

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO ACADÊMICO DE QUÍMICA  
CURSO DE LICENCIATURA EM QUÍMICA

KAREN PRISCILA DE LIMA

**CRIAÇÃO E ATUALIZAÇÃO DE MODELOS DE CALIBRAÇÃO  
MULTIVARIADA PARA DETERMINAÇÃO DE CARBONO  
ORGÂNICO NO SOLO POR ESPECTROSCOPIA NIR**

TRABALHO DE CONCLUSÃO DE CURSO

CAMPO MOURÃO

2015

KAREN PRISCILA DE LIMA

**CRIAÇÃO E ATUALIZAÇÃO DE MODELOS DE CALIBRAÇÃO  
MULTIVARIADA PARA DETERMINAÇÃO DE CARBONO ORGÂNICO NO  
SOLO POR ESPECTROSCOPIA NIR**

Trabalho de Conclusão de Curso de graduação, apresentado à disciplina de Trabalho De Conclusão de Curso 2, do Curso Superior de Licenciatura em Química do Departamento Acadêmico de Química – DAQUI – da Universidade Tecnológica Federal do Paraná – UTFPR, Campus Campo Mourão, como requisito parcial para obtenção do título de Licenciado em Química.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Valderrama

CAMPO MOURÃO, 2015

## **AGRADECIMENTOS**

Agradeço a Deus.

Agradeço ao meu pai Antonio Carlos de Lima, a minha mãe Eleni Aparecida Ribeiro de Lima, toda minha família e amigos pelo apoio, carinho e compreensão que dispensaram a mim ao longo de minha vida.

Agradeço especialmente a professora Patrícia Valderrama que me ajudou em todos os momentos. Agradeço-a por sempre estar presente, pela dedicação e orientação, por acreditar em mim e principalmente pelo aprendizado que me propiciou.

Agradeço ao André Marcelo de Souza e a Embrapa Solos do Rio de Janeiro pelo fornecimento dos dados para a realização deste trabalho.

Agradeço a todos os professores do curso de Licenciatura em Química da UTFPR que contribuíram para que eu chegasse até aqui.

Agradeço a minha amiga Dayane Aparecida dos Santos pela parceria e ajuda nos trabalhos.

Agradeço aos meus colegas de turma, que com certeza tornaram esses quatro de curso mais agradáveis.

Agradeço também a todos que de alguma forma contribuíram com esse trabalho.

Obrigada!

## Resumo

Lima, Karen Priscila de. Criação e atualização de modelos de calibração multivariada para determinação de carbono orgânico no solo por espectroscopia NIR. 2015. 39 f. Trabalho de Conclusão de Curso (Licenciatura em Química) – Departamento Acadêmico de Química – DAQUI – Universidade Tecnológica Federal do Paraná. Campo Mourão, 2015.

O carbono orgânico é um indicativo da qualidade do solo e atualmente sua quantificação é feita pelo método Walckley & Black modificado, que além de demorado consome muito reagente e gera resíduos tóxicos. Desse modo, pode-se adotar um modelo de calibração multivariada como alternativa aos problemas do método de referência e otimizar as análises. O modelo foi construído utilizando método linear de mínimos quadrados parciais (PLS), 1000 amostras de diferentes solos do Brasil e 17 variáveis latentes. Os *outliers* foram identificados e retirados. O modelo foi então usado para fazer a previsão de um conjunto externo de amostras que, no entanto, não foram bem previstas e isso indica que o modelo atual necessita ser atualizado para garantir sua durabilidade. Foram usadas estratégias como atualizar o modelo PLS com amostras representativas do conjunto externo, e aplicar a padronização através dos algoritmos DS e PDS, para averiguar qual seria a melhor opção. O desempenho dos modelos atualizados foi analisado a partir da comparação dos valores do erro de previsão (RMSEP) e do coeficiente de correlação para o ajuste das amostras externas ( $R_v$ ), e observou-se que a inclusão de uma porcentagem de amostras representativas do conjunto externo foi a melhor estratégia para atualização, comparada as demais.

**Palavras-chave:** carbono orgânico, calibração multivariada, atualização de modelos, quimiometria.

## Abstract

Lima, Karen Priscila. Creating and updating multivariate calibration models for determining soil organic carbon by NIR spectroscopy. 2015. 39 f. Work Completion of course (Degree in Chemistry) - Academic Department of Chemistry - DAQUI - Federal Technological University of Paraná. Campo Mourão, 2015.

The organic carbon is an indicator of soil quality and currently its quantification is made by the modified Walckley & Black method, which in addition to spend a lot of time, use reagent and then generates toxic waste. Thus, it is possible to use an alternative method based on multivariate calibration model. The model was constructed using a linear method of partial least squares (PLS), 1000 samples of different Brazilian soils and 17 latent variables. Outliers were identified and removed. This model was used to quantify the organic carbon in an external samples set and the results were not satisfactory indicating that the current model needs to be updated to ensure its durability. For this, strategies were used to update the PLS model with representative samples of the external set, the DS and PDS standardization algorithms were apply to verify its ability in the PLS model update. The updated models performance were analyzed by comparing the prediction errors (RMSEP) and correlation coefficient to adjust of external samples ( $R_v$ ), it was observed that the inclusion of a percentage of representative samples of the external set is the best strategy to update the model.

**Keywords:** organic carbon, multivariate calibration, model update, chemometrics.

## LISTA DE FIGURAS

- Figura 1 - a) Equipamento NIR (Spectrum 100N, PerkinElmer) empregado na aquisição dos espectros das amostras de solo; b) saída do feixe de radiação infravermelha do acessório de reflectância difusa (NIRA); c) placas de vidro de 90mm de diâmetro empregadas como amostradores; e) amostras no acessório de reflectância difusa.....18
- Figura 2 - Etapas principais do método da oxidação com dicromato de potássio empregado como referência nas análises de carbono orgânico no solo: a) pesagem das amostras de solo; b) digestão das amostras com solução de dicromato de potássio; c) digestão a quente; d) adição do indicador difenilamina; e) titulação do excesso de dicromato com solução de sulfato ferroso amoniacal.....20
- Figura 3 - A) Espectros NIR para as amostras de solo. B) correção da linha de base. C) Região empregada no desenvolvimento do modelo de calibração.....25
- Figura 4 - Ajuste: (A) Modelo PLS otimizado; (B) Previsão direta; (C) Atualização por PDS; (D) Atualização por DS; (E) Atualização por adição de 10%; (F) Atualização após remoção de *outliers*; (G) Atualização por adição de 30%. (●) amostras de calibração. (✖) amostras de validação.....27
- Figura 5 - *Outliers* das amostras do conjunto teste.....29
- Figura 6 - (A) Resíduos para o modelo A; (B) Resíduos para o modelo G.....32

## LISTA DE TABELAS

Tabela 1 - RMSEP e $R_v$ como comparativo entre os modelos atualizados.....	30
Tabela 2 - Comparativa entre as figuras de mérito do modelo A e do modelo G.....	31

## Sumário

<b>1 INTRODUÇÃO .....</b>	<b>9</b>
<b>2 REVISÃO TEÓRICA .....</b>	<b>12</b>
2.1 CALIBRAÇÃO MULTIVARIADA .....	12
2.2 ATUALIZAÇÃO DE MODELOS .....	13
<b>3 OBJETIVO .....</b>	<b>17</b>
3.1 OBJETIVO GERAL .....	17
3.2 OBJETIVOS ESPECÍFICOS .....	17
<b>4 MATERIAL E MÉTODOS.....</b>	<b>18</b>
4.1 DETERMINAÇÃO DE CARBONO ORGÂNICO DO SOLO PELO MÉTODO EMBRAPA MODIFICADO .....	19
4.2 PLS – MÍNIMOS QUADRADOS PARCIAIS .....	22
4.3 TRANSFERÊNCIA DE CALIBRAÇÃO .....	23
4.3.1 DS – PADRONIZAÇÃO DIRETA .....	23
4.3.2 PDS – PADRONIZAÇÃO DIRETA POR PARTES .....	24
<b>5 RESULTADOS E DISCUSSÕES .....</b>	<b>25</b>
<b>6 CONCLUSÕES .....</b>	<b>33</b>
<b>REFERÊNCIAS.....</b>	<b>35</b>



## 1 INTRODUÇÃO

Conhecer os teores de carbono orgânico do solo (COS) é de fundamental relevância, pois a partir desses valores pode-se estimar a qualidade, em termos de fertilidade, do mesmo. Atualmente a determinação da quantia de COS presente na matéria orgânica do solo (MOS) é feita mais comumente pelo método volumétrico do dicromato de potássio ( $K_2Cr_2O_7$  0,2 mol.L<sup>-1</sup>) descrito no Manual de Método de análise de Solos da Embrapa Solos (Embrapa, 2011). No entanto, o uso de tal metodologia consome quantidade elevada de reagentes e gera resíduos tóxicos e cumulativos, que podem por em risco o meio ambiente e a saúde do analista. Diante disso, nota-se a importância do desenvolvimento de métodos alternativos, como os que envolvem a atualização ou criação de modelos de calibração multivariada, que permitam quantificar os teores de COS por meios de espectroscopia no infravermelho próximo (NIR – *Near Infrared*), por exemplo. Essa outra opção de metodologia além de ser bastante eficiente, proporcionar resultados rápidos e eficazes, também evita a geração de resíduos – já que as amostras não precisam de pré-tratamento; o que vai de encontro com os pressupostos da Química Verde (PRADO, 2003).

