

TRANSFORMADA WAVELET E MÁQUINAS DE VETORES DE SUPORTE PARA PREDIÇÃO DE SST EM UVAS VINÍFERAS COM USO DE ESPECTROSCOPIA VIS-NIR

DANIELE SILVA REIS¹, ÁLISSON PEREIRA DOS SANTOS¹, THAÍS PEREIRA DE
AZEVEDO², DANIEL DOS SANTOS COSTA³, RODRIGO PEREIRA RAMOS⁴,
BÁRBARA JANET TERUEL MEDEROS⁵

¹Estudante de Graduação, Engenharia da Computação, UNIVASF/Juazeiro-BA

²Engenheira Agrícola e Ambiental, Profa. Assistente I, UNIVASF/Juazeiro-BA

³Engenheiro Agrícola e Ambiental, Prof. Adjunto I, UNIVASF/Juazeiro-BA, daniel.costa@univasf.edu.br

⁴Engenheiro Eletricista, Prof. Associado III, UNIVASF/Juazeiro-BA

⁵Engenheira Mecânica, Prof. Livre Docente, UNICAMP/Campinas-SP

Apresentado no
XLVIII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2019
17 a 19 de setembro de 2019 - Campinas - SP, Brasil

RESUMO: O presente estudo possui como objetivo a análise do desempenho de técnicas de processamento digital de sinais e aprendizagem de máquina para predição de SST em uvas da variedade Cabernet Sauvignon com uso da espectroscopia VIS-NIR. Os dados espectrais foram submetidos a diferentes pré-processamentos, dentre eles a transformada wavelet, filtro passa-baixa, filtro da mediana e da média móvel, com posterior diferenciação por Savitzky-Golay e normalização por MSC e SNV. Os modelos matemáticos de regressão utilizados foram a regressão por quadrados mínimos parciais (PLSR) e regressão por vetores de suporte (SVR) para correlação com o atributo de qualidade SST. A predição do atributo SST apresentou como melhor resultado um R^2 da validação externa igual a 0,9654 e RMSEP igual a 1,2553 °Brix, através da regressão por SVR e alisamento por transformada wavelet. O estudo demonstrou a superioridade das técnicas baseadas em aprendizagem de máquina e processamento digital de sinais em detrimento dos métodos convencionais.

PALAVRAS-CHAVE: atributo de qualidade, aprendizagem de máquina, processamento digital de sinais.

WAVELET TRANSFORM AND SUPPORT VECTOR MACHINES FOR TSS PREDICTION IN VINIFEROUS GRAPES USING VIS-NIR SPECTROSCOPY

ABSTRACT: This study aims to analyze the performance of digital signal processing techniques and machine learning to predict SST in Cabernet Sauvignon grapes using VIS-NIR spectroscopy. The spectral data were submitted to different pre-processing, among them the wavelet transform, low-pass filter, median and moving average filters, with subsequent differentiation by Savitzky-Golay and normalization by MSC and SNV. The mathematical regression models used were by partial least squares regression (PLSR) and by support vectors regression (SVR) for correlation with the SST quality attribute. The prediction of the SST attribute presented as the best result an external validation R^2 equal to 0.9654 and RMSEP equal to 1.2553 °Brix, through SVR regression and smoothing by wavelet transformation. The study demonstrated the superiority of techniques based on machine learning and digital signal processing over conventional methods.

KEYWORDS: quality attribute, machine learning, digital signal processing.

INTRODUÇÃO: A uva é uma das frutas mais cultivadas no mundo, alcançando a marca de aproximadamente 75 milhões de toneladas produzidas por ano, sendo a maior parte dessa

