

Indicação de fórmula para estimativa de teores de carbono em solos utilizando Máquina de Vetor de Suporte e refletância espectral

RESUMO

O carbono constitui-se em um importante atributo na capacidade produtiva do solo. Porém, as tradicionais metodologias empregadas para sua determinação geram problemas ambientais devido ao uso de reagentes químicos. A espectroscopia é uma das técnicas promissoras na Agricultura de Precisão para análises de solos e que pode trazer uma solução viável para análise de teor de carbono. Dentre suas vantagens, destaca-se a preservação da amostra, o não consumo de reagentes, além de sua eficiência na aquisição de dados provenientes de um grande número de amostras. O objetivo deste trabalho foi contribuir com um modelo de regressão capaz de prever a quantidade de carbono em amostras de solo utilizando a espectroscopia na região do visível e no infravermelho próximo. Para tanto, foi utilizada a técnica de Aprendizagem de Máquina SVM incorporada ao software WEKA como auxílio na criação do modelo. A SVM tem representado uma alternativa melhor aos já consagrados métodos de regressão multivariada por apresentar capacidade de generalização. Nos experimentos realizados foi utilizado um conjunto de amostras de solo coletadas na região dos Campos Gerais. A avaliação dos resultados teve como base os erros de previsão e os coeficientes de correlação entre os valores dos teores de carbono preditos pelo modelo. Foram encontrados coeficientes de correlação 0,89. Concluiu-se que a espectroscopia no vis-NIRS aliada à técnica SVM pode ser recomendada como uma alternativa aos métodos convencionais de análise de carbono em solos.

PALAVRAS-CHAVE: Agricultura de Precisão, WEKA, vis-NIRS;

ABSTRACT

Carbon is an important attribute in the soil productivity. However, due to the use of chemical reagents traditional methods used for their determination generate environmental problems. Spectroscopy is one of the promising techniques in Precision Agriculture for soil analysis and can become a viable solution for carbon analysis. Among its advantages, there is the preservation of the sample, no reagent consumption, as well as its efficiency in acquiring data from a large number of samples. The objective of this work was to determine a regression

model able to predict the amount of carbon in soil samples using spectroscopy in the visible and near infrared. SVM machine learning technique incorporated into the WEKA software was used to analyze the data. SVM has been considered a better alternative than others already established multivariate regression methods because of its generalization capability. A set of soil samples collected in the Campos Gerais region – Parana State - was used. The evaluation of the results was based on the forecast errors and the correlation coefficients taking into account the carbon levels predicted by the model. The best Correlation coefficient was 0.89. It was concluded that the vis-NIRS spectroscopy coupled to SVM technique can be recommended as an alternative to conventional carbon analysis methods in soils.

KEYWORDS: Precision Agriculture, WEKA, vis-NIRS;

INTRODUÇÃO

O aumento na produtividade das culturas é visível quando os atributos físicos, químicos e biológicos do solo estão equilibrados e suficientemente disponíveis. Para que haja este equilíbrio a utilização dos princípios e tecnologias da Agricultura de Precisão (AP) torna-se inerente. Hoje, especialmente no Brasil, as soluções existentes estão focadas na aplicação de fertilizantes e corretivos em taxa variável, porém não se deve perder de vista que AP é um sistema de gestão que considera a variabilidade espacial das lavouras em todos seus aspectos: produtividade, solo (características físicas, químicas, compactação, etc), infestação de ervas daninhas, doenças e pragas (MAPA, 2013).

Considerado um indicador de qualidade, o carbono é um importante atributo na capacidade produtiva do solo. Dentre as diversas metodologias empregadas para determinação de carbono destaca-se a desenvolvida por Walkley & Black em 1934, que utiliza o princípio da combustão úmida (SATO, 2013). Entretanto, problemas ambientais devido ao uso de cromo (BRUNETTO et al., 2006), tem estimulado a substituição desse procedimento por outros que geram menor ou nenhuma quantidade de resíduos potencialmente tóxicos.

Segundo Alkimin (2011), a espectroscopia é considerada uma técnica promissora para análises de solos. Dentre suas vantagens, destaca-se a preservação da amostra, o não consumo de reagentes, além de sua eficiência na aquisição de dados provenientes de um grande número de amostras. Adicionalmente, um único espectro pode ser usado para avaliar diferentes atributos do solo. Para Shepherd e Walsh (2007), é uma das técnicas analíticas mais eficientes e disponíveis do século 21. A espectroscopia gera um conjunto de dados após a leitura das amostras. De posse destes dados é possível extrair conhecimento aplicando-se técnicas de aprendizado de máquina supervisionada.

