

**UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE QUÍMICA**

**APLICAÇÃO DE TÉCNICAS MULTIVARIADAS EM ESPECTROS DE
INFRAVERMELHO PARA A DETERMINAÇÃO DE TEORES TOTAIS
DE CARBONO, OXIGÊNIO E HIDROGÊNIO EM AMOSTRAS DE
BIOMASSAS E BIOCÁRVÕES**

Yuri Araújo Borges

Orientador: Prof. Dr. Freddy Fernandes Guimarães

Goiânia
2015

Folha de Aprovação

Membros da Comissão Julgadora da Dissertação de Mestrado em Química, apresentada ao Instituto de Química da Universidade Federal de Goiás pelo discente Yuri Araújo Borges, em 19/06/2015:



Prof. Dr. Freddy Fernandes Guimarães (UFG)



Prof. Dr. Patterson Patrício de Souza (CEFET-MG)



Prof. Dr. Anselmo Elcana de Oliveira (UFG)

**UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE QUÍMICA**

**APLICAÇÃO DE TÉCNICAS MULTIVARIADAS EM ESPECTROS DE
INFRAVERMELHO PARA A DETERMINAÇÃO DE TEORES TOTAIS
DE CARBONO, OXIGÊNIO E HIDROGÊNIO EM AMOSTRAS DE
BIOMASSAS E BIOCÁRVÕES**

Yuri Araújo Borges

Orientador: Prof. Dr. Freddy Fernandes Guimarães

Goiânia
2015

Ficha catalográfica elaborada automaticamente
com os dados fornecidos pelo(a) autor(a), sob orientação do Sibi/UFG.

Borges, Yuri Araújo

Aplicação de técnicas multivariadas em espectros de infravermelho para a determinação de teores totais de carbono, oxigênio e hidrogênio em amostras de biomassas e biocarvões [manuscrito] / Yuri Araújo Borges. - 2015.
x, 99 f.: il.

Orientador: Prof. Dr. Freddy Fernandes Guimarães; co-orientadora Dra. Beata Eموke Madari.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Química (IQ), Programa de Pós-Graduação em Química, Goiânia, 2015.

Bibliografia. Apêndice.

Inclui siglas, abreviaturas, símbolos, gráfico, tabelas, algoritmos, lista de figuras, lista de tabelas.

1. Biocarvão. 2. Infravermelho. 3. Técnicas Multivariadas. I. Guimarães, Freddy Fernandes, orient. II. Madari, Beata Eموke, co orient. III. Título.

Primeiramente à Deus, e a família e amigos

AGRADECIMENTOS

Ao Professor Dr. Freddy Fernandes Guimarães pelas orientações e excessiva paciência durante o curso.

À Pesquisadora Dr. Beata Eموke Madari do Centro Nacional de Pesquisa de Arroz e Feijão (CNPAF/EMBRAPA) por viabilizar a realização deste estudo e contribuir com diversas informações.

Ao amigo Diego Mendes de Souza pela ajuda e conhecimentos prestados para o desenvolvimento do trabalho.

Ao Professores Dr. Anselmo Elcana Oliveira do Instituto de Química da UFG (IQ-UFG) e Patterson Patrício de Souza do CEFET de Minas Gerais (CEFET-MG), e a Professora Dra. Carolina Horta Andrade da Faculdade de Farmácia da UFG (FF-UFG) pelas sugestões e correções referentes ao trabalho.

Aos amigos do LTCC pelo companheirismo e dedicação.

SUMÁRIO

LISTA DE FIGURAS	i
LISTA DE TABELAS	vii
LISTA DE ABREVIATURAS DE SIGLAS	viii
RESUMO	x
CAPÍTULO 1 – INTRODUÇÃO	1
CAPÍTULO 2 – BIOCHAR	5
2.1 As Terras Pretas de Índios da Amazônia	5
2.2 Produção, estrutura e reatividade do carbono pirogênico	6
2.3 Interesse ambiental e agrícola do carbono pirogênico	8
CAPÍTULO 3 – ESPECTROSCOPIA NO INFRAVERMELHO	11
3.1 Introdução à espectroscopia no infravermelho	11
3.2 Espectroscopia no infravermelho em determinações quantitativas	13
3.3 Análise da biomassa pirolisada por espectroscopia no infravermelho	14
CAPÍTULO 4 – PRÉ-PROCESSAMENTO DE DADOS, SELEÇÃO DE VARIÁVEIS E TÉCNICAS MULTIVARIADAS	17
4.1 Multiplicative Scatter Correction: Pré-processamento de dados	18
4.2 SPA, GA e iPLS na seleção de variáveis	19
4.3 Técnicas de Calibração Multivariada	24
Calibração por MLR	24
Calibração por PLS	24
4.4 Cálculo dos erros Análise	26
4.5 Parâmetro RPD (Relative Predictive Determinant)	26
4.6 Análise de Componentes Principais (PCA)	27
4.7 Análise Hierárquica de Cluster (HCA)	28
CAPÍTULO 5 – Materiais e Métodos	30
5.1 Amostragem	30
5.2 Espectroscopia MID-IR	30
5.3 Recursos Computacionais	31
5.4 Separação de amostras para calibração e validação	31
5.5 Parâmetros utilizados nas técnicas de seleção de variáveis GA e iPLS	32
5.6 Tratamento Estatístico dos Dados	33
Proposta 1 – Seleção de Variáveis por SPA e calibração por MLR	33

Proposta 2 – Calibração por PLS	33
Proposta 3 – Seleção de Variáveis por GA com validação cruzada em MLR e calibração por MLR	34
Proposta 4 – Seleção de Variáveis por GA com validação cruzada em PLS e calibração por PLS	34
Proposta 5 – Seleção de Variáveis por iPLS e calibração por PLS	34
CAPÍTULO 6 – TRATAMENTO ESPECTRAL E ANÁLISE DE CLUSTER	35
6.1 Identificação de compostos das biomassas no espectro MID-IR	35
6.2 Pré-tratamento espectral	36
6.3 Análise Hierárquica de Cluster	37
6.4 Análise de Componentes Principais	42
CAPÍTULO 7 – DETERMINAÇÃO DE CARBONO TOTAL	45
7.1 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 1 (SPA-MLR)	45
7.2 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 2 (PLS)	47
7.3 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)	49
7.4 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)	50
7.5 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)	52
CAPÍTULO 8 – DETERMINAÇÃO DE OXIGÊNIO TOTAL	57
8.1 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 1 (SPA-MLR)	57
8.2 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 2 (PLS)	59
8.3 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)	60
8.4 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)	62
8.5 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)	63
CAPÍTULO 9 – DETERMINAÇÃO DE HIDROGÊNIO TOTAL	68
9.1 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 1 (SPA-MLR)	68
9.2 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 2 (PLS)	70
9.3 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)	71
9.4 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)	72
9.5 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)	75
CAPÍTULO 10 – CONCLUSÕES	79
REFERÊNCIAS BIBLIOGRÁFICAS	81
APÊNDICE A – DETERMINAÇÃO DE NITROGÊNIO TOTAL	89
A.1 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 1 (SPA-MLR)	89
A.2 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 2 (PLS)	91
A.3 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)	93
A.4 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)	94
A.5 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)	96

LISTA DE FIGURAS

Figura 2.1. Conformações estruturais do carbono pirogênico.	6
Figura 2.2. Representação do ciclo do carbono natural e por biocarvão.	9
Figura 3.1. Regiões do espectro eletromagnético e seus respectivos comprimentos de onda.	11
Figura 3.2. Tipos de vibrações permitidas entre 3 átomos em um composto com 5 átomos.	12
Figura 3.3. Regiões do espectro infravermelho e tipos de ligações ou grupos funcionais correspondentes.	13
Figura 3.4. 1-Lignina, 2-Celulose, 3-Hemicelulose.	15
Figura 3.5. Espectro de infravermelho médio de uma amostra de biocarvão.	16
Figura 4.1. Fluxograma resumido das etapas do Algoritmo Genético (GA).	20
Figura 4.2. Algoritmo das Projeções Sucessivas (SPA).	23
Figura 4.3. Modelo esquemático PLS.	25
Figura 4.4. Exemplo de dendrograma obtido pela análise cluster.	29
Figura 6.1. Espectros originais de infravermelho médio das amostras de biomassas e biocarvões.	35
Figura 6.2. Espectro de infravermelho médio das biomassas e biocarvões corrigido por MSC.	36
Figura 6.3. Dendrograma obtido pela HCA para as amostras de biomassas e biocarvões.	37
Figura 6.4. Gráfico comparativo do percentual de oxigênio entre as amostras dos grupos 1 e 2.	40
Figura 6.5. Gráfico comparativo do percentual de carbono entre as amostras dos grupos 1 e 2.	41

- Figura 6.6. Gráfico comparativo do percentual de hidrogênio entre as amostras dos grupos 3 e 4. 42
- Figura 6.7. Representação gráfica dos *scores* da componente principal 1. 43
- Figura 6.8. Representação gráfica dos *loadings* da componente principal 1. 44
- Figura 7.1. Espectro MID-IR médio das 44 amostras com destaque para as 12 variáveis selecionadas pelo SPA para a determinação de carbono. 45
- Figura 7.2. Valor previsto para carbono total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta predita e as barras a resposta experimental da análise elementar. 47
- Figura 7.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de carbono total. 48
- Figura 7.4. Valor previsto para carbono total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 49
- Figura 7.5. Valor previsto para carbono total pelo modelo de calibração MLR (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 50
- Figura 7.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de carbono total. 51
- Figura 7.7. Valor previsto para carbono total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 52
- Figura 7.8. Espectro MID-IR médio das 44 amostras com destaque para as 45 variáveis selecionadas pelo iPLS correspondente a 7 intervalos de números de onda para a determinação da concentração de carbono. 53

Figura 7.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de carbono total. 54

Figura 7.10. Valor previsto para carbono total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 55

Figura 8.1. Espectro MID-IR médio das 44 amostras com destaque para as 6 variáveis selecionadas pelo SPA para a determinação de oxigênio. 57

Figura 8.2. Valor previsto para oxigênio total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta predita e as barras a resposta experimental da análise elementar. 58

Figura 8.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de oxigênio total. 59

Figura 8.4. Valor previsto para oxigênio total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 60

Figura 8.5. Valor previsto para oxigênio total pelo modelo de calibração MLR (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 61

Figura 8.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de oxigênio total. 62

Figura 8.7. Valor previsto para oxigênio total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 63

Figura 8.8. Espectro MID-IR médio das 44 amostras com destaque para as 100 variáveis selecionadas pelo iPLS correspondente a 13 intervalos de números de onda para a determinação da concentração de oxigênio. 64

Figura 8.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de oxigênio total. 65

Figura 8.10. Valor previsto para oxigênio total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 66

Figura 9.1. Espectro MID-IR médio das 44 amostras com destaque para as 6 variáveis selecionadas pelo SPA para a determinação de hidrogênio total. 68

Figura 9.2. Valor previsto para hidrogênio total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta predita e as barras a resposta experimental da análise elementar. 69

Figura 9.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de hidrogênio total. 70

Figura 9.4. Valor previsto para hidrogênio total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 71

Figura 9.5. Valor previsto para hidrogênio total pelo modelo de calibração MLR (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 72

Figura 9.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de hidrogênio total. 73

Figura 9.7. Valor previsto para hidrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 74

Figura 9.8. Espectro MID-IR médio das 44 amostras com destaque para as 105 variáveis selecionadas pelo iPLS correspondente a 11 intervalos de números de onda para a determinação da concentração de hidrogênio. 75

Figura 9.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de hidrogênio total. 77

Figura 9.10. Valor previsto para hidrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 77

Figura A.1. Espectro MID-IR médio das 44 amostras com destaque para as 2 variáveis selecionadas pelo SPA para a determinação de nitrogênio. 89

Figura A.2. Valor previsto para nitrogênio total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta predita e as barras a resposta experimental da análise elementar. 91

Figura A.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de nitrogênio total. 92

Figura A.4. Valor previsto para nitrogênio total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 92

Figura A.5. Valor previsto para nitrogênio total pelo modelo de calibração MLR (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 94

Figura A.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de nitrogênio total. 95

Figura A.7. Valor previsto para nitrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 96

Figura A.8. Espectro MID-IR médio das 44 amostras com destaque para as 30 variáveis selecionadas pelo iPLS correspondente a 9 intervalos de números de onda para a determinação da concentração de nitrogênio. 97

Figura A.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de nitrogênio total. 98

Figura A.10. Valor previsto para nitrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elementar. (●) Calibração e (♥) Validação. Linha verde calibração, linha vermelha validação. 99

LISTA DE TABELAS

Tabela 5.1. Tabela de parâmetros utilizados para seleção de variáveis por GA.	32
Tabela 6.1. Tabela de análise elementar (%) das amostras de biomassas e biocarvões.	38
Tabela 6.2. Porcentagem de variância explicada pelas componentes principais.	42
Tabela 7.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de carbono total.	56
Tabela 8.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de oxigênio total.	66
Tabela 9.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de hidrogênio total.	78
Tabela A.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de nitrogênio total.	99

LISTA DE ABREVIATURAS E SIGLAS

GA – Genetic Algorithm

FIR – Far Infrared

FT-IR – Fourier Transform Infrared Spectroscopy

HCA – Hierarchical Cluster Analysis

iPLS – Interval Partial Least Square

LS-SVM – Least Squares Support Vector Machines

MID-IR – Mid Infrared

MSC – Multiplicative Scatter Correction

NIR-IR – Near Infrared

PCA – Principal Components Analysis

PCR – Principal Components Regression

PLS – Partial Least Squares

ppm – parte por milhão

MLR – Multiple Linear Regression

RMSE – Root Mean Square Error

RMSEC – Root Mean Square Error of Calibration

RMSEP – Root Mean Square Error of Validation

RMSECV – Root Mean Square Error of Cross-Validation

RPD – Residual Prediction Deviation

ANN – Artificial Neural Network

SH – Substâncias Húmicas

SNV – Standard Normal Variate

SPA – Sucessive Projection Algorithm

SVM – Support Vector Machine

RESUMO

Um meio eficiente para a manutenção da concentração de dióxido de carbono na atmosfera é a utilização do biocarvão na agricultura. Para utilizá-lo de forma correta em determinadas situações torna-se necessária a quantificação dos principais teores elementares nele contidos assim como na sua matéria-prima, a biomassa. Para quantificação de teores totais de carbono, nitrogênio, oxigênio e hidrogênio, a utilização das técnicas de calibração multivariada MLR e PLS associada aos métodos de seleção de variáveis SPA, GA e iPLS sobre espectros de infravermelho médio de amostras de biomassas e de seus biocarvões respectivos, resultaram na obtenção de modelos para a determinação de teores totais para esses elementos. Em diversas propostas foi obtido modelos viáveis para a quantificação dos teores elementares presentes nas amostras, com destaque a proposta SPA-MLR que apresentou sempre em todos os teores modelos com capacidade predição a outros conjuntos de amostras, com valores do parâmetro RPD de 4,152; 4,517; e 3,314 para carbono, oxigênio e hidrogênio, respectivamente. Além dos modelos multivariados também foram feitas análise de agrupamentos das amostras. Elas apresentaram padrões quando submetidas às análises PCA e HCA, havendo a separação de agrupamentos de amostras com relação direta às diferenças de porcentagem composicional de carbono e oxigênio, sendo esses os elementos com maior concentração nas amostras de biomassas e biocarvões e os mais sujeitos a mudança porcentual devido ao processo de pirólise.

CAPÍTULO 1

INTRODUÇÃO

Nos últimos 100 anos a poluição atmosférica tem aumentado de tal forma que fenômenos naturais que promovem a manutenção do ecossistema no planeta têm sofrido perturbações danosas. Um desses fenômenos naturais, e que é responsável pela manutenção energética do planeta, é o efeito estufa, sendo assim chamado pela sua capacidade de reter parte da radiação vinda do sol para promover o aquecimento do planeta (IPCC, 2007) possibilitando, assim, a existência de vida como conhecida. Entretanto, com a liberação excessiva de gases que propiciam a retenção de energia térmica, fenômenos anômalos, como o aquecimento global, provocam o aumento da temperatura do planeta.

O aquecimento global é causado pela emissão de gases como por exemplo CO₂, N₂O, CH₄ e vapor de água. Dentre os gases compostos por carbono o CO₂ se destaca por ter um grande volume de emissão, pois é resultado residual de diversos processos químicos ocorridos na era contemporânea, em indústrias, agricultura, produção e refino de combustíveis fósseis.

Estima-se que a quantidade segura para a concentração atmosférica de CO₂ deveria ser no máximo de 350 ppm, mas a concentração atual já se aproxima de 400 ppm (WMO, 2014). Enquanto muitos governos e organizações se preocupam em diminuir a taxa de emissão de CO₂, outros pensam em diminuir a concentração atmosférica já existente. Vários métodos foram avaliados por Fox e Champman (Fox; Champman, 2011) para a diminuição do aquecimento global. Dentre eles, e o que representou uma solução mais viável, é a utilização do biocarvão na agricultura, para a captação de carbono atmosférico e sua retenção no solo. A utilização do biocarvão na agricultura tem tanto benefícios para a diminuição da concentração de carbono na atmosfera quanto para a correção de deficiências dos solos (Day et al, 2009), tornando possível a produtividade em terras inférteis. Outra vantagem é a capacidade de ligação química entre o biocarvão e as estruturas inorgânicas do solo evitando desmoronamentos de terrenos em épocas de chuvas.

O biocarvão é produzido através do processo de pirólise, ou carbonização, em baixa atmosfera de oxigênio da biomassa, no caso, de madeiras de diversas plantas. Esse processo permite que a biomassa resulte em um carvão com estruturas orgânicas altamente estáveis e expressiva recalcitrância química promovendo, assim, uma degradação mais lenta e maior permanência no ambiente (Rezende et al., 2011). Deve ser observado que devido aos diferentes tipos de biomassa que podem ser utilizados e aos diversos métodos de pirólise existentes, o biocarvão resultante pode apresentar variações em suas características por se tratar de um material não homogêneo. Portanto, pode-se, em princípio, produzir biocarvões com propriedades específicas, voltadas à utilização em diferentes climas, solos e tipos de cultura agrícola conforme se desejar (Brownsort et al., 2011). Entretanto, vale salientar que ainda são necessários estudos para viabilização do uso de biocarvões com características específicas na agricultura, assim como no desenvolvimento dos métodos de produção.

Aspectos específicos na biomassa são buscados para a produção do biocarvão. Entre os principais estão o teor de lignina, pois dele pode-se avaliar o rendimento na produção do biocarvão (Sohi et al., 2010; Zhang et al., 2010), e o teor de cinzas, que ajuda na correção da acidez do solo pela presença de óxidos de potássio, cálcio e magnésio. Várias plantas e madeiras têm teores diferentes de carbono, nitrogênio, oxigênio e hidrogênio podendo, assim, refletir na qualidade final do carvão obtido. Portanto, é necessário o conhecimento desses teores para observância das temperaturas utilizadas durante o processo de pirólise, visto que esses componentes principais da biomassa, lignina, celulose e hemicelulose, têm degradação térmica diferentes (Maia, 2013).

Para a análise dos teores de carbono, nitrogênio, oxigênio e hidrogênio e a correlação com as razões entre esses elementos, podem-se utilizar diversos métodos instrumentais. Dentre esses, as espectroscopias no infravermelho médio e próximo (MID-IR e NIR-IR, respectivamente) se destacam pelas suas capacidades de análises simples, rápidas e de baixo custo operacional. A espectroscopia no infravermelho quando associada às técnicas de calibração multivariada possibilita determinações dos diferentes teores elementares do biocarvão (Madari et al., 2005). Os dados obtidos pela espectroscopia no infravermelho são de um conjunto de amostras com valor conhecido

por uma metodologia de referência como, por exemplo, um conjunto de porcentagens obtidas através de análise elementar. Assim, com a utilização de um método de calibração multivariado pode-se construir um modelo matemático e correlacionar o espectro de Infravermelho com teores específicos de substâncias químicas (Morgano et al., 2007).

Antes da utilização dos dados para a construção do modelo multivariado, devem ser utilizados métodos de pré-tratamento espectral, para diminuição de ruídos do espectro e desvios na linha de base (Rinnan et al., 2009). Também é possível, e desejável, a utilização de técnicas de seleção de variáveis, como por exemplo, GA (Algoritmo Genético), PCA (Análise de Componentes Principais) ou SPA (Algoritmo de Projeções Sucessivas), as quais proporcionam a obtenção de um subconjunto de variáveis que contém informações mais relevantes. Concluídas as etapas acima descritas pode-se construir um modelo multivariado através dos diversos métodos existentes como: a Regressão Linear Múltipla (MLR), PCR (Regressão de Componentes Principais), PLS (Mínimos Quadrados Parciais), Redes Neurais Artificiais (ANN) e LS-SVM (Máquina de Vetor de Suporte com Mínimos Quadrados), entre outros. Esses modelos multivariados são utilizados para propósitos analíticos desde a década de 70, fornecendo resultados estatísticos para os sistemas estudados, mas com a utilização dos microcomputadores o processo de cálculo desses modelos se tornou ainda mais rápido e eficiente, observando a possibilidade de escolha e a utilização de diferentes métodos de calibração multivariada, pré-processamento de dados e seleção de variáveis (Neto et al., 2006).

Este trabalho tem por objetivo a obtenção de modelos multivariados que permitam a quantificação de teores totais elementares de carbono, nitrogênio, oxigênio e hidrogênio em amostras de biomassas e biocarvões, sendo assim utilizadas técnicas de calibração multivariada em espectros de MID-IR em um conjunto de 11 amostras “in natura” de 11 tipos de biomassas e 33 amostras de biocarvão produzidos em 4 temperaturas distintas de pirólise dessas mesmas biomassas. Foi utilizado o método de pré-espectral MSC (Multiplicative Scatter Correction) para a diminuição de ruídos e amenização de espalhamento de luz no espectro, os métodos de seleção de variáveis SPA (Algoritmo de Projeções Sucessivas), GA (Algoritmo Genético) e iPLS (Interval Partial

Least Square) para a separação de subconjuntos de variáveis contendo máximo de informação relevante das amostras e as técnicas de calibração multivariada PLS e MLR para a construção dos modelos.

CAPÍTULO 2

BIOCARVÃO

2.1 As Terras Pretas de Índios da Amazônia

O solo amazônico é bastante conhecido pelas suas características estéreis, resultantes tanto pelo tipo de vegetação quanto pelo clima da região que não propiciam o cultivo de plantações como em outras regiões do país. As árvores de alta estatura e copas extensas, que são a maioria, não permitem que a luminosidade solar alcance as plantas de médio e pequeno porte mais próximas do solo, inibindo o crescimento das mesmas (Teixeira et al., 2009). Outro fator é a infertilidade natural do solo amazônico, que tem como única fonte de matéria orgânica uma fina camada do solo onde se decompõem animais, plantas e as folhas das copas de árvores grandes que caem sobre o chão (Ciência Hoje, 2011). Logo, se ocorrer o desmatamento para a implantação agrícola, essa camada se extingue com o tempo, e a agricultura não é mais possível tornando o solo árido e estéril (Teixeira et al., 2009).

Entretanto, existe o conhecimento desde o século IXX de localidades da região amazônica que apresentam solos de coloração escura e que têm uma capacidade fértil alta com grande concentração de carbono. Esses solos estão diretamente relacionados à presença de assentamentos humanos indígenas datados de antes do contato com colonizadores europeus. Por isso, esse tipo de solo foi chamado de *Terras Pretas de Índios da Amazônia* (Teixeira et al., 2009). A atividade humana é perceptível nesse solo pela presença de restos animais e vegetais assim como de fragmentos de cerâmica e ossos. Cinzas e carvões também encontrados nas amostras de solo revelam traços de elementos como chumbo, magnésio, estrôncio, cálcio, zinco e outros confirmando assim a assinatura geoquímica da ocupação humana (Novotny et al., 2009). Como nas Terras Pretas de Índios da Amazônia o carbono está presente na forma de carvão, o incentivo da criação de uma forma de produção de insumos agrícolas que tivessem as mesmas características desses solos, incluindo a sua excelente capacidade de captação de carbono atmosférico, chegou assim na idealização e criação do biocarvão ou carbono pirogênico (International Biochar Initiative, 2006).

2.2 Produção, estrutura e reatividade do carbono pirogênico

A produção do biocarvão se dá através do processo de pirólise da biomassa, no caso madeiras de árvores. A pirólise se dá pela queima da biomassa em temperaturas entre 300 e 550 °C em ausência, ou em baixas concentrações, de gás oxigênio permitindo, assim manter o teor de carbono no produto em mais de 50% em relação ao teor da biomassa inicial, ao contrário da combustão que mantém somente de 2 a 3% (Ciência Hoje, 2011). Esse processo leva a estruturas quimicamente mais reduzidas, com maior grau de aromaticidade e condensação em relação às estruturas da biomassa inicial. Portanto o carvão gerado pelo processo de pirólise é diferente do carvão utilizado pra fins energéticos, pois nas estruturas aromáticas do carbono pirogênico existem grupos funcionais oxigenados, como hidroxilas e carboxilas (Teixeira et al., 2009).

As estruturas aromáticas do biocarvão apresentam a conformação gráfitica com várias camadas superpostas de conjuntos poli aromáticos, sendo essas de tamanho e organização diferentes entre si, como visto na Figura 2.1. Embora apresente unidades estruturais gráficas, ele não é necessariamente gráfico, pois apresenta grandes quantidades de cadeias alifáticas e uma razoável quantidade de oxigênio, (Poirier et al., 2002).

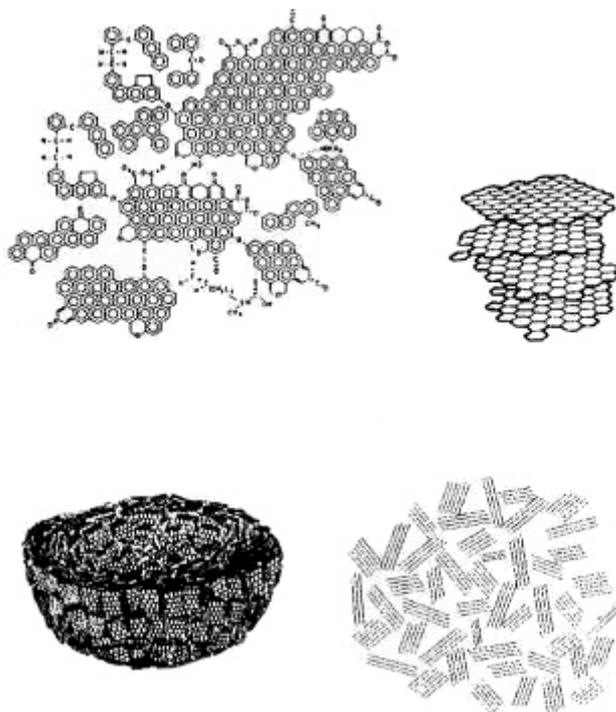


Figura 2.1. Conformações estruturais do carbono pirogênico
Fonte: adaptado de As Terras Pretas de Índios da Amazônia (2009).

As conformações gráficas se agrupam em conjuntos formando uma macroestrutura na forma esférica, com várias camadas sobrepostas. Os grupos oxigenados citados anteriormente, principalmente os grupos carboxila, que se encontram periféricamente no biocarvão (Kramer et al., 2004) podem explicar a boa capacidade de troca catiônica do mesmo, pois elementos catiônicos que influenciam na fertilidade do solo como cálcio, potássio e magnésio são retidos com facilidade.

Outra característica importante da oxidação periférica do carbono pirogênico é que as estruturas carboxiladas aromáticas são muito recalcitrantes e reativas promovendo a captura e retenção do carbono atmosférico (Novotny et al., 2006). Ainda em relação a sua estrutura, o carbono pirogênico possui uma elevada porosidade e área superficial, sendo assim capaz de abrigar microrganismos que ajudam na agregação dos solos (Teixeira et al., 2009) e, conseqüentemente, possibilita a melhor capacidade de absorção de nutrientes pelas plantas, o que ajuda no desenvolvimento de culturas agrícolas.

Outro fator importante relacionado ao carbono pirogênico é a sua provável capacidade de produção de substâncias húmicas (SH). As SH são uma fração orgânica do solo constituída de diversas substâncias oriundas da decomposição de material orgânico, seja ele vegetal ou animal, devido à ação de microrganismos além de fatores abióticos como, por exemplo, temperatura, irradiação de luz solar e pH do solo. As SH são importantes além da sua conhecida capacidade de fertilização, elas também ajudam na retenção de calor no solo devido a sua coloração escura e, assim, estimulam a germinação de sementes e o desenvolvimento de raízes, evitando a ocorrência de erosões devido à sua eficácia para a absorção de água. O carbono pirogênico, por ser um material carbonizado, pode sofrer oxidação química (Kumada, 1983; Haumaier & Zech, 1995) das cadeias laterais do carbono pirogênico originando grupos carboxílicos que aumentam a capacidade de troca de cátions e assim mantêm a estabilidade da matéria orgânica no solo (Glaser et al., 2000).

2.3 Interesse ambiental e agrícola do carbono pirogênico

A principal qualidade buscada na produção do carvão pirogênico é a reprodução e a melhoria das características encontradas nas terras pretas da Amazônia, a combinação de boa capacidade de captura de carbono atmosférico com a possibilidade de correção de pouco férteis melhorando a retenção de água e levando ao aumento da produtividade (Ciência Hoje, 2011).

Primeiramente, a captação de carbono atmosférico é a qualidade em caráter ambiental mais importante, pois as taxas de concentração de compostos contendo carbono nos dias atuais é muito alta devido aos problemas constantes de poluição e do aumento da mesma (Rezende et al., 2011). Entre esses compostos, o de maior concentração na atmosfera é o dióxido de carbono, e o carbono presente neste é redistribuído no meio ambiente pelo ciclo do carbono, que pode ser resumido em alguns passos:

Passo 1: O carbono entra na atmosfera como dióxido de carbono por respiração e combustão.

Passo 2: O dióxido de carbono é absorvido por seres produtores (que produzem seu próprio alimento, como plantas) para a produção de carboidratos na fotossíntese e liberação do oxigênio.

Passo 3: Animais se alimentam das plantas e o carbono é passado através da cadeia alimentar. Muito do carbono consumido é exalado como dióxido de carbono para a atmosfera pelo processo de respiração. Esses animais e plantas eventualmente morrem.

Passo 4: Os organismos mortos, tanto plantas quanto animais são decompostos no solo e o carbono presente em sua biomassa retorna à atmosfera na forma de dióxido de carbono (Lehmann, 2007).

A utilização do biocarvão no solo interfere diretamente no processo natural do ciclo do carbono e reduz o retorno de carbono na forma de dióxido de carbono para a atmosfera, como pode ser observado na Figura 2.2. Devido à sua recalcitrância o carbono pirogênico tem uma alta resistência à degradação química e biológica (Kuhlbusch, 1998). Portanto com utilização do mesmo no solo promove a captura do carbono atmosférico e, devido a essa estabilidade, sua meia vida é estimada em torno de 10^2 a 10^7 anos dependendo da quantidade de componentes alifáticos ou voláteis em sua estrutura (Zimmerman et al., 2010).

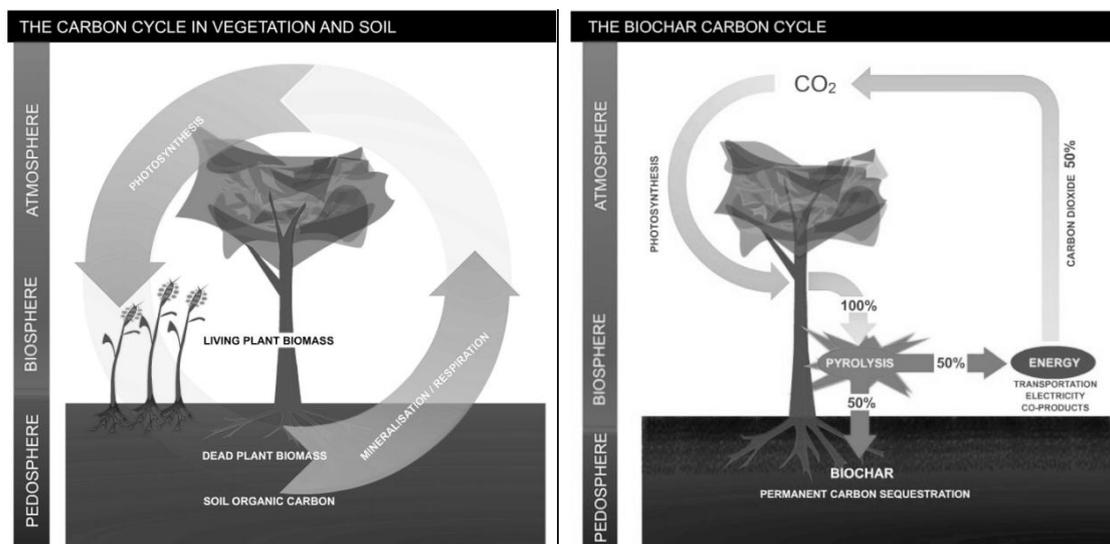


Figura 2.2. Representação do ciclo do carbono natural e por biocarvão
 Fonte: The Green Disc: New Technologies For a New Future (2009)

Observando o ciclo normal do carbono, o seu acúmulo e liberação na forma de dióxido de carbono na decomposição em ecossistemas não perturbados acontece em equilíbrio no meio ambiente, mas em sistemas perturbados o acúmulo é menor e ocorre o acúmulo de dióxido de carbono na atmosfera (The Green Disc, 2009). O biocarvão inserido no solo promove a maior retenção de carbono no solo durante o processo de decomposição da biomassa animal ou vegetal e uma menor liberação na forma de dióxido de carbono diminuindo a concentração do mesmo na atmosfera. Além da redução da concentração de carbono na atmosfera, o biocarvão também ajuda na redução de gases potencializadores do efeito estufa (*e.g.* óxido nítrico e metano), apesar deste efeito ser menos significativo. Com a adição de poucas quantidades de

carbono pirogênico no solo é possível reduzir as emissões de óxido nitroso em até 80% e extinguir as emissões de gás metano (Lehmann, J., 2007).

Aliada a boa qualidade de captação e retenção do carbono atmosférico está a capacidade de correção do solo na utilização do biocarvão. É importante observar as características e propriedades do biocarvão utilizado como, por exemplo, os teores de umidade, cinzas, carbono, e de diversos minerais para a utilização correta de um carbono pirogênico de determinada biomassa para atender às determinadas necessidades de um determinado tipo de solo. Então são necessárias análises físicas e de fertilidade do solo, da composição química do biocarvão e das exigências da cultura para que se possa estipular métodos e doses de aplicação.

As aplicabilidades de correção do solo do biocarvão são muito variadas, e dentre elas estão: melhoramento da qualidade da água, redução da lixiviação de nutrientes e acidez do solo, redução da irrigação e utilização de fertilizantes, (Day et al, 2009). Dentre as aplicabilidades, a retenção de nutrientes é extremamente favorável à capacidade de troca de cátions do solo, pois o biocarvão tem melhor eficácia de absorção de cátions por unidade de carbono do que a matéria orgânica presente no solo ajudando assim a manter minerais importantes para a fertilidade do solo, (Lehmann, J., 2007).

Ainda é necessário mais informações e dados científicos para que o biocarvão possa ser utilizado de forma completamente eficientemente em culturas agrícolas levando em conta as necessidades para cada região e de cada tipo de biocarvão. O desenvolvimento de técnicas e métodos de produção e aplicação correta de biocarvão originados de biomassas diferentes é fundamental para que se obtenha uma base sólida de utilização visando o desenvolvimento agrícola.

