, finalmente, os corpos-de-prova com as dimensões de 1 cm no sentido radial, 2 cm no sentido tangencial e 2 cm de comprimento para os corpos de prova destinados ao estudos sobre a correlações com a densidade e rugosidade e as dimensões de 2 cm no

sentido radial, 2 cm no sentido tangencial e 3 cm de comprimento para os corpos de prova destinados aos estudos sobre a correlações com a dureza e a resistência à compressão paralela. A Figura 18 ilustra o esquema geral de amostragem nas toras para a retirada das amostras para densidade e rugosidade. Os termos “amostra” e “corpo-de-prova” serão adotados como sinônimos no decorrer deste trabalho.

Os corpos-de-prova assim obtidos receberam uma numeração que identificava a árvore, a posição na altura do tronco e a posição no sentido medula-casca de onde foram retirados. Foram obtidos em média, 9 corpos de prova por bagueta radial, totalizando 256 amostras individuais.

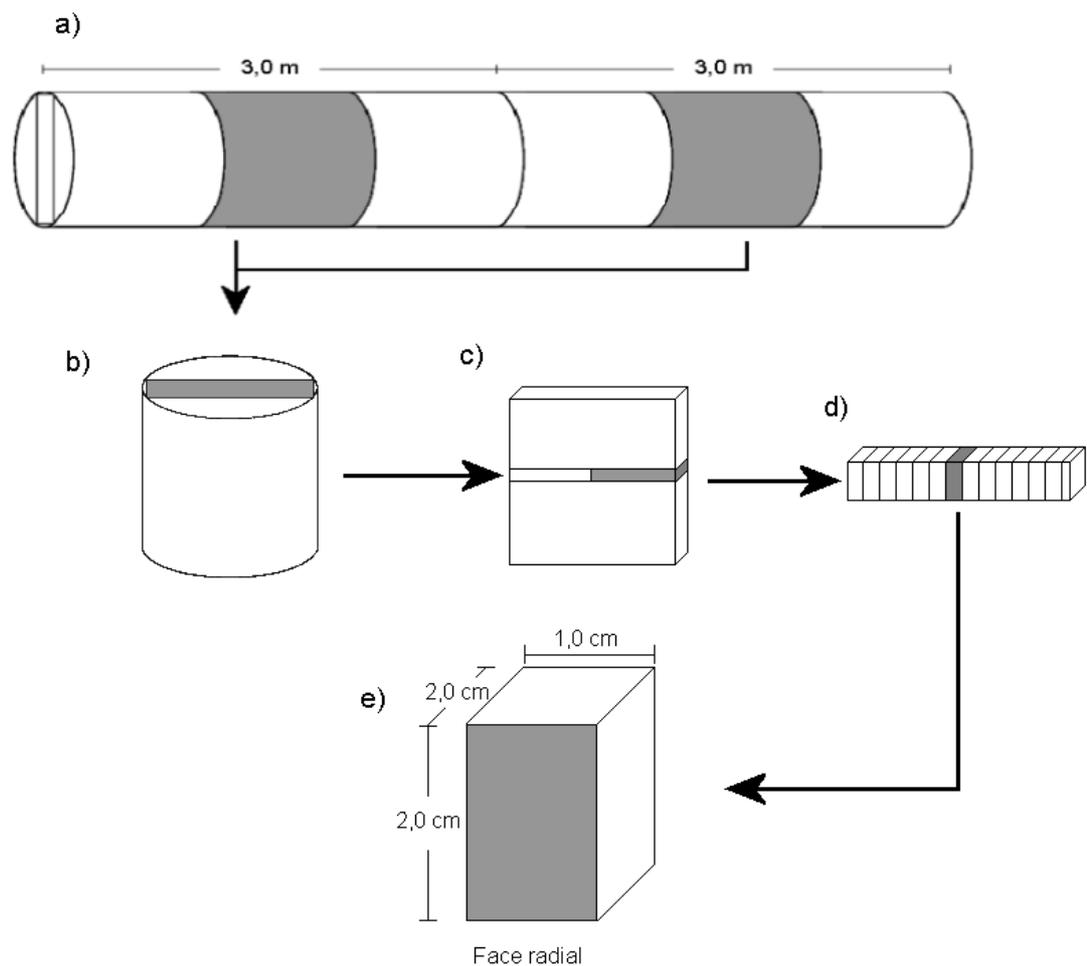


Figura 18 - Esquema geral de retirada das amostras para a retirada das amostras para os estudos sobre densidade e rugosidade. a) Tora evidenciando toretes de 1 m de comprimento; b) Torete evidenciando a prancha diametral; c) Prancha diametral evidenciando a bagueta radial; d) Bagueta radial subdividida em corpos-de-prova com as dimensões de 2 cm x 2 cm x 1 cm.

4.2. A determinação das propriedades da madeira

4.2.1. Os espectros de reflectância

No presente trabalho, foi registrada a reflectância total (difusa + especular) da madeira desde 190 até 1.000 nm em intervalos de 0,30 nm, o que resultou em curvas espectrais contendo 2.700 pontos.

Os espectros na região do visível e transição para o NIR (400 a 1.000 nm) foram tomados por um espectrômetro OceanOptics modelo USB2000 VIS-NIR com o auxílio de uma esfera integradora que possui uma abertura circular para amostragem com 13,5 mm de diâmetro. Tal abertura foi reduzida a um quadrado com 9 mm de lado com o auxílio de uma máscara de plástico polietileno preto e fosco. Os espectros foram tomados em 3 posições de cada uma das 2 faces radiais da amostra. Foram tomados 3 espectros em cada uma dessas posições (3 repetições) e à média aritmética dos valores da reflectância em cada comprimento de onda desses 6 espectros é o que se chama “espectros individuais” ou simplesmente “espectros da amostra” no decorrer deste trabalho. A Figura 19 ilustra o posicionamento das tomadas.

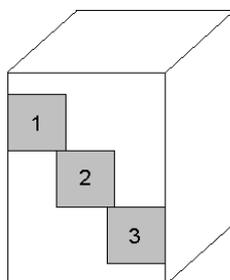


Figura 19 - Corpo-de-prova para os estudos sobre compressão e dureza evidenciando as posições (1-2-3) de tomada dos espectros. Note-se que a outra face radial onde foram tomados outros 3 espectros não pode ser vista nesta representação.

A reflectância no ultravioleta (190-400 nm) foi obtida por um espectrômetro OceanOptics USB2000 UV-VIS com o auxílio de uma sonda de reflectância que consiste de um cabo óptico composto por um conjunto de seis fibras ópticas dispostas circularmente em torno de uma fibra central. As fibras circundantes trazem a radiação desde uma fonte externa até a extremidade da sonda que se abre para a amostra, enquanto que a fibra central conduz a radiação refletida para ao espectrômetro.

Para maiores detalhes acerca dos equipamentos utilizados e sua operação, consultar o APÊNDICE II.

4.2.2. Propriedades físicas e mecânicas da madeira

Algumas propriedades físicas e mecânicas da madeira constituíram o conjunto das 4 variáveis dependentes para as quais foram ajustados os modelos de calibração. Assim, as amostras tiveram determinadas as suas seguintes propriedades:

a) Densidade básica e rugosidade

Cada um dos corpos-de-prova destinados aos testes de densidade e rugosidade tiveram determinados o seu peso e volume quando saturados de água. Após equilibrados em ambiente com umidade relativa de 75% e temperatura de 21 °C, foram tomados os espectros das faces radiais das amostras e foi determinada a rugosidade de sua superfície com o auxílio do rugosímetro marca Mitutoyo Modelo SJ-201P. Foram registrados pelo rugosímetro os valores da rugosidade média (Ra), da rugosidade máxima (Ry) e da rugosidade total (Rt) de cada uma das duas faces radiais das amostras. Em seguida, as amostras foram conduzidas à secagem em estufa para a determinação do peso seco.

b) Resistência à compressão paralela e dureza EQUOTIP

Após a sua retirada das pranchas e equilibradas em ambiente com umidade relativa de 75% e temperatura de 21 °C, as amostras destinadas aos testes de compressão tiveram determinados os espectros de suas faces radiais. Em seguida foi determinada a dureza da madeira com o auxílio de um durômetro de impacto portátil da marca Mitutoyo modelo Hardmatic HH-401. Para o registro da dureza foi utilizada a unidade “Equotip” representada pela letra “L” maiúscula. Tal unidade é típica dos aparelhos de determinação da dureza por impacto que se baseiam no disparo de um elemento metálico impactador contra um alvo (material a ser testado) registrando-se suas velocidades de avanço e retorno (Figura 20).

Imediatamente antes de atingir o alvo, o elemento impactador passa por duas bobinas magnéticas separadas por uma distância conhecida que registram dois pulsos elétricos em um intervalo de tempo que será utilizado para o cálculo de sua velocidade de avanço. Após atingir o alvo, o impactador ricocheteia passando novamente pelas bobinas. Assim, de posse das velocidades de avanço V_a e retorno V_r calcula-se a dureza EQUOTIP com base na expressão:

$$D(L) = 1000 \times V_r / V_a$$

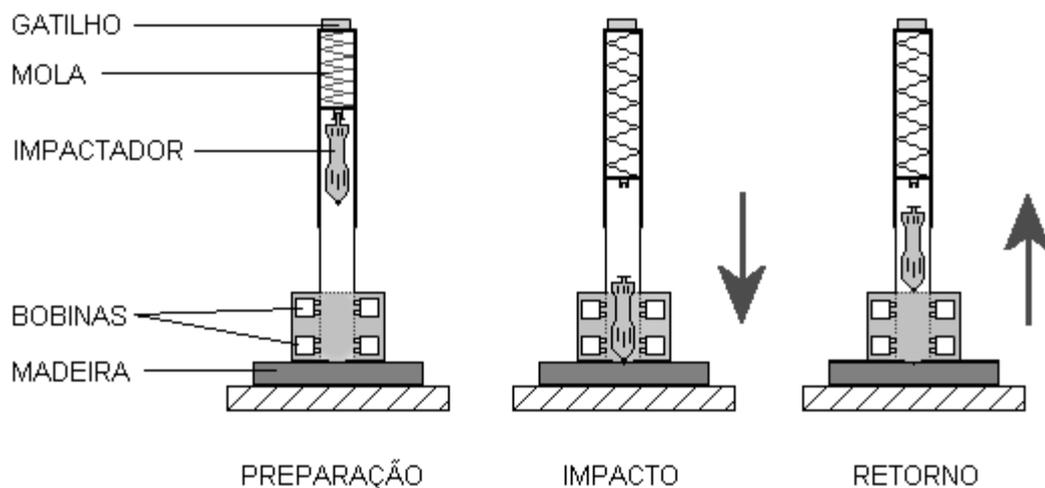


Figura 20 - Esquema de funcionamento de um durômetro de impacto.

Após os testes de dureza, as amostras foram conduzidas à máquina de ensaios universal para a determinação de sua resistência à compressão paralela às fibras e que seguiu o procedimento descrito na norma ABNT MB-26 (ABNT, 1940).

4.3. Pré-processamento dos dados

Antes do processamento propriamente dito, que consiste no ajuste e validação dos modelos de calibração, os dados passaram por manipulações preliminares que consistiram na filtragem e transformações matemáticas dos espectros e na compilação dos resultados na forma de matrizes adequadas para o ajuste dos modelos. O pré-processamento, bem como o ajuste e validação dos modelos, seguirão os procedimentos indicados por Marten e Næs (1996), Næs et al. (2002), Adams (1998) e Scheinost et al. (1998).

4.3.1. Filtragem dos espectros

Com o objetivo de se eliminar ruídos possivelmente presentes quando da sua obtenção, os espectros individuais devem ser submetidos a filtragens. A filtragem foi realizada empregando-se filtros passa-baixas produzidos a partir da técnica de filtragem baseada nos algoritmos de Savitzky-Golay e que consiste no ajuste de modelos polinomiais por mínimos quadrados aos valores da reflectância dentro de pequenos intervalos de comprimento de onda. Uma vez ajustados os modelos, os valores originais de reflectância são substituídos por outros calculados com os modelos. O grau dos

polinômios, bem como a largura dos intervalos, interferem também no grau de suavização das curvas espectrais, que, por sua vez, afetam a performance dos modelos de calibração.

4.3.2. Transformações sobre os espectros

Além dos espectros originais e filtrados, transformações matemáticas sobre os seus valores compuseram o conjunto das variáveis independentes para o ajuste dos modelos de calibração. As funções de transformação empregadas foram:

- a) $R' = (1 - R)^2 / 2R$ Função de Kubelka-Munk;
- b) $R' = \log\left(\frac{1}{R}\right)$ Absorbância;
- c) $R' = \sqrt{R}$ Raiz quadrada da reflectância;
- d) $R' = R_{\lambda} / \sigma_{\lambda}^2$ Divisão pela variância
- e) $R' = (R^b - 1) / b$ Normalização de Box-Cox
- f) $R' = \frac{dR}{d\lambda}$ Primeira derivada;
- g) $R' = \frac{d^2R}{d\lambda^2}$ Segunda derivada;

em que:

R = Reflectância (espectros originais ou filtrados);
 R' = Reflectância transformada (espectro transformado);
 λ = comprimento de onda;
 σ_{λ}^2 = variância da reflectância no comprimento de onda λ ;
 b = parâmetro da regressão de Box-Cox.

Note-se que as derivações foram aplicadas diretamente sobre os espectros (R) e também após as demais transformações.

As transformações a), b), f) e g) são empregadas principalmente para a redução da dispersão dos dados espectroscópicos devida a peculiaridades estruturais apresentadas por cada amostra, enquanto que as transformações c), d) e e) objetivam a normalização dos dados. A normalização de Box-Cox consiste no ajuste de modelos exponenciais entre os dados a serem normalizados e uma curva de distribuição normal com variância e média iguais às originais.

Foram verificados os efeitos de cada transformação sobre a acurácia dos modelos de calibração quando da sua validação.

4.3.3. Compilação dos dados

Nesta etapa foram montadas as matrizes para o cômputo dos coeficientes de regressão dos modelos de calibração. Uma compilação inicial resultou em uma matriz de dados contendo 2612 colunas, representando a reflectância em cada comprimento de onda e que compuseram as variáveis independentes X originais, por 252 linhas que representam as amostras. A matriz das variáveis independente, por sua vez, continha 252 linhas por 4 colunas.

As matrizes listadas acima serão repetidas para cada tipo e grau de filtragem e para cada tipo de transformação matemática a ser aplicada aos espectros. Adicionando-se as variáveis dependentes, as matrizes apresentaram o aspecto apresentado no Quadro 1.

Neste ponto, é importante notar que o número de variáveis independentes variou em função da resolução adotada no momento do ajuste, pois o efeito deste parâmetro também constituiu objeto de estudo do presente trabalho.

Quadro 1 - Aspecto geral das matrizes de dados no momento do ajuste dos modelos

	Variáveis independentes (reflectância ou suas transformações)					Variáveis dependentes (propriedades da madeira)			
	x 190 nm	x 190,31 nm	...	x 1000 nm		y1	y2	y3	y4
Amostra-1
Amostra-2
...
Amostra-252

4.4. Ajustes dos modelos de calibração e avaliação dos resultados

Os modelos de calibração foram ajustados sobre dois terços do número total de corpos-de-prova (amostras de calibração). O outro terço foi utilizado para a validação desses modelos (amostras de validação). Uma série de modelos foi ajustada adotando-se amostras de calibração procedentes de uma mesma localidade, sendo esses modelos validados com as amostras restantes da mesma localidade. Em outros momentos, ajustes foram feitos tomando-se amostras de um local e validando-se os modelos com amostras de outro. Ainda, outra série foi ajustada sobre amostras procedentes das duas localidade

e a validação realizada com as amostras restantes das duas árvores. A Figura 21 ilustra o procedimento de distribuição das amostras.

Tais séries de ajustes foram realizados tomando-se aleatoriamente por centenas de vezes as amostras que comporiam os grupos de calibração e validação dos modelos (Figura 28). A cada uma das séries chamou-se aqui “rodada”. A seleção de grupos de amostras que efetivamente vieram a compor os dados de calibração e validação em cada rodada baseou-se nas seguintes restrições:

As rodadas deveriam conter combinações de amostras diferentes entre si;

Os valores da variável dependente no grupo de validação deveriam estar contidos no intervalo entre os valores máximos e mínimos desta variável para o grupo de calibração. Tal procedimento foi adotado no intuito de se evitar a extrapolação do domínio dos modelos no momento da validação.

A média e o desvio-padrão dos grupos de calibração e validação não deveriam diferir em $\pm 10\%$.

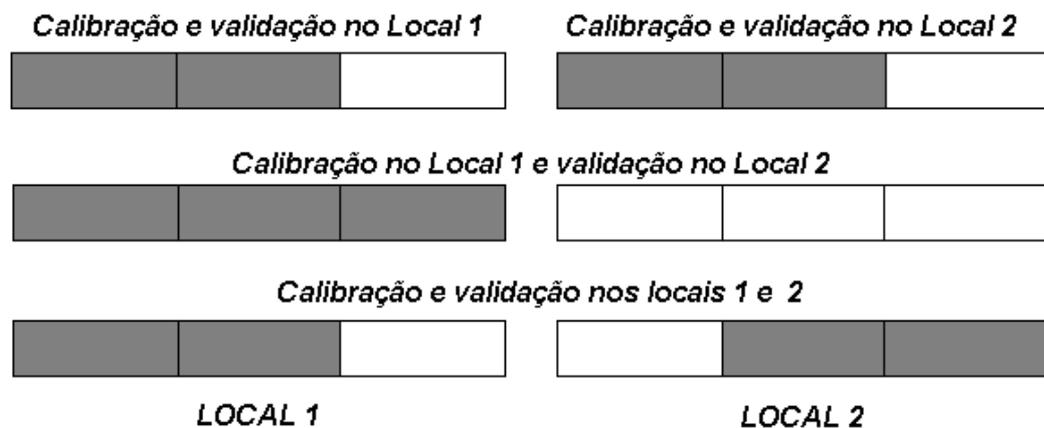


Figura 21 - Esquema de distribuição das amostras para ajuste e validação dos modelos. As áreas em cinza representam as amostras de calibração e, em branco, as de validação.

Assim, todos os modelos em estudo foram ajustados e validados considerando-se estas três distribuições de amostras. Ou seja, foram obtidos e testados três grandes grupos de modelos.

Constituíram a validação dos modelos, análises que revelaram a sua acurácia, bem como a realização de comparações entre esses parâmetros advindos de diferentes modelos. A acurácia ou exatidão dos modelos foi avaliada através da raiz quadrada do erro quadrático médio da predição (RMSEP) definida por:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

Nas equações acima:

\hat{y}_i = valor estimado de uma propriedade y da madeira para a amostra i ;

y_i = valor real da propriedade y da madeira para a amostra i ;

n = número de amostras utilizadas na validação do modelo.

Quanto menor a RMSEP, mais exato é o modelo.

4.4.1. Regressões lineares simples e múltiplas por MQO

Foram ajustados e validados, pelo método dos mínimos quadrados ordinários, modelos lineares simples da forma:

$$\hat{y}_p = \hat{\beta}_0 + \hat{\beta}_1 x_\lambda + \varepsilon$$

e múltiplos da forma

$$\hat{y}_p = \hat{\beta}_0 + \sum_{k=1}^n \hat{\beta}_k x_\lambda + \varepsilon,$$

em que y_p representa uma propriedade p da madeira e x_λ a reflectância no comprimento de onda λ ou suas transformações.

4.4.2. Regressão múltipla por componentes principais

Para cada matriz de dados foram computados os componentes principais das variáveis independentes e os escores dos componentes principais. Os escores computaram os conjuntos das variáveis independentes de novas matrizes sobre as quais foram realizadas as análises de regressão múltiplas por mínimos quadrados segundo o modelo:

$$\hat{y}_p = \hat{\beta}_0 + \sum_{k=1}^n \hat{\beta}_k R_k + \varepsilon$$

em que:

y_p = propriedade “p” da madeira;

n = número de componentes principais utilizados no modelo;

R_k = escores do componente principal “k”;

$k = 1 \dots n$;

ε = erro aleatório.

Foi avaliado o efeito do número de variáveis independentes utilizadas nos modelos sobre a acurácia e a precisão. Os modelos foram ajustados, ora adicionando-se

os componentes um a um a partir do primeiro – aquele que carrega a maior parte da variância dos dados originais, ora pelo método *forward*, no qual as variáveis eram adicionadas de acordo com a correlação que guardavam com as variáveis dependentes e com os resíduos dos modelos que continham o número imediatamente inferior de variáveis independentes.

4.4.3. Esquema geral de processamento dos dados

Para o processamento de todos os dados utilizados neste trabalho foi preparado um algoritmo para a plataforma MATLAB[®] que possibilita várias combinações entre as condições de análise, quais sejam, filtragem dos espectros, redução da resolução espectral, transformações matemáticas e métodos de regressão utilizados no ajuste dos modelos (APÊNDICE A). Os resultados obtidos pelo algoritmos em certas operações-chave como a determinação das matrizes de correlação e de covariância, dos autovetores e autovalores, por exemplo, foram comparados àqueles fornecidos por *softwares* estatísticos consagrados, como o STATISTICA e o SAS, confirmando seus valores. A Figura 22 apresenta um fluxograma para o esquema geral do processamento dos dados.

O algoritmo foi construído utilizando-se, em alguns casos, funções oferecidas pela plataforma MATLAB[®]. As principais funções utilizadas nos algoritmos foram:

$\text{cov}(X)$ → fornece a matriz de covariância das variáveis contidas na matriz X ;

$\text{pcacov}(C)$ → fornece a matriz dos autovetores, os autovalores e o percentual da variância carregado por cada componente principal da matriz de covariância C ;

$X \backslash y$ → fornece o produto $X^{-1}y$ que representa o vetor dos parâmetros estimados para o

$$\text{modelo } \hat{y} = \hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k x_k + \varepsilon \text{ por MQO.}$$

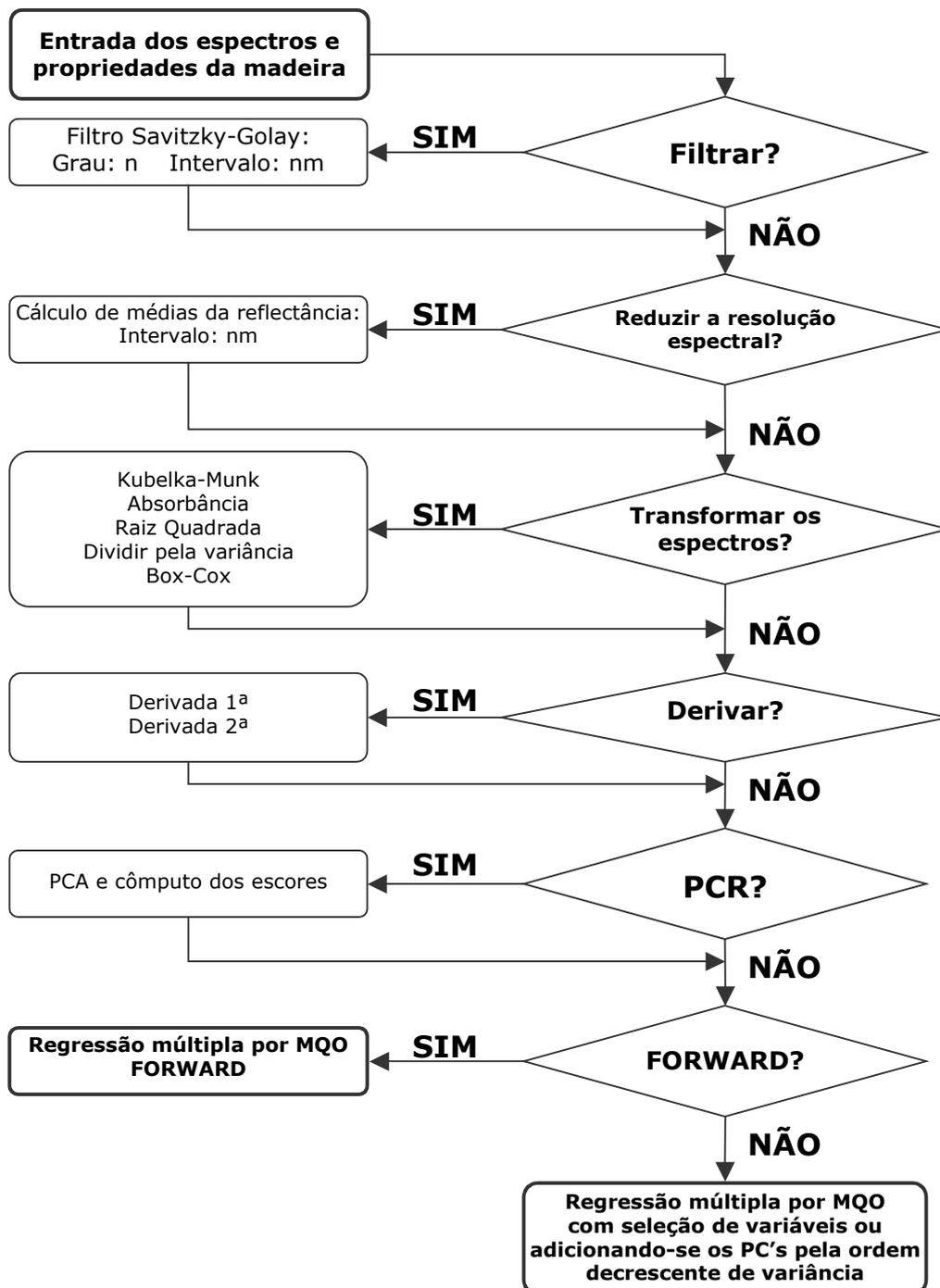


Figura 22 - Fluxograma para o algoritmo computacional elaborado para o processamento dos dados.

5. RESULTADOS E DISCUSSÃO

5.1. Colorimetria quantitativa

Dentre as propriedades da madeira a serem consideradas para a verificação de sua adequação ao uso na indústria moveleira, a cor merece lugar de destaque. Segundo Gonçalves e Macedo (2003), o impacto visual causado pela cor sobrepõem-se aos que são causados pela percepção dos demais atributos estéticos como padrões anatômicos e a própria forma do produto. Isto porque as tendências da moda em projetos de móveis estão sempre alternando entre cores diferentes, impondo preferências que limitam o uso de madeiras que encontram-se fora do que é exigido pelo mercado em um determinado momento.

Além das questões relacionadas aos modismos, a homogeneidade da cor da madeira em um lote a ser utilizado industrialmente é um quesito básico de qualidade, posto que refletirá na homogeneidade dos produtos resultantes de seu beneficiamento. A necessidade do uso de pigmentos e outros tipos de acabamento e suas dosagens são também determinados a partir do conhecimento dos padrões de cor identificados na matéria-prima.

Assim, torna-se clara a necessidade da determinação objetiva da cor da madeira e sua variabilidade como instrumento de tomada de decisão sobre a sua utilização como matéria-prima de um segmento industrial que deve sempre atentar para o valor estético de seus produtos.

Uma vez obtidos os espectros de reflectância da madeira, estes foram submetidos à análise matemática com o auxílio de um software especialmente desenvolvido para a determinação das coordenadas cromáticas $L^*a^*b^*$ de acordo com o sistema CIELab (*Comission Internationale de l'Eclairage, 1976*). Os cálculos basearam-se na conversão dos espectros através das equações tristimulares. Também foram determinados os parâmetros colorimétricos cromaticidade (C) e ângulo de tinta (h^*).

A identificação das amostras segundo a localidade, tora, árvore e posição medula-casca permitiram o conhecimento e a comparação dos valores das coordenadas cromáticas entre locais, árvores e entre posições dentro do tronco. O Quadro 2 apresenta as médias e os desvios-padrões para as coordenadas $L^*a^*b^*$, e para C e h^* discriminando-se a localidade de procedência e a posição medula-casca das amostras. As amostras retiradas da terça parte do raio mais próxima à medula foram agrupadas e designadas com a letra M; aquelas do terço intermediário são representadas por I e, finalmente, aquelas da terça parte mais externa do cerne são agrupadas sob a letra E.

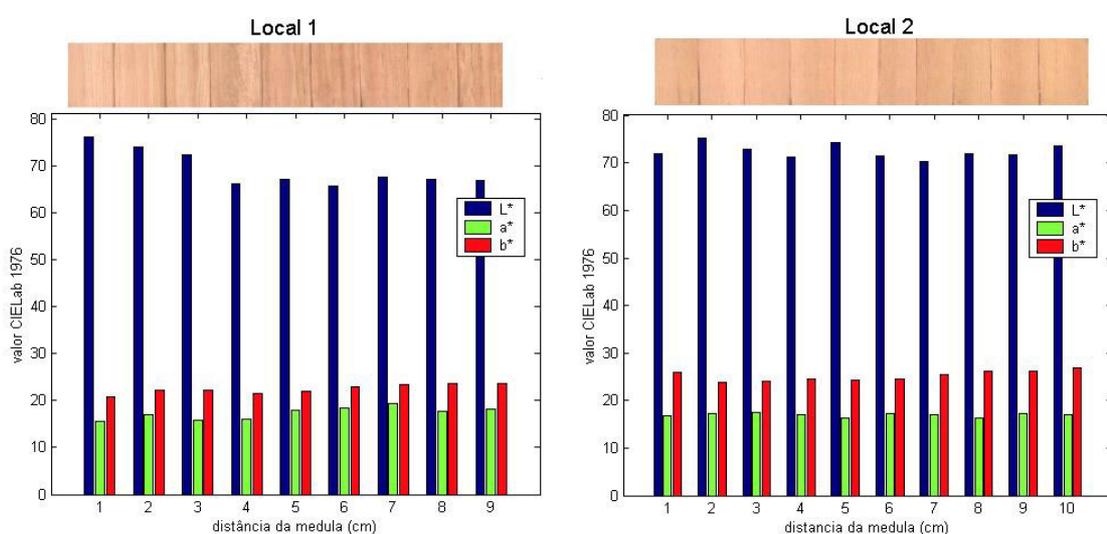


Figura 30 - Variação das coordenadas cromáticas $L^*a^*b^*$ a cada centímetro a partir da medula da madeira de uma árvore típica de cada localidade. As fotografias são das amostras que geraram as coordenadas e estão posicionadas de maneira a, aproximadamente, coincidir a posição real com o eixo medula-casca.

A análise apresentada no Quadro 2 revela que no Local 2 houve uma menor variação dos parâmetros colorimétricos considerando-se as posições dentro do tronco, pois as médias para L^* , a^* e h^* não diferiram, enquanto que para a madeira procedente do Local 1) somente para o parâmetro h^* houve igualdade entre as médias. Com base

nesses resultados pode-se afirmar que a madeira do Local 2 é mais homogênea que a do Local 1. As variações nas coordenadas cromáticas no sentido medula-casca, tomando-se médias a cada 1 cm, para uma das árvores de cada localidade são mostradas na Figura 30. Entretanto, a necessidade do uso de pigmentos para lograr-se um aspecto homogêneo nos produtos obtidos a partir da madeira do Local 1 deve ser estudada utilizando-se métodos mais subjetivos como a construção e avaliação visual de móveis e outros artefatos, pois, talvez os testes estatísticos sejam por demais rigorosos.

Quadro 2 - Médias e desvios-padrões para L*a*b*, C e h*

Local	Posição	L*		a*		b*		C		h*	
		Média	Desvio-padrão								
Local 1	M	72,20 ¹	2,93	16,82 ¹	1,19	22,57 ¹	1,14	19,91 ¹	0,98	55,84	1,92
	I	69,12 ²	2,75	17,65 ²	1,20	23,28 ²	1,02	20,67 ²	0,96	55,35	1,66
	E	68,58 ²	1,62	18,35 ³	1,02	23,90 ³	1,28	21,32 ³	0,96	54,96	1,81
	Média	69,97	2,44	17,61	1,14	23,25 ¹	1,15	20,63 ¹	0,97	55,38 ¹	1,80
Local 2	M	69,64	4,35	17,14	1,75	23,45 ¹	1,68	20,55 ¹	1,58	56,44	1,92
	I	70,53	2,66	17,96	2,12	23,80 ¹	1,15	21,11 ¹	1,41	55,58	2,74
	E	70,65	2,77	18,03	2,21	25,07 ²	1,38	21,87 ²	1,39	56,95	3,34
	Média	70,27	3,26	17,71	2,03	24,11 ²	1,41	21,17 ²	1,46	56,33 ²	2,66
Média geral		70,10	3,12	17,65	1,66	23,62	1,46	20,87	1,34	55,78	2,31

Dentro de uma mesma localidade ou comparando-se médias de localidades diferentes, as médias seguidas por índices sobrescritos diferentes são diferentes ao nível de 5% de probabilidade pelo teste *t*.

Em ambas as localidades foram detectados aumentos nos valores de a*, b* e C e reduções nos valores de L no sentido medula-casca (Figura 30). Tal comportamento deve-se à intensificação da cor da madeira à medida em que as amostras afastam-se da medula. Ou seja, há uma maior absorção geral e seletiva da luz pela presença de maior massa de madeira por unidade de área, já que a densidade também aumenta em direção à casca.

No que diz respeito à variação entre localidades, pode-se dizer que esta se mostrou bastante discreta. Apesar de terem sido detectadas diferenças estatisticamente significativas para os parâmetros b*, C e h*, os valores das médias não são discrepantes a ponto de suscitar algum tipo de segregação. Porém, mais uma vez, a necessidade do uso de agentes homogeneizadores deve ser investigada na prática.

5.2. Densidade básica

Os detalhes acerca da influência dos métodos de regressão e do pré-processamento dos dados nos resultados para os modelos de calibração serão apresentados somente para densidade básica, por ser esta a propriedade que forneceu as melhores possibilidades de ajuste. Para as demais propriedades, tais detalhes serão, às vezes, omitidos.

A variação dos valores para a densidade básica da madeira no sentido medula-casca pode ser observada na Figura 31. Os valores médios e a variabilidade, dentro de cada local e entre locais, são mostrados no Quadro 3.

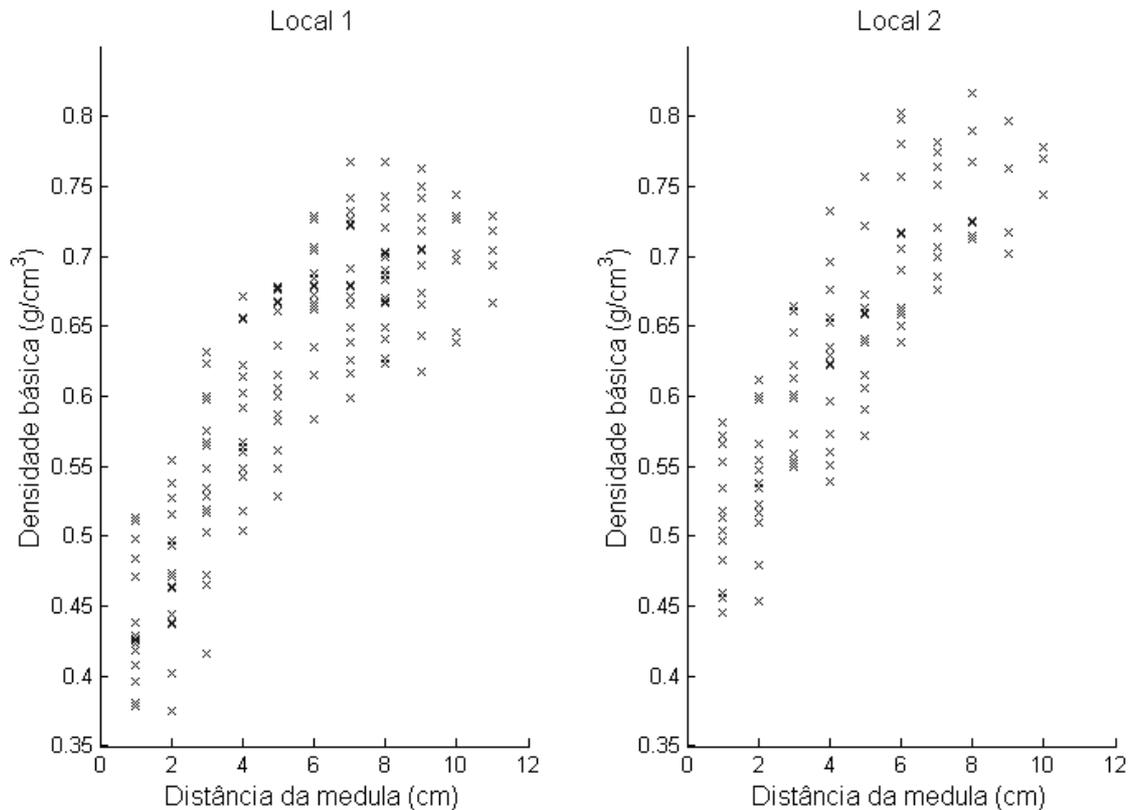


Figura 31 - Variação da densidade básica para a madeira de *E. urophylla* no sentido medula-casca procedente de dois locais no Estado de Minas Gerais.

A análise dos gráficos da Figura 31 permite constatar que as árvores procedentes do Local 1 (Paraopeba) produziram madeira, em geral, menos densa que aquelas do Local 2 (Turmalina). Tal observação pode ser confirmada no Quadro 3. Também é possível notar na figura que, tanto nas árvores do Local 1 quanto naquelas do Local 2, há uma certa tendência de estabilização nos valores da densidade a partir da distância de

6 cm da medula. Tal fato torna-se especialmente importante se considerarmos a grande diferença de idade entre as árvores dos diferentes locais. Sendo aquelas do Local 2 mais jovens, seus lenhos parecem atingir a maturidade muito precocemente. Esta observação possui valor somente se considerarmos que a transição entre madeira juvenil e adulta está bem demarcada apenas pela estabilização da densidade. Ainda sobre os gráficos, é notável a diferença na taxa de crescimento em diâmetro entre os locais, posto que as árvores do Local 1 com 23 anos de idade apresentaram cernes com, no máximo, 11 cm de diâmetro, enquanto que dentre aquelas do Local 2, apesar de 10 anos mais jovens, algumas já produziam cernes com até 10 cm de diâmetro. A comparação entre os valores da densidade da madeira em várias posições em relação a medula e entre os locais é apresentada na Figura 32.

Quadro 3 - Valores médios e desvio-padrão para a densidade da madeira seca ao ar

Distância da medula (cm)	Densidade			
	Local 1		Local 2	
	Média (g/cm ³)	Desvio-padrão (g/cm ³)	Média (g/cm ³)	Desvio-padrão (g/cm ³)
1	0,57	0,06	0,66	0,06
2	0,61	0,06	0,70	0,06
3	0,70	0,07	0,76	0,07
4	0,77	0,07	0,80	0,07
5	0,80	0,07	0,84	0,08
6	0,86	0,06	0,90	0,07
7	0,87	0,06	0,92	0,05
8	0,86	0,04	0,95	0,05
9	0,89	0,05	0,92	0,04
10	0,87	0,05	0,98	0,02
11	0,88	0,03	-	-
Média*	0,79		0,84	
Desvio-padrão	0,11		0,11	
CV(%)	14,55		13,03	

* As médias por local diferem significativamente a 5% de probabilidade.

A análise dos dados acerca da densidade da madeira e da sua variação medula-casca entre os locais de coleta, permite afirmar-se que os dois locais de procedência forneceram árvores com diferentes fenótipos para esta propriedade. Possivelmente, tais árvores possuíam diferentes materiais genéticos ou as condições ambientais onde vegetaram eram deveras distintas, ou ainda, ambas as situações ocorreram simultaneamente. Tais considerações são importantes para os propósitos do presente trabalho, pois

oferecem a oportunidade de testar a utilidade de modelos de calibração ajustados sobre amostras procedentes de um povoamento para a predição de propriedades da madeira produzida por árvores de distintos genótipos e, ou, submetidas a distintas condições de crescimento.

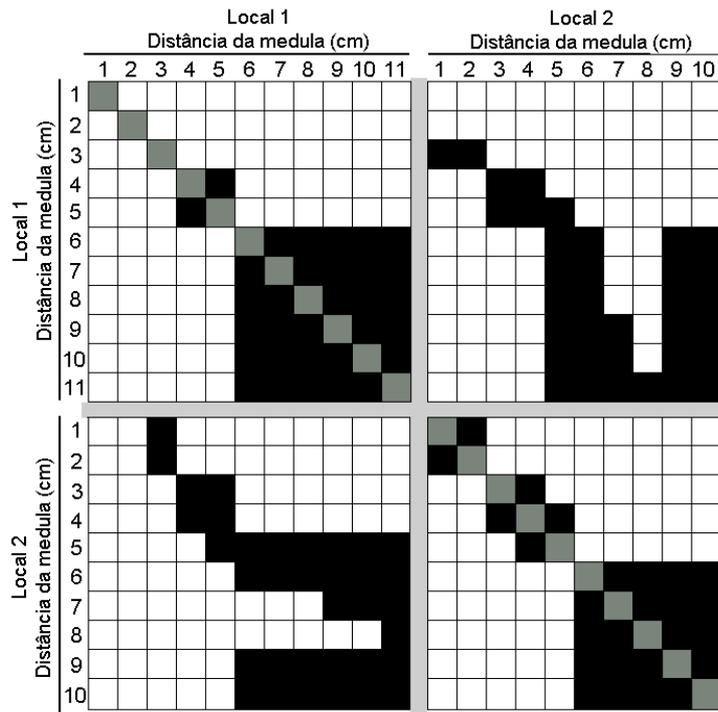


Figura 32 - Resultados do teste t para as médias da densidade em várias posições medula-casca e nos dois locais de coleta. Os quadrados brancos indicam que as médias correspondentes àquele cruzamento são diferentes e os pretos que elas são iguais a 5% de probabilidade.

5.2.1. Os espectros e suas correlações com a densidade básica

Os espectros médios dentro de cada posição medula-casca para os dois locais de coleta são mostrados na Figura 33. Nenhuma diferença significativa é prontamente notável entre os espectros das várias posições, nem tampouco entre os locais. Percebe-se em todos os casos, assinaturas espectrais que revelam grande absorção no ultravioleta e nos mais baixos comprimentos de onda do visível (região do azul) e, em seguida, uma invariável elevação da reflectância em direção ao infravermelho. Bastante notável, entretanto, é o salto observado na reflectância exatamente na transição entre o UV e o VIS e a sua relativa estabilização da resposta espectral na região do azul (≈ 400 a 500 nm).

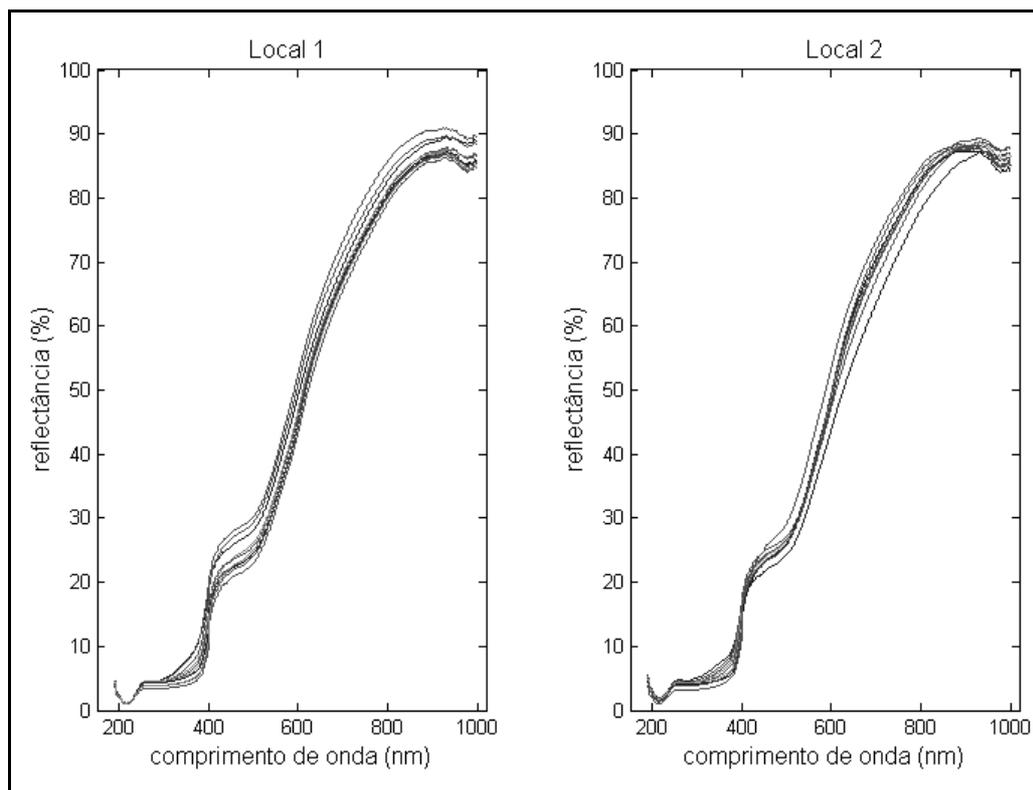


Figura 33 - Espectros de reflectância total das faces radiais de amostras da madeira de *E. urophylla* procedentes do Local 1 (Paraopeba) e do Local 2 (Turmalina), cada curva representando uma posição medula-casca (a cada centímetro).

É provável que a alta absorção da radiação no ultra-violeta e no azul esteja associada à resposta espectral dos cromóforos da lignina e que a sua correlação com a densidade da madeira esteja associada ao fato de que há uma maior concentração da “substância madeira” por unidade de área irradiada quando a densidade também é mais alta. Lebow et al. (1996) apresentam curvas espectrais semelhantes àquelas aqui obtidas em seu trabalho sobre a madeira sólida de *Pseudotsuga menziesii*.

Ao examinar, porém, o comportamento da variância da reflectância em cada comprimento de onda, observa-se uma grande diferença entre os valores, com picos bem marcados em torno de 370 e 650 nm (Figura 34).

Ao serem detectadas variáveis para as quais a variância é excepcionalmente mais elevada, é possível que sejam também reveladas correlações entre tais variáveis e outras tomadas como dependentes nos modelos de calibração. Ou seja, se não há variabilidade no valor de uma variável, dificilmente esta encontrará correlações com outras que, por sua vez, apresentem-se com uma certa dispersão de valores entre as

amostras de uma população. Entretanto, tal afirmação não quer dizer que maiores correlações com qualquer variável dependente serão encontradas obrigatoriamente para a independente que apresenta maior variância. Ao tomar-se a densidade da madeira como variável dependente, por exemplo, foi possível aqui se constatar que suas maiores correlações com a reflectância encontram-se por volta de 370 nm, como ilustra a Figura 35.

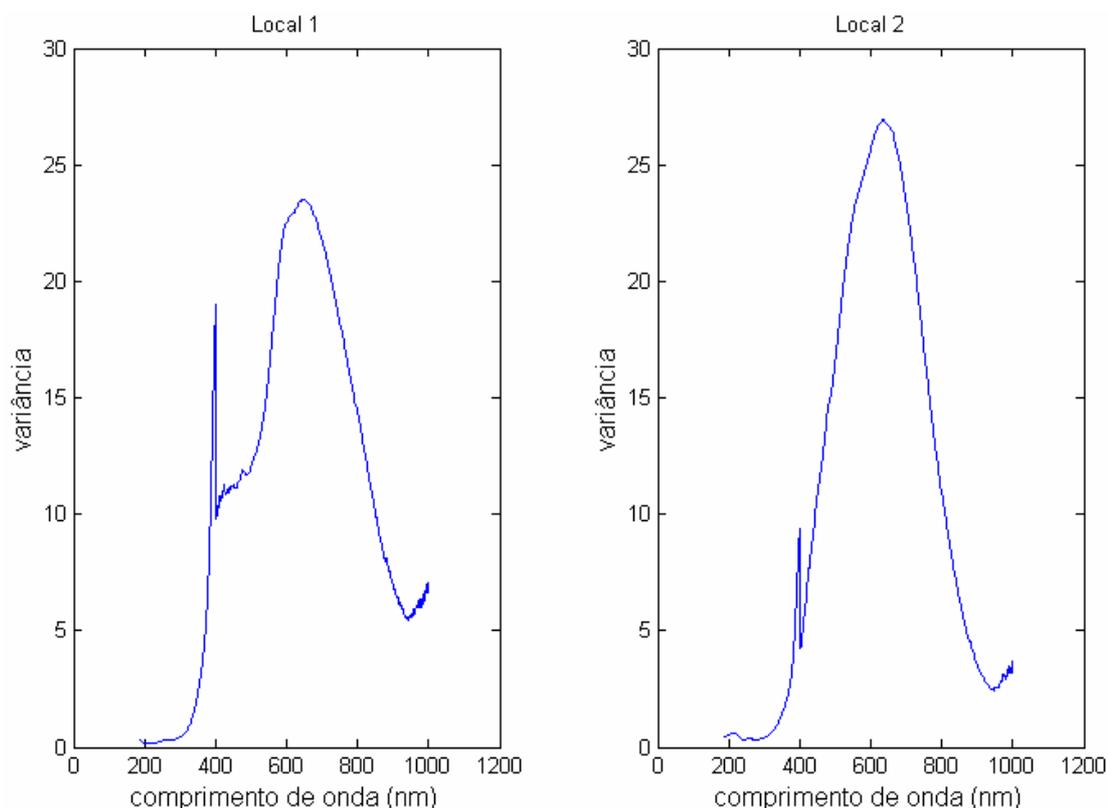


Figura 34 - Variância para reflectância total da superfície da face radial de amostras da madeira de *E. urophylla* procedente dos Locais 1 e 2.

Agrupando-se as amostras dos Locais 1 e 2 ou analisando-se cada local separadamente, os resultados são bastante semelhantes no que diz respeito à variância da reflectância e das suas correlações com a densidade da madeira. As tentativas para o ajuste de modelos univariados ou com poucas variáveis independentes diretamente a partir dos dados brutos (reflectância) para a previsão da densidade da madeira quedaram-se frustradas devido às correlações insatisfatórias encontradas quando são tomadas individualmente ou pela grande autocorrelação que existe entre as variáveis independentes. Assim, modelos univariados forneceram RMSEP elevados, enquanto

que modelos múltiplos ajustados por *forward*, limitaram-se a admissão de comprimentos de onda em número insuficiente para uma redução expressiva no erro de predição na validação (RMSEP). Foi também detectada uma grande instabilidade nos modelos no que diz respeito aos comprimentos de onda que eram adicionados para explicar-se os resíduos subseqüentes obtidos. Ou seja, ao tomar-se ao acaso diferentes grupos de amostras para calibração e validação, exceto para a primeira variável admitida nos modelos (sempre aquela de maior correlação com a independente), não houve constância na seqüência de variáveis que entravam por estarem mais bem correlacionadas aos resíduos do modelo anterior. Tal fato é ilustrado na Figura 36 que apresenta a freqüência de ocorrência de mais altas correlações entre as variáveis independentes e a densidade básica da madeira (X1) ou com os resíduos de modelos ajustados por *forward* em 50 vezes que foram escolhidas ao acaso amostras para calibração e validação agrupando-se os Locais 1 e 2.

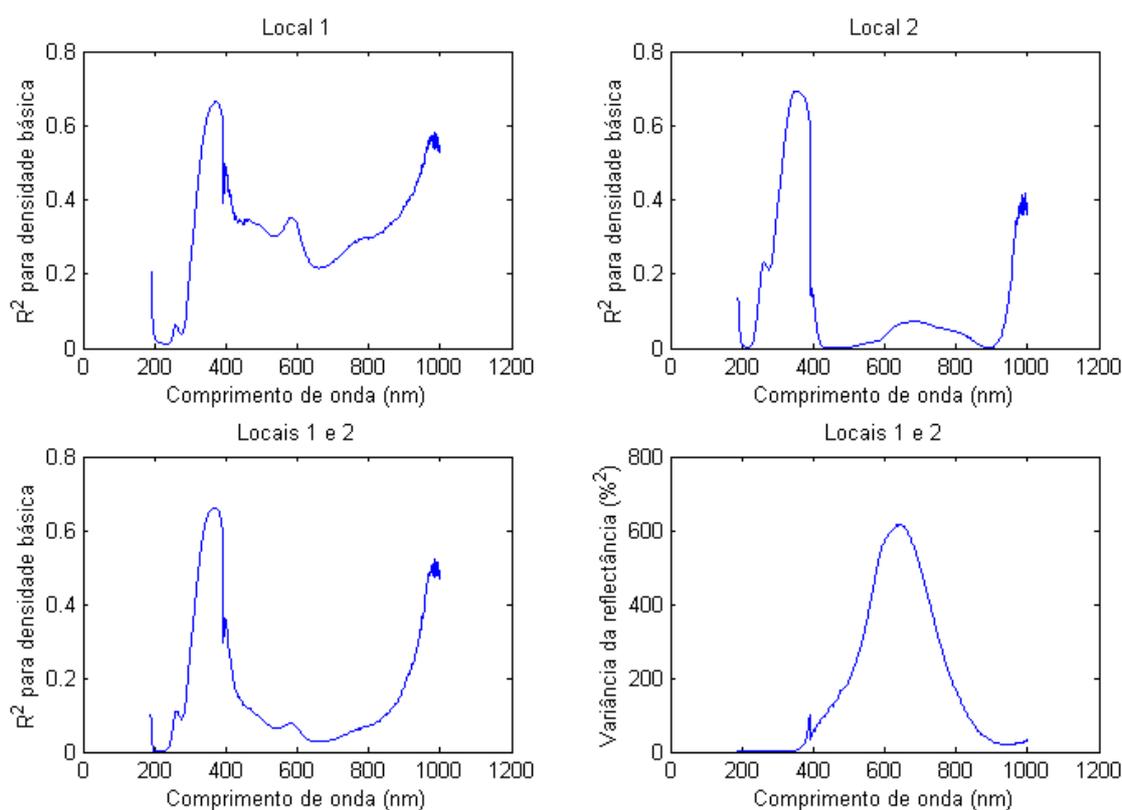


Figura 35 - Quadrado do coeficiente de correlação (R^2) entre a densidade básica da madeira de *E. urophylla* e a reflectância de sua face radial em vários comprimentos de onda no UV-VIS-tNIR (Locais 1 e 2 separadamente – gráficos superiores). Correlações com a densidade e variância da reflectância quando agrupados os Locais 1 e 2 (gráficos inferiores).