Uma técnica que vem sendo utilizada pelos pesquisadores é a Support Vector Machine (SVM) que contempla algoritmos de Aprendizado de Máquina para regressão. A SVM consiste na busca de uma função ótima na separação de um hiperplano de decisão na tentativa de maximizar a capacidade de generalização. Áreas diversas como medicina, química, agricultura e imagens têm utilizado a SVM como ferramenta de aprendizado de máquina, como nos trabalhos de Orru et al. (2012), Keqiang et al. (2014), Fonseca et al. (2010), Chen et al. (2012).

Neste contexto, o objetivo deste trabalho foi contribuir com um modelo de regressão capaz de prever a quantidade de carbono em amostras de solo utilizando a espectroscopia na região do visível e no infravermelho próximo. Para tanto foi utilizada a técnica de aprendizado de máquina SVM como auxílio na criação do modelo.

MATERIAL E MÉTODOS

Coleta das amostras

As amostras de solo foram coletadas em maio de 2004, na região dos Campos Gerais no Estado do Paraná (25°8' S, 50°15' W), com altitude média de 853 m, cujo solo é classificado como Latossolo Vermelho-Escuro distrófico textura média e com alta acidez (JORIS, 2011). Cinco diferentes camadas de profundidade foram coletadas. O estudo utilizou 38 amostras de 00-05 cm, 38 amostras de 05-10 cm, 38 amostras de 10-20 cm, 59 amostras de 20-40 cm e 58 amostras de 40-60 cm totalizando 231 amostras. Todas as amostras foram secas em estufa com circulação forçada de ar a 40 °C e passadas em peneira com malha de 2 mm. O carbono orgânico foi determinado de acordo com o método WALKLEY & BLACK descrito em Pavan et al. (1992).

Análise de reflectância das amostras

O conjunto de amostras foi analisado no espectrômetro marca FOSS, modelo XDS Near-infrared, preparado para leitura de espectros de refletância difusa na região do infravermelho próximo. A faixa de comprimento de onda lido foi de 400 a 2500 nanômetros com intervalo de 2 nm a cada varredura. O software utilizado para aquisição dos espectros foi o ISIScan versão 3.2 (PROENÇA, 2012).

Construção do conjunto de dados utilizado

O conjunto de dados analisado no espectrômetro continha as leituras de 231 amostras com os resultados das 1051 (um mil e cinquenta e um) faixas de comprimentos de onda. Para cada uma das amostras o correspondente teor de carbono foi adicionado ao conjunto, totalizando dessa forma 1052 atributos.

Geração de modelos de regressão multivariada

Os modelos de regressão multivariada foram gerados utilizando-se o software WEKA (*Waikato Environment for Knowledge Analysis*) que é uma ferramenta de KDD (Knowledge Discovery in Database) a qual contempla uma série de algoritmos de preparação de dados, de aprendizado de máquina e de validação de resultados (HALL et al, 2009). De posse do conjunto de dados formatado de acordo com o padrão utilizado pelo software WEKA versão 3.7.10 foi possível o carregamento e aplicação do algoritmo SMOReg. De acordo com Cornélio (2012), SMOReg é uma implementação do algoritmo de Otimização Sequencial Mínima (*Sequential Minimal Optimization - SMO*) para treinamento de um modelo de regressão de vetor de suporte. O algoritmo SMOReg consiste em uma implementação Java e está disponível no pacote de software WEKA.

Os espectros do visível (vis) e do infravermelho próximo (nir) foram analisados separadamente utilizando o algoritmo SMOReg. O Kernel utilizado foi o polinomial de grau 2 e com o parâmetro de regularização $C = 10^6$. Os dados não sofreram normalização. A validação cruzada em 10 blocos foi utilizada para avaliar a capacidade de generalização. Erros de previsão e os coeficientes de correlação serviram de parâmetro para avaliar a capacidade de predição dos modelos obtidos.