CAPÍTULO 3

ESPECTROSCOPIA NO INFRAVERMELHO

3.1 Introdução à espectroscopia no infravermelho

A radiação eletromagnética pode interagir com a matéria de formas diferentes dependendo da região de energia à qual é oriunda. Ondas de rádio, micro-ondas, infravermelho, visível, ultravioleta, raios-x e raios gama podem interagir com a matéria em níveis cinético (translacional e rotacional), vibracional ou eletrônico. Isso torna possível a utilização dos diferentes tipos de radiação do espectro eletromagnético na verificação de propriedades comportamentais da matéria e disso tirar informações valiosas como, por exemplo, em relação à sua identificação e quantificação. As técnicas de análise espectroscópica foram desenvolvidas, tanto para qualificação quanto para a quantificação de diversas substâncias ou misturas através da interação da radiação com a matéria.

Dentre as faixas do espectro eletromagnético, o infravermelho é dividido em três faixas características: infravermelho próximo (NIR-IR), médio (MID-IR) e distante (FIR) (Figura 3.1). O NIR-IR tem frequência entre 14000 a 4000 cm^{-1} , sendo o mais energético e provoca excitações vibracionais nas moléculas. O MID-IR tem frequência de aproximadamente entre 4000 a 400 cm^{-1} e provoca excitações vibracionais e roto-vibracionais em estruturas moleculares. E por fim o FIR com faixa de aproximadamente 400 a 10 cm^{-1} e interferindo nas energias rotacionais (Xu et al., 2013).

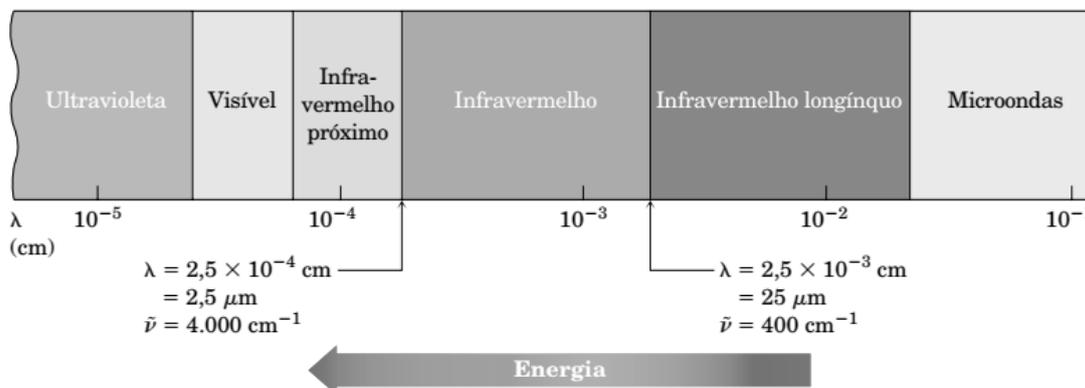


Figura 3.1. Regiões do espectro eletromagnético e seus respectivos comprimentos de onda
Fonte: McMurry (2008).

Excitações vibracionais na região do infravermelho ocorrem apenas para os modos vibracionais que apresentam variação no momento de dipolo. Os modos vibracionais que apresentam momento dipolo constantes não absorvem a radiação na forma de excitações vibracionais. Os modos vibracionais mais comuns são os estiramentos (simétricos e assimétricos) e dobramentos (scissoring, wagging, rocking e twisting) que são distorções angulares (Stuart, 2004). Alguns exemplos de estiramentos e dobramentos podem ser vistos na Figura 3.2.

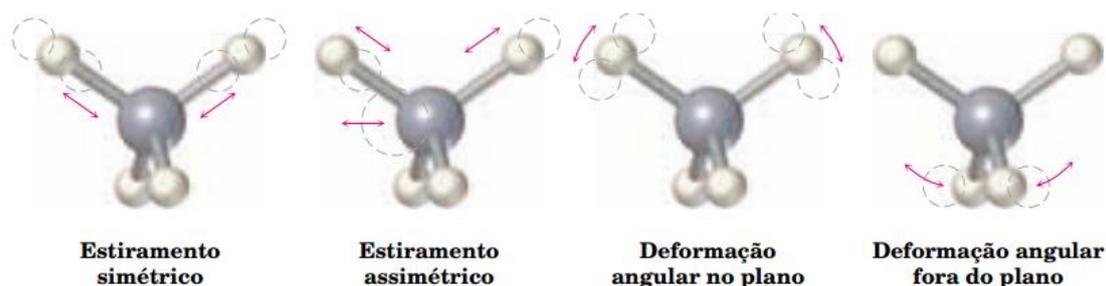


Figura 3.2. Tipos de vibrações permitidas entre 3 átomos em um composto com 5 átomos
Fonte: McMurry (2008).

É importante ressaltar que as vibrações decorrentes da absorção da radiação infravermelha ocorrem devido às transições eletrônicas entre estados vibracionais. Mas isso acontece somente em níveis de energia associados a frequências específicas (McMurry, 2008), sendo assim, os estados vibracionais absorvem quantidades de energia discretas para que ocorram transições entre níveis específicos. Isso se deve pela quantização dos níveis que pode ser exemplificada, dado a equação

$$E_n = \left(n + \frac{1}{2}\right) \hbar \omega_\alpha \quad (1)$$

que representa a resolução do oscilador harmônico quântico em uma dimensão pela equação de Schrödinger independente do tempo. Nessa equação E_n representa a energia do nível vibracional n , \hbar é a constante de Planck dividida por 2π e ω_α é a frequência angular do modo vibracional α . Assim, como as energias dos níveis vibracionais são bem definidas, somente valores específicos de energia na faixa infravermelha podem provocar excitações nos estiramentos e distorções em grupos funcionais, sendo que diferentes valores de frequências provocam excitações específicas de modos vibracionais α distintos.

3.2 Espectroscopia no infravermelho em determinações quantitativas

A espectroscopia no infravermelho é em princípio uma técnica qualitativa, ou seja, é utilizada para a identificação de compostos químicos desconhecidos em uma amostra. Com baixo custo operacional, análises simples e rápidas, e baixíssima geração de resíduos, é utilizada como uma técnica comumente recorrente para elucidação de compostos desconhecidos em solos, biomassas ou mesmo em laboratórios de controle qualidade em indústrias diversas. Através das bandas de absorção observadas no espectro podemos relacionar cada número de onda com um grupo funcional ou ligação entre átomos específicos (Figura 3.3) (Stuart, 2004).

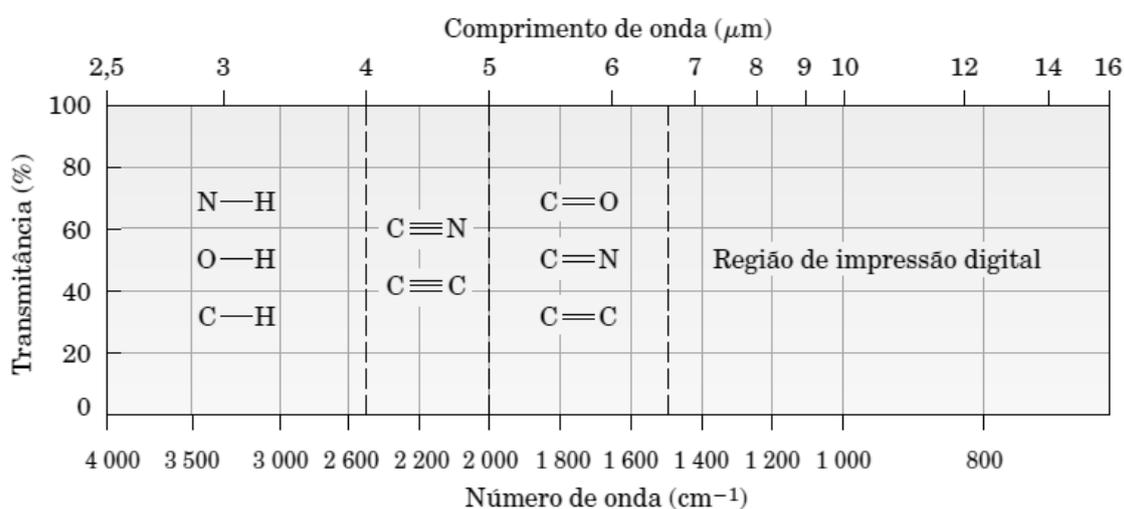


Figura 3.3. Regiões do espectro infravermelho e tipos de ligações ou grupos funcionais correspondentes
Fonte: McMurry (2008).

Entretanto, a espectroscopia no infravermelho tem sido muito utilizada de forma quantitativa quando associada a métodos de calibração multivariados. A utilização de NIR-IR e MID-IR tem sido cada vez mais comuns para análise quantitativa de, por exemplo, carbono, oxigênio e matéria orgânica em solos (Souza et al., 2012), também de adulteração em café (Tavares et al., 2012), ou leite (Ferrão et al., 2007).

Vantajosamente, a espectroscopia no infravermelho não necessita da abertura de amostra com a utilização de reagentes algumas vezes muito tóxicos ou em temperaturas muito altas. Outro ponto positivo é que a análise não destrói a amostra e necessita de uma pequena quantidade para a obtenção do espectro. Como em muitos

outros métodos espectroscópicos, está presente na análise ruídos experimentais e sobreposição de bandas. Algumas dessas desvantagens são minimizadas ou eliminadas pela utilização da transformada de Fourier acoplada a espectroscopia no infravermelho (FT-IR), que promove a medida de todas as frequências do infravermelho de modo simultâneo sem a necessidade de separação de sinais (Reeves et al., 2002).

3.3 Análise da biomassa pirolisada por espectroscopia no infravermelho

Para identificação e quantificação das amostras de biomassa a utilização da espectroscopia MID-IR é fundamental, pois à mesma corresponde à faixa de 4000 a 400 cm^{-1} , onde os grupos funcionais das substâncias orgânicas presentes na biomassa têm suas bandas características (Alisson, 2011). A partir da análise de infravermelho do carbono obtido no processo de pirólise, a observação de bandas específicas no espectro corresponde à presença de grupos funcionais de compostos lignocelulósicos comuns presentes na biomassa de partida. Com isso é possível identificar grupos funcionais associados a compostos presentes na biomassa e quantificar teores elementares através de métodos de calibração multivariada para avaliar as características do biocarvão estudado quanto às suas potencialidades de promoção da fertilidade do solo e captura de carbono atmosférico.

Os compostos lignocelulósicos que têm maior presença percentual na biomassa são celulose, hemicelulose e lignina (Figura 3.4). Esses três compostos têm uma porcentagem grande de composição da biomassa, sendo que 30 a 40% é de lignina, 40 a 50% é de celulose e aproximadamente 20% é de hemicelulose, (Tamanini; Haully, 2004). Há presença de diversos extrativos de baixo peso molecular como terpeno, ácidos graxos, óleos vegetais, entre outros que resultam na porcentagem total da biomassa.

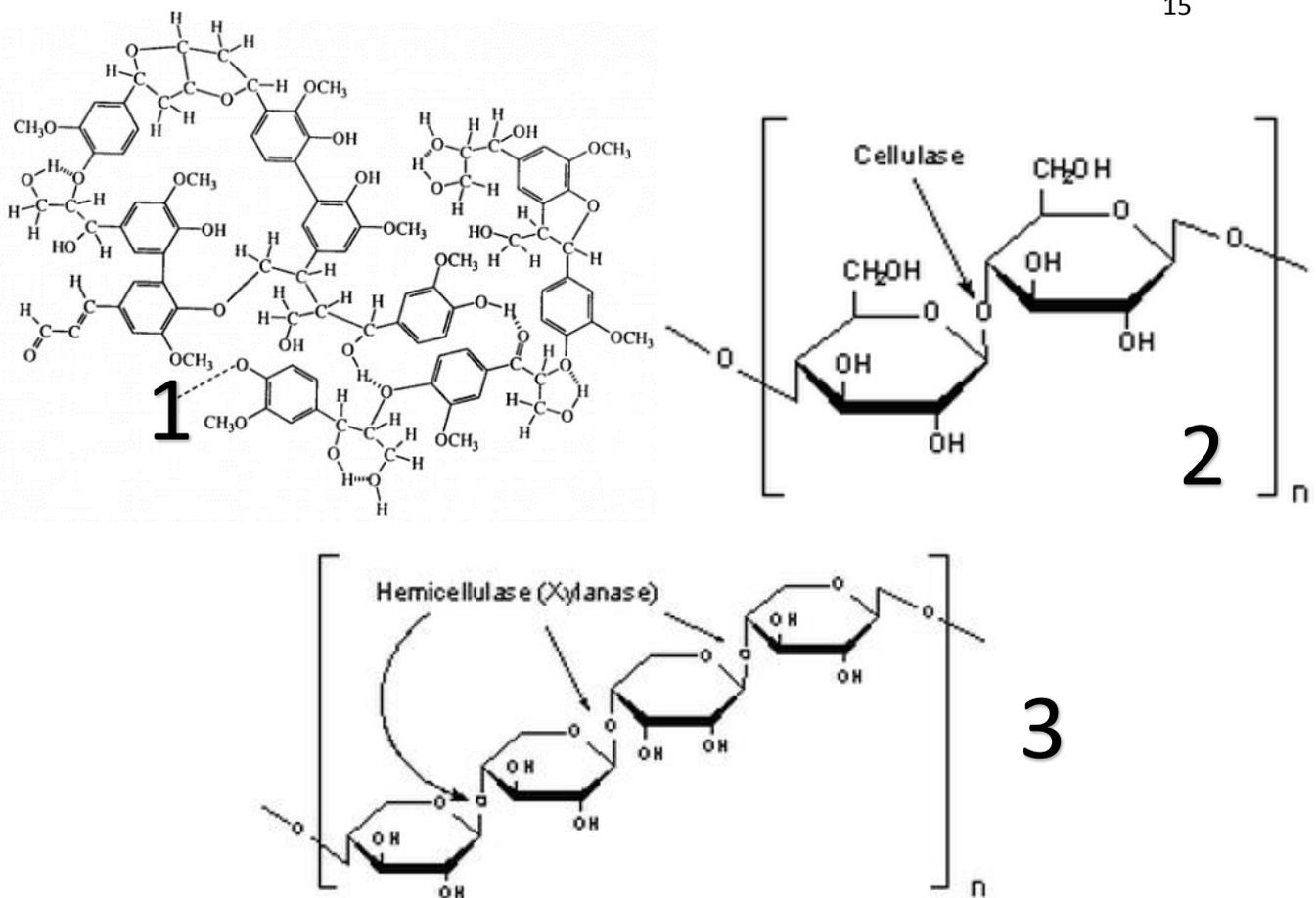


Figura 3.4. 1-Lignina, 2-Celulose, 3-Hemicelulose.

Fontes: <http://www.infoescola.com/compostos-quimicos/lignina>
http://www.bio.miami.edu/dana/226/226F07_3.html

Diversas bandas do espectro de biomassa têm relação direta a presença de traços de grupos funcionais, estiramentos e distorções de ligações C-H, C=O, C-C e O-H (Figura 3.5). Algumas bandas características como entre 2800 e 2900 cm^{-1} correspondem ao estiramento C-H, próximo a 3400 cm^{-1} resultante do estiramento O-H, e em 1500 cm^{-1} devido a vibração do anel aromático resultam de traços da presença de lignina. Próximo a 1700 cm^{-1} é possível observar os estiramento C=O de carbonila presente em grupos cetona, aldeído e éster proveniente de traços de hemicelulose, assim como também em 875 cm^{-1} é possível observar banda correspondente à ligação glicosídica. A celulose e a hemicelulose podem ser identificadas pelo estiramento assimétrico C-O-C de grupos éteres próximos a 1100 cm^{-1} , vibração de ligação C-H, e O-H no plano de curvatura ambos próximos a 1335 cm^{-1} (Xu et al., 2013).

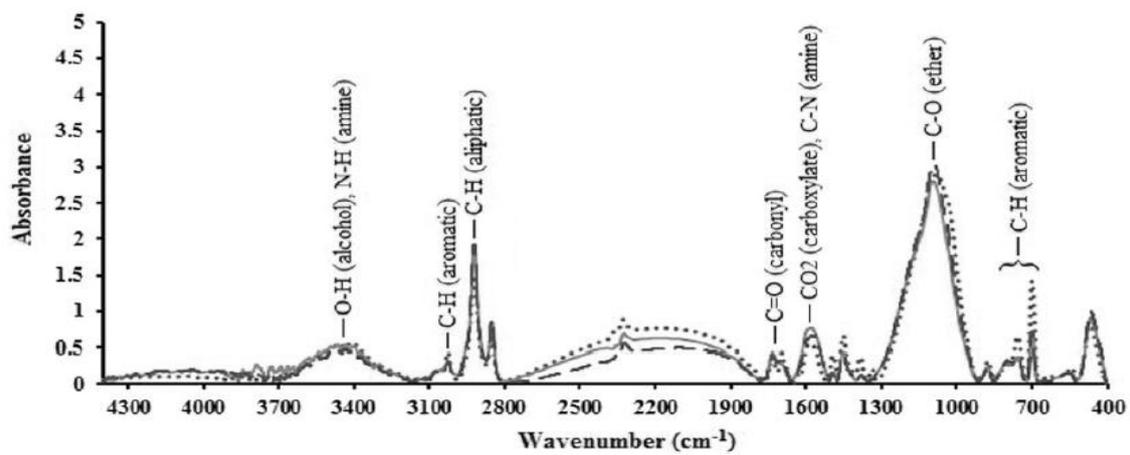


Figura 3.5. Espectro de infravermelho médio de uma amostra de biocarvão. Fonte: Chen (2014)

CAPÍTULO 4

PRÉ-PROCESSAMENTO DE DADOS, SELEÇÃO DE VARIÁVEIS E TÉCNICAS MULTIVARIADAS

O tratamento de dados estatísticos foi outrora um processo muitas vezes demorado e dispendioso quando havia a necessidade da resolução de problemas envolvendo conjunto dados complexos (complicados), tanto na esfera científica quanto na sua aplicação cotidiana. Até a década de 70 não havia computadores de mesa disponíveis para o tratamento de dados e os cálculos eram realizados por computadores de grande porte (Neto et al., 2006), tornando a obtenção de resultados dificultosa. Atualmente, a utilização dos recursos computacionais se tornou mais simples, possibilitando a utilização e o desenvolvimento de técnicas multivariadas, tornando o tratamento estatístico mais rápido e preciso.

Diversos modelos de calibração multivariada, como a MLR, LS-SVM ou ANN, podem ser encontrados e utilizados em uma infinidade de softwares de tratamento estatístico. Tais softwares têm ampla utilidade, apresentando modelos de calibração multivariada, como MLR e PLS, com aplicabilidade em muitas áreas científicas, tais como quimiometria, engenharia, economia, entre outras. Um exemplo disso é a sua utilização em Quimiometria, no qual métodos estatísticos são utilizados para a obtenção de informações de dados de origem química (Kowalski, 1980). Quando se diz dados de origem química, englobam-se todos os métodos instrumentais de análise, desde os mais clássicos, como análises volumétricas e gravimétricas, até as mais frequentemente utilizadas, como espectroscopia no infravermelho, ressonância magnética nuclear e cromatografia gasosa ou líquida.

Dentre as técnicas instrumentais mais atuais, a espectroscopia no infravermelho, tanto próximo quanto médio, fornecem análises de baixo custo e geração de resíduos nula. Essa excelente capacidade de análise somada aos modelos multivariados torna a obtenção de informações através dos dados de infravermelho uma tarefa mais simples para a avaliação das amostras estudadas.

Métodos de calibração multivariada como MLR, PCR e PLS são frequentemente utilizados, mas técnicas como ANN são muito úteis dependendo do tipo, do tamanho e da presença de não linearidade no conjunto de dados analisado. No auxílio aos métodos de calibração existem técnicas de pré-processamento de dados como MSC e SNV para o pré-tratamento do espectro, e redução de dimensionalidade como SPA.

4.1 Multiplicative Scatter Correction: Pré-processamento de dados

É pouco aconselhável realizar o tratamento dos dados espectroscópicos de forma direta em métodos multivariados devido à presença de diversos fatores que provocam a interferência, como por exemplo, a atividade do reagente de contraste, o posicionamento da amostra em relação ao sensor ou ruídos da linha de base. Esses exemplos são, respectivamente, de interferentes químicos, físicos e instrumentais, sendo que existem ainda muitos outros que podem provocar erros no tratamento. Assim, é necessário à utilização de técnicas de pré-processamento de espectro que minimizem a influência dos interferentes para melhorar a aplicação de modelos multivariados.

Existem muitas técnicas de pré-processamento de espectros, algumas mais simples como 1ª derivada e 2ª derivada, e algumas mais complexas, como alisamento Savitsky-Golay, SNV e MSC. Dentre esses, o mais utilizados para o tratamento de dados de infravermelho é o MSC em razão da sua boa capacidade de diminuição dos efeitos provocados pela dispersão da luz decorrente das partículas das amostras devido à utilização das mesmas na forma de pó, além do ajuste da linha de base (Rinnan et al., 2009).

A técnica de pré-processamento espectroscópico MSC se baseia na regressão do espectro medido contra o espectro de referência, que geralmente é o espectro médio, e, em seguida, corrigindo o espectro medido com a utilização do coeficiente angular e, em alguns casos, até mesmo o coeficiente linear, obtido da regressão (Rinnan et al., 2009).

$$\mathbf{x}_{org} = b_0 + \mathbf{b}_{ref,1} \mathbf{x}_{ref} + \mathbf{e} \quad (2)$$

A Eq. 2 é utilizada para estimativa dos coeficientes de correção b_0 e $b_{ref,1}$, onde x_{org} é o espectro experimental, x_{ref} é o espectro de referência, e é o resíduo. Com a obtenção dos coeficientes o espectro medido poder ser corrigido com base no mesmo nível de espalhamento do espectro de referência,

$$x_{corr} = \frac{x_{org} - b_0}{b_{ref,1}} = x_{ref} + \frac{e}{b_{ref,1}} \quad (3)$$

sendo x_{corr} o espectro corrigido (Rinnan et al., 2009).

4.2 SPA, GA e iPLS na seleção de variáveis

Os dados espectroscópicos de análises quantitativas em infravermelho necessitam de técnicas de calibração multivariada para a determinação simultânea de vários analitos. Para que essas técnicas multivariadas possam fornecer modelos bem ajustados é necessária a aplicação de métodos de seleção de variáveis.

Dentre os métodos de seleção de variáveis mais utilizados atualmente está o GA, proposto por John Henry Holland em 1975, que utiliza técnicas de busca randômica baseadas nos mecanismos da seleção natural de Darwin. O GA possui inúmeras variações a partir de seu algoritmo original, todos eles a fim de atender a seleção de variáveis de um determinado tipo de sistema ou conjunto de dados. O algoritmo original do GA, assim como a suas variações, utiliza operadores que representam os mecanismos da seleção natural para obtenção somente dos indivíduos mais adaptados, ou seja, os melhores dados, resultados ou informações possíveis a partir do conjunto de dados original (Ahn, 2006). O GA pode ser esquematizado no fluxograma da Figura 4.1 e resumido nas seguintes etapas:

1. Por meio do operador seleção natural, os melhores indivíduos da população inicial são escolhidos e mantidos.
2. Os melhores indivíduos realizam o cruzamento ou acasalamento entre si.
3. Uma pequena quantidade dos indivíduos gerados pelo cruzamento sofrem mutação, ou seja, ocorrem mudança em suas características.
4. O operador seleção natural escolhe novamente os indivíduos mais adaptados e remove os menos adaptados.

5. Se os indivíduos escolhidos pelo operador seleção natural forem os mais adaptados possíveis, o algoritmo termina, se não, o processo é realizado novamente com a nova população. (Mitchel, 1999; Ahn, 2006)

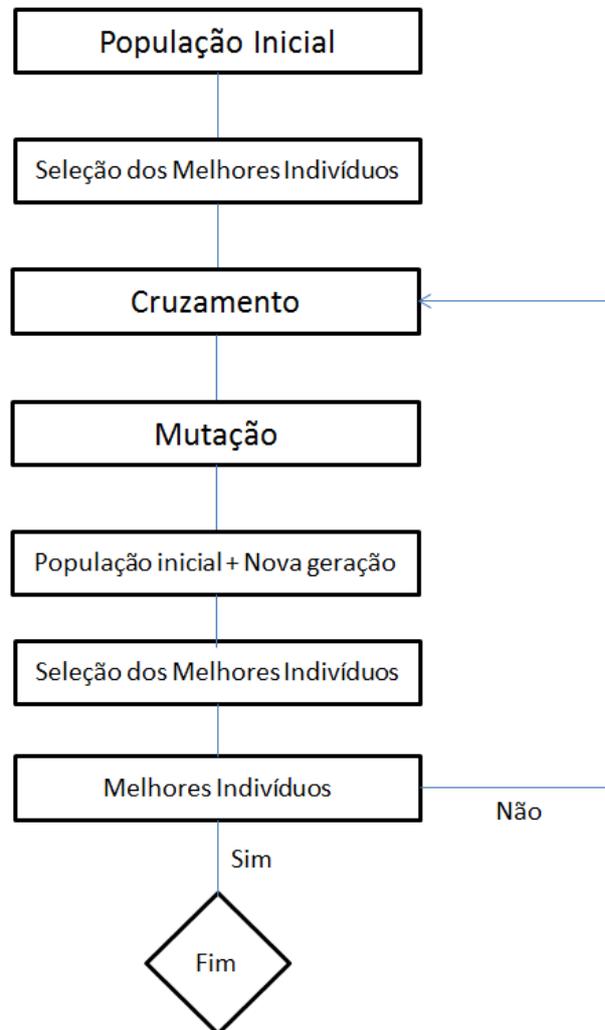


Figura 4.1. Fluxograma resumido das etapas do Algoritmo Genético (GA).

A idéia para a utilização desse método, em termos do tratamento estatístico dos dados, seria a busca e seleção de variáveis que forneçam os menores erros possíveis na etapa de validação cruzada da calibração multivariada. Entretanto, devido a essa própria busca randômica, os resultados são dependentes dessa mesma busca e a seleção de variáveis pode não ser reproduzível (Araújo et al., 2001).

Semelhantemente ao GA, iPLS é também um método de seleção de variáveis, mas não executa a busca dados de forma randômica, realizando a busca de variáveis de interesse ou a combinação dessas apenas de forma sequencial. O “intervalo” escolhido pelo iPLS não é somente um faixa específica de dados, como pode ser também um única variável, sendo assim o mesmo pode escolher um subconjunto formado por intervalos de variáveis e variáveis isoladas (Norgaard et al, 2000).

O algoritmo do iPLS se baseia na busca contígua de números de onda ou intervalos de números de onda que contribuam para a obtenção do melhor valor de RMSECV (Raiz do Erro Médio Quadrado de Validação Cruzada) do espectro total. É possível também a utilização de outros parâmetros como o coeficiente de determinação R^2 ou o coeficiente angular da reta obtida pela comparação entre a resposta predita e a resposta experimental para a avaliação da busca de variáveis (Norgaard et al, 2000). As buscas podem ser realizadas de duas formas: “para frente” (forward) e “reverso” (backward). No modo “para frente” as variáveis são selecionadas pela criação de modelos PLS individuais dentro de um intervalo pré-definido e avaliação dos melhores valores de RMSECV com cada conjunto possível de variáveis presentes no intervalo. Já no modo “reverso” todos os intervalos são inicialmente selecionados e o algoritmo escolhe somente um para ser descartado por vez, no qual com descarte deste se obtenha um melhor valor de RMSECV (Leardi; Norgaard, 2004).

Em relação ao GA e semelhantemente ao iPLS, o SPA é uma técnica de seleção de variáveis mais recente (Araújo et al., 2001), mas tem crescido em utilização ao longo dos últimos anos. É uma técnica que tem como vantagens o reduzido custo computacional quando comparado com GA, e a capacidade de ser acoplada com técnicas como MLR. O SPA é um método iterativo, que partindo do primeiro valor de número de onda incorpora um novo valor a cada iteração, sendo esses os valores que levem a um menor valor do RMSE e que possuam menor colinearidade entre si. Essa seleção é necessária para que os comprimentos de onda escolhidos não sofram com problemas de colinearidade e a informação obtida não seja redundante (Galvão et al., 1999).

Por se tratar de um algoritmo, o SPA segue um procedimento que pode ser resumido em passos a serem seguidos partindo-se do primeiro número de onda (Figura 4.2):

Passo 0: Antes da primeira iteração ($n = 1$), \mathbf{x}_j representa os valores da coluna j da matriz \mathbf{X} ; $j = 1, \dots, J$.

Passo 1: S será o conjunto de comprimentos de onda que não foram selecionados ainda. $S = \{j | 1 \leq j \leq J \text{ e } j \notin \{k(0), \dots, k(n-1)\}\}$.

Passo 2: Calcular a projeção de \mathbf{x}_j sobre o subespaço ortogonal para $\mathbf{x}_{k(n-1)}$ como

$$\mathbf{P}\mathbf{x}_j = \mathbf{x}_j - (\mathbf{x}_j^T \mathbf{x}_{k(n-1)}) \mathbf{x}_{k(n-1)} (\mathbf{x}_{k(n-1)}^T \mathbf{x}_{k(n-1)})^{-1}$$

para todo $j \in S$, onde $\mathbf{P}\mathbf{x}_j$ é o operador projeção.

Passo 3: Determinar o próximo número de onda com projeção máxima, $k(n) = \arg(\max \|\mathbf{P}\mathbf{x}_j\|, j \in S)$.

Passo 4: Igualar \mathbf{x}_j ao operador projeção $\mathbf{P}\mathbf{x}_j$; $\mathbf{x}_j = \mathbf{P}\mathbf{x}_j, j \in S$.

Passo 5: Faz $n = n + 1$. Se $n < N$, onde N é o número específico de comprimentos de onda a ser alcançada, volta-se ao passo 1.

Final: Os comprimentos de onda resultantes são $\{k(n); n = 0, \dots, N - 1\}$, (Araújo et al., 2001).

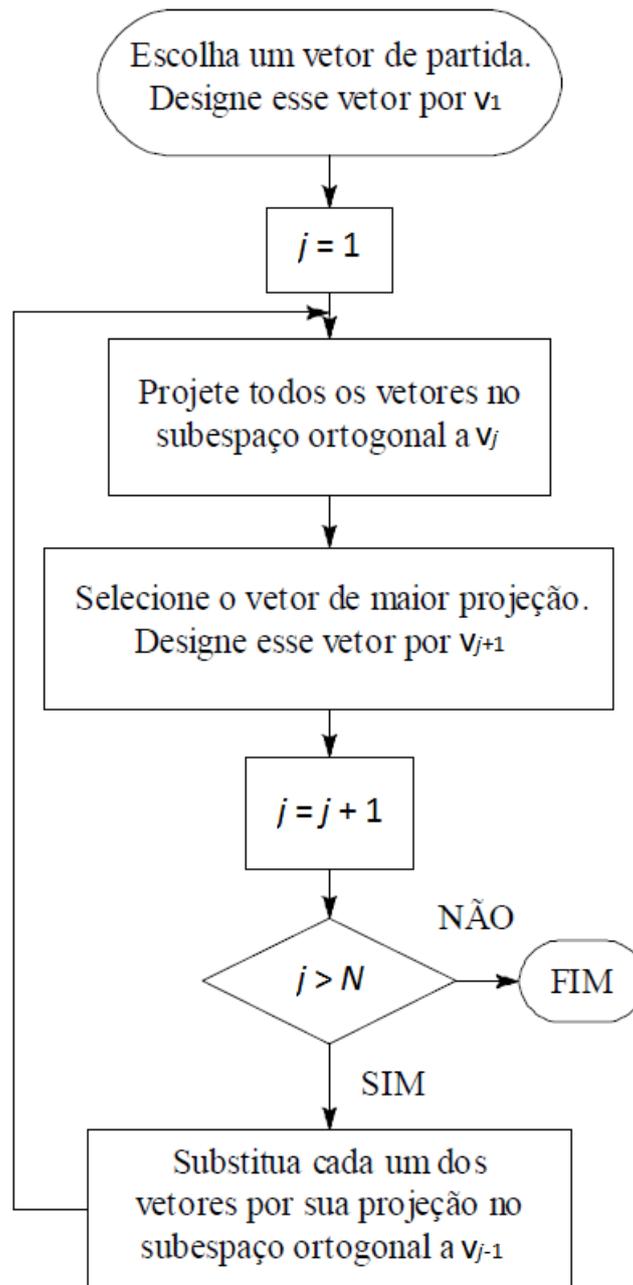


Figura 4.2. Algoritmo das Projeções Sucessivas (SPA)
Fonte: Galvão (1999).

Após o último passo do algoritmo os comprimentos de onda selecionados podem ser utilizados em uma técnica de calibração multivariada para a obtenção do modelo, predição através do conjunto de validação e cálculo da Raiz Quadrática do Erro Médio (RMSE – Root Mean Square Error).

4.3 Técnicas de Calibração Multivariada

Calibração por MLR

A MLR é o método de calibração multivariada mais simples e um dos mais utilizados para o tratamento de dados em geral. O método é uma extensão da regressão linear simples, onde ao invés de uma variável independente x podem existir várias variáveis independentes $x_1, x_2, x_3, \dots, x_n$ que se relacionam linearmente com uma variável dependente y . Por meio da regressão pode-se encontrar o modelo, que é combinação linear das variáveis independentes, que representa o sistema.

$$y = b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (6)$$

É importante frisar que MLR oferece bons resultados quando é utilizada em um conjunto de dados que apresenta a característica de projeção linear, ou seja, quando os dados se comportam de forma linear, quando existem baixas interferências experimentais e quando pouca colinearidade entre as variáveis independentes (Sena et al., 2000). Já para sistemas não lineares MLR não é recomendada visto que existem outras técnicas que fornecem resultados mais satisfatórios como a SVM. As desvantagens da MLR são que o número de amostras deve ser maior que o número de variáveis e que a regressão também pode acrescentar ocasionalmente ao modelo final informação de variáveis que tem pouca influência no sistema.

Calibração por PLS

Diferentemente da RLM, onde o número de amostras deve ser maior que o número de variáveis, a regressão por PLS não tem esse pré-requisito. Introduzido pelo estatístico sueco Herman Wold, o PLS é um método estatístico que obtêm o modelo de regressão linear projetando as variáveis experimentais e preditas em um novo espaço (Wold, 1985).

Assim como a regressão linear apresenta um modelo geral para o método multivariado, a regressão por PLS também possui um modelo genérico (Rosipal; Kramer, 2006),

$$X = TP^T + E \quad (7)$$

$$Y = UQ^T + F \quad (8)$$

onde X é a matriz de predições, Y a matriz de resposta, T e U são as matrizes de projeção de X e Y respectivamente, P e Q as matrizes ortogonais, E e F são as matrizes do erros. O algoritmo PLS é um processo iterativo, no qual os vetores de contagem, também chamados de variáveis latentes, T e U são sucessivamente extraídos das matrizes de predição X e resposta Y (Figura 4.3), presentes na amostra tal que a covariância entre eles seja maximizada, de modo a prever as respostas do conjunto de calibração (Tobias, 1997; Maitra, Yan, 2008). Assim as variáveis latentes serão utilizadas substituindo as variáveis originais avaliando-se até onde a troca ainda é representativa para o conjunto de dados estudado. Então, o número ótimo de variáveis latentes será aquele capaz de produzir o menor RMSECV (Raiz Quadrada do Erro Médio de Validação Cruzada).