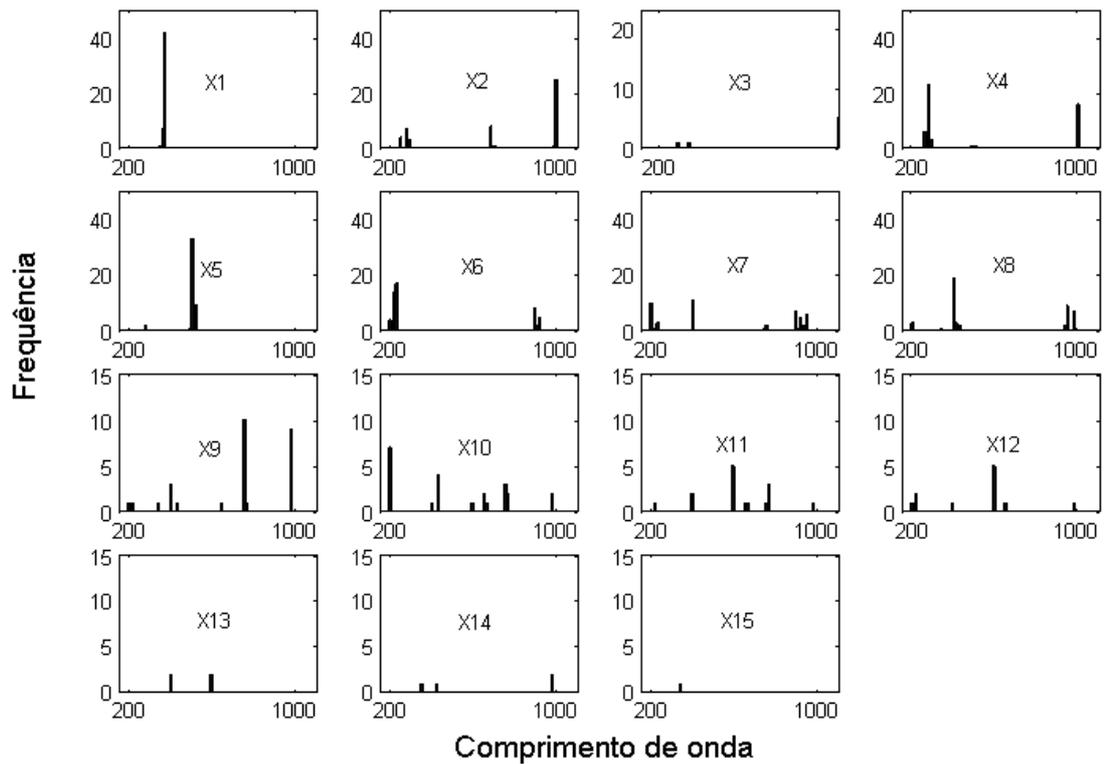


Figura 36 - Frequência com que os vários comprimentos de onda figuraram como variáveis independentes nos modelos múltiplos ajustados por MQO *forward* em 50 rodadas. (Atentar-se para a mudança de escala no eixo das ordenadas nos gráficos das duas últimas linhas).

De fato, uma certa consistência foi encontrada apenas para a entrada das variáveis correspondentes aos comprimentos de onda em torno de 365 nm, pois, apesar de uma discreta dispersão em torno desse valor (350 a 375 nm), 100% das rodadas tiveram esta faixa do espectro com sendo a que melhor se correlacionou com densidade básica da madeira. O modelo univariado ajustado para somente a reflectância em 350 nm e seus resíduos na predição da densidade é apresentado na Figura 37.

O modelo da Figura 37 apresentou EPE = 8,64%, RMSEP = 11,94%, R^2 para a calibração = 0,70 e R^2 para a validação = 0,54. Percebe-se que ainda há uma certa tendenciosidade na distribuição dos resíduos o que sugere a necessidade de inclusão de novas variáveis. Apesar da análise da variância ter demonstrado que o coeficiente de regressão e o intercepto eram significativos ao nível de 5% de probabilidade, o modelo não se presta a propósitos de predição, ou seja, seria inaceitável um erro de predição tão elevado em um modelo de calibração.

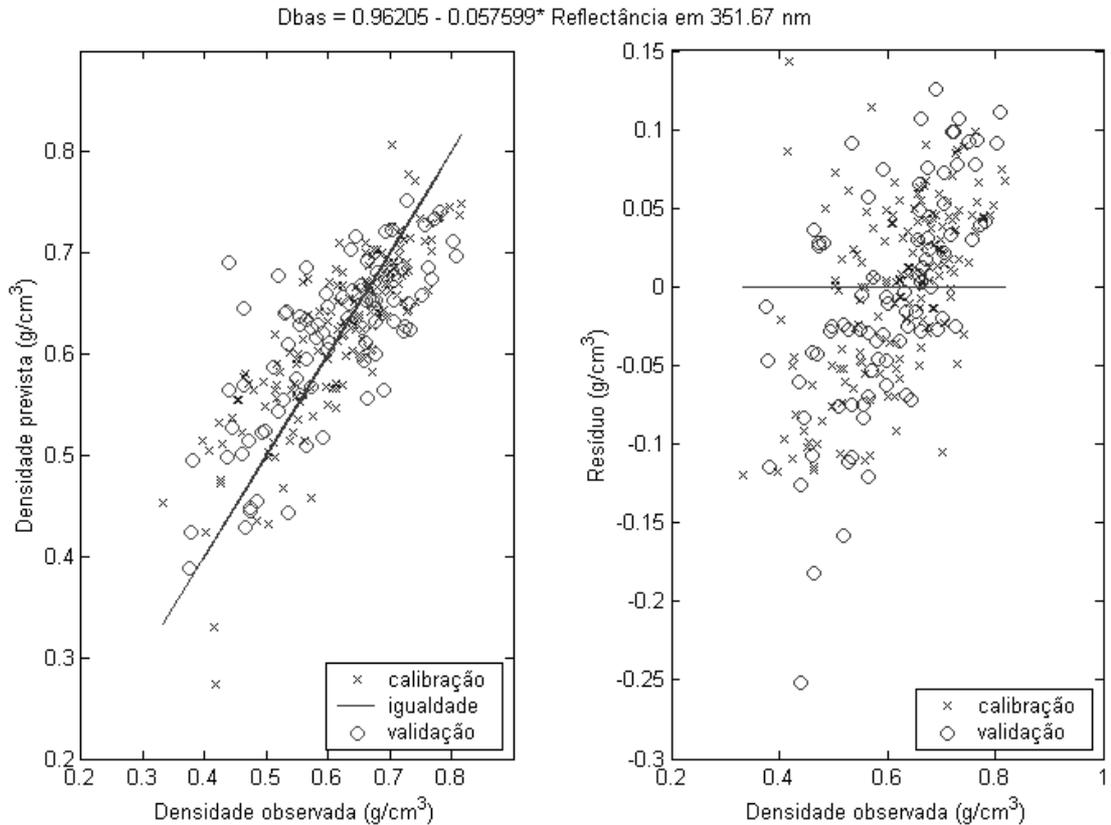


Figura 37 - Plotagem dos valores previstos e observados e os resíduos obtidos para a densidade básica da madeira de *E. urophylla* utilizando-se um modelo que possui apenas a reflectância em 351,57 nm como variável independente.

A agregação da reflectância em outros comprimentos de onda aos modelos univariados utilizando-se o método MQO *forward* levou a ganhos significativos na redução dos erros e na elevação dos coeficientes de correlação como pode ser observado na Figura 38. Note-se, porém, na figura que apresenta os resultados para 3 rodadas tomadas ao acaso, que o acréscimo das variáveis nos modelos não produz efeitos consistentes. Ou seja, em alguns casos os erros tendem a diminuir, noutros, a aumentar. Dentre as 50 rodadas acima referidas, os modelos admitiram, em média, 7 variáveis independentes antes que se manifestasse a redundância entre variáveis independentes. Um máximo de 15 variáveis foi admitido, não causando, entretanto, ganhos na estabilidade dos modelos. A admissão de variáveis em um modelo multivariado é realizada testando-se a redundância da informação que esta variável carrega. Tal teste é realizado através da análise da correlação de uma nova variável com aquelas que já figuram no modelo. Variáveis com altas correlações com aquelas originais são impedidas de entrar.

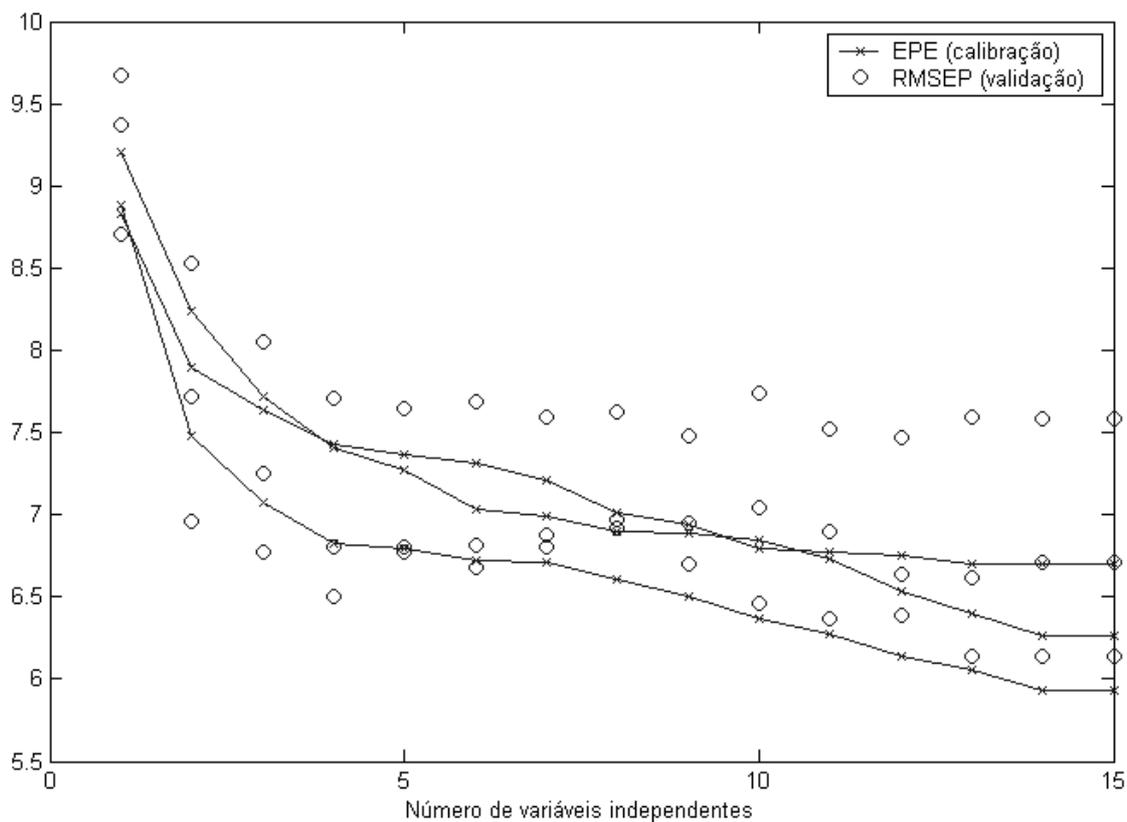


Figura 38 - Erro-padrão da estimativa (EPE) e raiz quadrada do erro quadrático médio de predição (RMSEP) para modelos ajustados por MQO *forward* para a densidade básica da madeira de *E. urophylla* em três grupos distintos de amostras tomadas ao acaso para validação e calibração.

O relativo insucesso obtido nos ajustes dos modelos univariados e múltiplos por MQO está certamente associado à alta colinearidade observada entre as variáveis independentes (reflectância). A Figura 39 apresenta os coeficientes de determinação entre as variáveis independentes. Note-se que, ao afastarem-se da linha diagonal, onde o coeficiente de correlação é naturalmente igual a 1, as correlações vão reduzindo seu coeficientes. Outro fato a ser observado na figura é que a reflectância em comprimentos de onda dentro do UV apresentam baixas correlações com a região do visível e sua transição para o infravermelho. De qualquer forma, os modelos obtidos nas regressões por componentes principais (PCR) não são susceptíveis às interferências geradas pelas altas correlações entre as variáveis independentes, já que estas serão transformadas por PCA em outras completamente descorrelacionadas.

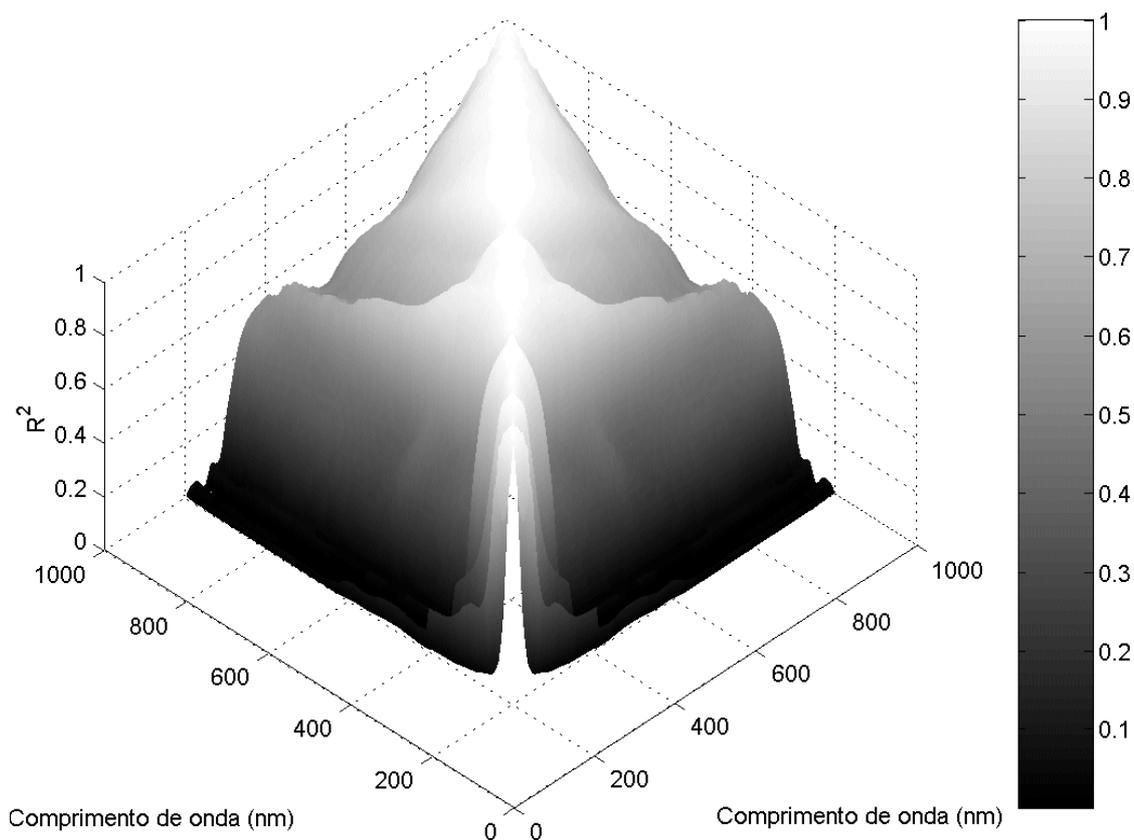


Figura 39 - Coeficientes de determinação entre as variáveis independentes, neste caso, a reflectância.

5.2.2. Modelos múltiplos ajustados por PCR

A seguir serão apresentados os resultados obtidos no ajuste de modelos de calibração por PCR para a predição da densidade básica da madeira. Foram tomadas como variáveis independentes as médias da reflectância a cada 5 nm, desde 200 nm até 1.000 nm. Sobre essas variáveis foi aplicada a PCA e, em seguida foram ajustados os modelos adicionando-se os escores dos componentes principais um a um, em ordem decrescente de variância. Ou seja, nesta etapa, foram obtidos modelos com número de variáveis independentes (PC's) variando de 1 a 97 (dois terços do número de amostras no Local 1 menos 1). A Figura 40 mostra como a PCA foi capaz de concentrar a variância dos dados originais da reflectância. Note-se que os primeiros três componentes carregam mais de 90% da variância global dos dados. Entretanto, a alta variância dos primeiros componentes não quer dizer que estes estejam arranjados hierarquicamente no que diz respeito à sua importância para a entrada em modelos para a

predição de valores uma certa variável independente. Ou seja, a utilidade dos componentes nos modelos de calibração depende essencialmente da correlação de seus escores com a variável independente que se deseja modelar. Ainda, a Figura 40 revela que o componente número 3 está melhor correlacionado com a densidade básica da madeira que o primeiro e o segundo, que são aqueles que carregam a maior parte da variância dos dados. Assim, modelos ajustados por *forward* poderão fornecer melhores resultados que aqueles em que os componentes são adicionado um a um em ordem decrescente de sua variância. Tal comparação poderá ser verificada mais adiante, quando da otimização dos modelos.

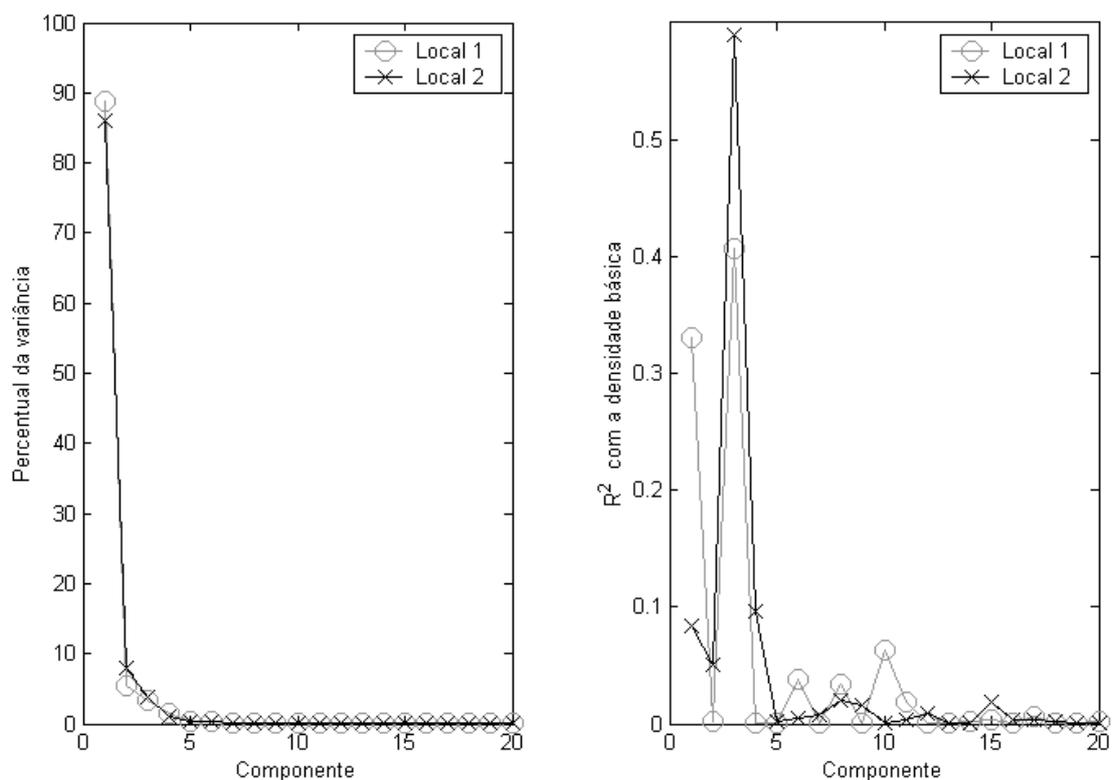


Figura 40 - Percentual da variância global carregado por cada componente principal e correlações entre os componentes e a variável independente (densidade básica) para os Locais 1 e 2.

Quando testados dentro de um mesmo local, os modelos foram ajustados sobre dois terços dos dados disponíveis, sendo o terço restante reservado para a validação. A separação das amostras para os grupos de calibração (ajuste dos modelos) e validação foi realizada aleatoriamente por 300 vezes não se permitindo que houvesse repetição.

Ou seja, os ajustes foram realizados 300 vezes utilizando-se grupos diferentes de amostras. A cada um desses grupos de modelos também chamaremos aqui “rodadas”. Os resultados para uma dessas rodadas sobre os dados do Local 1 são mostrados na Figura 41. Vale ressaltar que essa figura apresenta o comportamento típico para os parâmetros sob análise em todas as rodadas (erros e coeficientes de correlação).

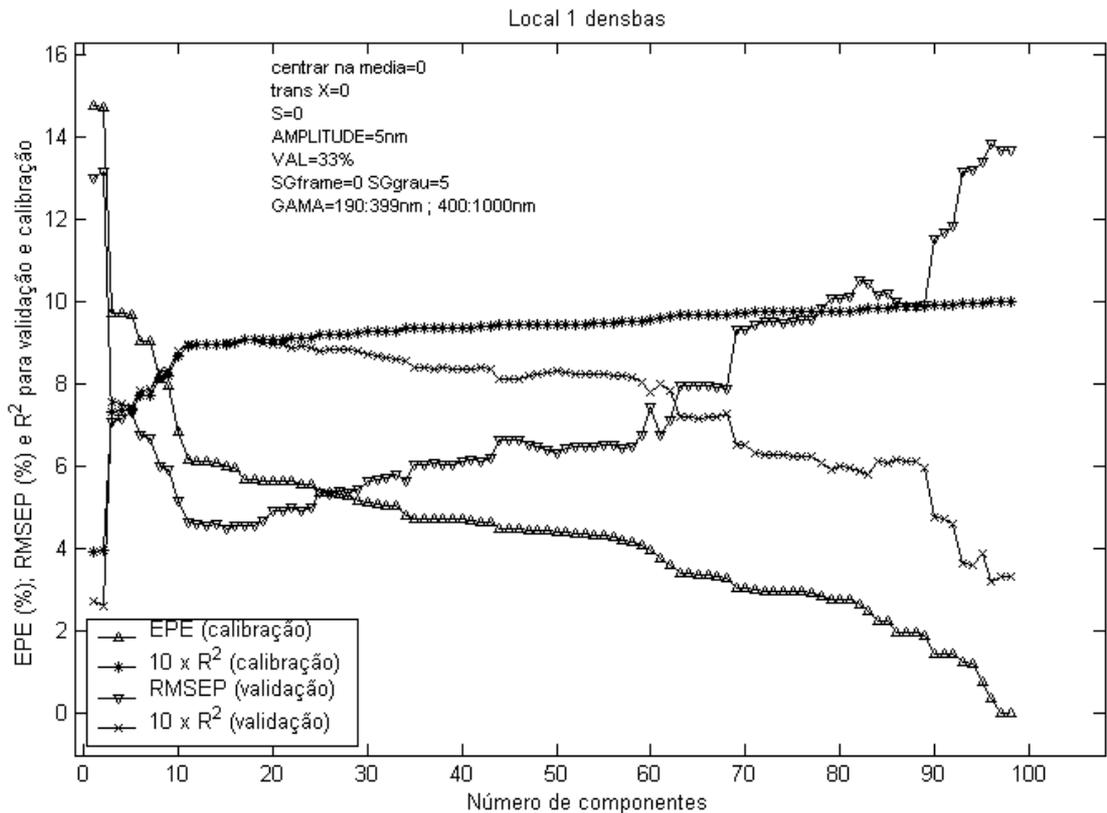


Figura 41 - Erro-padrão da estimativa (EPE), raiz quadrada do Erro-padrão da predição (RMSEP) e coeficientes de determinação (R²) para modelos ajustados e validados sobre os dados do Local 1.

Note-se na Figura 41 que, à medida que vão sendo acrescentados os componentes ao modelo, o EPE decresce atingindo o valor zero e o R² do modelo de calibração cresce até 1 quando o número de componentes adicionados torna-se igual ao número de amostras utilizadas na calibração. Neste ponto, o RMSEP e o R² da predição se estabilizam em valores que são característicos para os grupos de amostras tomadas para a calibração e predição em cada rodada. Mais objetivamente, ao se atingir o número de variáveis (neste caso os escores dos componentes principais) igual ao número de amostras, os coeficientes de regressão tornam-se zero, ou seja, não são

calculados, pois quando emprega-se MQO não é possível inverter-se a matriz $X'X$ para o cálculo dos coeficientes da regressão. A distribuição dos valores de RMSEP mínimos em cada uma das 300 rodadas e dos correspondentes números de componentes em que ocorreram tais valores são mostrados na Figura 42.

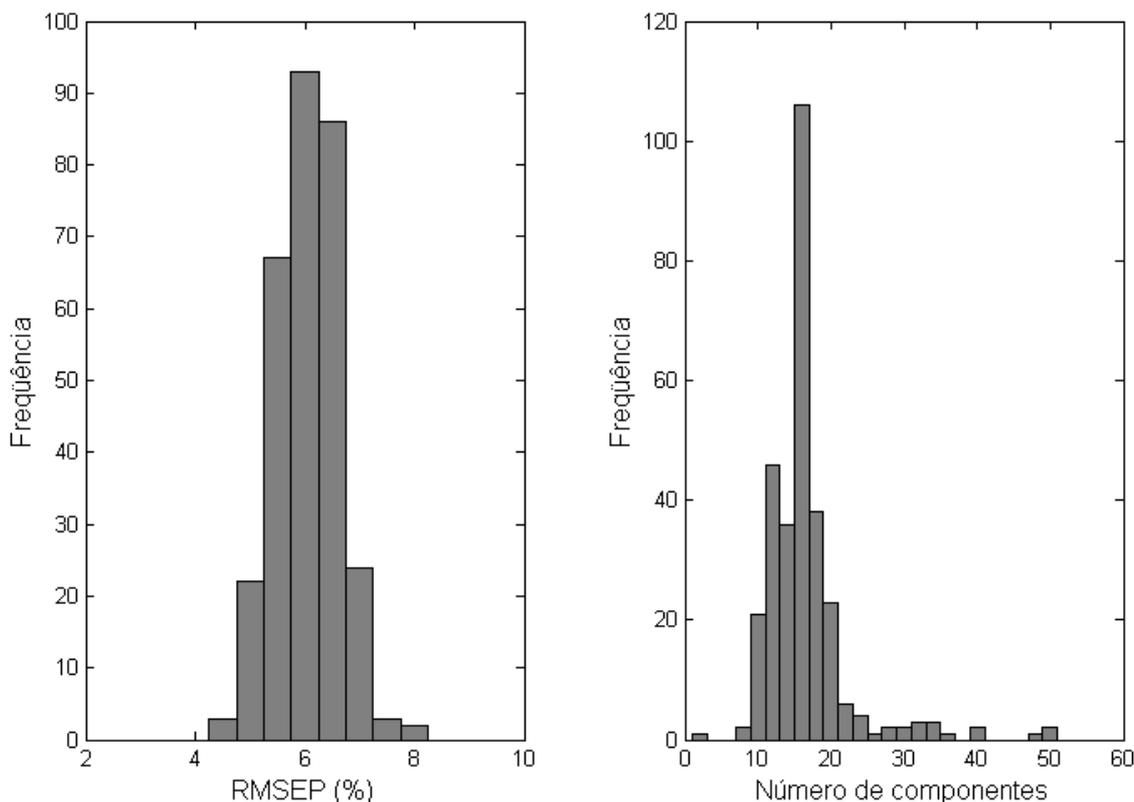


Figura 42 - Distribuição dos valores de RMSEP mínimos para a densidade básica da madeira do Local 1 em cada uma das 300 rodadas e dos correspondentes números de componentes com que ocorreram.

Ainda, a Figura 42 permite constatar que a maior parte dos erros mínimos obtidos nas 300 rodadas para o Local 1 se encontravam por volta de 6% e que foram obtidos com modelos que possuíam em média 17 variáveis independentes (componentes).

A Figura 43 apresenta a plotagem dos valores observados e previstos e resíduos de modelos ajustados sobre os valores da reflectância e da densidade básica de amostras de calibração e validação. Os gráficos mostram os modelos e o número de variáveis com que se obtiveram valores mínimos para o RMSEP em rodadas nas quais este valor mostrou-se o mais elevado, o mais próximo da média dos valores mínimos e para a rodada em que apresentou o seu menor valor. Note-se que os dois extremos para o

RMSEP foram obtidos com um número semelhante de componentes admitidos nos modelos (16 e 17, respectivamente). De fato, dentre os erros mínimos obtidos em cada rodada, não existe correlação entre o número de componentes e o valor de RMSEP. Tal informação pode ser confirmada observando-se a diferença apresentada pelas distribuições dos valores de RMSEP mínimos, com distribuição normal, e do número de componentes presentes nos modelos que obtiveram tais erros mínimos, que se apresenta de maneira assimétrica com o deslocamento da média para os menores valores (Figura 43).

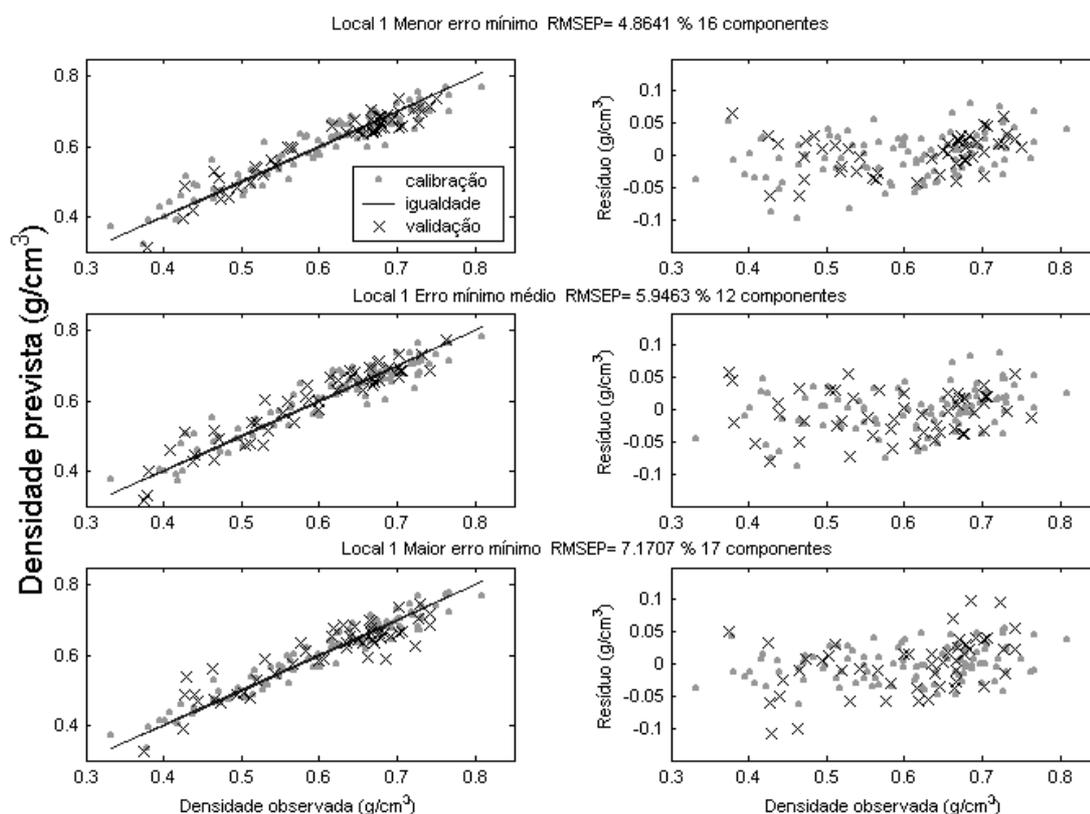


Figura 43 - Plotagem dos valores observados versus previstos e os resíduos de modelos que apresentaram RMSEP maior, menor e mais próximo da média dos erros mínimos na validação para a previsão da densidade básica das amostras do Local 1. São apresentados também os valores para RMSEP e o número de componentes principais utilizados em cada modelo.

O Quadro 4 apresenta a análise da variância para as regressões múltiplas por PCA realizadas para a obtenção dos modelos de calibração que obtiveram os erros mínimos na sua validação apresentados na Figura 43.

Quadro 4 - Análise da variância para as regressões múltiplas e erros padrão da estimativa (EPE) dos modelos de calibração que apresentaram os erros maior, menor e mais próximo da média dos RMSEP mínimos dentre 300 rodadas, na validação para a previsão da densidade básica após a remoção dos *outliers*. Os modelos foram ajustados tomando-se amostras para a calibração e validação dos Locais 1. O RMSEP também é apresentado

	Modelo de mínimo RMSEP			
	EPE = 6,4088	RMSEP = 4,8641	R ² da calibração = 0,8920	R ² da validação = 0,9159
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	16	0,9903	0,0619	41,8245
Resíduo	81	0,1199	0,0015	
Total	97	1,1101		
	Modelo de médio RMSEP			
	EPE = 5,9778	RMSEP = 5,9463	R ² da calibração = 0,8929	R ² da validação = 0,8994
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	12	0,9482	0,0790	59,0557
Resíduo	85	0,1137	0,0013	
Total	97	1,0619		
	Modelo de máximo RMSEP			
	EPE = 4,8199	RMSEP = 7,1707	R ² da calibração = 0,9401	R ² da validação = 0,8122
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	17	1,0631	0,0625	73,8069
Resíduo	80	0,0678	0,0008	
Total	97	1,1308		

A distribuição dos valores de RMSEP mínimos em cada uma das 300 rodadas e dos correspondentes números de componentes em que ocorreram tais valores são mostrados na Figura 44 para os ajustes realizados sobre as amostras do Local 2. Note-se o comportamento semelhante àquele apresentado para os modelos do Local 1. Apesar de apresentar uma maior dispersão, o número de componentes com os quais foram obtidos os erros mínimos em cada rodada apresentam-se também de maneira assimétrica e com uma grande concentração de valores entre 10 e 20 componentes. Para o Local 2 as média do REMSEP e do número de componentes para a obtenção do erro mínimo foram 5, 8 e 18, respectivamente. Ainda sobre a distribuição dos erros mínimos, tal qual observado para o Local 1, também não foram detectadas correlações entre o número de componentes e o valor do RMSEP mínimo em cada rodada.

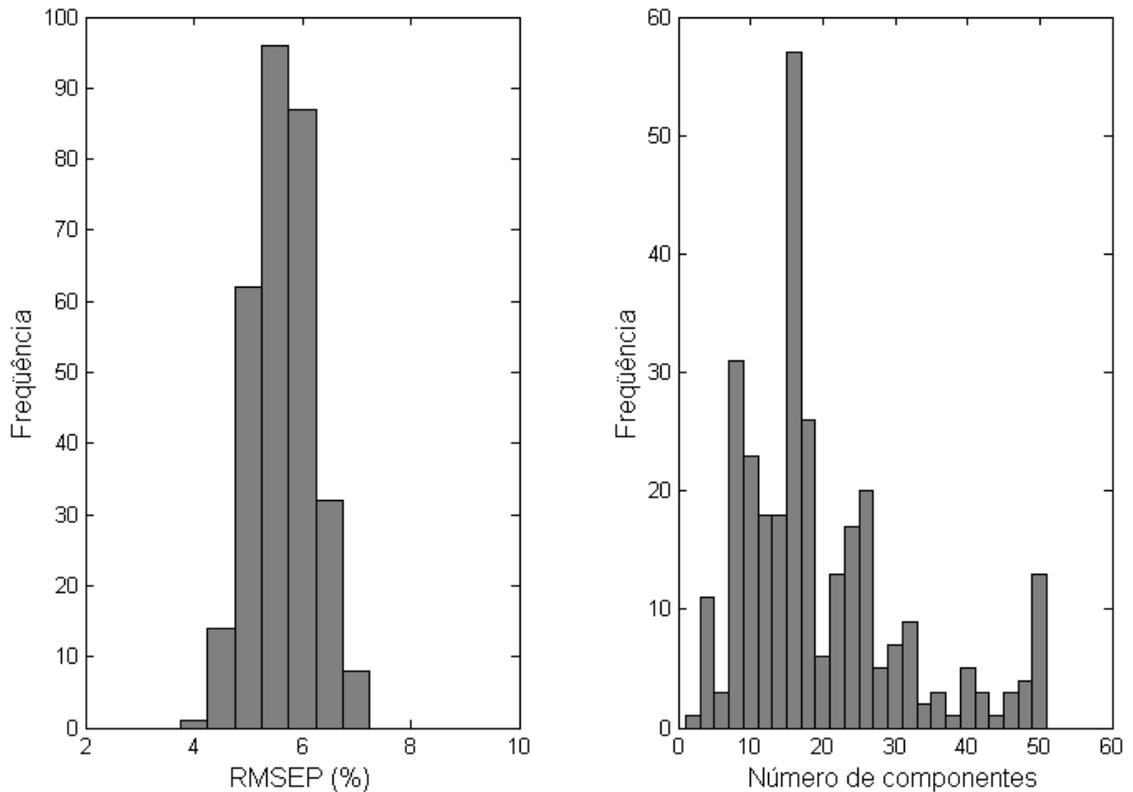


Figura 44 - Distribuição dos valores de RMSEP mínimos para a densidade básica da madeira do Local 2 nas 300 rodadas e dos correspondentes números de componentes com que ocorreram.

A Figura 45 apresenta a plotagem dos valores observados e previstos e resíduos de modelos ajustados sobre os valores da reflectância e da densidade básica de amostras de calibração e validação para o Local 2. Também à semelhança do que foi observado para o Local 1, os menores e maiores erros mínimos foram obtidos com números de componentes bastante próximos (10 e 8, respectivamente). O erro mínimo mais próximo da média dos erros mínimos das 300 rodadas foi obtido, porém, com um número excepcionalmente baixo de componentes (4) se comparado com os demais modelos. A análise da variância para as regressões que forneceram os modelos apresentados na Figura 45 é encontrada no Quadro 5. Os três modelos apresentaram coeficientes de regressão significativos se considerado o nível de 5% de probabilidade.

Ainda na Figura 45, é possível notar-se uma ligeira tendenciosidade na distribuição dos resíduos em relação aos valores da variável dependente.

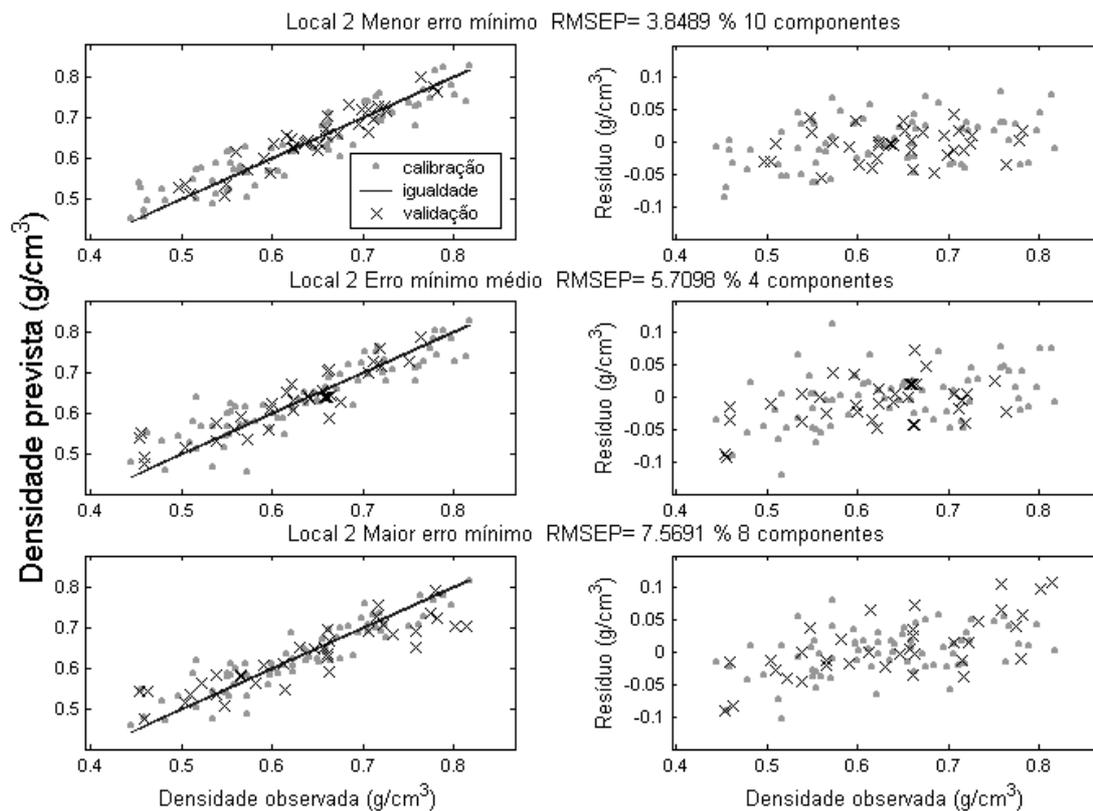


Figura 45 - Plotagem dos valores observados versus previstos e os resíduos de modelos que apresentaram RMSEP maior, menor e mais próximo da média dos erros mínimos na validação para a previsão da densidade básica das amostras do Local 2. São apresentados também os valores para RMSEP e o número de componentes principais utilizados em cada modelo.

Com os resultados apresentados até este ponto, pode-se constatar que é possível ajustar-se modelos para a predição da densidade básica da madeira de *Eucalyptus urophylla* utilizando-se amostras de um mesmo local para calibração e para a predição. A predição dos valores da densidade pode ser realizada com um nível de erro razoável para alguns modelos. Ver-se-á, a seguir, como se comportam modelos ajustados com todas as amostras de um local na tentativa de se prever a densidade das amostras do outro local.

As Figuras 46 e 47 apresentam os erros e coeficientes de determinação na calibração e validação dos modelos ajustados sobre as amostras do Local 1 e validados sobre as amostras do Local 2 e vice-versa. Está claro nas figuras que tal estratégia não produz bons resultados. Ou seja, utilizando-se a PCR não é possível prever-se a densidade básica da madeira procedente de um local aplicando-se modelos ajustados sobre a reflectância da madeira do outro local.

Quadro 5 - Análise da variância para as regressões múltiplas e erros padrão da estimativa (EPE) dos modelos de calibração que apresentaram os erros maior, menor e mais próximo da média dos RMSEP mínimos dentre 300 rodadas, na validação para a previsão da densidade básica após a remoção dos *outliers*. Os modelos foram ajustados tomando-se amostras para a calibração e validação do Local 2. O RMSEP também é apresentado

	Modelo de mínimo RMSEP			
	EPE = 6,6146	RMSEP = 3,8489	R ² da calibração = 0,8602	R ² da validação = 0,8938
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	10	0,6243	0,0621	36,3144
Resíduo	59	0,1014	0,0017	
Total	69	0,7257		
	Modelo de médio RMSEP			
	EPE = 6,8569	RMSEP = 5,7098	R ² da calibração = 0,8141	R ² da validação = 0,8284
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	4	0,5469	0,1367	71,18
Resíduo	65	0,1249	0,0019	
Total	69	1,3718		
	Modelo de máximo RMSEP			
	EPE = 6,0833	RMSEP = 7,5691	R ² da calibração = 0,8372	R ² da validação = 0,8061
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	8	0,4607	0,0576	39,2252
Resíduo	61	0,0895	0,0015	
Total	69	0,5502		

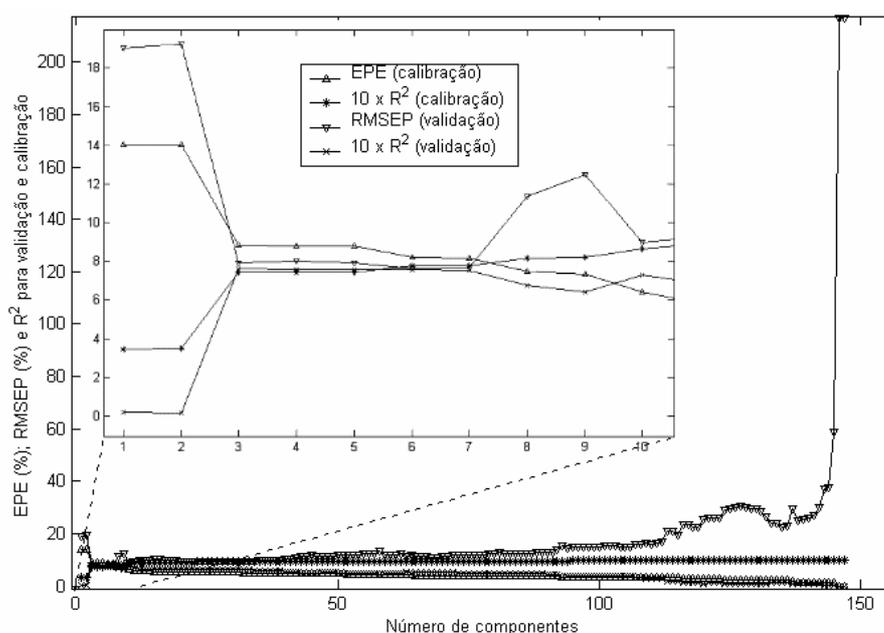


Figura 46 - Erros e coeficientes de determinação para modelos ajustados sobre a madeira do Local 1 e validados sobre as amostras do Local 2.

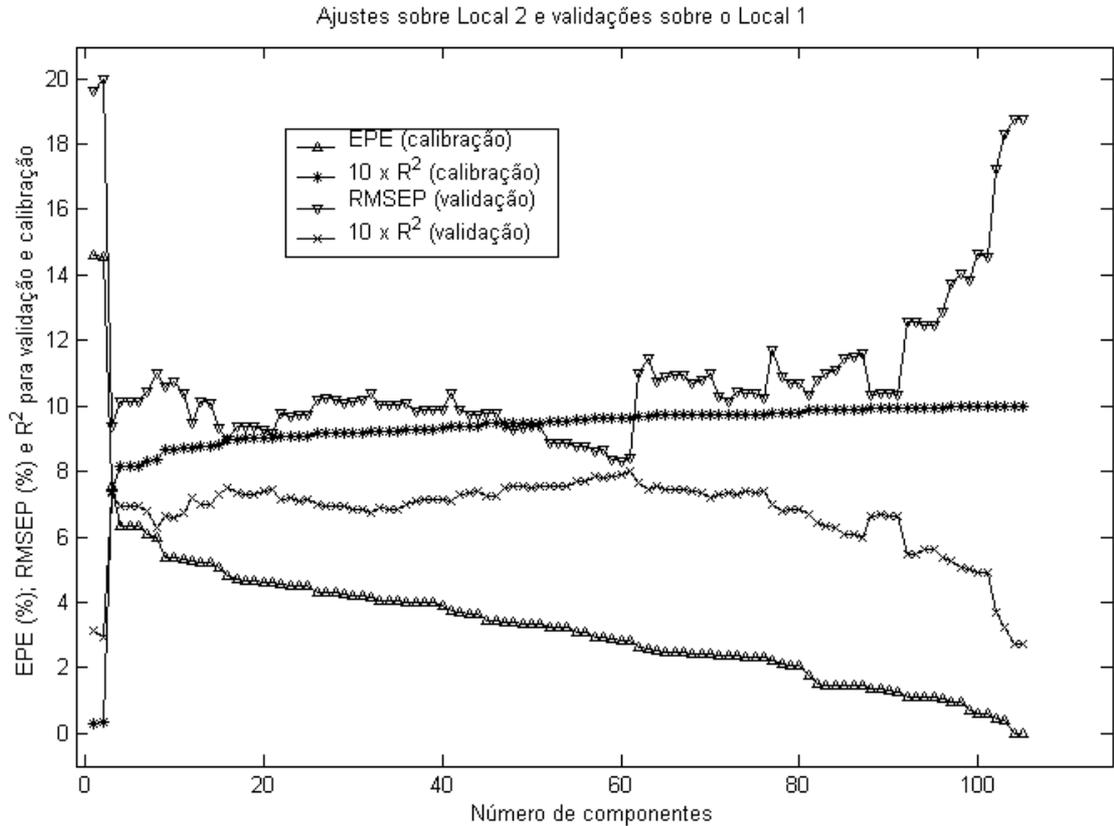


Figura 47 - Erros e coeficientes de determinação para modelos ajustados sobre a madeira do Local 2 e validados sobre as amostras do Local 1.

Finalmente, são apresentados a seguir os resultados para os modelos de calibração ajustados sobre grupos de amostras que representava tanto o Local 1 quanto o Local 2. À semelhança do que foi realizado sobre os locais individualmente, as amostras foram separadas ao acaso em 2 grupos, sendo que aquele destinado à calibração continha dois terços do número total de amostras, restando o outro terço à validação dos modelos.

A Figura 48 apresenta a distribuição dos valores dos erros de predição (RMSEP) mínimos e do número de componentes com que foram obtidos tais erros em cada uma das 200 rodadas em que modelos ajustados sobre os dados dos Locais 1 e 2 foram validados sobre amostras também procedentes de ambos os locais. Note-se que, neste caso, os erros mínimos apresentaram o valor médio de 6,34%, bastante próximo daqueles observados para as situações anteriores, enquanto que o número de componentes necessários para a obtenção dos erros mínimos apresentou uma média de 42 componentes, número bastante elevado se comparado ao que foi obtido até agora

com as combinações de amostras entre e dentro dos dois locais. Também, a distribuição do número de componentes necessários ao erro mínimo nas rodadas apresentou-se de forma bastante peculiar em um histograma bimodal, exibindo um hiato em torno da região centrada em 40 componentes e cujas sub-populações forneceram médias em torno de 23 e 65 componentes. As razões para tal comportamento permanecem não elucidadas, uma vez que não há correlação significativa entre os valores dos erros e números de componentes, nem entre si, nem tampouco com parâmetros relativos aos dados de calibração e validação utilizados em cada rodada. A saber, os parâmetros referidos foram: média, variância, amplitude, curtose e assimetria. Na verdade, a própria variabilidade encontrada para tais parâmetros nas diferentes rodadas mostrou-se um tanto reduzida e, somente, para a variância e a curtose (que são relacionadas entre si), observaram-se discretas correlações com o RMSEP.

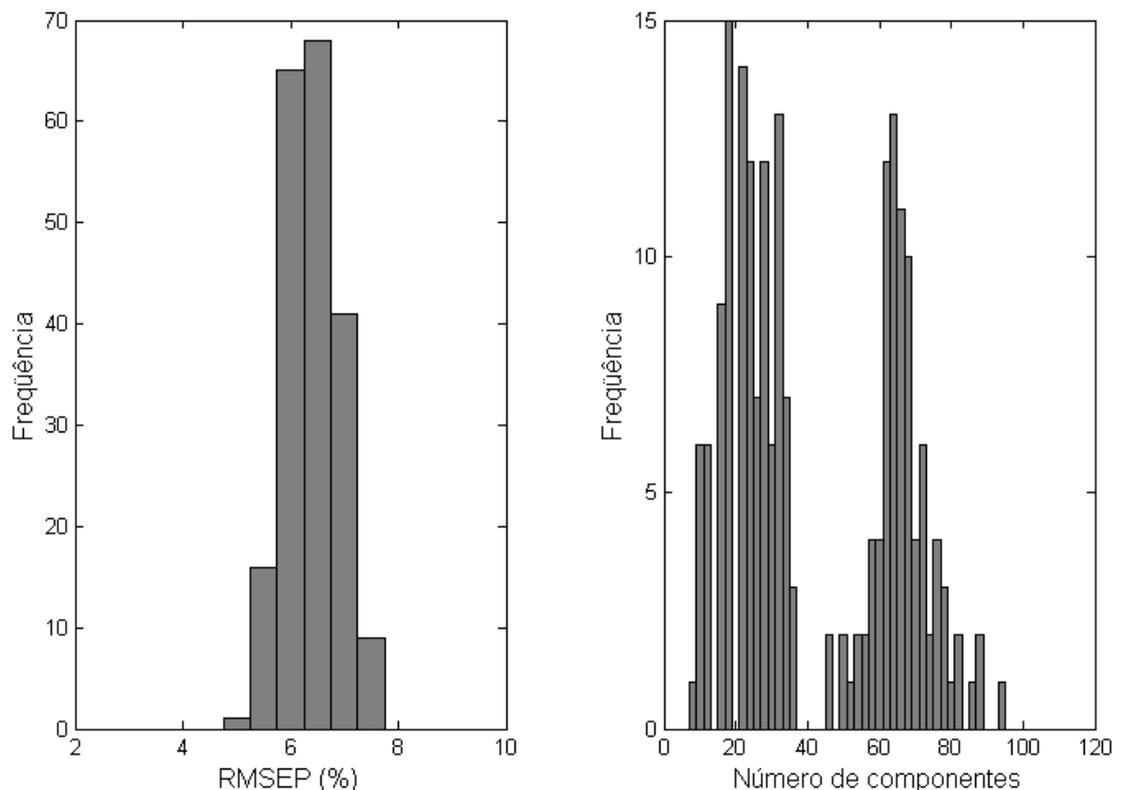


Figura 48. Distribuição dos valores de RMSEP mínimos para a densidade básica da madeira dos Locais 1 e 2 e dos correspondentes números de componentes com que ocorreram em cada uma das 200 rodadas.

Pode-se observar na Figura 49 o comportamento do RMSEP para algumas rodadas tomando-se amostras de ambos os locais para calibração e validação. Note-se

que, nas rodadas em que o menor RMSEP foi obtido com modelos compostos por um grande número de variáveis, em alguns casos foi possível obter-se erros ligeiramente maiores que o mínimo, porém com um número muito inferior de variáveis. Isso quer dizer que, a depender do nível de erro admitido em uma aplicação dos modelos de calibração, a distribuição dos erros apresentada na Figura 48 poderia ser modificada, resultando em um histograma unimodal para o número de componentes.