RESULTADOS E DISCUSSÃO

De posse dos dados foi aplicada a redução da dimensionalidade para selecionar um conjunto de até dez espectros mais relevantes, visando gerar uma equação de regressão de tamanho menor, uma vez que, um dos resultados apresentados pelo algoritmo SMOReg é uma equação tendo como parâmetro de entrada as leituras dos espectros e como saída a predição em questão. Para a seleção dos atributos foi utilizado o filtro de seleção baseado em correlação o *CfsSubsetEval* (*Correlation-based Feature Selection*) com algoritmo *GreedyStepwise*. Esse seletor analisa a capacidade de predição de cada atributo no subconjunto juntamente com o grau de redundância entre os atributos. Um subconjunto é considerado bom se os atributos contidos nele são altamente correlacionados com a classe e contém atributos não correlacionados entre si. Para os conjuntos de dados vis e nir, com a redução da dimensionalidade foram selecionadas as bandas espectrais mais relevantes para o estudo, conforme apresentado na Tabela 1.

Tabela 1 – Espectros selecionados nos conjuntos de dados, após redução da dimensionalidade.

Conjuntos	Bandas Espectrais (nm)
Conjunto de dados /vis	438, 470, 490, 502, 518, 566, 596, 630, 662, 694
Conjunto de dados /nir	748, 962, 1096, 1310, 1520, 1726, 1900, 2118, 2302, 2496

A Tabela 2 apresenta os resultados da aplicação do algoritmo de regressão nas dez bandas espectrais selecionadas (Tabela 1) nos espectros vis e nir, tendo como atributo de referência o teor de carbono.

Tabela 2 - Resultado da predição para o conjunto de dados com os espectros vis e nir após a redução de dimensionalidade.

Parâmetros	VIS	NIR
Nº de Amostras	231	231
Atributo Meta	Carbono	Carbono
Nº de Bandas Espectrais	10	10
Coefficiente de Correlação	0,891	0,890
Erro médio absoluto	1,394	1,405
Raiz quadrada do erro médio	1,782	1,783
Erro absoluto relativo	44,25%	44,61%
Raiz quadrada do erro relativo	45,40%	45,42%

Os resultados obtidos, tanto para o intervalo VIS quanto o NIR alcançaram valores satisfatórios de coeficiente de correlação de 0,89. Porém o erro relativo de 45,4% indica que o modelo deve ser ajustado para minimizar o erro. Esse ajuste poderá ser feito a medida que mais dados amostrais sejam inseridos na base. As equações geradas por meio do algoritmo SMOReg para os espectros vis (Equação 1) e nir (Equação 2) são apresentadas a seguir.

$$\begin{aligned} \text{Conjunto de dados /vis} = & (- 314,6805 * r438 - 887,7766 * r470 - 242,2445 * r490 \\ & - 36,7462 * r502 + 960,5514 * r518 + 391,4396 * r566 + 682,9357 * r596 + \\ & 925,6285 * r630 - 3881,1982 * r662 + 2207,5724 * r694 + 8,7281) \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Conjunto de dados/nir} = & (- 82,8045 * r748 + 1028,3264 * r962 - 1268,0595 * \\ & r1096 + 117,4458 * r1310 - 168,5211 * r1520 + 1102,7026 * r1726 + 162,9387 \\ & * r1900 - 1177,1477 * r2118 + 156,2091 * r2302 + 203,3498 * r2496 + 21,218) \end{aligned} \quad (2)$$

Para validar os modelos de regressão obtidos, todo o conjunto de dados foi utilizado como entrada em um teste de validação. Os resultados de 197 amostras mais representativas, ou seja, 85% da base, serviu como fonte para que fosse gerado um gráfico de dispersão com o respectivo coeficiente de determinação tanto para a região do Visível (Figura 1), quanto para a região do Infravermelho Próximo (Figura 2).

Figura 1 - Gráfico de dispersão com teor de carbono estimado versus teor de carbono de referência para espectros VIS do conjunto de dados.