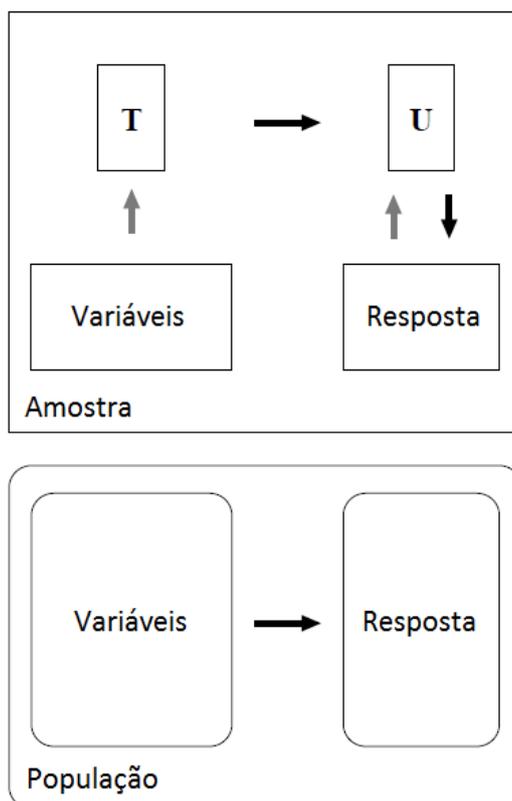


Figura 4.3. Modelo esquemático PLS
Fonte: adaptado de Tobias (1997).

O algoritmo PLS original tem uma forma e resolução simples, mas existem diversas variações desse algoritmo. Algumas variações do algoritmo PLS são utilizadas para estimar os fatores pela combinação linear entre as variáveis originais (Jong, 1993).

4.4 Cálculo dos erros

Para determinar o nível de eficiência dos modelos de calibração multivariada é necessário o cálculo da Raiz Quadrada do Erro Médio (RMSE – Root Mean Square Error),

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^{(\text{modelo})} - y_i^{(\text{real})})^2}{n}} \quad (9)$$

onde $y_i^{(\text{modelo})}$ e $y_i^{(\text{real})}$ são respectivamente os valores de predição e real para a variável de importância e n é o número de amostras. Esse valor expressa a exatidão do modelo, sendo assim compara a proximidade do valor calculado pelo modelo, o $y_i^{(\text{modelo})}$, com o valor verdadeiro ou obtido por um método de referência, $y_i^{(\text{real})}$. O RMSE pode ser calculado na calibração, sendo chamado de RMSEC (Root Mean Square Error of Calibration – Raiz Quadrada do Erro Médio de Calibração), na etapa de validação cruzada sendo chamado de RMSECV (Root Mean Square Error of Cross Validation – Raiz Quadrada do Erro Médio de Validação Cruzada), e na predição, sendo chamado de RMSEP (Root Mean Square Error of Prediction - Raiz Quadrada do Erro Médio de Predição).

4.5 Parâmetro RPD (Residual Prediction Deviation)

O parâmetro RPD é utilizado para avaliar a precisão e exatidão de um modelo obtido por determinada técnica de calibração multivariada, servindo assim como um método de controle de qualidade. Ele determina quanto um modelo de calibração é capaz de prever outros conjuntos de validação seguintes (Igne, Hurburgh, 2010).

O RPD pode ser calculado através do desvio padrão entre os valores medidos e os valores preditos, podendo também relacionado com o coeficiente de determinação do grupo de validação estudado, dado por:

$$\frac{dp(y)}{dp(\hat{y}-y)} = RPD \approx \sqrt{\frac{1}{1-r^2}} \quad (10)$$

onde dp é o desvio padrão, \hat{y} é o valor predito pelo modelo, y é o valor medido e r^2 é o coeficiente de determinação da etapa de predição (Igne, Hurburgh, 2010).

O valor obtido a partir da Equação 10 pode ser avaliado utilizando-se a escala fornecida por Williams, onde valores maiores que 5 são aceitáveis para controle de qualidade, com valores maiores que 6,5 o modelo utilizado pode ser utilizado para monitoramento de processos e com valores maiores que 10 o modelo pode ser utilizado em diversas aplicações (Williams, 2001).

4.6 Análise de Componentes Principais (PCA)

Igualmente às técnicas GA e SPA, que visam a diminuição de um conjunto de dados existentes para a obtenção de um grupo reduzido que tenham melhor capacidade de representar o sistema, o PCA tem por objetivo a redução da dimensionalidade dos dados diminuindo a multicolinearidade entre as variáveis existentes, além da capacidade do reconhecimento de padrões entre as amostras estudadas (Correia; Ferreira, 2007), complementando assim técnicas como HCA (Análise Hierárquica de Cluster) que propiciam a separação das variáveis existentes no conjunto total de dados em diferentes grupos com características específicas.

O PCA se baseia na redução da dimensionalidade através da compressão dos dados (Mello, 1998), resultando assim em variáveis chamadas PCs (Principal Components – Componentes Principais) que substituem as variáveis originais para a construção do modelo multivariado por conterem maior capacidade de informação que explique o sistema estudado. As PCs são obtidas através da diagonalização da matriz de dados original sendo assim ortogonais (não são correlacionadas), onde a primeira PC contém a maior parte da variância total e as PCs subsequentes valores menores (Ferreira, 2008). A PCA se baseia na seguinte equação que representa a diagonalização da matriz original:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T \quad (11)$$

sendo \mathbf{X} a matriz original de dados, \mathbf{T} a matriz de *scores* e \mathbf{P} a matriz de *loadings*. A matriz de *scores* é a projeção das amostras no novo sistema de coordenadas e os *loadings* são os coeficientes que determinam o peso de cada variável da matriz original em cada PC obtida.

Na representação gráfica da matriz de *scores* é apresentada a separação das variáveis em grupos diferentes, revelando assim um padrão entre as mesmas, podendo promover a distinção entre as amostras e as caracterizando em importância para o conjunto total estudado (Correia; Ferreira, 2007). A matriz de *loadings* apresenta o peso de cada PC sobre a variância total do sistema, podendo ser representada pelo somatório do produto do peso p pelas suas variáveis correspondentes x_i , dada a Equação 12 para n variáveis do espaço residual.

$$PC_i = x_{i1}p_1 + x_{i2}p_2 + \dots + x_{in}p_n \quad (12)$$

4.7 Análise Hierárquica de Cluster (HCA)

A Análise Hierárquica de Cluster (HCA) tem por objetivo realizar o agrupamento de amostras em um conjunto de dados levando-se em conta a suas similaridades e propriedades, formando assim grupos com diferentes padrões entre si. O método é baseado nas diferentes distâncias entre cada variável individualmente ou agrupamento de variáveis. O resultado da análise é mostrado na forma de um dendrograma, observando a separação de grupos entre as amostras e as ligações entre elas dentro do grupo (Chatfield; Collins, 1980). Existem diversos métodos que servem como avaliação para o agrupamento das amostras, tendo por princípio a ponderação das distâncias, tanto primeiramente entre amostras quanto posteriormente entre aglomerados. Na Figura 4.4 é exemplificada a estrutura de um dendrograma hipotético, onde as letras no eixo das abscissas representam cada amostra do conjunto total, e os números presentes no eixo das ordenadas seriam os valores referentes ao tipo de método utilizado para o cálculo de avaliação das distâncias.

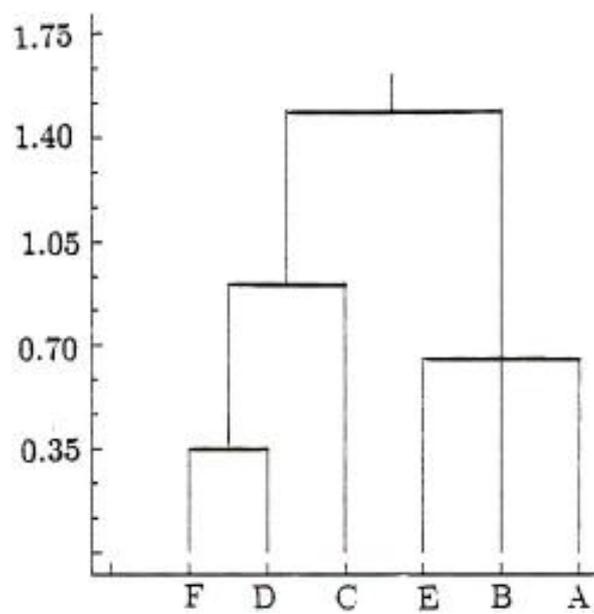


Figura 4.4. Exemplo de dendrograma obtido pela análise cluster.
Fonte: Introdução à Análise de Agrupamentos (1990)

CAPÍTULO 5

Materiais e Métodos

5.1 Amostragem

No presente trabalho foi realizado o tratamento de dados de espectros de 11 amostras de biomassas (material lignino-celulósico, madeiras) in natura e de 33 tipos de biocarvões produzidos dessas mesmas biomassas. As amostras de biomassa antes da pirólise foram secas em estufa a 110°C por 24 horas, moídas e peneiradas em peneiras à 2 milímetros. A pirólise para a produção dos carvões foi conduzida em forno mufla adaptado, em baixa concentração de oxigênio, utilizando as seguintes temperaturas de 350, 450 e 550°C, taxa de aquecimento 10°C min⁻¹, isoterma de 60 minutos na temperatura final. As biomassas utilizadas foram:

- Aguapé, origem: Bocaiúva do Sul, Paraná.
- Bambu Guadua, origem: Guarapuava, Paraná.
- Bambu Phyllostachys aurea, origem: Guarapuava, Paraná.
- Candeia Eremanthus erythropappus, origem: Minas Gerais.
- Casca de Macadâmia, origem: Rio de Janeiro.
- Eucalyptus dunnii, origem: Colombo, Paraná.
- Eucalyptus urophylla, origem: Seropédica, Rio de Janeiro.
- Pinus caribea, origem: Colombo, Paraná.
- Pinus taeda, origem: Seropédica, Rio de Janeiro.
- Serragem de madeiras tropicais nativas, origem: Rio Branco, Acre.
- Serragem de madeiras tropicais nativas, origem: Sinop, Mato Grosso.

5.2 Espectroscopia MID-IR

Os espectros de absorção de FT-IR na região do infravermelho médio foram obtidos pela utilização de um espectrômetro Varian 660-IR FT-IR (Agilent) utilizando o acessório EasyDiff (Pike). A região utilizada do espectro foi de 4000~400 cm⁻¹ (MID-IR), realizando 64 varreduras por espectro com resolução de 2 cm⁻¹. As amostras (1%) foram diluídas em KBr grau espectroscópico. Os espectros foram registrados por absorbância.

Todas as análises foram realizadas na EMBRAPA Arroz e Feijão, na Rodovia GO-462, Km 12, Fazenda Capivara, município de Santo Antônio de Goiás, no estado de Goiás.

5.3 Recursos Computacionais

Para o pré-tratamento dos dados, agrupamentos de espectros, calibrações multivariadas e identificação de padrões foram utilizados: um microcomputador com Processador Intel® i3-540 3.0 GHz, 4.0 Gbytes de memória RAM; e os softwares MATLAB™ 2014 (The MathWorks, Natick, EUA) com o emprego do pacote PLS Toolbox 2.0 (Wise et al., 2001) e Unscrambler 9.7 ® Free Trial (CAMO Software, Norway), presente nos computadores do Laboratório de Química Teórica e Computacional do Instituto de Química da Universidade Federal de Goiás (IQ-UFG), e SPA (Galvão et al., 2012) obtido do website da Divisão de Engenharia Eletrônica do Instituto Tecnológico de Aeronáutica (IEE-ITA).

5.4 Separação de amostras para calibração e validação

Os grupos de calibração e validação foram separados utilizando o algoritmo de Kennard-Stone (Kennard, Stone, 1969) obtido do toolbox TOMCAT (Daszykowski et al., 2007), que separa os dados de acordo com a distância euclidiana existente entre eles em relação à média, podendo-se partir da maior para a menor distância, ou vice-versa até que o número desejado para a validação seja obtido. Para o presente trabalho 70% do dados (31 amostras) foram utilizados para a calibração e validação cruzada, e 30% (13 amostras) foram utilizados para a validação. Os conjuntos de calibração e validação obtidos pela utilização do algoritmo de Kennard-Stone para os teores elementares foram os mesmos para todas as propostas de quantificação estudadas, exceto quando houve a utilização do SPA.

Na utilização do SPA na seleção de variáveis, a separação por Kennard-Stone foi aplicada novamente sobre o conjunto de calibração previamente obtido como descrito no parágrafo anterior para a divisão do mesmo em um novo grupo de calibração e um grupo para a realização de testes pelo SPA obtendo assim a seguinte separação das amostras em relação ao número total (44 amostras): 40% dos dados para a calibração

(18 amostras), 30% para testes realizados durante a seleção de variáveis (13 amostras) e 30% para a validação dos modelos obtidos (13 amostras).

5.5 Parâmetros utilizados nas técnicas de seleção de variáveis GA e iPLS

O método utilizado para a seleção de variáveis no GA foi a avaliação dos melhores valores de RMSECV obtidos dos espectros através das regressões feitas (MLR ou PLS). Os parâmetros utilizados para a seleção de variáveis por GA podem ser observados na Tabela 5.1 para os dois tipos de regressão utilizados.

Tabela 5.1. Tabela de parâmetros utilizados para seleção de variáveis por GA.

Parâmetros	Método de Regressão	
	MLR - CV	PLS - CV
População inicial	100	100
Quantidade inicial de variáveis	10%	30%
Número máximo de gerações	100	100
Tipo de cruzamento	Duplo	Duplo
Taxa de mutação	0,5%	0,5%
Validação cruzada	Aleatória em blocos de 5 amostras	Aleatória em blocos de 5 amostras
Número de Máximo de Variáveis Latentes	-	15
Pré-processamento	Não	Não

O iPLS realizou a busca de variáveis através da avaliação e escolha do número de onda que contribuíssem para a obtenção do melhor valor total de RMSECV de absorbâncias correspondentes obtidas do espectro completo. Para a utilização do iPLS, 5 parâmetros foram utilizados: modo de busca, número de intervalos, tamanho dos intervalos, tamanho do passo e número máximo de variáveis latentes. O modo de busca utilizado foi o modo “*forward*”, o número de intervalos entre 20 e 30, o tamanho dos intervalos entre 1 e 5, o tamanho dos passos foi utilizado em modo automático e o número máximo de variáveis latentes foi 10.

5.6 Tratamento Estatístico dos Dados

Os espectros de infravermelho médio foram tratados e avaliados através do método de pré-tratamento espectral MSC com a utilização do GA, SPA e iPLS para a seleção de variáveis e técnicas de calibração multivariada MLR e PLS. Assim, os dados foram tratados através das seguintes propostas:

1. SPA – MLR
2. PLS
3. GA – MLR
4. GA – PLS
5. iPLS

Proposta 1 – Seleção de Variáveis por SPA e calibração por MLR

Essa proposta consiste na redução da dimensionalidade dos dados pela seleção de variáveis utilizando o Algoritmo das Projeções Sucessivas (SPA). O critério do número máximo de variáveis selecionadas pelo SPA foi 15. A MLR foi realizada através da ferramenta fornecida pelo pacote SPA. Para essa proposta e para as demais seguintes foram utilizados como método de avaliação o RMSEC, RMSEP, RMSECV, os coeficientes R^2 da calibração e da validação, valor obtido de Bias e o parâmetro RPD.

Proposta 2 – Calibração por PLS

Nesta proposta, os dados são tratados diretamente pela regressão em PLS sem a prévia seleção de variáveis. Para a calibração por PLS, e para os demais métodos envolvendo a utilização do PLS para a obtenção do modelo foi utilizada a validação cruzada para a averiguação do modelo mais adequado com a opção de subconjuntos aleatórios. O máximo de variáveis latentes selecionadas foi de 10, sendo os dados separados em 5 e o número de interações em 5.

Proposta 3 – Seleção de Variáveis por GA com validação cruzada em MLR e calibração por MLR

É utilizada a técnica de seleção de variáveis GA com o método de avaliação de busca por validação cruzada da MLR. A calibração multivariada foi realizada utilizando a MLR.

Proposta 4 – Seleção de Variáveis por GA com validação cruzada em PLS e calibração por PLS

O GA para a seleção de variáveis também foi realizado utilizando como método de avaliação a validação cruzada por PLS. A calibração multivariada foi realizada utilizando a regressão por PLS, comparando os valores de RMSEC e RMSECV para definição do modelo mais adequado.

Proposta 5 – Seleção de Variáveis por iPLS e calibração por PLS

Para essa proposta, a seleção de variáveis foi realizada com a utilização do iPLS (Interval Partial Least Square), com o número máximo de 10 variáveis latentes escolhidas para a avaliação dos intervalos.

CAPÍTULO 6

TRATAMENTO ESPECTRAL E ANÁLISE DE CLUSTER

6.1 Identificação de compostos das biomassas no espectro MID-IR

A caracterização dos compostos principais presentes na biomassa foi realizada a partir do espectro total, contendo todos os espectros individuais de todas as amostras de biomassas e biocarvões analisadas pela espectroscopia no Infravermelho Médio. Os espectros originais podem ser observados na Figura 6.1.

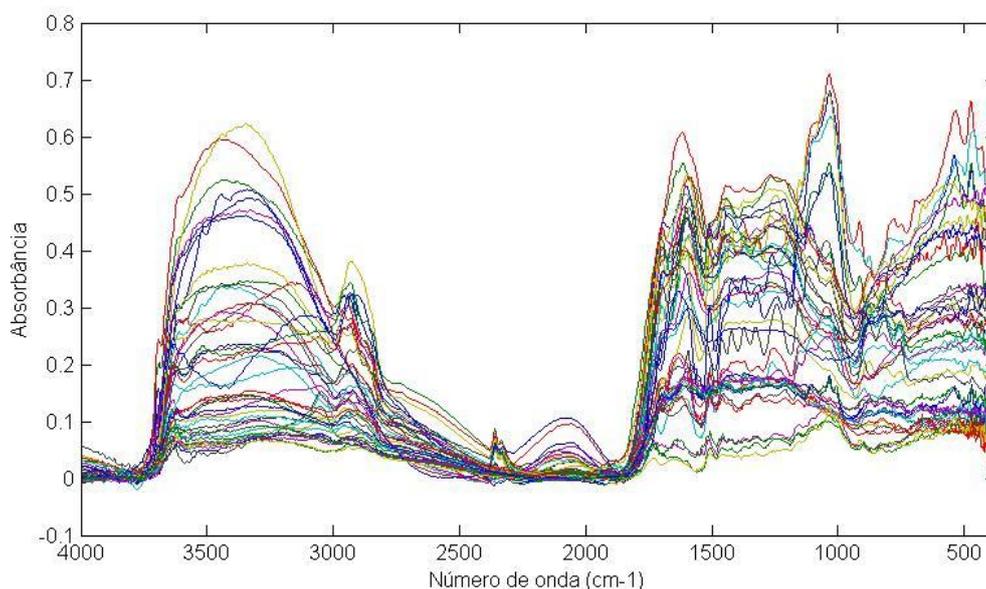


Figura 6.1. Espectros originais de infravermelho médio das amostras de biomassas e biocarvões.

Como mencionado no capítulo 3, os principais componentes das biomassas e que influenciam nas propriedades finais do biocarvão produzido correspondem lignina, celulose e hemicelulose. Ao observar o espectro, pode-se encontrar uma banda próxima a 3400 cm⁻¹ associada ao estiramento da ligação O-H presente na lignina. A lignina também pode ser caracterizada por uma banda de menor intensidade com número de onda próximo a 2900 cm⁻¹ associado a um estiramento de ligação C-H, além de uma outra banda entre 1600 e 1500 cm⁻¹ oriunda de um modo vibracional do anel aromático (Xu et al., 2013).

Em 2100 cm^{-1} também é associado à presença da celulose pela banda representativa da combinação das bandas de O-H e C-O. Próximo a 1400 cm^{-1} ocorre o indicativo da presença de lignina, celulose e hemicelulose pela deformação angular assimétrica no plano da ligação O-H e na ligação C-H. Outra banda característica está entre 1300 e 1200 cm^{-1} correspondente a deformação angular simétrica na ligação CH_2 da celulose e da hemicelulose. Outras caracterizações da celulose e da hemicelulose podem ser indicadas pela presença de uma em 1100 cm^{-1} devido ao estiramento assimétrico C-O-C de grupos éteres. A ligação glicosídica da hemicelulose é caracterizada pela absorção no número de onda de 800 cm^{-1} (Xu et al., 2013).

6.2 Pré-tratamento espectral

Os dados espectrais de todas as amostras obtidas por espectroscopia de infravermelho médio foram previamente tratados antes da utilização do método de seleção de variáveis por SPA, GA e iPLS e dos métodos de calibração multivariadas MLR e PLS. Para o tratamento dos espectros foi utilizada a técnica MSC, para correção de erros provocados por problemas de dispersão de luz, muito comuns em varreduras espectrais por refletância difusa, onde a radiação é incidida diretamente na amostra em pó, além da correção de desvios de linha de base (Rinnan et al., 2009). Os espectros corrigidos por MSC podem ser observados na Figura 6.2.

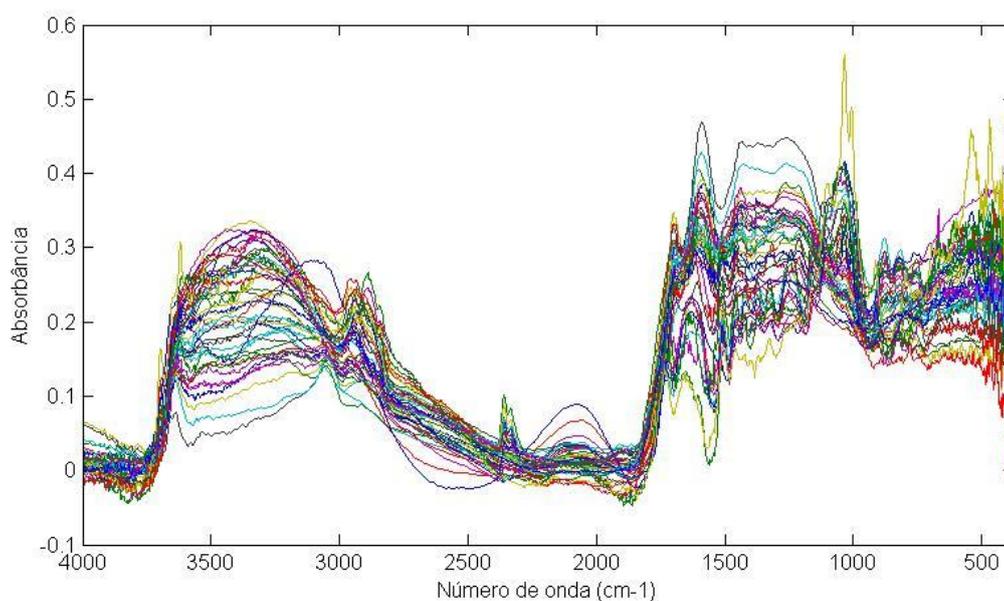


Figura 6.2. Espectro de infravermelho médio das biomassas e biocarvões corrigido por MSC.

6.3 Análise Hierárquica de Cluster

Os espectros corrigidos por MSC foram submetidos a uma Análise Hierárquica de Cluster (HCA) para a avaliação de padrões existentes entre as amostras de biomassas e de biocarvões estudadas. A HCA foi realizada utilizando o Método de Ward, em que os agrupamentos são aglomerados no sentido de minimizar a soma dos quadrados os desvios de cada centroide em relação ao próprio grupo, obtendo assim dendrogramas bem estruturados (Neto et al., 2006). Na Figura 6.3 mostra o dendrograma obtido a partir dos espectros de infravermelho das amostras de biomassas e biocarvões.

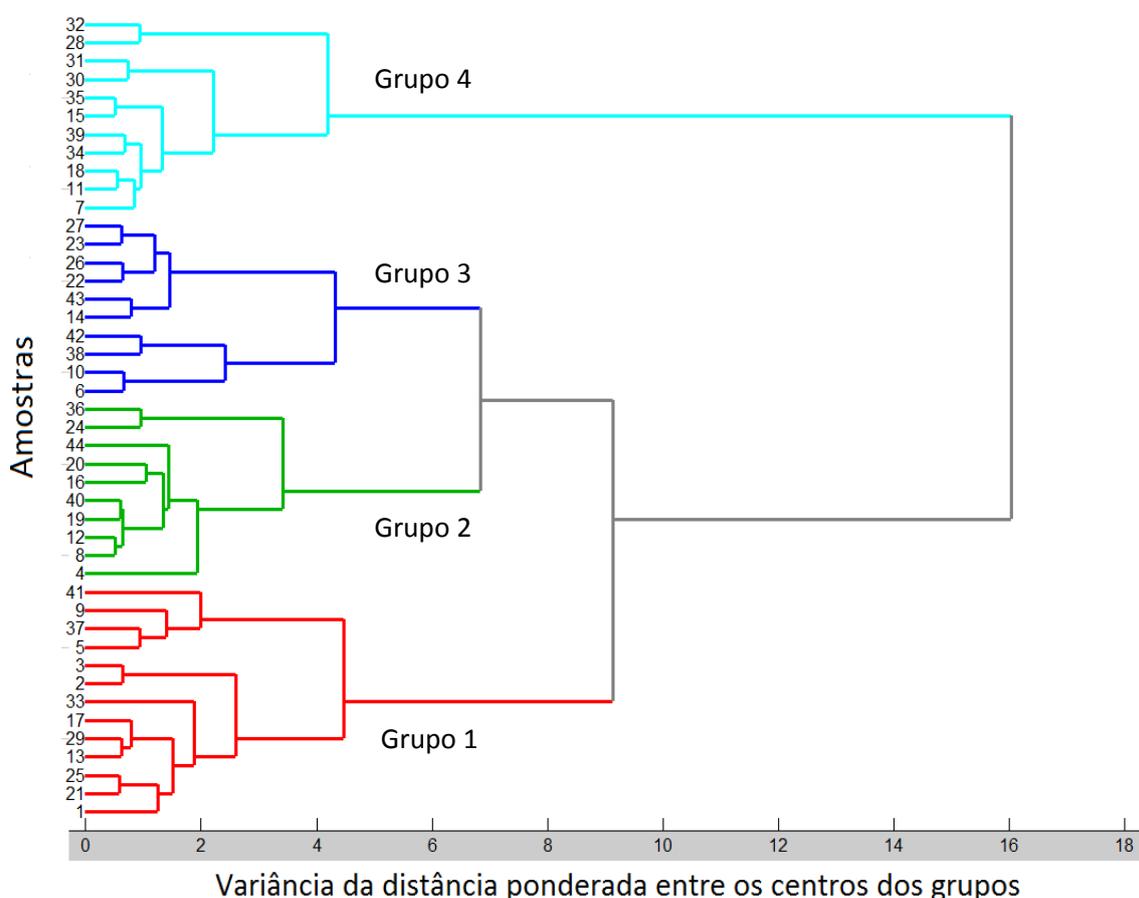


Figura 6.3. Dendrograma obtido pela HCA para as amostras de biomassas e biocarvões.

Através da análise do dendrograma, é possível observar a divisão do conjunto de amostras em 4 grupos distintos e bem estruturados. Os números existentes no eixo das ordenadas no dendrograma representam as amostras de biomassas e biocarvões listadas na Tabela 6.1.

Tabela 6.1. Tabela de análise elementar (%) das amostras de biomassas e biocarvões.

Fonte: adaptado de Maia (2013).

Número	Amostras	Análise Elementar (%)			
		C	H	N	O
1	Aguapé	34,45	5,25	0,82	59,48
2	Aguapé 350 °C	38,26	3,11	1,61	57,02
3	Aguapé 450 °C	39,04	2,41	1,5	57,05
4	Aguapé 550 °C	38,71	1,59	1,33	58,37
5	Bambu guadua sp	39,4	5,21	0,54	54,85
6	Bambu guadua sp 350 °C	67,66	4,23	1,03	27,08
7	Bambu guadua sp 450 °C	69,34	3,83	0,93	25,9
8	Bambu guadua sp 550 °C	68,81	3,34	0,9	26,96
9	Bambu phyllostachys aurea	45,46	5,77	0,24	48,52
10	Bambu phyllostachys aurea 350 °C	69,39	4,32	0,4	25,88
11	Bambu phyllostachys aurea 450 °C	71,88	3,92	0,32	23,88
12	Bambu phyllostachys aurea 550 °C	74,99	3,56	0,26	21,19
13	Candeia	49,67	6,1	0,13	44,1
14	Candeia 350 °C	73,17	3,9	0,24	22,69
15	Candeia 450 °C	77,2	3,08	0,27	19,45
16	Candeia 550 °C	84,36	2,68	0,25	12,71
17	Casca macadamia	47,86	5,8	0,39	45,95
18	Casca macadamia 350 °C	70,27	4,38	0,74	24,61
19	Casca macadamia 450 °C	76,57	3,44	0,71	19,28
20	Casca macadamia 550 °C	81,47	2,76	0,6	15,17
21	Eucalyptus dunnii	38,2	4,83	0,07	56,9
22	Eucalyptus dunnii 350 °C	70,61	4,37	0,09	24,93
23	Eucalyptus dunnii 450 °C	70,9	4,17	0,08	24,85
24	Eucalyptus dunnii 550 °C	78,14	2,8	0,09	18,97
25	Eucalyptus urophylla	45,63	6,09	0,08	48,2
26	Eucalyptus urophylla 350 °C	67,52	4,44	0,07	27,97
27	Eucalyptus urophylla 450 °C	75,29	3,85	0,13	20,73
28	Eucalyptus urophylla 550 °C	81,94	3,14	0,2	14,72
29	Pinus caribea	39,71	5,12	0,07	55,1
30	Pinus caribea 350 °C	72,38	4,41	0,26	22,95
31	Pinus caribea 450 °C	77,34	3,66	0,52	18,48
32	Pinus caribea 550 °C	83,21	3,08	0,93	12,78
33	Pinus taeda	44,5	5,78	0,07	49,65
34	Pinus taeda 350 °C	71,89	4,34	0,96	22,81
35	Pinus taeda 450 °C	76,59	3,65	0,86	18,9
36	Pinus taeda 550 °C	82,68	3,13	0,58	13,61
37	Serragem AC	46,22	5,71	0,25	47,82
38	Serragem AC 350 °C	57,96	4,98	0,26	36,79
39	Serragem AC 450 °C	68,68	4,03	0,31	26,97
40	Serragem AC 550 °C	73,33	3,48	0,33	22,85
41	Serragem MT	42,87	5,46	0,37	51,29
42	Serragem MT 350 °C	57,54	4,92	0,52	37,02
43	Serragem MT 450 °C	58,37	3,79	0,69	37,14
44	Serragem MT 550 °C	62,69	3,11	0,63	33,57

A Tabela 6.1 foi obtida e adaptada do trabalho da pesquisadora Claudia Maria Branco de Freitas Maia da Embrapa Florestas submetido para a X Encontro Brasileiro de Substâncias Húmicas (Maia, 2013) contendo as porcentagens elementares de carbono, nitrogênio, oxigênio e hidrogênio para cada tipo de biocarvões e biomassas através da aplicação da avaliação composicional por análise elementar. O erro médio da análise elementar é aproximadamente 0,3% para cada teor presente nas amostras (Perkin Elmer, 2011). As amostras de biomassas e biocarvões respectivos contidos na Tabela 6.1 são as mesmas amostras utilizadas para a análise por MID-IR e construção do dendrograma obtido na Figura 6.3 através da utilização dos dados espectrais.

Através da análise do dendrograma é possível perceber alguns padrões existentes entre determinados espectros quando comparados com os valores previamente obtidos de análise elementar. O Grupo 1 tem aglomerado sobre ele todos os espectros relacionados às biomassas estudadas. Neste mesmo grupo existem ainda 2 biocarvões correspondentes ao Aguapé à 350°C e a 450°C. Pode-se concluir que entre todos os componentes que existe uma alta concentração de oxigênio e uma baixa concentração de carbono, como pode ser observado no gráfico da Figura 6.4 que apresenta a comparação das porcentagens de oxigênio entre os Grupos 1 e 2.

A presença dos biocarvões de Aguapé à 350°C e à 450°C no grupo das biomassas é um indício que essas duas amostras tiveram uma pirólise parcial ou não sofreram pirólise nas temperaturas indicadas, indicando assim uma certa resistência da amostras de biomassa de aguapé ao processo de pirólise. Analisando as posições dos patamares das amostras de biocarvões de Aguapé à 350°C e à 450°C no dendrograma, sendo essas correspondentes aos números 2 e 3, respectivamente, observa-se que essas amostras não se encontram relacionadas a nenhuma outra amostra do grupo 1, contendo assim características distintas em relação às demais amostras.

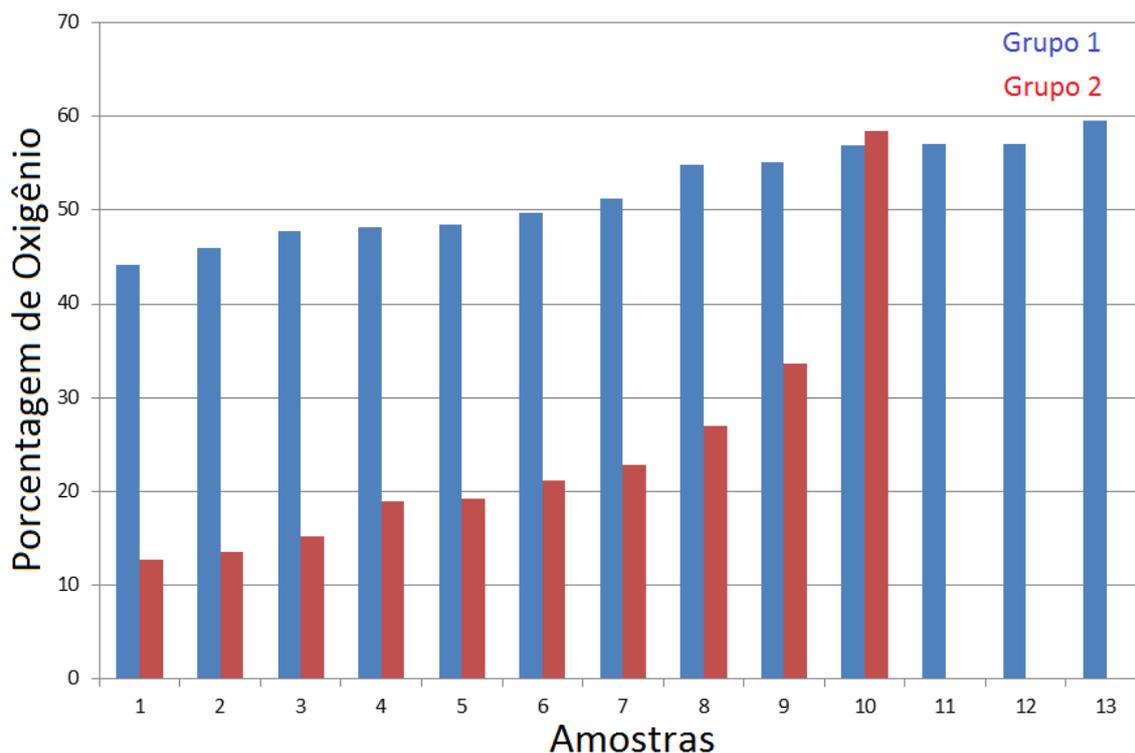


Figura 6.4. Gráfico comparativo do percentual de oxigênio entre as amostras dos grupos 1 e 2

O Grupo 2 teve a peculiaridade de reunir 9 das 11 amostras de biocarvão produzidas na temperatura de 550°C. A característica principal desse grupo é a baixa concentração de oxigênio e a alta concentração de carbono, como pode ser observado comparando-se os gráficos das Figuras 6.4 e 6.5. Isso se deve ao próprio processo de produção do biocarvão, a pirólise em baixa atmosfera de oxigênio, portanto para haver a combustão da biomassa é necessário o consumo do oxigênio da amostra.