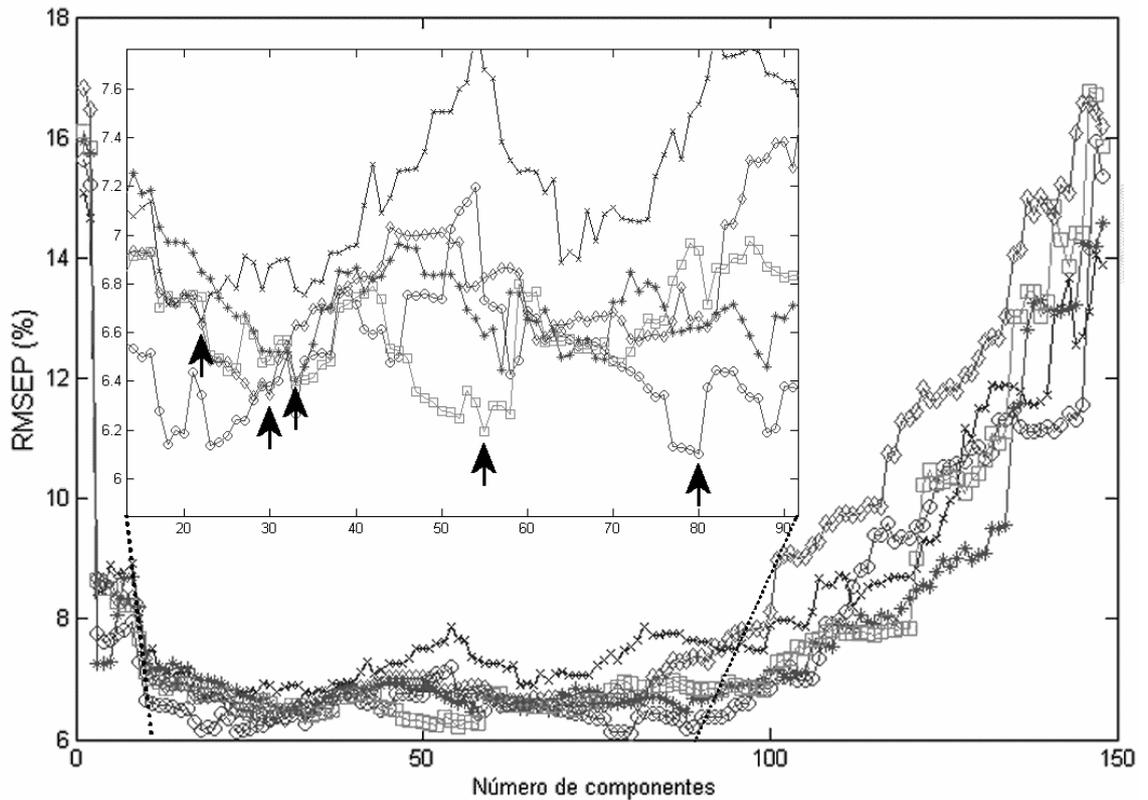


Figura 49 - RMSEP para modelos ajustados em cinco das 200 rodadas, tomando-se amostras para validação e calibração procedentes de ambos os locais. As setas indicam o RMSEP mínimo em cada rodada.

A Figura 50 apresenta a plotagem dos valores observados e previstos e resíduos de modelos que obtiveram o menor, o maior e o valor médio para o RMSEP mínimo nas 200 rodadas em que foram ajustados sobre os valores da reflectância e da densidade básica de amostras de calibração e validação tomadas dos locais 1 e 2. A análise da variância para estes modelos é apresentada no Quadro 6.

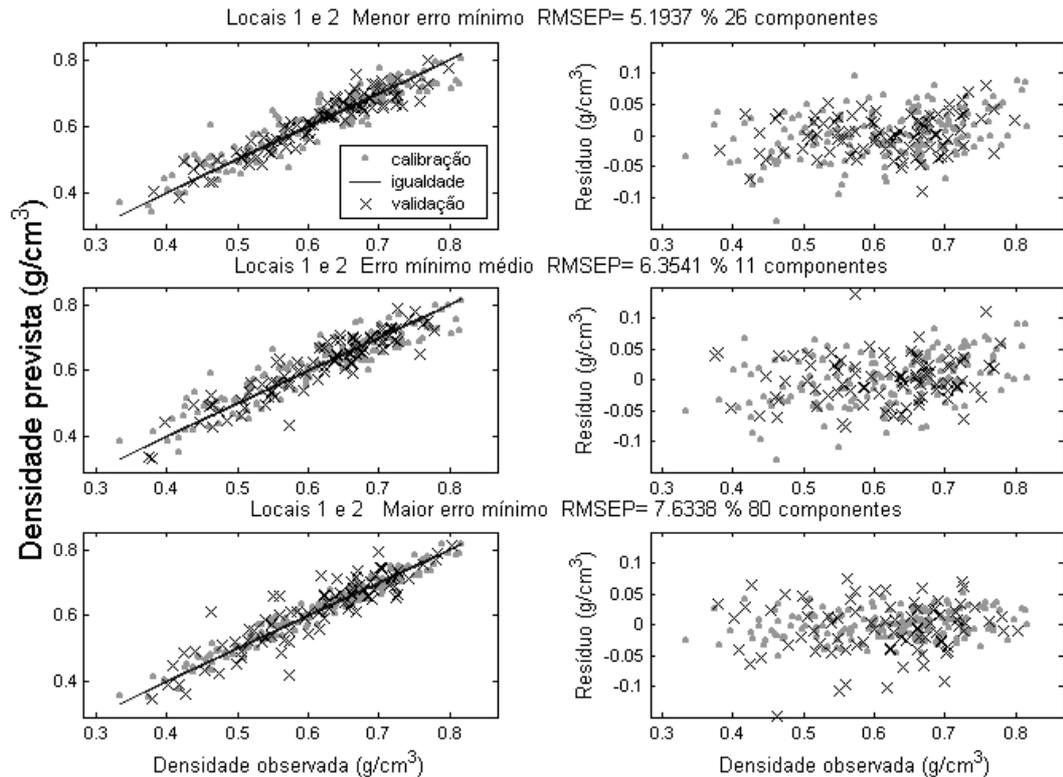


Figura 50 - Plotagem dos valores observados versus previstos e os resíduos de modelos que apresentaram RMSEP maior, menor e mais próximo da média dos erros mínimos na validação para a previsão da densidade básica das amostras. Os modelos foram ajustados tomando-se amostras para a calibração e validação dos Locais 1 e 2. São apresentados também os valores para RMSEP e o número de componentes principais utilizados em cada modelo.

Pode-se notar que os modelos apresentados na Figura 50 possuem erros e coeficientes de determinação semelhantes àqueles obtidos quando os modelos eram ajustados e validados sobre amostras da madeira de um mesmo local. Tal fato indica que, ao serem tomadas amostras de ambos os locais, as correlações entre os fenômenos físicos (densidade versus reflectância) voltam a ser detectadas e uma compensação para as peculiaridades intrínsecas às diferentes localidades é realizada pela PCA. De fato, como ilustrado na Figura 35, a maior correlação entre a reflectância e a densidade encontra-se por volta de 370 nm, tanto para o Local 1, quanto para o Local 2 e também quando são agregadas as amostras de ambos os locais.

Quadro 6 - Análise da variância para as regressões múltiplas e erros padrão da estimativa (EPE) dos modelos de calibração que apresentaram os erros maior, menor e mais próximo da média dos RMSEP mínimos dentre 200 rodadas, na validação para a previsão da densidade básica após a remoção dos *outliers*. Os modelos foram ajustados tomando-se amostras para a calibração e validação dos Locais 1 e 2. O RMSEP também é apresentado

	Modelo de mínimo RMSEP			
	EPE = 6.7443	RMSEP = 5.1937	R ² da calibração = 0.8690	R ² da validação = 0,8861
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	26	1,6149	0,0623	35,96
Resíduo	141	0,2442	0,0017	
Total	167	1,8636		
	Modelo de médio RMSEP			
	EPE = 6.7555	RMSEP = 6.3541	R ² da calibração = 0.8518	R ² da validação = 0.8468
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	11	1,5572	0,1416	81,52
Resíduo	156	0,2709	0,0017	
Total	167	1,8280		
	Modelo de máximo RMSEP			
	EPE = 4.8520	RMSEP = 7.6338	R ² da calibração = 0.9556	R ² da validação = 0.8149
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	80	1,6854	0,0211	23,38
Resíduo	87	0,0784	0,0009	
Total	167	1,7637		

5.2.3. Otimização dos modelos

5.2.3.1. Remoção de pontos discrepantes

Foi utilizada a técnica da validação cruzada para a identificação de pontos discrepantes no conjunto de dados. Tais pontos serão referidos aqui com o termo inglês *outliers*. Assim, ainda adotando-se a resolução de 5 nm, tomaram-se todas as amostras de ambos os locais, reservando-se apenas uma delas para a validação e foram ajustados modelos por PCR, incluindo os componentes principais pela ordem decrescente de variância e computando-se os valores de RMSEP obtidos para a amostra de validação com até 20 componentes figurando nos modelos. Em seguida, foram determinadas as médias para os valores de RMSEP assim obtidos e, também, o desvio-padrão dos

valores dessas médias. Este procedimento foi realizado tantas vezes quanto era o número de amostras total, incluindo-se ambos os locais. Uma vez calculados o desvio-padrão e a média do RMSEP, foram aprovadas para permanecer no conjunto de dados apenas as amostras cujo erro era maior que a média mais duas vezes o desvio-padrão dos erros.

A Figura 51 apresenta a distribuição dos valores de RMSEP médio para cada uma das amostras e sua situação em relação ao limite de rejeição e à média geral dos erros. Não houve nenhuma tendência quanto à distribuição dos erros fornecidos pelas amostras dos Locais 1 ou 2. Note-se na figura que a maior parte das amostras forneceram erros abaixo da média e que algumas amostras apresentaram erros que eram apenas ligeiramente superiores ao limite de rejeição. A permanência de tais amostras no conjunto dos dados poderia contribuir para a robustez dos modelos de calibração, porém, foram excluídas, já que não existe uma definição consensual acerca dos critérios de rejeição. Desta forma, aproximadamente 10% das amostras foram consideradas *outliers* e removidas do banco de dados e novos modelos foram ajustados, fornecendo diferentes resultados em termos de sua acurácia, como será visto a seguir.

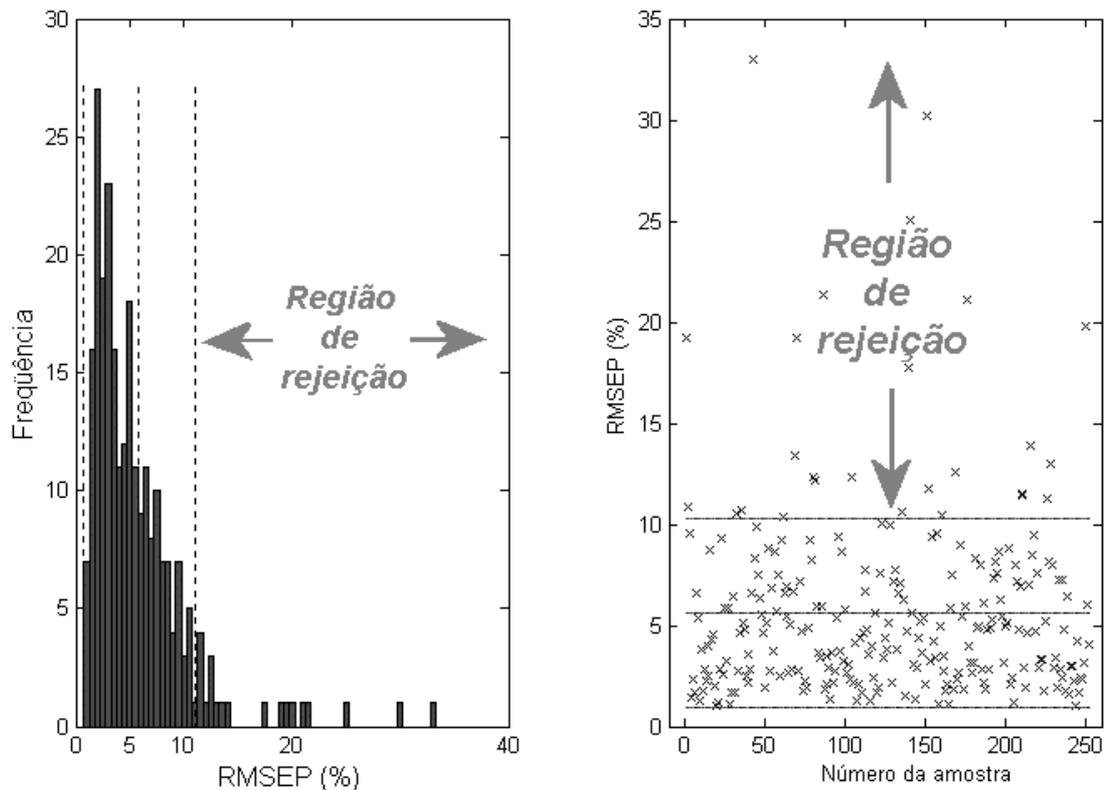


Figura 51 - Distribuição dos valores de RMSEP para cada uma das amostras na validação cruzada e sua situação em relação aos limites de rejeição e à média geral dos erros (linhas pontilhadas).

Optou-se por continuar o trabalho de otimização dos modelos agrupando os dados dos dois locais em uma única base, uma vez que, mesmo com a remoção dos *outliers*, os resultados obtidos ao se validar os modelos ajustados em um local sobre dados do outro local continuaram ruins, ou seja, fornecendo elevados RMSEP's e baixos coeficientes de determinação.

A Figura 52 apresenta o percentual da variância global dos espectros carregado por cada um dos vinte primeiros componentes principais e a sua correlação com a densidade básica da madeira após a remoção dos *outliers*. Não houve praticamente nenhuma alteração na variância dos componentes, o que significa que a PCA é uma técnica pouco susceptível à presença de *outliers* que se encontrem dentro de um intervalo razoável. Uma perceptível elevação, porém, pode ser observada no coeficiente de correlação entre os componentes de número 1 do Local 1 e número 3 do Local 2 com a densidade da madeira. A elevação dos coeficientes de correlação entre os componentes principais individualmente e a variável dependente pode significar a possibilidade de se obter melhores modelos múltiplos. O que, de fato, foi observado logo a seguir, analisando-se o Erro-padrão da estimativa e o RMSEP dos modelos ajustados após a remoção dos *outliers*.

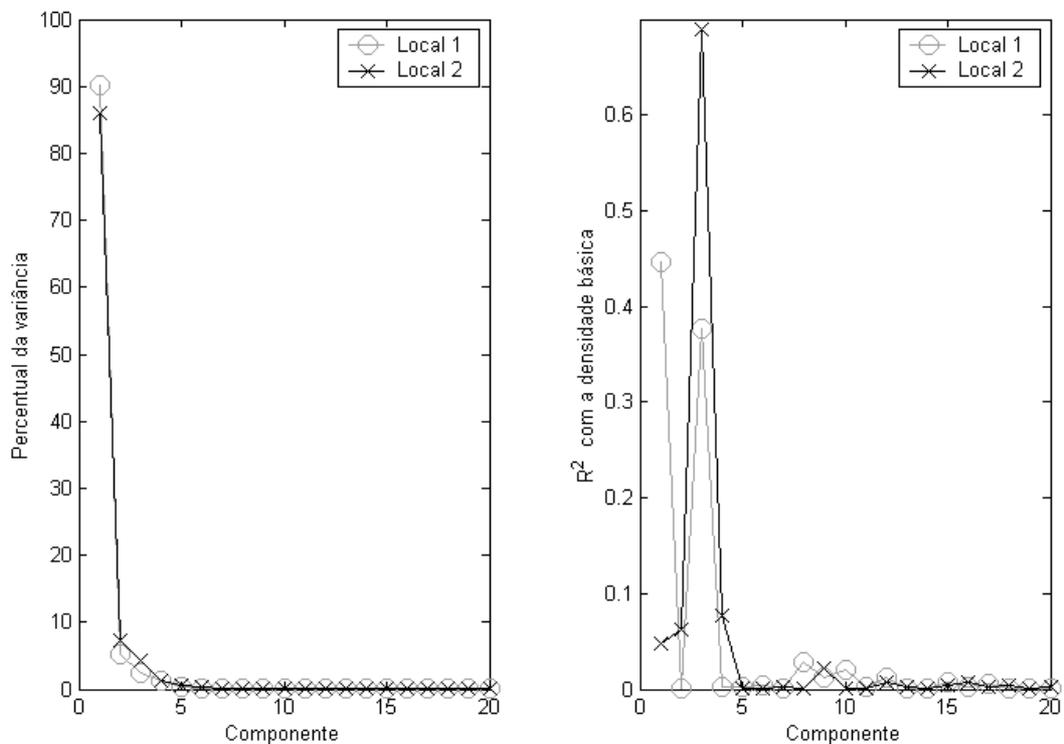


Figura 52 - Percentual da variância global carregado por cada um dos componentes principais (*esquerda*) dos espectros de reflectância após a remoção dos *outliers* e suas correlações individuais com a densidade básica da madeira.

A Figura 53 apresenta a distribuição dos erros mínimos e do número de componentes com que foram obtidos em modelos ajustados a partir da escolha aleatória de dados em 150 rodadas após a remoção dos *outliers*. Note-se que houve uma redução de 6,34 para 4,81% na média dos valores do RMSEP mínimo e que o número médio de componentes necessários para a obtenção de tais erros caiu de 42 para 31.

A Figura 53 apresenta ainda, o histograma para a distribuição do número de componentes correspondentes aos erros mínimos, que também sofreu mudanças em seu aspecto, apresentando-se não mais na forma bimodal, como antes da remoção dos *outliers* (Figura 49), exibindo uma acentuada assimetria, com pequena ocorrência de valores altos e concentração das maiores frequências em torno da média. Tal concentração revela um importante efeito da remoção dos *outliers* sobre os modelos, qual seja, a estabilização do número de componentes úteis quando comparados modelos ajustados sobre diferentes grupos de amostras de calibração e validação.

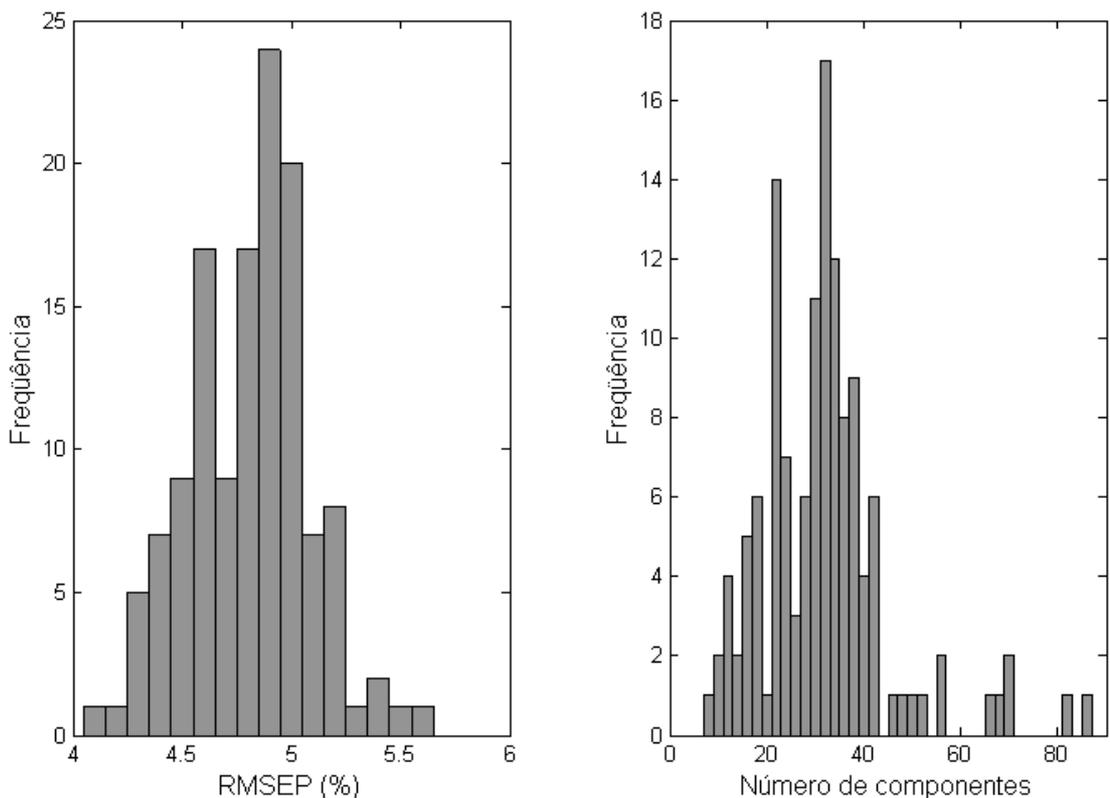


Figura 53 - Distribuição dos valores de RMSEP mínimos para a densidade básica da madeira dos Locais 1 e 2 e dos correspondentes números de componentes com que ocorreram em cada uma de 150 rodadas realizadas após a remoção dos *outliers*.

A Figura 54 exibe detalhes do efeito de estabilização do número de componentes principais necessários para a obtenção do RMSEP mínimo em 150 rodadas. A linha preta pontilhada representa a média dos erros obtidos com cada número de componentes nas várias rodadas. Note-se que todas as rodadas forneceram modelos bastante semelhantes no que diz respeito ao comportamento do erro, ao serem acrescentados os componentes pela ordem decrescente de variância. Tal comportamento aproxima-se mais daquele que é previsto teoricamente na validação de modelos ajustados por PCR, o que não era tão notável antes da remoção dos *outliers*.

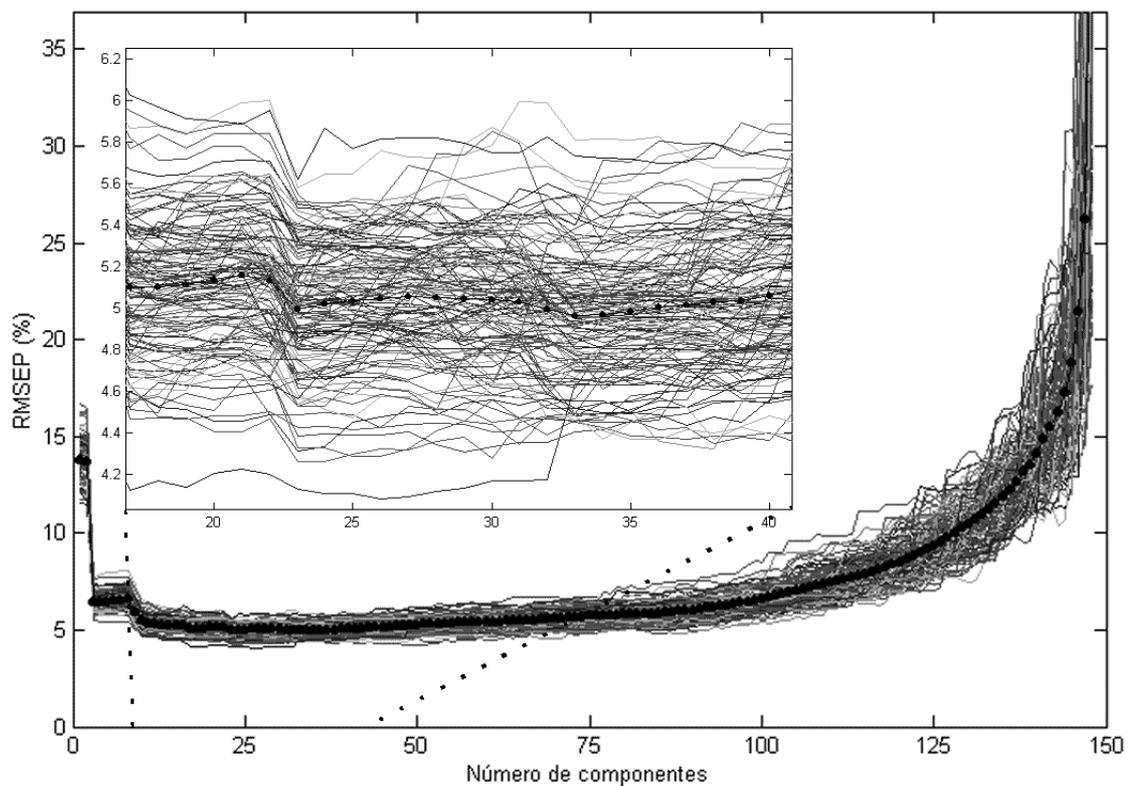


Figura 54 - RMSEP para modelos ajustados em 150 rodadas, tomando-se amostras para validação e calibração procedentes de ambos os locais. A linha pontilhada representa a média para cada número de componentes.

Ainda na Figura 54 é possível perceber-se que, aproximadamente, entre 15 e 80 componentes, o valor do RMSEP permanece praticamente inalterado, o que sugere não ser necessário o uso de números elevados de componentes para obterem-se erros razoáveis. Mais uma vez faz-se necessária a observação de que modelos que utilizam um menor número de variáveis independentes são sempre mais desejáveis, posto que, a

probabilidade de que o valor da estatística F na análise de variância da regressão seja significativo será sempre maior quanto menor for o número de variáveis independentes. Assim, ao escolher o número de componentes e, conseqüentemente, o nível de erro admitido, deve ser considerado o tamanho da amostra que foi extraída de uma população. Caso a amostragem seja ampla o bastante, como é o caso dos bancos de dados espectroscópicos permanentemente alimentados, pode-se optar pelos modelos de maior número de componentes sem, com isso, comprometer a sua robustez. Tal procedimento é adotado em laboratórios que mantêm serviços de espectroscopia e utilizam amostras de monitoramento dos modelos para a retroalimentação de seus modelos.

A Figura 55 apresenta a plotagem dos valores observados versus aqueles previstos por modelos que forneceram o menor, o maior e o RMSEP mínimo mais próximo da média de 130 rodadas, após a remoção dos *outliers*.

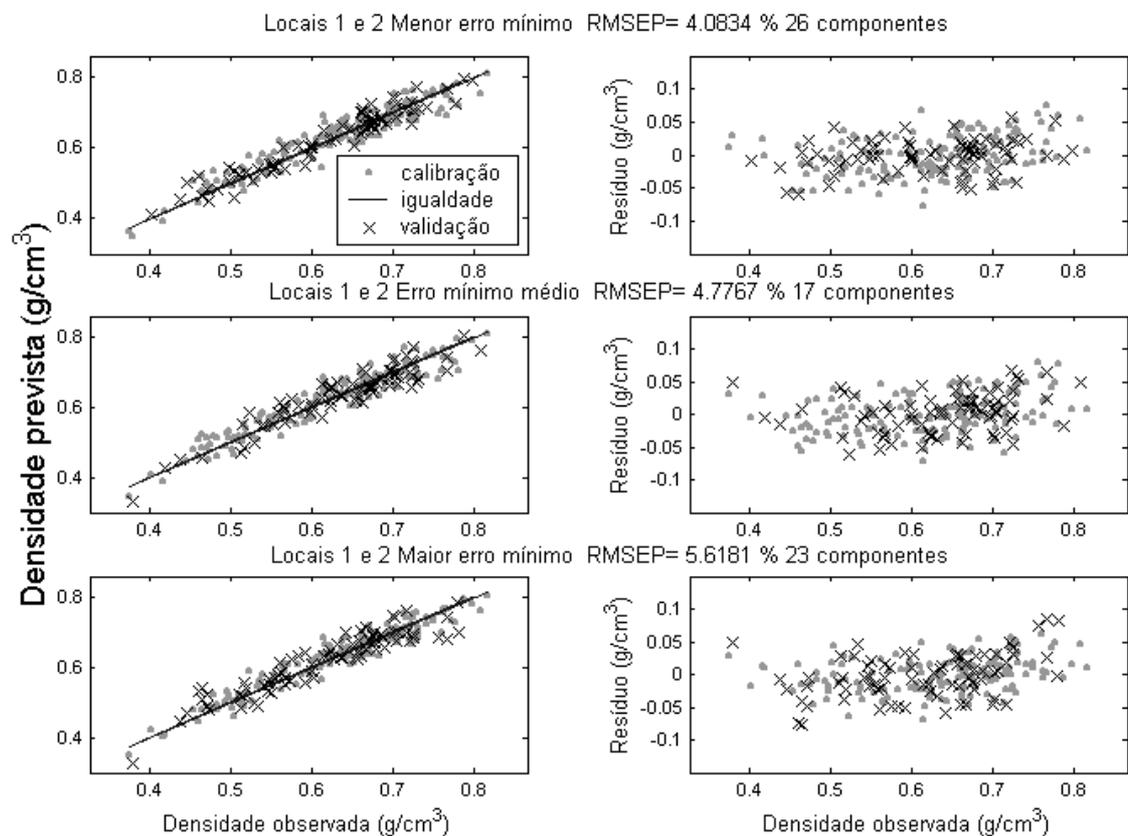


Figura 55 - Plotagem dos valores observados e previstos e os resíduos de modelos que apresentaram os erros de valor maior, menor e mais próximo da média dos RMSEP mínimos dentre 130 rodadas na validação para a previsão da densidade básica após a remoção dos *outliers*. Apresentados também os valores para RMSEP e o número de componentes principais utilizados em cada modelo. Amostras para a calibração e validação dos Locais 1 e 2.

O Quadro 7 apresenta a análise da variância para os modelos de calibração apresentados na Figura 55. Notem-se os elevados valores de F.

Quadro 7 - Análise da variância para as regressões múltiplas e erros padrão da estimativa (EPE) dos modelos de calibração que apresentaram os erros maior, menor e mais próximo da média dos RMSEP mínimos dentre 130 rodadas, na validação para a previsão da densidade básica após a remoção dos *outliers*. São apresentados também os coeficientes de determinação e erros. Amostras para a calibração e validação dos Locais 1 e 2

	Modelo de mínimo RMSEP			
	EPE = 5,1815	RMSEP = 4,0834	R ² da calibração = 0,8943	R ² da validação = 0,9265
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	26	1,1192	0,0430	40,3459
Resíduo	124	0,1323	0,0011	
Total	150	1,2514		
	Modelo de médio RMSEP			
	EPE = 5,3397	RMSEP = 4,7955	R ² da calibração = 0,8793	R ² da validação = 0,8941
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	11	1,1366	0,1033	92,0626
Resíduo	139	0,1560	0,0011	
Total	150	1,2926		
	Modelo de máximo RMSEP			
	EPE = 4,8520	RMSEP = 7,6338	R ² da calibração = 0,9179	R ² da validação = 0,8600
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	23	1,1455	0,0498	61,7358
Resíduo	127	0,1025	0,0008	
Total	150	1,2480		

5.2.3.2. Resolução espectral

Uma vez removidos os *outliers*, para testar-se a influência da resolução espectral sobre a capacidade de predição dos modelos, foram selecionadas, ao acaso, amostras que compuseram dois grupos fixos e distintos: o grupo da calibração e o grupo da validação. Tal procedimento tornou-se necessário pelo imenso incremento na quantidade de dados a serem processados e, conseqüentemente, no consumo de memória do computador, causados pela variação da resolução, ao mesmo tempo em que variavam as amostras (rodadas). Assim, as rodadas foram aqui abolidas e modelos com

diferentes números de componentes principais foram ajustados sobre o mesmo grupo de amostras. Tais modelos tiveram computados os seus erros mínimos de calibração e validação (S_{xy} ou EPE e RMSEP), os coeficientes de determinação na calibração e validação e o valor das estatísticas p e F da análise da variância das regressões que originaram cada curva de calibração.

Assim, foram tomadas como variáveis independentes, as médias da reflectância em diferentes intervalos de comprimento de onda em cada um dos espectros, como ilustrado na Figura 56. Para obter-se uma resolução espectral de 50 nm, o intervalo entre 350 e 1000 nm foi dividido em 13 janelas, dentro das quais foram calculadas as médias para a reflectância e seus valores atribuídos ao comprimento de onda central.

Considerando-se que a resolução do espectrômetro utilizado era de 0,31 nm e que foi registrada a reflectância em uma faixa de amplitude igual a 810 nm, conclui-se que, o número de comprimentos de onda a serem computados era de 2375. Lembrando-se ainda, que este é também o número de colunas da matriz de variáveis independentes originais, pode-se imaginar o esforço computacional para o processamento dos dados quando se deseja utilizar a resolução máxima do aparelho.

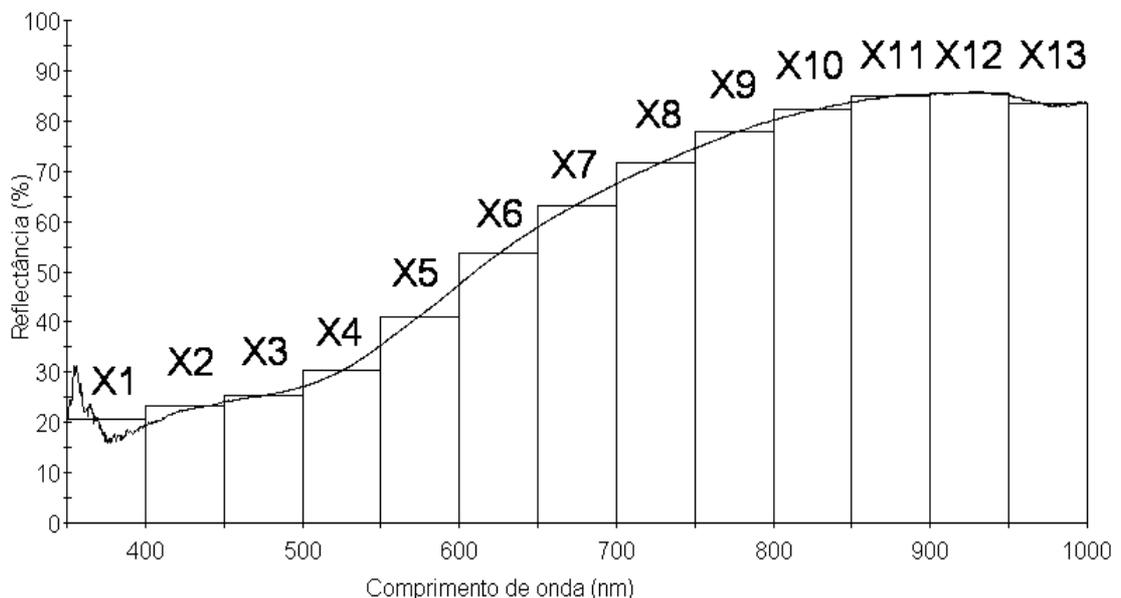


Figura 56 - Espectro praticamente contínuo com resolução de 0,31 nm sendo exageradamente discretizado ao assumir-se uma resolução de 50 nm.

De fato, um computador com processador de 2,4 GHz levou, em média, um minuto para ajustar um único modelo contendo os primeiros vinte componentes

principais que, por sua vez, carregam a informação contida em 2375 variáveis independentes. Lembre-se que todos os cálculos necessários à realização deste trabalho foram realizados por meio de algoritmos especialmente desenvolvidos para este propósito sobre a plataforma MATLAB[®]. Certamente, alguma otimização dos algoritmos para se incrementar a velocidade de processamento poderia ser realizada, dado que não se conta com peritos em informática na equipe do laboratório. Apesar de tal consideração, é notável o consumo de tempo e memória nesse tipo de processamento de dados.

Outro inconveniente trazido pela utilização da máxima resolução espectral vem do fato de que, nesse nível de medição da reflectância, a susceptibilidade do sensor a ruídos aleatórios fica mais evidenciada. Ou seja, a precisão é reduzida a medida que aumenta a resolução. Os ruídos aleatórios, provocados pelo próprio circuito eletrônico do aparelho, podem interferir negativamente na capacidade de predição dos modelos.

Por outro lado, ao reduzir-se em demasiado a resolução espectral, perdem-se detalhes dos espectros que poderiam fornecer informações valiosas para os modelos. Assim, ao tomar-se a média da reflectância em um determinado intervalo de comprimentos de onda, ocorre a deformação do espectro, ocultando picos de absorção ou reflexão que, muitas vezes, são determinantes para o sucesso da aplicação da espectroscopia. A Figura 56 ilustra a deformação sofrida por um espectro ao serem substituídas a reflectâncias a cada 0,3 nm, pelas médias de intervalos de 50 nm.

Apesar dos possíveis inconvenientes citados anteriormente, pode-se observar na Figura 57 que a redução da resolução, desde a máxima (0,30 nm) até, aproximadamente, 20 nm, não causou prejuízos apreciáveis para a performance dos modelos. Note-se que os valores dos coeficientes de determinação e dos erros dos modelos permaneceram estáveis neste intervalo.

Ainda na Figura 57, é possível notar-se que a estatística F sofre um incremento até atingir um valor máximo por volta da resolução de 130 nm para, em seguida, decrescer com o incremento na amplitude da janela de médias. Tal fato não parece estar relacionado à capacidade de predição dos modelos, já que, nesta região, os modelos de calibração não apresentaram nada de especial em seus erros ou coeficientes de correlação. O decréscimo do valor da estatística F observado quando a janela de média continua aumentando está relacionado ao simultâneo incremento no número de componentes principais necessários para a obtenção do RMSEP mínimo. Ou seja, a probabilidade de que alguns dos coeficientes de regressão sejam estatisticamente nulos aumenta, o que sugere o aumento na instabilidade dos modelos.

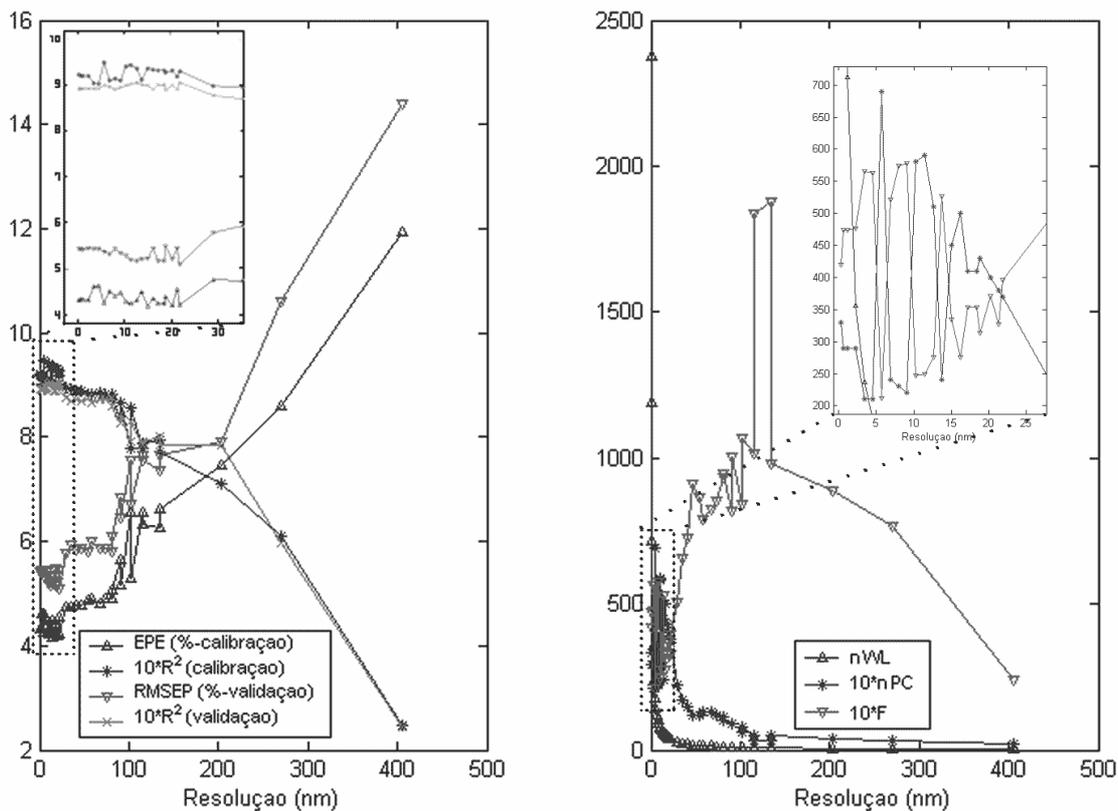


Figura 57 - Variação dos valores de R^2 e dos erros de predição (EPE e RMSEP) na calibração e validação de modelos que forneceram os menores RMSEP mínimos para a densidade básica, utilizando-se vários níveis de resolução espectral. nWL= número de comprimentos de onda; nPC= número de componentes principais utilizados nos modelos de menor erro mínimo, F= estatística do teste de Fisher para as análises de regressão que forneceram os modelos de menor erro mínimo.

5.2.3.3. Suavização das curvas espectrais

A Figura 61 apresenta o efeito da suavização das curvas espectrais realizada através do filtro polinomial de Savitzky-Golay sobre os espectros quando adotados polinômios do 3º, 6º e 9º graus ajustados sobre intervalos de comprimento de onda contendo 10, 100 e 1.000 pontos (janelas de filtragem). Em todos os casos foram utilizados os mesmos grupos de amostras de calibração e validação sobre os quais foram testados os níveis de resolução.

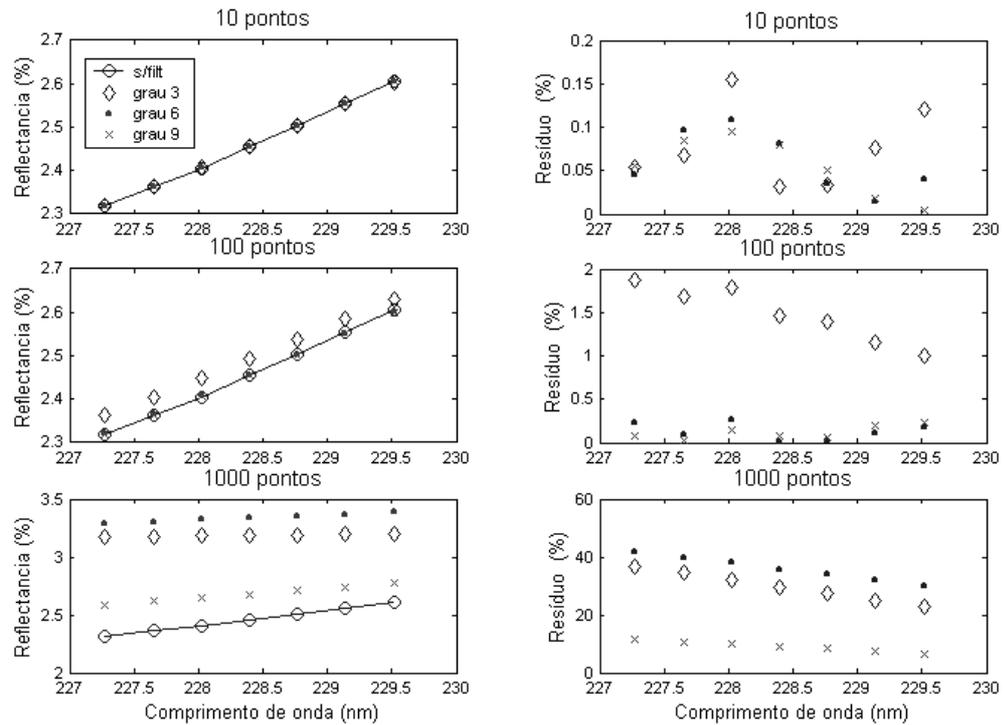


Figura 61 - Efeito da filtragem polinomial sobre uma faixa do espectro de uma amostra. Os gráficos da esquerda apresentam os valores originais e aqueles calculados com os filtros para reflectância. Os gráficos da direita mostram os resíduos em relação aos valores originais de cada ponto em termos percentuais.

Note-se na Figura 61 que, independente do grau adotado para o polinômio, quanto maior a janela de filtragem, maior é a transformação sofrida pelo espectro, ou seja, maiores são os resíduos em relação aos valores originais da reflectância. O grau do polinômio, por sua vez, fornece resultados que estão relacionados ao tamanho do intervalo de filtragem. De maneira geral, como era de se esperar, os polinômios de grau 9 produziram as menores distorções nos espectros para todas as janelas de filtragem testadas. Já os filtros de graus 3 e 6 apresentaram oscilações para os desvios entre os espectros originais e filtrados ao variar-se o tamanho das janelas. Este último fato pode ser explicado pela própria forma original dos espectros que, ora era mais bem descrita pelas curvas de 3º grau, ora pelas de 6º, a depender da região do espectro em que se encontrava a janela de filtragem. De qualquer forma, ao se aplicar filtros sobre os espectros, o que se deseja é remover ruídos aleatórios que, porventura, estejam ali presentes, com a menor deformação possível. A determinação do tipo e da configuração dos filtros e sua eficácia estão condicionados aos efeitos que produzem sobre a capacidade de predição dos modelos, como será visto adiante.

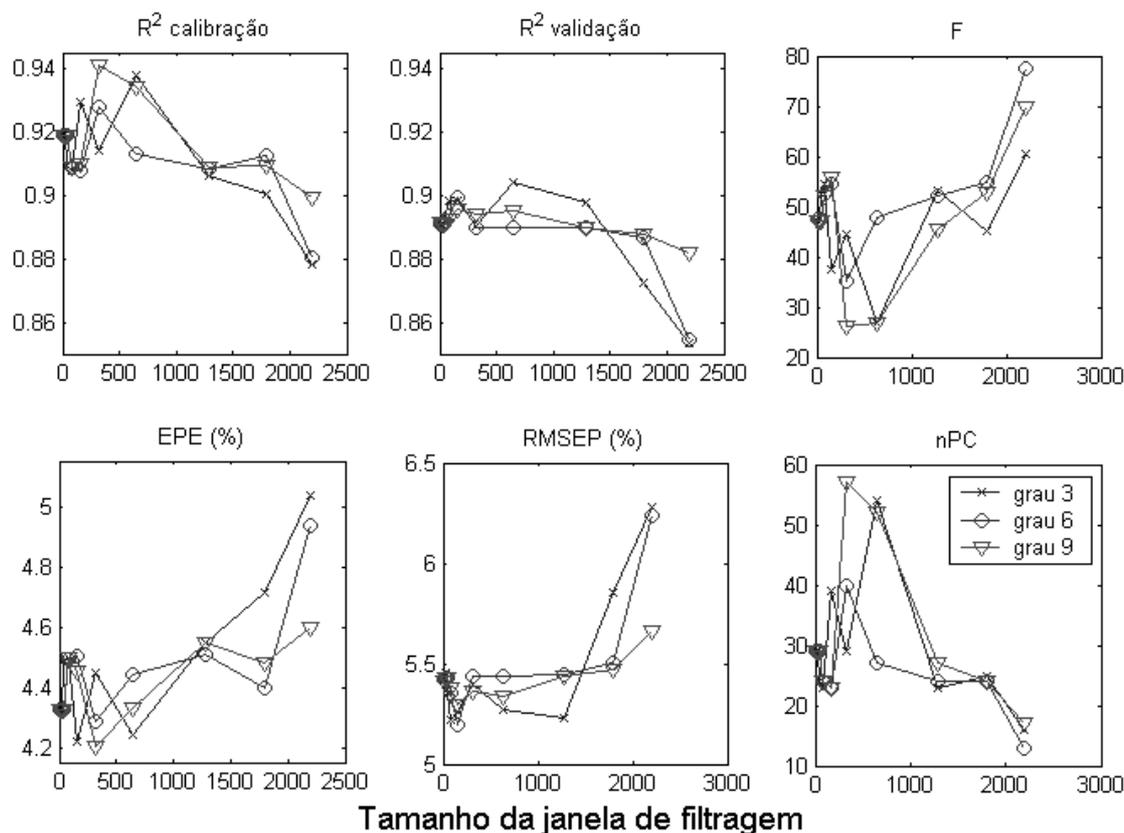


Figura 62 - Efeito da filtragem polinomial sobre algumas características dos modelos ajustados para a previsão da densidade básica a partir da reflectância. Foram testados filtros de graus 3, 6 e 9 aplicados sobre janelas de variados tamanhos.

A Figura 62 apresenta o efeito da filtragem polinomial sobre os coeficientes de determinação, erros de calibração e validação, número de componentes necessários à obtenção do erro mínimo e valores da estatística F dos modelos. Para os três graus de polinômios testados, ao aumentar-se o tamanho da janela de filtragem até um intervalo de aproximadamente 150 pontos, observaram-se ligeiras quedas nos valores de RMSEP, do R^2 para a calibração e do número de componentes necessários ao RMSEP mínimo, enquanto que, por outro lado, observaram-se ligeiros incrementos nos valores de EPE, do R^2 para a validação e da estatística F. Após o referido intervalo, os valores de RMSEP retornam para os patamares originais mantendo-se relativamente constantes até que o intervalo de filtragem se aproxime de 1500, quando começam a crescer.

De maneira geral, a capacidade de predição dos modelos revelou-se mais sensível às variações no tamanho da janela de filtragem quando são aplicados polinômios do 3º grau, enquanto que aqueles de 9º grau provocam menores variações nos parâmetros de qualidade dos modelos. Tal fato deve-se, provavelmente, à menor

deformação causada nos espectros originais quando são utilizados filtros polinomiais de grau elevado, já que as curvas apresentam menores resíduos em relação aos valores anteriores à filtragem.

Em resumo, a filtragem polinomial não se mostrou eficaz ou mesmo necessária no tratamento dos espectros quando a reflectância é tomada como variável independente, posto que seus efeitos sobre a acurácia e precisão dos modelos são, ora negativos, ora nulos ou desprezíveis. Tais fatos podem ser explicados pela maneira com que foram tomados os dados. Ou seja, chamou-se, aqui, de “espectro individual” à média de 6 espectros tomados de uma amostra (item 3.2.1.1.4.). Como os ruídos são aleatórios, devem apresentar média igual a zero. Desta forma, os espectros sofreram um tipo de filtragem já durante a aquisição e compilação dos dados e as tentativas de aplicação de filtros subseqüentes podem ter ocasionado perdas de informações relevantes contidas em pequenas sinuosidades da curva espectral.

5.2.3.4. Transformações sobre os espectros

5.2.3.4.1. Derivação das curvas espectrais

Com os objetivos de realçar o efeito das pequenas sinuosidades acima referidas, de anular possíveis erros aleatórios ou sistemáticos na intensidade da reflectância provocados pela peculiar dispersão da radiação em cada amostra e de reduzir a autocorrelação entre as variáveis independentes, procedeu-se à extração das derivadas das curvas espectrais. Assim, tomando-se a reflectância como função do comprimento de onda [$R=f(\lambda)$] e assumindo-se que o espectro é uma curva contínua e derivável em todo a faixa de trabalho, calcularam-se as derivadas a partir da razão $\Delta R/\Delta\lambda$ em cada ponto do espectro. A derivada segunda, por sua vez, é calculada dividindo-se a razão $\Delta R/\Delta\lambda$ novamente pela variação do comprimento de onda [$\Delta R/(\Delta\lambda)^2$]. Pela maneira com que são calculadas as derivadas, percebe-se imediatamente que, tanto a resolução espectral, quanto a filtragem das curvas, afetam sensivelmente os seus valores e, conseqüentemente, a utilidade desses valores como variáveis preditoras.

5.2.3.4.1.1. Derivada primeira

A Figura 63 apresenta o aspecto da curva espectral e de sua primeira derivada após sofrer uma filtragem polinomial do terceiro grau e redução da resolução para 5 nm

em toda a faixa de trabalho. Note-se que, apesar da filtragem e redução da resolução, a curva da primeira derivada apresenta picos em regiões onde a curva espectral apresentava-se aparentemente suave. Os ruídos na região de transição para o infravermelho (700 a 1000 nm) são também muito evidenciados por esta transformação.

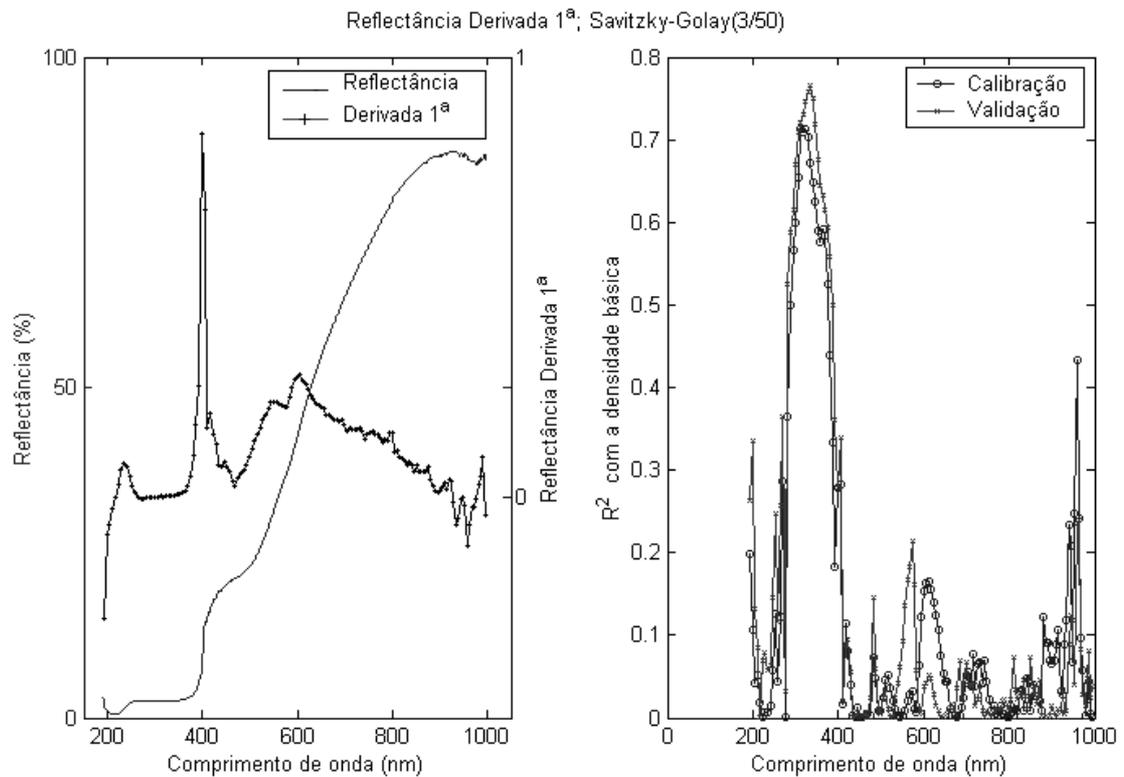


Figura 63 - O espectro de reflectância de uma amostra e sua respectiva derivada primeira (*esq.*). Correlações entre os valores da derivada primeira e a densidade básica (*dir.*).

A transformação dos espectros em suas primeiras derivadas promoveu uma sensível redução no nível de autocorrelação entre as variáveis independentes, como pode ser observado na Figura 64. O surgimento de profundos vales onde, antes da derivação, existiam superfícies elevadas e contínuas (Figura 42) evidencia a decorrelação provocada pela derivação. Tal efeito deve-se à supressão da influência da reflectância média de cada amostra que pode conter erros sistemáticos na tomada da linha de base e que podem ser causados por pequenas oscilações na tensão de alimentação da lâmpada do aparelho.

É possível que modelos ajustados diretamente sobre os valores das derivadas dos espectros por regressão múltipla com base no método dos mínimos quadrados ordinários *forward* possam fornecer melhores resultados que aqueles ajustados sobre os

espectros, já que a colinearidade entre as variáveis independentes foi reduzida pela derivação. Tal procedimento, entretanto, não foi aqui testado.