A matéria orgânica, um dos componentes mais importantes do solo, é uma mistura de todos os resíduos animais e vegetais em estados de decomposição distintos, que apresenta um papel de destaque “como a geração de cargas elétricas negativas, disponibilização de nutrientes e a agregação do solo”; (SATO, 2013, p. 3), além de ser responsável por concentrar o carbono orgânico. O COS pode ser encontrado no solo na forma de aminoácidos, polissacarídeos, polifenóis e ácidos graxos; e ele constitui a maior parte da MOS. A variação da quantia de carbono orgânico é um indicativo da qualidade do solo. Essa qualidade faz referência à fertilidade, que é um fator bastante importante na agricultura, pois se pressupõe que quanto maior a fertilidade do solo (grande quantidade de COS) maior será a produtividade e maior o valor agrônomo (LUZ, FERREIRA e BEZERRA, 2002). A partir dessa relevância que o COS apresenta é necessário quantificá-lo.

Quanto às maneiras de se determinar a quantidade de carbono orgânico presente no solo Nelson e Sommers (1996), indicam que elas podem ser feitas de três modos: analisando o carbono total do solo (CT), analisando o carbono inorgânico (CI) e fazendo a subtração do CI da parte total; determinar o CO; ou avaliar o CO através da oxidação do COS por dicromato ( $\text{Cr}_2\text{O}_7^{2-}$ ) e posterior determinação do dicromato não reduzido pela titulação de oxirredução com  $\text{Fe}^{2+}$ . Porém, a mais utilizada atualmente no Brasil é a última, que faz referência ao método Embrapa.

A partir da titulação percebe-se que a quantidade de CO é igual a de dicromato reduzido durante a reação com o solo. Apesar da simplicidade desse método de referência, existem alguns problemas que podem por em risco não só a saúde do analista, como também do meio ambiente, já que são gerados resíduos tóxicos, como o cromo; o emprego de muito reagente; além de ser uma análise que consome tempo significativo. Frente a essa situação surge a necessidade de se empregar uma metodologia alternativa, que seja tão eficiente quanto o método da oxidação com dicromato, gerando menos resíduos e apresentando um tempo de análise mais adequado (COHEN et al., 2005; FERNANDES et al., 2010), como, por exemplo, a espectroscopia NIR aliada a quimiometria. Esta combinação permite a criação de um modelo de calibração multivariada para determinação de carbono orgânico no solo. Esse método espectroscópico, assim como outros, vem sendo muito utilizado em outros países devido as vantagens que apresenta. (CHATTERJEE et al., 2009)

A quimiometria é uma parte da química analítica iniciada por volta das décadas de 60 e 70 com o desenvolvimento de microcomputadores e o avanço de equipamentos de análises, de modo que o acoplamento desses instrumentos propiciou o surgimento de uma grande quantidade de dados. Pensando nesse volume de informações fornecidas pelos métodos analíticos computadorizados a técnica espectroscópica NIR pode ser conciliada a quimiometria, que “pode ser definida como pesquisa e utilização de métodos matemáticos e estatísticos para o tratamento de dados químicos de forma a extrair uma maior quantidade de informações, podendo levar a melhores resultados analíticos.” (VALDERRAMA, 2005, p. 37)

Com base na quimiometria surgiu a calibração multivariada, que pode

oferecer metodologias bastante eficazes para determinações simultâneas e resolução de superposição de sinais analíticos. A calibração multivariada se tornou popular com o emprego do método dos Mínimos Quadrados Parciais (PLS - *Partial Least Squares*) (BRERETON, 2000), que é uma metodologia de ajuste linear das variáveis.

Aliando a quimiometria a espectroscopia NIR para determinação da quantidade de COS sugere-se a criação de um modelo matemático que forneça resultados de COS a partir de espectros gerados na região do infravermelho próximo. Este modelo seria uma excelente alternativa ao método clássico, pois apresenta rapidez e eficiência no fornecimento de resultados, não proporciona geração de resíduos tóxicos e cumulativos. Todos esses fatores indicam uma capacidade em potencial substituinte da metodologia tradicionalmente empregada em laboratórios de análise de solo do Brasil para a determinação de COS.

Contudo, a manutenção desses modelos multivariados está cercada pela problemática de que se for inserido uma amostra completamente diferente do que o modelo possui, o resultado gerado será ineficiente. Tal fato cria a necessidade de se atualizar o modelo multivariado por meio de algumas estratégias disponíveis. Essa atualização (*model update*) se justifica e faz-se necessária para propiciar o uso constante de um modelo já existente e evitar a criação de um novo modelo sempre que surgirem novas amostras. Para que essa atualização ocorra, este trabalho propõe o uso da “metodologia de transferência de calibração, que foi originalmente desenvolvida para transferir e corrigir modelos multivariados desenvolvidos em diferentes espectrofotômetros por meio de cálculos matriciais”. (PEREIRA, 2013, p.1) Para isso, a construção do modelo de calibração multivariada, será baseada no método de Walckley & Black modificado para determinar carbono orgânico de solos brasileiros por espectroscopia NIR.

## 2 REVISÃO TEÓRICA

### 2.1 CALIBRAÇÃO MULTIVARIADA

A calibração multivariada é uma metodologia bastante difundida e empregada atualmente em quimiometria, por possibilitar a calibração mesmo na presença de interferentes - a partir de operações, como funções, que relacionam muitas variáveis. Essas variáveis se dividem em dependentes (**Y**) e independentes (**X**): (ZENI, 2005. VALDERRAMA, 2009. PEREIRA, 2013)

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}) = \mathbf{Xb} \quad (1)$$

Nesse modo de calibração é usado um número bastante elevado de variáveis independentes para se determinar outra variável (**Y**), que dá origem a um conjunto de dados chamado de conjunto de calibração. A calibração multivariada apresenta vantagens quando métodos univariados não podem ser empregados. Isso acontece principalmente porque nos métodos univariados não é tolerado a presença de interferentes enquanto que na calibração multivariada de primeira ordem os interferentes não prejudicam os resultados, desde que estes estejam presentes na etapa de calibração, conceito conhecido como vantagem de primeira ordem (BOOKSH e KOWALSKI, 1994).

Para a construção de tais modelos de calibração observa-se que há diferentes métodos, entretanto, o mais utilizado para elaboração de um modelo com dados de primeira ordem é o método de Regressão por Mínimos Quadrados Parciais (PLS). Este emprega um procedimento que ajusta linearmente as variáveis obtidas (GELADI e KOWALSKI, 1986). O PLS é o método mais comumente utilizado na calibração multivariada por apresentar melhores resultados de previsão quando comparado aos demais métodos. Isso porque o PLS leva em consideração os valores de **y** no cálculo do vetor dos coeficientes de regressão do modelo multivariado (WOLD, SJÖSTRÖM e ERIKSSON, 2001). O método PLS será descrito com maiores detalhes na

seção Métodos.