A única exceção ao padrão existente no Grupo 2 é presença do biocarvão de Aguapé à 550°C, que tem uma concentração baixa de carbono e alta de oxigênio quando comparado com as outras amostras presentes no grupo, podendo assim apresentar mais resistência ao processo de pirólise. Assim como ocorre para as duas amostras de biocarvões de presentes no Grupo 1, o biocarvão de Aguapé à 550°C encontra-se isolado das outras amostras do Grupo 2 no dendrograma, sugerindo assim que apesar de ser produzido em alta temperatura, contém ainda uma alta concentração de oxigênio e não se encaixa perfeitamente no grupo podendo assim ter sofrido pirólise ou uma pirólise parcial.

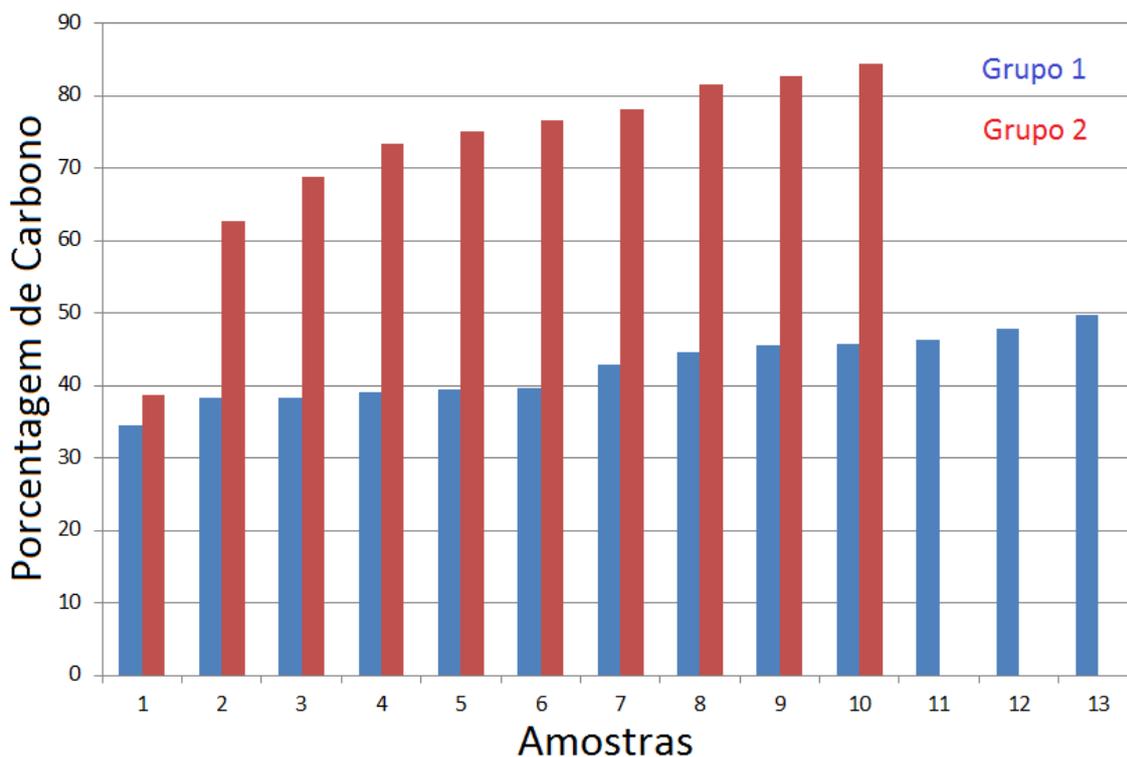


Figura 6.5. Gráfico comparativo do percentual de carbono entre as amostras dos grupos 1 e 2

O Grupo 3 contém amostras de biocarvões produzidas em temperaturas 350 e 450 °C, e apresentando valores intermediários entre as concentrações de carbono e oxigênio, contendo concentrações de hidrogênio em geral mais altas que as outras amostras de biocarvões quando comparados com o Grupo 4, como pode ser observado no gráfico da Figura 6.6. Ainda neste grupo existem 3 amostras de biocarvões produzidos a partir de serragem que apresentam concentração de mais baixa de carbono e concentração mais alta de oxigênio em relação às outras amostras do grupo: Serragem MT à 350 °C e à 450 °C, e Serragem à AC 350 °C.

O Grupo 4, assim como o Grupo 3, contém em sua maioria amostras de biocarvões produzidos nas temperaturas de 350 e 450°C, com exceção de duas amostras produzidas em temperaturas de 550°C: *Pinus caribea* e *Eucalyptus urophylla*. As amostras desse grupo contém uma alta concentração de carbono, de oxigênio quando comparadas com outras amostras dos grupos 2 e 3. As duas amostras produzidas à temperatura de 550 °C no Grupo 4 contém a concentração de carbono alta e de oxigênio baixa, semelhantemente às amostras do Grupo 3, entretanto por terem uma alta concentração de oxigênio após o processo de pirólise não se encaixam no Grupo 2.

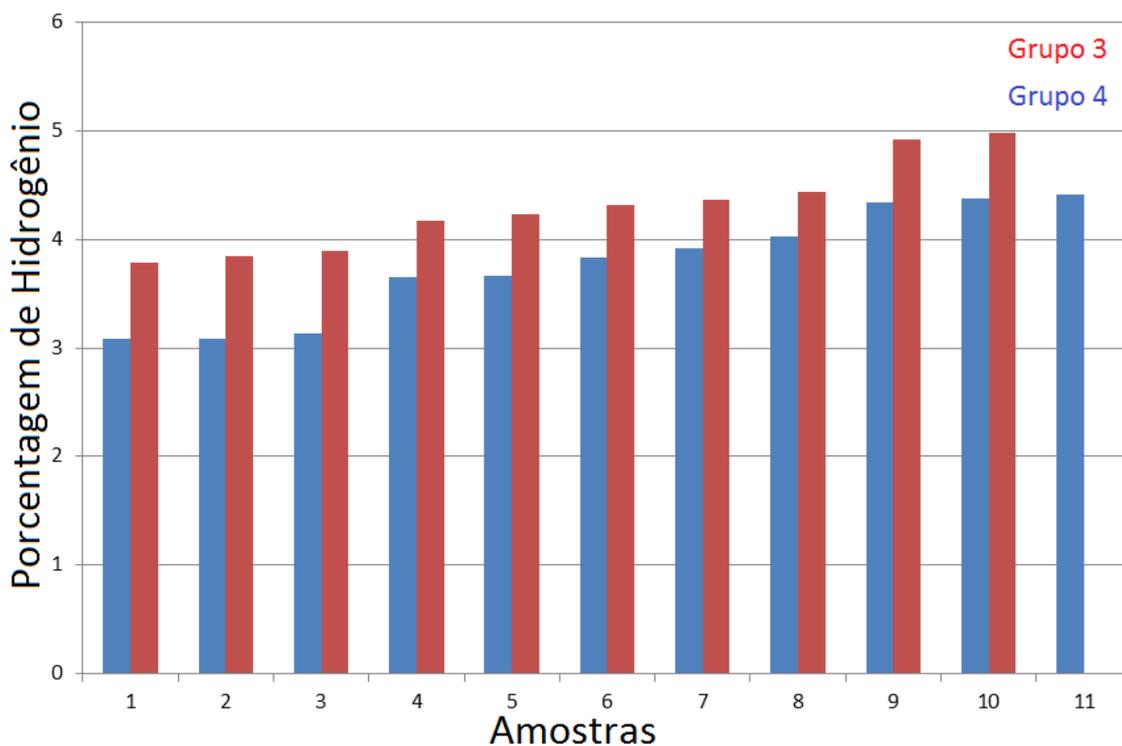


Figura 6.6. Gráfico comparativo do percentual de hidrogênio entre as amostras dos grupos 3 e 4.

6.4 Análise de Componentes Principais

O conjunto original de dados foi submetido à análise por PCA. A melhor descrição para o conjunto de amostras estudado foi com a utilização de 4 componentes principais, ou 4 PCs, como visto na Tabela 6.2, sendo que a primeira componente (PC1) contém a maior porcentagem de variância individual. As componentes subsequentes contêm variância individual muito inferior em relação à primeira componente, sugerindo assim que a PC1 pode explicar a maior parte da variância dos dados.

Tabela 6.2. Porcentagem de variância explicada pelas componentes principais.

Componentes	Porcentagem de Variância	
	Individual	Acumulada
PC1	95,47	95,47
PC2	3,03	98,50
PC3	0,59	99,08
PC4	0,44	99,52

Na Figura 6.7 está a representação o gráfico dos *scores* da componente principal 1. Cada amostra de biomassa e de biocarvão no gráfico de *scores* está representada por um ponto com seu número respectivo no eixo da abscissa de acordo com a Tabela 6.1. Analisando o gráfico de *scores* pode-se observar, através da variação do valor de *scores*, que as primeiras amostras aproximadamente até a 25ª amostra em geral não apresentam grandes diferenças entre as concentrações de carbono e oxigênio de acordo com o aumento da temperatura de pirólise, ocorrendo assim um aumento da concentração de carbono e não havendo um consumo extenso de oxigênio da amostra. A partir da 25ª amostra, em geral, é observável um maior consumo de oxigênio com aumento da concentração de carbono em acordo com o aumento da temperatura de pirólise para cada amostras. Assim é possível observar a separação das amostras de acordo com a sua resistência ao processo de pirólise nas temperaturas utilizadas, e as suas capacidade de retenção do oxigênio presente nas biomassas de partida.

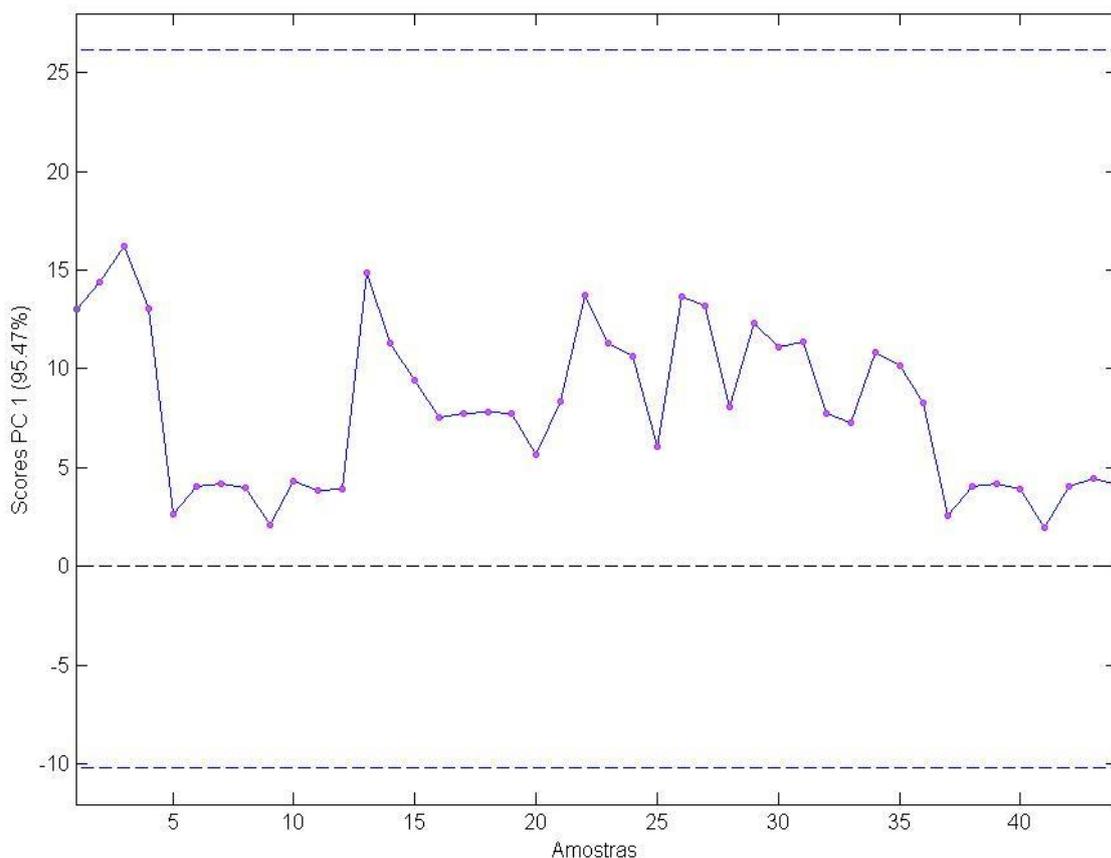


Figura 6.7. Representação gráfica dos *scores* da componente principal 1.

Na Figura 6.8 está a representação o gráfico dos *loadings* da componente principal 1. O gráfico dos *loadings* tem a representação semelhante do espectro médio resultante dos 44 espectros amostrais. Isso pode ser explicado pela porcentagem de variância muito superior detida pela PC1 em relação à PC2, PC3 e PC4, no qual mais de 95% das informações espectrais referentes às amostras está concentrada na componente principal 1, mostrando assim que o peso desta em relação a cada variável explica as características do espectro obtido associada à natureza das amostras.

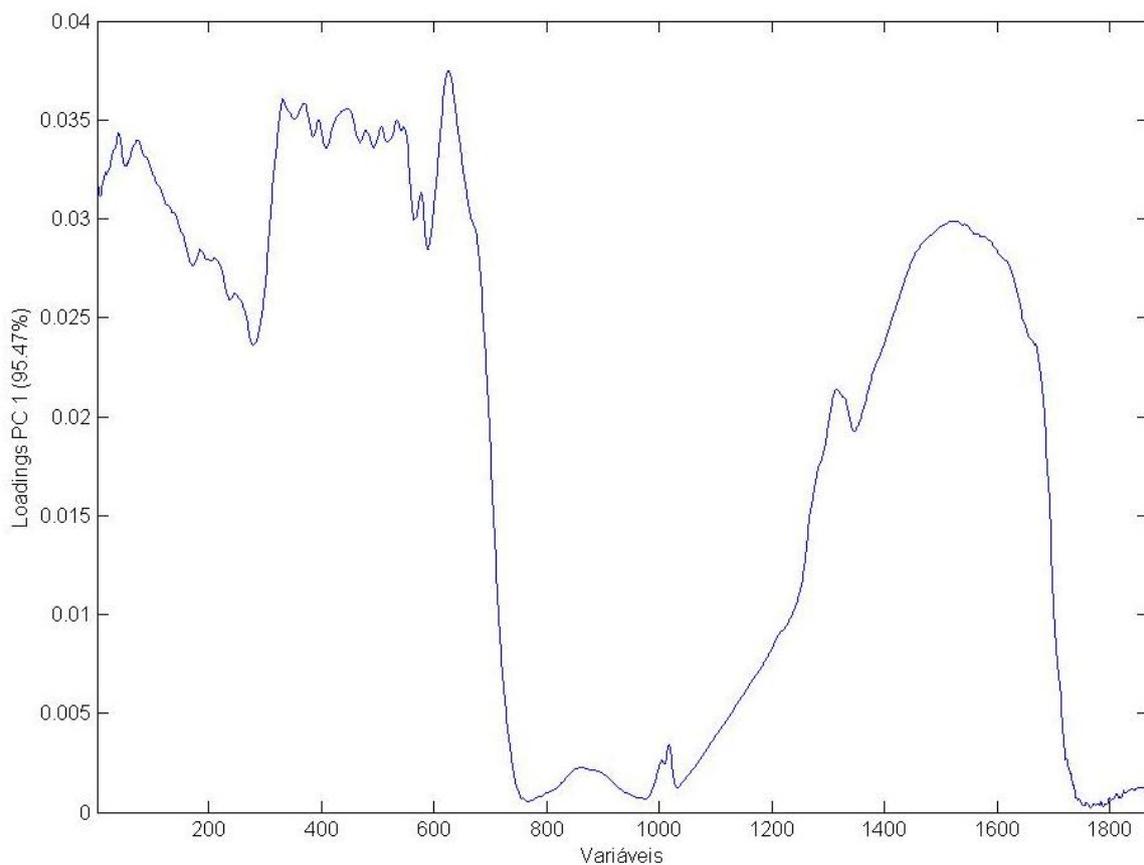


Figura 6.8. Representação gráfica dos *loadings* da componente principal 1.

CAPÍTULO 7

DETERMINAÇÃO DE CARBONO TOTAL

7.1 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 1 (SPA-MLR)

Na seguinte proposta foi utilizado o Algoritmo de Projeções Sucessivas (SPA) para a seleção de variáveis a partir das variáveis independentes nos espectros originais. Foram selecionadas 12 variáveis pelo SPA em um universo de 1869 pontos do espectro. Na Figura 7.1 são apresentados o espectro médio contendo e as variáveis selecionadas em destaque referentes aos número de onda de 404,97; 420,4; 470,54; 989,3; 1035,5; 1471,4; 1571,6; 1708,6; 2127; 2360,4; 3411,4 e 3617,7 cm^{-1} .

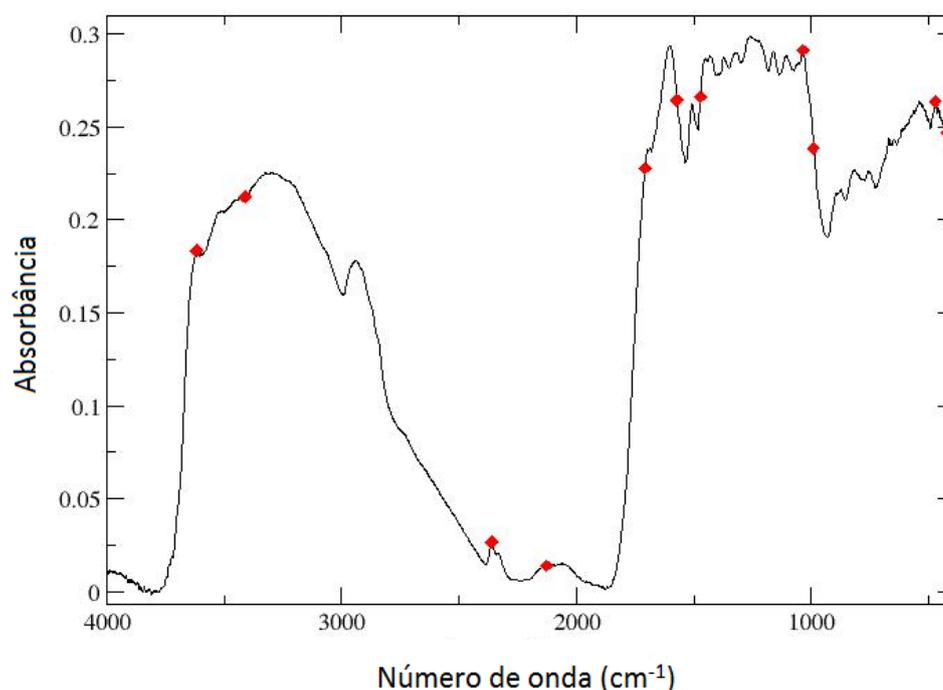


Figura 7.1. Espectro MID-IR médio com destaque para as 12 variáveis selecionadas pelo SPA para a determinação de carbono.

O número de onda em 989,3 cm^{-1} selecionado pelo SPA está associado a presença do modo vibracional de deformação dos grupos C-O-C da celulose, em 1035,5 cm^{-1} verifica-se a presença de estiramentos das ligações C-O, C=C e C-C-O na celulose, hemicelulose e lignina, respectivamente. A variável escolhida em 1471,4 cm^{-1} corresponde à ocorrência de deformação das ligações C-H de grupos alquílicos na

lignina. O número de onda escolhido de $1571,6 \text{ cm}^{-1}$ é associado a ligações aromáticas C=C na lignina. Em $1708,6 \text{ cm}^{-1}$ ocorre absorção atribuída ao estiramento de ligações C=O em aldeídos e cetonas na hemicelulose (Xu et al., 2013). Algumas das variáveis selecionadas não puderam ser diretamente relacionadas com modos vibracionais devido a presença de carbono nas amostras de biocarvões ou biomassas. O SPA seleciona variáveis que apresentam baixa multicolinearidade com as variáveis já selecionadas. Por isto, as variáveis escolhidas nesta proposta não precisam estar diretamente relacionadas a modos vibracionais relacionados a espécie quantificada. É possível que as variáveis com números de onda $404,97$; $420,4$; $470,54$; 2127 ; $2360,4$; $3411,4$ e $3617,7 \text{ cm}^{-1}$ estejam correlacionadas com outros grupos funcionais presentes nos biocarvões e biomassas. Por exemplo, o número de onda em $3617,7 \text{ cm}^{-1}$ pode ser relacionado ao estiramento O-H de álcoois e fenóis presentes na celulose, lignina e hemicelulose. Já o número de onda em $3411,4 \text{ cm}^{-1}$ é, em geral, relacionado com o estiramento de ligações N-H em aminas primárias.

Na seção 6.3 é apresentada a Tabela 6.1 com os teores elementares percentuais de cada uma das amostras de biomassa e biocarvão. Os teores elementares podem ser interpretados como a quantidade de gramas do teor elementar a cada 100 gramas de amostras. No decorrer deste trabalho nas apresentações dos resultados serão utilizados valores de concentração na unidade de g/kg, a qual é obtida multiplicando os valores dos teores da Tabela 6.1 por 10.

As 12 variáveis selecionadas pelo método SPA foram utilizadas como dados de entrada para a Regressão Linear Múltipla (RLM), conseguindo assim obter um modelo com RMSEP de $55,2 \text{ g/kg}$, RMSEC de $21,1 \text{ g/kg}$ e RMSECV de $44,6 \text{ g/kg}$. O fator de correlação R^2 obtido foi de $0,9421$ para a validação e $0,9909$ para a calibração. O erro de calibração foi muito inferior ao erro de predição, mas o ajustamento da curva de calibração e validação com a resposta medida foi muito superior, como pode ser observado analisando os coeficientes de determinação nas etapas de calibração e predição, $0,9909$ e $0,9421$, respectivamente. A análise da Figura 7.2 mostra que no centro da reta ocorre uma baixa concentração de pontos referentes às amostras contendo concentrações variando de 50 a 70 % de carbono. Este resultado mostra que deve-se ter cuidado ao aplicar este modelo na determinação de concentração de

carbono de biomassas e biocarvões nesta faixa de concentração. Uma das razões da ocorrência de uma baixa concentração de amostras na região citada acima está relacionado ao conjunto de amostras disponíveis para a construção do modelo.

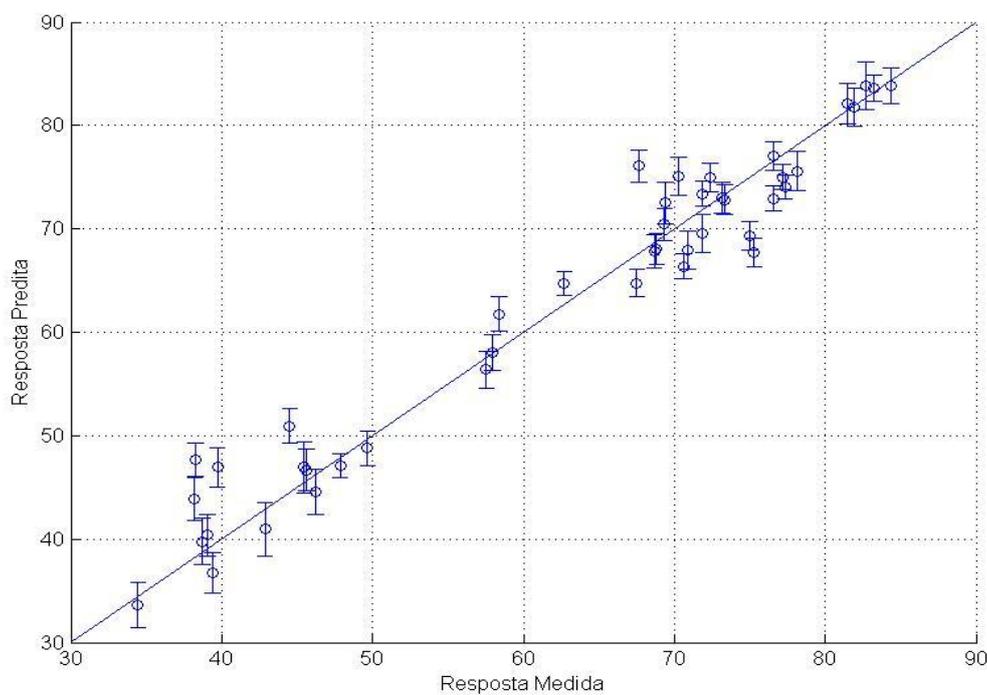


Figura 7.2. Valor previsto para carbono total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta predita e as barras a resposta experimental da análise elementar.

7.2 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 2 (PLS)

Como visto anteriormente no capítulo 4, as variáveis latentes substituem as variáveis originais do conjunto de dados até o ponto em que ainda se tornem representativas para o sistema estudado. Isso se dá observando as quantidades de variáveis latentes, escolhendo, assim, o modelo que utiliza o menor número de variáveis latentes, mas que contenha os valores do coeficiente de correlação R^2 de calibração e predição próximos entre si, concluindo assim que o modelo obtido representa bem os conjuntos de calibração e predição.

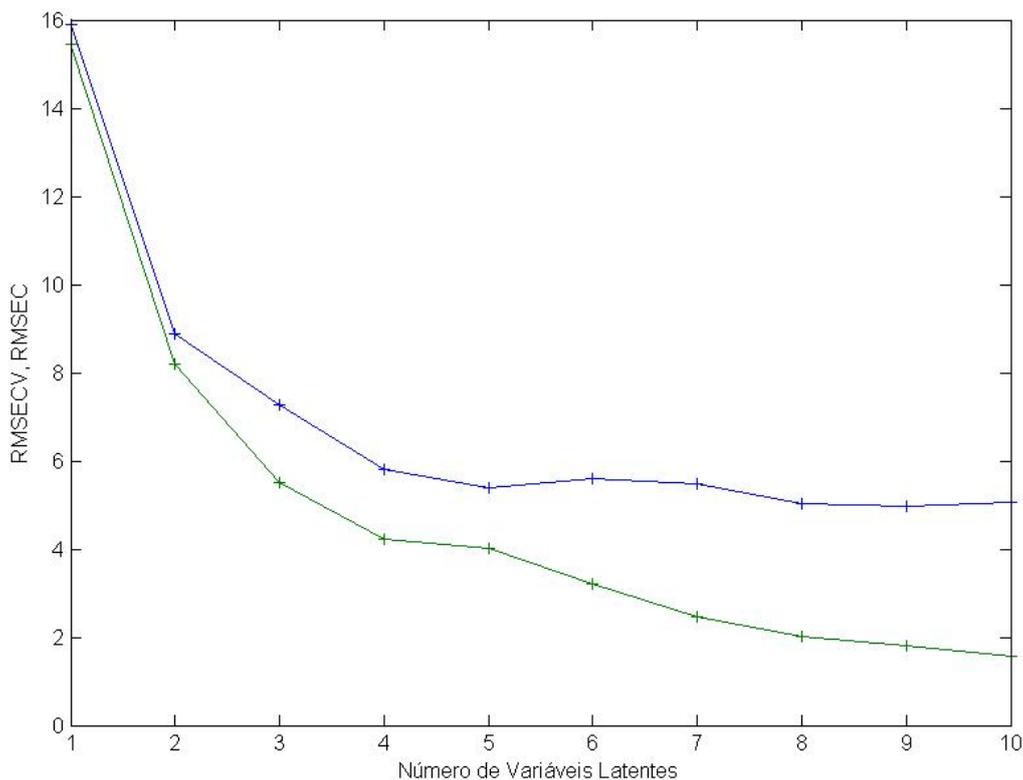


Figura 7.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de carbono total.

Dentre os resultados obtidos, o modelo que apresentou o menor valor RMSECV foi com a utilização de 5 variáveis latentes (Figura 7.3), com os coeficientes de determinação de 0,9362 e 0,9123 para a calibração e validação, respectivamente. Os resultados deste modelo são apresentados na Figura 7.4. A variância acumulada dos blocos de variáveis independentes X foi de 99,62% e no bloco de variáveis dependentes Y foi de 99,61%. Os valores de RMSEC, RMSEP e RMSECV foram respectivamente de 40,0 g/kg; 53,16 g/kg e 63,25 g/kg. A correlação entre os pontos da resposta medida e a resposta predita apresentados na Figura 7.4, caracterizam o ajustamento entre a curva das amostras e a curva obtida do modelo. Os valores de RMSEC e RMSEP se mostram aproximados, assim como os ajustes das curvas nas etapas de calibração e predição como pode ser visto comparando os valores dos coeficientes de determinação das duas etapas. Assim, o modelo obtido apresenta capacidade de quantificação de carbono.

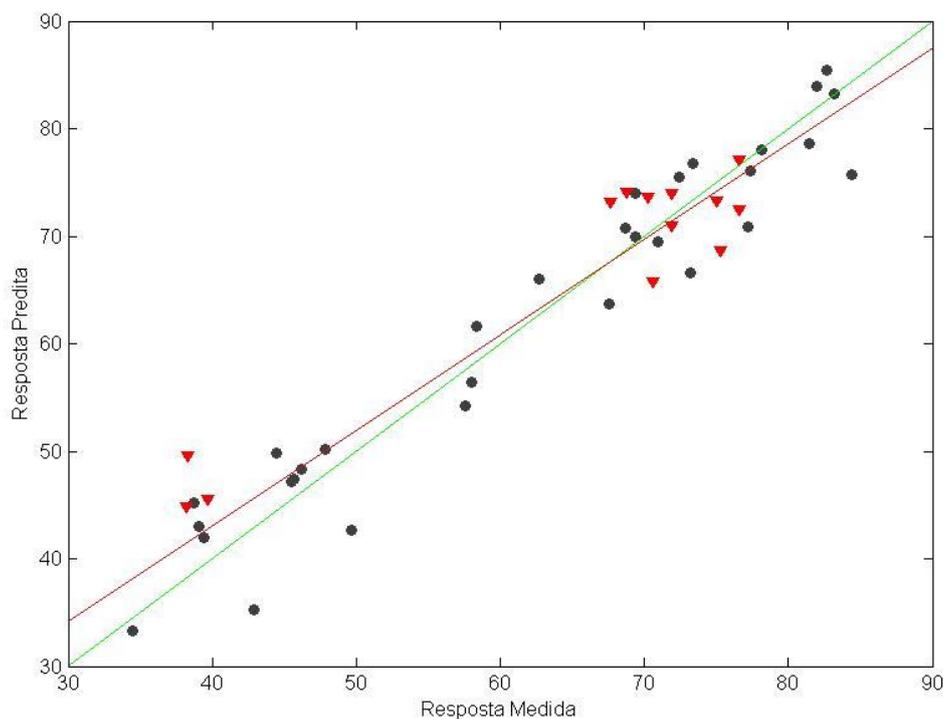


Figura 7.4. Valor previsto para carbono total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elementar. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

7.3 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)

A busca de variáveis através do Algoritmo Genético foi realizada primeiramente utilizando a Validação Cruzada da Regressão Linear Múltipla (RLM) com parâmetro para avaliação e seleção subconjunto de variáveis para a obtenção dos modelos. No total foram selecionadas 130 variáveis que apresentaram o melhor RMSECV em relação ao conjunto total de variáveis. Dentre as 130 variáveis selecionadas pelo AG, as seguintes foram selecionadas em comum pelas outras técnicas de seleção de variáveis: 406,9; 809,95; 979,65 e 1257,3 cm^{-1} . Algumas dessas variáveis foram identificadas em relação à presença direta de carbono nas amostras, como pode ser visto nas discussões das propostas 1 e 5 nas seções 7.1 e 7.5. Portanto, essas variáveis, em princípio, contém informação relevante para a construção de modelos multivariados, apesar de algumas delas não conterem informação química relacionada diretamente com modos vibracionais devido a presença de carbono nas amostras.

Com os valores selecionados pelo AG, foi utilizado o método RLM para a regressão dos dados. Os valores de RMSEC, RMSEP e RMSECV foram respectivamente 42,0 g/kg, 51,5 g/kg e 57,1 g/kg. O coeficiente de determinação R^2 para a calibração foi de 0,9313 e para a validação foi de 0,8776 (Figura 7.5). O valor de RMSEC é inferior ao valor de RMSEP e ambos os erros deste modelo são maiores que os obtidos pela utilização do modelo SPA-MLR. Portanto, podemos considerar que o modelo GA-MLR é menos preciso que o SPA-MLR.

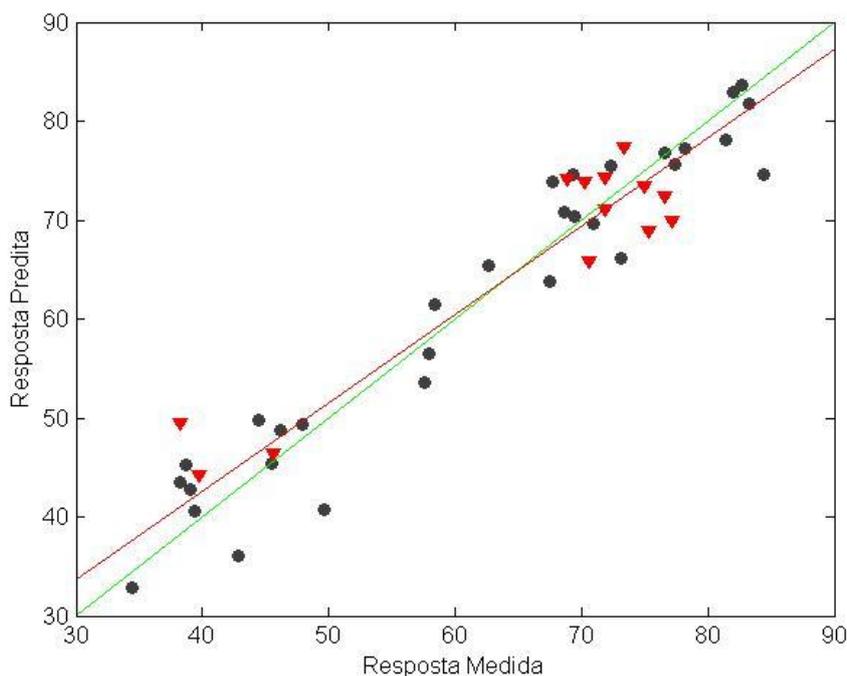


Figura 7.5. Valor previsto para carbono total pelo modelo de calibração RLM (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

7.4 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)

O Algoritmo Genético também foi utilizado para a realização de busca de variáveis através do emprego da Validação Cruzada no método PLS. O princípio de avaliação e seleção é o mesmo utilizado para a Regressão Linear Múltipla, a busca de um subconjunto de valores de absorvância que resultem no melhor valor de RMSECV de todo o espectro original. Foi selecionado um subconjunto de 528 variáveis em relação ao espectro original. Dentre essas variáveis, os números de onda em 804,16; 806,09; 1186 e 1708,6 cm^{-1} também foram selecionados pelos métodos SPA e/ou iPLS,

significando assim a possibilidade de conter informação de importância para a construção do modelo multivariado além de identificar a presença de direta de carbono nas amostras, como pode ser visto nas seções 7.1 e 7.5.