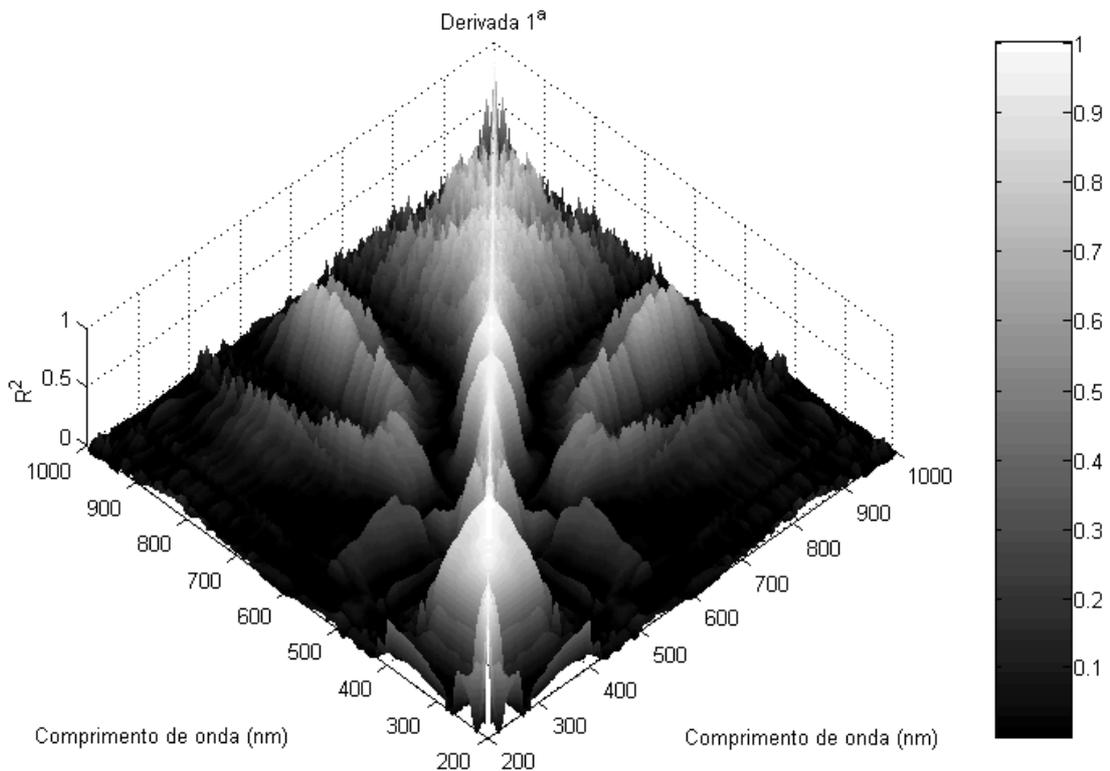


Figura 64 - Coeficientes de determinação entre as variáveis independentes, neste caso, a primeira derivada da reflectância.

A Figura 65 apresenta o percentual da variância global dos dados carregada pelos primeiros 30 componentes principais das derivadas dos espectros, bem como os coeficientes de determinação entre estes componentes e a densidade básica da madeira. Aqui é interessante notar que os primeiros componentes carregam um percentual muito menor da variância global do que antes observado para os componentes extraídos diretamente da curva espectral. Tal fato deve-se à contribuição dos ruídos realçados pela derivação da reflectância, os quais interferem nas correlações entre os componentes principais e a variável independente.

Assim, ao utilizarem-se as derivadas como variáveis independentes em modelos de calibração em espectroscopia é de crucial importância considerar-se o nível da relação sinal/ruído, pois, como anteriormente comentado, os ruídos realçados pela operação de derivação podem acrescentar informações irrelevantes ao conjunto de dados. A aplicação de filtros de suavização (passa-baixas) e a redução da resolução muitas vezes se tornam necessárias para que a relação sinal/ruído não seja reduzida.

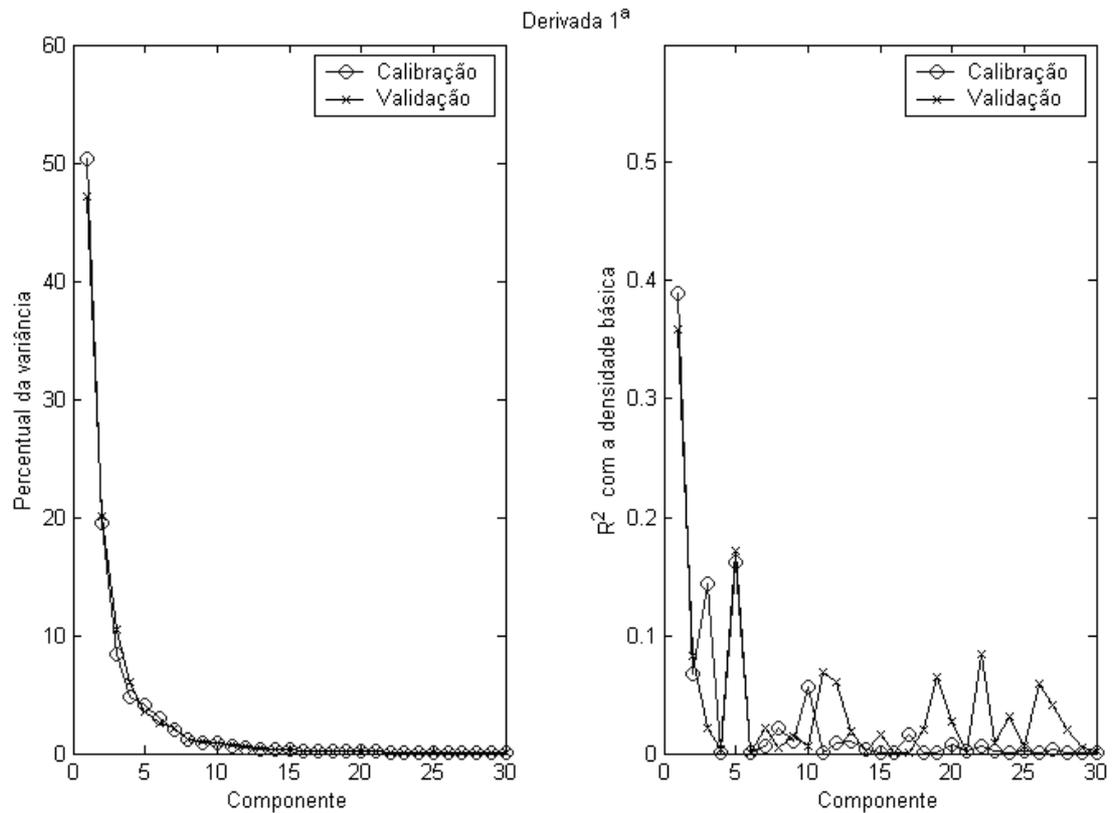


Figura 65 - Percentual da variância global carregado por cada um dos 30 primeiros componentes principais da derivada primeira dos espectros (*esquerda*) e suas correlações individuais com a densidade básica da madeira.

A Figura 66 apresenta o efeito da resolução e de filtros polinomiais sobre a qualidade de modelos ajustados para a previsão da densidade básica a partir da primeira derivada da reflectância. Note-se que os resultados apresentados pelos modelos ajustados sobre a derivada dos espectros não filtrados foram os que sofreram maior influência da variação da resolução espectral. Isso se deve, provavelmente, ao efeito de filtragem obtido com o cálculo das médias da reflectância nos intervalos que determinaram as resoluções testadas (0,31-1,00-3,00-5,00 e 10,00 nm).

Assim, para os espectros não filtrados, ao ser reduzida a resolução, pôde-se observar que os valores do coeficiente de determinação da validação, os respectivos RMSEP's, o valor de F e do número de componentes necessários à obtenção do erro mínimo na validação sofreram variações consideráveis. Os valores de R^2 e EPE, por sua vez, foram afetados de forma positiva apenas inicialmente, com a variação da resolução de 0,31 para 1nm, assumindo valores mais ou menos estáveis a partir da resolução de 3 nm.

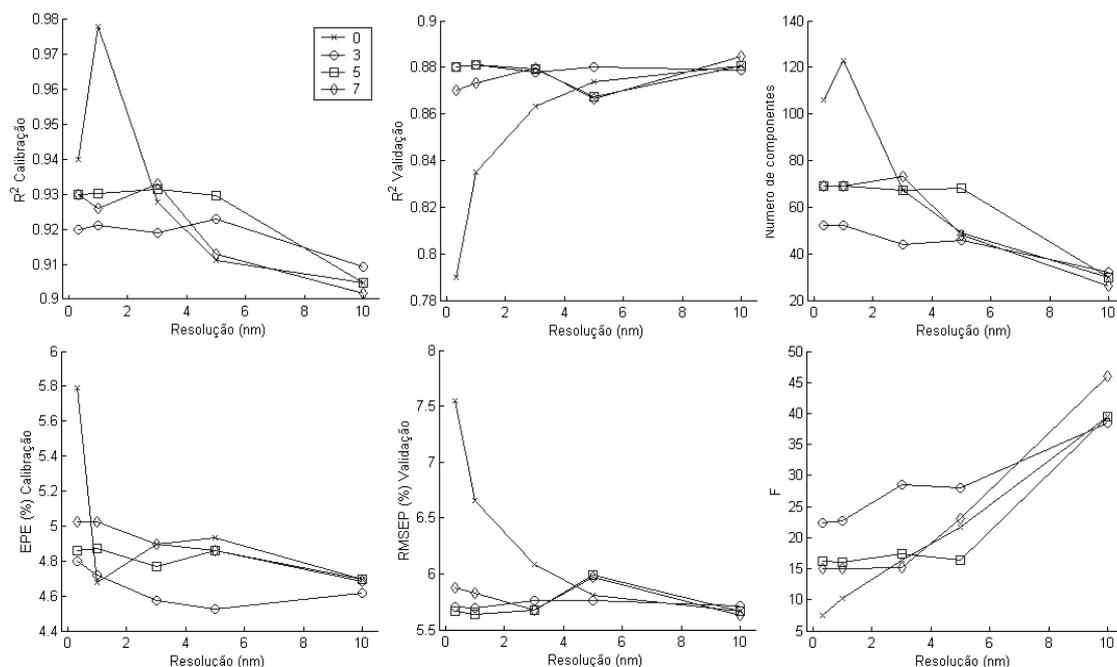


Figura 66 - Efeito da resolução variando com os valores 0,31-1,00-3,00-5,00 e 10,00 nm e de filtros polinomiais sobre a qualidade de modelos ajustados para a previsão da densidade básica a partir da primeira derivada da reflectância. As curvas de grau 0 indicam espectros não filtrados. Foi adotada uma janela de filtragem igual 50 pontos.

De maneira geral, para os modelos ajustados sobre a primeira derivada dos espectros, o filtro polinomial do terceiro grau apresentou os melhores resultados, posto que seus parâmetros de qualidade não se mostraram susceptíveis a bruscas variações quando variava a resolução espectral. Assim, a combinação do filtro polinomial de grau 3 com a resolução de 5 nm parece ser a alternativa mais razoável para uso com a derivada primeira, pois resoluções menores implicam em maior tempo de processamento e consumo de memória, além da elevação do nível de ruídos e suas conseqüentes perdas em acurácia e precisão. Resoluções ainda menores que 5 nm, por sua vez, acarretam perdas de informações relevantes contidas em pequenas nuanças dos espectros e suas derivadas.

De fato, a filtragem do terceiro grau associada à resolução de 5 nm forneceram resultados bastante satisfatórios quando fez-se variar por 100 vezes as amostras tomadas como grupos de calibração e validação, para o ajuste de modelos por PCR sobre as primeiras derivadas dos espectros como preditoras da densidade básica da madeira. A Figura 67 apresenta a distribuição dos erros mínimos na validação e no número de componentes principais necessários à sua obtenção.

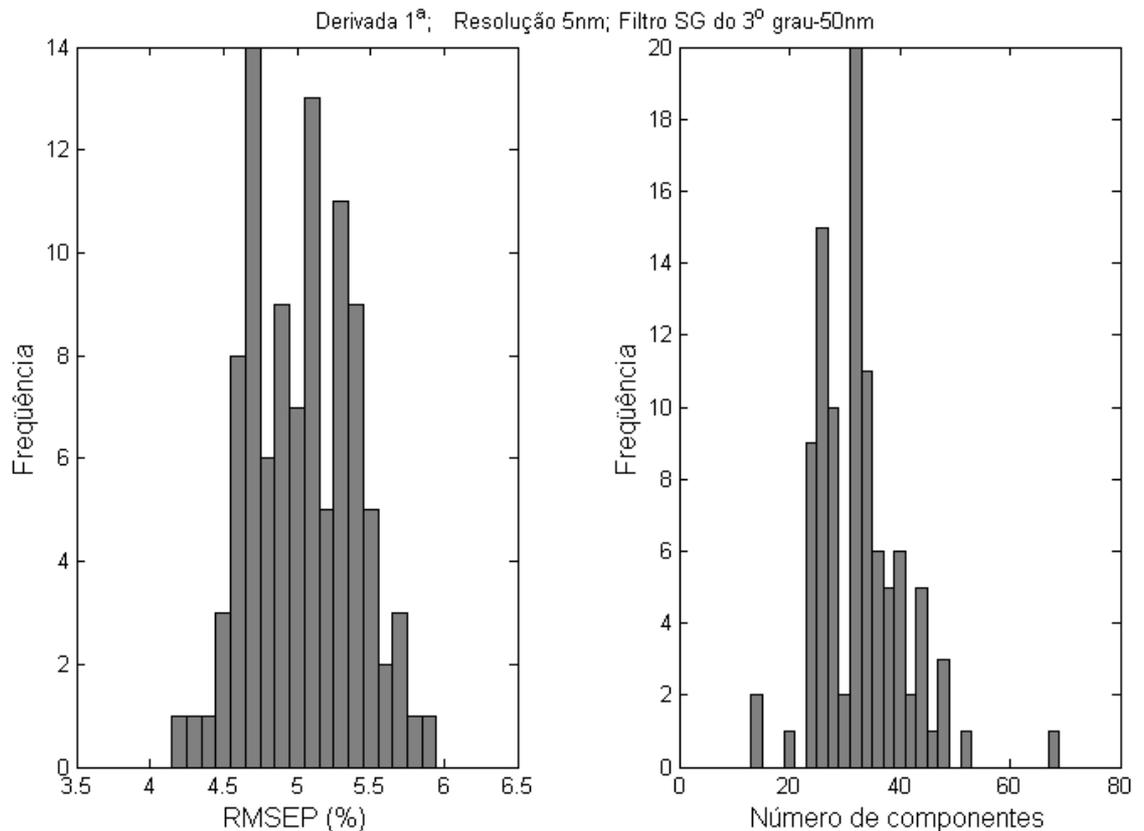


Figura 67 - Distribuição de freqüências para os erros mínimos na validação (RMSEP) e para o número de componentes principais necessários à sua obtenção quando utilizada a primeira derivada dos espectros na predição da densidade básica em 100 rodadas.

Com essas condições de análise a média do RMSEP mínimo e o número de PC's necessários à obtenção de tais erros, apresentaram os valores de 5,04% e 33,34, respectivamente. Lembre-se que os valores médios destes parâmetros obtidos ao ajustarem-se modelos diretamente sobre a reflectância, após a remoção dos *outliers* encontravam-se em 4,8% para o RMSEP mínimo e 31 componentes. Ou seja, a transformação dos espectros pela primeira derivada parece ser uma opção ligeiramente mais vantajosa que a utilização direta da reflectância para se prever a densidade básica da madeira.

A Figura 68 apresenta, para a densidade básica da madeira, a plotagem dos valores observados e previstos pelos modelos ajustados sobre os valores da primeira derivada da reflectância e que apresentaram o valor maior, o menor e o mais próximo da média dos RMSEP's mínimos obtidos em 100 rodadas. Pode-se notar na figura que os resíduos permanecem distribuídos aleatoriamente.

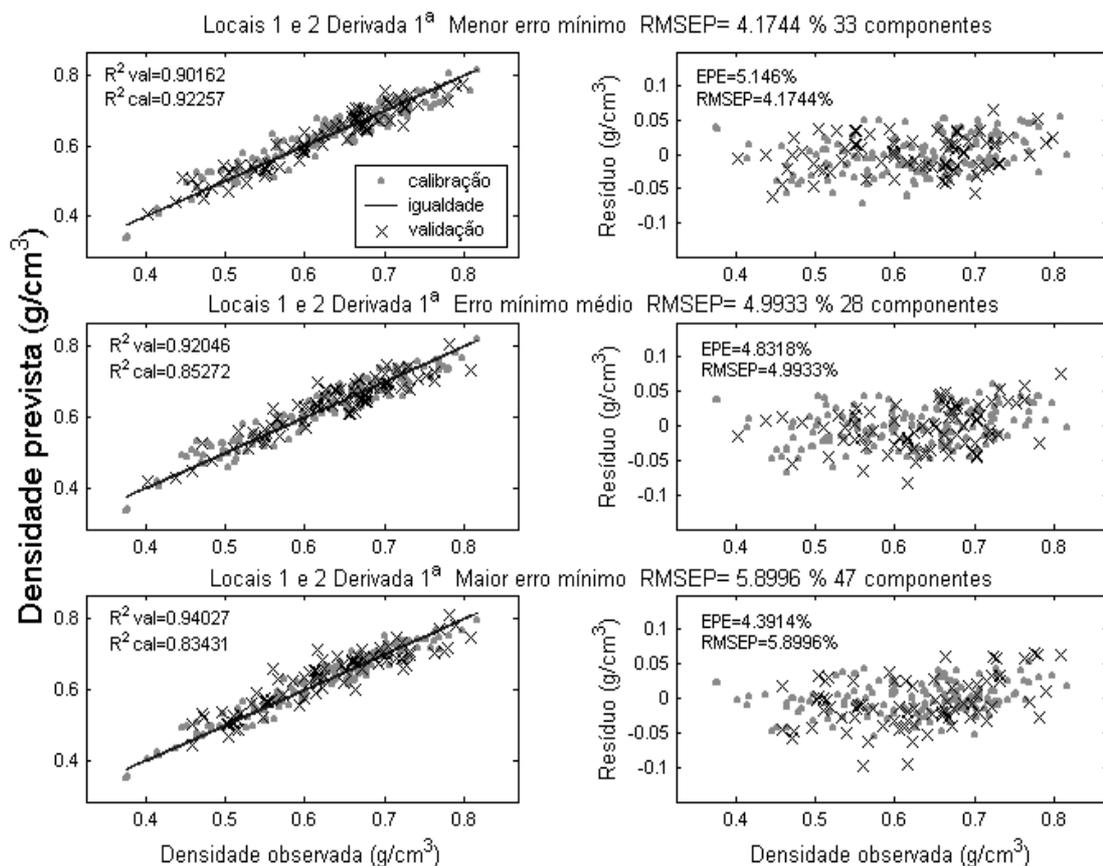


Figura 68 - Plotagem dos valores observados versus previstos e os resíduos de modelos ajustados sobre a primeira derivada dos espectros que apresentaram RMSEP maior, menor e mais próximo da média dos erros mínimos na validação para a previsão da densidade básica. São apresentados também os valores do R^2 na calibração e validação, além do EPE, do RMSEP e o número de componentes principais utilizados em cada modelo.

5.2.3.4.1.2. Derivada segunda

A Figura 69 apresenta o aspecto da curva espectral e de sua segunda derivada após sofrer uma filtragem polinomial do terceiro grau e redução da resolução para 5 nm em toda a faixa de trabalho. Note-se que a curva da segunda derivada apresenta-se ainda mais sinuosa que a da primeira (Figura 63) e que os ruídos na região de transição para o infravermelho (700 a 1000 nm) são ainda mais acentuados pela segunda derivação.

A extração da segunda derivada dos espectros provocou uma queda na colinearidade das variáveis independentes ainda maior que aquela observada para a primeira derivada, como pode ser visto na Figura 70. Entretanto, aqui é importante observar que o efeito de ambas as derivações sobre as autocorrelações pode ser

meramente fruto do acaso e que o aspecto extremamente rugoso apresentado pela curva de correlações entre as derivadas e a variável independente pode significar a inclusão de informações irrelevantes ao conjunto de dados. Este último fato pode ter como consequência a instabilidade do número de componentes necessário à obtenção do erro mínimo quando se variam as amostras de validação e calibração entre várias rodadas, como será visto mais adiante.

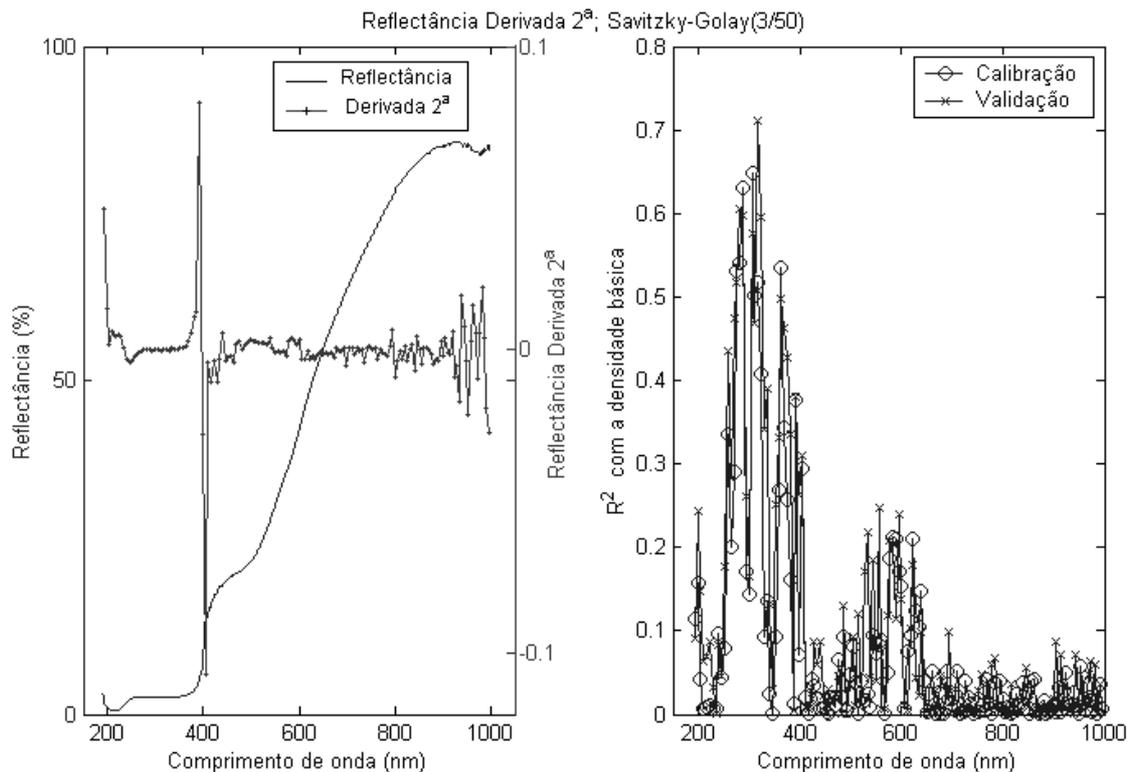


Figura 69 - O espectro de reflectância de uma amostra e sua respectiva derivada segunda (*esq.*). Correlações entre os valores da derivada primeira e a densidade básica (*dir.*).

A Figura 71 apresenta o percentual da variância global dos dados carregada pelos primeiros 20 componentes principais das segundas derivadas dos espectros, bem como os coeficientes de determinação entre estes componentes e a densidade básica da madeira. Note-se que os primeiros componentes voltaram a carrear um alto percentual da variância global, como antes observado para os componentes extraídos diretamente da curva espectral (Figura 43). Tal fato, porém, não afetou sensivelmente correlações entre os componentes e a variável dependente.

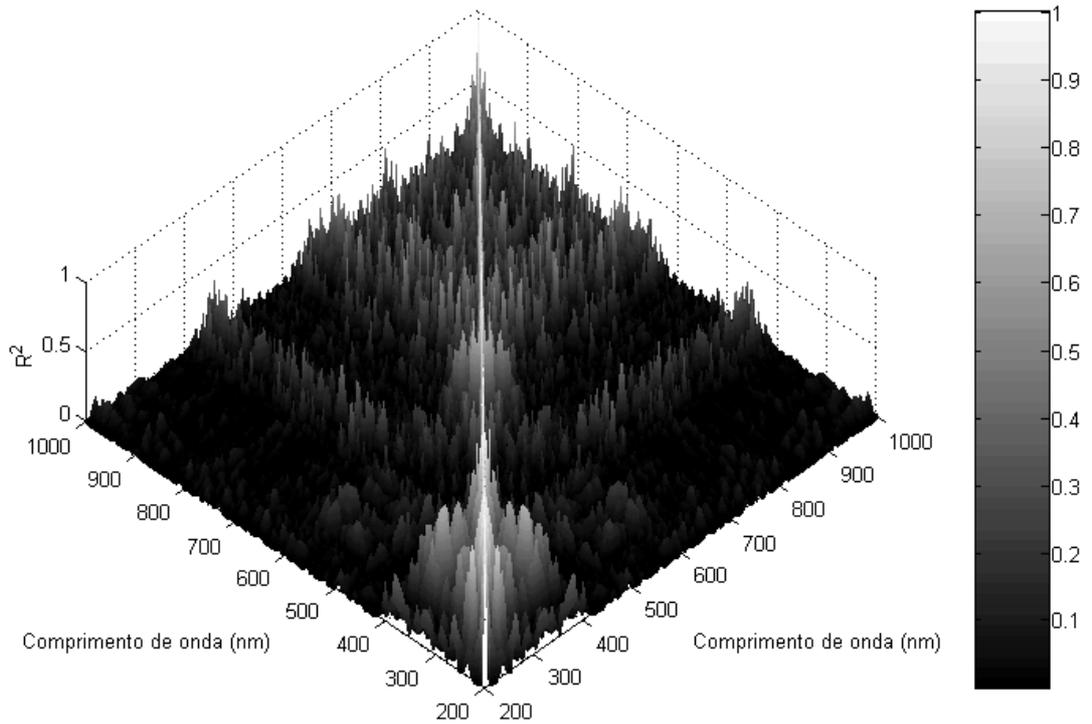


Figura 70 - Coeficientes de determinação entre os valores da segunda derivada em cada comprimento de onda.

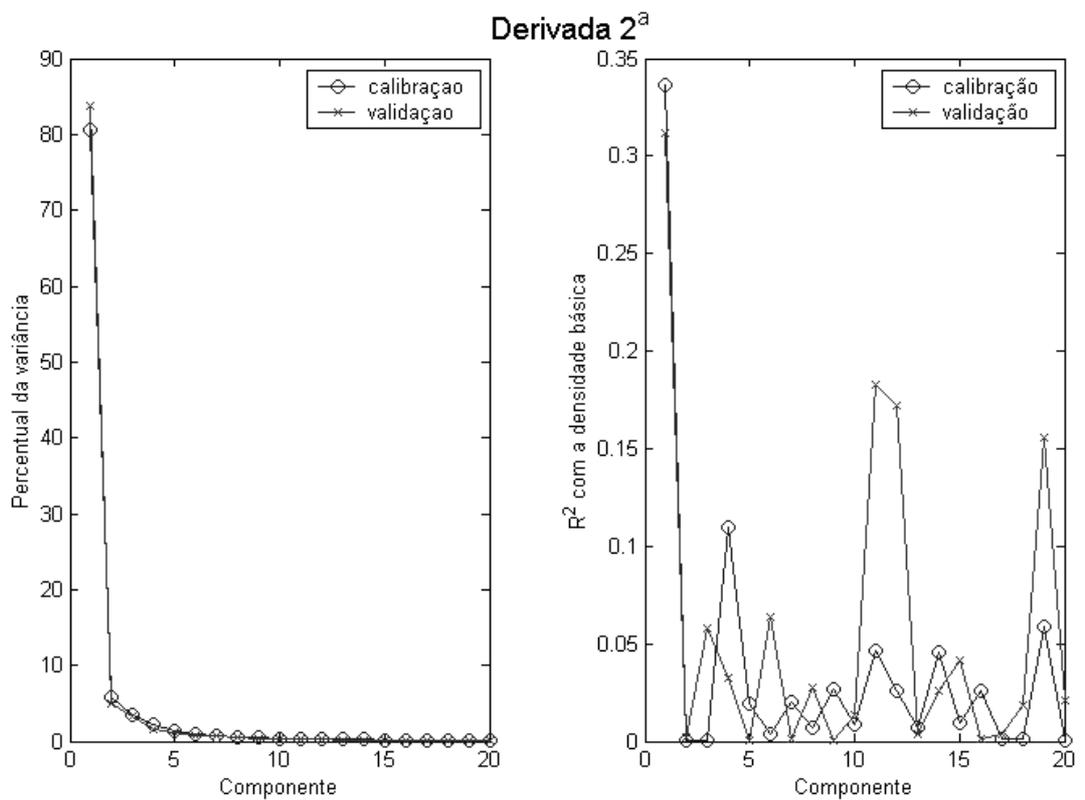


Figura 71 - Percentual da variância global carregado por cada um dos 20 primeiros componentes principais da derivada segunda dos espectros (*esquerda*) e suas correlações individuais com a densidade básica da madeira.

A Figura 72 apresenta o efeito da resolução e de filtros polinomiais sobre a qualidade de modelos ajustados para a previsão da densidade básica a partir da segunda derivada da reflectância. À semelhança do que foi observado para a primeira derivada, os resultados apresentados pelos modelos ajustados sobre a segunda derivada dos espectros não filtrados também sofreram grande influência da variação da resolução espectral. Mais uma vez, acredita-se que tal fato seja explicado pelo efeito de filtragem obtido com o cálculo das médias da reflectância nos intervalos que determinaram as resoluções testadas (0,31-1,00-3,00-5,00 e 10,00 nm).

Dentre os modelos obtidos a partir da segunda derivada de espectros que sofreram algum tipo de filtragem, aqueles suavizados com o filtro do terceiro grau parecem ter fornecido os melhores resultados, pois os modelos apresentaram consistentemente maiores coeficientes de determinação e menores erros na calibração e validação. Por outro lado, os filtros do terceiro grau fizeram com que, para a obtenção dos erros mínimos (RMSEP), fosse necessária a utilização de um grande número de componentes nos modelos, fornecendo também baixos valores para razão F.

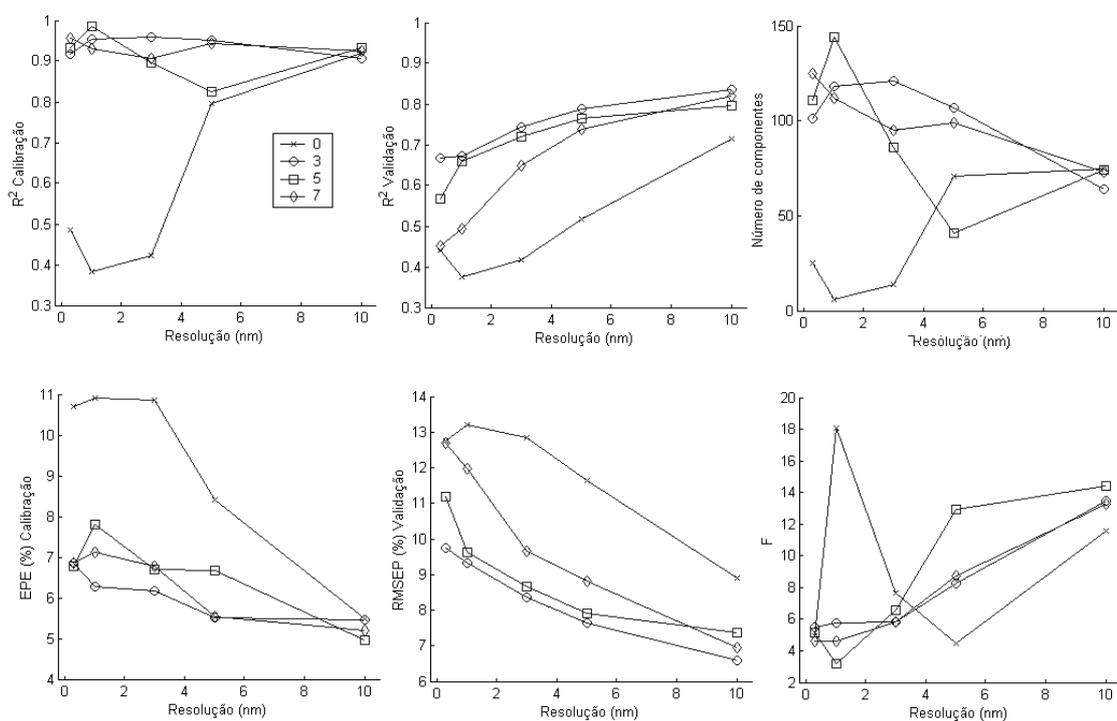


Figura 72 - Efeito da resolução variando com os valores 0,31-1,00-3,00-5,00 e 10,00 nm e de filtros polinomiais sobre a qualidade de modelos ajustados para a previsão da densidade básica a partir da primeira derivada da reflectância. As curvas de grau 0 indicam espectros não filtrados. Foi adotada uma janela de filtragem igual 50 pontos.

Assim, de maneira geral, o uso da segunda derivada dos espectros como variável preditora da densidade básica não forneceu bons resultados em termos de valores dos erros de calibração e validação, nem tampouco, do número de componentes necessários para a obtenção do RMSEP mínimo e da razão F para os grupos de amostras escolhidos para a calibração e validação.

Para os modelos ajustados sobre a segunda derivada dos espectros sobre os quais foram aplicados filtros polinomiais do terceiro grau e subsequente redução da resolução espectral para 5 nm, em um conjunto de 100 rodadas tomando-se aleatoriamente diferentes grupos de amostras para calibração e validação, obtiveram-se os valores médios para o RMSEP mínimo e para o número de componentes necessários à obtenção do erro mínimo os valores de 6,62% e 46,39 componentes, respectivamente. Tais valores são superiores àqueles fornecidos pelos modelos ajustados diretamente sobre os valores da reflectância e também àqueles ajustados sobre a primeira derivada dos espectros.

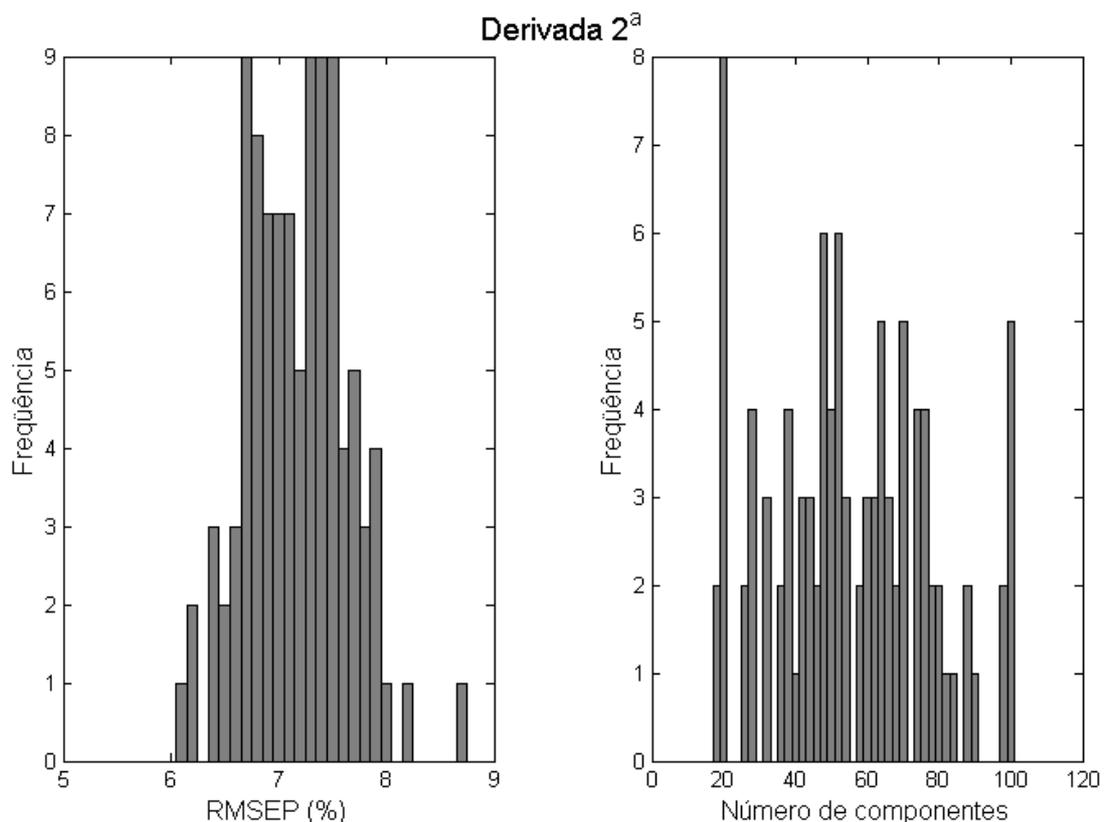


Figura 73 - Distribuição de freqüências para os erros mínimos na validação (RMSEP) e para o número de componentes principais necessários à sua obtenção quando utilizada a segunda derivada dos espectros na predição da densidade básica.

A Figura 73 apresenta a distribuição dos erros mínimos na validação e no número de componentes principais necessários à sua obtenção em cada uma das 100 rodadas. Note-se a instabilidade no número de componentes correspondentes aos erros mínimos expressada pela baixa curtose apresentada pelo seu histograma.

5.2.3.4.2. A absorbância

A transformação dos espectros de reflectância realizada pelo cálculo de seu cologaritmo, em espectros de absorbância, resulta em curvas que apresentam um aspecto complementar à reflectância. A ausência de picos isolados na curva de correlações entre os valores da absorbância em cada comprimento de onda e a variável dependente é um indicativo de que pouca informação irrelevante foi computada (Figura 74).

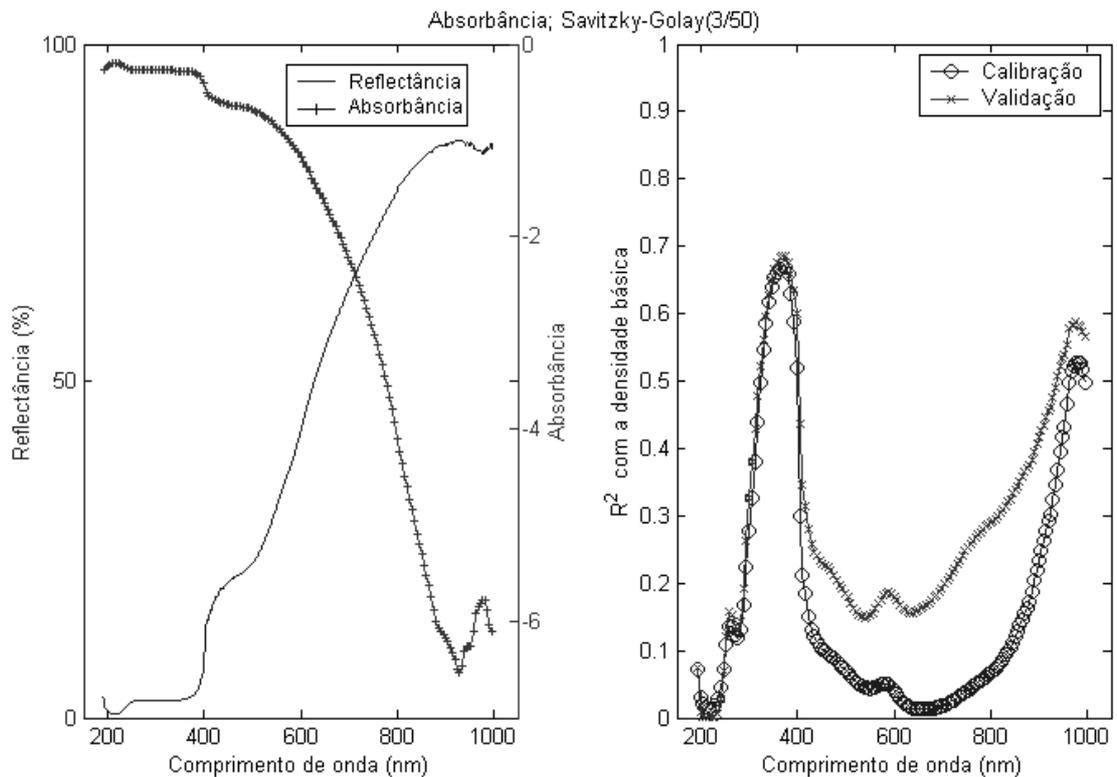


Figura 74 - Os espectros de reflectância e absorbância de uma amostra (*esq.*). Correlações entre os valores da absorbância e a densidade básica (*dir.*).

De fato, a transformação dos espectros de reflectância em absorbância tem o objetivo de compensar erros originados pela maneira peculiar com que cada amostra realiza a dispersão da radiação sobre ela incidente, fazendo com que parte da energia

seja dispersa em direções que aumentam ou reduzem as chances de ser capaz de retornar ao sensor na forma de radiação difusa. Assim, tal fração pode ser absorvida da maneira prevista pela teoria quântica, com a transição de elétrons para orbitais diferentes dos originais ou ser refletida de volta ao meio externo. A dispersão depende da geometria e das diferenças entre os índices de refração dos constituintes sob análise e da matriz onde estão inclusos. No caso da espectroscopia aplicada à análise química de materiais pulverizados (pós), é marcante a influência das diferenças na distribuição das dimensões e de variações na geometria de suas partículas. No caso de emulsões, suspensões e soluções, esta mesma afirmação é válida para as gotículas, partículas ou moléculas de cada amostra (NÆS et al. 2002).

Na espectroscopia de reflectância aplicada à superfície da madeira sólida, é possível que as diferenças na dispersão da radiação estejam relacionadas a variações anatômicas e as conseqüentes variações químicas e físicas que acarretam. Assim, sendo cada amostra de madeira única em sua estrutura anatômica, é possível que haja realmente a necessidade de se realizar a correção da dispersão dos dados espectroscópicos se o objetivo da análise for a quantificação de um determinado componente químico. Por outro lado, em uma investigação acerca das relações entre a reflectância e propriedades físicas do material, talvez a contribuição da dispersão seja relevante para o sucesso do ajuste de modelos de calibração.

A PCA dos espectros de absorvância revelou uma grande concentração da variância nos dois primeiros componentes principais, o que não havia sido observado para nenhuma das transformações apresentadas anteriormente. A consistência apresentada pelas correlações entre os primeiros dez componentes principais e a densidade básica, quando tomados ao acaso diferentes grupos de amostras, foi também um efeito notável desta transformação. Tais efeitos podem ser visualizados na Figura 75.

Outros importantes efeitos da utilização da absorvância nos modelos de calibração para a previsão da densidade básica da madeira foram a redução e a estabilização do número de componentes principais necessários à obtenção dos erros mínimos. Tais efeitos, provavelmente, estão também relacionados à correção da dispersão espectroscópica. Assim, em uma centena de rodadas, tomando-se ao acaso amostras para calibração e validação, o RMSEP mínimo apresentou uma média de 5,18%, enquanto que o número médio de componentes para a sua obtenção concentrou-se por volta de 22. As distribuições de frequências para os valores de RMSEP mínimos e seus respectivos números e componentes podem ser observados na Figura 76.

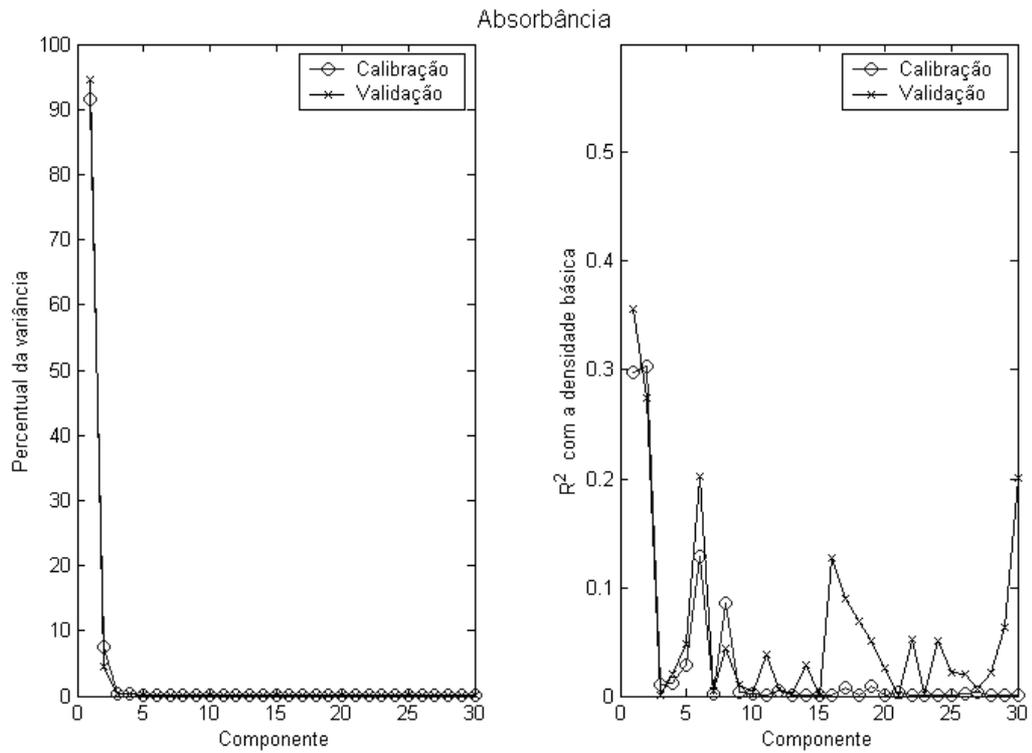


Figura 75 - Percentual da variância global carregado por cada um dos 20 primeiros componentes principais dos espectros de absorvância (*esquerda*) e suas correlações individuais com a densidade básica da madeira.

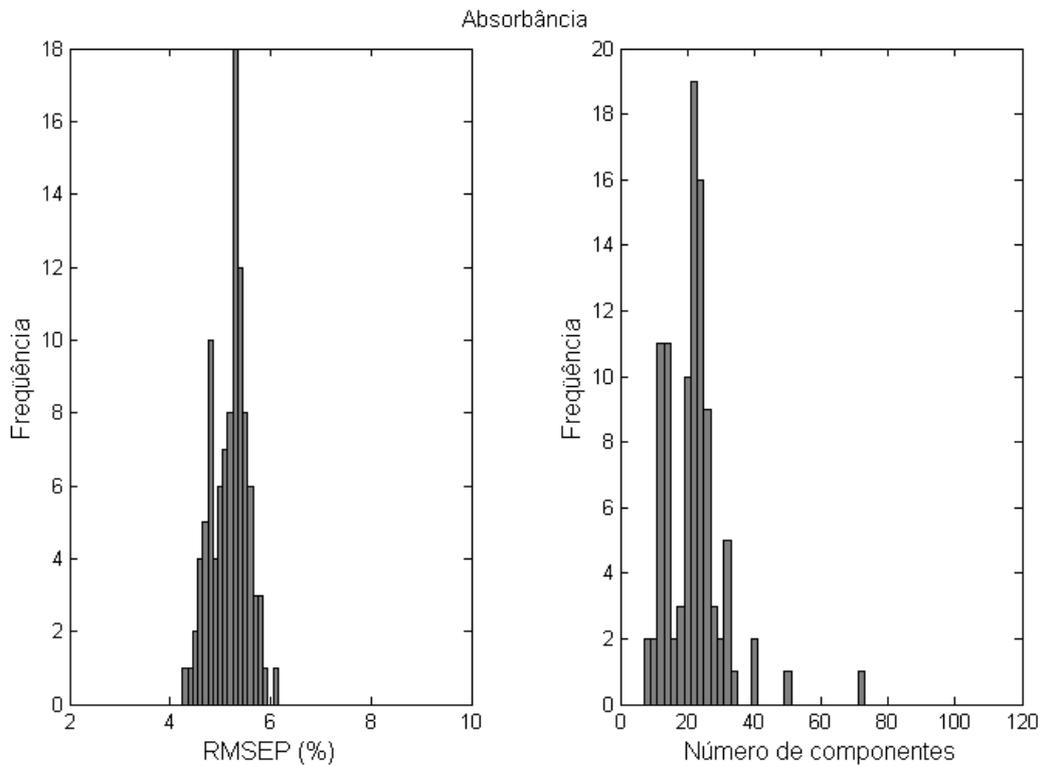


Figura 76 - Distribuição de frequências para os erros mínimos na validação (RMSEP) e para o número de componentes principais necessários à sua obtenção quando utilizados os espectros de absorvância na predição da densidade básica.

Na Figura 77 são apresentadas as plotagens dos valores observados da densidade básica contra aqueles previstos por modelos que forneceram o valor maior, o menor e o mais próximo da média dos RMSEP's mínimos em 100 rodadas tomando-se aleatoriamente amostras para a calibração e validação dos modelos ajustados sobre os valores da absorvância.

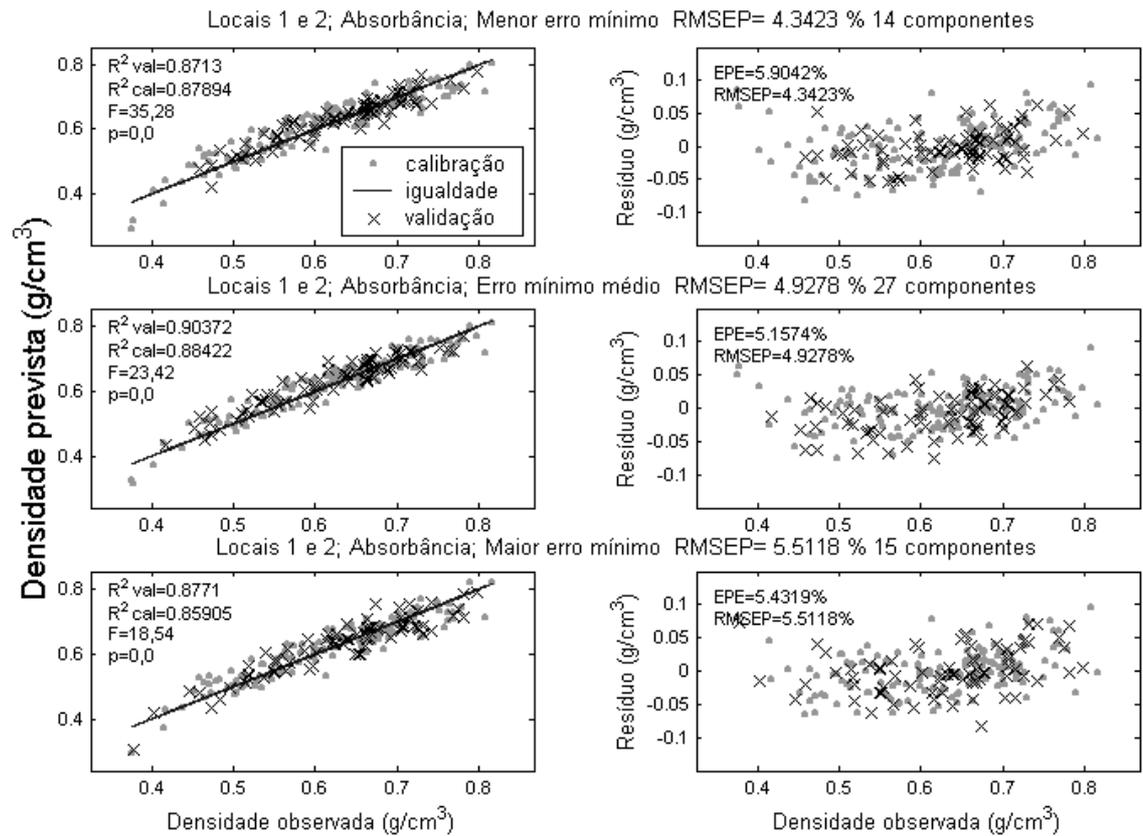


Figura 77 - Plotagem dos valores observados versus previstos e os resíduos de modelos ajustados sobre os espectros de absorvância, que apresentaram RMSEP maior, menor e mais próximo da média dos erros mínimos na validação para a previsão da densidade básica. São apresentados também os valores de R^2 na calibração e validação, além do EPE, do RMSEP, o número de componentes principais utilizados e as estatísticas F e p relativas às regressões realizadas para obtenção dos modelos.

O cômputo das derivadas primeira e segunda dos espectros de absorvância não trouxeram ganhos à qualidade dos modelos, como será visto mais adiante.

5.2.3.4.3. A função de Kubelka-Munk

A transformação dos espectros de reflectância na função de Kubelka-Munk origina curvas de aspecto um tanto diverso e complementar aos espectros, já que são uma expressão da absorbância com a correção da dispersão (Figura 78). De maneira similar ao que foi observado para a absorbância, a curva de correlações entre os valores da função de Kubelka-Munk e a variável dependente apresentou-se de forma suave, sem picos isolados que podem significar a presença de interferências espúrias.