Apesar das vantagens que os modelos de calibração multivariada apresentam como a dispensa de uso de reagentes e preparo de amostras, a rapidez e eficiência; observa-se que a manutenção do bom funcionamento de tais modelos pode enfrentar uma problemática, caso não seja feita a constante atualização do modelo. Dentre os problemas destaca-se a previsão equivocada de resultados decorrentes da não introdução de variações, na fase de calibração, dos novos espectros, levando a invalidação do modelo. (HONORATO et al, 2007)

Segundo Honorato et al, 2007, existem três causas para a desatualização dos modelos: (1) mudanças na constituição química e/ou física das amostras (viscosidade, tamanho de partículas, textura da superfície etc.); (2) mudanças no ambiente do equipamento (variações de temperatura e umidade); e (3) mudanças na função de resposta do equipamento (que farão com que o espectro de uma dada amostra apresente o mesmo perfil, mas intensidades diferentes). Nesse último caso, as mudanças podem ser causadas por substituição de partes do equipamento, envelhecimento da fonte, instabilidade do detector e/ou, pela aquisição dos espectros em outro equipamento.

Todavia, a nulidade do modelo pode ser evitada fazendo-se a atualização (ou transferência de calibração) a partir de alguns métodos descritos na literatura, permitindo que o modelo se mantenha robusto, e capaz de prover resultados confiáveis. (HONORATO et al, 2007)

## 2.2 ATUALIZAÇÃO DE MODELOS

A atualização de modelos (DESPSGNE, MASSART e CHABOT, 2000) é uma estratégia, que permite a adição, quando conhecidas, de fontes influentes de mudanças no modelo pelo acréscimo de novas amostras no momento da calibração. Ou quando não se conhece a fonte de interferência,

um *model update* (atualização do modelo) pode ser considerado. Nesse caso, a função de resposta de um conjunto de dados secundário é modificada para se tornar parecida com a função de resposta do conjunto de dados primário, onde o modelo de calibração foi desenvolvido. Isto pode ser feito através de funções matemáticas nas respostas espectrais ou nos valores preditos. (HONORATO, et al, 2007) Como se observa no trabalho de Feudale et al, 2002, existem algumas estratégias de transferência, que fazem uma adaptação do modelo ou dos espectros obtidos, que são denominados métodos de padronização.

Os métodos de padronização são ferramentas de cálculo matricial que modificam o sinal analítico de uma situação secundária (também chamada de filha ou escrava), na qual é feita a análise para que ele seja semelhante ao sinal analítico da situação primária (pai ou mestre) na qual foi construído o modelo. Para realizar essa padronização, um grupo de amostras mais representativas dentro do conjunto de calibração é selecionado como o conjunto de amostras de transferência e elas são analisadas nas duas condições. Assim, geram-se duas matrizes de dados:  $\mathbf{X}_1(n_{\text{trans}} \times q)$ , da condição primária, e  $\mathbf{X}_2(n_{\text{trans}} \times p)$ , da condição secundária. Dessa forma, as matrizes são relacionadas por uma matriz de transformação  $\mathbf{F}$ , como mostrado na equação:

$$\mathbf{X}_1 = \mathbf{X}_2 * \mathbf{F} \quad (2)$$

Estimando-se  $\mathbf{F}$  pela relação das amostras de transferência, pode-se usar essa estimativa ( $\hat{\mathbf{F}}$ ) para ajustar o sinal analítico de uma amostra na condição secundária ( $\mathbf{x}_2$ ), de modo que esse sinal seja semelhante ao que seria observado, caso essa amostra fosse analisada na condição primária. Com isso, obtém-se o sinal analítico ajustado à condição primária. (PEREIRA, 2013)

A literatura traz uma explanação dos diversos métodos de padronização que existem, tais como: a padronização reversa, a univariada, correção de inclinação e intercepto (*Slope and bias correction*), a Padronização Direta (DS – *Direct Standardization*) e a Padronização Direta por Partes (PDS – *Piecewise Direct Standardization*).

A padronização reversa é um método bastante parecido com a DS e a

PDS, já que faz uso das mesmas equações; diferindo apenas pelo fato de que a matriz de transformação ajusta os espectros de todas as amostras do conjunto de dados primários, de maneira a parecer que elas foram feitas no conjunto de dados secundários. Com a padronização desse conjunto é possível calcular um novo modelo, que servirá para prever as características das amostras por meio dos espectros fornecidos pelo conjunto de dados secundários.

A padronização reversa foi avaliada por Lima e Borges, 2002, como muito eficiente, já que fornece resultados mais flexíveis no sentido de possibilitar o uso de métodos quimiométricos para reconhecer e excluir valores anormais (*outliers*) e para escolher o melhor conjunto de calibração e de variáveis.

A padronização univariada foi proposta por Shenk e Westerhaus, 1985, com o objetivo de adequar diferenças de intensidades, porém os autores consideravam que não havia deslocamento nos comprimentos de onda ( $\lambda$ ) entre os equipamentos nos quais eram realizadas as medidas. No entanto, ao mudar de equipamento é necessária uma correção, realizada a partir da correlação dos  $\lambda$  do conjunto de dados primários com o  $\lambda$  do conjunto de dados secundários. De modo geral, os dados obtidos no segundo equipamento são ajustados a partir de uma função quadrática, na qual o máximo corresponde ao  $\lambda$  equivalente ao do conjunto de dados primários. Quanto à intensidade espectral, ela pode ser ajustada utilizando-se equação 3:

$$\mathbf{x}_{iP} = \mathbf{a}_i + \mathbf{b}_i \cdot \mathbf{x}_{jS} \quad (3)$$

Onde  $\mathbf{x}_{iP}$ : é o vetor resposta no comprimento de onda das amostras do conjunto de dados primário;  $\mathbf{x}_{jS}$ : é o vetor resposta das mesmas amostras no comprimento de onda do conjunto de dados secundário equivalente ao do primário;  $\mathbf{a}_i$ : coeficiente linear (“*offset*”);  $\mathbf{b}_i$ : coeficiente angular da reta de regressão determinada a partir das amostras de transferência)

Para melhorar a metodologia de padronização univariada, corrigindo diferenças de comprimentos de onda e de intensidade, autores como Bouveresse et al, 1995, propuseram uma modificação na correção da intensidade considerando a natureza diversa das amostras de transferência; e

Forina et al, 1999, recomendou a correção dos comprimentos de onda por meio do ajuste da função quadrática às distâncias Euclidianas e não aos coeficientes de correlação, uma vez que estes são mais influenciados pelos ruídos.

Na padronização dos valores preditos utiliza-se a equação 4. Assim, as amostras de transferência têm suas características (**y**) aferidas e previstas nos dois equipamentos. Para isso é usado o modelo de calibração, criado no conjunto de dados primários, e os valores obtidos são usados para ajustar a equação de regressão. Os coeficientes angular (**b**) e linear (**a**), da equação são capazes de corrigirem os desvios da linha de base e inclinações decorrentes da transferência.

$$y_p = a + b \cdot y_s \quad (4)$$

Quanto a padronização direta (DS) e a padronização direta por partes (PDS) tem-se que, a DS faz uso do espectro todo para o ajuste do sinal em cada comprimento de onda, enquanto, a PDS emprega o modo de janela móvel com certos comprimentos de onda próximos da variável ajustada. Essas metodologias foram propostas por Wang et al em 1991, e serão discutidas na seção Métodos.

Utilizar regiões espectrais, confere à PDS a vantagem de apresentar um posto local (número de “espécies” presente) menor em relação ao espectro como um todo. Isto possibilita que menos amostras sejam necessárias para a transferência. (BARBOZA, 2003)



### 3 OBJETIVO

#### 3.1 OBJETIVO GERAL

Criar e atualizar modelos de calibração multivariada para determinar carbono orgânico nos solos brasileiros.

#### 3.2 OBJETIVOS ESPECÍFICOS

- Determinar carbono orgânico em solos brasileiros pelo método volumétrico do dicromato de potássio ( $K_2Cr_2O_7$  0,2 mol.L<sup>-1</sup>) descrito no Manual de Método de análise de Solos da Embrapa Solos (Embrapa, 2011).

- Construir um modelo de calibração multivariada empregando a regressão por quadrados mínimos parciais (PLS).

- Avaliar o potencial de dois métodos de transferência de calibração na atualização do modelo desenvolvido: padronização direta (DS) e a padronização direta por partes (PDS).

- Atualizar o modelo de calibração multivariada com 10% e 30% das amostras do conjunto externo.

- Atualizar o modelo de calibração multivariada empregando a estratégia de remoção de *outliers* com base no *leverage* e no resíduo em **X** e **Y** (Resíduos Q).

- Comparar o desempenho dos modelos atualizados.