Na Figura 7.6 é apresentado gráfico do RMSECV e RMSEC em relação ao número de variáveis latentes para o modelo GA-PLS considerando as 528 variáveis selecionadas. A Figura 7.6 mostra que o número de 3 variáveis latentes apresenta o menor valor de RMSECV aos demais, e portanto, ele será selecionado para as análises dos resultados. O modelo GA-PLS com 3 variáveis latentes apresenta um valor de RMSEC de 54,2 g/kg, de RMSEP de 63,6 g/kg e de RMSECV de 79,3,0 g/kg. Os valores dos coeficiente de determinação R^2 para a calibração e para a validação foram de 0,8893 e 0,8037, respectivamente (ver Figura 7.7). A variância acumulada dos blocos de variáveis independentes X foi de 98,42% e no bloco de variáveis dependentes Y foi de 99,28%. O modelo obtido apresentou um menor valor de RMSEC e um melhor ajuste na etapa de calibração, e um maior valor de RMSEP e menor ajuste na etapa de predição. Esses resultados indicam um sobre ajuste do modelo obtido, tornando-o assim tendencioso à etapa de calibração e menor capacidade de predição dos dados.

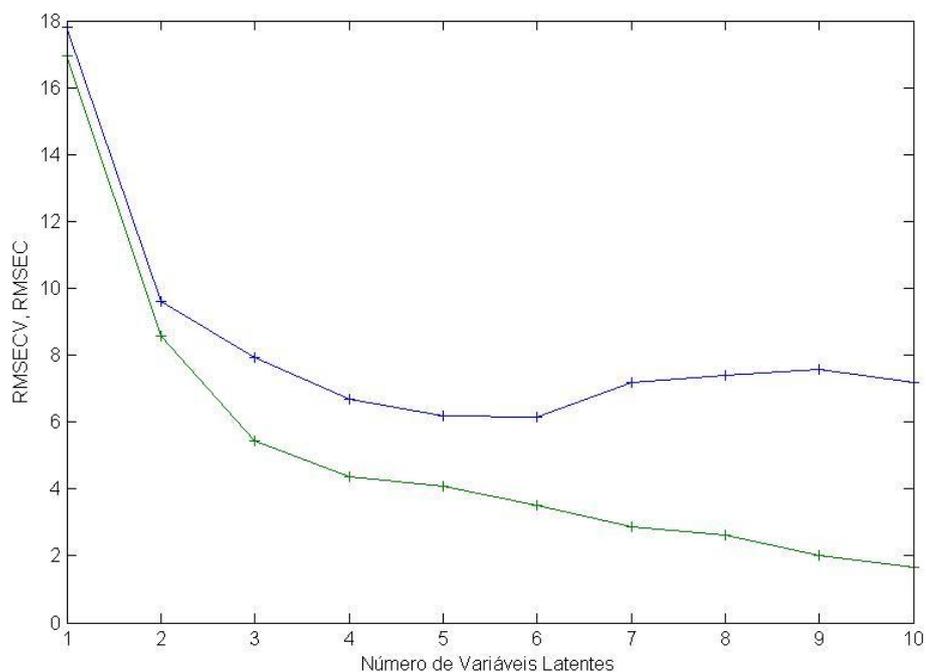


Figura 7.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de carbono total.

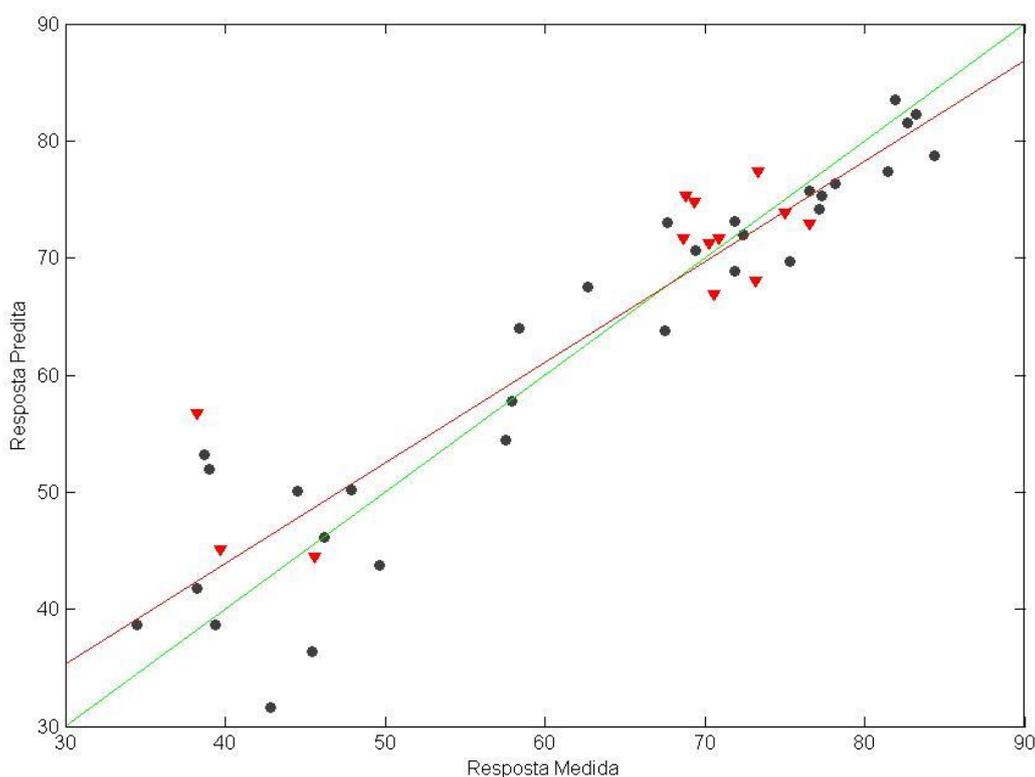


Figura 7.7. Valor previsto para carbono total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

7.5 Determinação de carbono total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)

A determinação de carbono também foi realizada pela seleção de variáveis Interval Partial Least Square (iPLS ou IPLS) acoplada com a regressão por PLS. O método de seleção do iPLS é semelhante ao do AG no sentido de utilizar o RMSECV como método de avaliação, mas a realiza a busca de forma contígua em intervalos de número de onda que contenham absorbâncias que conduzam um melhor valor de RMSECV. Para o análise do carbono, o iPLS selecionou 45 variáveis em 7 intervalos distintos em relação ao conjunto de dados totais. Os seguintes intervalos foram selecionados: entre 399,19~406,9; 804,16~811,88; 977,72~985,44; 1180,2~1187,9; 1257,3~1265; 2530,1~2557,1; e 3176,1~3183,8 cm^{-1} . O espectro médio com os 7 intervalos contendo as 45 variáveis selecionadas pode ser observado na Figura 7.8.

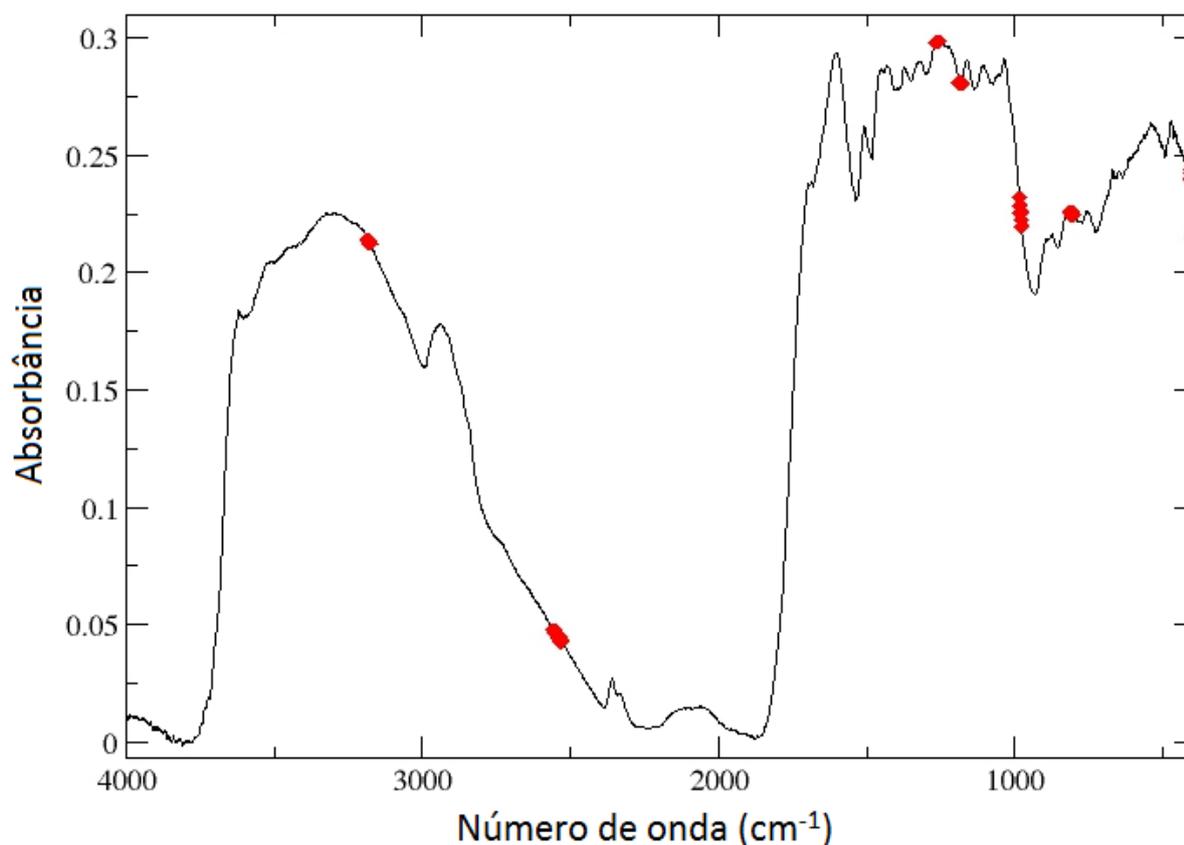


Figura 7.8. Espectro MID-IR médio das 44 amostras com destaque para as 45 variáveis selecionadas pelo iPLS correspondente a 7 intervalos de frequências para a determinação da concentração de carbono.

O intervalo de 804,16~811,88 cm^{-1} está associado com o estiramento de ligações C-H de estruturas aromáticas dissustituídas, que podem ser encontradas na lignina. De 977,72~985,44 cm^{-1} é associado com presença de vibrações de valência em ligações C-O da celulose. A presença de ligações O-H de álcoois e o estiramento assimétrico em ligações C-O-C de grupos ésteres na celulose e hemicelulose identifica as variáveis escolhidas de 1180,2~1187,9 cm^{-1} . De 1257,3~1265 cm^{-1} é associado ao estiramento de ligações C-C alquílicas e C-O de ácidos carboxílicos e ésteres na lignina (Xu et al., 2013). Os números de onda de 399,19~406,9; 2530,1~2557,1; e 3176,1~3183,8 cm^{-1} não puderam ser identificados relacionando diretamente com o teor de carbono total.

A regressão por PLS forneceu o modelo mais ajustado com a utilização de 5 variáveis latentes, como pode ser observado na Figura 7.9. A variância acumulada dos blocos de variáveis independentes X foi de 99,73% e no bloco de variáveis dependentes Y foi de 99,65%, e o RMSEC apresentou o valor de 37,7 g/kg, o RMSEP o valor de 40,7 g/kg e o RMSECV o valor de 67,3 g/kg. O R^2 da calibração e da validação foram 0,9433 e

0,9685, respectivamente (Figura 7.10). O modelo obtido pela utilização do iPLS se mostrou muito promissor para a quantificação de carbono em amostras de biomassas e biocarvões. Os valores de RMSEC e RMSEP são aproximados, havendo a diferença de aproximadamente 3,0 g/kg entre os mesmos, e o modelo se ajusta de forma semelhante aos conjuntos de calibração e validação, como pode ser visto pelo coeficientes de determinação R^2 obtidos.

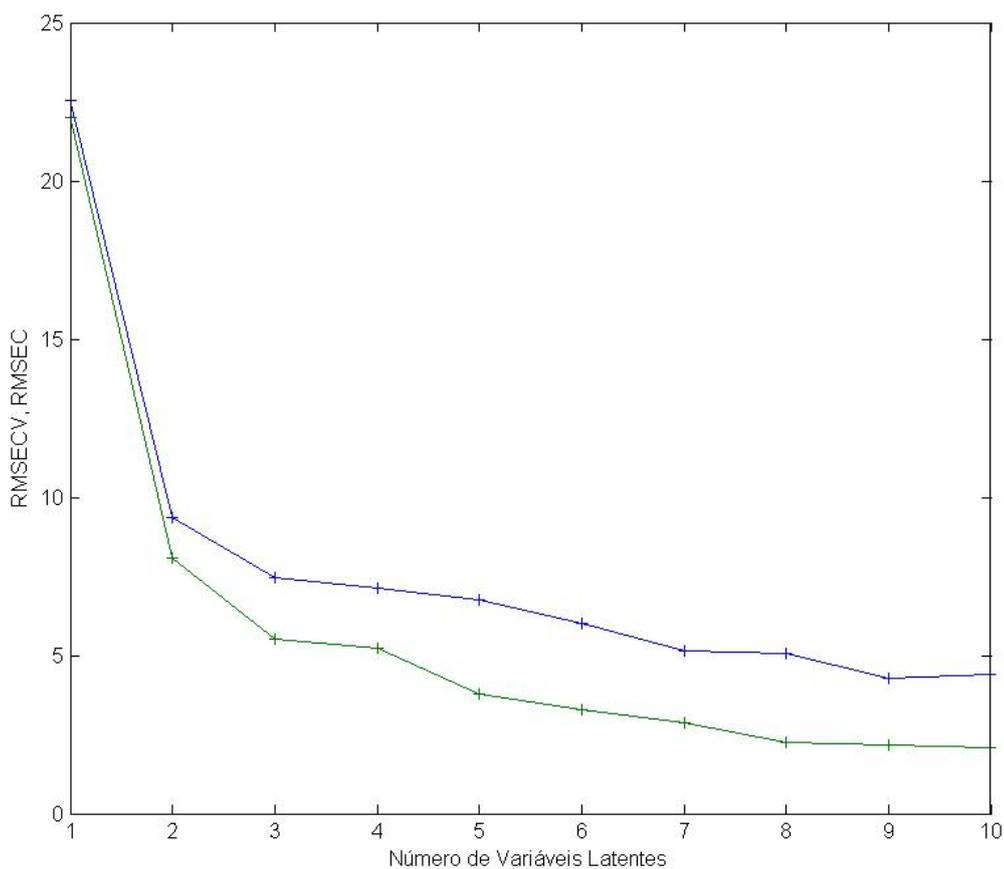


Figura 7.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de carbono total.

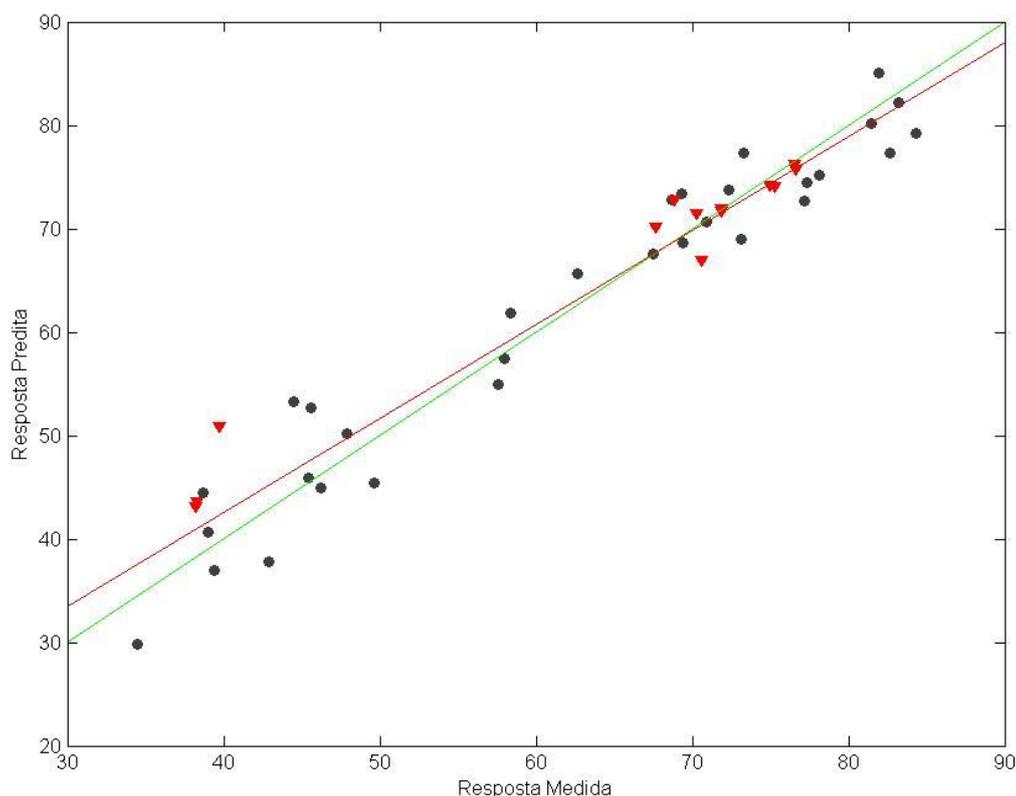


Figura 7.10. Valor previsto para carbono total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elementar. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

Na Tabela 7.1 estão contidos os resultados correspondentes a cada proposta apresentada para a determinação das concentrações totais de carbono. As propostas SPA-MLR e iPLS apresentaram os melhores resultados em relação às outras propostas, havendo um ajuste mais adequados desses modelos tanto na etapa de calibração quanto na etapa de predição dos dados, como visto pelo coeficientes de determinação de ambas etapas e também pelo parâmetro RPD calculado para as duas propostas. Entretanto, o SPA-MLR apresentou um valor de RMSEC muito baixo e um valor de RMSEP muito alto, sugerindo assim um possível sobre ajuste do modelo na etapa de calibração. Diferentemente disso, o iPLS apresenta valores de RMSEC e RMSEP, e coeficientes de determinação R^2 próximos uns dos outros, indicando assim menor erros referentes ao modelo e maior capacidade de quantificar novos dados espectrais de biomassa e biocarvões pirólicos, podendo o iPLS, deste modo, ser mais preciso que o SPA-MLR.

Tabela 7.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de carbono total

Proposta	Nº de variáveis selecionadas	RMSEC (g/kg)	R ² Calibração	RMSEP (g/kg)	R ² Validação	RPD
SPA-MLR	12	21,1	0,990	55,2	0,942	4,152
PLS	-	40,0	0,936	53,1	0,912	3,370
GA-MLR	130	42,0	0,931	51,5	0,877	2,851
GA-PLS	528	54,2	0,889	63,6	0,803	2,253
iPLS	45	37,7	0,943	40,7	0,968	5,590

Podemos também observar que dentre as proposta utilizadas que contém técnicas de seleção de variáveis, o SPA-MLR e o iPLS apresentam o menor número de variáveis selecionadas, contendo assim uma carga de informação referente às amostras mais relevante em comparação às outras propostas. Nas propostas utilizando o GA foram utilizados um número superior de variáveis, que podem ser considerados excessivo e desnecessários à construção do modelo, contendo informações irrelevantes a determinação da concentração de carbono. Isto leva a construção de modelos menos adaptados ou sobre adaptados a determinação da concentrações de carbono em amostras de biomassas e biocarvões.

CAPÍTULO 8

DETERMINAÇÃO DE OXIGÊNIO TOTAL

8.1 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 1 (SPA - MLR)

Na determinação do oxigênio, a utilização do Algoritmo das Projeções Sucessivas resultou na seleção de um total de 6 variáveis de interesse que são um subconjunto do total de variáveis que guardam o mínimo de colinearidade entre si. Essas variáveis selecionadas correspondem aos seguintes números de onda na faixa do espectro de MID-IR: 470,54; 989,3; 1035,5; 1571,6; 2360,4 e 3411,4 cm^{-1} . O espectro médio contendo a 6 variáveis selecionadas é apresentado na Figura 8.1.

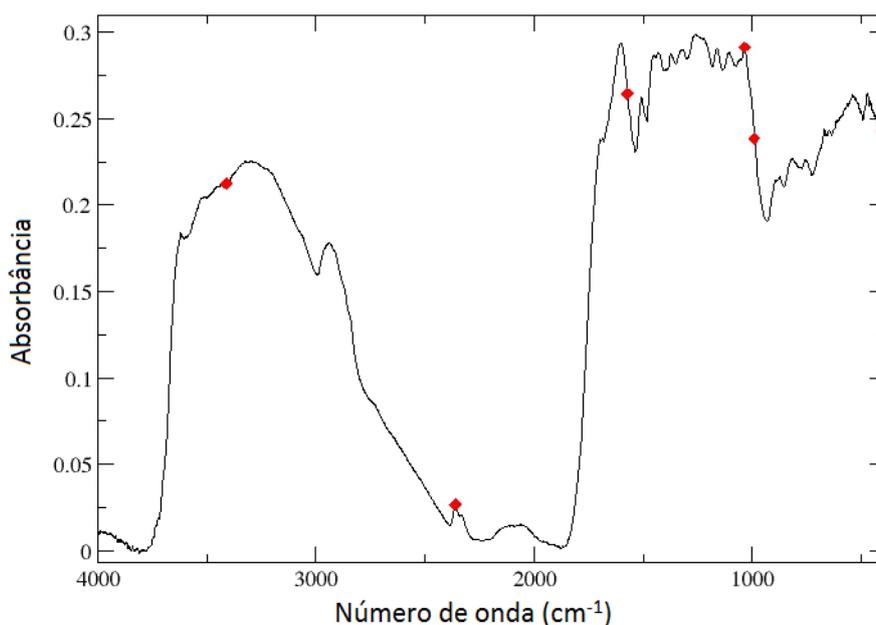


Figura 8.1. Espectro MID-IR médio das 44 amostras com destaque para as 6 variáveis selecionadas pelo SPA para a determinação de oxigênio.

O número de onda em 989,3 cm^{-1} pode ser associado com as vibrações de valência em ligações C-O da celulose, e em 1035,5 com o estiramento de ligações C-O e C-C-O na celulose, hemicelulose e lignina (Xu et al., 2013). O número de onda escolhido de 1571,6 cm^{-1} identifica estiramentos de ligações aromáticas C=O de ácidos carboxílicos e/ou derivados desses (Solomons; Fryhle, 2008). Em 1708,6 cm^{-1} ocorre o estiramento de ligações C=O em aldeídos e cetonas na hemicelulose. O número de onda em 3411,4

cm^{-1} está associado com o estiramento de ligações O-H de álcoois e fenóis na lignina (Xu et al., 2013). As variáveis em $470,54$ e $2360,4 \text{ cm}^{-1}$ não puderam ser identificadas de forma a ser relacionadas diretamente com o teor total de oxigênio. Os comprimentos de onda de $470,54$; $989,3$; $1035,5$; $1571,6$ e $2360,4 \text{ cm}^{-1}$, como visto no capítulo de determinação de carbono também foram selecionados pela seleção de variáveis por SPA, indicando assim que essas variáveis podem apresentar baixa colinearidade entre si, sendo que em $989,3$; $1035,5$ e $1571,6 \text{ cm}^{-1}$ estão relacionadas diretamente com ligações carbono-oxigênio.

Os valores encontrados de RMSEC, RMSEP e RMSECV foram, respectivamente, $25,7 \text{ g/kg}$; $52,0 \text{ g/kg}$ e $36,5 \text{ g/kg}$. O coeficiente de correlação R^2 apresentou o valor de $0,9516$ para a validação e $0,9852$ para a calibração. Comparando os valores de RMSEC e RMSEP em relação aos dados de análise elementar para o oxigênio na Tabela 6.1, os erros apresentam um margem aceitável em relação à maioria das amostras. A qualidade do modelo para a predição de outros dados pode também ser assegurada pelo ajustamento das respostas medidas e preditas, dada a Figura 8.2.

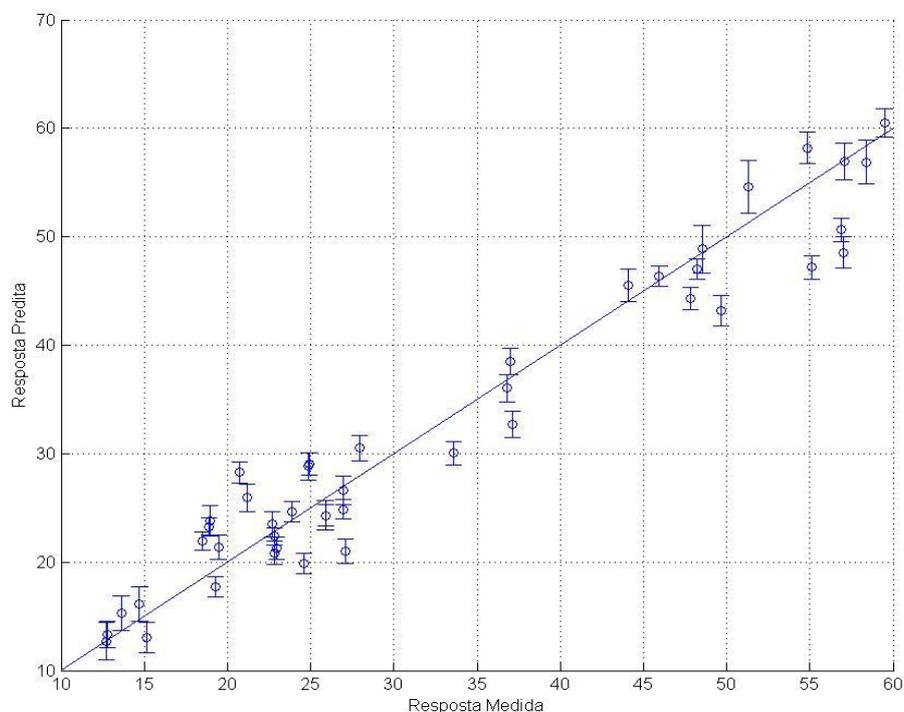


Figura 8.2. Valor previsto para oxigênio total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta predita e as barras a resposta experimental da análise elementar.

É possível observar a ausência de densidade de pontos preditos pelo modelo na região central do gráfico, como também foi observado no capítulo de determinação do carbono na seção referente ao SPA-MLR. A explicação para o comportamento visto nos gráficos de resposta para o carbono é a mesma para o oxigênio, a quantidade reduzida de dados utilizados para a obtenção do modelo sem uma variedade maior de amostras decorre em um conjunto de dados pouco variado, ou amplo, resultando em uma densidade de dados imperfeita.

8.2 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 2 (PLS)

Para a construção do modelo por PLS foram utilizadas 5 variáveis latentes, como observado na Figura 8.3, onde se mantém o limite onde não ocorre um desvio muito grande do modelo analisando os valores de RMSEC e RMSECV em relação a quantidade de variáveis latentes. A variância acumulada dos blocos de variáveis independentes X foi de 99,43% e no bloco de variáveis dependentes Y foi de 98,92%. O valores de RMSEC, RMSEP e RMSECV foram de 38,0 g/kg; 50,9 g/kg e 55,7 g/kg, respectivamente.

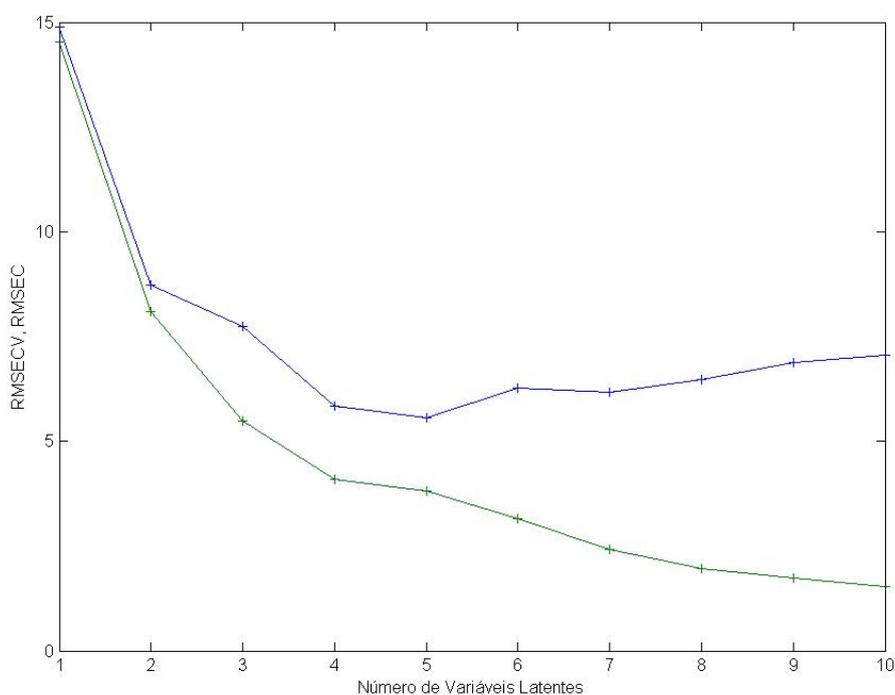


Figura 8.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de oxigênio total.

Os coeficientes de determinação R^2 obtidos para a calibração e validação foram de 0,936 e 0,939. A utilização do PLS forneceu um modelo para a quantificação de oxigênio com precisão ligeiramente inferior à proposta por SPA-MLR, mas que ainda contém uma capacidade preditiva significativa, dado o valor de RMSEP e o ajustamento da curva na etapa de predição (ver Figura 8.4).

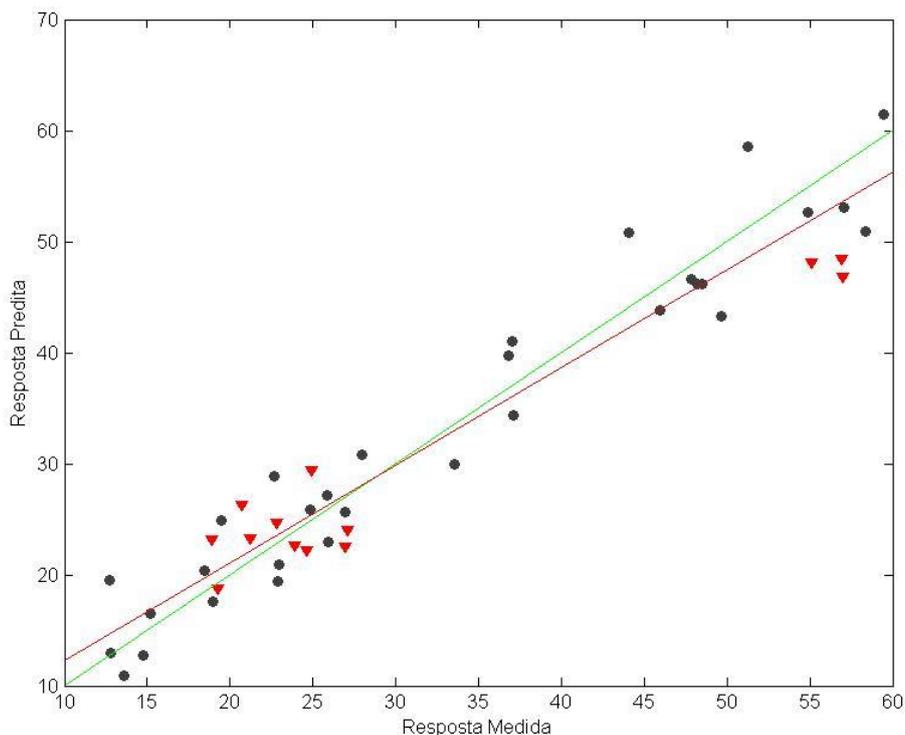


Figura 8.4. Valor previsto para oxigênio total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elementar.
 (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

8.3 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)

O Algoritmo Genético para a seleção de variáveis com validação cruzada por RLM selecionou um subconjunto de 168 variáveis do conjunto total do espectro. Deste subconjunto selecionado, as seguintes variáveis escolhidas pelo AG foram selecionadas em comum pelos outros métodos de seleção de variáveis, SPA e iPLS: 1226,5; 1249,6; 1251,5; 1259,2; 1378,8; 2790,4; 3315; 3357,4; 3392,1; 3442,3 e 3764,3 cm^{-1} .

Portanto, pode-se afirmar que essas variáveis portam informação significativa para a construção de um modelo multivariado mais adequado, sendo que algumas delas possuem identificação, sob o aspecto químico, relacionada diretamente com a presença de grupos funcionais contendo oxigênio nas amostras. A identificação em relação à presença de oxigênio nas amostras são apresentadas nas discussões das seções 8.1 e 8.5 relativas as propostas 1 e 5.

Os valores de RMSEC, RMSEP e RMSECV para o modelo foram de 39,7 g/kg; 45,8 g/kg e 54,5 g/kg, respectivamente. O coeficiente de determinação para a calibração foi de 0,934 e para a validação foi de 0,907. A proposta GA-MLR não obteve um modelo capaz de prever outros conjuntos de amostras quando comparado com os modelo obtidos por PLS e SPA-MLR, com um valor de RMSEP superior ao valor de RMSEC e o ajustamento da curva na etapa de predição inferior ao obtido na etapa de calibração (ver Figura 8.5).

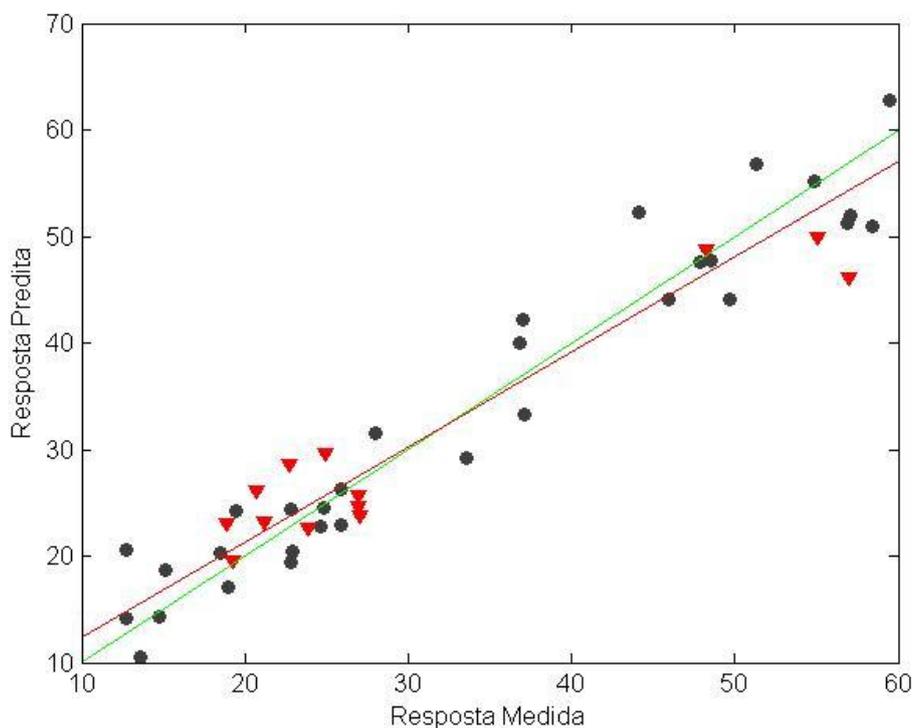


Figura 8.5. Valor previsto para oxigênio total pelo modelo de calibração MLR (%) com seleção de variáveis por GA contra o valor obtido pela análise elemental.
(●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

8.4 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)

Na utilização da validação cruzada por PLS para a seleção de variáveis, o AG selecionou um total de 502 comprimentos de onda em relação ao espectro total. Dentre essas 502 variáveis, das quais pode-se destacar as seguintes: 973,87; 1261,2; 1265; 2308,3; 2790,4; 3054,6; 3272,5; 3276,4; 3301,5; 3309,2; 3318,8; 3411,4; 3436,5 e 3440,3 cm^{-1} . Essas variáveis também foram selecionadas em comum pelas outras técnicas de seleção de variáveis, SPA e iPLS, como também algumas delas identificam diretamente a presença de oxigênio nas amostras de biomassa e biocarvões. A identificação em relação à presença de oxigênio nas amostras são apresentadas nas seções 8.1 e 8.5.