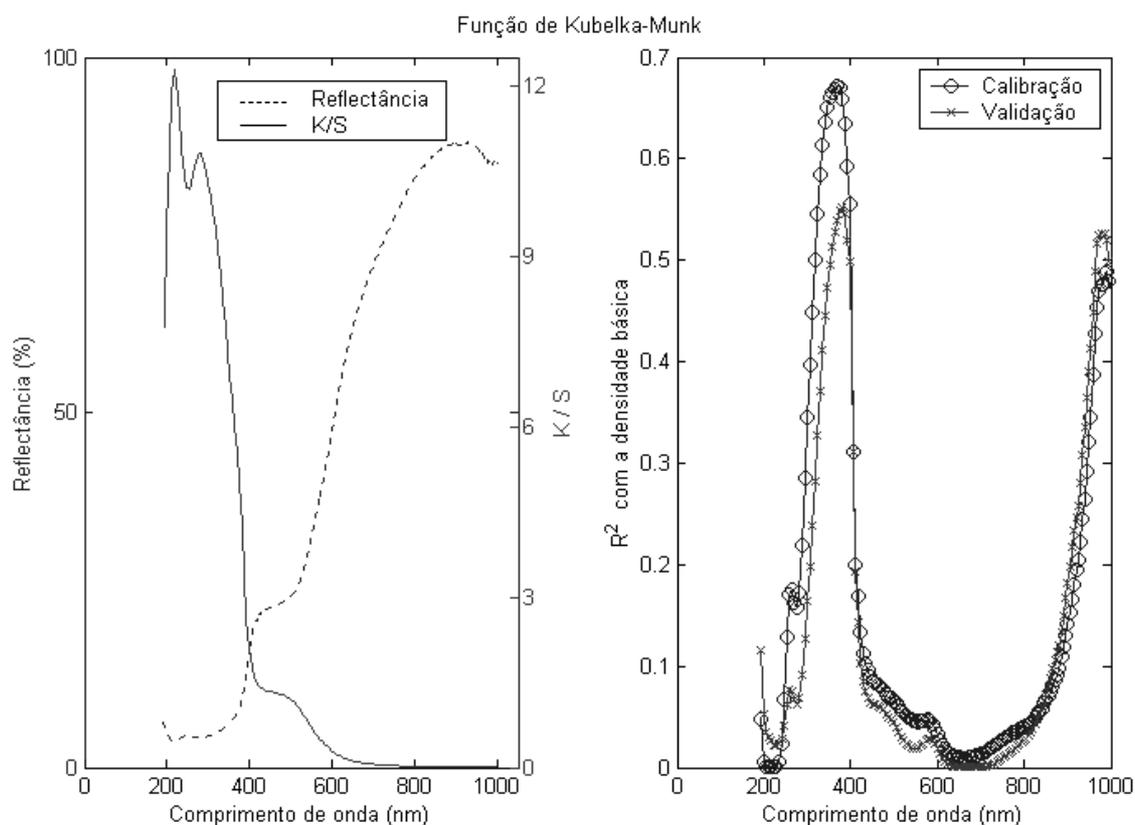


Figura 78 - Os espectros de reflectância de uma amostra e sua transformada na função de Kubelka-Munk (*esq.*). Correlações entre os valores da absorbância e a densidade básica (*dir.*).

Como observado para a absorbância, outro efeito interessante da função de Kubelka-Munk foi a concentração da maior parte da variância nos primeiros componentes principais. Também aqui foi notável a coerência com que os primeiros componentes extraídos a partir das transformadas dos valores de diferentes grupos de amostras (validação e calibração) mostraram-se coerentes no percentual da variância

que carregam e em suas correlações com a variável dependente. Estes efeitos são visualizados na Figura 79.

A aplicação da função de Kubelka-Munk, entretanto, manteve as variáveis independentes com alto grau de correlação, o que sugere a PCA em seguida a essa transformação para a obtenção de modelos de calibração estáveis.

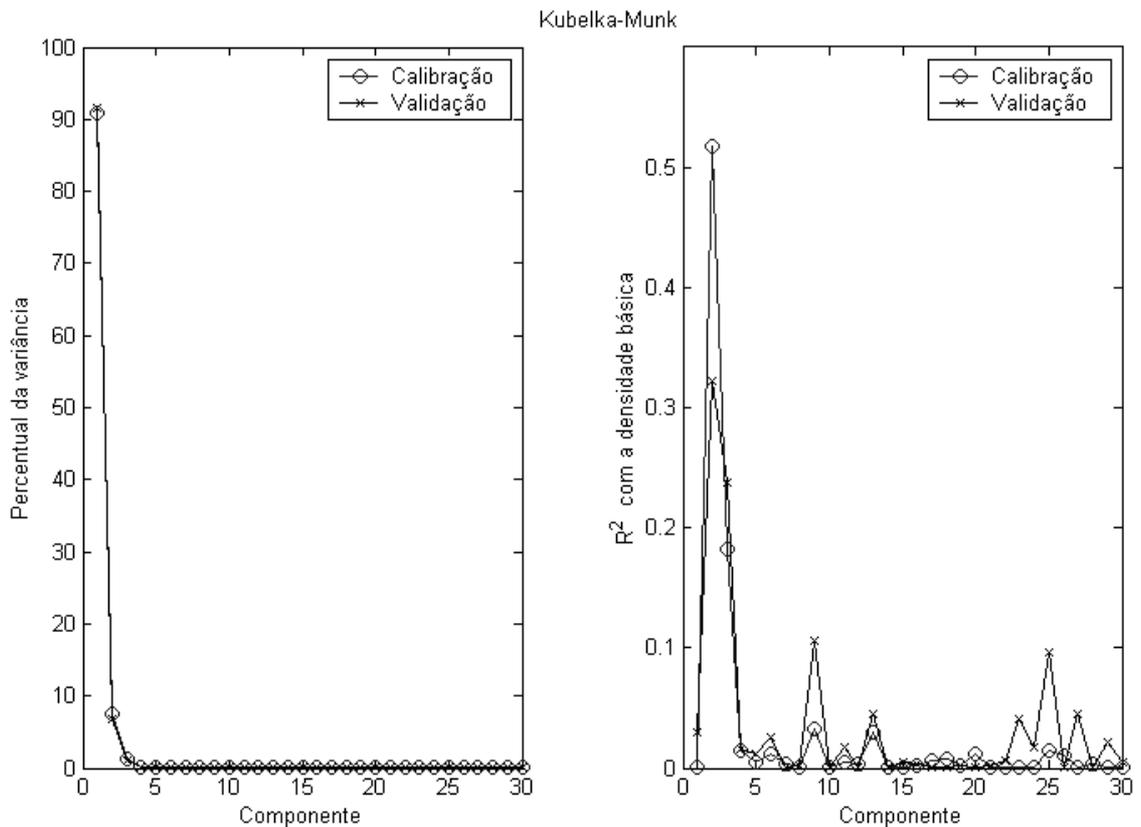


Figura 79 - Percentual da variância global carregado por cada um dos 30 primeiros componentes principais da função de Kubelka-Munk aplicada aos espectros de reflectância (*esquerda*) e suas correlações individuais com a densidade básica da madeira.

Um efeito negativo da transformação dos espectros de reflectância para a função de Kubelka-Munk pôde ser notado na distribuição dos números de componentes principais necessários à obtenção do erro mínimo em cada uma das 100 rodadas aqui testadas. Note-se na Figura 80 o histograma bimodal apresentado para esta variável. Na mesma figura, porém, é possível perceber-se a distribuição quase normal dos RMSEP's mínimos obtidos. Os valores médios para o RMSEP mínimo e para o número de componentes correspondentes situaram-se em torno de 6,08% e 44 componentes respectivamente.

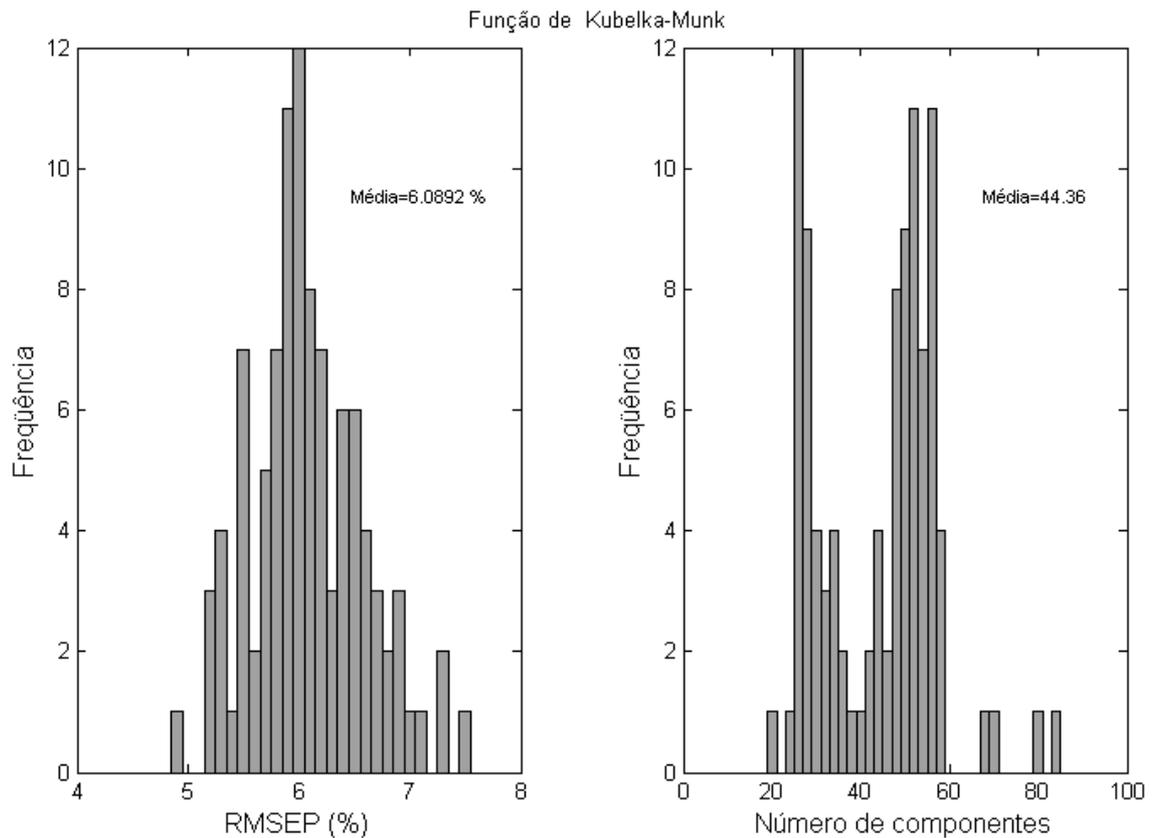


Figura 80 - Distribuição de frequências para os erros mínimos na validação (RMSEP) e para o número de componentes principais necessários à sua obtenção quando utilizados os valores da função de Kubelka-Munk aplicada sobre os espectros de reflectância na predição da densidade básica.

Conclui-se que a função de Kubelka-Munk não apresentou vantagens significativas para a qualidade dos modelos de calibração para a predição da densidade básica da madeira.

5.2.3.4.4. Transformações para normalização dos dados espectroscópicos

As transformações realizadas no intuito de se normalizar os dados fornecido pelo espectrômetro foram a extração da raiz quadrada da reflectância, a divisão da reflectância pela sua variância e a função de Box-Cox.

Utilizando-se o teste de Lilliefors foi possível constatar que os valores da reflectância em apenas 45% dos comprimentos de onda apresentam distribuição normal. Note-se na Figura 81 que isto ocorre em largas faixas do espectro. Além disso, é possível perceber na figura que as faixas que apresentam distribuição normal são praticamente contínuas, ou seja, parece existir um padrão para o fenômeno que pode estar relacionado a algum aspecto físico da madeira. Ao realizar-se a extração da raiz

quadrada da reflectância, houve um ligeiro incremento no número de comprimentos de onda que passaram no teste de normalidade (46%); esta transformação, porém, conferiu uma certa alternância entre as faixas do espectro que apresentaram distribuição normal e aquelas que não apresentaram.

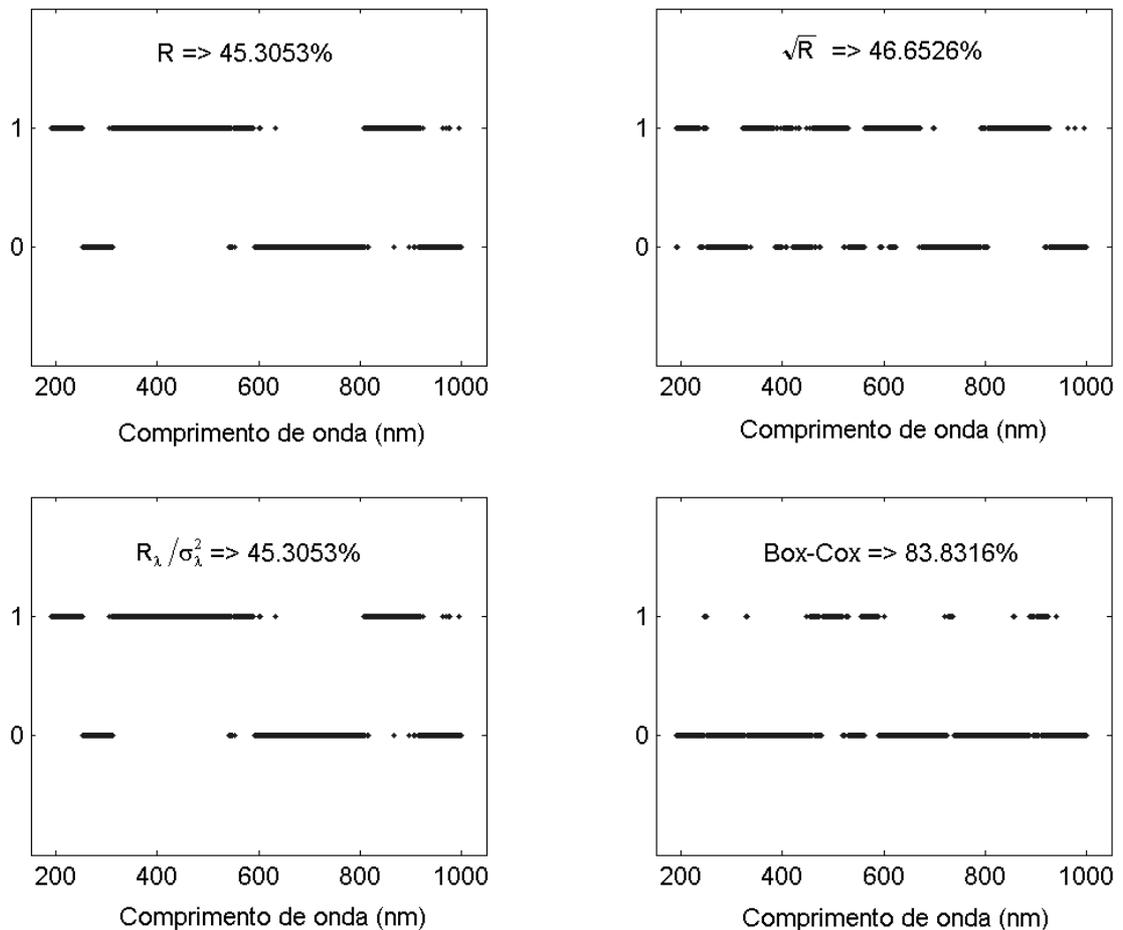


Figura 81 - Resultados do teste de Lilliefors para a normalidade da reflectância antes e após algumas transformações matemáticas aplicadas sobre seus valores em cada comprimento de onda. O valor 0 significa que, naquele comprimento de onda, a variável apresenta distribuição normal; o valor 1 indica não-normalidade. R = reflectância.

A divisão pela variância da reflectância não causou nenhum efeito sobre o percentual dos comprimentos de onda que apresentavam distribuição normal, nem tampouco, houve intercâmbio dessa característica entre as variáveis. Ou seja, nenhum efeito foi observado sobre a normalidade dos dados ao se dividir pela variância. A transformação pela função de Box-Cox, por sua vez, provocou um significativo aumento no número de variáveis que passaram no teste de Lilliefors (83%), entretanto,

algumas variáveis que apresentavam normalidade o deixaram de apresentar. De qualquer forma, no que diz respeito à eficiência na normalização dos dados, a transformação pela função de Box-Cox foi a que apresentou os melhores resultados. A Figura 82 apresenta os histogramas para a reflectância em 375 nm antes e após a realização das transformações.

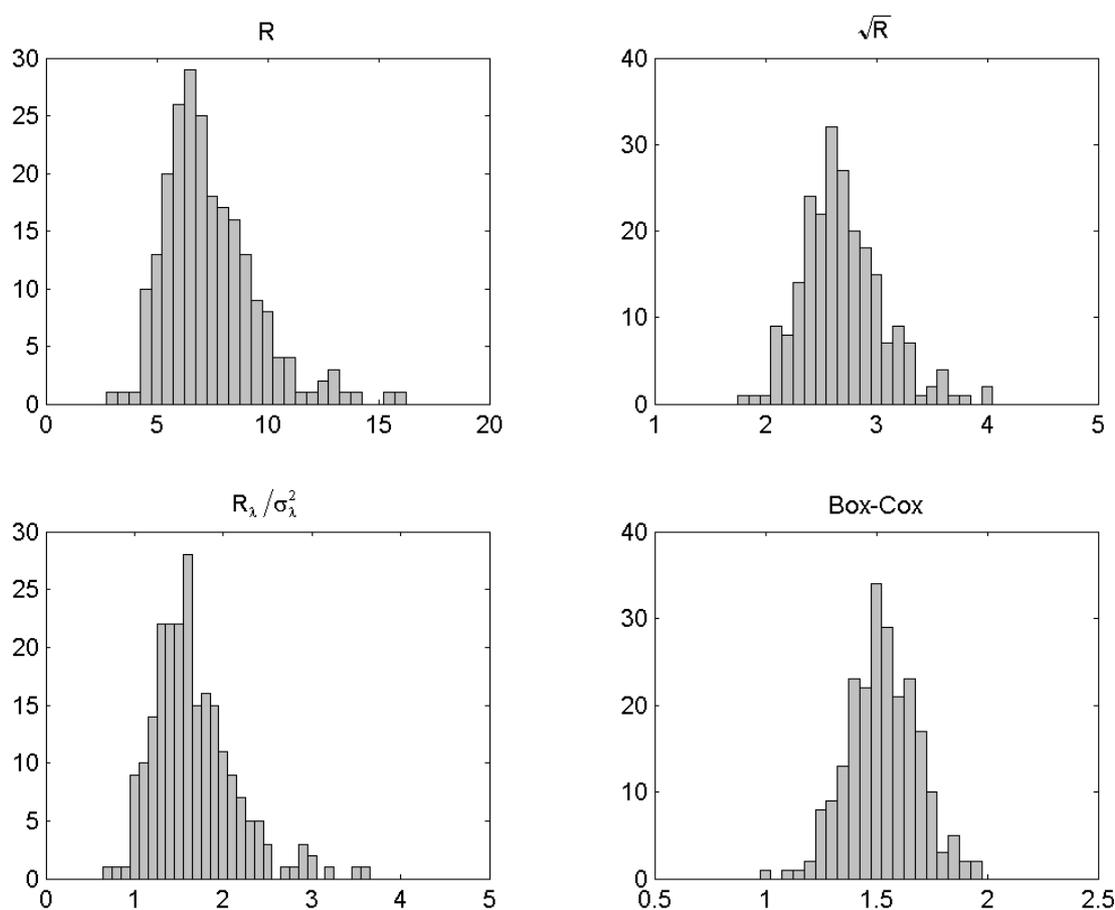


Figura 82 - Distribuição de frequências para a reflectância em 375 nm e para os valores desta variável após a aplicação das transformações: extração da raiz quadrada, divisão pela variância e função de Box-Cox.

A divisão pela variância e a extração da raiz quadrada, bem como o cálculo da primeira e segunda derivadas subsequentemente à essas transformações, levaram a resultados semelhantes no que diz respeito às correlações entre as variáveis independentes e a densidade básica, como pode ser observado nas Figuras 83 e 84. Tais correlações não se apresentaram significativamente distintas daquelas observadas para a absorbância, para a função de Kubelka-Munk ou para a reflectância e suas derivadas. Para os valores transformados e para a primeira derivada desses valores, permaneceram

com os mais altos coeficientes de correlação, as variáveis associadas aos comprimentos de onda do ultravioleta, com picos de base larga por volta de 375 nm.

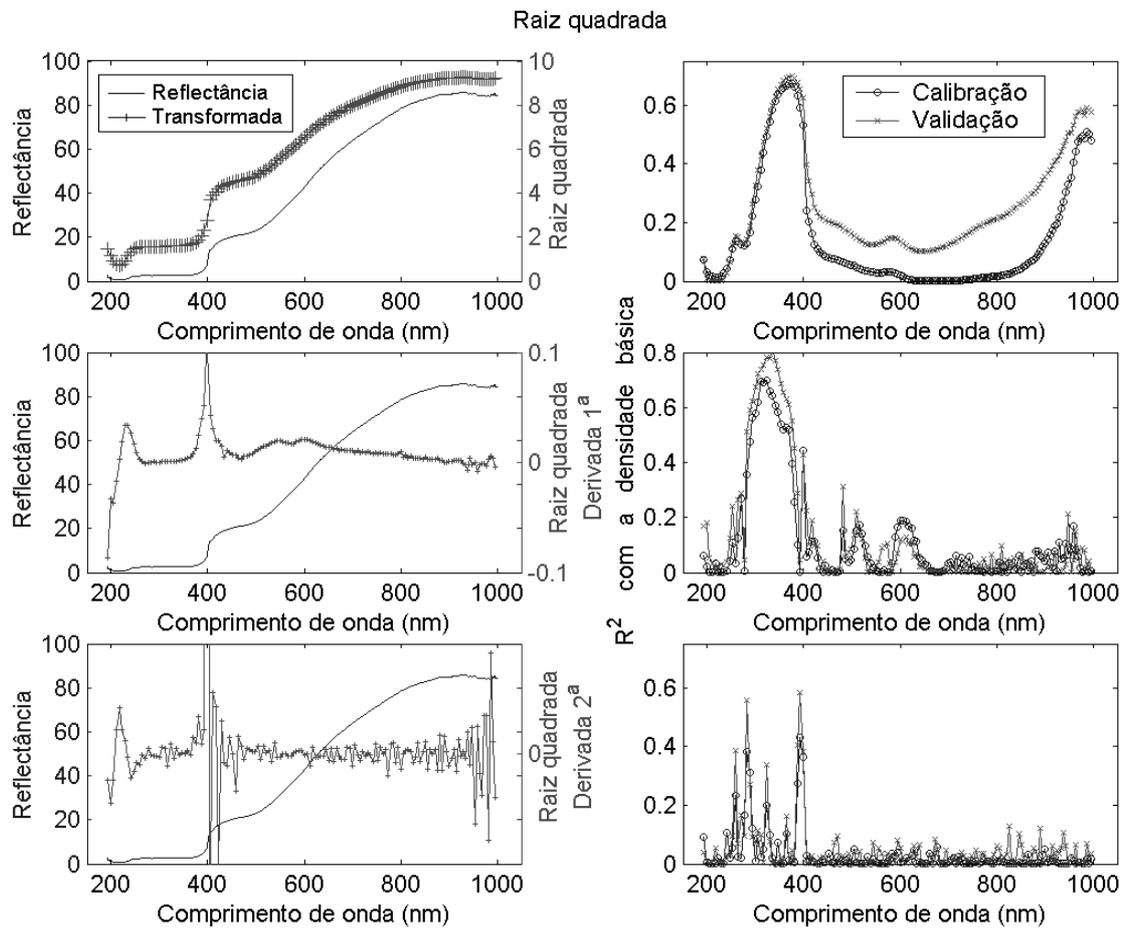


Figura 83 - Raiz quadrada da reflectância, suas derivadas primeira e segunda e suas correlações com a densidade básica.

A divisão pela variância levou a uma drástica queda nos coeficientes de correlação na região de transição para o infravermelho, o que não foi observado para a transformação pela extração da raiz quadrada da reflectância.

As segundas derivadas das transformações pela extração da raiz-quadrada e pela divisão pela variância conduziram a situações em que as maiores correlações também permaneceram no ultravioleta, porém com picos isolados.

A função de Box-Cox, por sua vez levou a expressivas modificações nas correlações entre as variáveis dependentes e a densidade básica. Os coeficientes de correlação foram reduzidos em toda a extensão do espectro, tornando-se constantes em toda a faixa do visível até a transição para o infravermelho (Figura 85).

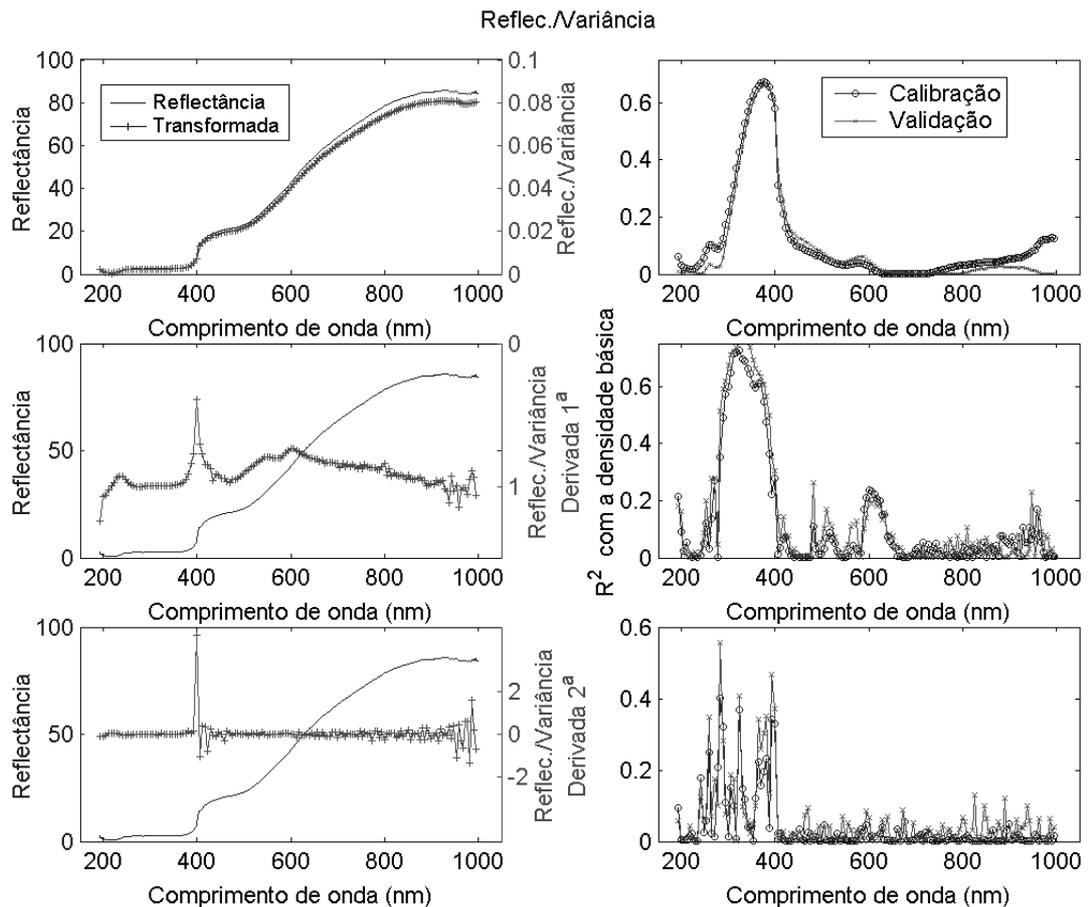


Figura 84 - Curva resultante da divisão da reflectância pela sua variância em cada comprimento de onda, suas derivadas primeira e segunda e suas correlações com a densidade básica.

Já a primeira derivada da função de Box-Cox provocou uma elevação generalizada nos valores das correlações entre a densidade básica e as variáveis correspondentes os comprimentos de onda da faixa do visível.

A segunda derivada da função de Box-Cox, por sua vez, provocou uma drástica redução dos coeficientes de correlação com a densidade básica em toda a faixa de trabalho.

No que diz respeito aos componentes principais, a divisão pela variância e a extração da raiz quadrada da reflectância, bem como suas derivadas, provocaram a distribuição da variância global entre os primeiros componentes, havendo casos em que o primeiro componente principal carrega apenas 30% da variância. Um outro efeito notável da função de Box-Cox, porém, foi a concentração de praticamente 100% da variância no primeiro componente principal, tanto para os valores simplesmente transformados, quanto para as suas derivadas primeira e segunda, como pode ser visto na Figura 86.

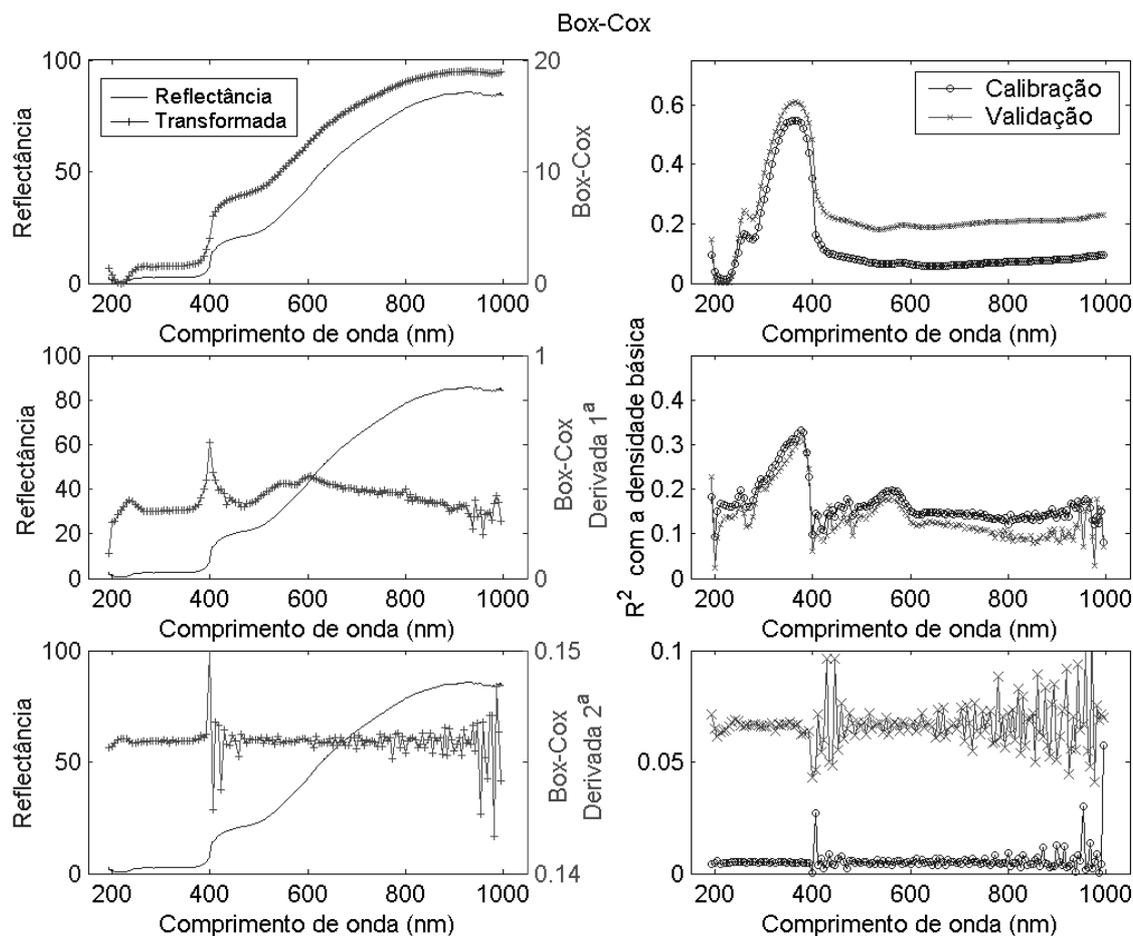


Figura 85 - Curva resultante da aplicação da função de Box-Cox sobre os valores da reflectância, suas derivadas primeira e segunda e suas correlações com a densidade básica.

A segunda derivada da função de Box-Cox provocou quedas generalizadas nas correlações entre os valores de seus escores e a densidade básica.

Estão comparados na Figura 87 os efeitos das transformações normalizadoras, bem como daquelas que objetivaram a correção da dispersão dos dados espectroscópicos (absorbância e função de Kubelka-Munk) sobre a capacidade de predição dos modelos ajustados para a densidade básica como variável dependente. Estes resultados foram obtidos a partir de 500 rodadas para cada tipo de transformação, dentro das quais foram registrados os coeficientes de determinação e erros (EPE e RMSEP) na calibração e validação para os modelos que forneceram o menor RMSEP mínimo. Também foram computadas as variâncias e curtoses para os erros mínimos além do número de componentes principais necessários à sua obtenção.

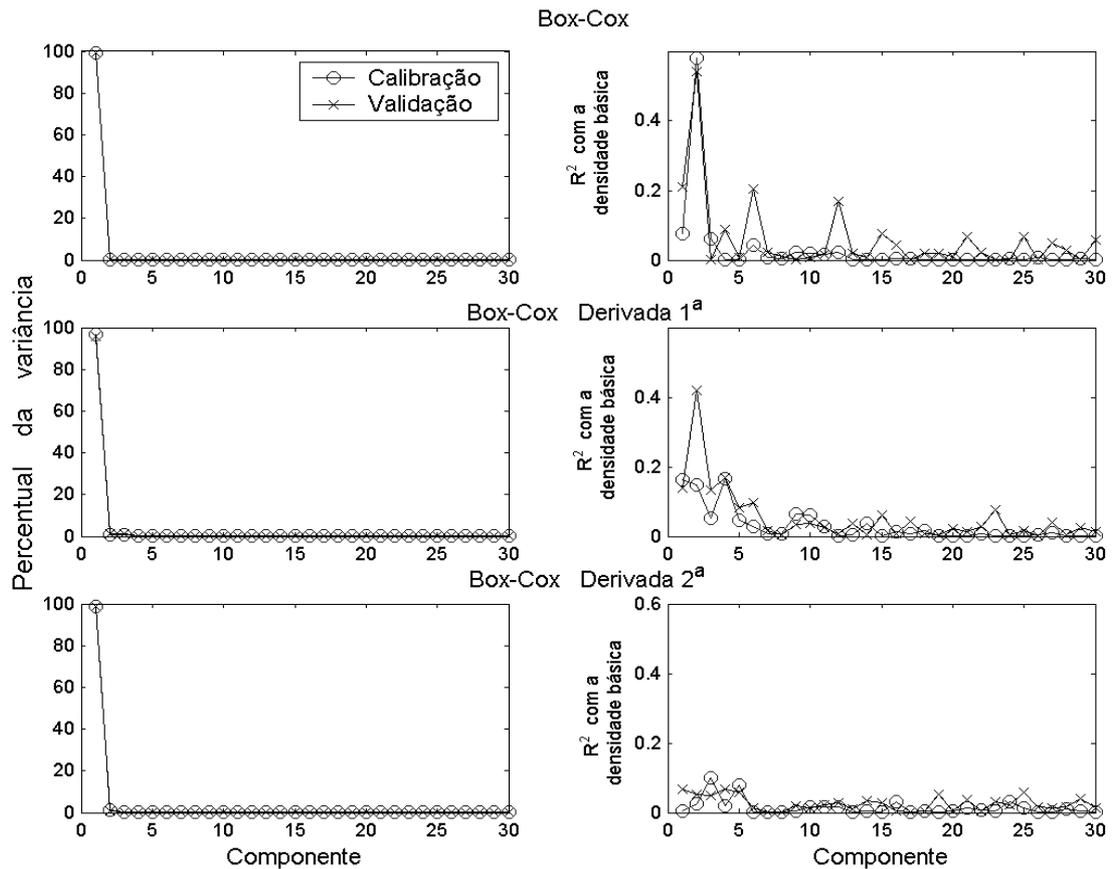
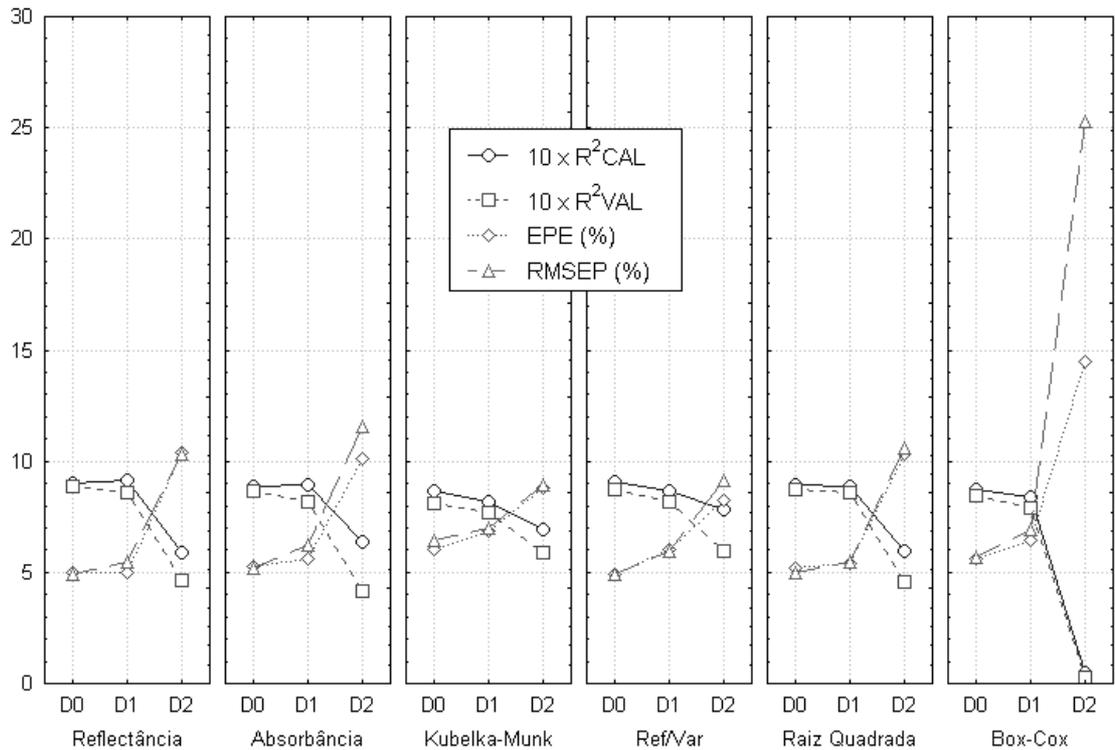


Figura 86 - Percentual da variância global carregado por cada um dos 30 primeiros componentes principais da função de Box-Cox aplicada aos espectros de reflectância, derivadas primeira e segunda (*esquerda*). Correlações individuais entre cada componente e a densidade básica da madeira (*direita*).

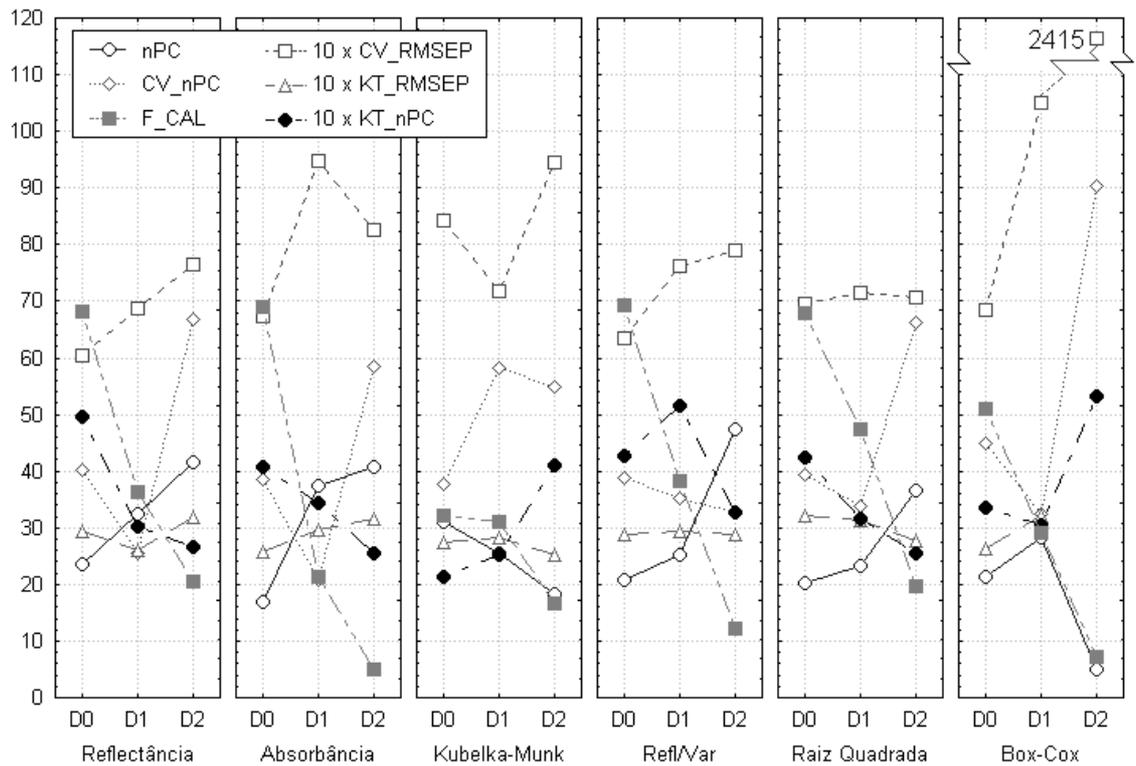
É notável na Figura 87 que, para todas as transformações, a utilização dos valores da derivada primeira como variáveis independentes levou a ligeiras quedas nos coeficientes de determinação, com correspondentes incrementos nos erros. O mesmo foi observado para a derivada segunda, porém, de forma mais acentuada. Tal efeito torna-se ainda mais pronunciado ao derivarem-se os valores da função de Box-Cox; por outro lado, manifesta-se mais discretamente na função de Kubelka-Munk. Sabe-se que a derivação possui efeitos positivos sobre a correção da dispersão espectroscópica, já que elimina as diferenças na linha de base provocadas por peculiaridades estruturais de cada amostra. A função de Kubelka-Munk, porém, já realiza tal correção de forma bastante eficiente, o que poderia explicar o menor efeito das derivadas sobre seus valores na capacidade de predição dos modelos.



D0: sem derivar; D1: derivada primeira; D2: derivada segunda.

Figura 87 - Coeficientes de determinação, erros na calibração (R^2_{CAL} e EPE) e validação (R^2_{VAL} e RMSEP) apresentados pelos modelos ajustados para previsão da densidade básica da madeira que obtiveram RMSEP's mínimos em 500 rodadas para cada tipo de transformação.

De maneira geral, o uso da reflectância sem nenhuma transformação, ainda parece ser a opção mais interessante na calibração para a densidade básica, pois com um mínimo de processamento, são obtidos modelos que superam ou se equiparam àqueles que apresentaram os melhores resultados. Assim, considerando apenas os coeficientes de correlação e os erros poder-se-ia optar pela adoção da reflectância. Estes, porém, não são os únicos critérios para realizar-se a escolha do mais adequado pré-processamento dos dados. Há que se considerar, ainda, o número de componentes envolvidos e a estabilidade dos resultados quando variam as amostras tomadas para calibração e validação. Tais considerações podem ser amparadas pela análise de parâmetros como os coeficientes de variação para o número de componentes principais, para os erros e para a estatística F, além de características mais específicas de suas distribuições entre as diversas rodadas, como a curtose de seus histogramas. A Figura 88 apresenta o efeito das transformações sobre alguns desses parâmetros.



O eixo das ordenadas possui várias unidades.

Figura 88 - Efeitos das transformações sobre a curtose (KT) e coeficientes de variação (CV) para número de componentes principais (nPC) e valores do RMSEP mínimo em modelos ajustados em 500 rodadas. O valor de F para os modelos também é apresentado.

De maneira geral, o valor da estatística F decresce com as derivações subseqüentes. Tal fato pode ser explicado pelo típico realce dos ruídos que representam informações irrelevantes para a modelagem da variável dependente. Os demais parâmetros não se comportam de maneira tão regular. Uma maneira de se tornar mais clara a interpretação dos parâmetros de qualidade dos modelos é a sua aglutinação em um único índice, cujo valor é diretamente proporcional aos coeficientes de determinação (R^2_{CAL} e R^2_{VAL}), à curtose das distribuições do número de componentes utilizados nos modelos (KU_{nPC}) e dos erros mínimos obtidos na validação (KU_{RMSEP}) e ao valor de F e inversamente proporcional ao número de componentes e erros de calibração e validação e de seus coeficientes de variação. Tal valor é aqui chamado Índice de Qualidade (IQ), e foi obtido através da expressão:

$$IQ = \frac{R^2_{CAL} \times R^2_{VAL} \times KU_{nPC} \times KU_{RMSEP} \times F}{nPC \times EPE \times RMSEP \times CV_{nPC} \times CV_{RMSEP}}$$

A Figura 90 apresenta os índices de qualidade calculados a partir dos resultados apresentados nas Figuras 87 e 88.

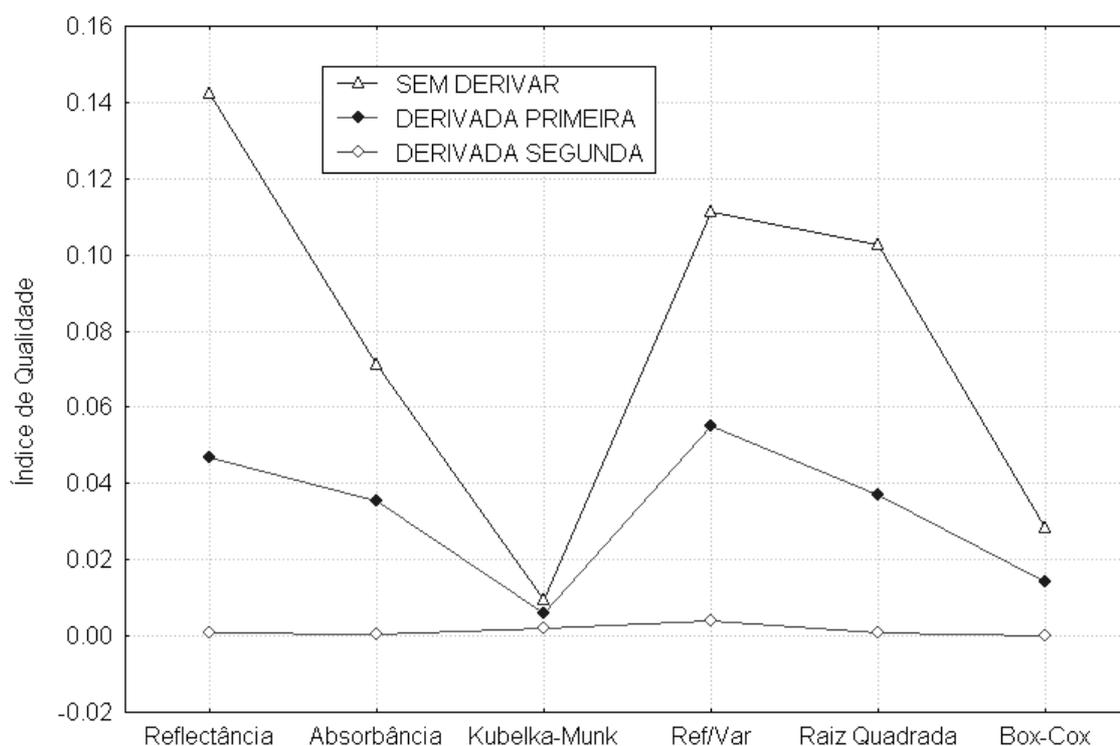


Figura 90 - Plotagem dos valores de índice de qualidade para os modelos ajustados por PCR sobre a reflectância e suas transformações.

Para a reflectância, o número médio de PC's necessários para a obtenção do RMSEP mínimo aumenta linearmente ao realizarem-se as primeiras e segundas derivações, enquanto que seu coeficiente de variação encontra um valor mínimo na primeira derivada e sua curtose parece aí também se estabilizar. Considerando-se ainda que o valor do RMSEP mínimo e de seu coeficiente de variação ainda sofrem ligeiros prejuízos, e que, apesar de se manter em um nível alto, cai também o valor de F, pode-se concluir que a reflectância é a melhor opção dentre todas as outras transformações para o desenvolvimento de modelos de calibração utilizando-se a regressão por componentes principais adicionados pela ordem decrescente de variância para a predição da densidade básica. Tal conclusão pode ser extraída diretamente da Figura 90.

5.2.3.5. A regressão por PCR *forward*

A inclusão dos componentes principais pela ordem decrescente de variância pode resultar em modelos com um número muito grande de variáveis independentes

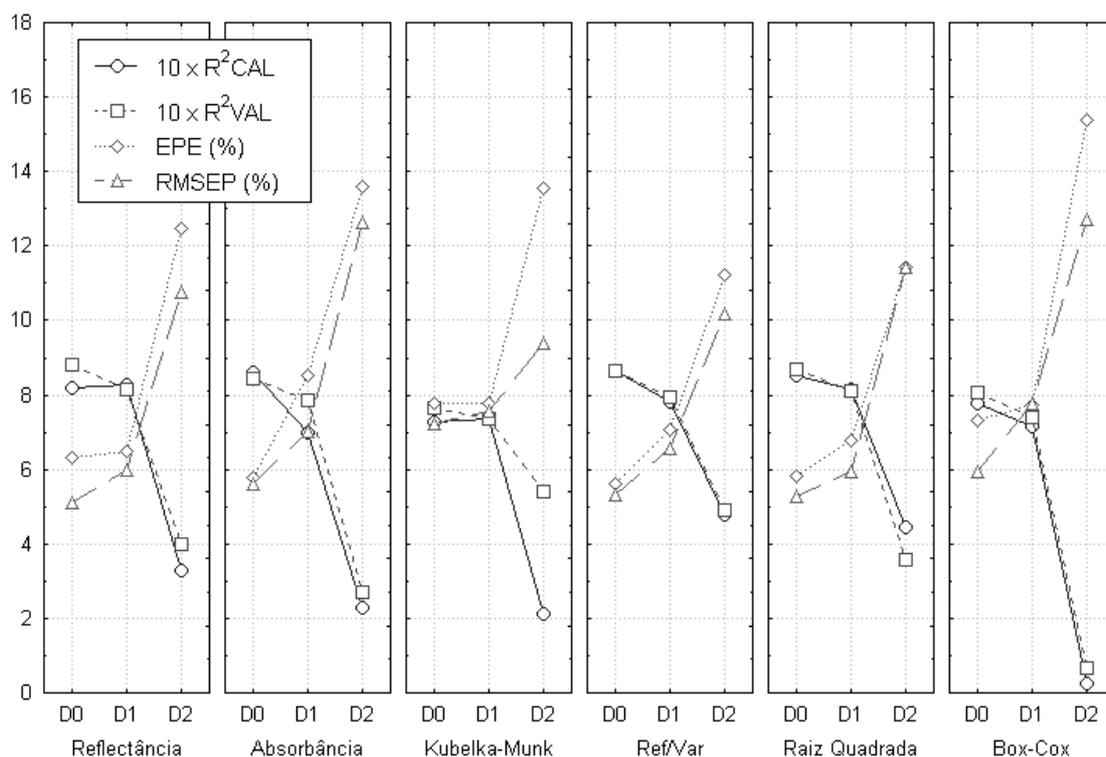
para se obter o RMSEP mínimo e, para as quais, nem sempre são significativos os valores dos coeficientes de regressão. Tais modelos podem conter muita informação irrelevante, o que reduz a probabilidade da rejeição da hipótese de nulidade dos parâmetros da regressão, seja para a totalidade dos parâmetros, seja para cada parâmetro individualmente.

Uma maneira de se contornar este problema é a aplicação da regressão por componentes principais *forward* (PCR_FWD), em que a primeira variável independente a ser adicionada ao modelo é aquela que guarda a maior correlação com a variável dependente. Em seguida, são calculados os resíduos deste modelo e identificada a segunda variável independente que guarda maior correlação com estes resíduos. Esta será a próxima variável a ser adicionada e assim, sucessivamente, até que se atinja um número de variáveis independentes igual ao número de amostras menos a unidade. Uma vez realizada esta série de ajustes, identifica-se qual é o número de variáveis independentes que fornece o RMSEP mínimo, de maneira similar ao procedimento aplicado no método de ajuste pela inclusão das variáveis (PC's) pela ordem decrescente de variância (PCR_ODV).

Assim, em 500 rodadas escolhendo ao acaso grupos de amostras para calibração e validação, foram ajustados modelos por *forward* para reflectância e cada uma das suas transformações referidas anteriormente.

Nestas séries de ajustes foi realmente possível observar-se uma redução generalizada no número de componentes necessários à obtenção do RMSEP mínimo, que, por sua vez, teve seus valores ligeiramente ampliados se comparados aos resultantes da validação dos modelos ajustados por PCR_ODV. Foram recorrentes os prejuízos causados aos modelos pelas derivações, seja da reflectância, seja de suas transformações.

Na Figura 91 é possível notar-se que os efeitos das transformações das variáveis antes da realização da PCR_FWD levaram a resultados bastante semelhantes àqueles observados com a PCR_ODV, apesar desta última ter sofrido variações de menor amplitude nos coeficientes de correlação e erros na calibração e validação provocados pelas derivações subsequentes. De qualquer forma, em ambos os métodos de inclusão de variáveis, fica descartado o uso da derivada segunda da reflectância e suas transformações, já que esta operação provocou inaceitáveis reduções nos coeficientes de correlação, com conseqüentes ampliações nos erros.



D0: sem derivar; D1: derivada primeira; D2: derivada segunda.

Figura 91 - Coeficientes de determinação, erros na calibração (R^2_{CAL} e EPE) e validação (R^2_{VAL} e RMSEP) apresentados pelos modelos ajustados por PCR-FWD para previsão da densidade básica da madeira que obtiveram RMSEP's mínimos em 500 rodadas para cada tipo de transformação.

Menos sensíveis às derivações revelaram-se os números de componentes principais necessários à obtenção dos RMSEP's mínimos e as curtoses de suas distribuições entre as rodadas, à exceção da função de Box-Cox, como pode ser observado na Figura 92.

A análise dos demais parâmetros de qualidade dos modelos que estão apresentados na Figura 92 torna-se novamente confusa, dada a ausência de tendências ou confluências entre os valores. Para prosseguir na análise, lançou-se mão do cômputo dos índices de qualidade dos modelos, cujos valores estão plotados na Figura 93, onde se torna clara a superioridade dos modelos ajustados sem o uso das derivadas no pré-processamento. Uma exceção a esta observação foi detectada para a função de Kubelka-Munk, na qual o valor do índice de qualidade para os modelos sem derivação, praticamente coincide com aquele fornecido por aquele ajustado sobre a primeira derivada. Tal fato não traria maiores conseqüências, já que foram ambos os grupos de modelos classificados entre os que seriam descartados.