produção destinada à fabricação de vinho (FAO-OIV, 2016). Dentre as variedades destinadas a esse mercado, destaca-se a Cabernet Sauvignon, que apresentou a maior área de cultivo dentre as demais variedades nos últimos anos (OIV, 2017). Para alcançar uma boa qualidade do vinho produzido, a indústria vinícola deve observar os atributos de qualidade das uvas utilizadas no processo. Sendo assim, torna-se essencial avaliar o atributo sólidos solúveis totais (SST), que indica a quantidade de açúcares existentes no fruto. Entretanto, a avaliação das qualidades internas das uvas é realizada através de processos complexos, laboriosos e demorados, em que um trabalho manual intenso é exigido. Ademais, as amostras utilizadas precisam ser destruídas para que as análises químicas sejam realizadas. A espectroscopia de refletância VIS-NIR tem se destacado na área de ciência de alimentos para a determinação dos atributos de qualidade por ser uma técnica simples, não destrutiva e por apresentar resultados rápidos. Entretanto, o grande volume de dados fornecido pela espectroscopia VIS-NIR contém informações não desejadas, além de ruído. Assim, para que as informações analíticas contidas nos sinais espectrais das uvas possam ser associadas à SST, torna-se necessário realizar determinados tratamentos matemáticos sobre os dados. Após isso, esses podem então ser utilizados para a calibração de modelos de regressão multivariada. Assim, o presente estudo possuiu como objetivo determinar o desempenho da transformada *wavelet* e máquinas de vetores de suporte para predição de SST em uvas viníferas com uso de espectroscopia VIS-NIR.

MATERIAL E MÉTODOS: As amostras de uva (*Vitis vinifera* L.) da variedade Cabernet Sauvignon foram coletadas de um vinhedo localizado na região do Vale do Submédio São Francisco (Lagoa Grande, Pernambuco, Brasil). Foram coletadas 576 bagas em diferentes estádios de maturação, sendo cada submetida a um espectrorradiômetro FieldSpec 3 (*Analytical Spectral Devices*, Boulder, Colorado, USA), com campo de visão do sensor óptico de 8°, faixa de 350 a 2500 nm, resolução de 3 a 10 nm, tempo de varredura de 100 ms e fonte de luz de quartzo-tungstênio-halogênio de 70 watts. Os valores de refletância foram transformados em absorbância. As bandas de ondas ruidosas em ambas as extremidades do espectro foram removidas e apenas dados espectrais nas regiões de 450 a 1800 nm foram utilizados na análise de dados. A determinação de sólidos solúveis totais (SST) foi realizada a partir da leitura direta em refratômetro digital (HI 96804, *Hanna Instruments*, USA), utilizando-se duas gotas da amostra triturada e homogeneizada, com os resultados expressos em °Brix. Quatro diferentes técnicas para alisamento foram aplicadas com o objetivo de remover ruídos aleatórios e aumentar a relação sinal/ruído dos espectros (CEN et al., 2007), sendo eles o filtro da mediana, filtro da média móvel, filtro passa-baixa, e a transformada *wavelet*. Os filtros da mediana e média móvel foram testados com valores de janela entre 7 e 27 pontos, enquanto o filtro passa-baixa foi construído para ordens igual a 10, 20 e 26, com frequências de corte entre 200 e 1000 Hz. Os filtros *wavelet* utilizou-se de duas técnicas para limiarização, sendo o *global threshold* e o *scale-adaptive threshold* (DONOHO et al., 1994). Foram utilizadas as funções-mãe *daubechies*, *symlets* e *coiflets*. Nos experimentos com transformada *wavelet*, optou-se por testar a técnica de alisamento em partes do sinal, assim como no espectro inteiro, pois notou-se a deterioração de faixas do sinal que não continham ruído. As faixas de espectro selecionadas foram 450-600 nm e 1400-1800 nm. Para a diferenciação, foi aplicada a primeira e segunda derivada por Savitzky-Golay, com polinômio de segundo grau e janelas de tamanho 7 e 21 pontos. Para a normalização utilizou-se MSC e SNV. Os modelos preditivos utilizados foram a regressão por quadrados mínimos parciais (PLSR) e regressão por máquinas de vetores de suporte (SVR). Os modelos de regressão foram calibrados através das duas técnicas (PLSR e SVR), e validados através da validação cruzada 10-*fold*. Por fim, os melhores modelos obtidos foram submetidos à validação externa

para a avaliação de sua capacidade preditiva. A avaliação de desempenho dos modelos foi realizada através das métricas R^2 e RMSE das validações cruzada e externa.

RESULTADOS E DISCUSSÃO: O modelo SVR apresentou melhores resultados que o modelo PLSR em todas as métricas, conforme a Tabela 1. Apesar da SVR utilizar a maioria das amostras para a construção da função *epsilon-intensive*, o modelo apresentou uma boa capacidade de generalização, com RMSECV (root means standard error cross-validation) de 0,8565 °Brix. Os resultados para validação externa mostraram-se bons da mesma forma, com R^2 alcançando 0,9654 e RMSEP (root means standard error prediction) de 1,2553 °Brix. Além disso, o modelo alcançou um custo C igual a 1, indicando uma baixa complexidade do modelo e uma tolerância média a erros.