As 502 variáveis foram utilizadas para a regressão por PLS, e o modelo mais satisfatório obtido foi com a utilização de 5 variáveis latentes (Figura 8.6), sendo a variância acumulada dos blocos de variáveis independentes X e de variáveis dependentes Y para esse modelo de 99,06% e 99,19%, respectivamente.

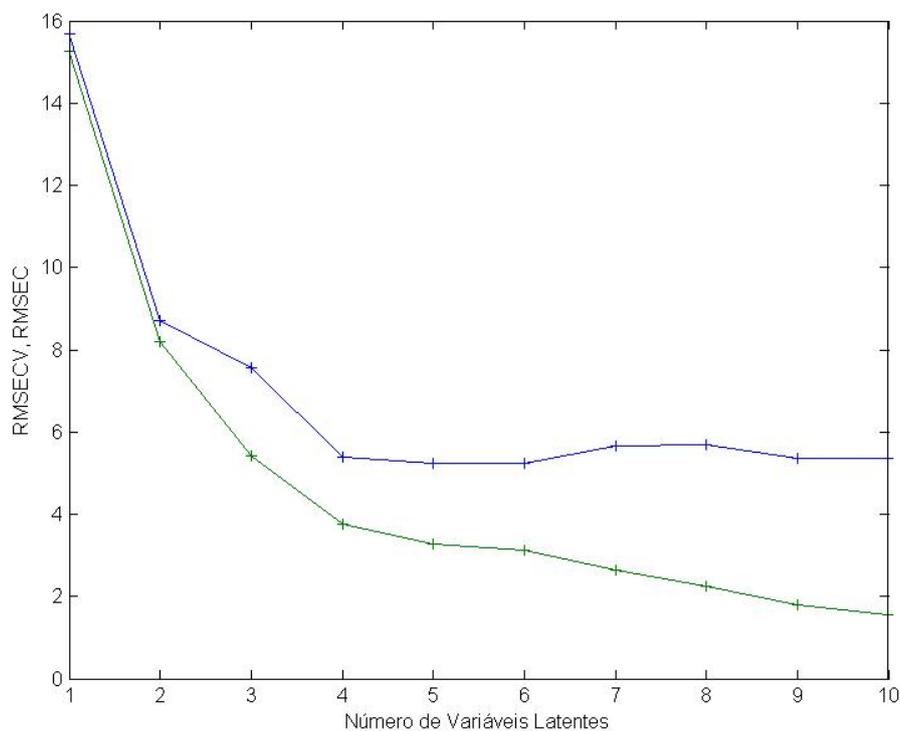


Figura 8.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de oxigênio total.

O modelo obtido forneceu os valores de RMSEC, RMSEP e RMSECV de 32,7 g/kg; 51,2 g/kg e 52,5 g/kg, respectivamente. Os coeficientes de determinação para a calibração foi de 0,954 para a validação foi de 0,957 (Figura 8.7). O modelo obtido pelo PLS apresentou um ajuste semelhante nas etapas de calibração e validação, como pode ser visto analisando os coeficientes de determinação nas duas etapas. Entretanto o valor de RMSEC foi muito inferior em relação ao valor do RMSEP, sugerindo assim a ocorrência de um sobre ajuste na etapa de calibração.

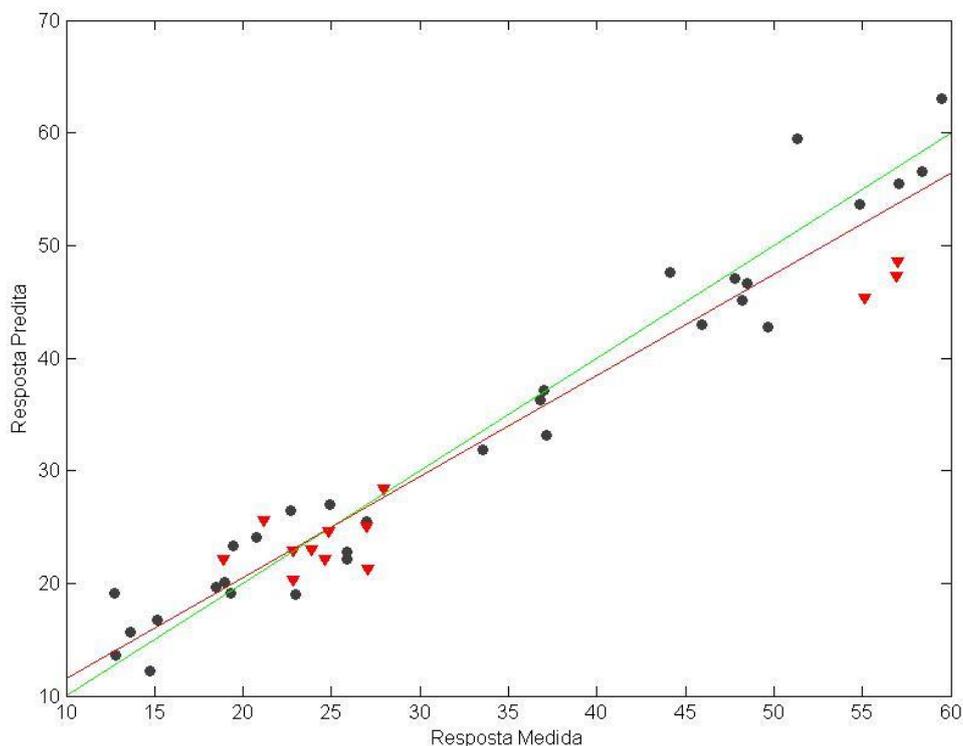


Figura 8.7. Valor previsto para oxigênio total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elemental.
 (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

8.5 Determinação de oxigênio total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)

Para a determinação de oxigênio nas amostras de biomassa e de biocarvão o iPLS selecionou um subconjunto de 100 variáveis incluídos em 13 intervalos diferentes. Os intervalos selecionados foram os seguintes: de 968,08~975,8; de 1257,3~1265; de 1373~1380,7; de 2308,3~2335,3; de 2790,4~2798,1; de 2829~2836,7; de

3031,5~3068,1; de 3272,5~3280,3; de 3301,5~3318,8; de 3349,7~3357,4; de 3388,3~3396; de 3436,5~3444,2; e de 3764,3~3781,7 cm^{-1} . Na Figura 8.8 é possível observar o espectro médio com os 13 intervalos contendo as 100 variáveis selecionadas pelo iPLS.

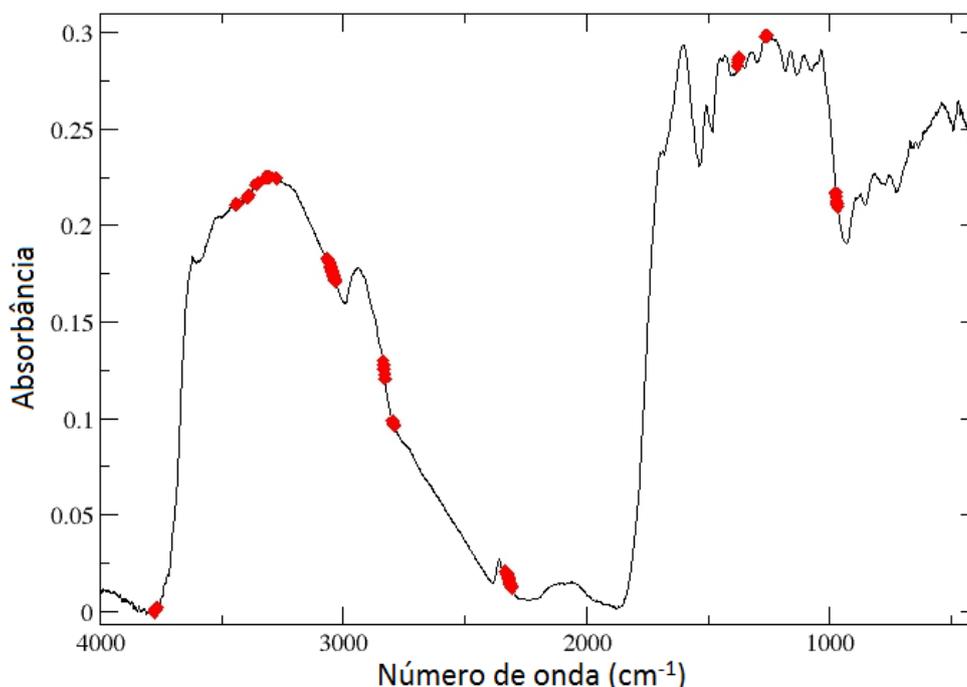


Figura 8.8. Espectro MID-IR médio das 44 amostras com destaque para as 100 variáveis selecionadas pelo iPLS correspondente a 13 intervalos de números de onda para a determinação da concentração de oxigênio.

O intervalo de 1257,3~1265 cm^{-1} indica a presença de ligação C-O de ácidos carboxílicos e/ou éteres, tanto alifáticos quanto aromáticos. O intervalo entre 1373~1380,7 cm^{-1} pode ser associado à presença de ligações N-O de compostos nítricos. Ligações O-H de ácidos carboxílicos, álcoois e fenóis são associados nos intervalo entre 3031,5~3068,1; 3272,5~3280,3; 3301,5~3318,8; 3349,7~3357,4; 3388,3~3396; e 3436,5~3444,2 cm^{-1} . No último intervalo escolhido, de 3764,3~3781,7 cm^{-1} também pode indicar a presença de ligações O-H de álcoois e fenóis (Solomons; Fryhle, 2008). As variáveis de 968,08~975,8; 2308,3~2335,3 e de 2790,4~2798,1 cm^{-1} não puderam ser identificadas de forma a serem relacionadas diretamente com a presença de oxigênio nas amostras, mas algumas delas apresentaram relações com outros teores estudados. Um exemplo ocorre no intervalo de 2790,4~2798,1 cm^{-1} , que pode ser identificado pelo estiramento de ligações C-H em aldeídos.

O melhor modelo obtido pela utilização dos intervalos obtidos pelo iPLS foi com a presença de 6 variáveis latentes (Figura 8.9). A variância acumulada dos blocos de variáveis independentes X e de variáveis dependentes Y para esse modelo de 99,99% e 98,80%, respectivamente.

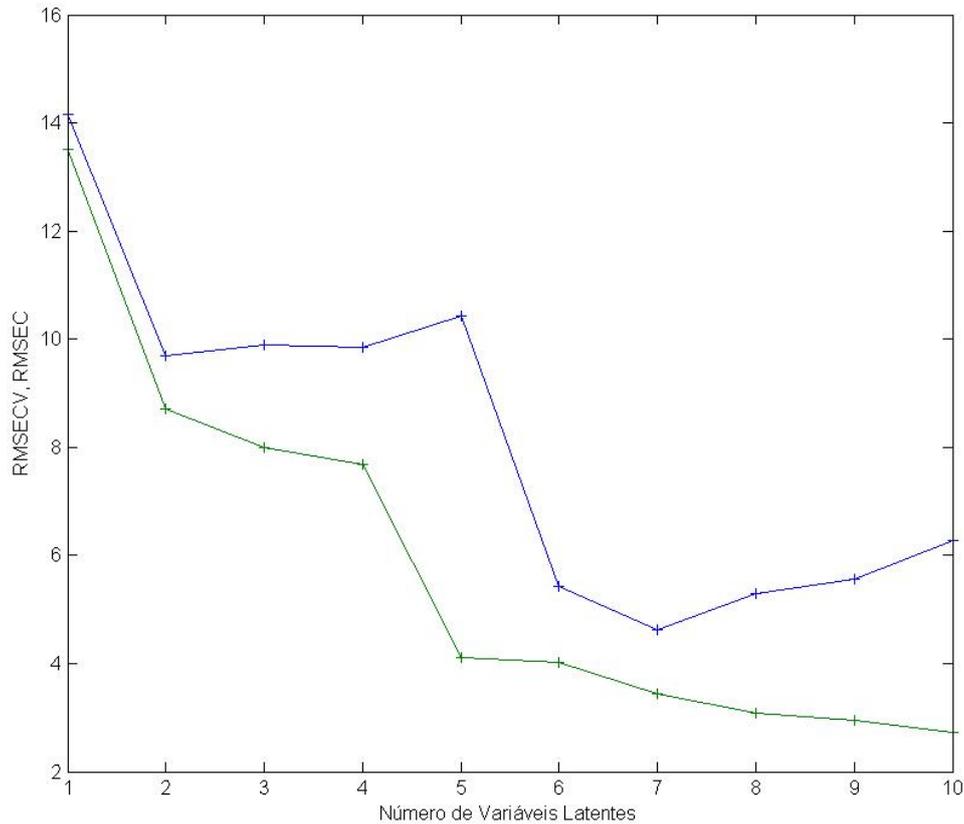


Figura 8.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de oxigênio total.

O RMSEC para o modelo utilizado apresentou o valor de 40,1 g/kg, o RMSEP o valor de 55,1 g/kg e o RMSECV o valor de 54,2 g/kg. Os coeficientes de correlação para a calibração e a validação foram, respectivamente, 0,9292 e 0,9145 (Figura 8.10). Os coeficientes R^2 nas etapas de calibração e predição apresentaram ajuste semelhantes, mas ocorre uma pequena disparidade entre os valores de RMSEC e RMSEP, sugerindo assim um leve sobre ajuste do modelo na etapa de calibração dos dados.

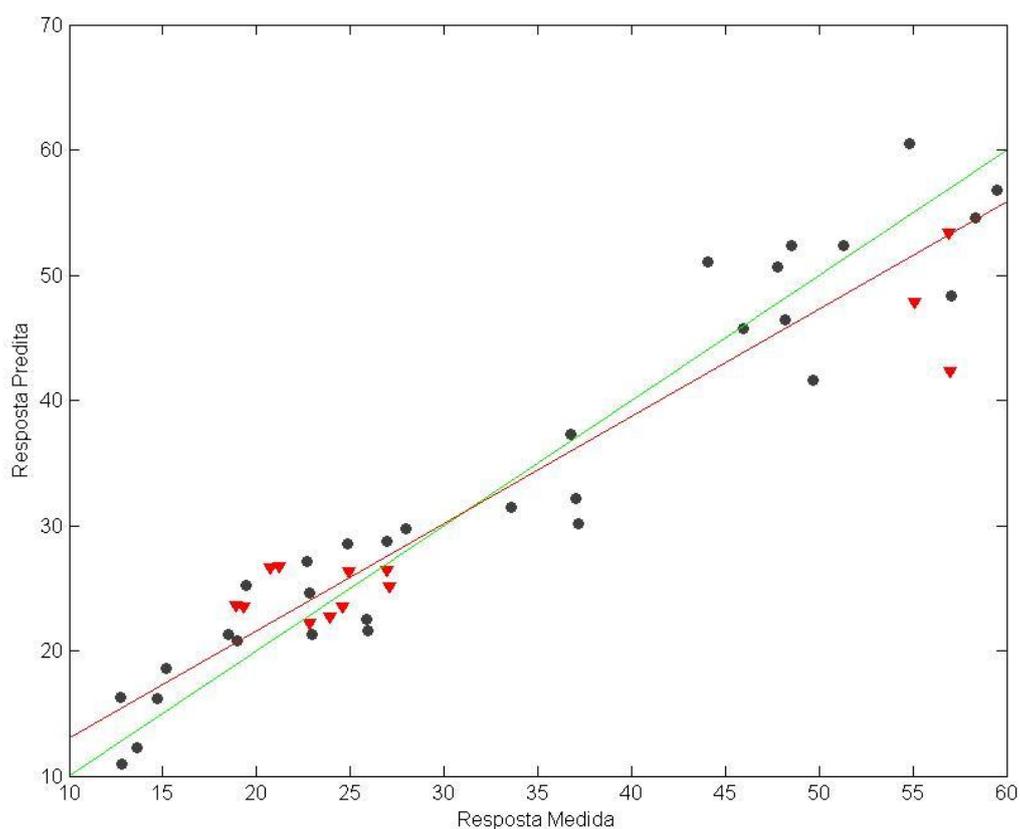


Figura 8.10. Valor previsto para oxigênio total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elementar. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

Tabela 8.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de oxigênio total.

Proposta	Nº de variáveis selecionadas	RMSEC (g/kg)	R ² Calibração	RMSEP (g/kg)	R ² Validação	RPD
SPA-MLR	6	25,7	0,985	52,0	0,951	4,517
PLS	-	38,0	0,936	50,9	0,939	4,048
GA-MLR	168	39,7	0,934	45,8	0,907	2,374
GA-PLS	502	32,7	0,954	51,2	0,957	4,822
iPLS	100	40,1	0,929	55,1	0,914	3,409

Na Tabela 8.1 estão contidos os resultados correspondentes a cada proposta apresentada para a determinação de oxigênio. Dentre as propostas utilizadas, o SPA-MLR e GA-PLS apresentam os melhor do modelo nas etapa de calibração e predição. Já a proposta GA-MLR apresenta valores menores e mais próximos de RMSEC e RMSEP, e coeficientes R^2 relativamente aproximados, tornando assim também um modelo mais preciso para a quantificação de oxigênio em amostras de biomassas e biocarvões, apesar do ajuste das curvas de calibração e predição inferiores em relação ao SPA-MLR e ao GA-PLS.

CAPÍTULO 9

DETERMINAÇÃO DE HIDROGÊNIO

9.1 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 1 (SPA-MLR)

Para a análise do hidrogênio por Regressão Linear Múltipla, o Algoritmo das Projeções Sucessivas selecionou 6 variáveis de frequência do espectro total. As variáveis selecionadas foram: 582,39; 1035,5; 1508; 1565,9; 3087,4 e 3411,4 cm^{-1} . A Figura 9.1 apresenta o gráfico médio dos espectros de infravermelho das amostras contendo as 6 variáveis selecionadas pelo SPA.

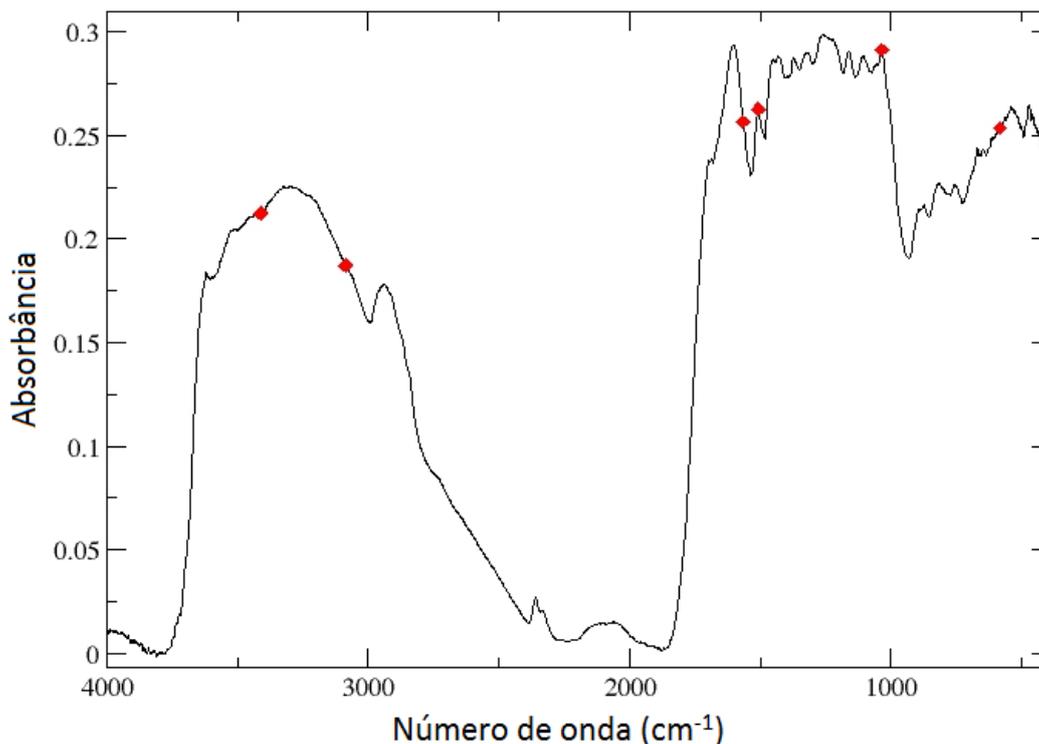


Figura 9.1. Espectro MID-IR médio das 44 amostras com destaque para as 6 variáveis selecionadas pelo SPA para a determinação de hidrogênio total.

De acordo com a literatura os comprimentos de onda em 989,3 e 3087,4 cm^{-1} estão associados a presença de ligações C-H de grupos vinílicos. O número de onda escolhido em 1565,9 cm^{-1} é proveniente de ligações N-H de aminas primárias (Solomons; Fryhle, 2008). O número de onda em 3411,4 cm^{-1} pode ser associado com o estiramento de ligações O-H de álcoois e fenóis que estão presentes na lignina que está presente nas

amostras de biomassa (Xu et al., 2013). Os números de onda 1035,5 e 3411,4 cm^{-1} foram também selecionados pelo SPA na proposta 1 para a determinação de carbono total, sendo assim a possibilidade de existir baixa multicolinearidade nestes postos selecionados. Os números de onda de 582,39; 1035,5 e 1508 cm^{-1} não puderam ser interpretados diretamente com a presença de grupos funcionais contendo hidrogênio nas amostras. O número de onda em 1508 cm^{-1} pode ser relacionado com o carbono, pois identifica o estiramento de ligações C=C em anéis aromáticos.

Os valores de RMSEC, RMSEP e RMSECV obtidos foram 4,6 g/kg, 4,4 g/kg e 7,2g/kg, respectivamente. O coeficiente de correlação R^2 obtido apresentou o valor de 0,9096 para a validação e 0,9233 para a calibração. O modelo obtido apresenta capacidade de quantificação de teores de hidrogênio nas amostras quando comparado os valores de RMSEC e RMSEP com os dados da Tabela 6.1. Na Figura 9.2 pode-se observar que os pontos preditos pelo modelo se ajustam bem a curva experimental, havendo poucas exceções de pontos afastados e uma dispersão mais uniforme.

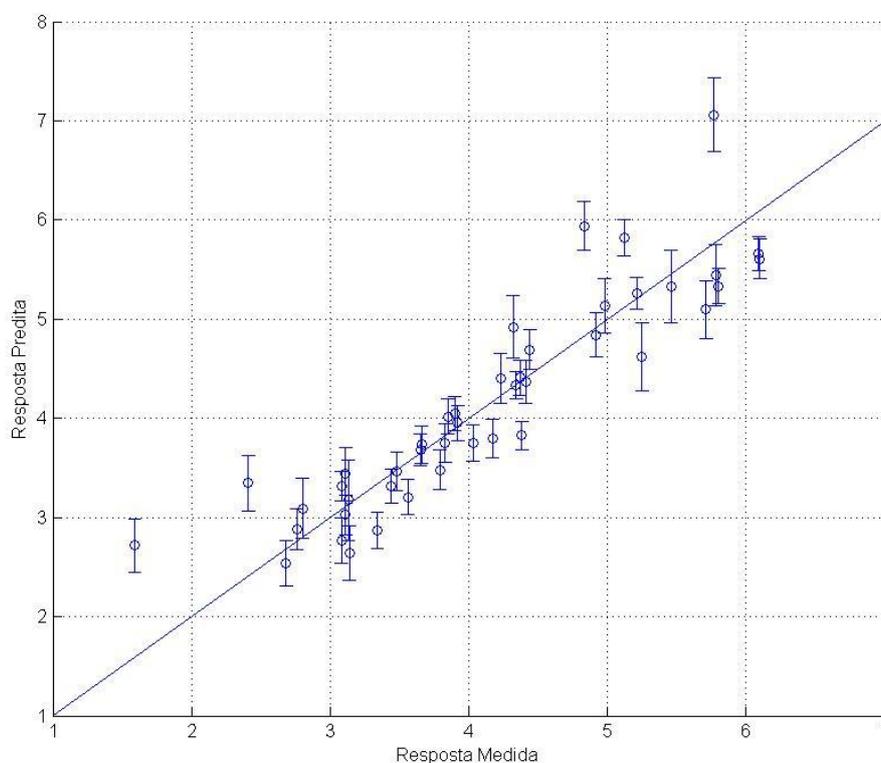


Figura 9.2. Valor previsto para hidrogênio total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta predita e as barras a resposta experimental da análise elementar.

9.2 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 2 (PLS)

Na regressão por PLS foram utilizadas 5 variáveis latentes. Na Figura 9.3, pode-se observar que após utilizar de 5 variáveis latentes, A variância acumulada dos blocos de variáveis independentes X foi de 99,55% e no bloco de variáveis dependentes Y foi de 99,14%. A RMSEC apresentou o valor de 4,0 g/kg, o RMSEP de 4,6 g/kg e o RMSECV de 6,1 g/kg. O coeficiente de determinação da calibração foi de 0,889 e da validação foi de 0,858 (Figura 9.4). O valores de RMSEC e RMSEP para o modelo obtido pelo PLS apresentam valores aproximadamente semelhantes, assim como os coeficientes obtidos pelo ajuste nas etapas de calibração e predição, caracterizando assim a capacidade para a quantificação de teores de hidrogênio em amostras de biomassas e biocarvões.

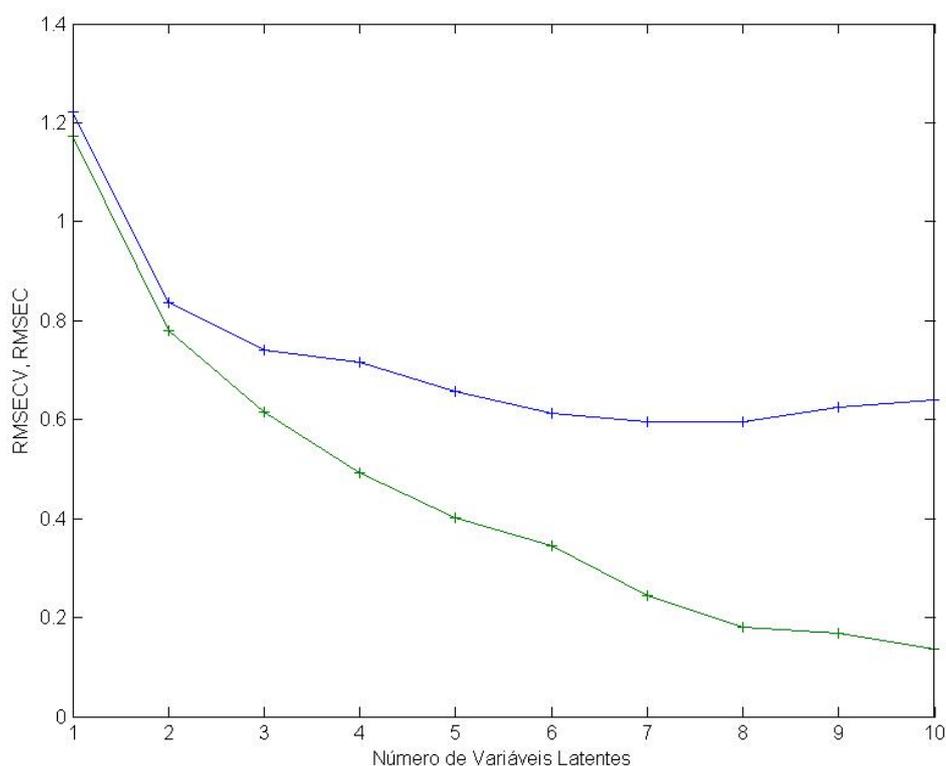


Figura 9.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de hidrogênio total.

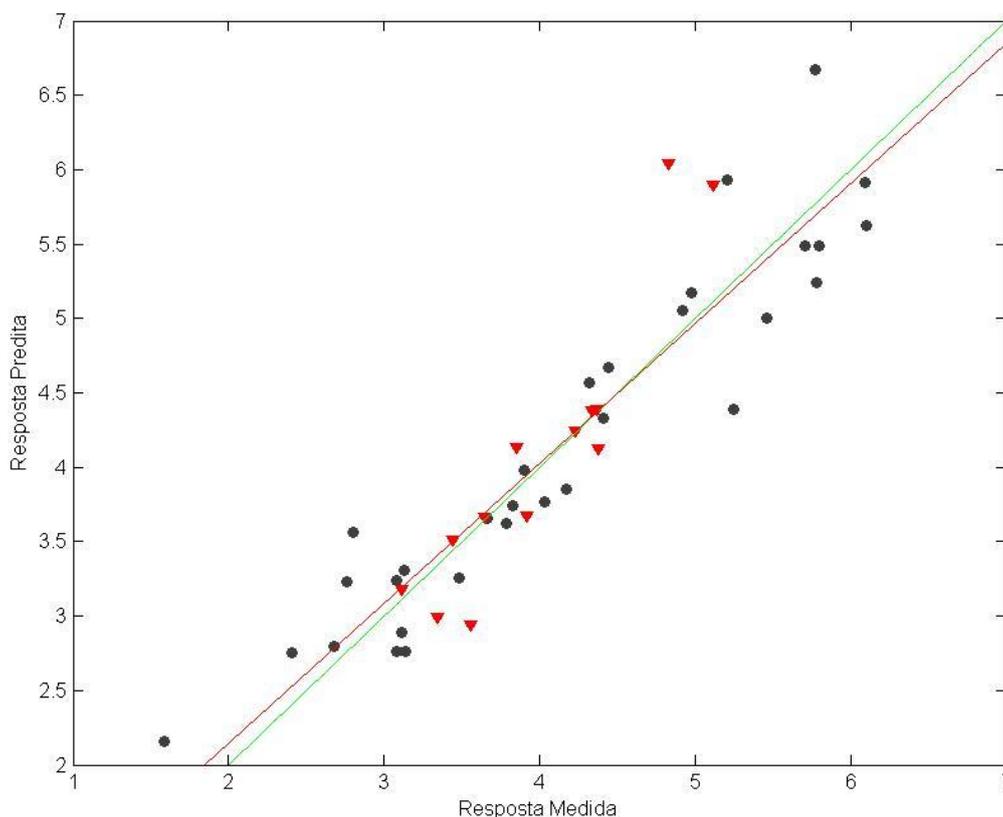


Figura 9.4. Valor previsto para hidrogênio total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elemental.
 (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

9.3 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)

O Algoritmo Genético com validação cruzada em RLM selecionou um subconjunto de 159 variáveis do conjunto do espectro total. Das 159 variáveis escolhidas 12 variáveis (806,09; 811,88; 1286,2; 1500,3; 2304,5; 2321,8; 2343; 2832,9; 2838,6; 2854,1; 3446,1 e 3507,8 cm^{-1}) são comuns ao GA e as técnicas SPA e iPLS. Algumas destas variáveis propiciam suas identificações diretamente relacionadas com modos vibracionais do átomo de hidrogênio e foram apresentadas nas seções 9.1 e 9.5 das propostas propostas 1 e 5. A identificação química relacionado ao modos vibracionais das bandas de absorção do espectro de infravermelho mostra que essas variáveis são significativa para a construção de modelos multivariados mais precisos para a quantificação de oxigênio em amostras de biomassas e biocarvões.

Os valores de RMSEC, RMSEP e RMSECV para o modelo foram de 2,4 g/kg, 6,2 g/kg, e 5,7 g/kg, respectivamente. O coeficiente de determinação para a calibração foi de 0,957 e para a validação foi de 0,839 (Figura 9.5). Como ocorrido para o PLS, o modelo obtido se mostra sobre ajustado para a etapa de calibração, como ocorre na proposta anterior pela calibração por PLS. O valor de RMSEC foi muito inferior ao valor de RMSEP, significando assim um melhor ajuste do modelo para explicar o conjunto de dados de calibração mas pouco preciso para a predição de outros conjuntos de dados. Portanto, o modelo obtido está sobre ajustado em relação aos dados de predição.

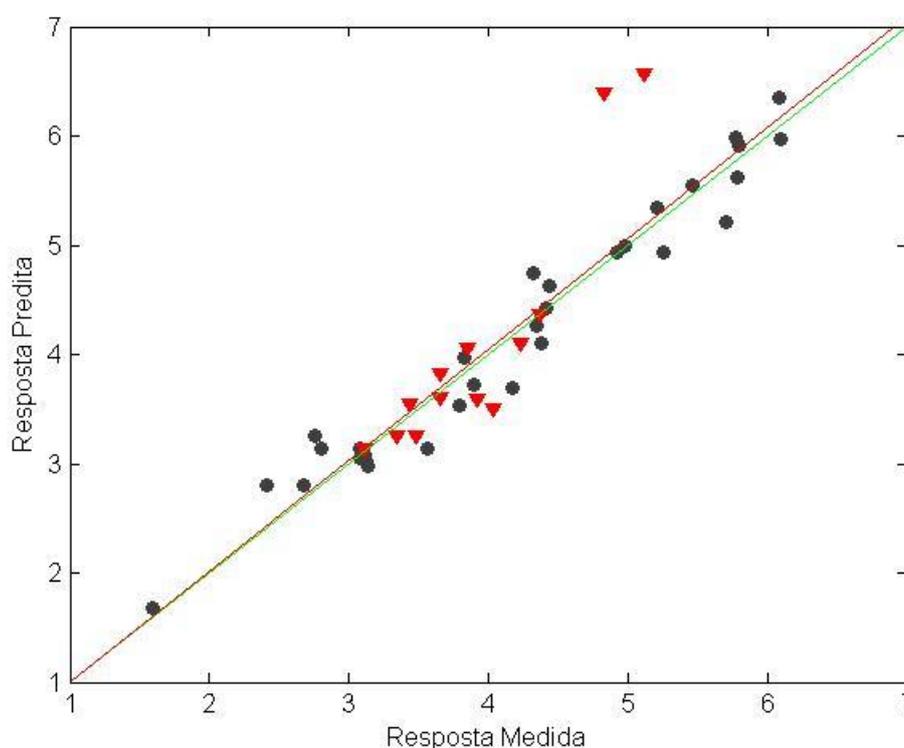


Figura 9.5. Valor previsto para hidrogênio total pelo modelo de calibração MLR (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

9.4 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)

Com a utilização do Algoritmo Genético pela validação cruzada em PLS, foram selecionados um subconjunto de 493 variáveis. Dentre essas variáveis selecionadas, pode-se evidenciar as seguintes: 401,11; 804,16; 806,09; 809,95; 811,88; 1070,2; 1286,2; 1565,9; 1581,3; 1583,2; 1590,9; 1592,9; 1602,5; 2300,6; 2304,5; 2321,8; 2335,3;

2341,1; 2842,5; 2846,4; 2848,3; 2852,1; 2854,1; 3448; 3457,7; 3459,6 e 3511,7 cm^{-1} , que também foram selecionadas em comum com as técnicas SPA e iPLS. A repetição do aparecimento destas frequências em técnicas distintas de seleção de variáveis indica que elas tem maior relevância para a construção do modelo multivariado em relação as demais variáveis presentes no conjunto de dados. Algumas delas são identificadas e diretamente relacionadas com a presença de hidrogênio nas amostras, como pode ser visto nas propostas 1 e 5 das seções 9.1 e 9.5.