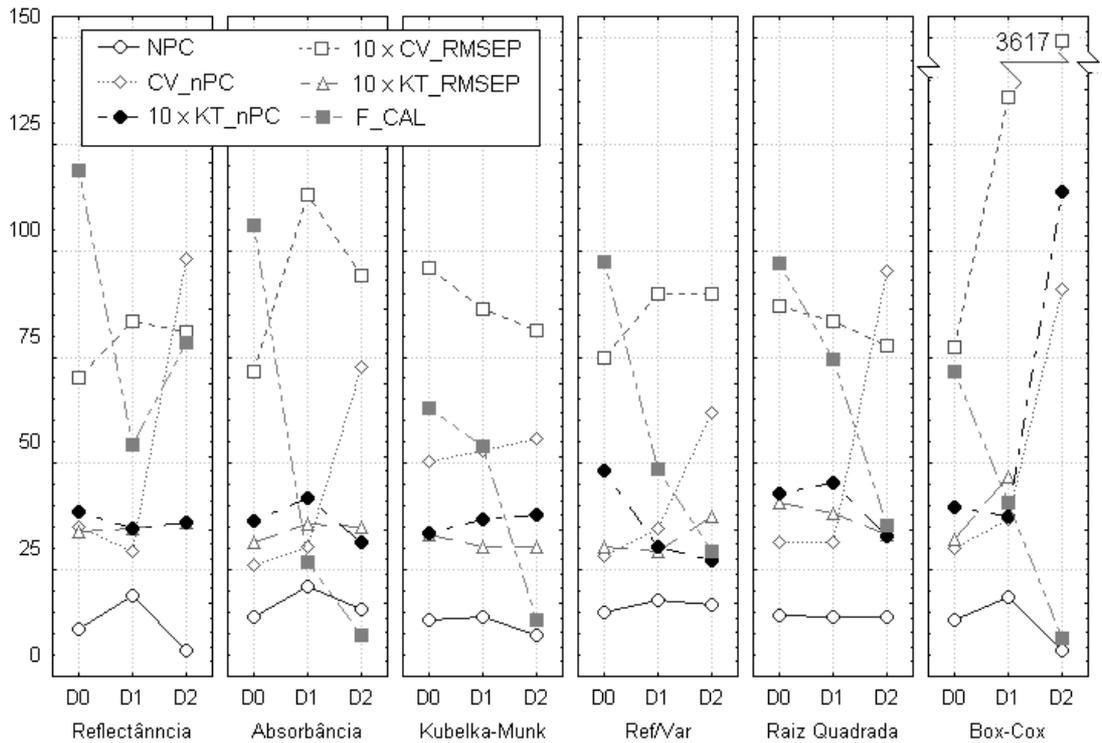


Figura 92 - Efeitos das transformações sobre a curtose (KT) e coeficientes de variação (CV) para número de componentes principais (nPC) e valores do RMSEP mínimo em modelos ajustados por PCR-FWD em 500 rodadas. O valor de F para os modelos também é apresentado.

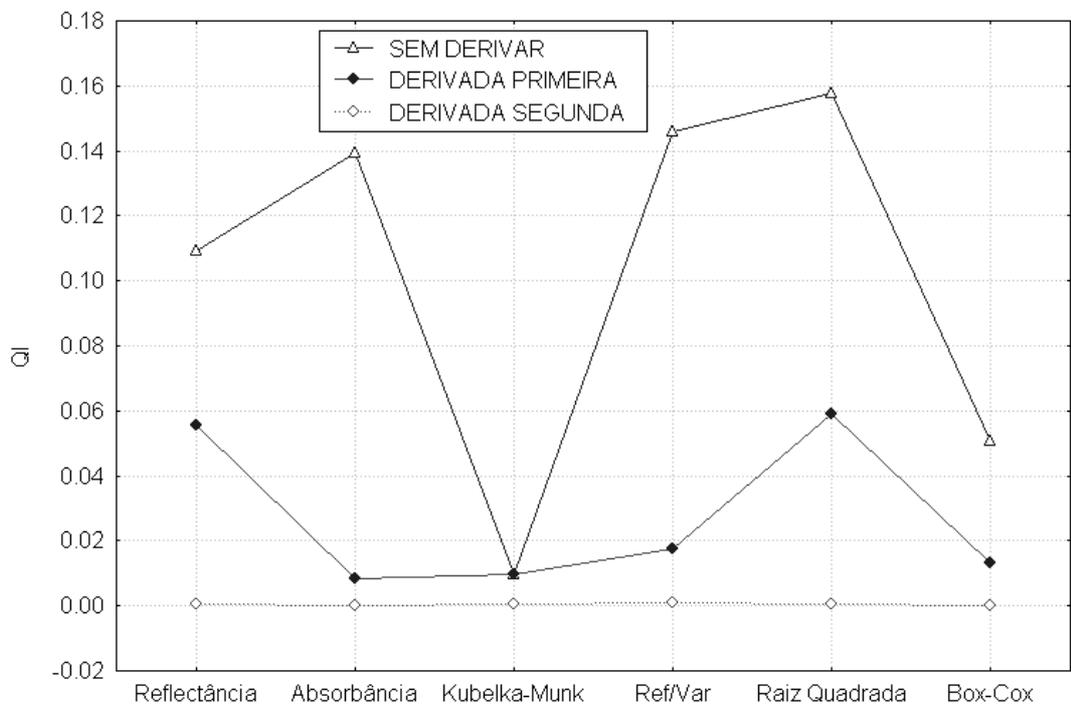


Figura 93 - Valores médios para o índice de qualidade dos modelos ajustados por PCR_FWD sobre a reflectância e suas transformações para a predição da densidade básica.

Ainda na Figura 93, destacam-se os modelos ajustados diretamente sobre os valores da reflectância e de suas transformações em absorbância, divisão pela variância e raiz quadrada da reflectância. Para estes modelos, os valores médios do índice de qualidade não diferem significativamente ao nível de 5% de probabilidade.

De maneira geral, as regressões por PCR_FWD forneceram modelos melhores que aqueles obtidos por PCR_ODV em termos de acurácia e precisão.

5.3. A rugosidade da madeira

A rugosidade da madeira, representada pelos diversos parâmetros de rugosidade fornecidos diretamente pelo rugosímetro, apresentou-se debilmente correlacionada à posição medula-casca, à densidade e aos espectros de reflectância e suas transformações. A Figura 94 apresenta a variação no sentido medula-casca dos parâmetros Ra e Rt.

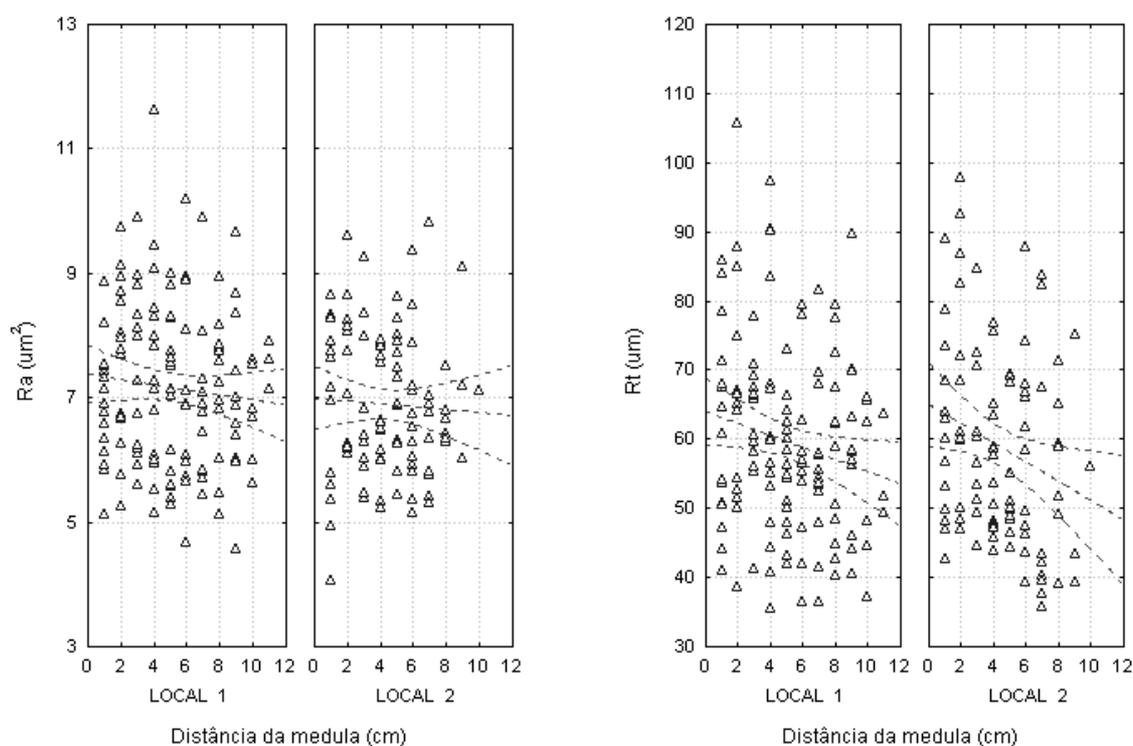


Figura 94 - Variação no sentido medula-casca para os parâmetros de rugosidade Ra e Rt.

Ao agruparem-se as amostras dos dois locais de procedência, as correlações continuam baixas, porém, à exceção do Ra, para os demais parâmetros assumem valores

significativos ao nível de 5% de probabilidade, tanto entre si, quanto com a densidade básica e a posição medula-casca. O Quadro 8 apresenta a matriz de correlação para estas variáveis, ao agrupar-se as amostras de ambos os locais.

Quadro 8 - Matriz de correlação para os valores dos parâmetros de rugosidade, a posição medula-casca e a densidade básica.

	Distância da medula	Densidade básica	Ra	Ry	Rt
Distância da medula	-	0,79*	-0,06	-0,18*	-0,20*
Densidade básica	0,79*	-	-0,21*	-0,25*	-0,26*
Ra	-0,06	-0,21*	-	0,54*	0,57*
Ry	-0,18*	-0,25*	0,54*	-	0,98*
Rz	-0,20*	-0,32*	0,80*	0,81*	0,84*
Rt	-0,20*	-0,26*	0,57*	0,98*	-

* Valores significativos a 5% de probabilidade.

As pequenas correlações guardadas com a densidade básica, variável modelada com relativo sucesso até agora, não representam um fato que por si só, seja desencorajador para a investigação de correlações entre os parâmetros de rugosidade e os espectros de reflectância e suas transformações. Outros componentes principais que não aqueles bem correlacionados à densidade perderiam apresentar correlações satisfatórias com a rugosidade, como é o caso do primeiro componente principal, aquele que carrega a maior parte da variância. Entretanto, nem a reflectância e suas transformações, nem tampouco os componentes principais de suas matrizes de covariância apresentaram boas correlações com o Ra ou com qualquer outro parâmetro de rugosidade. A Figura 95 revela como os coeficientes de determinação entre os componentes e o Ra apresentaram valores extremamente baixos e de comportamento nada coerente quando se comparam as correlações obtidas para os diferentes grupos de amostras (calibração e validação).

Na Figura 96 pode-se observar o comportamento dos erros, dos coeficientes de correlação, do número de componentes necessários à obtenção do erro mínimo e da estatística F para modelos ajustados por PCR seguindo-se a ordem decrescente de variância dos componentes.

O panorama geral mostrado nas Figuras 95 e 96 repetiu-se para todos os outros parâmetros de rugosidade, inclusive quando se aplicou a regressão *forward*. Ou seja, não foi possível obter-se bons modelos de calibração para a estimação da rugosidade da madeira utilizando-se a reflectância total de sua superfície na região do UV-VIS-tNIR.

Os coeficientes de determinação mostraram-se sempre baixos e os erros por demais elevados, apresentando-se o RMSEP com valores que variaram de 13,96 a 158,52% em relação aos parâmetros médios. Entretanto, cerca de 60% dos modelos apresentaram valores de F significativos a 5% de probabilidade, o que quer dizer que, em muitos casos, é possível obter-se modelos que, dentre as amostras de calibração são capazes de relacionar os efeitos das variáveis independentes sobre as dependentes, e que, ao serem validados, entretanto, não são aplicáveis a propósitos de predição. Recomenda-se que, para aumentar-se as chances de sucesso na modelagem da rugosidade, outras transformações matemáticas sejam aplicadas aos dados espectroscópicos, ou que a reflectância total seja auxiliada pela presença da reflectância difusa e da especular no conjunto de variáveis independentes.

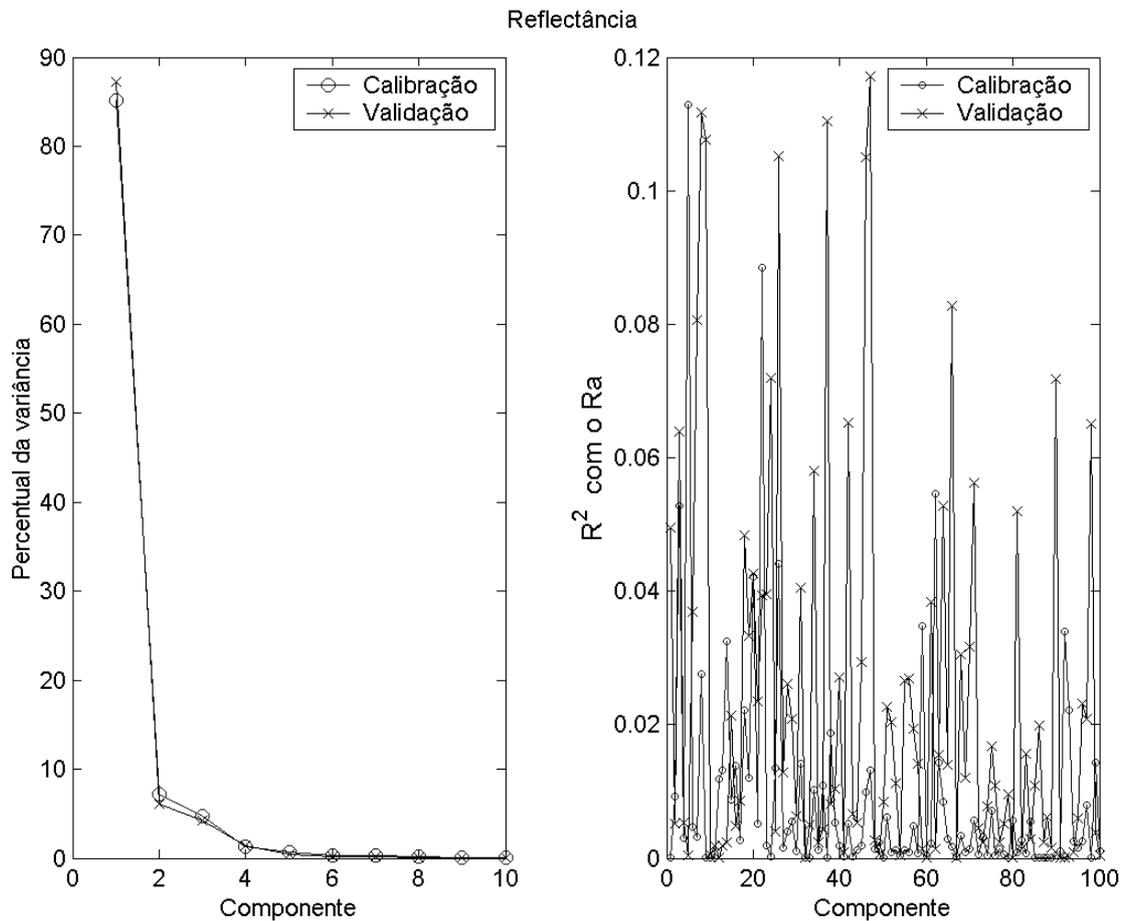


Figura 95 - Percentual da variância global carregado por cada componente principal da reflectância e correlações entre os componentes e a variável independente (Ra), agregadas as amostras dos Locais 1 e 2.

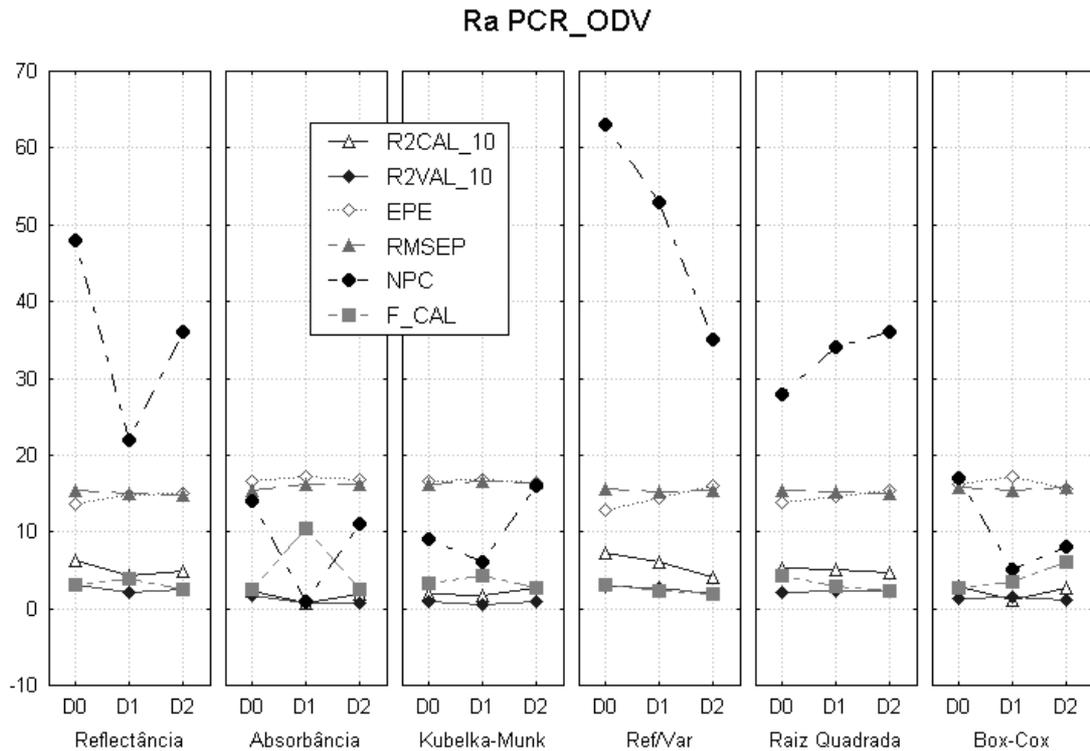


Figura 96 - Valores de R^2 na calibração e validação multiplicados por 10 (R2CAL_10 e R2VAL_10; dos erros na calibração e validação (EPE e RMSEP), do número de componentes necessários à obtenção do RMSEP mínimo e da estatística F para modelos ajustados por PCR-ODV para a predição da rugosidade média Ra.

5.4. A resistência à compressão e a dureza

A variação medula-casca para os valores da dureza e da resistência à compressão da madeira nos dois locais de procedência pode ser visualizada na Figura 97. Pode-se perceber que, para as árvores aqui testadas, a dureza apresenta um comportamento um tanto imprevisível, com uma grande variabilidade dentro de cada posição medida a partir da medula. A resistência à compressão, por sua vez, mostrou-se melhor correlacionada à posição de retirada das amostras, apesar de também apresentar altos coeficientes de variação.

O comportamento inesperado da dureza pode ser explicado pela possível inadequação do método EQUOTIP à aplicações sobre a madeira. Devido a pequena abrangência da área atingida pelo elemento impactador do durômetro, pode se levantar a dúvidas sobre a representatividade de suas leituras, já que, desta forma, os valores desta propriedade ficam extremamente susceptíveis à influências locais como a presença de elementos de vaso ou de outras variações físicas e anatômicas devidas à constituição

dos anéis de crescimento. Assim, as tomadas em três posições de cada face radial das amostras, parece não ter sido suficiente para representar a dureza EQUOTIP. É possível que um número maior de tomadas em posições uniformemente distribuídas ao largo das faces radiais possa aumentar a representatividade das leituras da dureza.

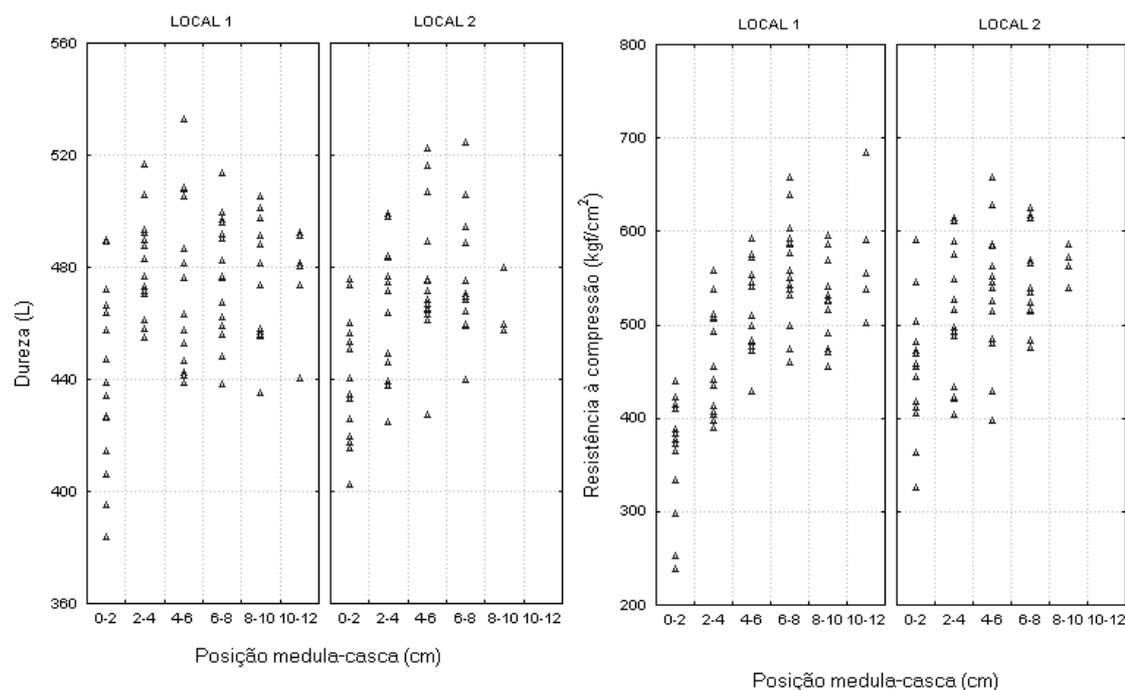


Figura 97 - Variação da dureza e da resistência à compressão no sentido medula-casca para as madeiras dos Locais 1 e 2.

A Figura 98 apresenta as relações entre a dureza e resistência à compressão e a densidade básica das amostras.

Note-se na Figura 98 que a resistência à compressão está mais bem correlacionada à densidade básica da madeira que a dureza. Os coeficientes de correlação para estas propriedades, para a média da reflectância e para a posição medula-casca, separando-se e agrupando-se as amostras dos diferentes locais de procedência são apresentados no Quadro 9. Note-se que, no Local 2, a reflectância média não encontra correlação nem com a posição medula-casca, nem com nenhuma das outras propriedades da madeira.

De fato, a madeira do Local 2 revelou-se anatomicamente mais heterogênea, com algumas árvores apresentando anéis de crescimento bem marcados e de pequenos diâmetros, inclusive nas regiões da próximas à medula, o que, provavelmente está

relacionado à adversidades nas condições de crescimento ali encontradas, como secas prolongadas e deficiências minerais ou à características genéticas de cada indivíduo que os levaram a condição de árvores suprimidas no povoamento. Além disso, muitas vezes a região da medula apresentou-se escurecida o que pode ser atribuído à intensa oxidação de polifenóis ou ao ataque de fungos. Estas pequenas variações locais certamente contribuíram para que a reflectância não encontrasse correlação nem com a posição medula-casca, nem às outras propriedades, as quais guardam uma certa correlação entre si.

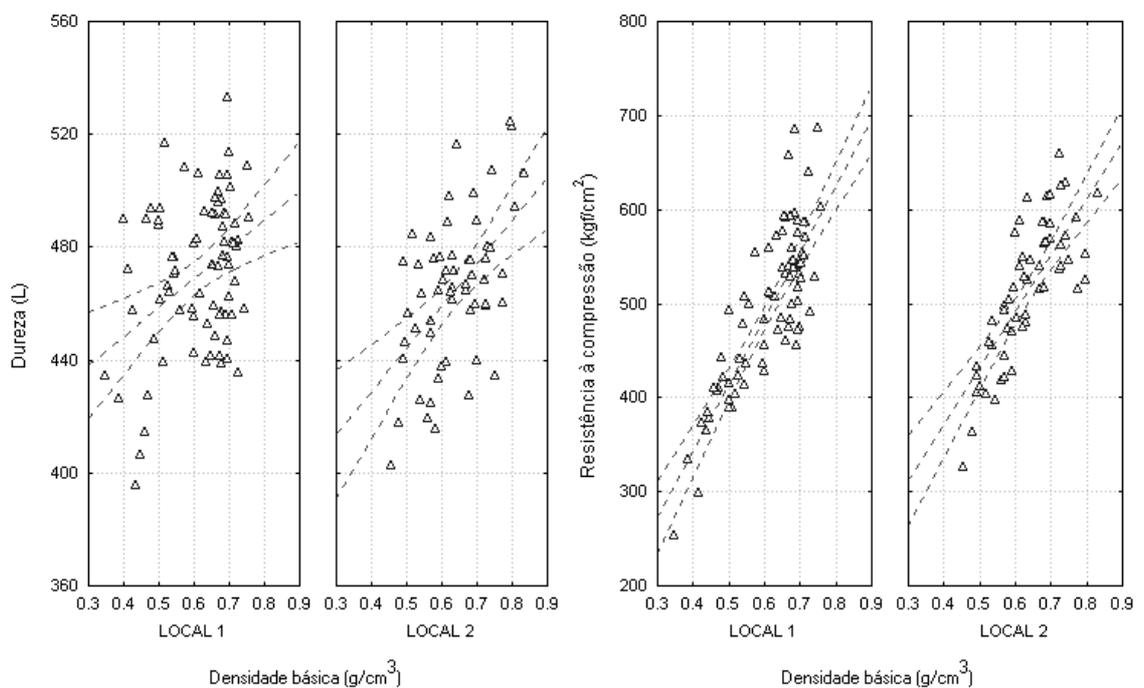


Figura 98 - Plotagem dos valores médios da dureza EQUOTIP e da resistência à compressão das amostras em função da sua densidade básica.

Por outro lado, a madeira do Local 1 apresentou a mais baixa correlação entre a dureza EQUOTIP e reflectância. Tal fato pode ser também explicado pela deficiência revelada pelo equipamento em determinar de forma representativa esta propriedade em amostras que apresentam variações anatômicas, como é o caso da madeira do Local 2, que possui anéis de crescimento mais largos.

Quadro 9 - Coeficientes de correlação entre a posição medula-casca, a dureza EQUOTIP a resistência à compressão e à reflectância média das amostras dos Locais 1 e 2 quando separadas e quando agrupadas

	Posição	Dureza	Res. Comp.	Dbas	Reflectância
Local 1					
Posição		0,31	0,63	0,82	-0,47
Dureza	0,31		0,41	0,37	0,02
Res. Comp.	0,63	0,41		0,81	-0,30
Dbas	0,82	0,37	0,81		-0,53
Reflectância	-0,47	0,02*	-0,30	-0,53	
Local 2					
Posição		0,45	0,49	0,66	0,06*
Dureza	0,45		0,39	0,53	0,21*
Res. Comp.	0,49	0,39		0,76	0,26*
Dbas	0,66	0,53	0,76		0,14*
Reflectância	0,06*	0,21*	0,26*	0,14*	
Locais 1 e 2					
Posição		0,37	0,54	0,72	-0,27
Dureza	0,37		0,38	0,41	0,10*
Res. Comp.	0,54	0,38		0,79	-0,09*
Dbas	0,72	0,41	0,79		-0,26
Reflectância	-0,27	0,10*	-0,09*	-0,26	

* Significativo a 5% de probabilidade.

A Figura 99 apresenta a variação dos coeficientes de determinação entre a dureza EQUOTIP, a resistência à compressão e a reflectância em cada comprimento de onda e com seus componentes principais. Os primeiros 5 PC's da reflectância destas amostras carregam 99% da variância global. Os problemas relacionados à representatividade da dureza EQUOTIP certamente interferiram negativamente nas suas correlações com a reflectância ou com seus componentes principais, levando a uma situação em que grupos diferentes de amostras apresentam diferentes correlações, além de serem estas extremamente baixas. Por outro lado, apesar de serem também baixos os coeficientes de determinação, a resistência à compressão apresentou uma certa consistência em suas correlações com a reflectância ou com seus componentes principais. Estas observações podem justificar a melhor performance dos modelos de calibração ajustados para a resistência a compressão se comparados àqueles obtidos para a dureza, nos quais o insucesso pode ser considerado total.

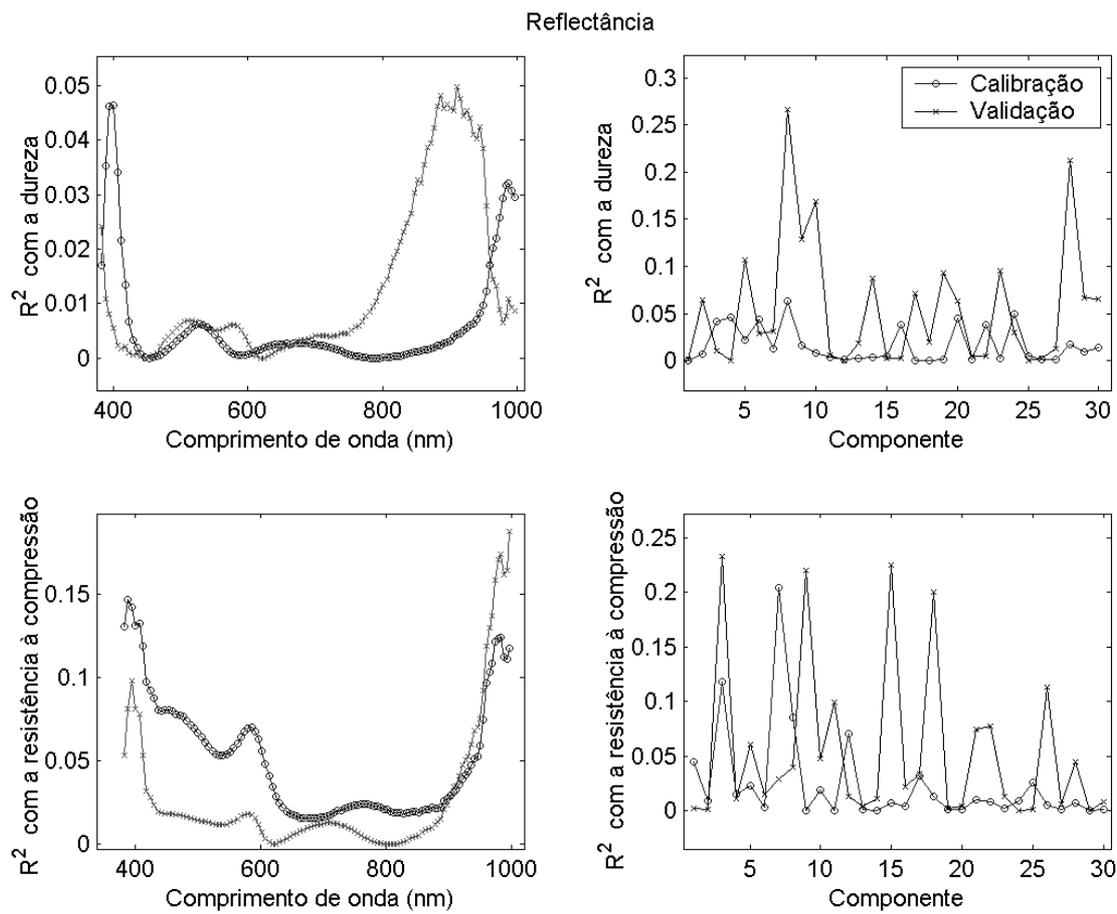


Figura 99 - Coeficientes de determinação entre a dureza EQUOTIP e resistência à compressão e a reflectância em cada comprimento de onda e seus primeiros 30 componentes principais, agrupando as amostras dos Locais 1 e 2 e tomando-se ao acaso dois grupos de amostras (validação e calibração).

O comportamento dos coeficientes de correlação observado na Figura 98 repetiu-se para todas as transformações aplicadas sobre a reflectância e suas derivadas. Ou seja, mesmo com exaustivas tentativas de pré-processamento dos dados, a dureza permaneceu não se deixando modelar de forma alguma. A resistência à compressão, por sua vez, revelou-se mais promissora, fornecendo modelos de melhor performance, apesar de ainda inadequados para propósitos preditivos, como será visto a seguir.

A Figura 100 apresenta os índices de qualidades para os modelos ajustados para a predição da dureza EQUOTIP e da resistência à compressão da madeira, agrupando-se as amostras dos dois locais de procedência. O pré-processamentos dos dados deu-se de maneira similar àquele aplicado para os dados da densidade básica, ou seja, foi realizada a redução da resolução espectral para 5 nm, aplicaram-se filtros polinomiais de Savitzky-Golay do terceiro grau sobre janelas de filtragem de 50 nm de

intervalo e foram removidos *outliers* com base na validação cruzada, descartando-se amostras que apresentaram erros maiores que duas vezes o desvio-padrão dos erros.

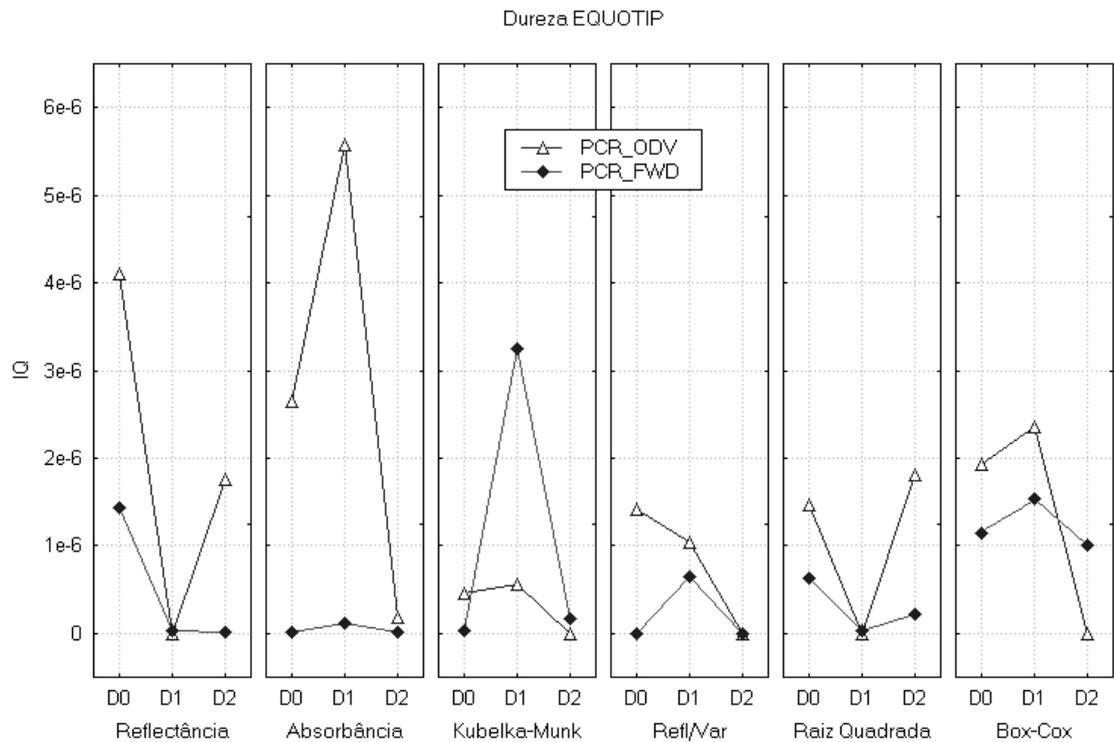


Figura 100 - Valores médios para o índice de qualidade dos modelos ajustados por PCR_ODV e PCR_FWD sobre a reflectância e suas transformações para a predição da dureza EQUOTIP.

O índices de qualidade dos modelos para a estimação da dureza mostraram-se extremamente baixos se comparados àqueles obtidos para a densidade básica. Tais valores são consequência, principalmente, dos baixos coeficientes de determinação e dos altos valores dos erros registrados na calibração e validação. Os modelos aqui obtidos mostraram-se simplesmente incapazes de prever a dureza.

Por outro lado, os índices de qualidade para os modelos sobre a resistência à compressão mostraram-se ligeiramente mais elevados que aqueles para a dureza, como pode ser observado na Figura 101. Os melhores índices foram fornecidos pelos modelos ajustados por PCR_ODV sobre a reflectância e sua primeira derivada, sobre a absorbância (inclusive pelo método *forward*), sobre as primeiras derivadas da razão entre a reflectância e sua variância e da raiz quadrada e sobre os valores da função de Box-Cox. Naturalmente, somente a análise este índice não é suficiente para escolha de um modelo. Sua vantagem está na fusão dos parâmetros em um único número que tem,

entretanto, uma função comparativa. Ou seja, dentre os modelos que apresentaram os melhores índices, há sempre a necessidade de se verificar os valores de cada um dos parâmetros e, além disso, o comportamento dos resíduos de cada modelo.

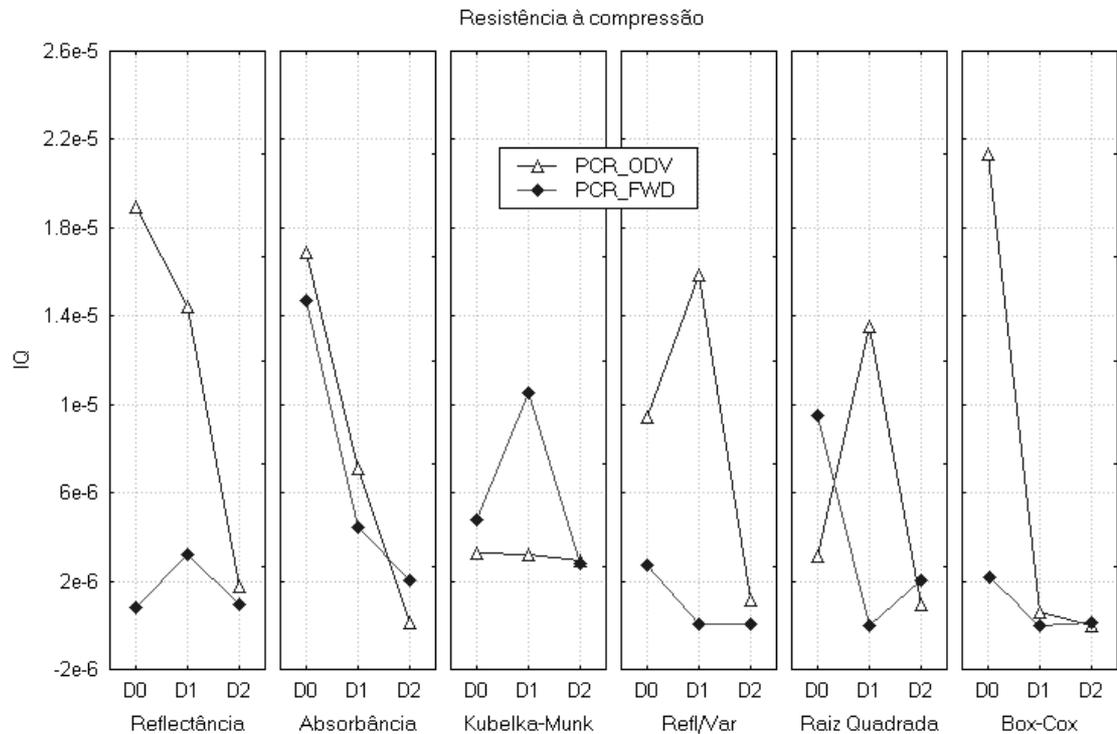


Figura 101 - Valores médios para o índice de qualidade dos modelos ajustados por PCR_ODV e PCR_FWD sobre a reflectância e suas transformações para a predição da resistência à compressão.

De fato, os modelos indicados como sendo os melhores na série ajustada para a resistência à compressão, todos apresentaram baixos valores de R^2 e altos erros na calibração e validação, além de desvios de comportamento tendencioso e erros sistemáticos. Na Figura 102, pode-se observar detalhadamente as características do modelo ajustado sobre os valores da função de Box-Cox, que apresentou o maior índice de qualidade para a resistência à compressão.

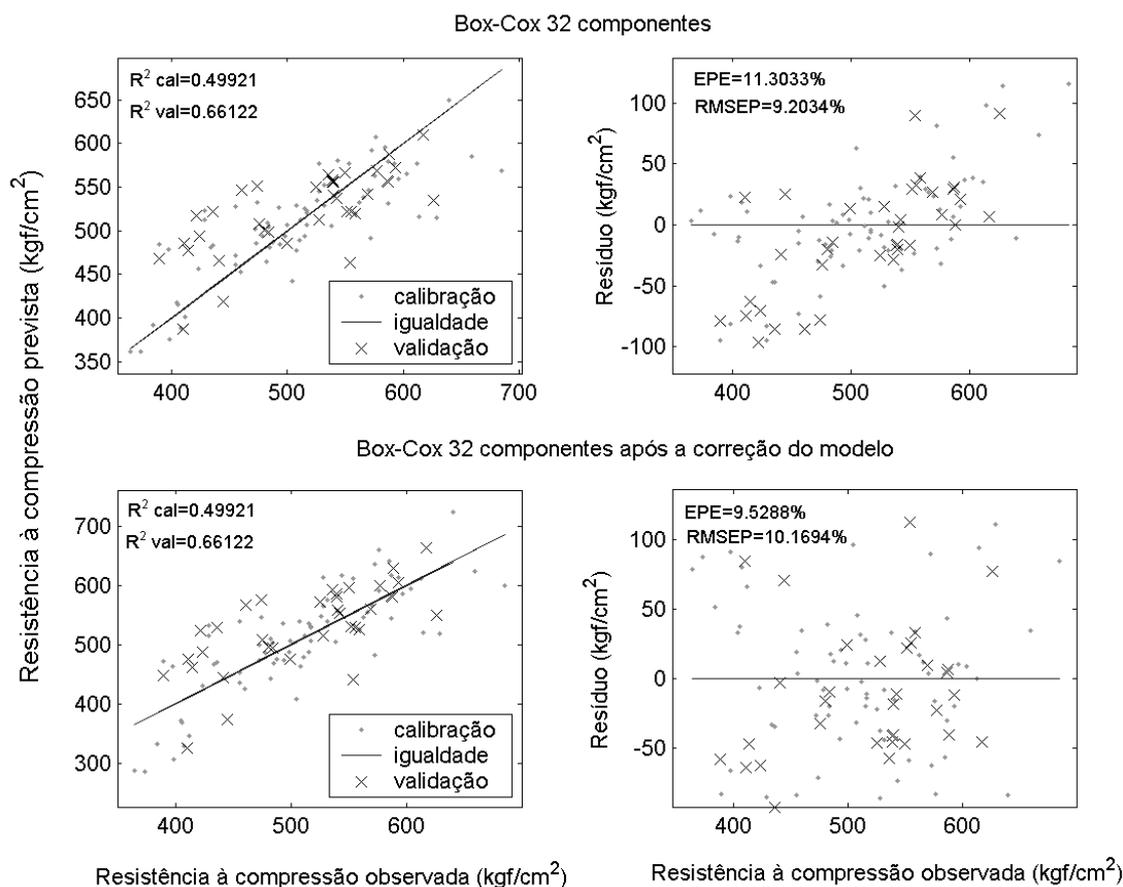


Figura 102 - Plotagem dos valores previstos e observados e resíduos para a resistência à compressão prevista por um modelos ajustado sobre os valores da função de Box-Cox aplicada à variância. Embaixo, apresenta-se o modelo após a correção do erro sistemático e da tendenciosidade dos desvios.

Fica clara na Figura 102 a necessidade de correção da tendenciosidade dos desvios e do erro sistemático representados pelo intercepto da reta previsto versus observado. Tal correção é possível se é também possível ajustar-se um modelo que correlacione de forma eficiente os valores previstos (y_m) e observados (y_o) e invertendo-se tal modelo para o cálculo de novos valores estimados:

$$y_m = a + by_o$$

$$y_o = (y_m - a) / b$$

$$y_{mcor} = (y_m - a) / b$$

Esta transformação elimina o intercepto e torna os desvios aleatórios, porém, não altera o valor do coeficiente de determinação. De posse do novo modelo de calibração, este pode ser testado na previsão dos valores da variável dependente no

grupo de amostras de validação. Caso o modelo de validação seja realmente válido, será também capaz de prever as propriedades das amostras de validação, agora, porém sem a tendenciosidade e sem a presença do erro sistemático.

De qualquer forma, a resistência à compressão também apresentou sérias limitações para a obtenção de modelos de calibração com base nos espectros de reflectância da superfície das amostras. Ao contrário do que foi observado para a dureza EQUOTIP, acredita-se que, no caso da resistência à compressão, as limitações sejam devidas à relativa ineficiência das tomadas dos espectros em representar a amostra como um todo. Neste ponto é oportuno lembrar que as amostras utilizadas para as calibrações para a densidade básica eram de dimensões bastante reduzidas (1 x 2 x 2 cm) se comparadas às de compressão (2 x 2 x 3cm), o que pode ter contribuído para o relativo sucesso das calibrações. Assim, sendo a resistência uma propriedade altamente dependente da estrutura anatômica da madeira, é possível que somente os espectros das faces radiais não tenham sido capazes de revelar o que acontece no interior da amostra.

6. CONCLUSÕES E SUGESTÕES

A densidade básica da madeira de *Eucalyptus urophylla* apresenta forte correlação com a reflectância de sua superfície radial nos comprimentos de onda em torno de 375 nm. Isto é válido quando se analisa as amostras de um mesmo local e também quando são agrupadas as amostras dos dois locais. Tal correlação, porém, não é suficientemente alta para fornecer a possibilidade de ajuste de modelos de calibração para a predição da densidade básica, pois apresentam erros de predição em torno de 8%.

Quando as amostras de calibração e validação são retiradas de um mesmo local, é possível obter-se modelos de calibração para a densidade com erros de predição em torno de 6% utilizando o método de regressão por componentes principais adicionando-se, em média, 17 componentes segundo a ordem decrescente de variância.

Quando as amostras de calibração e validação são retiradas de locais diferentes, não é possível obter-se modelos para a densidade com erros de predição menores que 8% utilizando o método de regressão por componentes principais adicionando-se os componentes segundo a ordem decrescente de variância.

Quando as amostras de calibração e validação são retiradas de locais diferentes, não é possível obter-se modelos para a densidade com erros de predição menores que 8% utilizando o método de regressão por componentes principais adicionando-se os componentes segundo a ordem decrescente de variância.

Se as amostras de calibração e validação são obtidas misturando-se as amostras de locais diferentes, torna-se novamente possível obter-se modelos de calibração para a

densidade com erros de predição em torno de 6% utilizando o método de regressão por componentes principais adicionando-se os componentes segundo a ordem decrescente de variância. O número de componentes necessários para a obtenção desse nível de erro é, porém, variável.

A detecção de *outliers* pelo método da validação cruzada, com a posterior remoção de tais amostras do conjunto de dados, contribuiu significativamente para a estabilidade e exatidão dos modelos.

É possível otimizar-se os modelos de calibração aplicando-se técnicas de filtragem dos espectros, redução da resolução espectral e realizando-se transformações matemáticas sobre os espectros. Dentre todas as formas de pré-processamento dos dados espectroscópicos testadas, o uso da reflectância, da absorbância, da razão entre a reflectância e a sua variância e da extração da raiz quadrada da reflectância apresentaram os melhores resultados em termos de acurácia e precisão na previsão da densidade básica. O uso do método de regressão por componentes principais *forward* trouxe, em alguns casos, ganhos na capacidade de predição dos modelos de calibração para a densidade básica.

Não foi possível obter-se bons modelos de calibração para a estimação da rugosidade da madeira utilizando-se a reflectância total de sua superfície na região do UV-VIS-tNIR, mesmo aplicando-se sobre os espectros as várias transformações matemáticas citadas anteriormente. Os coeficientes de determinação mostraram-se sempre baixos e os erros por demais elevados, apresentando-se o RMSEP com valores que variaram de 13,96 a 158,52% em relação aos parâmetros médios. Entretanto, cerca de 60% dos modelos apresentaram valores de F significativos a 5% de probabilidade, o que quer dizer que, em muitos casos, é possível obter-se modelos que, dentre as amostras de calibração são capazes de relacionar os efeitos das variáveis independentes sobre as dependentes, e que, ao serem validados, entretanto, não são aplicáveis a propósitos de predição. Recomenda-se que, para aumentar-se as chances de sucesso na modelagem da rugosidade, outras transformações matemáticas sejam aplicadas aos dados espectroscópicos, ou que a reflectância total seja auxiliada pela presença da reflectância difusa e da especular no conjunto de variáveis independentes.

Não se obteve sucesso no ajuste de modelos para a estimação da dureza EQUOTIP. Os problemas encontrados foram atribuído à deficiência do método EQUOTIP de determinação da dureza em fornecer resultados que representem bem esta propriedade ao ser utilizado sobre materiais heterogêneos. A aplicação deste método

sobre a madeira talvez requeira uma distribuição mais ampla e uniforme dos pontos de teste dentro das faces das amostras, dadas as diminutas dimensões da ponta de prova do aparelho.

A resistência à compressão também apresentou sérias limitações para a obtenção de modelos de calibração com base nos espectros de reflectância da superfície das amostras. Ao contrário do que foi observado para a dureza EQUOTIP, acredita-se que, no caso da resistência à compressão, as limitações sejam devidas à relativa ineficiência das tomadas dos espectros em representar a amostra como um todo. Assim, sendo a resistência uma propriedade altamente dependente da estrutura anatômica da madeira, é possível que somente os espectros das faces radiais não tenham sido capazes de revelar o que acontece no interior da amostra.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, J. M. **Chemometrics in analytical spectroscopy**. Cambridge: Royal Society of Chemistry. 1995. 216 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. **MB-26/40**: ensaios físicos e mecânicos da madeira. Rio de Janeiro, ABNT, 1940. 16 p.

BAILLERES, H.; DAVRIEUS, F.; PICHAVANT, F. H. Near infrared analysis as a tool for rapid screening of some major wood characteristics in a *Eucalyptus* breeding program. **Annals of Forest Science**, v. 59, n. 5-6, p. 479-490, 2002.

BEKHTA, P. A. The study of sound propagation in the wood-based composite materials. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

BINDZI, I.; SAMSON, M. A new method of grading lumber for spiral grain. **Forest Products Journal**, v. 45, n. 2, p. 63-66, 1995.

BODIG, J. The process of NDE research for wood and wood composites. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

BROWN, H. P.; PANSHIN, A. J.; FORSAITH, C. C. **Text book of wood technology**. New York: The Maple Press Company. 1949. v.1. 652 p.

BUCUR, V. Acoustics as a tool for the nondestructive testing of wood. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

BURDZIK, W. M. G.; NKWERA, P. D. Transverse vibration tests for predictions of stiffness and strength properties of full size *Eucalyptus grandis*. **Forest Products Journal**, v. 52, n. 6, p. 63-67, 2002.

BUTLER, D. A.; BRUNNER, C. C.; FUNCK, J. W. Wood surface feature classification via extended-color imagery. **Forest Products Journal**, v. 52, n. 6, p. 80-84, 2002.

CAREY, F. A. A.; SUNDBERG, R. J. **Advanced organic chemistry: structure and mechanisms (part A)**. New York: Kluwer Academic Publishers. 832 p. 1990.

CAVALCANTI, P. R. **Introdução à computação gráfica**. Notas de aula do Curso de Computação Gráfica I. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2000. 178 p.

CIENFUEGOS, F.; VAITSMAN, D. **Análise instrumental**. Rio de Janeiro: Editora Interciência, 2000, 610 p.

CLARK, R. N.; ROUSH, T. L. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. **Journal of Geophysics Research**, v. 89, p. 6329-6340, 1984.

COATES, J. Encyclopedia of analytical chemistry. In: MEYERS, R. A. (Ed.). **Interpretation of infrared spectra, a practical approach**. Chichester: John Wiley & Sons Ltd, 2000. p. 10815-10837.

FARDIM, P.; FERREIRA, M. M. C.; DURAN, N. Multivariate calibration for quantitative analysis of eucalypt Kraft pulp by NIR spectrometry. **Journal of Wood Chemistry and Technology**, v. 22, n. 1, p. 67-81, 2002.

FULLER, J. J.; ROSS, R. J. DRAMM, J. R. Honeycomb and surface check detection using ultrasonic nondestructive evaluation. USDA Forest Service. FPL. Madison: Research Note FPL-RN-0621, 1994. 6 p.

FUNT, B. V.; BRYANT, E. C. Detection of internal log defects by automatic interpretation of computer tomography images. **Forest Products Journal**, v. 37, n. 1, p. 56-62, 1987.

GINDL, W.; TEISCHINGER, A. The potential of VIS and NIR spectroscopy for the nondestructive evaluation of grain-angle in wood. **Wood and Fiber Science**, v. 34, n. 4, p. 651-656, 2002.

GINDL, W.; TEISCHINGER, A.; SCHWANNINGER, M.; HINTERSTOISSER, B. The relationship between near infrared spectra of radial wood surfaces and wood mechanical properties. **Journal of Near Infrared Spectroscopy**, v. 9, n. 4, p. 255-261, 2001.

GONZALES, R. C.; WOODS, R. E. **Digital image processing**. Massachusetts: Addison-Wesley, 1993. 716 p.

GRIMBERG, R.; SAVIN, A.; LUPU, A.; ROTUNDU, C. A method to determine the debonding zones in multilayers wood materials. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

HEIJDEN, W. A. M. G.; POLDER G.; GEVERS, T. Comparison of multispectral images across the Internet. In: BERETTA, G. B.; SCHETTINI, R. (Eds.). **Proceedings of SPIE, volume 3964: internet imaging.** San Jose: 2000. p. 196-206.

HYVARINEN, T. L.; HERRALA, E.; DALL'AVA, A. Direct sight imaging spectrograph: a unique add-on component brings spectral imaging to industrial applications. In: IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology (EI98), in Conference 3302: Digital Solid State Cameras: Design and Applications, Paper 3302-21, January 25-30, 1998, San Jose Convention Center, San Jose: 1998.

JACKSON, E. J. **A user's guide to principal components.** New York: John Wiley & Sons. 1991. 570 p.

KHOURY JUNIOR, J. K.; PINTO, F. A. C.; SANTOS, N. T.; DELLA LÚCIA, R. M.; MAEDA, E. E. Análise discriminante paramétrica para reconhecimento de defeitos em tábuas de eucalipto utilizando imagens digitais. **Revista Árvore**, v. 29, n. 2, p. 299-309, 2005.