TABELA 1. Desempenho dos modelos PLSR e SVR para predição de SST em uvas viníferas.

Atributo	Modelo	Pré-tratamento	VL/VS	Validação cruzada		Validação externa	
				R^2	RMSECV	R^2	RMSEP
SST	PLSR	Nenhum	9	0,9199	1,8518	0,9168	1,9736
	SVR	Nenhum	370	0,9048	2,0715	0,9266	1,9605
	PLSR	Mediana 9/SG1st2x7/MS*	7	0,9679	1,1703	0,9483	1,5232
	SVR	R1-ST-symlets 3/SG1st2x21/SNV**	344	0,9828	0,8565	0,9654	1,2553

* Alisamento por filtro da mediana com janela de tamanho 9, diferenciação por primeira derivada com polinômio de ordem 2 e janela de tamanho 7 e normalização por MSC. ** Alisamento por transformada *wavelet*, utilizando a função-mãe symlets 3, com *threshold scale-adaptive*, aplicada no espectro inteiro, diferenciação por primeira derivada com polinômio de ordem 2 janela de tamanho 21 e normalização por SNV.

A necessidade do emprego de processamento dos sinais espectrais é evidenciada ao observar as métricas obtidas para os dados brutos. O modelo PLSR apresentou complexidade maior ao utilizar 9 variáveis latentes para explicar a covariância entre os dados, alcançando coeficientes de determinação menores que aqueles obtidos após o pré-processamento, enquanto os valores para RMSE encontrados foram maiores. Este mesmo comportamento foi observado no modelo SVR, que necessitou de uma maior quantidade de vetores de suporte para a construção da função de regressão. O emprego do espectro inteiro para alisamento mostrou-se a melhor abordagem para alisamento, apesar de que com a utilização de apenas duas faixas obteve-se um desempenho ligeiramente inferior. A predominância da transformada *wavelet* no alisamento dos sinais espectrais já era esperada, pois ela permite a atenuação ou remoção de coeficientes menos significantes de acordo com um limiar. Da mesma forma, a SVR mostrou-se superior à PLSR, devido à sua busca pelos vetores de suporte que melhor constroem a função *epsilon-intensive*. O modelo PLSR construído mostrou-se mais simples que o construído por HERRERA et al. (2003), que construíram um modelo de 20 variáveis com base na espectroscopia de transmitância para a faixa 650-1100 nm. Esses mesmos autores alcançaram um valor de R^2 para validação interna igual a 0,942, inferior ao obtido no presente estudo. Os erros médios quadráticos para validação externa mostraram-se menores, alcançando o valor de 1,06 °Brix, mas com RMSECV maior (2,96 °Brix).

CONCLUSÕES: Os resultados demonstraram a superioridade da SVR para predição de SST em uvas da variedade Cabernet Sauvignon, ao apresentar um modelo de regressão robusto e generalista, alcançando altos coeficientes de determinação e RMSE baixos. Demonstrou-se também que técnicas baseadas em processamento digital de sinais são superiores às demais técnicas comumente adotadas, tendo em vista que o emprego da transformada *wavelet* para alisamento resultou em uma redução da complexidade dos modelos SVR, ao utilizar valores para C inferiores, além de contribuir para um desempenho superior.

REFERÊNCIAS

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v. 18, n. 2, p. 72-83, 2007.

DONOHO, D. L.; JOHNSTONE, Iain M. *Threshold selection for wavelet shrinkage of noisy data*. In: Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE. IEEE, 1994. p. A24-A25 vol. 1.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Table and Dried Grapes: FAO-OIV Focus 2016. Disponível em: <<http://www.fao.org/statistics/en/>>. Acesso em: 02 abr. 2018.

HERRERA, J.; GUESALAGA, A.; AGOSIN, E. Shortwave–near infrared spectroscopy for non-destructive determination of maturity of wine grapes. **Measurement Science and Technology**, v. 14, n. 5, p. 689, 2003.

INTERNATIONAL ORGANISATION OF VINE AND WINE. 2017 World Vitiviniculture Situation: OIV Statistical Report on World Vitiviniculture. Disponível em: <<http://www.oiv.int/en/technical-standards-and-documents>>. Acesso em: 02 abr. 2018.