## 4 MATERIAL E MÉTODOS

Toda a parte experimental do presente projeto, como coleta e pré-tratamento das amostras de solo; a análise das mesmas através do método da oxidação pelo dicromato e a obtenção dos espectros por NIR foi realizada por bolsistas do Projeto MP3 (NIRSOLOS: “A Espectroscopia Vis-NIR-MIR e a Quimiometria na avaliação simultânea de vários atributos do solo: rumo a inovação e adequação dos laboratórios de solos aos preceitos da Química Verde”) sob Liderança da Embrapa Solos, na unidade Embrapa Solos – Rio de Janeiro.

Amostras de solos de perfis completos, representativas do Brasil, foram selecionadas a partir da coleção nacional de amostras de terra (Soloteca) da Embrapa Solos. O critério de seleção das amostras considerou sua representatividade ambiental dos solos brasileiros de tal forma que foram selecionadas todas as classes de solos em nível de subordem taxonômica (segundo nível categórico), de acordo com o Sistema Brasileiro de Classificação de Solos (SiBCS). São 13 as classes de solos do Brasil descritas no SiBCS: Argissolos, Cambissolos, Chernossolos, Espodossolos, Gleissolos, Latossolos, Luvissolos, Neossolos, Nitossolos, Organossolos, Planossolos, Plintossolos e Vertissolos, sendo que as classes Latossolos e Argissolos correspondem a 60% dos solos do Brasil. Essas amostras de solos foram secas em estufa a 40° C durante 48 horas, destorradas com um martelo de borracha e passadas em peneira (10 mesh) obtendo-se a fração conhecida como terra fina seca ao ar (SANTOS et al., 2013). A Figura 1 ilustra o equipamento de espectroscopia na região do infravermelho próximo e as principais etapas do processo de aquisição dos espectros das amostras de solo.



**Figura 1 - a) Equipamento NIR (Spectrum 100N, PerkinElmer) empregado na aquisição dos espectros das amostras de solo; b) saída do feixe de radiação infravermelha do acessório de reflectância difusa (NIRA); c) placas de vidro de 90mm de diâmetro empregadas como amostradores; e) amostras no acessório de reflectância difusa.**

**Fonte: Souza, 2014.**

Quanto à parte computacional do tratamento dos dados, esta foi realizada a partir do uso do *software Matlab* e de metodologias como o PLS (PLS-*toolbox*), o DS e o PDS (rotinas escritas em ambiente *Matlab* no próprio laboratório).

#### 4.1 DETERMINAÇÃO DE CARBONO ORGÂNICO DO SOLO PELO MÉTODO EMBRAPA MODIFICADO

O método Embrapa (Embrapa, 2011) consiste na determinação da quantidade de carbono orgânico no solo por um método de via úmida. Nesse método são coletados cerca de 20g de solo, tritura-se e homogeneiza-se em uma peneira de 80 *mesh*. Em seguida, pesa-se 0,5g de amostra de solo triturado, transferindo-se para um erlenmeyer de 250mL. Então, adiciona-se 10mL de solução de dicromato de potássio 0,4 mol L<sup>-1</sup>. Para comparação coleta-se 10mL de solução de dicromato de potássio 0,4 mol L<sup>-1</sup> em outro

erlenmeyer, que será a prova branca (a prova branca é uma solução que contém todos os constituintes da reação menos o analito, e ela serve para achar o valor que será acrescentado no resultado final da análise devido a presença de interferentes – contaminantes – nos reagentes usados). Após esse procedimento deve-se colocar um tubo de ensaio de 25mm de diâmetro e 250mm de altura cheio de água até a boca do erlenmeyer contendo o solo e a solução de  $K_2Cr_2O_7$ ; formando um sistema semelhante a um condensador. Depois de montado, o sistema é aquecido por uma placa elétrica até a fervura branda do conteúdo do recipiente, por cinco minutos. Após o resfriamento da amostra contida no erlenmeyer, deve-se acrescentar 80ml de água destilada, 2mL de ácido ortofosfórico e 3 gotas do indicador difenilamina. A seguir é realizada uma titulação nos erlenmeyers contendo a prova branca e a amostra de solo, utilizando como titulante uma solução de sulfato ferroso amoniacal  $0,1 \text{ mol L}^{-1}$  até que a coloração mude de azul para verde, anotando-se os volumes gastos na titulação. A Figura 2 ilustra os passos empregados na determinação do carbono orgânico a partir do método Embrapa.



**Figura 2 - Etapas principais do método da oxidação com dicromato de potássio empregado como referência nas análises de carbono orgânico no solo: a) pesagem das amostras de solo; b) digestão das amostras com solução de dicromato de potássio; c) digestão a quente; d) adição do indicador difenilamina; e) titulação do excesso de dicromato com solução de sulfato ferroso amoniacal.**

**Fonte: adaptado de Souza, 2014.**

Para o cálculo da quantidade de carbono orgânico usa-se a equação 5:

$$C \text{ (g/ kg)} = (40 - \text{volume gasto na titulação do solo}) \times f \times 0,6 \quad (5)$$

Em que C representa a concentração de carbono orgânico; 40 e 0,6 são constantes e f é igual a 40 / volume sulfato ferroso gasto na prova em branco.

## 4.2 PLS – MÍNIMOS QUADRADOS PARCIAIS

A Regressão por Mínimos Quadrados Parciais (PLS – *Partial Least Squares*), segundo Geladi e Kowalski, 1986, é um método de ajuste linear bastante utilizado na construção de modelos de calibração multivariada com dados de primeira ordem.

No PLS uma matriz com informações espectrais (matriz **X**) e um vetor contendo as informações retiradas do conjunto de dados de referência (**y**) são decompostas ao mesmo tempo num produto de duas matrizes, chamadas de *scores* e *loadings* (BRERETON, 2000. MORAIS, 2014. FERRARINI, 2004), como demonstrado nas equações 6 e 7:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (6)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (7)$$

Em que **T** e **U** são as matrizes *scores* de **X** e **Y**; **P** e **Q** são as matrizes *loadings* das matrizes **X** e **Y**, e **E** e **F** são os resíduos.

A correlação entre as matrizes **X** e **Y** pode ser obtida pelo coeficiente de regressão linear, tal como descrito a seguir pela equação 8:

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h \quad (8)$$

“para “h” variáveis latentes, sendo que os valores de **b<sub>h</sub>** são agrupados em uma matriz diagonal **B**, que contém os coeficientes de regressão entre a matriz de *scores* **U** de **Y** e a matriz de *scores* **T** de **X**.” (SENA et al, 2000, p. 551)

Conhecendo o valor de **u<sub>h</sub>**, e conseqüentemente a matriz diagonal **B**, pode-se calcular a matriz **Y** utilizando-se a equação 9:

$$\mathbf{Y} = \mathbf{TBQ}' + \mathbf{F} \quad (9)$$

Desta forma, a quantidade de carbono orgânico de novas amostras pode ser prevista a partir dos novos *scores*, **T\***, substituídos na equação anterior, como mostra a equação 10:

$$\mathbf{Y} = \mathbf{T}^* \mathbf{BQ}' \quad (10)$$

### 4.3 TRANSFERÊNCIA DE CALIBRAÇÃO

A transferência de calibração é a adaptação do modelo de calibração construído em um determinado instrumento para analisar novas amostras oriundas de outro instrumento. O potencial dessa estratégia foi avaliado na atualização do modelo de calibração multivariada para previsão de espectros de amostras que não fizeram parte da etapa de calibração mas que, entretanto, foram obtidos no mesmo equipamento. A transferência pode ser realizada por diferentes modos, tais como os métodos de padronização DS e PDS. Estes são métodos de cálculo matricial, que alteram o sinal analítico de um conjunto de dados secundários para que ele se torne parecido ao sinal analítico do conjunto de dados primários, no qual o modelo foi criado. (PEREIRA, 2013)

#### 4.3.1 DS – PADRONIZAÇÃO DIRETA

Nessa metodologia de atualização, para que o conjunto de dados secundários se torne semelhante ao conjunto primário é necessário relacionar as matrizes respostas através da equação 11:

$$\mathbf{X}_P = \mathbf{X}_S \mathbf{F} \quad (11)$$

Em que  $\mathbf{F}$  é uma matriz de dimensão  $(p \times q)$ , onde  $q$  representa o número de comprimentos de onda obtidos no equipamento primário, e  $p$  é o número de comprimentos de onda para os espectros registrados no equipamento secundário. A matriz  $\mathbf{F}$  é estimada com o intuito de se produzir, a partir de um espectro  $\mathbf{x}_S$  registrado para uma nova amostra no equipamento secundário, um espectro ajustado,  $\mathbf{x}_a$ , que seja parecido ao que seria obtido se a mesma amostra tivesse sido analisada no equipamento primário (HONORATO, 2007), como mostra a equação 12:

$$\mathbf{X}_a = \mathbf{x}_S \hat{\mathbf{F}} \quad (12)$$

Em que  $\hat{\mathbf{F}}$  é uma estimativa de  $\mathbf{F}$  obtido na equação 11.