KLINE, E. D.; WIDOYOKO, A.; WIEDENBECK, J. K.; ARAMAN, P. A. Performance of color camera machine vision in automated furniture rough mill systems. **Forest Products Journal**, v. 48, n. 3, p. 88-45, 1998.

KOLLMANN, F. **Tecnología de la madera y sus aplicaciones.** Madrid: Tomo I. Instituto Forestal de Investigaciones y Experiencias y Servicio de la Madera. 1959. 674 p.

LEBOW, P. K.; BRUNNER, C. C.; MARISTANY, A. G.; BUTLER, D. A. Classification of wood surface features by spectral reflectance. **Wood and Fiber Science**, v. 28, n. 1, p. 74-90, 1996.

LIN, S. Y.; DENCE, C. W. **Methods in lignin chemistry.** Berlin: Springer-Verlag. 1992, 578 p.

LISSETTE, G.; FERENC, D. Possibility of glue-defect detection in solid wood panels with ultrasound. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

MARKLUND, A.; PAPER, M.; HAUKSSON, J. B.; SJOSTROM, M. Prediction of strength parameters for softwood Kraft pulps. Multivariate data analysis based on orthogonal signal correction and near infrared spectroscopy. **Nordic Pulp & Paper Research Journal**, v. 14, n. 2, p. 140-148, 1999.

MARTEN, H.; NÆS, T. **Multivariate calibration**. New York: John Wiley & Sons, 1996. 420 p.

MICHELL, A. J.; HIGGINS, H. G. **Infrared spectroscopy in australian forest products research**. Melbourne: CSIRO Forestry and Forest Products, 2002. 56 p.

MOSCHLER, W. W.; WINISTORFER, P. W. **Radiation based densitometry: the history and evolution of a unique program at UT**. Universidade do Tennessee. 2000. Disponível em: <<http://web.utk.edu/~tfpc/person/densres.htm>>. Acesso em: 20 mar. 2005.

NÆS, T.; ISAKSSON, T.; FEARM, T.; DAVIES, T. **A user-friendly guide to multivariate calibration and classification**. NIR Publications. Chichester, UK. 2002. 344 p.

NISHINO, Y.; SHINAGAWA, R. A scanner could work like a colorimeter. **Wood Industry**, v. 54, n. 4, p 167-169, 1999.

OJA, J.; GRUNDBERG, S.; GRONLUND, A. Predicting the strength of sawn products by X-ray scanning of logs: A preliminary study. **Wood and Fiber Science**, v. 32, n. 2, p. 203-208, 2000.

PANSHIN, A. J.; DE ZEEUW, C. **Textbook of wood technology**. 3.ed. New York: Mc Graw-Hill, 1970. 705 p.

PERKAMPUS, H. H. **UV-VIS spectroscopy and its applications**. Berlin: Springer-Verlag, 1992. 244 p.

PHAM, D. T.; ALCOCK, R. J. Automated grading and defect detection. **Forest Products Journal**, v. 48, n. 4, p. 34-42, 1998.

RAYMOND, C. A.; SCHIMLECK, L. R. Development of near infrared reflectance analysis calibration for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. **Canadian Journal of Forest Research**, v. 32, n. 2, p. 170-176, 2002.

ROSS, R. J. Using sound to evaluate standing timber. **International Forestry Review**, v. 1, n. 1, p. 43-44, 1999.

ROSS, R. J.; WANG, X.; MATTSON, J. A.; ERICKSON, J. R.; FORSMAN, J. W.; GESKE, E. A.; WEHR, M. A. Comparison of several nondestructive evaluation techniques for assessing stiffness and MOE of small-diameter logs. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

RUST, S. A new tomographic device for the non - destructive testing of trees. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

- SAMSON, M. Measuring general slope of grain with slope-of-grain indicator. **Forest Products Journal**, v. 34, n. 7-8, p. 27-32, 1984.
- SAMSON, M. Transverse scanning for automatic detection of general slope of grain in lumber. **Forest Products Journal**, v. 38, n. 7-8, p. 33-38, 1988.
- SAMSON, M.; TREMBLAY, C.; LANGLAIS, P. A. Measuring slope of grain by electrical capacitance at moisture contents above fiber saturation. **Forest Products Journal**, v. 43, n. 2, p. 58-60, 1993.
- SCHAD, K. C.; SCHMOLDT, D. L.; ROSS, R. J. **Nondestructive methods for detecting defects in softwood logs**. U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. Madison: Res. Pap. FPL-RP-546, 1996. 13 p.
- SCHAJER, G. S. Lumber strength grading using x-ray scanning forest products society member. **Forest Products Journal**, v. 51, v. 1, p. 43-50, 2001.
- SCHIMLECK, L. R.; EVANS, R. Estimation of microfibril angle of increment cores by near infrared spectroscopy. **IAWA Journal**, v. 23, n. 3, p. 225-234, 2002.
- SCHIMLECK, L. R.; MICHELL, A. J.; RAYMOND, C. A.; MUNERI, A. Estimation of basic density of *Eucalyptus globulus* using near-infrared spectroscopy. **Canadian Journal of Forest Research**, v. 29, n. 2, p. 194-201, 1999.
- SEARS, F. W.; ZEMANSK, M. W. **Física. V. 3. eletricidade, magnetismo e tópicos de física moderna**. Rio de Janeiro: Livros Técnicos e Científicos, 1977. 685 p.
- SILVA, J. C. **Caracterização da madeira de *Eucalyptus grandis* Hill ex. Maiden, de diferentes idades visando a sua utilização na indústria moveleira**. 2002. 160 f. Tese (Doutorado em Ciências Florestais) – Universidade Federal do Paraná, Curitiba, 2002.
- SOLOMONS, T. W. G. **Química orgânica V1**. Rio de Janeiro: Livros Técnicos e Científicos, 1992. 532 p.
- SUNDQVIST, B. **Colour changes and acid formation in wood during heating**. 2004. 154 f. Thesis (Ph.D.) – Lulea University of Technology, Lulea, 2004.
- SWAIN, P. H.; DAVIS, S. M. **Remote sensing: the quantitative approach**. New York : McGraw-Hill, 1978. 396 p.
- STEELE, P. H.; NEAL, S. C.; MCDONALD, K. A.; CRAMER, S. M. The slope-of-grain indicator for defect detection in unplaned hardwood lumber. **Forest Products Journal**, v. 41, n. 1, p. 15-20, 1991.
- SZYMANI, R.; MCDONALD, K. A. Defect detection in lumber - state of the art. **Forest Products Journal**, v. 31, n. 11, p. 34-44, 1981.
- THOMAS, M. J. K. **Ultraviolet and visible spectroscopy**. Chichester: John Wiley & Sons, 1996. 230 p.

THUMM, A.; MEDER, R. Stiffness prediction of radiata pine clearwood test pieces using near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, v. 9, n. 2, p. 117-122, 2001.

TREMBLAY, C. Longitudinal and radial variation of slope of grain in black spruce lumber. **Forest Products Journal**, v. 45, n. 1, p. 79-83, 1995.

TROUGHTON, G. E.; CLARKE M. R. Development of a new method to measure moisture content in unseasoned veneer and lumber. **Forest Products Journal**, v. 37, n. 1, p. 13-19, 1987.

WANG, X.; ROSS, R. J.; ERICKSON, J. R.; GARY, J. W. F.; MCGINNIS, D.; PELLERIN, R. F. Nondestructive evaluation of standing trees with stress wave methods. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

WIBERG, P.; LINDGREN, O. Moisture gradient development during wood drying followed by CT-scanning and image processing. In: INTERNATIONAL SYMPOSIUM ON NDT OF WOOD, 10., 1996, Switzerland. **Proceedings...** Switzerland: Lausanne, 1996. p. 16-28.

YAMAMOTO, K. Nondestructive test on the evaluation of basic wood properties of fast growing plantation tree in vietnam. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

ZOMBORI, B. "In situ" nondestructive testing of built in wooden members. In: INTERNATIONAL SYMPOSIUM ON NONDESTRUCTIVE TESTING OF WOOD, 12., 2000. **Proceedings...** Sopron: University of Western Hungary, 2000.

APÊNDICE A

A.1. A regressão por componentes principais

A regressão por componentes principais (PCR, do inglês: *Principal Components Regression*) deriva de uma importante técnica da estatística multivariada, a análise por componentes principais (PCA), também conhecida como transformada de Hotelling. A PCA consiste na transformação de um grupo de variáveis possivelmente correlacionadas em novas variáveis, chamadas componentes principais (PC's), totalmente descorrelacionadas. Além disso, as novas variáveis criadas são hierarquizadas de maneira tal que a maior parte da variância original é concentrada nos primeiros componentes, o que permite a redução do número de variáveis independentes (por seleção) a serem computadas no ajuste de modelos para predição de uma variável dependente. Para um conjunto de p variáveis independentes $[X_1, X_2 \dots X_p]$, obter-se-ão p componentes principais $[PC_1, PC_2 \dots PC_p]$. A PCA consiste, portanto, na mudança de base dos dados antes pertencentes ao espaço gerado pelos eixos X para o espaço dos eixos PC. Tal transformação é de grande utilidade no ajuste de modelos em situações que:

- a) O número de variáveis independentes (p) é maior ou igual ao número de amostras (n). Nesta situação, a regressão por mínimos quadrados ordinários (MQO) não pode ser aplicada diretamente, pois o Erro-padrão da estimativa (EPE ou $S_{y.x}$) tenderia ao infinito se $p=n$ ou não seria definido se $p>n$. O Erro-padrão da estimativa representa o desvio-padrão das diferenças (resíduos: $y - \hat{y}$) entre os valores reais da variável dependente (y) e os valores preditos pelo modelo (\hat{y}) e é definido pela expressão:

$$S_{y.x} = \sqrt{\frac{y'y - b'X'y}{n - p - 1}} \quad (\text{Eq. 5})$$

em que:

y = vetor dos valores da variável independente;

b = vetor dos valores estimados para os parâmetros do modelo;

X = matriz ($n \times p$) dos valores da variável independente.

Com a transformação é possível reduzir-se o número de variáveis independentes que participarão do modelo, reduzindo-se também $S_{y.x}$, o que torna o modelo mais preciso. O critério para a escolha dos PC's que participarão da regressão,

geralmente, é o percentual da variância que carregam. Ou seja, os primeiros componentes são incluídos preferencialmente nos modelos. Porém, tal procedimento não é uma regra, podendo até mesmo ser perigoso para a capacidade de predição dos modelos. Jakcson (1991) relatou vários casos em que a inclusão de PC's de menor módulo pode melhorar a acurácia de modelos. Adams (1995) e Marten e Næs (1996) sugeriram o emprego de técnicas do tipo *forward*, *backward* e *stepwise* para a escolha das variáveis (PC's) que participarão do modelo.

b) As variáveis independentes apresentam-se correlacionadas entre si. Tal situação implica em duas conseqüências importantes para a acurácia do modelo:

Primeiro, se as colunas da matriz X estão correlacionadas, a matriz $X'X$ também apresentará colunas correlacionadas e seu determinante será próximo de zero (se houverem pelo menos duas colunas perfeitamente correlacionadas o determinante será igual a zero). Ou seja, lidar-se-á com situações em que a matriz $X'X$ é uma matriz singular ou quase-singular, o que impede ou dificulta a sua inversão para a resolução da equação:

$$b = [X'X]^{-1} X'y \quad (\text{Eq. 6})$$

Tal resolução é necessária para a determinação do vetor dos coeficientes b por MQO. Algoritmos computacionais têm sido empregados no cálculo de inversas de matrizes quase singulares, porém os resultados são instáveis e a confiabilidade dos modelos é pequena (JACKSON, 1991).

Em segundo lugar, a presença de correlações altas entre variáveis independentes implica também na ampliação dos erros padrões para os valores dos coeficientes da regressão (S_{bi}) a níveis inaceitáveis. Tais erros são calculados a partir da expressão:

$$S_{bi} = \sqrt{\frac{S_{y,x}^2}{s_{xi}^2(n-1)(1-R_i^2)}} \quad (\text{Eq. 7})$$

em que:

i = identificador da variável independente x ;

$S_{y,x}^2$ = Erro-padrão da estimativa (Eq. 1) elevado ao quadrado;

s_{xi}^2 = variância da variável x_i ;

R_i = coeficiente de correlação múltiplo entre a variável x_i e as demais variáveis independentes.

Na Eq. 7 a razão $[1 / (1 - R_i^2)]$ é denominada *fator de inflação da variância* e, quanto mais próximo da unidade apresenta-se o valor absoluto do coeficiente R_i , mais próximo do infinito estará este fator, fazendo com que o Erro-padrão para os coeficientes b também tenda para o infinito. O emprego da PCA garante que as variáveis criadas (PC's) serão perfeitamente ortogonais eliminando-se o problema da correlação entre variáveis independentes.

A.1.1. As direções dos eixos PC's

A seguir será apresentado um exemplo de aplicação da PCR que enfocará um problema com apenas duas variáveis independentes (X1 e X2) e uma dependente (Y) que são plotadas na Figura A1.1. Este número reduzido de variáveis não permite expor as potencialidades da técnica, porém permite a possibilidade de visualização gráfica das transformações.

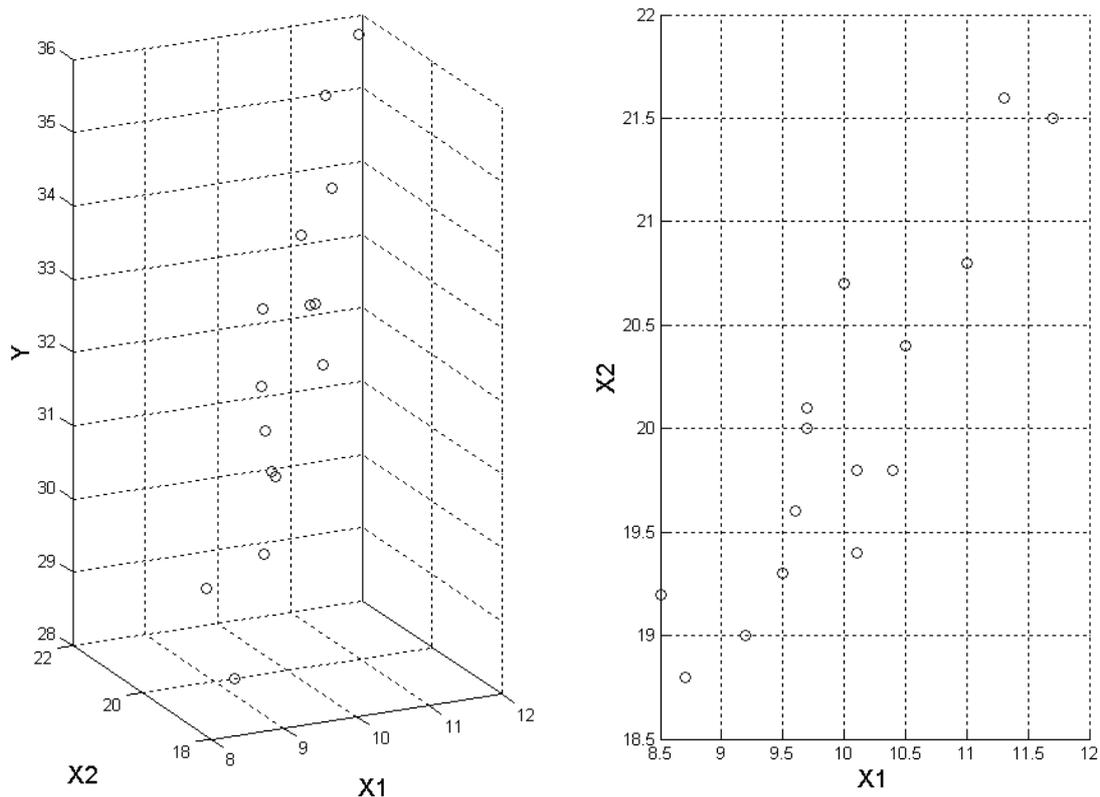


Figura A1.1 - Plotagem dos valores de X1, X2 e Y.

Como pode ser notado na Figura A1.1, a correlação entre X1 e X2 é bastante expressiva, apresentando um coeficiente $R = 0,77$. Tal fato sugere o uso da PCR para

predição de Y em função de X_1 e X_2 . Como exposto anteriormente, a PCR consiste de uma transformação das variáveis independentes com a criação de novas variáveis PC's que são combinações lineares das originais X . Ou seja, os eixos X_1 e X_2 darão lugar a novos eixos PC1 e PC2 sobre os quais será finalmente realizada a regressão para se encontrar um estimador para Y pelo método dos mínimos quadrados ordinários (MQO).

Ver-se-a inicialmente o cálculo de PC1, pois a partir dele pode-se chegar a PC2, considerando-se a premissa de que são ortogonais. A literatura sobre PCA comumente apresenta a solução para o cálculo dos PC's a partir de técnicas matriciais, o que é extremamente útil quando o número de variáveis X é grande, porém se pode introduzir o conceito de PCA com o auxílio da geometria analítica, o que auxiliará na visualização das posições dos novos eixos em relação aos eixos originais. Este enfoque analítico foi aqui desenvolvido a partir da exposição de Jackson (1991) sobre o conceito de PCA. Este autor chama a técnica para a obtenção dos PC's de "regressão ortogonal", já que ela visa obter uma reta (no caso de apenas duas variáveis X) que minimize os desvios perpendiculares a si mesma. Esta definição deu origem ao desenvolvimento exposto a seguir.

Na Figura A1.2. são apresentados os valores de X_1 e X_2 plotados em um sistema cartesiano convencional. Se se deseja obter uma reta "R" para a estimação de X_2 em função de X_1 , tal reta deverá apresentar parâmetros "a" e "b" que minimizem a soma dos quadrados dos desvios "I". Para obter tais parâmetro basta aplicar MQO tomando-se X_1 como variável independente. Se, por outro lado, se se deseja obter um estimador para X_1 em função de X_2 , aplica-se MQO tomando-se X_2 como variável independente, o que minimizará a soma dos quadrados dos desvios "C". Ao se minimizar "I" ou "C" obtêm-se as chamadas, por razões históricas, *calibrações inversa* e *clássica*, respectivamente (NÆS et. al., 2002).

A obtenção de PC1, por sua vez, requer que seja minimizada a soma dos quadrados dos desvios "O" perpendiculares à reta R utilizando-se a é chamada técnica de regressão ortogonal para o cálculo dos coeficientes "a" e "b". Nesta técnica, tanto assumindo X_1 , quanto X_2 como variável independente, os resultados serão idênticos. Prossegue-se com a descrição dos estimadores de "a" e "b" na regressão ortogonal tomando X_1 como variável independente.

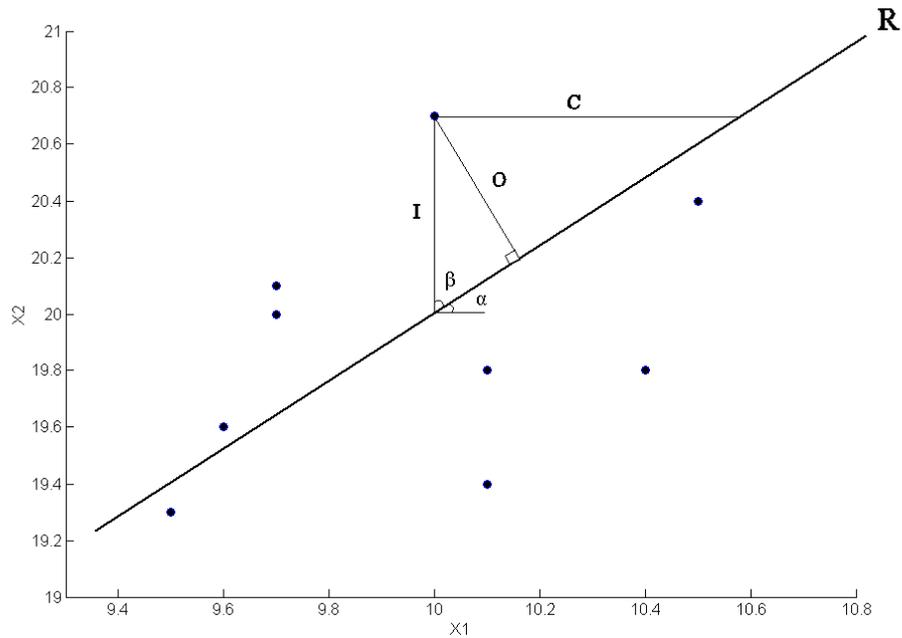


Figura A1.2 - Gráfico de X1 versus X2, evidenciando os desvios em relação a uma reta R que correlaciona estas variáveis.

Na Figura A1.2, o ângulo α é definido como $\arctg(b)$, ou seja, o arco cuja tangente é igual ao coeficiente angular da reta. O ângulo β , por sua vez, é o complemento de α . O desvio “O” para o i -ésimo ponto amostral pode ser expresso em função de “a”, “b” e dos valores de X1 e X2 pela equação:

$$O_i = I_i \text{sen}(\beta) \quad (\text{Eq. 8})$$

Como:

$$\text{sen}(\beta) = \cos(\alpha) = \cos(\arctg(b));$$

$$I_i = (X2_i - \hat{X}2_i) \text{ e}$$

$$(X2_i - \hat{X}2_i) = (X2_i - a - bX1_i),$$

substituindo as expressões acima na Eq. 8 tem-se:

$$O_i = (X2_i - a - bX1_i) \cos(\arctg(b)) \quad (\text{Eq. 9})$$

Assim, para encontrar os valores de “a” e “b” que minimizam a soma dos quadrados dos desvios “O” basta resolver o sistema:

$$\left\{ \begin{array}{l} \frac{d}{da} \sum_{i=1}^n [(X2_i - a - bX1_i) \cos(\arctg(b))]^2 = 0 \\ \frac{d}{db} \sum_{i=1}^n [(X2_i - a - bX1_i) \cos(\arctg(b))]^2 = 0 \end{array} \right. \quad (\text{Eq. 10})$$

$$\left\{ \begin{array}{l} \frac{d}{da} \sum_{i=1}^n [(X2_i - a - bX1_i) \cos(\arctg(b))]^2 = 0 \\ \frac{d}{db} \sum_{i=1}^n [(X2_i - a - bX1_i) \cos(\arctg(b))]^2 = 0 \end{array} \right. \quad (\text{Eq. 11})$$

A partir das equações 10 e 11 obtêm-se os estimadores para a constante “a” que é o mesmo estimador utilizado no MQO ($a = \bar{X}_2 - b \bar{X}_1$) e para o coeficiente angular “b”, respectivamente. A resolução da Eq. 11, no caso de duas variáveis independentes, apresentará duas raízes para “b”, uma indicando o ponto de máximo e outra o mínimo para o resíduo “O”. Neste ponto é oportuno observar que a complexidade da resolução algébrica das equações 10 e 11 cresce com o número de variáveis envolvidas na análise. A Figura A1.3. apresenta as retas $X_2 = a + bX_1$ ajustadas pelas regressões por MQO inversa e clássica e pela regressão ortogonal e o comportamento da soma de quadrado dos resíduos (SQO) obtidas com cada técnica fazendo-se variar arbitrariamente os parâmetros “a” e “b”. Note-se como o valor da SQO é minimizado para certos valores de “b” nas três técnicas de ajuste e que somente para a regressão ortogonal existe um ponto de máximo.

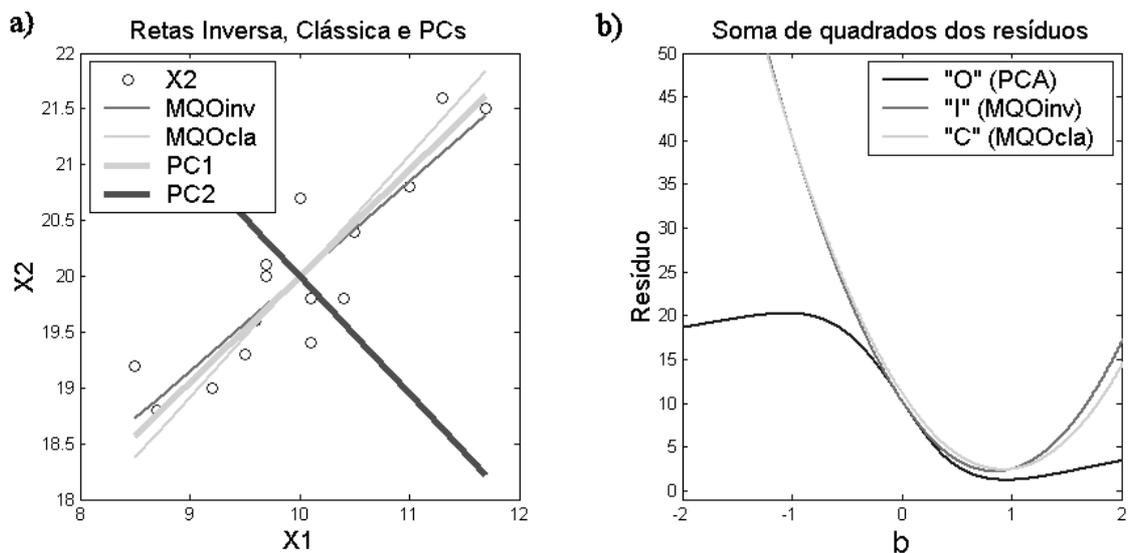


Figura A1.3 - a) Retas ajustadas por MQO inversa e clássica e por regressão ortogonal (PC’s). b) Comportamento da soma de quadrados dos resíduos das regressões de a) em função do valor do parâmetro “b”.

A regressão ortogonal fornece o intercepto da reta PC1 quando se assume para o modelo $\hat{X}_2 = \hat{a} + \hat{b}X_1$, o valor de “b” que minimiza os desvios ortogonais. Para PC2, por sua vez, o intercepto é obtido quando se emprega o valor de “b” que maximiza os desvios ortogonais ($b_{PC2} = -1/b_{PC1}$, já que PC2 é perpendicular a PC1).

Uma vez conhecidas as direções dos eixos PC1 e PC2, pode-se então calcular os novos valores sobre os quais serão ajustados modelos para prever Y. Tais valores são chamados “scores” dos componentes principais e são obtidos a partir das projeções

ortogonais dos pontos de X2 sobre os eixos PC1 e PC2 como mostrado na Figura A1.4. Note-se na figura que a regressão ortogonal, ao mesmo tempo que minimiza a soma dos quadrados dos desvios “O”, maximiza a amplitude das projeções ortogonais dos pontos sobre o eixo PC1. Isto equivale a se atribuir a maior variância possível aos valores dos *scores* de PC1. A amplitude das projeções ortogonais dos pontos sobre PC2, por sua vez, é minimizada. Como consequência, os *scores* de PC2 apresentarão a menor variância.

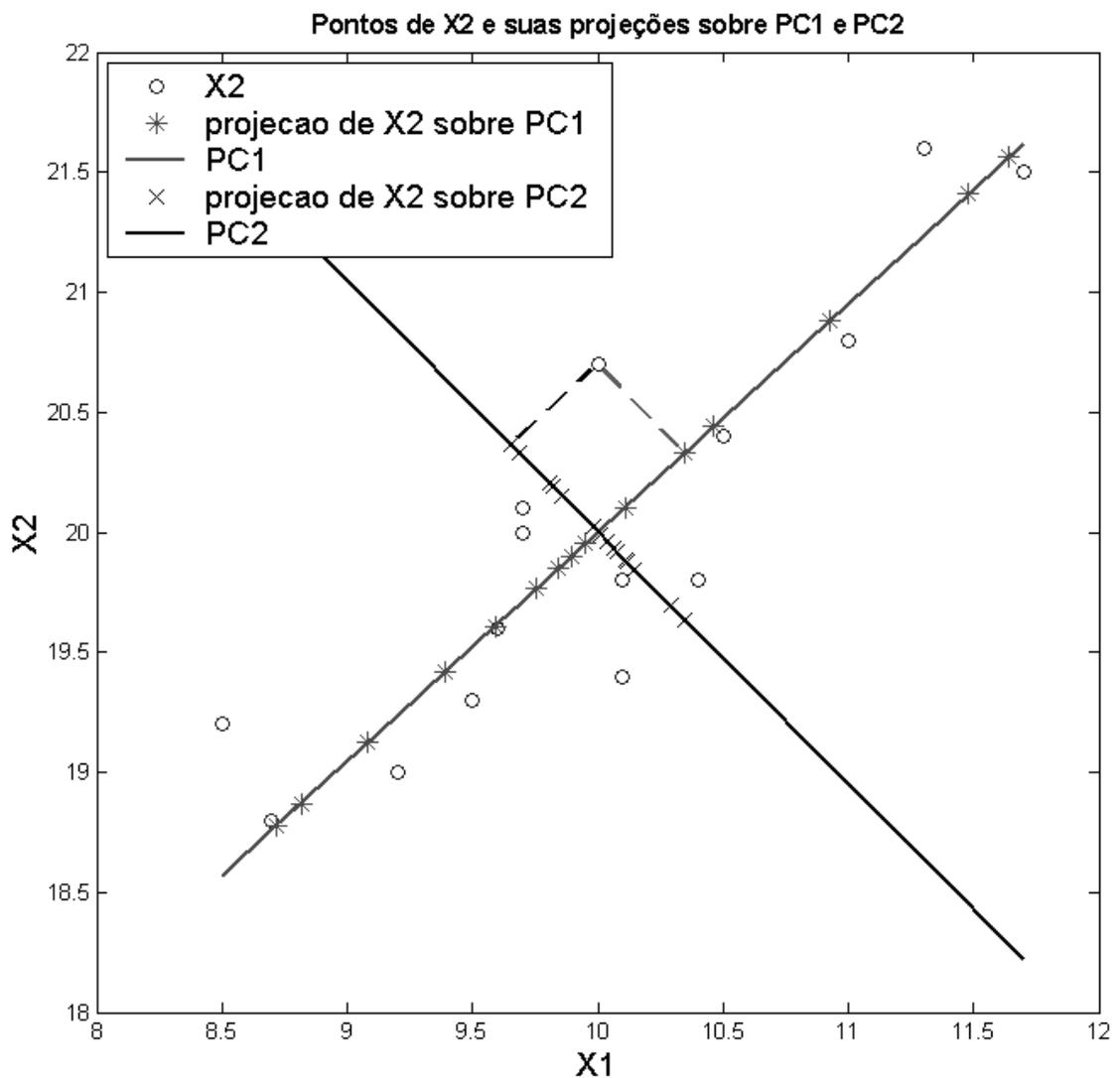


Figura A1.4 - Obtenção dos *scores* de PC1 e PC2 através das projeções de X2 sobre os eixos PC's.

Uma vez conhecidas as coordenadas das projeções, são calculadas as distâncias entre esses pontos ao longo dos eixos PC's. Atribui-se o valor zero para o primeiro

ponto e os demais serão obtidos somando cumulativamente as distâncias subsequentes. Posteriormente, dos valores assim obtidos, são subtraídas as suas médias. Caso as variáveis originais tenham sido padronizadas para possuírem média igual a zero e variância igual a 1, divide-se ainda o resultado da subtração pelo desvio-padrão dos *scores*. Desta maneira são obtidos os eixos PC1 e PC2 ortogonais e com a origem nas médias dos respectivos valores de seus *scores*. Com isto está completa a tarefa de mudança de base dos dados. A Figura A1.5. apresenta gráficos que evidenciam alguns dos resultados importantes da transformação por PCA. Ao contrário da situação original em que X1 e X2 eram altamente correlacionadas (Figura A1.5-a), as novas variáveis independentes PC1 e PC2 são descorrelacionadas (Figura A1.5-d). Conseqüentemente, o erro (Sb_1) associado ao coeficiente b_1 (coeficiente de X1 e PC1) foi reduzido.

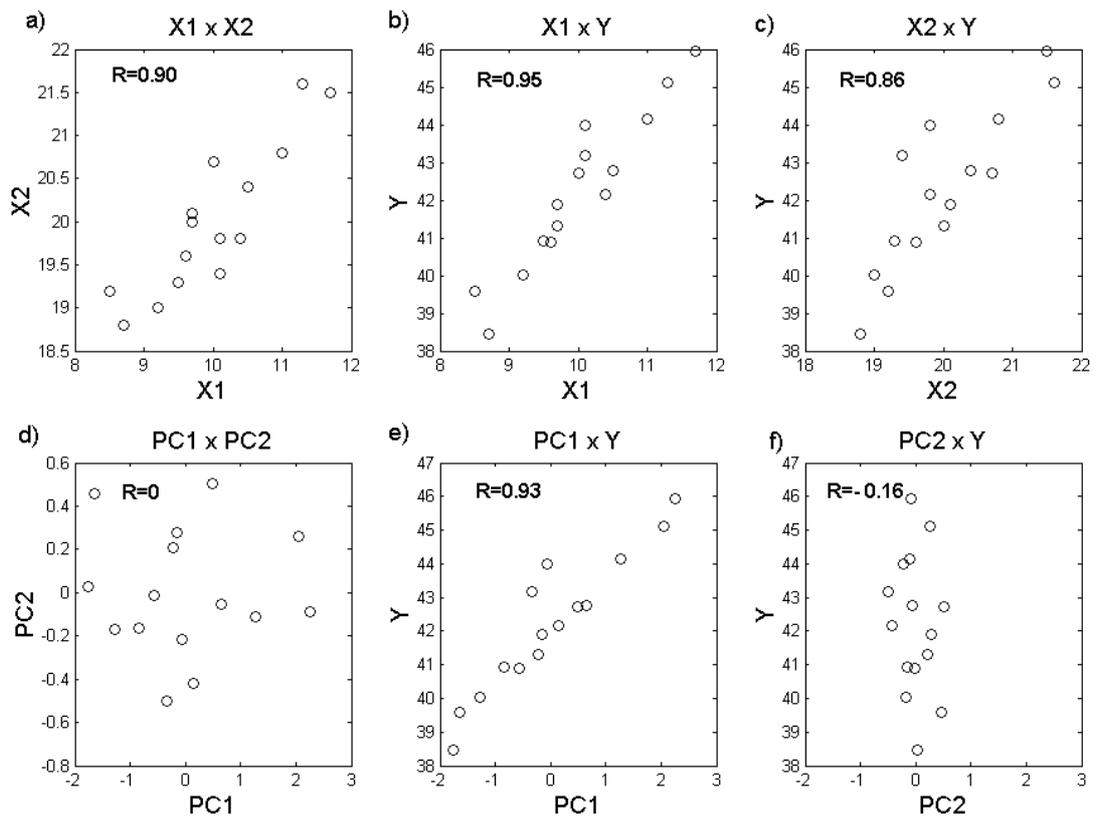


Figura A1.5 - Relações entre as variáveis originais e transformadas por PCA, entre si e com a variável dependente.

Como comentado anteriormente, o presente conjunto de dados apresentado como exemplo não é capaz de revelar com clareza as potencialidades da PCR. O valor de Sb_1 , por exemplo, seria afetado com maior intensidade caso X1 e X2 apresentassem

um coeficiente de correlação mais alto entre si. O pequeno número de variáveis independentes do exemplo também não permite que seja vislumbrada a capacidade de compressão das variáveis apresentada pela PCA.

A.1.2. Procedimento matricial para o cálculo dos PC's

O procedimento acima descrito é útil para a compreensão geométrica da mudança e base dos dados. Porém, na prática se tornaria muito laborioso a sua aplicação em problemas que envolvam um grande número de variáveis independentes. Neste caso lança-se mão da técnica matricial para a determinação das direções dos eixos dos PC's e cálculo dos *scores*. Tal técnica baseia-se na determinação dos autovetores e autovalores da matriz de covariância das variáveis independentes. O mesmo poderá ser aplicado a matriz de correlação que será idêntica à matriz de covariância caso as variáveis estejam padronizadas com média igual a zero e variância igual a um.

Assim, seja $S_{(p \times p)}$ a matriz de covariância de p variáveis independentes X sobre as quais se deseja realizar uma regressão para se obter um modelo que estime os valores de uma variável dependente Y . Da Álgebra Linear tem-se que: se S é quadrada, simétrica e não singular, existem as matrizes U e L , também quadradas e de dimensões $p \times p$, que tornam verdadeiras as expressões:

$$U^T S U = L \quad \text{ou} \quad S U = L U \quad (\text{Eq. 12})$$

em que:

$$S = \begin{vmatrix} s_{x_1}^2 & \text{cov}(x_2, x_1) & \text{cov}(x_p, x_1) \\ \text{cov}(x_1, x_2) & s_{x_2}^2 & \text{cov}(x_p, x_2) \\ \text{cov}(x_1, x_p) & \text{cov}(x_2, x_1) & s_{x_p}^2 \end{vmatrix}$$

$$U = \begin{vmatrix} u_{11} & u_{21} & u_{p1} \\ u_{12} & u_{22} & u_{p2} \\ u_{1p} & u_{2p} & u_{pp} \end{vmatrix}$$

$$L = \begin{vmatrix} l_{11} & 0 & 0 \\ 0 & l_{22} & 0 \\ 0 & 0 & l_{pp} \end{vmatrix}$$

A matriz L é uma matriz diagonal cujos elementos $l_{11}, l_{22}, \dots, l_{pp}$ são chamados autovalores (também são empregados os termos *valores latentes* e *raízes latentes*) da

matriz S. Tais valores podem ser agrupados em um vetor $l = [l_1, l_2 \dots l_p]$ de maneira a satisfazer:

$$l \cdot I = L \quad (\text{Eq. 13})$$

em que I é uma matriz identidade de dimensões $p \times p$. Os valores de $l_1, l_2 \dots l_p$ representam as variâncias dos *scores* de PC1, PC2 ... PCp, respectivamente, e estão arranjados em ordem decrescente. Os autovalores de S podem ser então calculados substituindo-se a Eq. 10 na Eq. 9 e resolvendo-se para l:

$$\begin{aligned} SU &= lIU \\ SU - lIU &= 0 \\ (S - lI)U &= 0 \end{aligned}$$

Supondo-se $U \neq 0$ obtém-se a chamada *equação característica*:

$$\det(S - lI) = 0 \quad (\text{Eq. 14})$$

Uma vez determinados os autovalores l_i , estes são aplicados na expressão

$$(S - l_i I)t_i = 0 \quad (\text{Eq. 15})$$

com $i = 1, 2, \dots, p$, para o cálculo dos vetores t_i que fornecerão, finalmente, os autovetores u_i através da transformação:

$$u_i = \frac{t_i}{\sqrt{t_i' t_i}} \quad (\text{Eq. 16})$$

No caso de apenas duas variáveis X, a matriz dos autovetores fica:

$$U = \begin{vmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{vmatrix}$$

Os autovetores devem satisfazer as condições de ortonormalidade, ou seja:

$$u_1' u_1 = u_2' u_2 = 1 \quad \text{e} \quad u_1' u_2 = 0$$

O valor de u_{11} equivale ao coseno do ângulo de inclinação do eixo PC1 em relação ao eixo original X1, enquanto que u_{21} é o coseno do ângulo entre PC2 e X2. A Figura A1.6. apresenta as relações entre os valores dos elementos de U e os ângulos formados entre os PC's e os eixos originais.

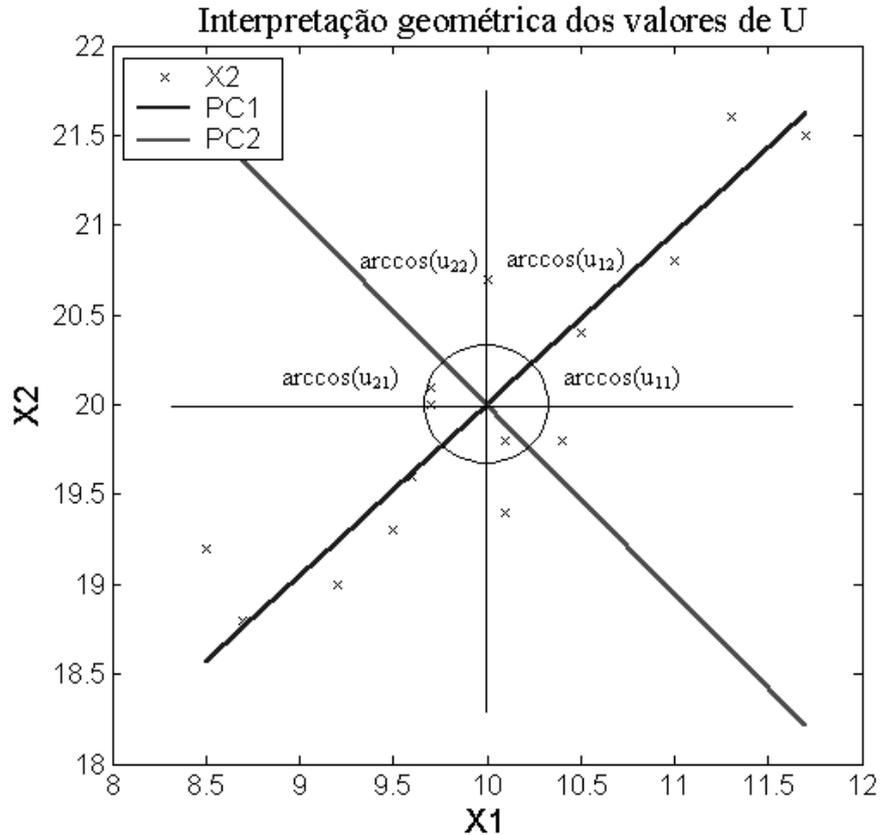


Figura A1.6 - Relações entre os valores dos elementos da matriz dos autovetores e os ângulos formados entre os eixos originais e os PC's.

Finalmente, a partir dos autovetores $U = [u_1 \ u_2 \ \dots \ u_p]$ e dos valores originais das variáveis $X = [X_1 \ X_2 \ \dots \ X_p]$ pode ser calculada a matriz Z dos *scores* através do produto:

$$Z = [X - \bar{X}] \cdot U \quad (\text{Eq. 17})$$

A matriz Z possui as mesmas dimensões de X e os elementos de suas colunas correspondem aos valores das novas variáveis PC's sobre as quais será realizada a regressão para obter-se o estimador de Y . A coluna Z_1 (PC1) apresenta maior variância que Z_2 (PC2) que, por sua vez, possui maior variância que Z_3 (PC3) e assim sucessivamente até o último componente Z_p (PCp).

Uma vez de posse dos valores dos scores dos PC's, pode-se finalmente ajustar os modelos pelo método dos mínimos quadrados ordinários para estimar Y . As regressões realizadas com os scores podem ser submetidas à análise de variância, cômputo dos erros, testes de significância e intervalos de confiança como normalmente se faria com outros tipos de variáveis aleatórias. Para se calcular o valor de Y a partir dos valores de X fornecidos por uma determinada amostra, basta converter os valores de X em *scores* dos PC's através da Eq.17 e aplicar os valores transformados no modelo.

A Figura A1.7. apresenta a dispersão dos valores preditos por modelos ajustados por MQO antes (funções de X) e depois (funções dos PC's) da realização da transformação de bases promovida pela PCA.

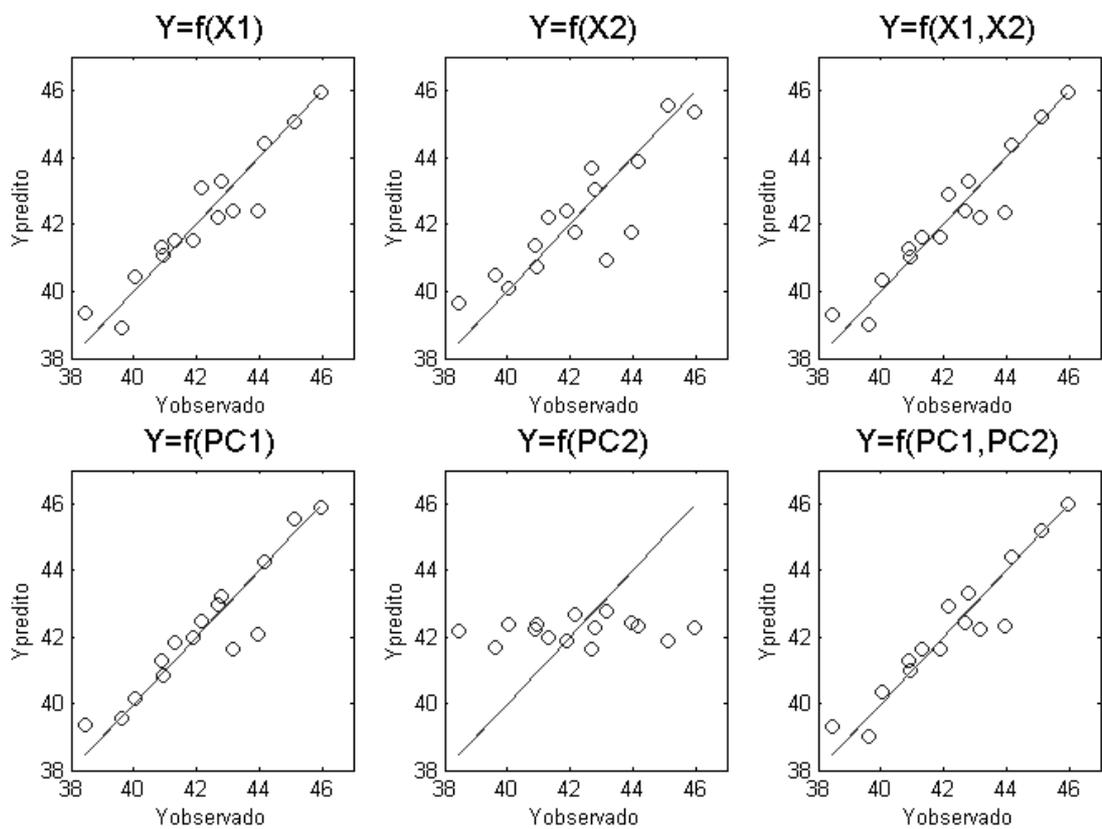


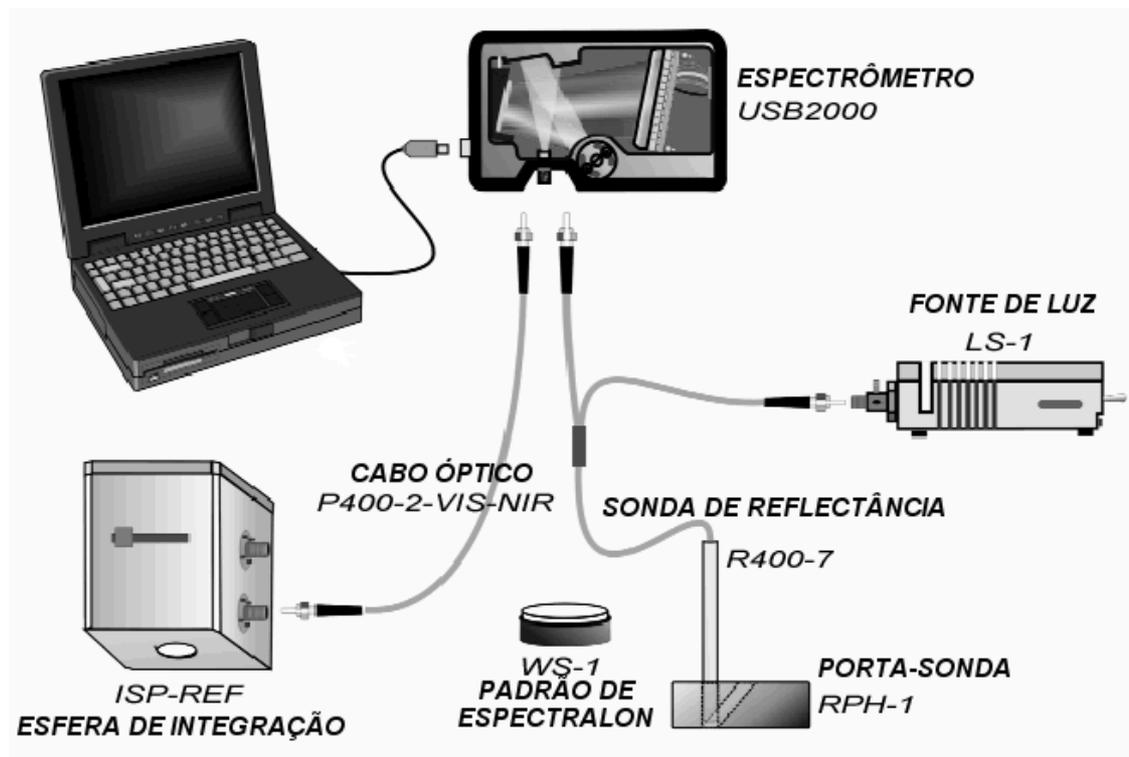
Figura A1.7 - Valores de Y observados e preditos pelos modelos ajustados por MQO antes e depois da PCA.

Como pode ser observado na Figura A1.7, ao se utilizar todos os PC's em uma PCR obtêm-se os mesmos resultados que são obtidos por MQO em termos de valores preditos pelo modelo. A confiabilidade do modelo fornecido pela PCR, porém, é maior já que os erros associados aos coeficientes da regressão são menores.

A.2. O equipamento e sua operação

A.2.1. O sistema óptico

O sistema óptico utilizado neste trabalho consiste de dois espectrômetros USB2000 da marca OceanOptics, de fabricação norte-americana com os acessórios: esfera integradora e cabo óptico para a sua conexão com o espectrômetro; sonda de reflectância e fonte de luz conectadas entre si e com o espectrômetro através de um cabo óptico que conduz a luz da fonte à sonda e da sonda, após ser refletida pela madeira, para o espectrômetro. A esfera e a sonda possuem funções e performances equivalentes. Integram ainda o sistema, o porta-sonda, o padrão de spectralon e o microcomputador. A Figura A2.1. apresenta um esquema geral da montagem do sistema.



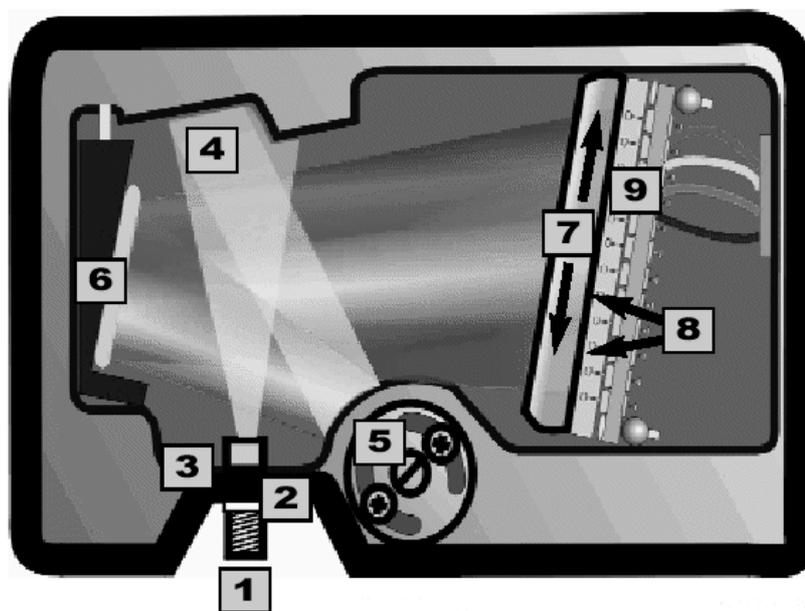
Fonte: Modificado de OceanOptics (2005).

Figura A2.1 - Esquema geral do sistema óptico utilizado.

A.2.2. Os espectrômetros

O espectrômetro USB2000 é fruto de um conceito de equipamentos portáteis e de baixo custo produzidos para operar em condições de campo ou de laboratório que requeiram mobilidade. Ou seja, o aparelho vai até a amostra e não o contrário, como se observa nos equipamentos de bancada. Tal conceito também possibilita o trabalho com amostras de formatos e dimensões diversos, o que amplia as possibilidades de uso e economiza o tempo que seria despendido com a preparação de amostras.

A construção do equipamento é muito simples, apesar de utilizar alta tecnologia em sua mecânica, óptica e eletrônica. O aparelho não apresenta partes móveis. A Figura A2.2. apresenta a anatomia do espectrômetro USB2000.



Fonte: OceanOptics (2005).

Figura A2.2 - Anatomia do espectrômetro USB2000. Vide texto para a descrição das partes.

Assim, na Figura A2.2. as partes numeradas representam:

Conector para a acoplagem da fibra óptica que traz o sinal luminoso desde a amostra.

Logo após a entrada do sinal encontram-se filtros para a eliminação de efeitos de segunda e terceira ordem produzidos no cabo. Deve estar de acordo com a configuração da grade de difração (5).

Fenda de entrada da luz. Deve estar de acordo com a faixa espectral do sinal a ser analisado.

Espelho colimador para tornar paralelos os raios de luz.

Grade de difração. Esta é uma das partes mais importantes do equipamento. Aqui um sinal que contém vários comprimentos de onda sobrepostos é decomposto revelando o seu chamado espectro ou assinatura espectral. O ângulo no qual é posicionada a grade em relação aos outros componentes é de crucial importância para o funcionamento do equipamento. O USB2000 utiliza uma grade de difração holográfica que representa um avanço na tecnologia de produção de grades.

Espelho de focalização, cuja função é espalhar o feixe que emerge da grade projetando a radiação de cada comprimento de onda em seu devido lugar no detector (9).

Lente do detector, cuja função é concentrar a radiação sobre os diminutos elementos sensíveis do detector.

Detector CCD. Trata-se de um semicondutor fotossensível subdividido em 2048 seções (pixels) com dimensões de 14 x 200 μm . Cada pixel recebe a radiação correspondente a um comprimento de onda em intervalos de 0,3 nm. A sensibilidade do detector está entre 200 e 1100 nm, porém a faixa útil para trabalho encontra-se entre 350 e 1000 nm. O CCD está conectado a um chip que estabelece a comunicação com o microcomputador que registra os espectros segundo o desejo do usuário. Na Figura A2.3. pode-se observar o aspecto físico do detector CCD com sua lente coletora e o *chip* a ele acoplado.