Essas variáveis aplicadas na regressão por PLS forneceram um modelo mais adequado com a utilização de 5 variáveis latentes. Esse modelo apresentou a variância acumulada do bloco de variáveis independentes X em 99,54% e do bloco variáveis dependentes Y em 99,15%. Na Figura 9.6, podemos observar que a partir de 5 variáveis latentes, a utilização de mais variáveis leva a dispersão do modelo comparando-se o RMSEC com o RMSECV.

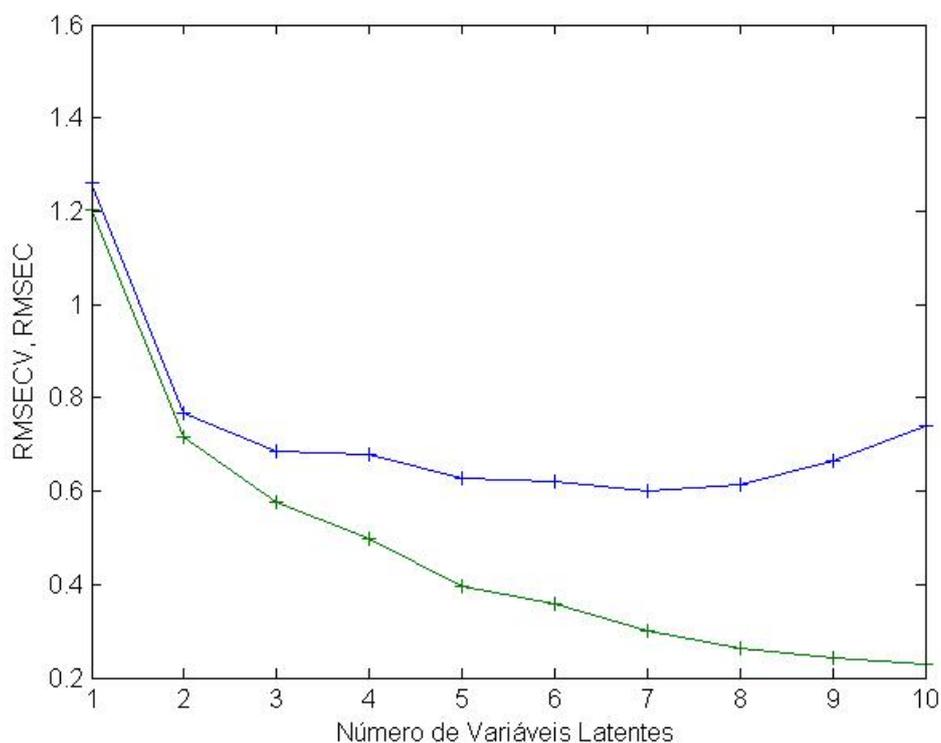


Figura 9.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de hidrogênio total.

Os valores de RMSEC, RMSEP e RMSECV obtidos pelo modelo com 5 variáveis latentes foram 3,9 g/kg, 3,1 g/kg, e 6,0 g/kg, respectivamente. Os coeficientes de determinação R^2 para esse modelo na calibração e validação foram 0,884 e 0,871 (Figura 9.7), respectivamente. O modelo obtido pela utilização do GA-PLS apresentam a capacidade de quantificação dos teores de hidrogênio. Isso pode ser explicado devido aos valores aproximadamente semelhantes entre os valores de RMSEC e RMSEP, e também em relação ao ajuste das curvas nas etapa de calibração e predição, 0,884 e 0,871, sugerindo assim a capacidade do modelo de predição dos dados aproximadamente igual a etapa de calibração.

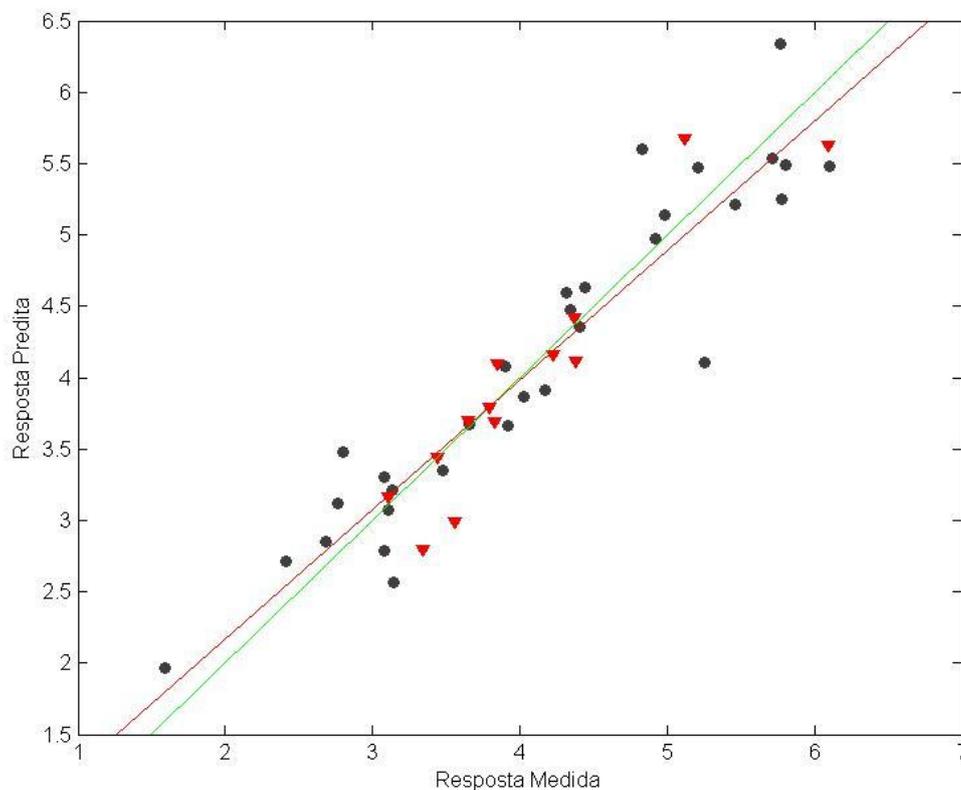


Figura 9.7. Valor previsto para hidrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

9.5 Determinação de hidrogênio total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)

A utilização do iPLS para a seleção de variáveis para a determinação de hidrogênio nas amostras resultou na seleção de 105 variáveis contidas em 11 intervalos distintos. Os intervalos selecionados foram: entre 399,19~406,9; 804,16~811,88; 1064,5~1072,2; 1286,2~1303,6; 1459,8~1467,5; 1498,4~1506,1; 1585,1~1602,5; 2298,7~2354,6; 2829~2856; 3446,1~3463,5; 3504~3511,7 cm^{-1} . O espectro médio com os 11 intervalos contendo as 105 variáveis selecionadas podem ser visualizadas na Figura 9.8.

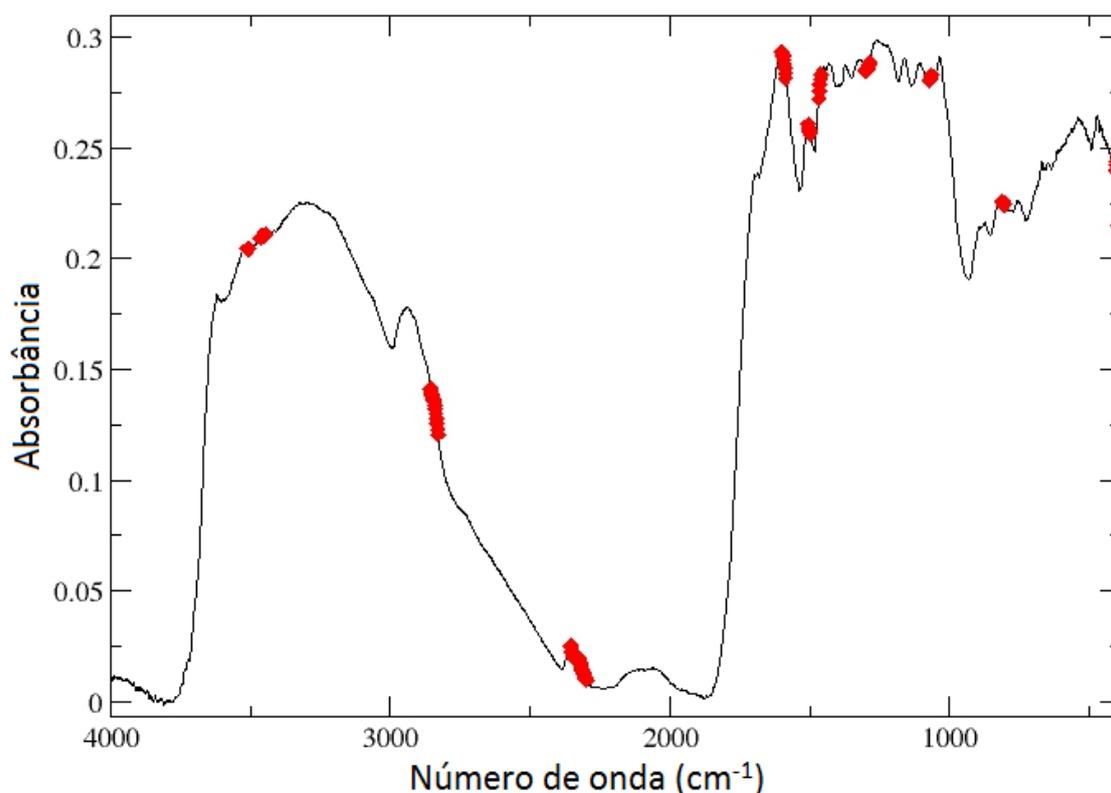


Figura 9.8. Espectro MID-IR médio das 44 amostras com destaque para as 105 variáveis selecionadas pelo iPLS correspondente a 11 intervalos de números de onda para a determinação da concentração de hidrogênio.

No intervalo entre 804,16~811,88 cm^{-1} pode ser associado com o estiramento de ligações C-H de estruturas aromáticas dissubstituídas, que podem ser encontradas na lignina (Xu et al., 2013). Os intervalos entre 1286,2~1303,6; 1459,8~1467,5; 1498,4~1506,1; e 2829~2856 cm^{-1} identificam ligações C-H em grupos alquilas, mas apresentam um pouco de deslocamento. A ligação N-H de aminas primárias podem ser

associadas nos intervalos de 1585,1~1602,5; 3446,1~3463,5; e 3504~3511,7 cm^{-1} (Solomons; Fryhle, 2008). Os números de onda dos intervalos entre 399,19~406,9; 1064,5~1072,2; e 2298,7~2354,6 cm^{-1} não puderam ser interpretados de forma a serem correlacionadas de forma direta com o teor total de hidrogênio.

A regressão por PLS foi feita com a utilização das 105 variáveis selecionadas pelo iPLS, e o modelo mais adequado obtido foi com a 3 variáveis latentes (Figura 9.9). Esse modelo apresentou a variância acumulada do bloco de variáveis independentes X em 99,32% e do bloco variáveis dependentes Y em 98,15%. O RMSEC para o modelo apresentou o valor de 5,8 g/kg, o RMSEP de 4,4 g/kg e o RMSECV de 7,8 g/kg. O coeficiente de determinação para a calibração foi de 0,769 e para a validação foi de 0,739 (Figura 9.10). O RMSEP apresentou um valor ligeiramente inferior em relação ao RMSEC, sugerindo que o modelo obtido apresenta um menor erro quando aplicado sobre o conjunto de dados de predição. Entretanto, o coeficiente de determinação na etapa de calibração foi maior do que na etapa de predição, sugerindo que o modelo se ajusta melhor aos dados de calibração. O modelo obtido pelo iPLS, devido aproximação entre os valores obtidos na etapas de calibração e predição, têm a capacidade de quantificação dos teores de hidrogênio em amostras de biomassas e biocarvões.

Uma possível causa para os sobre ajustes observados nas propostas PLS, GA-MLR, GA-PLS e iPLS pode estar relacionado com a baixa concentração de hidrogênio nas amostra de biomassas e biocarvões. Apesar da concentração de hidrogênio ainda ser significativa, a presença de sinais provenientes da análise por espectroscopia no infravermelho pode ser sobreposta pelos sinais mais claros provenientes do carbono e do oxigênio, que possuem concentrações muito maiores em todas as amostras. Deste modo os sinais de hidrogênio nos espectros são superpostos e não são gerados nos espectros informações significativas para a construção de modelos multivariado robustos e que permitam a quantificação mais precisa de concentração de hidrogênio nas amostras.

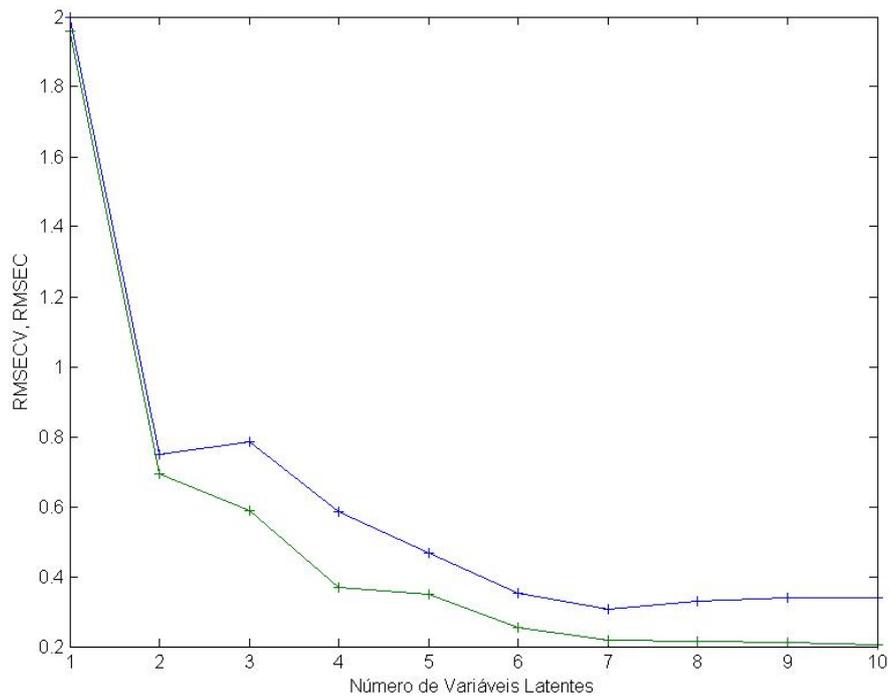


Figura 9.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de hidrogênio total.

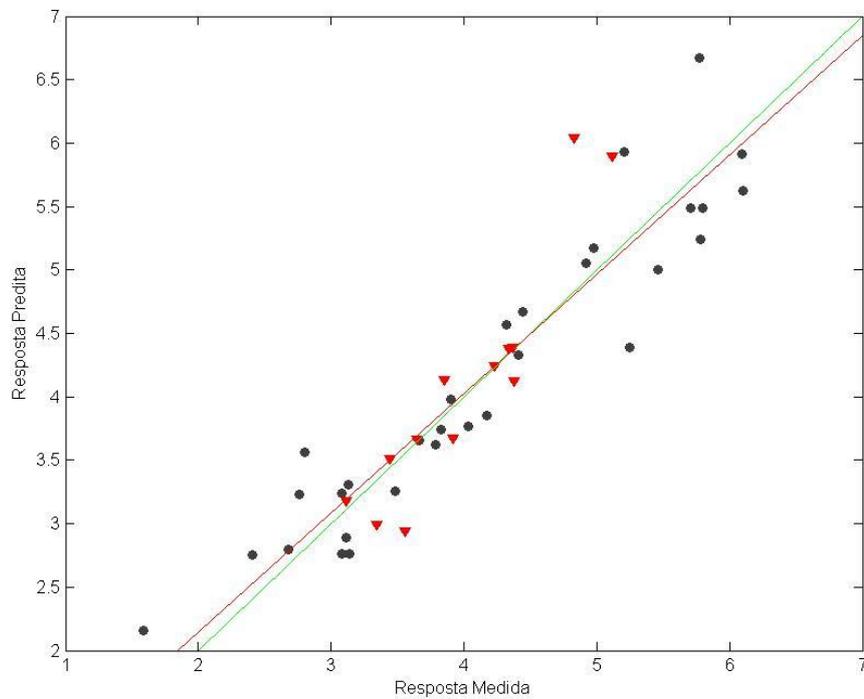


Figura 9.10. Valor previsto para hidrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elemental. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

Na Tabela 9.1 estão contidos os resultados correspondentes a cada proposta apresentada para a determinação de hidrogênio. Analisando a tabela, as propostas SPA-MLR, PLS e GA-PLS apresentam os menores valores de RMSEC e RMSEP, e contêm um bom ajuste nas etapas de calibração e predição, tornando-os assim aptos para quantificação de teores de hidrogênio em amostras de biomassas e biocarvões. O GA-MLR apresenta sobreajuste devido ao valor de RMSEC ser muito inferior ao valor de RMSEP, e o ajuste da curva do modelo na etapa de calibração ser muito superior em relação à etapa de predição. O iPLS apresenta um valor de RMSEC maior do que o valor de RMSEP, tornando assim questionável a qualidade do modelo para a predição de outros dados.

Tabela 9.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de hidrogênio total.

Proposta	Nº de variáveis selecionadas	RMSEC (g/kg)	R ² Calibração	RMSEP (g/kg)	R ² Validação	RPD
SPA-MLR	6	4,6	0,923	4,4	0,909	3,314
PLS	-	4,0	0,889	4,6	0,858	2,663
GA-MLR	159	2,4	0,957	6,2	0,839	2,492
GA-PLS	493	3,9	0,884	3,1	0,871	2,784
iPLS	105	5,8	0,769	4,4	0,739	1,957

Os resultados obtidos para o teor elementar de hidrogênio, assim como para o carbono e oxigênio, reforçam a capacidade da utilização dos dados espectrais de infravermelho para a quantificação de teores elementares presentes em um determinado tipo de amostra. As amostras de biomassas e biocarvões apresentam uma alta complexidade em relação a outros tipos de amostras passíveis de análise quantitativa de infravermelho, como por exemplo determinação de carbono em solos (Madari et al., 2005; Reeves et al., 2002), adulterantes em leite (Ferrão et al., 2007), matéria orgânica em solos (Souza et al., 2012) e adulterantes em café (Tavares et al., 2012), entretanto os resultados obtidos apresentaram qualidade e confiabilidade semelhantes quando comparado com esses outros tipos de amostras.

CAPÍTULO 10

CONCLUSÕES

As propostas apresentadas e utilizadas resultaram na obtenção de diversos modelos capazes de quantificar os teores elementares presentes nas amostras de biomassas e biocarvões: carbono, hidrogênio e oxigênio. Para cada teor elementar as propostas utilizadas forneceram alguns modelos mais precisos em relação aos outros obtidos. Dentre os modelos quimiométricos estudados, a proposta SPA-MLR apresentou sempre resultados satisfatórios para modelos de quantificação de carbono, oxigênio e nitrogênio. A utilização do SPA na seleção de variáveis se mostrou de grande valia para a obtenção de modelos precisos, e que forneçam maior capacidade preditiva para a quantificação de outros conjuntos de dados.

Dos modelos obtidos para a quantificação de carbono, as propostas SPA-MLR e iPLS apresentaram os melhores modelos, tanto pela observação do ajuste da curva na etapa de predição quanto pelo parâmetro RPD calculado. Para a quantificação do oxigênio, os melhores modelos foram obtidos pelas propostas SPA-MLR, PLS, GA-MLR e GA-PLS, destacando-se o GA-MLR que apesar de ter um parâmetro RPD baixo em relação às outras propostas possui os valores de RMSEC e RMSEP menores e mais próximos. Para o hidrogênio, os melhores modelos de quantificação foram obtidos com as propostas SPA-MLR e PLS. O PLS, apesar do ajuste inferior das curvas de calibração e predição em relação ao SPA, apresentou valores de RMSEC e RMSEP mais próximos, revelando assim que o modelo obtido tem a capacidade de quantificação semelhante nas etapas de calibração e predição.

A utilização das técnicas HCA e PCA revelaram os padrões ocorrentes nas amostras estudadas com relação direta entre as concentrações de carbono e oxigênio. O HCA realizou a separação de 4 grupos, sendo que dois deles foi possível observar características bem definidas. Em um grupo foram separadas todas as amostras de biomassas estudadas, contendo alto teor de oxigênio e baixo teor de carbono, e em outro grupo foram separadas a maioria das amostras de biocarvão produzidas à temperatura de 550°C, contendo alto teor de carbono e baixo teor de oxigênio.

Já o PCA, a separação das amostras se deu em relação às diferenças de concentrações entre carbono e oxigênio no decorrer do aumento da temperatura de pirólise nas amostras. Algumas amostras apresentaram maior variação entre a redução da concentração de oxigênio e o aumento da concentração de carbono de acordo com o aumento da temperatura de pirólise, enquanto outras amostras apresentaram uma menor variação. Assim, as características das amostras ficam bem definidas com relação direta às concentrações de carbono e oxigênio, sendo estes os elementos de maior porcentagem nas amostras.

A utilização das propostas para o nitrogênio (ver Apêndice A) não resultaram na obtenção de um modelo capazes de quantificar esse teor elementar em amostras de biomassas e biocarvões, possuindo erros muito maiores que a ordem de grandeza observada para porcentagem de nitrogênio nas amostras por análise elementar. Isso pode ter ocorrido em decorrência da baixa concentração de nitrogênio existentes nas amostras de biomassas e biocarvões quando comparado com as concentrações superiores de hidrogênio e principalmente carbono e oxigênio, indicando assim que os dados referentes ao nitrogênio no espectro de infravermelho são poucos e também ocultados devido à carga muito maior de informação provenientes de hidrogênio, carbono e oxigênio.

REFERÊNCIAS BIBLIOGRÁFICAS

ALLISON G. G. Application of Fourier Transform Mid-Infrared Spectroscopy (FTIR) for Research into Biomass Feed-Stocks. *Fourier Transforms - New Analytical Approaches and FTIR Strategies*, p.520, 2011. Disponível em: <<http://www.intechopen.com/books/fouriertransforms-new-analytical-approaches-and-ftir-strategies/application-of-fourier-transform-mid-infrared-spectroscopy-ftir-for-research-into-biomass-feed-stock>> Acesso em 22 abr. 2014.

ARAÚJO, M. C. U.; SALDANHA, T. C. B.; GALVÃO, R. K. H.; YONEYAMA, T.; CHAME, H. C.; VISANI, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems*, v.57, n.2, p.65-73, 2001.

BROWNSORT, PA; CROMBIE, K; MASEK, O; TURRION-GOMEZ, J. Pyrolysis Biochar Syatems: Pilot Scale Pyrolysis Plant for 'Specified Biochar' Production. 19th European Biomass Conference and Exhibition, Berlin, Germany. 2011. Disponível em: <<http://www.biochar.ac.uk/abstract.php?id=44&pr=a>> Acesso em: 02 mar. 2014.

BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. F. Introdução à Análise de Agrupamentos. 9º Simpósio Nacional de Probabilidade e Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, 1990.

AHN, C. W. Advances in Evolutionary Algorithms. *Studies in Computational Intelligence*, v.18, p.167, 2006.

CHATFIELD, C.; COLLINS, A. J. *Introduction to Multivariate Analysis*, Cambridge, 1 ed, 1980.

CHEN, W.; MATTERN, D. L.; OKINEDO, E.; SENTER, J. C.; MATTEI, A. A.; REDWINE, C. W. Photochemical and Acoustic Interactions of Biochar with CO₂ and H₂O: Applications in Power Generation and CO₂ Capture. *AIChE Journal*, v.60, n.3, 2014.

CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: Explorando procedimentos quimiométricos para o tratamento de dados analíticos. *Química Nova*, v. 30, n. 2, p. 481-487, 2007.

DASZYKOWSKI, M.; SERNEELS, S.; KACZMAREK, K.; ESPEN, P. V.; CROUX, C.; WALCZAK, B. TOMCAT: a MATLAB toolbox for multivariate calibration techniques. *Chemometrics and Intelligent Laboratory Systems*, v.85, n.2, p.269-277, 2007.

DASZYKOWSKI, M.; SERNEELS, S.; KACZMAREK, K.; ESPEN, P. V.; CROUX, C.; WALCZAK, B. Toolbox for Multivariate Calibration Techniques (TOMCAT). The University of Silesia - Institute of Chemistry, 2007. Disponível em <<http://www.chemometria.us.edu.pl/index.php?goto=downloads>> Acesso em: 13 mar. 2014.

DAY, D.; EVANS, R. J.; LEE, J. W.; REICOSKY, D. Economical CO₂, SO_x, and NO_x capture from fossil-fuel utilization with combined renewable hydrogen production and large-scale carbon sequestration. *Energy*, v.30, p.2558-2579, 2009.

FERRÃO, M. F.; MELLO, C.; BORIN, A.; MARETTO, D. A.; POPPI, R. J. LS-SVM: uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. *Química Nova*, v.30, n.4, p.852-859, 2007.

FERREIRA, D. F. *Estatística Multivariada*. 2.ed. Lavras, Editora UFLA, p.661, 2008.

FOX, T. A.; CHAPMAN, L. *Engineering geo-engineering. Meteorological Applications*, v.18, n.1, p.1-8, 2011.

GALVÃO, R. K. H.; SALDANHA, T. C. B.; YONEYAMA, T.; ARAÚJO, M. C. U. Uma Técnica para a Redução da Cardinalidade de Padrões de Entrada de uma Rede Neural Artificial. IV Congresso Brasileiro de Redes Neurais, Instituto Tecnológico da Aeronáutica, São José dos Campos, p.124-129, 1999.

GLASER, B.; GUGGENBERGER, G.; HAUMAIER, L.; ZECH, W. Persistence of soil organic matter in archaeological soils (Terra Preta) of Brazilian Amazon region. Sustainable management of soil organic matter. Wallingfort: CAB International, p.190-194, 2000.

HAMMOND, J.; SHACKLEY, S.; SOHI, S.; BROWNSORT, P. Prospective life cycle carbon abatement for pyrolysis biochar systems in the UK. Energy Policy, v.39, n.5, p.2646-2655, 2011.

HAUMAIER, L.; ZECH, W. Black carbon – possible source of highly aromatic components of soil humic acids. Organic Chemistry, v.23, p.191-196, 1995.

IGNE, B.; HURBURGH, C. R. JR. Local chemometrics for samples and variables: optimizing calibration and standardization processes. Journal of Chemometrics, v.24, n.2, p.75-86, 2010.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC). Frequently Asked Question 1.3 What is the Greenhouse Effect?. Genebra, Suíça. Disponível em: <http://www.ipcc.ch/publications_and_data/ar4/wg1/en/faq-1-3.html>

INTERNATIONAL BIOCHAR INITIATIVE. About IBI. Philadelphia, EUA. Disponível em <<http://www.biochar-international.org/about>> Acesso em: 10 nov. 2014.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SOARES, S. F. C.; PAIVA, H. M. Signal processing and variable selection for multivariate calibration. Instituto Tecnológico de Aeronáutica (ITA) - Divisão de Engenharia Eletrônica, 2012. Disponível em <<http://www.ele.ita.br/~kawakami/spa/>> Acesso em: 25 fev. 2014.

JONG, S. SIMPLS: an alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems, v.18, n.3, p.251-263, 1993.

KENNARD, R.W.; STONE, L.A. Computer aided design experiments. Technometrics, v.11, p.137-148, 1969.

KOWALSKI, B. R. Chemometrics. Analytical Chemistry, v.52, n.5, p.112R-122R, 1980.

KRAMER, R. W.; KUJAWINSKI, E. B.; HATCHER, P. G. Identification of black carbon derived structures in a volcanic ash soil humic acid by fourier transform ion cyclotron resonance mass spectrometry. *Environmental Science & Technology*, v.38, n.12, p.3387-3395, 2004.

KUHLBUSCH, T. A. J. Black carbono and the carbono cycle. *Science*, v.280, p.1903-1904, 1998.

KUMADA, K. Carbonaceous materials as possible source of soil humus. *Soil Science and Plant Nutrition*, v.29, p.383-386, 1983.

LEARDI, R.; NORGAAD, L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral relevant. *Journal of Chemometrics*, v. 18, p. 486-497, 2004.

LEHMANN, J. A handful of carbon. *Nature*, v.447, n.7141, p.143-144, 2007.

LEHMANN, J. Bio-energy in the black. *Frontiers in Ecology and the Environment*, v.5, n.7, p.381-387, 2007.

MAIA, C. M. B. F. Produção de biocarvões por pirólise lenta: potencial agrícola e ambiental, X Encontro Brasileiro de Substâncias Húmicas, p.283-295, 2013.

MADARI, B. E.; REEVES, J. B.; COELHO, M. R.; MACHADO, P. L. O. A.; DE-POLLI, H.; COELHO, R. M.; BENITES, V. M.; SOUZA, L. F.; MCCARTY, G. W. Mid- and Near-Infrared Spectroscopic Determination of Carbon in a Diverse Set of Soils from the Brazilian National Soil Collection. *Spectroscopy Letters*, v.38, n.6, p.721-740, 2005.

MAITRA, S.; YAN, J. Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression. *Casualty Actuarial Society, Discussion Paper Program*, p.79-90, 2008. Disponível em: <<http://www.casact.org/pubs/dpp/dpp08/08dpp76.pdf>> Acesso em: 19 mar. 2014.

MANGRICH, A. S.; MAIA C. M. B. F.; NOVOTNY, E. H. Biocarvão: As terras pretas de índios e o sequestro de carbono. *Ciência Hoje*, v.47, n.281, p.48-52, 2011.

MCMURRY, J. Determinações de Estruturas: Espectrometria de Massa e Espectroscopia na região do Infravermelho. In: Química Orgânica. 6.ed. São Paulo, Cengage Learning, p.388-418, 2008.

MELLO, C. A. Redes Neurais e técnicas de pré-processamento de sinais em espectroscopia de reflectância no infravermelho próximo. Campinas, 1998, p.138. Tese de Doutorado – Instituto de Química de Campinas, Universidade Estadual de Campinas.

MITCHELL, M. An Introduction to Genetic Algorithms. Massachusetts Institute of Technology, 5.ed. p.158, 1999.

MORGANO, M. A.; FARIA, C. G.; FERRÃO, M. F.; FERREIRA, M. M. C. Determinação de açúcar total em café cru por espectroscopia no infravermelho próximo e regressão por mínimos quadrados parciais. Química Nova, v.30, n.2, p.346-350, 2007.

NETO, B. B.; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de Quimiometria no Brasil. Química Nova, v.29, n.6, 1401-1406, 2006.

NORGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B.; Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. Applied Spectroscopy, v.54, n.3, p.413-419, 2000.

NOVOTNY, E. H.; HAYES, M. H. B.; AZEVEDO, E. R.; BONAGAMBA, T. J. Characterisation of black carbon-rich samples by ^{13}C solid-state nuclear magnetic resonance. Naturwissenschaften, v.93, n.9, p.447-450, 2006.

NOVOTNY, E. H.; HAYES, M. H. B.; MADARI, B. E.; BONAGAMBA, T. J.; DE AZEVEDO, E. R.; DE SOUZA, A. A.; SONG, G.; NOGUEIRA, C. M.; MANGRICH, A. S. Lessons from the Terra Preta de Índios of the Amazon Region for the Utilization of Charcoal for Soil Amendment. Journal of the Brazilian Chemical Society, v.20, n.6, p.1003-1010, 2009.

PERKIN ELMER. 2400 Series II CHNS/O Elemental Analysis. Waltham, Massachusetts, EUA. Disponível em: <http://www.perkinelmer.com/CMSResources/Images/44-74386BRO_2400_SeriesII_CHNSO_Elemental_Analysis.pdf> Acesso em: 15 abr. 2015.

POIRIER, N.; DERENNE, S.; BALESIDENT, J.; ROUZAUD, J.; MARIOTTI, A.; LARGEAU, C. Abundance and composition of the refractory organic fraction of an ancient, tropical soil (Pointe Noire, Congo). *Organic Geochemistry*, v.33, n.3, p.383-391, 2002.

REEVES, J. B.; MCCARTY, G.; MIMMO, T. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environmental Pollution*, v.116, n.1, p.S277-S284, 2002.

REZENDE, E. I. P.; ANGELO, L. C.; DOS SANTOS, S. S.; MANGRICH, A. S. Biocarvão (Biochar) e Sequestro de Carbono. *Revista Virtual de Química*, v.3, n.5, p.426-433, 2011.

RINNAN, A; BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, v.28, n.10, p.1201-1222, 2009.

ROSIPAL R., KRAMER N. *Overview and Recent Advances in Partial Least Squares*. Springer-Verlag Berlin Heidelberg, p. 34–51, 2006.

SENA, M. M.; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. *Química Nova*, v.23, n.4, p.547-556, 2000.

SOHI, S. P.; KRULL, E.; LOPEZ-CAPEL, E.; & BOL, R. A review of biochar and its use and function in soil. *Advances in Agronomy*, v. 105, p.47–82, 2010. Doi:10.1016/S0065-2113(10)05002-9.

SOLOMONS, T. W. G.; FRYHLE, C. B. *Química Orgânica*, 9.ed, v1., p.43-75, LTC, 2008.

SOUZA, D. M.; MADARI, B. E.; GUIMARÃES, F. F. Aplicação de técnicas multivariadas e inteligência artificial na análise de espectros de infravermelho para determinação de matéria orgânica em amostras de solos. *Química Nova*, v.39, n.9, p.1738-1745, 2012.

STEINER, C. *Biochar Carbon Sequestration. The Green Disc: New Technologies for a New Future - Chapter 17*. UN Climate Change Conference, Copenhagen, 2009. Disponível em: <<http://www.biochar.org/joomla/images/stories/Steiner%20Chapter%2017%202009.pdf>> Acesso em: 15 mar. 2014.

STUART, B. H. Infrared spectroscopy: Fundamentals and applications. 1.ed. Chichester, John Wiley & Sons, p.224, 2004.

TAMANINI, C.; HAULY, M. C. O. Resíduos agroindustriais para produção biotecnológica de xilitol. *Semina: Ciências Agrárias*, v. 25, n.4, p.315-330, 2004.

TAVARES, K. M.; PEREIRA, R. G. F. ALVARENGA; NUNES, C. A.; PINHEIRO, A. C. M.; RODARTE, M. P.; GUERREIRO, M. C. Espectroscopia no infravermelho médio e análise sensorial aplicada à detecção de adulteração de café torrado por adição de cascas de café. *Química Nova*, v.35, n.6, p.1164-1168, 2012.

TEIXEIRA, W. G.; KERN, D. C.; MADARI, B. E.; LIMA, H. N.; WOODS, W. As Terras Pretas de Índios da Amazônia: Sua Caracterização e Uso deste Conhecimento na Criação de Novas Áreas. Manaus, Embrapa Amazônia Ocidental, p.420, 2009.

TOBIAS, R. D. An Introduction to Partial Least Squares Regression. Proceedings of the Twentieth Annual SAS Users Group International Conference. SAS Institute Inc., Cary, North Carolina, US, 1997. Disponível em: <<http://www.sascommunity.org/sugi/SUGI95/Sugi-95-210%20Tobias.pdf>> Acesso em: 10 mar. 2014.

WILLIAMS P; NORRIS K. Implementation of Near-Infrared Technology. In *Near-Infrared Technology in the Agricultural and Food Industries*, 2^o edition, St. Paul, Minnesota, USA, 2001.

WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M.; WINDIG, W.; KOCH, R. S. PLS Toolbox 2.1 for use with MATLAB™, Eigenvector Research Inc.: Manson, 2001. Disponível em <http://www.eigenvector.com/software/pls_toolbox.htm> Acesso em: 1 mar. 2014.

WOLD, H. Partial least squares. *Encyclopedia of Statistical Sciences*, v.6, p.581-591, 1985.

WMO. World Meteorological Organization Greenhouse Gas Bulletin, n.10, p.8, 2014. Disponível em <
https://www.wmo.int/pages/mediacentre/press_releases/documents/1002_GHG_Bulletin.pdf> Acesso em: 20 jun. 2015.

XU, F.; YU, J.; TESSO, T.; DOWELL, F.; WANG, D. Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. *Applied Energy*, v.104, p.801–809, 2013.

ZHANG, L.; XU, C. (CHARLES); & CHAMPAGNE, P. Overview of recent advances in thermochemical conversion of biomass. *Energy Conversion and Management*, v.51, n.5, p.969–982. 2010. Doi:10.1016/j.enconman.2009.11.038.

ZIMMERMAN A. Abiotic and microbial oxidation of laboratory-produced black carbon (biochar). *Environmental Science & Technology*, v.44, n.4, p.1295-1301, 2010.

APÊNDICE A

DETERMINAÇÃO DE NITROGÊNIO TOTAL

A.1 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 1 (SPA – MLR)

Na determinação do nitrogênio, o Algoritmo de Projeções Sucessivas realizou a seleção de 5 variáveis a partir das 1869 frequências existentes no espectro de infravermelho. As variáveis selecionadas com números de onda: 422,33; 985,44; 1569,7; 3411,4 e 3654,4 cm^{-1} são apresentadas abaixo na Figura A.1 juntamente com o espectro médio das amostras. As variáveis selecionadas em 1569,7 e 3411,4 cm^{-1} são identificadas com sendo provenientes do modo vibracional de estiramento de ligações N-H de amins primárias (Solomons; Fryhle, 2008). As variáveis em 422,33; 985,44; e 3654,4 cm^{-1} não podem ser identificadas com relação direta à presença de modos vibracionais envolvendo o átomo de nitrogênio. Entretanto, a variável selecionada em 3654,4 cm^{-1} pode ser identificada e relacionada com a ocorrência de estiramentos de ligações O-H em álcoois e fenóis.

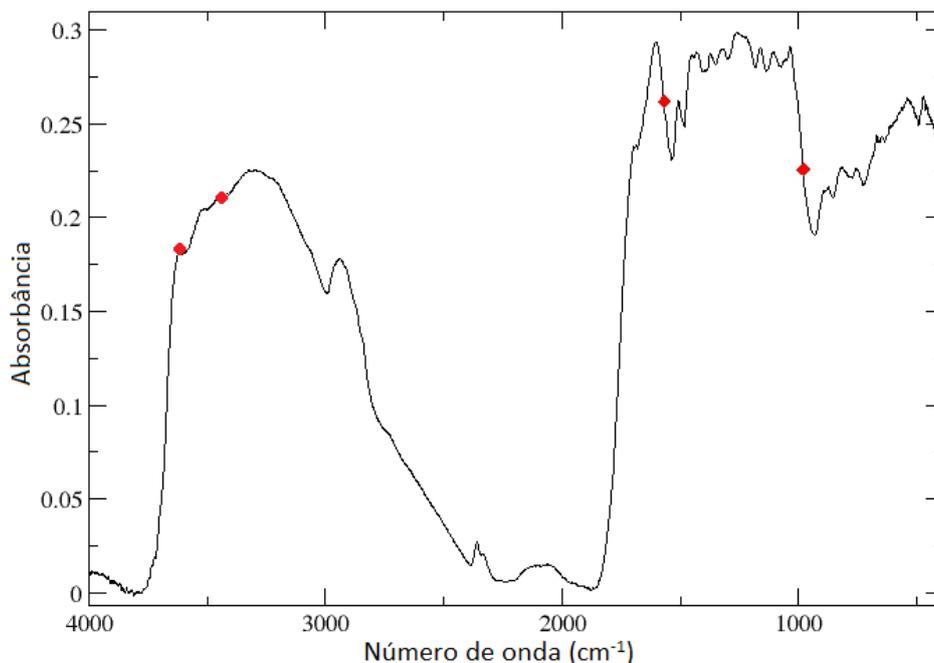


Figura A.1. Espectro MID-IR médio das 44 amostras com destaque para as 2 variáveis selecionadas pelo SPA para a determinação de nitrogênio.

Os valores de RMSEC, RMSEP e RMSECV deste modelo são, respectivamente, 1,5 g/kg, 3,8 g/kg e 2,0 g/kg. O coeficiente de determinação R^2 entre o modelo obtido com os dados encontrados na predição apresentou o valor de 0,7086, e o coeficiente de determinação para a fase de calibração foi de 0,9014 (Figura A.2). Os valores obtidos na etapa de calibração e predição se apresentaram insatisfatórios, pois o valores de RMSEP e RMSEC quando comparados com os valores obtidos pela análise elementar na Tabela 6.1 e se encontram na mesma faixa de concentração de nitrogênio presente nas amostras estudadas, levando a erros em certas amostra com valores acima de 100%. Portanto a utilização desta proposta para a quantificação de nitrogênio é inviável. Além disto, a curva de resposta na predição não se ajusta corretamente com a resposta medida, como pode ser concluído pelo o valor de R^2 de predição obtido. Na realidade nenhuma das propostas de modelos de calibração foi capaz de quantificar N nas amostras de biocarvão e biomassa e apesar deste modelo ter fornecido um ajuste de má qualidade ele foi o mais preciso quando comparado com as outras técnicas utilizadas para a quantificação de nitrogênio, que serão apresentadas nas próximas seções.

Uma possível causa do problema da quantificação de nitrogênio em amostras de biomassa e biocarvão são os valores de porcentagem na análise elementar para o nitrogênio no biocarvão, que se encontram na casa das unidades ou muitas vezes nas casas decimais em relação à porcentagem total, mostrando assim uma concentração baixíssima de nitrogênio nas amostras de biomassa e de biocarvão. Isso resulta em baixos sinais de resposta obtido da análise por infravermelho médio, tornando assim a construção de um modelo multivariado impreciso.

Na Figura A.2 pode se perceber o isolamento de 3 pontos na parte superior do gráfico de resposta. Enquanto, na parte inferior do gráfico observa-se o agrupamento de diversos pontos. Apesar da grande quantidade de dados nesta faixa de concentração a figura mostra uma dispersão entre a resposta do modelo e a experimental. Isto significa que o modelo apresenta uma baixa precisão para a quantificação de concentrações totais de nitrogênio, já que o mesmo não se ajusta apropriadamente com os dados experimentais.

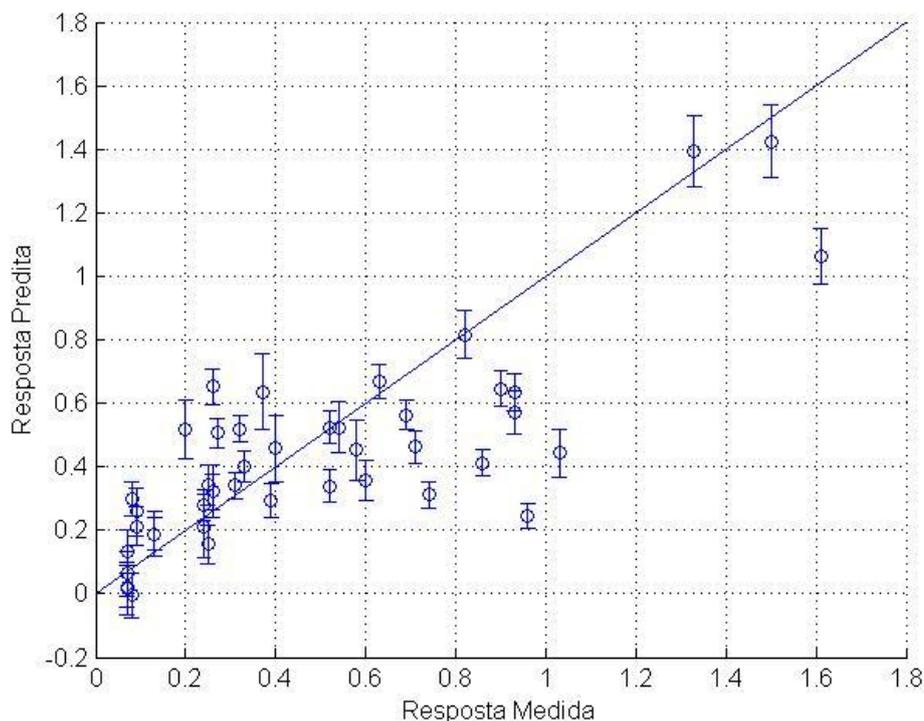


Figura A.2. Valor previsto para nitrogênio total pelo método SPA-MLR (%) contra o valor obtido pela análise elementar. A linha azul identifica resposta prevista e as barras a resposta experimental da análise elementar.

A.2 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 2 (PLS)

Os dados espectroscópicos foram submetidos à calibração por PLS sem a utilização de algum método de seleção de variáveis para a determinação do nitrogênio. O modelo mais apropriado foi construído com 5 variáveis latentes (ver Figura A.3), e os valores de RMSEC, RMSEP e RMSECV gerados são de 2,0 g/kg, 4,0 g/kg e 3,1 g/kg, respectivamente. A variância acumulada dos blocos de variáveis independentes X foi de 99,50% e no bloco de variáveis dependentes Y foi de 86,69%. O coeficiente de correlação para a calibração foi de 0,649 e de validação foi de 0,412 (ver Figura A.4). Os resultados obtidos para construção de um modelo de quantificação para nitrogênio por PLS apresentaram o mesmo problema observado na seção anterior, erros médios na ordem de grandeza das concentrações de nitrogênio nas amostras, portanto sendo inviáveis para quantificação de nitrogênio devido às baixas concentrações deste elemento nas amostras.

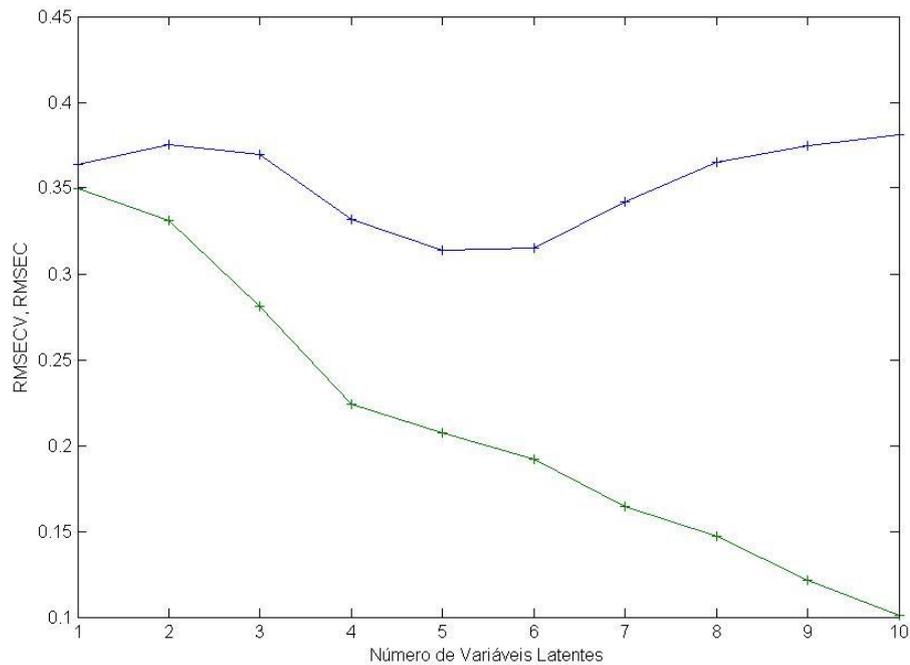


Figura A.3. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes pela calibração em PLS. Determinação de nitrogênio total.

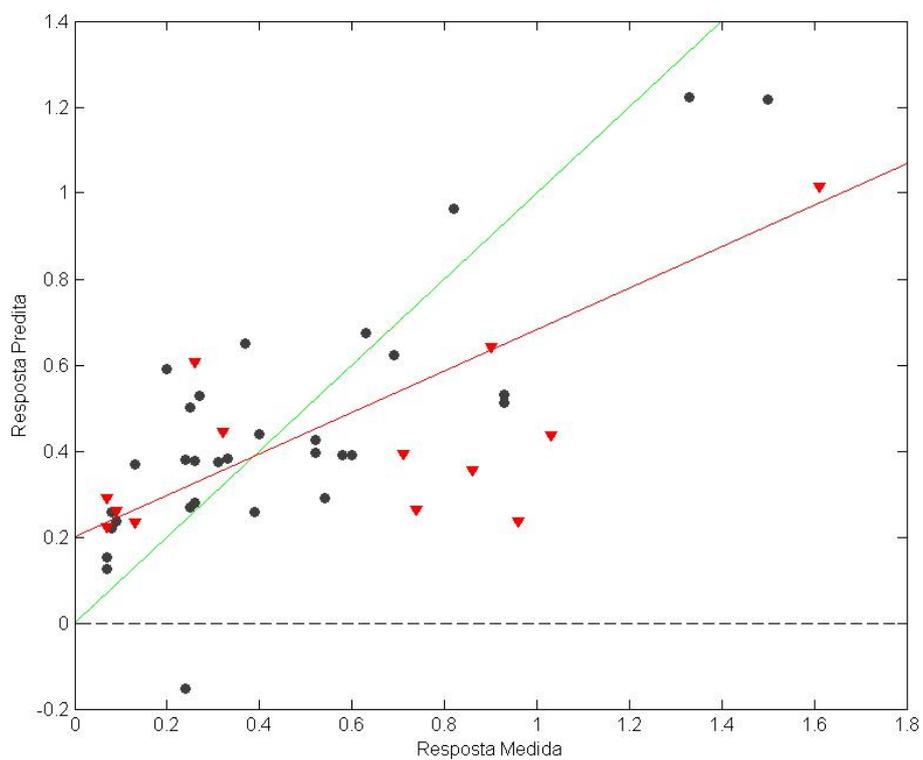


Figura A.4. Valor previsto para nitrogênio total pelo modelo de calibração PLS (%) contra o valor obtido pela análise elementar.

(●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

A.3 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 3 (GA-MLR)

Nesta proposta a determinação de nitrogênio no biocarvão e nas biomassas in natura foi realizada a busca de variáveis de importância utilizando o Algoritmo Genético primeiramente com a Validação Cruzada por Regressão Linear Múltipla. O número de variáveis selecionadas em um subconjunto a partir do conjunto de dados total foi de 197 variáveis. As variáveis que foram identificadas por conter informação relevante para a construção do modelo, quando comparadas com as variáveis escolhidas pelas técnicas de seleção SPA e iPLS, e correlacionadas com a presença de carbono nas amostras, foram as seguintes: 512,97; 1166,7; 1407,7; 1722,1 e 1724 cm^{-1} . Algumas dessas variáveis foram identificadas em relação à presença de carbono nas amostras, como pode ser visto nas discussões das propostas 1 e 5 nas seções A.1 e A.5.

As 197 variáveis encontradas foram utilizadas para a obtenção do modelo pela regressão por RLM. Foram obtidos os seguintes valores de RMSEC, RMSEP e RMSECV, respectivamente: 2,3 g/kg, 3,2 g/kg, e 3,3 g/kg. O coeficiente de determinação para a calibração foi de 0,5655 e para a validação foi de 0,5438 (ver Figura A.5). Igualmente como ocorreu com construção de um modelo através das técnicas SPA-MLR e PLS, os valores de erros de calibração e predição, RMSEC e RMSEP, assim como o ajustamento da curva de entre as respostas medidas e preditas, mostraram inviabilidade do modelo obtido por GA-MLR para a quantificação de nitrogênio em amostras de biomassa e biocarvões. Novamente, essa incapacidade na construção do modelo é resultante das baixas concentrações de nitrogênio nas amostras estudadas.

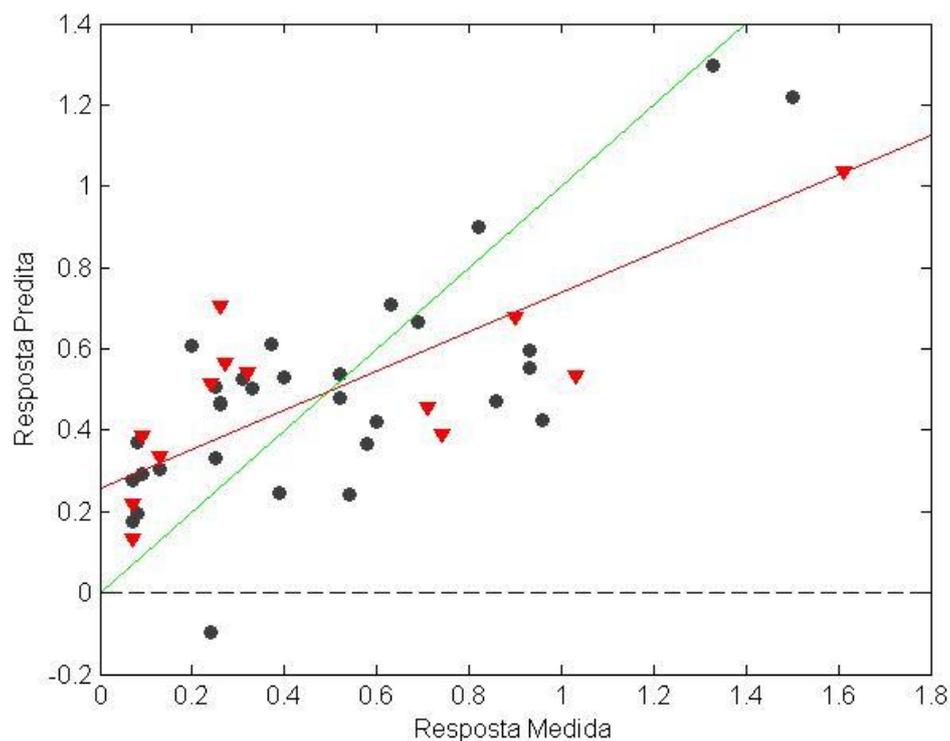


Figura A.5. Valor previsto para nitrogênio total pelo modelo de calibração MLR (%) com seleção de variáveis por GA contra o valor obtido pela análise elemental.
 (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

A.4 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 4 (GA-PLS)

A seleção de variáveis pelo AG com a utilização da Validação Cruzada por PLS obteve o número total de 493 variáveis selecionadas em relação ao conjunto de dados original. As variáveis selecionadas pelo AG em comum com os métodos SPA e iPLS foram: 528,39; 981,58; 1018,2; 1022; 1027,8; 1403,9; 1405,8; 1569,7 e 3637 cm^{-1} . Portanto, essas variáveis, algumas delas com identificação química relacionada ao nitrogênio nas amostras. A identificação em relação à presença de nitrogênio nas amostras pode ser vista nas discussões das seções A.1 e A.5.

Na regressão por PLS foi escolhido o modelo com 5 variáveis latentes (Figura A.6) que apresentou a variância acumulada dos blocos de variáveis independentes X foi de 99,55% e no bloco de variáveis dependentes Y foi de 84,69%. O modelo forneceu um RMSEC de 2,2 g/kg, um RMSEP de 3,3 g/kg, e um RMSECV de 3,0 g/kg. Os coeficientes

de determinação para a calibração e predição com valores de 0,6097 e 0,6225 (ver Figura A.7), respectivamente. O modelo obtido por GA-PLS não apresenta a capacidade de quantificação precisa de nitrogênio em amostras de biomassas e biocarvões, apresentando os mesmos problemas apontados em decorrência das amostras como dito nas seções anteriores.

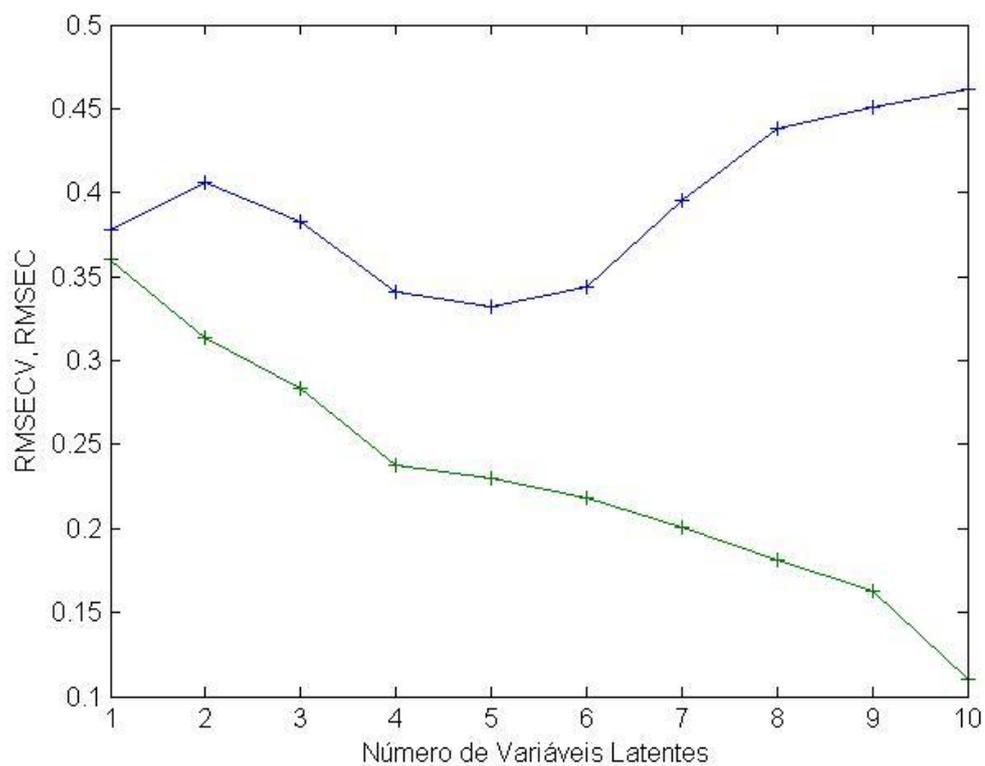


Figura A.6. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por GA. Determinação de nitrogênio total.

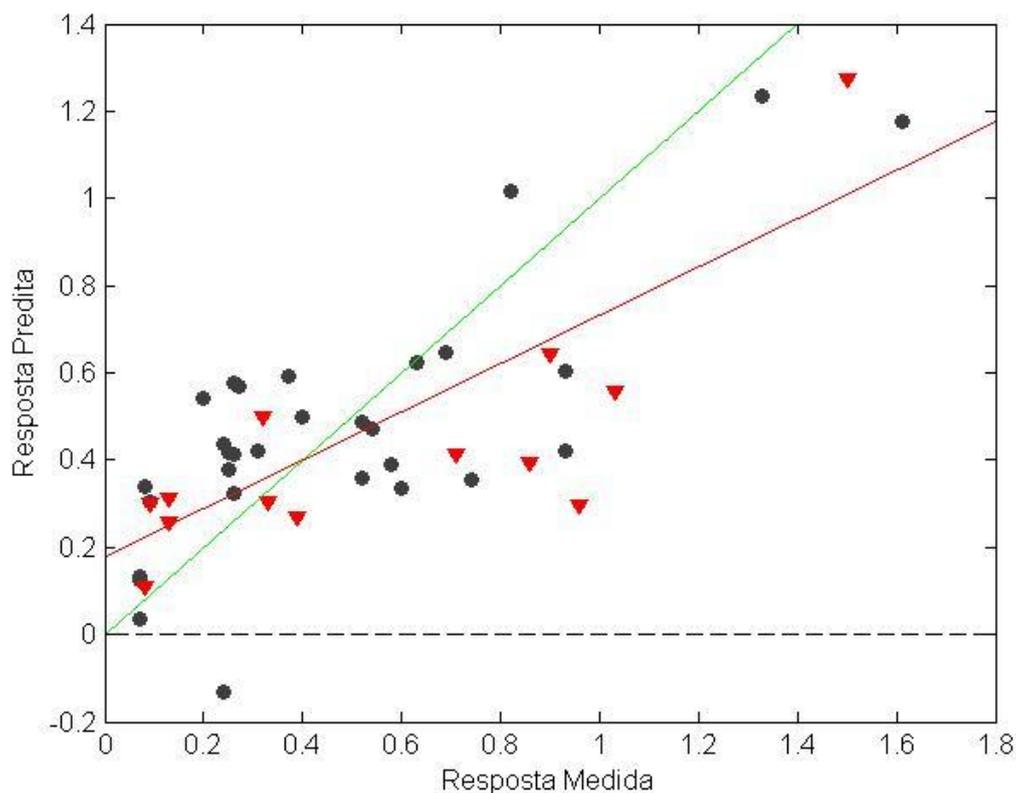


Figura A.7. Valor previsto para nitrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por GA contra o valor obtido pela análise elementar.
 (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

A.5 Determinação de nitrogênio total em amostras de biomassa e biocarvão pela proposta 5 (iPLS)

Na seleção de variáveis por iPLS foram escolhidas 30 variáveis em 8 intervalos diferentes de todo o espectro original. As variáveis selecionadas foram as seguintes: 495,61; 499,47~505,25; 512,97; 526,47~532,25; 1016,2~1027,8; 1160,9~1168,6; 1403,9~1407,7; 1722,1~1727,9; e 3637 cm^{-1} . Na Figura A.8 é possível visualizar o espectro médio com os 8 intervalos que contém as 30 variáveis selecionadas.

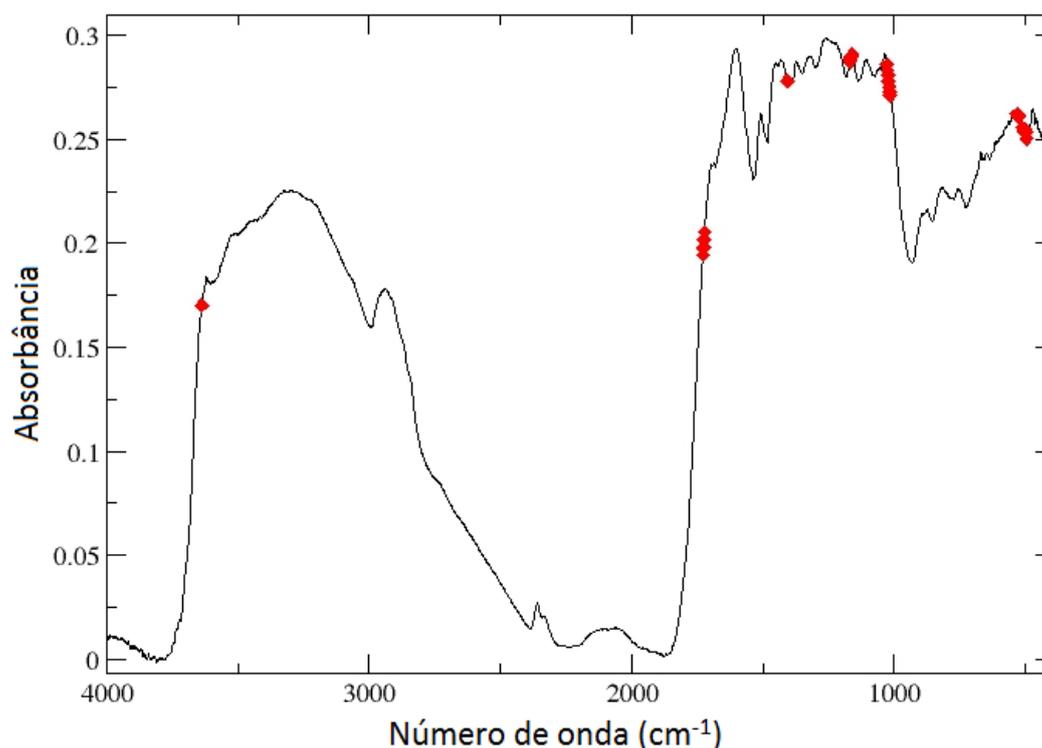


Figura A.8. Espectro MID-IR médio das 44 amostras com destaque para as 30 variáveis selecionadas pelo IPLS correspondente a 9 intervalos de frequências para a determinação da concentração de nitrogênio.

Entre 1160,9~1168,6 cm^{-1} correspondem à presença de ligações C-N de aminas alifáticas (Solomons; Fryhle, 2008). Os números de onda de 495,61; 499,47~505,25; 512,97; 526,47~532,25; 1016,2~1027,8; 1403,9~1407,7; 1722,1~1727,9 e 3637 cm^{-1} não puderam ser identificados sob o aspecto químico relacionando com o teor de nitrogênio total. Entretanto, algumas dessa faixas pode ser relacionadas com os outros teores elementares estudados, como por exemplo, entre 1403,9~1407,7 cm^{-1} é resultante do estiramento de ligações C-C em anéis aromáticos. Outra faixa, entre 1722,1~1727,9 cm^{-1} é identificada pelo estiramento de ligações C=O em aldeídos e cetonas.

Com a regressão por PLS foi escolhido o modelo com a utilização de 8 variáveis latentes (Figura A.9) e a variância acumulada dos blocos de variáveis independentes X foi de 99,98% e no bloco de variáveis dependentes Y foi de 88,08%. Como observado na Figura 8.10, a dispersão do modelo pode ser correlacionada com os métodos

anteriormente usados. O RMSEC apresentou o valor de 1,9 g/kg, o RMSEP o valor de 5,4 g/kg e o RMSECV o valor de 3,1 g/kg.

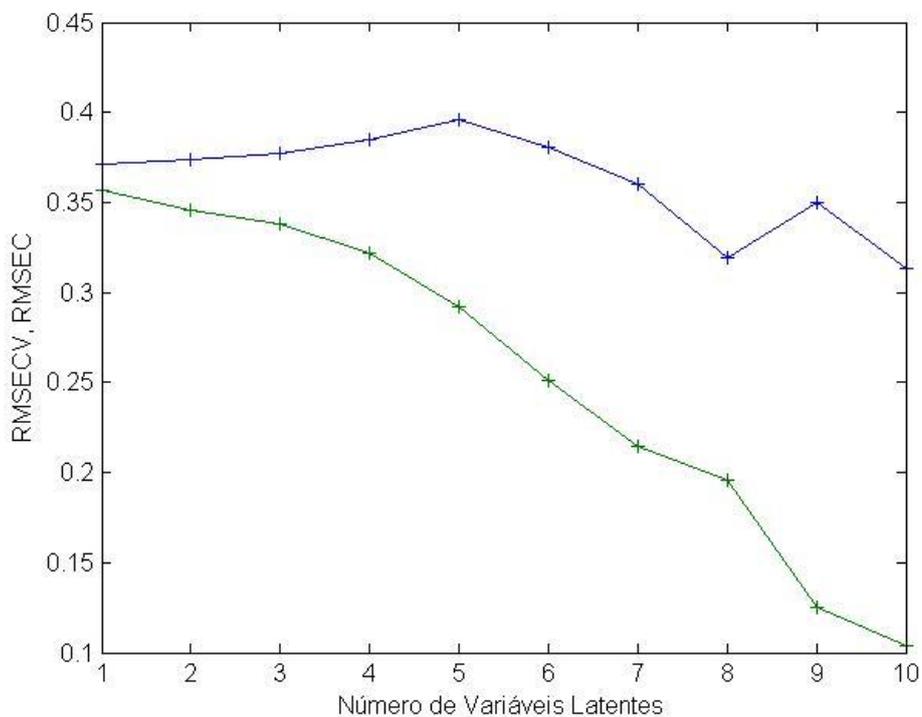


Figura A.9. Gráfico dos valores de RMSECV (linha azul), RMSEC (linha verde) em função do número de Variáveis Latentes para a calibração em PLS pela seleção de variáveis por iPLS. Determinação de nitrogênio total.

Os coeficientes de determinação para a calibração e validação foram, respectivamente, 0,6865 e 0,1124 (ver Figura A.10). Os resultados obtidos pela utilização do iPLS também resultaram em um modelo multivariado que não quantifica adequadamente teores de nitrogênio nas amostras.

Na Tabela A.1 estão contidos os resultados correspondentes a cada proposta apresentada para a determinação de nitrogênio. Os modelos obtidos apresentaram baixíssima precisão para a quantificação de nitrogênio, com erro na ordem de grandeza das porcentagens de nitrogênio observadas na Tabela 6.1, sendo assim inviáveis para a determinação de nitrogênio em amostras de biomassas e biocarvões.

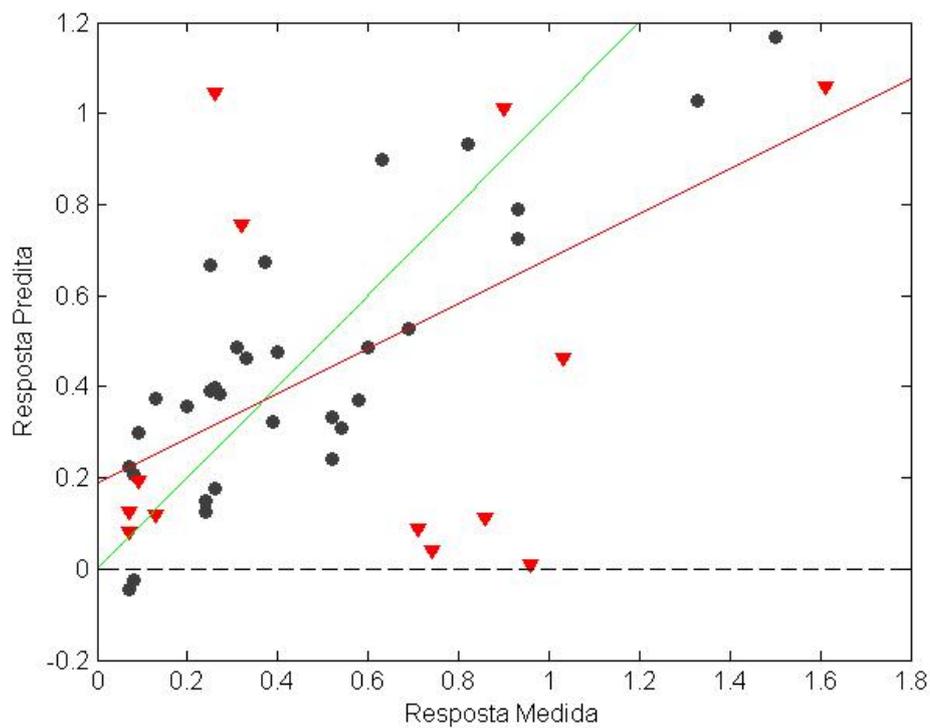


Figura A.10. Valor previsto para nitrogênio total pelo modelo de calibração PLS (%) com seleção de variáveis por iPLS contra o valor obtido pela análise elemental. (●) Calibração e (▼) Validação. Linha verde calibração, linha vermelha validação.

Tabela A.1. Características dos modelos multivariados para as 5 propostas estudadas para a determinação de nitrogênio total.

Proposta	Nº de variáveis selecionadas	RMSEC (g/kg)	R ² Calibração	RMSEP (g/kg)	R ² Validação	RPD
SPA-MLR	5	1,5	0,901	3,8	0,708	1,850
PLS	-	2,0	0,653	4,0	0,412	1,304
GA-MLR	197	2,3	0,565	3,2	0,543	1,479
GA-PLS	493	2,2	0,609	3,0	0,622	1,626
iPLS	30	1,9	0,686	5,4	0,112	1,061