Considerando a equação anterior observa-se que se  $\mathbf{X}_s$  for inversível a matriz  $\mathbf{F}$  pode ser obtida por meio da multiplicação da equação 11 por  $(\mathbf{X}_s)^{-1}$  como mostra a equação 13:

$$\mathbf{F} = (\mathbf{X}_s)^{-1} \cdot \mathbf{X}_p \quad (13)$$

Entretanto, situações como essa ocorrem muito poucas vezes, já que o número de amostras é menor que o de variáveis, gerando uma matriz  $\mathbf{X}_s$  retangular. Para resolver esse impasse,  $\mathbf{F}$  pode ser calculada usando métodos multivariados, como o PLS. Nessa metodologia as matrizes são projetadas sobre a coluna reduzida da matriz  $\mathbf{X}_p$ , que é obtida por decomposição. Então os *scores* gerados nas projeções é que são utilizados tanto para a padronização, quanto para a calibração.

Após a obtenção dos *scores* a padronização é feita a partir de  $\mathbf{F}$ , calculada através da equação 13. A multiplicação pela matriz  $\mathbf{F}$  possibilita que o vetor  $\mathbf{x}_s$ , que contém a amostra desconhecida registrada no conjunto de dados secundários seja padronizado para o vetor esperado para o conjunto de dados primário,  $\mathbf{x}_a$ .

#### 4.3.2 PDS – PADRONIZAÇÃO DIRETA POR PARTES

Na Padronização Direta por Partes não é usado todo o espectro do conjunto secundário como na DS, e sim os comprimentos de onda de uma parte do espectro. Este é relacionado ao comprimento de onda obtido no conjunto de dados primários através de uma janela espectral móvel para o comprimento de onda em estudo.

Quanto as matrizes espectrais de transferência, elas também se relacionam pela equação 10, contudo, é preciso calcular, para cada janela, um vetor de regressão pelo método PLS. Os valores dos vetores serão dispostos em uma matriz  $\mathbf{F}$ , cuja estrutura se apresenta de acordo com a equação 14:

$$\mathbf{F} = [f_1 \ f_2 \ \dots \ f_j \ \dots \ f_q] \quad (14)$$



Cada coluna desta matriz representa um vetor, como na equação 15:

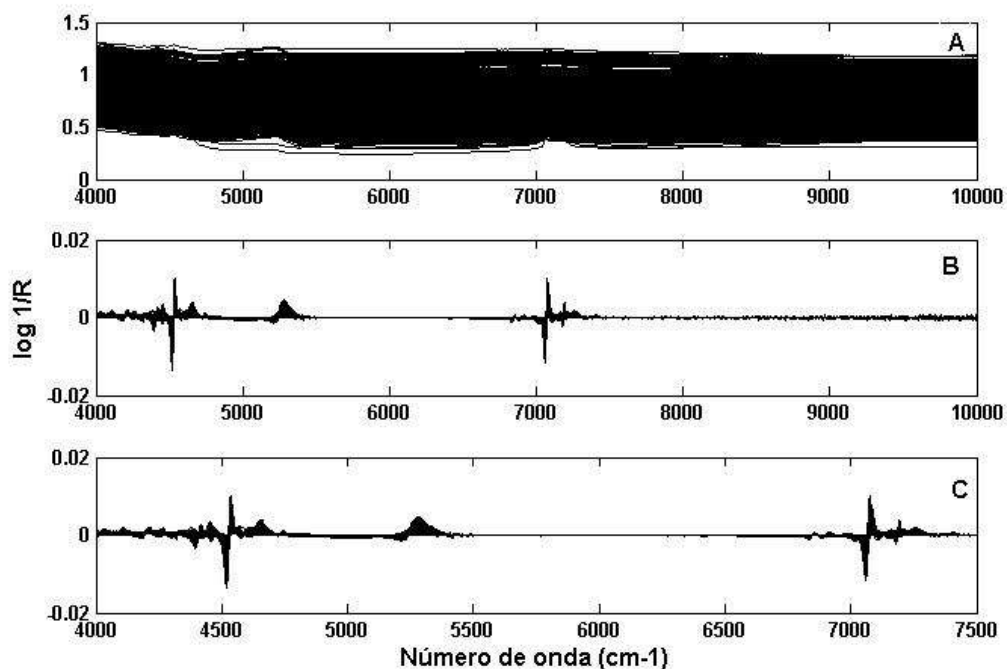
$$\mathbf{F}_j = [0 \dots f_{j-k,j} \dots f_{jj} \dots f_{j+l,j} \dots 0]^T \quad (15)$$

Nessa representação, a janela usada para a padronização é definida pela seção de elementos que vai de  $f_{j-k,j}$  a  $f_{j+l,j}$ .

Segundo Honorato et al, 2007 para tornar os espectros de uma nova amostra do conjunto de dados secundários parecido com um medido no equipamento primário, pode-se fazer a padronização utilizando a equação 12. No entanto, deve-se considerar que o índice “j” indica a que comprimento de onda do vetor  $x_a$  o vetor  $f_j$  se refere, sendo a absorbância nesse comprimento de onda padronizada empregando-se uma janela de largura  $k+l + 1$ .” (HONORATO et al, 2007, p. 1302)

## 5 RESULTADOS E DISCUSSÕES

A Figura 3 retrata os espectros na região do infravermelho próximo (NIR) para as amostras de solo (1.000 amostras) e os resultados da correção de linha de base a partir aplicação da primeira derivada. Para fazer esse pré-processamento foi utilizada uma janela de 15 pontos e o algoritmo de Savitsky-Golay, com um polinômio de segunda ordem com capacidade de remover efeitos de deslocamentos verticais e de linha de base (ou *off-set*). (SAVITSKY e GOLAY, 1964). Após a correção de linha de base observou-se que na região posterior a  $7500 \text{ cm}^{-1}$  não havia mais informações relevantes sobre as amostras, existia apenas ruídos; logo estes foram removidos.



**Figura 3 – A) Espectros NIR para as amostras de solo. B) correção da linha de base. C) Região empregada no desenvolvimento do modelo de calibração.**

**Fonte: autoria própria.**

A separação do conjunto de calibração e de validação foi feita a partir do algoritmo de Kennard-Stone (KENNARD e STONE, 1969) e resultou num total de 750 e 250 amostras para calibrar e validar, respectivamente. O modelo construído foi centrado na média e empregou um total de 17 variáveis latentes (VLs), escolhidos através dos resultados da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) para as amostras de calibração, em blocos contínuos de 10 amostras.

O primeiro modelo PLS construído apresentou valores de erro médio quadrático da calibração (RMSEC) 8,9410 e previsão (RMSEP) 7,4342 g/Kg. Esses valores estimados pelo modelo multivariado não apresentaram uma boa concordância com os valores do método de referência (EMBRAPA, 1997). Diante disso, para otimizar o modelo foi necessário identificar e eliminar as amostras anômalas (*outliers*) – que eram aquelas com comportamento muito diferente das demais amostras do conjunto de dados.

A otimização dos conjuntos de calibração e validação, pela eliminação

dos *outliers*, foi realizada de acordo com a norma da ASTM E1655-05 (ASTM,2005). Nesta otimização são avaliadas as amostras anômalas do conjunto de calibração com relação ao *leverage*, ao resíduo espectral e o resíduo na variável dependente. Para as amostras de validação o teste é aplicado com relação ao *leverage* e aos resíduos espectrais.

Depois da eliminação das amostras anômalas, o novo modelo (denominado A) de calibração multivariada foi construído com 670 amostras de calibração e 208 amostras de validação, e apresentou valores de RMSEC = 2,7134 g/Kg e RMSEP = 2,3020 g/Kg de carbono orgânico, respectivamente. Tais valores são considerados como um dos indicadores de exatidão do modelo multivariado e revelaram que os valores estimados concordam com os valores do método de referência.