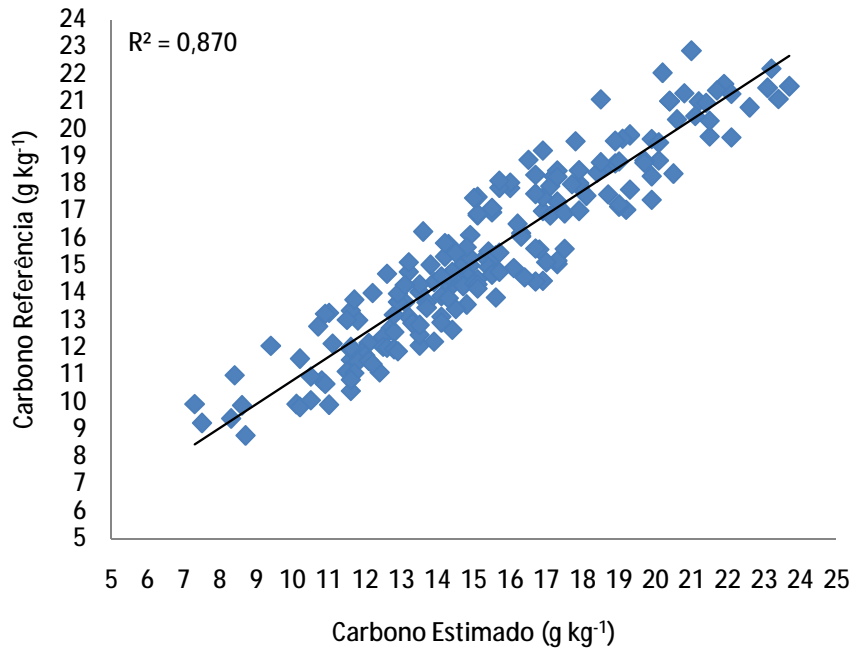
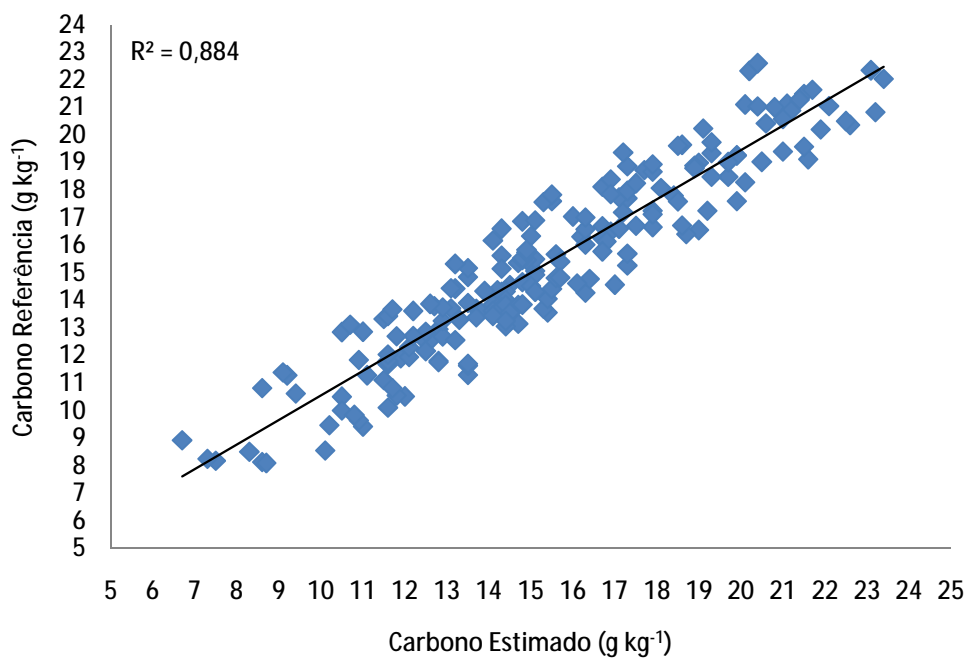


Figura 2 - Gráfico de dispersão com teor de carbono estimado versus teor de carbono de referência para espectros NIRS do conjunto de dados.



Os resultados obtidos com a determinação do carbono são semelhantes aos encontrados por Sato (2013) (R^2 0,86) na pesquisa feita em 64 amostras de solo do Cerrado brasileiro utilizando a faixa de 1100 nm a 2500 nm. Já Marchão et al.(2011) obteve um R^2 0,91 utilizando um conjunto de 390 amostras e os comprimentos de onda compreendidos entre 1100 nm e 2500 nm. Os autores realizaram a regressão pelo método dos mínimos quadrado parciais modificado (mPLS) para desenvolver os modelos de calibração.

Fora do Brasil MCCarty et al.(2010) no oeste da África realizando uma comparação entre espectroscopia na região do infravermelho próximo (nir) e médio (mir) na predição de carbono obteve um R^2 de 0,90, também satisfatório.