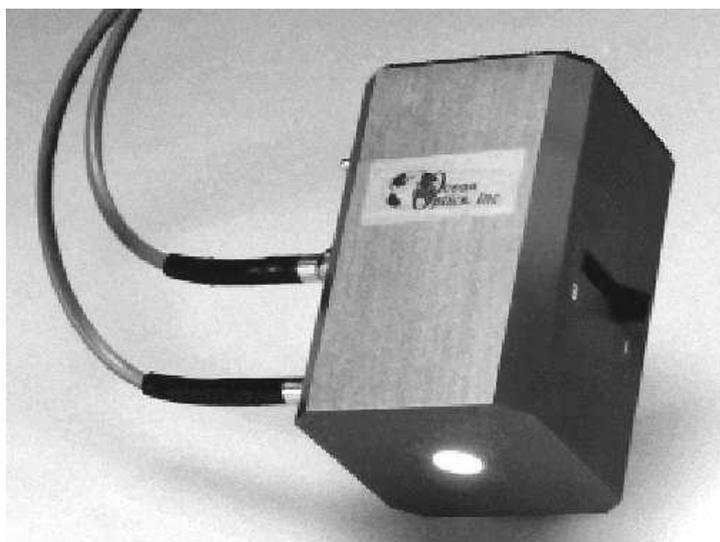


Fonte: OceanOptics (2005).

Figura A2.3 - Fotografia do detector Sony CCD ILX511 utilizado no espectrômetro USB2000.

A.2.3. A esfera integradora e a sonda de reflectância

A esfera integradora é um dispositivo utilizado para a determinação da reflectância da superfície de materiais sólidos. Consiste de uma câmara esférica cujas paredes são revestidas por um material de alta reflectância (sulfato de bário). Existem ainda três orifícios nessa câmara. Um deles, onde são posicionadas as amostras, se abre para o exterior. O segundo dá entrada à luz que irá interagir com a amostra. E o terceiro está conectado ao espectrômetro através de um cabo óptico. O segundo cabo que se vê conectado à esfera na Figura A2.4. destina-se à entrada da luz de uma fonte externa. Há ainda um dispositivo interno acionado por uma chave que permite a captação da reflectância especular além da difusa.



Fonte: OceanOptics (2005).

Figura A2.4 - Esfera integradora modelo ISP-REF da OceanOptics.

Como mostra a Figura A2.5. ao incidir sobre uma superfície (I), a luz é refletida, parte de forma especular (E - com um ângulo de reflexão α igual ao de incidência) e parte de forma difusa (D - em todas as direções). Ainda, uma terceira parte é absorvida (A) e uma quarta pode atravessar o material (T). Considerando que existe uma seletividade espectral nas interações, o conhecimento de cada uma delas pode fornecer diferentes tipos de informação acerca do material que foi irradiado.

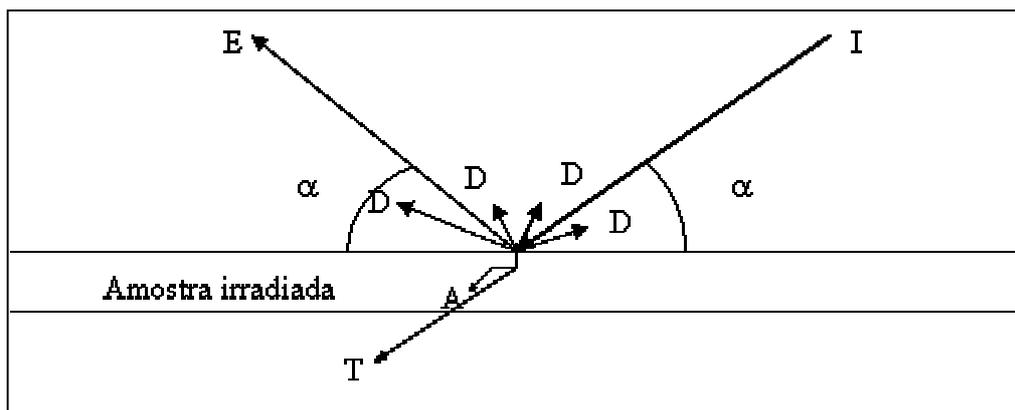


Figura A2.5 - Possíveis interações da radiação com uma superfície sólida.

Normalmente, em espectroscopia aplicada à propósitos quimiométricos e na colorimetria quantitativa, utilizam-se os espectros de transmitância e de reflectância difusa por estes conterem informações derivadas das interações nos níveis moleculares e eletrônicos. A radiação refletida de forma especular, por sua vez, é pobre em informações deste tipo e os espectros assim obtidos são muito semelhante àqueles da própria radiação incidente. A intensidade da reflexão especular, porém, está intimamente relacionada à textura da superfície do material. Assim, em investigações acerca de parâmetros físicos que são determinantes da textura, como no caso das dimensões de elementos anatômicos da madeira, a reflectância especular pode se apresentar como muito importante.

Tanto a esfera integradora quanto a sonda de reflectância podem captar, ou apenas a reflectância difusa, ou a combinação das duas. A esfera apresenta a vantagem de amostrar uma área maior que aquela possibilitada pela sonda e possuir iluminação própria. Por outro lado, a sonda possibilita maior mobilidade durante o trabalho e aplicações sobre pós e líquidos.

A.2.4. A aquisição dos espectros

A configuração dos parâmetros de aquisição dos espectros é de suma importância para o sucesso de qualquer trabalho em espectroscopia. O espectrômetro USB2000 é controlado pelo *software* OOIBASE 32, que controla também outros instrumentos do mesmo fabricante.

A aquisição dos espectros com o uso da esfera integradora ou com a sonda de reflectância devem seguir os mesmos procedimentos. Inicialmente, devem ser estabelecidos os parâmetros de amostragem, o que é feito através de caixas de diálogo de configuração do *software*. Tais parâmetros são:

Tempo de integração (box: *Integration time*). Este parâmetro significa por quanto tempo o detector deve estar ativo para obter o espectro. Ou seja, qual é a quantidade de energia que ele deve receber, à semelhança do tempo de exposição em uma câmera fotográfica. O tempo de integração pode variar de 1 a 65.000 milisegundos. Tempos de integração mais altos são indicados para amostras que apresentam baixa reflectância em toda a faixa de trabalho do aparelho, como é o caso do carvão vegetal. Ao aumentar-se o tempo de exposição do detector aumenta-se também a relação sinal/ruído. O que é positivo para a exatidão dos espectros obtidos. Porém, se o tempo é muito elevado, corre-se o risco de levar o sensor à saturação no momento de se tomar a linha de base, que é o espectro do padrão de spectralon. Assim, o limite superior deste parâmetro é determinado pela sensibilidade do detector.

Número de tomadas (Box: *Average*). Pode variar de 1 a 10.000. Significa o número de espectros a serem tomados para compor um espectro médio. Esta função também é útil na ampliação da relação sinal/ruído. Pois, considerando-se que os ruídos são oscilações aleatórias em torno do valor real da reflectância em um determinado comprimento de onda, é de se supor que sua média populacional seja igual a zero. Assim, aumentando-se o número de espectros a serem tomados de uma mesma amostra e, posteriormente, calculando-se a sua média, pode-se obter uma representação mais exata do espectro real. A ampliação excessiva do número de tomadas tem como consequência apenas o prolongamento do tempo para a obtenção dos dados. Tal prática pode ser uma solução no caso de não se poder elevar o tempo de integração.

Amplitude do filtro de média-móvel (Box: *Boxcar*). Pode variar de 1 a 500. O filtro de média-móvel consiste em se atribuir à reflectância em um ponto da curva espectral valores que são médias dos valores de pontos adjacentes. Trata-se também de mais um artifício para a eliminação de ruídos. No caso do espectrômetro USB2000, que apresenta uma resolução espectral de 0,3 nm, o uso moderado do filtro de média-móvel é aceitável. Porém, amplitudes muito grandes podem deformar completamente a curva

espectral, anulando qualquer possibilidade de se obter informações úteis a partir dos dados.

Correção do ruído elétrico (Box: *Correct for electrical noise*). Os primeiros 24 pixels do detector não respondem à luz, mas produzem um sinal elétrico que se soma aos outros ruídos aleatórios presentes no sistema. Acionando a correção do ruído elétrico, a média da tensão existente nos primeiros 24 pixels é subtraída em todos os demais. Assim, é interessante que esta filtragem esteja sempre acionada.

Uma vez determinados os parâmetros de aquisição, pode-se dar início à calibração do instrumento. A calibração consiste em se tomar os espectros do padrão de spectralon, que é um polímero sintético de alta reflectância em toda a extensão da faixa de trabalho do instrumento; e da referência escura, que é representada pelo espectro de um objeto escuro e fosco com a luz do instrumento desligada. O espectro da referência escura deve ser subtraído dos espectros das amostras obtidos com a luz acesa. O resultado desta subtração dividido pelo espectro do spectralon e multiplicado por 100 fornece o espectro da amostra.

A calibração é realizada no modo “Scope” (menu Spectrum=>Scope Mode), que é ativado quando o programa de controle do aparelho é iniciado. Sempre que, por qualquer motivo, o programa for reiniciado, se faz necessária a abertura dos arquivos onde foram gravados os espectros da referência clara e da escura. Assim, para realizar a calibração do instrumento e registrar espectros de amostras sob estudo, o usuário deve certificar-se de que o modo Scope do programa de controle está ativado e, em seguida, adotar o seguinte procedimento:

Ligar a luz da esfera ou da fonte de luz da sonda;

Posicionar o padrão de spectralon bem junto ao orifício de amostragem da esfera ou da extremidade da sonda devidamente encaixada no porta-sonda;

Aguardar por alguns segundos para que o sinal se estabilize e armazenar a referência clara (menu Spectrum=>Store Reference); gravar a referência clara para posterior utilização (menu File=>Save=>Reference); a referência clara é chamada linha de base (*baseline*) por alguns autores;

Desligar a luz da esfera ou da fonte da sonda, posicionar para amostragem o objeto escuro e fosco (um pedaço de borracha preta é uma boa opção) e armazenar a referência escura (menu Spectrum=>Store Dark); gravar a referência escura para posterior utilização (menu File=>Save=>Dark);

Ativar a opção de subtração da referência escura (menu Spectrum=>Scope Mode Minus Dark);

Abandonar o modo Scope, ativando o modo Transmission (menu Spectrum=>Transmission Mode);

Acender a luz da esfera ou da fonte da sonda;

Posicionar uma amostra do material a ser estudado no orifício de amostragem da esfera ou diante da extremidade da sonda;

Aguardar a estabilização do sinal e gravar o espectro (menu File=>Save=>Processed). Este procedimento armazena os dados em um arquivo *.*sample* que pode ser lido pelo OOIbase32 ou por editores de texto. Caso se queira analisar os dados em outro aplicativo, basta copiá-los para a área de transferência (menu Edit=>Copy Spectral Data=>All Spectrometer Channels) e colá-los no outro aplicativo (Ctrl+V). Caso se queira apenas exportar o gráfico, basta copiá-lo para a área de transferência (menu Edit=>Copy Graphical Spectra) e colá-lo em seu destino (um editor de textos, um editor de imagens, etc).

Pode-se notar pelo exposta acima que a operação do espectrômetro é bastante simples, porém, para que haja consistência nos dados de uma série de espectros, é de suma importância manter-se a padronização dos procedimentos. Especialmente os registros de abertura dos arquivos das referências claras e escuras devem receber atenção durante o trabalho, pois em espectroscopia os resultados são sempre expressos de maneira relativa à tais padrões. Ou seja, um conjunto de espectros para terem validade em análises qualitativas ou quantitativas, devem ter sido obtidos sempre com o mesmo padrão.

A.3. O algoritmo computacional

A seguir, encontra-se transcrito em sua íntegra, o algoritmo computacional construído sobre a plataforma MATLAB[®] para a realização dos cálculos e confecção dos gráficos apresentados no decorrer deste trabalho. É importante notar que é necessário um razoável conhecimento prévio da linguagem de programação MATLAB para a compreensão e utilização deste algoritmo. Especial atenção deve ser dada ao momento de entrada de dados.

Para que o algoritmo funcione corretamente basta copiar literalmente o código transcrito a seguir para um arquivo de script do MATLAB (*.m ou *m-file*) ou diretamente para a janela de comando.

O símbolo% indica uma linha de comentário ou comandos que podem ser ora utilizados, ora não utilizados. Por exemplo, os comandos relativos à confecção de alguns gráficos são anulados quando se deseja somente a obtenção dos resultados finais, a saber, as análises de variância e os erros dos modelos.

```
% ALGORITMO PARA O AJUSTE DE MODELOS DE CALIBRAÇÃO
% UTILIZANDO A REGRESSAO POR COMPONENTES PRINCIPAIS COM OS
% COMPONENTES
% SENDO ADICIONADOS PELA ORDEM DECRESCENTE DE VARIANCIA OU
% PELA MAIOR
% CORRELAÇÃO COM OS RESIDUOS (METODO FORWARD)

% CHRISTOVAO PEREIRA ABRAHAO
% UNIVERSIDADE FEDERAL DE VICOSA

clear
tic

for met=2:2% fecha no fim do loop dos metodos de regressao
    max_trans=1;
    max_transXD=1;
    for test=1:1% fecha no fim do loop dos testes
        RESUMOmin=ones(18,6,3);
        RESUMOmed=ones(18,6,3);
        RESUMOmax=ones(18,6,3);
        for cnttrans=1:max_trans% fecha no fim do loop das trnsformações
            if cnttrans==1; trans_X=0; end% reflectancia
            if cnttrans==2; trans_X=3; end% absorbancia
            if cnttrans==3; trans_X=4; end% kubelka-munk
            if cnttrans==4; trans_X=5; end% raiz quadrada da reflectancia
            if cnttrans==5; trans_X=6; end% dividir pela variancia
            if cnttrans==6; trans_X=7; end% Box-Cox

            for cntderiv=1:max_transXD% fecha no fim do loop das derivada
                if cntderiv==1; trans_XD=0; end% sem derivar
                if cntderiv==2; trans_XD=1; end% derivada primeira
                if cntderiv==3; trans_XD=2; end% derivada segunda

                load totalmed_ppfisar1
                load totalmed_ppfisar2
                load uv_especmed
                load uv_difmed
```

```

% carregando o vetor dos comprimentos de onda
load CVout1
load WL
load WL_uv

% RODADAS Número de vezes que o programa rodara tomando
aleatoriamente amostras para a calibração e validacao

nrodadas=100;
%%%%%%%%%%%%% DEFININDO CONDICOES DE
ANALISE %%%%%%%%%%%%%%

local=0;% se=0 => mistura L1 e L2, se=1=> Local 1, se=2=> Local 2

%DEFINIDO O METODO DE ENTRADA DAS VARIAVEIS (PC's) NOS
MODELOS
% metodo_regr=0;% 0=pela ordem (variancia dos PCs); 1=pela avaliacao da
correlação com o residuo (FORWARD)
if met==1;metodo_regr=0; end
if met==2;metodo_regr=1; end
% ==>=>=>=>=> PELA ORDEM
if metodo_regr==0; entrada='PO';end
% ==>=>=>=>=> FORWARD
if metodo_regr==1; entrada='FW';end

centrar_na_media=0;% 0=nao 1=sim

%TRANSFORMACOES SOBRE AS VARIAVEIS X
% trans_X=0 ;
%% 0=> sem transformar 3=> Absorbancia; 4=> Kubelka;
%% 5=> raiz quadrada;6=> dividir pela varia^ancia 7=> Box-Cox
%
% trans_XD=0 ;
% %0=>sem derivar;1=>derivada1;2=>derivada2

% correção da dispersao espectroscopica
SC=0;% 0=> sem corecao; 1=> MSC; 2=> PMSC
WINDOW=20;% Número de janelas para PMSC
window=(WINDOW/0.31)/2;

%%%%%%%%%%%%% LARGURA DO FILTRO DE MEDIAS - SMOOTH"
S=0;% DEVE SER IMPAR; se S=0 ou S=1 => nao suavizar;
if S==1 | S==0
s=0;
end
if S>0 & ((S/2)-fix(S/2))==0;
% beep
warning('S deve ser impar e foi tomado como S+1');

```

```

        S=S+1;
    end
    if S>0
        s=(S-1)/2;% METADE DA LARGURA MENOS 1 DO FILTRO DE
MEDIAS - SMOOTH
    end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% FILTRAGEM POLINOMIAL POR
SAVITZKY-GOLAY

% for SGgrau=1:3
%   cont=0
%   for SGframe=7:2:15

        SGframe = 0;% DEVE SER IMPAR; se S>1 => A FILTRAGEM 'E FEITA
PELO FILTRO DE MEDIA MOVEL;
        if SGframe>0 & ((SGframe/2)-fix(SGframe/2))==0
            %   beep
            warning('S deve ser impar e foi tomado como SGframe+1');
            SGframe=SGframe+1;
        end
        SGgrau=3;% DEVE SER MENOR QUE SGframe

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% AMPLITUDE (em nm) DA JANELA DE MEDIAS
PARA CALCULO DOS PONTOS X

        RESOLUCAO=5;% para RESOLUCAO=0 =>m=0 => nao calcula medias;
        m=fix(RESOLUCAO/0.31) ;%; m=0 => nao calcula medias;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%PERCENTUAL DAS AMOSTRAS QUE SERA
UTILIZADO NA VALIDACAO
        VAL=33 ;% percentual das amostras a serem utilizadas para validacao

        azul_uv=190 ;%mais baixo lambda a ser computado
        vermelho_uv=400 ;%mais alto lambda a ser computado

        azul=401 ;%mais baixo lambda a ser computado
        vermelho=1000;%mais alto lambda a ser computado

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% DEFININDO Y
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        %% definindo a variavel dependente a ser modelada:
        % 88747022 marln
        % y_cal=y_dados(sel_cal,dependente);

```

```

%%definindo os valores da variavel dependente para validar o modelo
% y_val=y_dados(sel_val,dependente);

dependente=10;
if dependente==9
    dep='densar';
end
if dependente==10
    dep='densbas';
end
if dependente==11;
    dep='dsec'
end
% clear CondAnalise
% CondAnalise=char([dep,' centrar na media= ',num2str(centrar_na_media),...
% ' trans X= ',num2str(trans_X),' S=num2str(S),' RESOLUCAO=
,num2str(RESOLUCAO),...
% ' VAL= ',num2str(VAL),'%',' azul= ',num2str(azul),' vermelho=
,num2str(vermelho)]);
clear str1
str1(1)={'centrar na media=',num2str(centrar_na_media)};
str1(2)={'trans X=',num2str(trans_X)};
str1(3)={'S=',num2str(S)};
str1(4)={'RESOLUCAO=',num2str(RESOLUCAO),'nm'};
str1(5)={'VAL=',num2str(VAL),'%'};
str1(6)={'SGframe=',num2str(SGframe),' SGgrau=',num2str(SGgrau)};
str1(7)={'GAMA=',num2str(azul_uv),':',num2str(vermelho_uv),'nm
,num2str(azul),':',num2str(vermelho),'nm'};

% str1(6)={'azul=',num2str(azul),'nm'};

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% azul=401 ;%mais baixo lambda a ser computado
% vermelho=1000;%PARA O UV-VIS
limite_inferior=azul;
limite_superior=vermelho;
[linf ljinf]= find(wl>limite_inferior);
[lsup ljsup]= find(wl>limite_superior);
liminf=linf(1,1);
limisup=lsup(1,1);
WL=wl(liminf:limisup,:);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%PARA UV%%%%%%%%
limite_inferior_uv=azul_uv;

```



```

x_dados=x_dados;
[Lx_dados Cx_dados]=size(x_dados);

[Lyps Cyps]=size(y_dados_dep);

clear matXY_bruta wl wl_uv bruta_transp

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
%%          %%%%%%%%%%-----%%%%%%%%%
$$$ $$$ $$$ $$$ $$$ $$$ $$$ $$$ %%%%%%%%%%
%%          %%%## # ## ### ### ##### INICIO DO PROCESSAMENTO # ### #####
### #####%%%%%%%%%%
%%          %%%%%%%%%%-----%%%%%%%%%
$$$ $$$ $$$ $$$ $$$ $$$ $$$ $$$ %%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%          %%%%%%%%%%-----%%%%%%%%%

% selecionando as linhas (amostras) que comporao os conjuntos de calibracao
e validacao
% clear rand_y_cal Rod
% tic toc
cal=Cx_dados-fix(Cx_dados/(100/VAL))+1;
val=fix(Cx_dados/(100/VAL))+1;

rand_y_cal=-100*ones(cal,nrodadas);
rand_y_val=-100*ones(cal,nrodadas);
rand_ym_cal=-100*ones(cal,cal,nrodadas);
rand_ym_val=-100*ones(val,val,nrodadas);
rand_betas=-100*ones(cal,cal,nrodadas);
rand_desviocal=-100*ones(cal,cal,nrodadas);
rand_desvioval=-100*ones(val,val,nrodadas);
% save RANDS rand_ym_cal rand_y_cal rand_ym_val rand_y_val rand_betas
rand_desviocal rand_desvioval
% clear rand_ym_cal rand_y_cal rand_ym_val rand_y_val rand_betas
rand_desviocal rand_desvioval
% end
%
% if metodo_regr==1
% rand_y_cal_fw=-100*ones(cal,nrodadas);
% rand_y_val_fw=-100*ones(cal,nrodadas);
% rand_ym_cal_fw=-100*ones(cal,cal,nrodadas);
% rand_ym_val_fw=-100*ones(val,val,nrodadas);
% rand_betas_fw=-100*ones(cal,cal,nrodadas);
% rand_desviocal_fw=-100*ones(cal,cal,nrodadas);
% rand_desvioval_fw=-100*ones(val,val,nrodadas);
% end

```

```

for Rod=1:nrodadas
    per=randperm(Cx_dados);
    [Lp Cp]=size(per);
    sel_cal=per(1:fix(Cp-(Cp/(100/VAL))));
    [Lselcal Cselcal]=size(sel_cal);
    sel_val=per(Cselcal+1:Cp);
    [Lselval Cselval]=size(sel_val);
    if Cx_dados-Cselcal<Cselval
        sel_val=per(Cselcal+1:Cp);
        [Lselval Cselval]=size(sel_val);
    end
    y_cal=y_dados_dep(sel_cal,1);
    y_val=y_dados_dep(sel_val,1);
    mesel=[mean(sel_cal),mean(sel_val)];
    while abs(mesel(1,1)-mesel(1,2))>2 |min(y_val)<min(y_cal) |...
        max(y_val)>max(y_cal) | abs(mean(y_val)-
mean(y_cal))>((mean(y_dados(:,1)))/100
        per=randperm(Cx_dados);
        [Lp Cp]=size(per);
        sel_cal=per(1:fix(Cp-Cp/(100/VAL))));
        [Lselcal Cselcal]=size(sel_cal);
        sel_val=per(Cselcal+1:Cp);
        [Lselval Cselval]=size(sel_val);
        if Cx_dados-Cselcal<Cselval
            sel_val=per(Cselcal+1:Cp);
            [Lselval Cselval]=size(sel_val);
        end
        mesel=[mean(sel_cal),mean(sel_val)];
        y_cal=y_dados_dep(sel_cal,1);
        y_val=y_dados_dep(sel_val,1);
    end

    randsel_cal(Rod,:)=sel_cal;
    randsel_val(Rod,:)=sel_val;
    [Lycal Cycal]=size(y_cal);
    [Lyval Cyval]=size(y_val);

    %%%%%%%%%%
ESCOLHENDO UMA RODADA  # % # % # % # % # % # % # % # % # % # %
# % # % # % # % # % # %
    if nrodadas==1
        load sel_cal
        load sel_val
    end
    y_cal=y_dados_dep(sel_cal,1);
    y_val=y_dados_dep(sel_val,1);
    %%%%%%%%%sel_cal=randsel_cal(155,:);
    %%%%%%%%%sel_val=randsel_val(155,:);

```



```

        X_cal=X_cal.^0.5;
        X_val=X_val.^0.5;
    end
    % dividir pela vari^ancia
    if trans_X==6
        str4(1)={'Reflec./Variância'};
        for irq=1:Cx_cal
            X_cal(:,irq)=X_cal(:,irq)/var(X_cal(:,irq));
        end
        for irqv=1:Cx_val
            X_val(:,irqv)=X_val(:,irqv)/var(X_val(:,irqv));
        end
    end
    [Lx_cal Cx_cal]=size(X_cal);
    [Lx_val Cx_val]=size(X_val);

    %*****
*****

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
MEDIAS%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%suavizando as curvas espectrais com um filtro de média móvel de
amplitude=2s+1
% for s=10:10:50;
% X_cals=ones(Lx_cal-2*s,Cx_cal); X_vals=ones(Lx_val-2*s,Cx_val);
% WLs=ones(Lx_cal-2*s,1); is=ones(1);
clear X_cals X_vals WLs
if s==0%=> nao suavizar
    X_cals=X_cal;
    X_vals=X_val;
    WLs=WL;
end

if s > 0 & SGframe == 0 & SGgrau == 0
    for is=s+1:Lx_cal-s;
        X_cals(is-s,:)=mean(X_cal(is-s:is+s,:));
        X_vals(is-s,:)=mean(X_val(is-s:is+s,:));
        WLs(is-s,1)=(WL(median(is-s:is+s),1));
    end
end

%suavizando as curvas espectrais com filtros de Savitzky-Golay

if s==0 & SGframe > 0 & SGgrau >0
    str4(3)={'Savitzky-Golay(',num2str(SGgrau),',',num2str(SGframe-
1),')'};
    X_cals= sgolayfilt(X_cal,SGgrau,SGframe);
    X_vals= sgolayfilt(X_val,SGgrau,SGframe);
    WLs=WL;
end

```

```

[Lx_calcs Cx_calcs]=size(X_calcs);
[Lx_vals Cx_vals]=size(X_vals);

% figure(60), plot(WL,X_cal(:,1:2),WLS,X_calcs(:,1:2))
% legend('X_cal','X_calcs',num2str(s))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% CORRECAO DA DISPERSAO ESPECTROSCOPICA
EM_X_calcs=(mean(X_calcs)');
EM_X_vals=(mean(X_vals)');
if SC == 0
    str4(4)={' '};
end
if SC == 1
    str4(4)={'CEA'};
    for jsc=1:Cx_calcs
        betas_sc=( [ones(Lx_calcs,1) EM_X_calcs] \ X_calcs(:,jsc);
        X_calcs_sc(:,jsc)=(X_calcs(:,jsc)-betas_sc(1,1))/betas_sc(2,1);
    end
    X_calcs=X_calcs_sc;
    for jscv=1:Cx_vals
        betas_scv=( [ones(Lx_vals,1) EM_X_vals] \ X_vals(:,jscv);
        X_vals_sc(:,jscv)=(X_vals(:,jscv)-betas_scv(1,1))/betas_scv(2,1);
    end
    X_vals=X_vals_sc;

end

%
figure(350),subplot(1,2,1),plot(EM_X_calcs,X_calcs(:,1:5),'+',EM_X_calcs,EM_X_calcs,'+'
)
% subplot(1,2,2),plot(EM_X_calcs,X_calcs_sc(:,1:10))

if SC == 2
    % while abs(Lx_calcs/window-fix(Lx_calcs/window))>0 |
abs(Lx_vals/window-fix(Lx_vals/window))>0
    % window=window+1
    % end
    wd=fix(1)% fix(Lx_calcs/)
    str4(4)={'CEM',num2str(wd),'nm'};

    for jsc_cal=1:Cx_calcs
        for isc_cal=wd+1:Lx_calcs-wd
            betas_psc_cal=( [ones((2*wd+1),1) EM_X_calcs((isc_cal-
wd):(isc_cal+wd),1))] \ X_calcs(isc_cal-wd:isc_cal+wd,jsc_cal);

```

```

        X_cals_psc(isc_cal-wd,jsc_cal)=(X_cals(isc_cal,jsc_cal)-
betas_psc_cal(1,1))/betas_psc_cal(2,1);
        % end
    end
end
for jsc=1:Cx_vals
    for isc=wd+1:Lx_cals-wd
        betas_psc=(ones((2*wd+1),1) EM_X_vals((isc-
wd):(isc+wd),1))\X_vals(isc-wd:isc+wd,jsc);
        X_vals_psc(isc-wd,jsc)=(X_vals(isc,jsc)-
betas_psc(1,1))/betas_psc(2,1);
        % end
    end
end
end
[Lxcal_psc Cxcal_psc]=size(X_cals_psc);
[Lxval_psc Cxval_psc]=size(X_vals_psc);
X_cals=X_cals_psc;
X_vals=X_vals_psc;
end
%calculando as derivadas dos espectros suavizados

```

%%
PRIMEIRA DERIVADA

```

if trans_XD==1 | trans_XD==2
    clear Xcald1 Xvald1 id Lxcald1 Cxcald1 Lxvald1 Cxvald1
    Xcald1=ones(Lx_cals-1,Cx_cals);
    Xvald1=ones(Lx_vals-1,Cx_vals);
    id=ones(1);
    for id=1:Lx_cals-1
        Xcald1(id,:)=(X_cals(id+1,:)-X_cals(id,:))/(WLs(id+1,1)-WLs(id,1));
        Xvald1(id,:)=(X_vals(id+1,:)-X_vals(id,:))/(WLs(id+1,1)-WLs(id,1));
    end
    [Lxcald1 Cxcald1]=size(Xcald1);
    [Lxvald1 Cxvald1]=size(Xvald1);
end

```

%%
SEGUNDA DERIVADA

```

% clear Xcald2n Xcald2n id1 Lxcald2 Cxcald2 Lxvald2 Cxvald2
if trans_XD==2
    Xcald2n=ones(Lxcald1-1,Cx_cals);
    Xvald2n=ones(Lxvald1-1,Cx_vals);
    id1=ones(1);
    for id1=1:Lxcald1-1
        Xcald2n(id1,:)=(Xcald1(id1+1,:)-Xcald1(id1,:))/(WLs(id1+1,1)-
WLs(id1,1));
        Xvald2n(id1,:)=(Xvald1(id1+1,:)-Xvald1(id1,:))/(WLs(id1+1,1)-
WLs(id1,1));
    end

```



```

for k=1:(fix(Lx_entracal/m));
    X_calcm(k,:)=mean(X_entracal((k-1)*m+1:k*m,:));
    X_valcm(k,:)=mean(X_entraval((k-1)*m+1:k*m,:));
    WLsm(k,1)=WLS(fix(median((k-1)*m+1:k*m)),1);
end
end

Xcal=X_calcm';
[Lxcal Cxcal]=size(Xcal);
Xval=X_valcm';
[Lxval Cxval]=size(Xval);
%Box-Cox

if trans_X==7
    str4(1)={'Box-Cox'};
    Xcal=Xcal+(-min(min(Xcal)))+1;
    Xval=Xval+(-min(min(Xval)))+1;
    for Lbx=1:Lxcal
        Xcal(Lbx,:)=(boxcox((Xcal(Lbx,:))))';
    end
    for Lbxv=1:Lxval
        Xval(Lbxv,:)=(boxcox((Xval(Lbxv,:))))';
    end

end

%*****

%%%%%%%%%%
%%%%%%%%%%
%*****

MYCAL=mean(y_cal);
MYVAL=mean(y_val);
MXCAL=mean(Xcal);
MXVAL=mean(Xval);

%CENTRANDO NA MEDIA
if centrar_na_media==1
    Xcalcm=Xcal-((ones(Lxcal,1))*(mean(Xcal)));
    Xvalcm=Xval-((ones(Lxval,1))*(mean(Xval)));
    y_calcm=y_cal-((ones(Lycal,1))*(mean(y_cal)));
    y_valcm=y_val-((ones(Lyval,1))*(mean(y_val)));
    Xcal=Xcalcm;
    Xval=Xvalcm;
    y_cal=y_calcm;
    y_val=y_valcm;
end
[Lxcal Cxcal]=size(Xcal);
[Lxval Cxval]=size(Xval);

```

```

% figure(61), plot(WLsm,Xcal)% ,WLs,X_cals(:,1:2))
% legend('Xcal-1','Xcal-2')

% agrupando o vetor da variavel dependente a ser modelada a matriz X
clear dataXYcal dataXYval
dataXYcal=[y_cal Xcal];
dataXYval=[y_val Xval];
%% wk1write('datacal.wk1',dataXYcal);
%% wk1write('dataval.wk1',dataXYval);

%calculando os coeficientes de correlacao entre as variaveis
CORR_cal=corrcoef(dataXYcal);
[Lcal Ccal]=size(CORR_cal);

CR2_cal=CORR_cal.^2;%calculando os coeficientes de determinacao R2
CR2Y_cal=CR2_cal(2:Lcal,1);%isolando uma coluna que contem os
coeficients de determinacao entre y e as variaveis X
WLsm_cal=WLsm(1:Cxcal);
CORR_val=corrcoef(dataXYval);
[Lval Cval]=size(CORR_val);
CR2_val=CORR_val.^2;%calculando os coeficientes de determinacao R2
CR2Y_val=CR2_val(2:Lval,1);%isolando uma coluna que contem os
coeficients de determinacao entre y e as variaveis X

clear CORR_cal CR2_cal CORR_val CR2_val
% figure,plot(CR2Y_cal,CR2Y_val,'x')
WLsm_val=WLsm(1:Cxval);
% size(WLsm_cal)
cor_R2caR2val=corrcoef([CR2Y_cal CR2Y_val]);
cor_R2caR2val=cor_R2caR2val(1,2);
Varcacal=var(dataXYcal);
Varval=var(dataXYval);
Varcacal_R=Varcacal(1,2:Lcal);
Varval_R=Varval(1,2:Lval);

% figure(2),plot(WLsm_cal,CR2Y_cal,'-r',WLsm_val,CR2Y_val,'-b')
% title(['Derivada ',num2str(trans_X),'^a'])
% xlabel('Comprimento de onda (nm)')
% ylabel('R^2 com a densidade basica')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%

% Calculando a matriz de covariancia,os componentes principais e seus
scores
covx = cov(Xcal);
covxval=cov(Xval);

```

```

if local==1
    Xcal_L1=Xcal;
    Ycal_L1=y_cal;
    save Xcal_L1 Xcal_L1;
    save Ycal_L1 Ycal_L1;
end

if local==2
    Xcal_L2=Xcal;
    Ycal_L2=y_cal;
    save Xcal_L2 Xcal_L2;
    save Ycal_L2 Ycal_L2;
end

[pc,variances,explained] = pcacov(covx);
[pcval,variancesval,explainedval] = pcacov(covxval);

score_cal=Xcal*pc;
randSCORE_cal(:,:,Rod)=score_cal;
score_val=Xval*pc;
randSCORE_val(:,:,Rod)=score_val;
scoreCORR_cal=corrcoef([y_cal score_cal]);
scoreCORR_val=corrcoef([y_val score_val]);

%          regstats(y_cal,score_cal(:,1:20))
%          %%%%%%%%%% PLOTANDO SCORES VS INDEPENDENTE X
XXX XXXXXX X XXXXXXXXXXXXXXX XXXXXXXXXXXXXXX XXXXXXXXXXX
%
% compon=30;
%      figure(50),subplot(1,2,1),plot(1:compon,explained(1:compon,1),'-
o',1:compon,explainedval(1:compon,1),'-xk')
% title([str4,])
% xlabel('Componente')
% ylabel('Percentual da variância')
% legend('Calibração','Validação')
%      subplot(1,2,2),plot(1:compon,(scoreCORR_cal(2:compon+1,1)).^2,'-
o',1:compon,(scoreCORR_val(2:compon+1,1)).^2,'-xk')
% xlabel('Componente')
% ylabel('R^2 com a densidade básica')
% axis([0 compon 0 0.6])
% legend('Calibração','Validação')
%
%          figure(3),subplot(1,2,1),[AX,H1,H2] =
plotyy(WL,dataXcal(:,1),WLsm,Xcal(1,:),'plot');
% set(get(AX(1),'Ylabel'),'String',[str4(1)])
% set(get(AX(2),'Ylabel'),'String',[str4(1);str4(2)])
% xlabel('Comprimento de onda (nm)')
% set(H1,'LineStyle','-')
% set(H2,'LineStyle','-')
% set(H1,'Marker','-')
% set(H2,'Marker','+')

```

```

%% legend('espectro',[trans_X,num2str(trans_X),'^o'])
% subplot(1,2,2),plot(WLsm_cal,CR2Y_cal,'-o',WLsm_cal,CR2Y_val,'-x')
% title([str4,])
% xlabel('Comprimento de onda (nm)')
% axis([150 1050 0 1])
% ylabel('R^2 com a densidade básica')
% legend('Calibração','Validação')
%
% R2_X=(corrcoef(Xcal)).^2;
% figure(200),surf(WLsm_cal,WLsm_cal,R2_X)
% axis([200 1000 200 1000 0 1])
% xlabel('Comprimento de onda (nm)')
% ylabel('Comprimento de onda (nm)')
% zlabel('R^2')
% colormap(gray)
% colorbar
% shading interp
%
% figure(201), imshow(R2_X)
% title([str4])
% colormap(gray)
% colorbar
%
% [counts,x] = hist(R2_X);
% counts=100*(counts./(sum(sum(counts))));
% figure(202), stem(x,counts)
% title([str4(1);str4(2);str4(3);str4(4)])
% xlabel('R^2 com a densidade básica')
% clear dataXcal dataXval
% end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
% ajustando modelos por forward

if metodo_regr==1
clear Lisc COMP S_cal_fw S_val_fw S_cal_fw S_val_fw s_cal_fw
s_val_fw betas_fw
clear ym_cal_fw ym_val_fw desviocal_fw desvioval_fw
clear scoreCORR_cal_fw scoreCORR_val_fw Lisc_fw EPE_fw
R2cal_fw
clear R2calorig_fw RMSEP_fw R2val_fw R2valorig_fw

[Lisc
Cisc]=find(abs(scoreCORR_cal(2:Cxcal+1,1))==max(abs(scoreCORR_cal(2:Cxcal+1,1
)))));
COMP(1,1)=Lisc;
for NC=1:Cxcal-1;
S_cal_fw=score_cal(:,COMP);

```

```

S_val_fw=score_val(:,COMP);
s_cal_fw=[ones(size(y_cal)), S_cal_fw];
s_val_fw=[ones(size(y_val)), S_val_fw];
betas_fw=s_cal_fw\y_cal;
ym_cal_fw=s_cal_fw*betas_fw;
ym_val_fw=s_val_fw*betas_fw;
desviocal_fw=(y_cal-ym_cal_fw);
desvioval_fw=(y_val-ym_val_fw);
scoreCORR_cal_fw=corrcoef([desviocal_fw score_cal]);
scoreCORR_val_fw=corrcoef([desvioval_fw score_val]);
[Lisc_fw Cisc_fw]=find(abs(scoreCORR_cal_fw(2:Cxcal+1,1))==...
    max(abs(scoreCORR_cal_fw(2:Cxcal+1,1))));
COMP(NC+1,1)=Lisc_fw;
EPE_fw(NC,1)=((sum((y_cal-ym_cal_fw).^2))/Lxcal)^0.5;
R2cal_fw=(corrcoef([y_cal,ym_cal_fw])).^2;
R2calorig_fw(NC,1)=R2cal_fw(1,2);
RMSEP_fw(NC,1)=((sum((y_val-ym_val_fw).^2))/Lxval)^0.5;
R2val_fw=(corrcoef([y_val,ym_val_fw])).^2;
R2valorig_fw(NC,1)=R2val_fw(1,2);

[Ly_cal Cy_cal]=size(y_cal);
rand_y_cal(1:Ly_cal,Rod)=y_cal;

[Ly_val Cy_val]=size(y_val);
rand_y_val(1:Ly_val,Rod)=y_val;

[Lym_cal Cym_cal]=size(ym_cal_fw);
rand_ym_cal(1:Lym_cal,NC,Rod)=ym_cal_fw;

[Lym_val Cym_val]=size(ym_val_fw);
rand_ym_val(1:Lym_val,NC,Rod)=ym_val_fw;

[Lbetas Cbetas]=size(betas_fw);
rand_betas(1:Lbetas,NC,Rod)=betas_fw;

[Ldesviocal Cdesviocal]=size(desviocal_fw);
rand_desviocal(1:Ldesviocal,NC,Rod)=desviocal_fw;

[Ldesvioval Cdesvioval]=size(desvioval_fw);
rand_desvioval(1:Ldesvioval,NC,Rod)=desvioval_fw;
[b1,bint1,r1,rint1,stats1] = regress(y_cal,s_cal_fw);

randF(NC,Rod)=stats1(1,2);
rand_prob(NC,Rod)=stats1(1,3);

end
figure(396),plot(COMP,'-x')

```

```

figure(397),plot(1:NC,(100*EPE_fw/mean(y_cal)),1:NC,(100*RMSEP_fw/mean(y_val
)))

```

```

        legend('EPE_fw','RMSEP_fw')
        randCOMP(:,Rod)=COMP;
        % figure(396),plot(randCOMP,'x')
        %
figure(397),plot(1:NC,(100*EPE_fw/mean(y_cal)),1:NC,(100*RMSEP_fw/mean(y_val
)))
        % legend('EPE_fw','RMSEP_fw')
        nc_cal_fw=(1:NC)';
        nc_val_fw=(1:NC)';

        % rand_ym_cal_fw(:,Rod)=ym_cal_fw;
        % rand_ym_val_fw(:,Rod)=ym_val_fw;
        % rand_betas_fw(:,Rod)=betas_fw;
        % rand_desviocal_fw(:,Rod)=desviocal_fw;
        % rand_desvioval_fw(:,Rod)=desvioval_fw;

        EPEporcent_fw=100*(EPE_fw./mean(y_cal));
        randEPE_fw(:,Rod)=EPEporcent_fw;
        randR2cal_fw(:,Rod)=R2calorig_fw;

        RMSEPporcent_fw=100*(RMSEP_fw./mean(y_val));
        randRMSEP_fw(:,Rod)=RMSEPporcent_fw;
        randR2val_fw(:,Rod)=R2valorig_fw;

        [LR2max
CR2max]=find(randR2val_fw==(max(randR2val_fw(:,Rod))));
        [LMin
CMin]=find(randRMSEP_fw==(min(randRMSEP_fw(:,Rod))));%determinando o
menor RMSEP dentro da rodada Rod
        [LszLmin CszLmim]=size(LMin);
        if LszLmin>1 & LMin(1,1)==LMin(2,1)
            LMin=LMin(1,1);
        end
        RMSEPMIN(Rod,1)=randRMSEP_fw(LMin,CMin);%acumulando os
menores valores de RMSEP de cada rodada
        RMSEPMIN(Rod,2)=LMin;%acumulando os numeros de componentes
que geraram os menores valores de RMSEP de cada rodada
        [test cnttrans cntderiv Rod]

        end% FIM DO LOOP DAS REGRESSOES FOWARD

        %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
ADICIONANDO OS SCORES PELA ORDEM

        % Ajustando modelos contendo números de componentes
        % variando de 1 ao total de variaveis X (p=Cxcal)
        % adicionando os escores pela ordem

        if metodo_regr==0

```

```

clear nc_cal nc_val S_cal S_val s_cal s_val betas ym_cal
clear ym_val desviocal R2valorig R2val RMSEP R2calorig R2cal EPE
clear desvioval_m desviocal_m desviocal betadesv desviocaldes

RMSEPdel

clear ym_caldes EPEdel ym_valdes
for c=1:fix(Cxcal);
    S_cal=score_cal(:,1:c);
    S_val=score_val(:,1:c);
    s_cal=[ones(size(y_cal)), S_cal];
    s_val=[ones(size(y_val)), S_val];
    betas=s_cal\y_cal;
    ym_cal=s_cal*betas;
    ym_val=s_val*betas;
    desviocal=(y_cal-ym_cal);
    desvioval=(y_val-ym_val);

    EPE(c,1)=((sum((y_cal-ym_cal).^2))/Lxcal)^0.5;
    R2cal=(corrcoef([y_cal ym_cal])).^2;
    R2calorig(c,1)=R2cal(1,2);
    RMSEP(c,1)=((sum((y_val-ym_val).^2))/Lxval)^0.5;
    R2val=(corrcoef([y_val ym_val])).^2;
    R2valorig(c,1)=R2val(1,2);

% load RANDS

[Ly_cal Cy_cal]=size(y_cal);
rand_y_cal(1:Ly_cal,Rod)=y_cal;

[Ly_val Cy_val]=size(y_val);
rand_y_val(1:Ly_val,Rod)=y_val;

[Lym_cal Cym_cal]=size(ym_cal);
rand_ym_cal(1:Lym_cal,c,Rod)=ym_cal;

[Lym_val Cym_val]=size(ym_val);
rand_ym_val(1:Lym_val,c,Rod)=ym_val;

[Lbetas Cbetas]=size(betas);
rand_betas(1:Lbetas,c,Rod)=betas;

[Ldesviocal Cdesviocal]=size(desviocal);
rand_desviocal(1:Ldesviocal,c,Rod)=desviocal;

[Ldesvioval Cdesvioval]=size(desvioval);
rand_desvioval(1:Ldesvioval,c,Rod)=desvioval;
[b1,bint1,r1,rint1,stats1] = regress(y_cal,s_cal);

randF(c,Rod)=stats1(1,2);
rand_prob(c,Rod)=stats1(1,3);

```

```

end
% clear randEPE randR2cal randRMSEP randR2val

EPEporcent=100*(EPE./mean(y_cal));
randEPE(:,Rod)=EPEporcent;
randR2cal(:,Rod)=R2calorig;

RMSEPporcent=100*(RMSEP./mean(y_val));
randRMSEP(1:c,Rod)=RMSEPporcent;
randR2val(:,Rod)=R2valorig;
nc_cal=(1:c)';
nc_val=(1:c)';

[LR2max CR2max]=find(randR2val==(max(randR2val(:,Rod))));
[LMin
CMin]=find(randRMSEP==(min(randRMSEP(:,Rod))));%determinando o menor
RMSEP dentro da rodada Rod
[LszLmin CszLmim]=size(LMin);
if LszLmin>1 & LMin(1,1)==LMin(2,1)
    LMin=LMin(1,1);
end
RMSEPMIN(Rod,1)=randRMSEP(LMin,CMin);%acumulando os
menores valores de RMSEP de cada rodada
RMSEPMIN(Rod,2)=LMin;%acumulando os numeros de componentes
que geraram os menores valores de RMSEP de cada rodada
[test cnttrans cntderiv Rod]

end% FIM DO LOOP DAS REGRESSOES PELA ORDEM
end% FIM DO LOOP DAS RODADAS

[Lmin Cmin]=find(RMSEPMIN(:,1)==(min(RMSEPMIN(:,1))));
Lmin=Lmin(1,1); Cmin=Cmin(1,1);
[Lmax Cmax]=find(RMSEPMIN(:,1)==(max(RMSEPMIN(:,1))));
Lmax=Lmax(1,1); Cmax=Cmax(1,1);
% clear Lmed Cmed
[Lmed Cmed]=find((RMSEPMIN(:,1)>(mean(RMSEPMIN(:,1))-0.05)) &
(RMSEPMIN(:,1)<((mean(RMSEPMIN(:,1))+0.05))));
% [Lmed Cmed]=find((RMSEPMIN(:,1)>(mean(RMSEPMIN(:,1))-
(mean(RMSEPMIN(:,1))/20)) &
RMSEPMIN(:,1)<((mean(RMSEPMIN(:,1))+mean(RMSEPMIN(:,1))/20))));
if isempty(Lmed)==1
    [Lmed Cmed]=find((RMSEPMIN(:,1)>(min(RMSEPMIN(:,1))) &
(RMSEPMIN(:,1)<((max(RMSEPMIN(:,1))))));
end
RMSEPMIN_minmedium=RMSEPMIN(Lmed,:);
[Lmediun
Cmediun]=find(RMSEPMIN(Lmed,2)==min(RMSEPMIN(Lmed,2)));
RMSEPMIN_minimorumedium=RMSEPMIN_minmedium(Lmediun,:);

```

```

[Lmediunmin
Cmediunmin]=find(RMSEPMIN_minimorumedium(:,1)==min(RMSEPMIN_minimoru
medium(:,1)));
[Lmed_1
Cmed_1]=find(RMSEPMIN(:,.)==RMSEPMIN_minimorumedium(Lmediunmin,Cmedi
unmin));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% ANOVA
% randsel_cal(Rod,:)=sel_cal;
% randsel_val(Rod,:)=sel_val;
if Lmin>0 & Lmed_1>0 & Lmax>0
    for ano=1:3
        if ano==1
            ana=Lmin;
        end
        if ano==2
            ana=Lmed_1;
        end
        if ano==3
            ana=Lmax;
        end
        clear XU yrec
        XU=randSCORE_cal(:,1:RMSEPMIN(ana,2),ana);
        yrec=rand_y_cal(1:Ly_cal,ana);
        ymrec_cal=rand_ym_cal(1:Lycal,(RMSEPMIN(ana,2)),ana);
        yrec_val=rand_y_val(1:Ly_val,ana);
        ymrec_val=rand_ym_val(1:Ly_val,(RMSEPMIN(ana,2)),ana);
        Xvalo=[ones(size(yrec_val)) yrec_val];
        [bm,bintm,rm,rintm,statsm] = regress(ymrec_val,Xvalo);
        XU1=[ones(size(yrec)) XU];
        [b,bint,r,rint,stats] = regress(yrec,XU1);
            % size(ymrec_val)
        % size(Xvalo)
        % size(
        %
figure(5353),subplot(1,2,1),plot(yrec,ymrec_cal,'x',yrec_val,ymrec_val,'x',yrec_val,rm,'
x',yrec,yrec,'x')
        % legend('calibra','valida')
        % subplot(1,2,2),plot(yrec,r,'x',yrec_val,rm,'x')
        % legend('calibra','valida')

            %b,bint,r,rint,
        p=stats(1,3);
        R2_val=statsm(1,1);

        BU=XU1\yrec;
        [nu pu]=size(XU);
        %analise de variancia da regressao simples
        u=ones(nu,1);% vetor de 1s de comprimento nu

```

```

cu=((1/nu)*yrec'*u*u'*yrec);%correcao para a media(eh necessario se
houve padronizacao???)
SQTu=yrec'*yrec-cu;% soma de quadrados total para y
SQRegu=BU'*XU1'*yrec-cu;% soma de quadrados da regressao para y
SQResu=yrec'*yrec-BU'*XU1'*yrec;% soma de quadrados do residuo
para y
R2u=SQRegu/SQTu;% coeficiente de determinacao multiplo para y
ru=(R2u)^0.5;% coeficiente de corre3lacao multiplo para y-cu
Syx=((yrec'*yrec)-((BU'*XU1'*yrec)))/(nu-pu-1)^0.5;% erro padrao da
estimativa de y
GLRegu=pu;% grau de liberdade da regressao
GLResu=nu-pu-1;% grau de liberdade do residuo
GLTotu=nu-1;% grau de liberdade total
QMRegu=SQRegu/GLRegu;% quadrado medio da regressao
QMResu=SQResu/GLResu;% quadrado medio do residuo
QMTotu=SQTu/GLTotu;% quadrado medio total
Fu=QMRegu/QMResu;% F calculado
% figure(557),plot(yrec,XU1*BU,'o')
% Quadro da analise de variancia da regressao
result=[Syx R2u];

ANOVAu(:,:,ano)=[R2u      R2_val      100*(Syx/mean(yrec))
(RMSEPMIN(ana,1));
GLRegu SQRegu QMRegu Fu;
GLResu SQResu QMResu p;
GLTotu SQTu 0 0];

end% FIM DO LOOP DA ANOVA

%[num do teste mimdmx tipodad derivada numdePC/RMSEPmin
R2_calibracao R2_validacao EPE (calibra) RMSEP(valida) RMSEP & nPC
MEDIA_RMSEP&nPC CV(%) RMSEP&nPC F_calibracao p_calibracao
RESUMOMin(:,cnttrans,cntderiv)=[test 1 trans_X trans_XD
(ANOVAu(2,1,1)) ANOVAu(1,1,1) ANOVAu(1,2,1) ANOVAu(1,3,1)
RMSEPMIN(Lmin,:) mean(RMSEPMIN(:,1)) mean(RMSEPMIN(:,2))
100*(std(RMSEPMIN(:,1))./mean(RMSEPMIN(:,1)))
100*(std(RMSEPMIN(:,2))./mean(RMSEPMIN(:,2))) kurtosis(RMSEPMIN(:,1))
kurtosis(RMSEPMIN(:,2)) ANOVAu(2,4,1) ANOVAu(3,4,1)];
RESUMOMed(:,cnttrans,cntderiv)=[test 2 trans_X trans_XD
(ANOVAu(2,1,2)) ANOVAu(1,1,2) ANOVAu(1,2,2) ANOVAu(1,3,2)
RMSEPMIN(Lmed_1,:) mean(RMSEPMIN(:,1)) mean(RMSEPMIN(:,2))
100*(std(RMSEPMIN(:,1))./mean(RMSEPMIN(:,1)))
100*(std(RMSEPMIN(:,2))./mean(RMSEPMIN(:,2))) kurtosis(RMSEPMIN(:,1))
kurtosis(RMSEPMIN(:,2)) ANOVAu(2,4,2) ANOVAu(3,4,2)];
RESUMOMax(:,cnttrans,cntderiv)=[test 3 trans_X trans_XD
(ANOVAu(2,1,3)) ANOVAu(1,1,3) ANOVAu(1,2,3) ANOVAu(1,3,3)
RMSEPMIN(Lmax,:) mean(RMSEPMIN(:,1)) mean(RMSEPMIN(:,2))
100*(std(RMSEPMIN(:,1))./mean(RMSEPMIN(:,1)))
100*(std(RMSEPMIN(:,2))./mean(RMSEPMIN(:,2))) kurtosis(RMSEPMIN(:,1))
kurtosis(RMSEPMIN(:,2)) ANOVAu(2,4,3) ANOVAu(3,4,3)];

```

```

rotResumo=char('MnMdMx','tipodad','derivada','numdePC/RMSEPmin','R2_calibracao'
              'R2_validacao','EPE
(calibra)','RMSEP(valida)','nPC','MEDIA_RMSEP','MEDIA_nPC','CV(%)_RMSEP','C
V(%)_nPC','kurtosis_RMSEP','kurtosis_nPC','F_calibracao','p_calibracao');
resres=[RESUMOmin(:,1); RESUMOmin(:,2);RESUMOmin(:,3);...
        RESUMOmed(:,1); RESUMOmed(:,2);RESUMOmed(:,3);...
        RESUMOmax(:,1); RESUMOmax(:,2);RESUMOmax(:,3)'];
wk1 write(['resresfimt',entrada,num2str(test),'.wk1'],resres);% gravando
resultados num arquivo do Lotus
end
end% FIM DO LOOP DAS DERIVADAS
end% FIM DO LOOP DAS TRANSFORMAÇÕES
end % FIM DO LOOP DOS TESTES
end % FIM DO LOOP DOS METODOS DE REGRESSAO
beep
toc

```