A partir desse modelo A, testou-se a sua capacidade de previsão direta do conjunto externo de amostras e avaliou-se o potencial dos métodos de transferência de calibração através da padronização direta e a padronização direta por partes. A inclusão de um percentual representativo de amostras do conjunto externo ao modelo também foi avaliada, bem como uma estratégia de eliminação de *outliers* com base *no leverage* e nos resíduos do conjunto externo.

O ajuste do modelo A, Figura 4(A), foi avaliado com base no gráfico que representa o teor de carbono orgânico estimado pelo modelo PLS contra os valores obtidos através de metodologia de referência (EMBRAPA, 1997). O coeficiente de correlação para o ajuste, no modelo foi da ordem de 0,85.

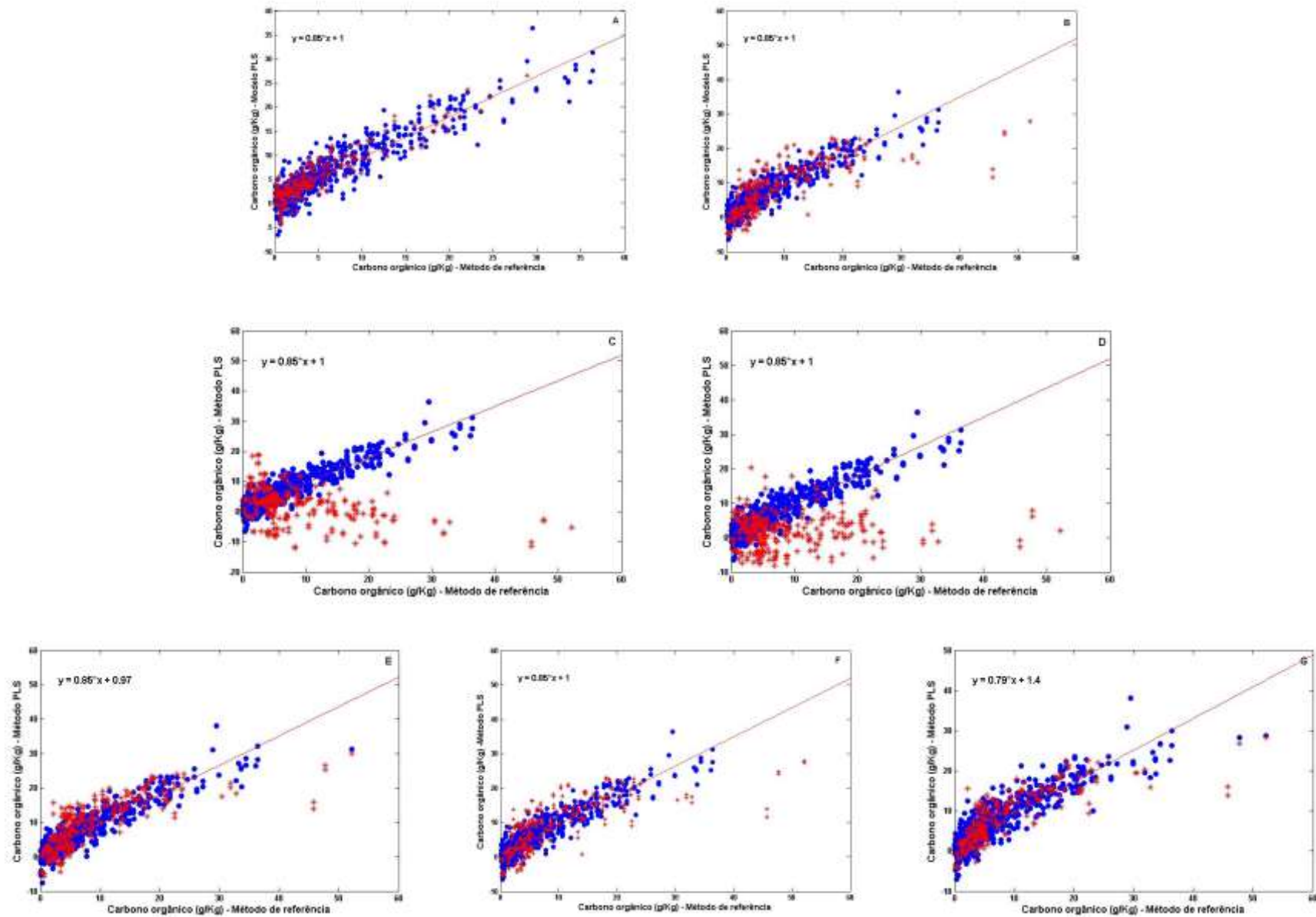


Figura 4 - Ajuste: (A) Modelo PLS otimizado; (B) Previsão direta; (C) Atualização por PDS; (D) Atualização por DS; (E) Atualização por adição de 10%; (F) Atualização após remoção de *outliers*; (G) Atualização por adição de 30%. (●) amostras de calibração. (×) amostras de validação. Fonte: autoria própria.

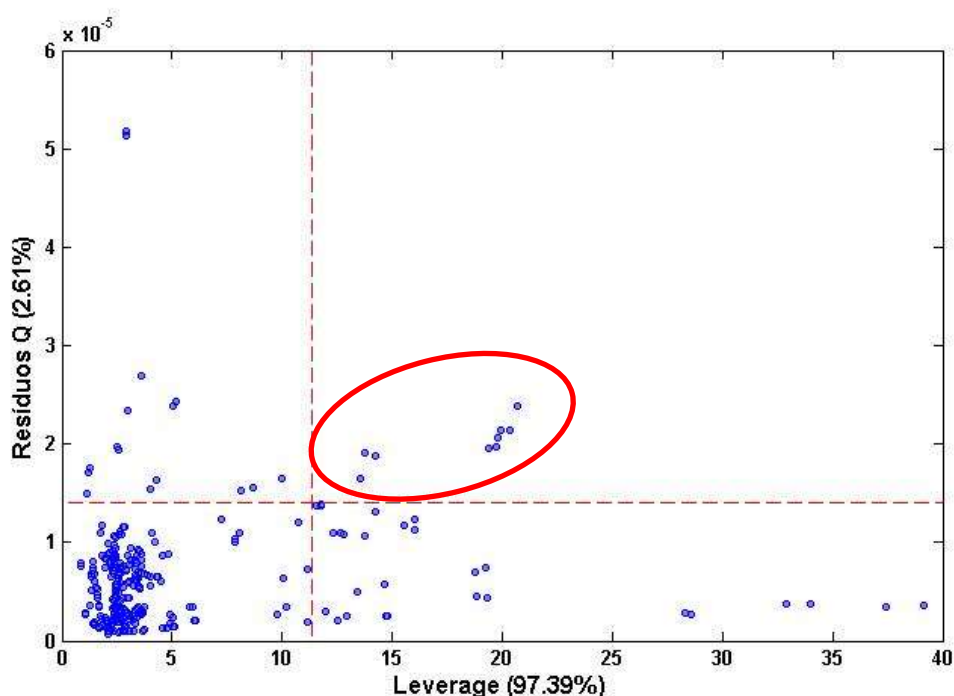
Após a otimização do modelo A, este foi usado para fazer a previsão direta de 300 amostras, cujos espectros foram obtidos no mesmo equipamento NIR empregado na obtenção dos espectros das amostras utilizadas na construção do modelo. Os resultados são apresentados na figura 4(B) e revelam que a previsão dessas amostras (representadas pelo símbolo “x” em vermelho) não foram adequadas visto que estão deslocadas da reta. Quanto ao valor de RMSEP este apresentou resultado de 5,9814 g/Kg.

O fato de o modelo A não conseguir prever adequadamente as amostras inseridas revela que ele necessita de uma atualização (*model update*) visando incluir possíveis interferentes dessas novas amostras no modelo. Primeiramente, a partir dos algoritmos DS e PDS (janela 5), buscou-se a atualização do modelo PLS para que este pudesse ser utilizado na previsão adequada das amostras do conjunto teste. As Figuras 4(C) e 4(D) apresentam o ajuste para a previsão do conjunto teste no modelo PLS após atualização por PDS e DS, respectivamente.