CONCLUSÕES

Os resultados alcançados na aplicação da técnica de aprendizado de máquina SVM em conjuntos de dados de espectros na região do visível e do infravermelho próximo, obtidos da leitura de amostras de solo para predição de carbono, foram satisfatórios.

Com os modelos de regressão gerados foi possível verificar que a espectroscopia no vis-NIR associada a técnica SVM podem vir a ser empregadas na análise de teores de carbono em solos, destacando-se por ser mais limpa, ou seja não poluente, e viável também em termos de rapidez no resultado.

AGRADECIMENTOS

A Fundação ABC por ceder a infraestrutura de laboratório e equipamento para a realização das análises.

REFERÊNCIAS

ALKIMIM, A. F., et al. Avaliação da refletância espectral de solos representativos da bacia do rio Benevente com o emprego da análise de componentes principais. Anais XV Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Curitiba, PR, Brasil, 30 de abril a 05 de maio de 2011, INPE página 9064, 2001.

BRUNETTO, G.; MELO, G.W.; KAMINSKI, J.; FURLANETTO, V.; FIALHO, F.B. Avaliação do método de perda de peso por ignição na análise de matéria orgânica em solos da Serra Gaúcha do Rio Grande do Sul. Ciência Rural, 36(6):1936-1939, 2006.

CHEN, C.; YANG, E.; JIANG, J.; LIN T.; An imaging system for monitoring the in-and-out activity of honey bees. *Computers and Electronics in Agriculture*, n. 89, p. 100-109, 2012.

FONSECA, A. de S. et al. Síntese e caracterização estrutural do ligante isatina-3-(N4-benziltiossemicarbazona) e do seu complexo de mercúrio(II). *Quím. Nova, São Paulo*, v. 33, n. 7, 2010.

HALL, Mark et al. The WEKA Data Mining Software: An Update; *SIGKDD Explorations, Witten*. Volume 11, Issue 1, 2009.

JORIS, H. A. W., Calagem Superficial, umidade do solo e comportamento do milho cultivado em sistema plantio direto. Dissertação (Mestrado em Agronomia) – Universidade Estadual de Ponta Grossa, 63 páginas, 2011.

KEQIANG, Y.; YANRU Z.; XIAOLI L.; YONGNI S.; FENGLE Z.; YONG H. Identification of crack features in fresh jujube using Vis/NIR hyperspectral imaging combined with image processing. *Computers and Electronics in Agriculture*, v. 103, p. 1-10, 2014.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. Agricultura de Precisão. Boletim Técnico. Brasília, 36 páginas, 2013. Disponível em: http://www.agricultura.gov.br/arq_editor/Boletim%20T%C3%A9cnico%20AP.pdf, Acessado em 10 de Dezembro de 2013.

MARCHÃO, R.L.; BECQUER, T.; BRUNET, D. Predição dos teores de carbono e nitrogênio do solo utilizando espectroscopia de infravermelho próximo. Planaltina, DF: Embrapa Cerrados, página 21, 2011.

MCCARTY, G.W.; REEVES III, J.B.; YOST, R.; DORAISWAMY, P.C.; DOUMBIA, M. Evaluation of methods for measuring soil organic carbon in West African soils. *African Journal of Agricultural Research*, página 2177, 2010.

ORRÙ G.; PETTERSSON-YEO W.; MARQUAND A.F.; SARTORI G.; MECHELLI A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*, n. 36(4), 1140-52, 2012.

PAVAN, M.A.; BLOCH, M.F.; ZEMPULSKI, H.C.; MIYAZAWA, M.; ZOCOLER, D.C. Manual de análise química do solo e controle de qualidade. Londrina: Instituto Agronômico do Paraná (Circular, 76), p. 38, 1992.

PROENÇA, C. A. Redes Neurais Artificiais para predição dos teores de matéria orgânica e argila do solo na região dos Campos Gerais utilizando Espectroscopia de Reflectância Difusa. Dissertação (Mestrado em Computação Aplicada) - UEPG, Ponta Grossa – PR., 2012.

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

SHEPHERD, K. D.; WALSH, M. G. Infrared spectroscopy - enabling an evidence based diagnostic surveillance approach to agricultural and environmental management in developing countries: *Journal of Near Infrared Spectroscopy*, Charlton, v.15, p.1-19, 2007.