Por meio dos ajustes apresentados nas Figuras 4(C) e 4(D) verifica-se que os métodos PDS e DS não foram eficientes para atualizar o modelo PLS já que as amostras testes não demonstram uma linearidade. A ineficiência da Padronização Direta e Direta por Partes pode estar relacionada ao fato de que essas estratégias devem ser empregadas em dados obtidos em equipamentos distintos ou em um equipamento que tenha sofrido alguma mudança nas suas condições, como ter alguma parte substituída; e neste caso as amostras foram submetidas a um único equipamento NIR que não sofreu alteração, logo não é possível fazer a atualização empregando ferramentas utilizadas para fazer transferência de calibração, uma vez que DS e PDS funcionam bem quando há mudança no perfil espectral, e, nesse caso, com a previsão do conjunto externo ocorreu alteração no perfil químico. Diante disso, nota-se a necessidade de empregar outras alternativas para atualizar o modelo de calibração multivariada.

O próximo passo consistiu na construção de outro modelo, denominado de E, onde foi adicionado 10% das amostras teste ao modelo anterior. Na seleção dos 10% de amostras desse conjunto teste (30 amostras) foi novamente empregado o algoritmo de Kennard-Stone (KENNARD e STONE,

1969). Segundo os resultados apresentados na Figura 4(E), percebe-se que o ajuste deste modelo, construído com 20 VLs, foi melhor quando comparado ao ajuste alcançado empregando os métodos de atualização DS e PDS. Contudo, espera-se que esses resultados sejam superiores com o uso de outras estratégias de atualização. Assim, com o intuito de melhorar a atualização do modelo fez-se uma Análise de Componentes Principais (PCA – *Principal Component Analysis*) nas amostras do conjunto teste afim de identificar *outliers* (Figura 5) e eliminaram-se todas aquelas amostras que continham alto leverage e alto resíduo, amostras no quadrante superior direito, circuladas em vermelho.



**Figura 5 – Outliers das amostras do conjunto teste.**

**Fonte: autoria própria.**

Fazendo-se a previsão das amostras do conjunto teste, após remoção de *outliers*, no modelo A e avaliando o gráfico do ajuste, Figura 4(F), este revelou que esta estratégia também não se mostrou muito eficiente. Observando o modelo E, em que foram acrescentadas 10% das amostras do conjunto teste, e comparando com os demais (todos atualizados) percebeu-se que este foi quem apresentou o melhor resultado. Assim, acrescentou-se um

número maior de amostras do conjunto teste, 30%, para atualização do modelo. Novamente a seleção de 30% de amostras do conjunto teste (90 amostras) foi realizada pelo algoritmo de Kennard-Stone (KENNARD e STONE, 1969) e esse modelo, denominado G, foi construído com 15 VIs, apresentando melhor capacidade preditiva, como mostra o gráfico do ajuste na Figura 4(G).

Comparando os valores de RMSEP e coeficiente de correlação ( $R_v$ ) para os ajustes das amostras teste previstas em cada um dos modelos que passaram pelo *update*, se percebe que o modelo G foi o que demonstrou melhores resultados (Tabela 1).

**Tabela 1 – RMSEP e  $R_v$  como comparativo entre os modelos atualizados**

	RMSEP (g/Kg)	$R_v$
<b>Previsão direta</b>	5,9814	0,5316
<b>Atualização por PDS</b>	14,3480	0,1594
<b>Atualização por DS</b>	12,1992	0,0056
<b>Modelo E (adição 10%)</b>	5,7503	0,5687
<b>Remoção de <i>outliers</i></b>	6,2257	0,5283
<b>Modelo G (adição 30%)</b>	5,4201	0,6168

Fonte: autoria própria.

Avaliando as figuras de mérito para os modelos A e G, percebe-se que houve uma diminuição dos limites de detecção e de quantificação, o que é positivo visto que o modelo atualizado pelo acréscimo de 30% das amostras teste consegue quantificar uma concentração menor de carbono orgânico quando comparado ao modelo A, como mostra a Tabela 2.

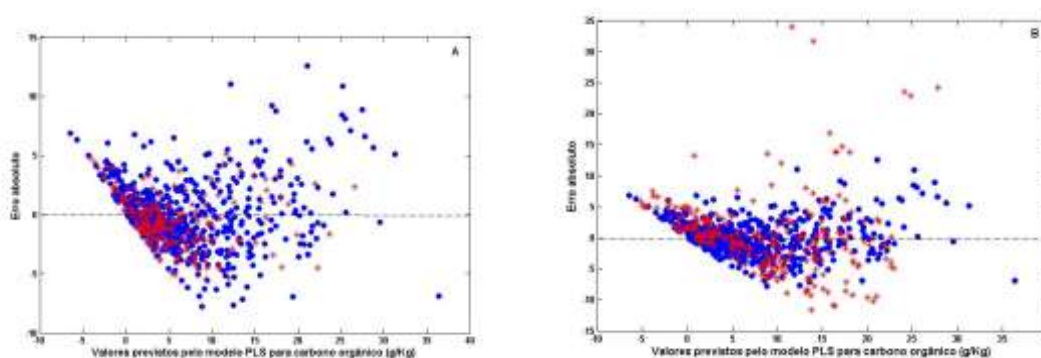
**Tabela 2 – Comparativa entre as figuras de mérito do modelo A e do modelo G**

Figuras de mérito		Modelo A (g/Kg)	Modelo G (g/Kg)
Exatidão	RMSEC	2,7134	3,3642
	RMSEP	2,3020	5,4201
Sensibilidade Analítica <sup>-1</sup>		1,5354	2,1619
Limite de Detecção		2,1493	1,5264
Limite de Quantificação		6,5131	4,6255

Fonte: autoria própria.

O inverso da sensibilidade analítica, que representa a menor diferença de concentração que o modelo é capaz de distinguir, mostrou-se menor no modelo A em relação ao modelo G.

O gráfico dos resíduos, Figura 6, mostrou um comportamento aleatório, indicando linearidade nos modelos construídos.



**Figura 6 - (A) Resíduos para o modelo A; (B) Resíduos para o modelo G.**

Fonte: autoria própria.

Comparando todas as estratégias de atualização testadas acredita-se que quanto maior o número de amostras do conjunto externo adicionado ao modelo, melhor será sua capacidade de previsão. Todavia, esse número não



pode ser muito grande, já que para tanto se necessita da concentração de carbono orgânico obtido a partir do método de referência o que implica em custos e demanda de tempo. Assim, um total de 30% do conjunto que necessita ser previsto promove um modelo atualizado em relação a capacidade preditiva.

## 6 CONCLUSÕES

Os modelos de calibração multivariada podem ser boas ferramentas de análise, já que fornecem resultados confiáveis rapidamente, não empregam reagentes e não possuem custo muito elevado; com essas características eles demonstram grande potencial em substituir os métodos de referência, como o Walckley & Black modificado (EMBRAPA, 1997). Contudo para que esses modelos matemáticos possam ser utilizados por um longo período de tempo, existe a necessidade de adotar alguma ferramenta que permita atualiza-lo sempre que preciso.

É útil fazer um *model update* sempre que se desejar fazer a previsão de amostras desconhecidas pelo modelo, principalmente amostras oriundas de outras regiões ou outros tipos de solo. Nesse sentido, a atualização realizada a partir de estratégias de transferência de calibração como PDS e DS não se mostraram satisfatórias. Os gráficos de ajuste, e parâmetros como o RMSEP e o  $R_v$ , confirmam que metodologias como a Padronização Direta e a Padronização Direta por Partes não foram eficientes para atualizar o modelo PLS, visto que os dados foram adquiridos em um mesmo equipamento, e segundo a literatura (HONORATO et al, 2007 e WANG et al, 1991) esse processo é eficiente quando se obtém os espectros em aparelhos distintos ou que tenham sofrido alterações (como por exemplo a troca de lâmpada).

A adição de uma porcentagem representativa de amostra do conjunto teste no modelo de calibração multivariada se revelou como a forma mais eficaz de atualização. Isso porque, nesse caso, as amostras (com espectros

adquiridos em um mesmo equipamento e que não passou por nenhuma mudança, o que acarretaria mudanças no perfil espectral) apresentam diferenças no perfil químico. Assim, a adição de 30% das amostras teste ao modelo multivariado promoveu uma atualização mais eficiente do que a adição de 10% das amostras teste ao modelo, com redução nos limites de detecção e quantificação. No entanto, mesmo tendo melhorado o modelo PLS com o aumento de dados externos, o número de acréscimos não pode ser muito elevado, pois isso implica que os valores do método de referência devem ser conhecidos, o que representaria um custo e demanda de tempo. A intenção é que com o método alternativo empregando NIR e calibração multivariada o número de análises com o método Walckley & Black modificado seja reduzido, para evitar a geração de resíduos tóxicos e o consumo de reagentes, por exemplo.

De modo geral, usar um modelo PLS é uma boa opção em relação à metodologia Walckley & Black modificado, entretanto para garantir a durabilidade do modelo de calibração é preciso atualiza-lo. Diante do estudo realizado a melhor escolha, entre as estratégias testadas, como ferramenta de fazer o *model update*, foi por meio da inserção de uma quantidade representativa de amostras do conjunto externo no modelo de calibração multivariada, uma vez que, nesse caso, as mudanças do conjunto externo em relação ao conjunto de calibração eram relacionadas ao perfil químico das amostras.

## REFERÊNCIAS

Annual Book of ASTM Standards. **Standards practices for infrared multivariate quantitative analysis**. E1655-05. ASTM International, West Conshohocken, Pennsylvania, USA, 2005.

BARBOZA, Fernando D.; **Desenvolvimento e a aplicação de modelos de transferência de calibração em espectroscopia no infravermelho próximo de ondas curtas (SW-NIR)**. 2003. 102 f. Dissertação (Mestrado em Química), Universidade Estadual de Campinas, Campinas, 2003.

BOOKSK, Karl S.; KOWALSKI, Bruce R; **Theory in analytical chemistry**. Analytical Chemistry, v. 66 (15), 782A – 7791A p., 1994.

BOUVERESSE, Eric; MASSART, Desire L.; DARDENNE, Pierre.; **Modified algorithm for standardization of near-infrared spectrometric instruments**. Analytical Chemistry, v. 67, 1381 - 1389 p., 1995.

BRERETON, Richard G.; **Introduction tom multivariate calibration in analytical chemistry**, Analyst, v. 125, 2125-2154 p., 2000.

CHATTERJEE, Ajay; LAL. Ramji.; WIELOPOLSKI, Lucian; MARTIN, Madhavi Z.; EBINGER, Michael H. **Evaluation of different soil carbon determination methods**. Critical Reviews in Plant Science, v. 28, 164-178 p., 2009.

COHEN, Matthew J.; PRENGER, Joseph P.; DeBUSK, Willian F. **Visible-near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality**. Journal of Environmental Quality, v. 34, 1422-1434 p., 2005.

DESPSGNE, Frédéric.; MASSART, D. Luc; CHABOT, Paul; **Development of a robust calibration model for nonlinear in-line process data**. Analytical Chemistry, v. 72 (7), 1657-1665 p., 2000.

EMBRAPA. **Manual de métodos de análise de solo** / Centro Nacional de Pesquisa de Solos. – 2. ed. rev. atual. – Rio de Janeiro, 1997. 212 p.

FERNANDES, Fernando A.; FERNANDES, Ana H.B.M.; BUENO SOBRINHO, Antonio A.; MONTEIRO, Hernandes de C.; SILVA, Augusto C.G. **Uso de espectrometria de refletância no infravermelho próximo (NIRS) na análise de carbono de Neossolos do Pantanal**. Embrapa. Comunicado Técnico, ISSN 1981-7231, dezembro 2010. Disponível em: <<http://www.cpap.embrapa.br/publicacoes/online/COT86.pdf>>. Acesso em: 09 jun. 2014.

FERRARINI, Hair. **Determinação de teores nutricionais do milho pro espectroscopia no infravermelho e calibração multivariada**. 107 f. 2004. Dissertação (Mestrado em Química). Universidade Federal do Paraná, Curitiba, 2004.

FEUDALE, Robert N., et al.; **Transfer of multivariate calibration models: a review**. Chemometrics and Intelligent Laboratory Systems. V. 64(2), 181-192 p., 2002.

FORINA, Michele; CASOLINO, Chiara; **Improvement of slope bias correction for the intercalibration of different instruments**. Quimica Analitica. V. 18(1), 49 – 59 p. 1999.

GELADI, Paul; KOWALSKI, Bruce R.; **Partial Least-Squares Regression: A Tutorial**. Anal. Chim. Acta, v. 185, 1-17 p., 1986.

HONORATO, Fernanda A., et al. **Transferência de calibração em métodos multivariados**. Química Nova, Vol. 30, No. 5, 1301-1312 p., 2007.

KENNARD, R.W.; STONE, L.A. **Computer aided desing of experiments**. Technometrics, vol. 11 (1), 137 -148 p., 1969.

LIMA, Francisco S. G.; BORGES, Luiz E. P.; **Evaluation of standardisation methods of near infrared calibration models**. Journal of Near Infrared Spectroscopy, v. 10, 269-278, 2002.

LUZ, Maria J.S.; FERREIRA, Gilvan B.; BEZERRA, José R.C. **Adubação e correção do solo: procedimentos adotados em função dos resultados da análise do solo**. Embrapa. Circular Técnica 63, Campina Grande – PB. 32 p. Outubro, 2002.

MORAIS, Cristiane S. **Aplicação de calibração multivariada em dados de espectroscopia UV-visível para previsão da acidez total em vinhos**. 2014. 21 f. Trabalho de Conclusão de Curso (Graduação) – Curso Superior de Tecnologia de Alimentos. Universidade Tecnológica Federal do Paraná, Campo Mourão, 2014.

NELSON, Davi W. & SOMMERS, Lee E. **Total carbon, organic carbon, and organic matter**. In: BLACK, C.A., ed. Methods of soil analysis. Part 3. Chemical methods. Madison, Soil Science of America and American Society of Agronomy, 961-1010 p., 1996.

PEREIRA, Leandro S.A. **Transferência de calibração de misturas de pós para comprimidos intactos: um novo uso em análise farmacêutica para uma ferramenta quimiométrica conhecida**. 2013. 51 f. Dissertação (Mestrado em Química) - Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

PRADO, Alexandre. G. S., **Química Verde, os desafios da química do novo milênio**. Química Nova, Vol. 26, No. 5, 738-744 p., 2003.

SANTOS, Humberto Gonçalves dos, et al. **Sistema Brasileiro de Classificação de Solos**. 3a ed. Embrapa, Brasília - Distrito Federal, 2013.

SATO, Juliana H. **Métodos para determinação do carbono orgânico em solos do Cerrado**. 2013. 79 f. Dissertação (Mestrado em Agronomia) - Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, 2013.

Savitsky, Abraham e Golay, Marcel J. E.; **Smoothing and differentiation of data by simplified least squares procedures**. Analytical Chemistry, v. 36, 1627 – 1639 p., 1964.

SENA, Marcelo M., et al. **Avaliação do uso de métodos quimiométricos em análise de solos**. Química Nova, v. 23(4), 547-556 p., 2000.

SHENK, John S.; WESTERHAUS, Mark O.; TEMPLETON Jr, Willian C.; **Calibration Transfer Between near Infrared Reflectance Spectrophotometers**. Crop Sci., v. 25, 159-161 p., 1985.

SOUZA, André M. **Avaliação do potencial da espectroscopia no infravermelho próximo como método de rotina para a determinação de carbono orgânico do solo**. 2014. 158 f. Tese (Doutorado em Ciências) – Universidade Estadual de Campinas, Campinas, 2014.

VALDERRAMA, Patrícia. **Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo**. 2005. 128 f. Dissertação (Mestrado em Química) - Universidade Estadual de Campinas, Campinas, 2005.

VALDERRAMA, Patrícia. **Calibração multivariada de primeira e segunda ordem e figuras de mérito na quantificação de enantiômeros por espectroscopia**. 2009. 219 f. Tese (Doutorado em Química) – Universidade Estadual de Campinas, 2009.

WANG, Yongdong; VELTKAMP, David J.; KOWALSKI, Bruce R. **Multivariate Instrument Standardization**. Analytical Chemistry, v. 63, 2750-2756 p., 1991.

WOLD, Svante; SJÖSTRÖM, Michael; ERIKSSON, Lennart; **PLS-regression: a basic tool of chemometrics**. ELSEVIER. Chemometrics and Intelligent Laboratory Systems. V. 58. 109–130 p. 2001.

ZENI, Diego. **Determinação de cloridrato de propranol em medicamentos por espectroscopia no infravermelho com calibração multivariada (PLS)**. 2005. 54 f. Dissertação (Mestrado em Química) - Universidade Federal de Santa Maria, Santa Maria, 2005.