



Influência do pré-processamento e escolha das variáveis latentes na construção de modelos PLS para quantificação de fenólicos totais em própolis

Matheus Augusto Calegari¹ Anaclara Prasniewski² Cleidiane da Silva³
Tatiane Luiza Cadorin Oldoni⁴

17 maio 2017

Resumo – A própolis é uma substância resinosa produzida pelas abelhas e sua principal função é a defesa da colmeia, pois as abelhas a utilizam para vedar pequenas frestas, protegendo-a contra micro-organismos e das intempéries, mumificar corpos invasores e desinfecção. Possui, ainda diversas propriedades farmacológicas que atuam em prol da saúde humana como a capacidade antioxidante, que inibe a oxidação de macromoléculas. Os principais compostos relacionados à atividade antioxidante são os compostos fenólicos, sendo comumente determinados em extratos de própolis, como maneira de medir a qualidade do material. Objetivou-se desenvolver modelos de calibração multivariada, baseado em análises utilizando o infravermelho próximo (NIR) combinado à ferramenta quimiométrica por PLS, para determinação de compostos fenólicos totais, verificando qual pré-processamento apresentou melhor resposta para a construção do modelo de calibração. Dentre os pré-processamentos analisados, *Constant Offset Elimination* foi o que apresentou maior capacidade preditiva e robustez para o modelo PLS para a quantificação de compostos fenólicos.

Palavras-chave: Antioxidantes. *Apis mellifera*. Calibração multivariada. Eliminação de deslocamento constante. Quimiometria.

1. INTRODUÇÃO

A própolis é um material resinoso natural coletado por abelhas *Apis mellifera* a partir de diversas fontes vegetais. É conhecida principalmente por apresentar diversas propriedades biológicas como: anticâncer, antimicrobiana, anti-inflamatória, antioxidante e antifúngica (LI et al., 2016).

A presença de compostos fenólicos totais (CFT) atua

diretamente como colaborador das propriedades farmacológicas, contribuindo principalmente na atividade antioxidante da própolis. Estas propriedades são comumente quantificadas por ensaios colorimétricos e cromatográficos, entretanto essas técnicas necessitam de pessoal treinado e apresentam custo elevado (WAGH, 2013; USMAN et al., 2016). Na busca por uma técnica que minimize os gastos com procedimentos analíticos, estudos recentes (YE

¹ matheus_augustocalegari@hotmail.com, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.

² anaclaraprasniewski@hotmail.com, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.

³ cleidi_dsv@hotmail.com, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.

⁴ tatieneoldoni@utfpr.edu.br, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.



et al., 2016; VIEGAS et al., 2016) mostram que a espectroscopia do infravermelho próximo (NIR, do inglês Near Infrared) associado à calibração multivariada vem sendo utilizado para quantificações analíticas.

Entre as técnicas multivariadas destaca-se a regressão por Mínimos Quadrados Parciais (PLS, do inglês "Partial Least Squares"), por apresentar como principal vantagem à alta capacidade para modelar regressões com múltiplas variáveis resposta e produz fatores com alto poder de predição (MORELLATO, 2010). Um fator muito importante que implica diretamente na capacidade preditiva dos modelos multivariados é a escolha do número de variáveis latentes (VL), o número de variáveis deve conter informações suficientes que permitam modelar o sistema de forma a não ser um modelo super ajustado (MORGANO, 2005).

O objetivo deste trabalho foi desenvolver modelos de calibração multivariada, baseado em espectros coletados no NIR combinado à ferramenta quimiométrica por PLS, para determinação de compostos fenólicos totais, verificando qual pré-tratamento apresentou melhor resposta para a construção do modelo de calibração.

2. MATERIAL E MÉTODOS

A determinação de compostos fenólicos totais foi realizada seguindo o método espectrofotométrico de Folin-Ciocalteu, sugerido por Singleton e Rossi (1965). Um total de 82 extratos hidroalcoólicos foram analisados em NIR da Bruker (Alemanha) modelo MPA, com leitura em absorvância com resolução de 8 cm^{-1} e 32 acumulações na região de comprimento de onda de $4000 \text{ a } 12000 \text{ cm}^{-1}$.

O modelo de calibração multivariada foi desenvolvido através do método estatístico de regressão por mínimos quadrados parciais (PLS) em software OPUS 7.2 quant2. As informações espectrais foram organizadas em uma matriz de dados e correlacionadas a um vetor que armazena dados da concentração de cada espectro. Para a minimização

de erros e consequente melhoria da capacidade preditiva do modelo, foram empregados pré-tratamentos como: 1^a derivada com suavização Savitzky-Golay de 17 janelas (1 D + SG), Constant offset elimination (COE), Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), escolha do número de variáveis latentes e remoção de outliers.

A validação do modelo foi realizada por dois métodos distintos: validação externa por grupo de teste (test set) e validação interna cruzada (cross validation). Segundo Li; Qu Haibin (2010) Root Mean Standard Error of Calibration (RMSEC), Root Mean Standard Error of Cross Validation (RMSECV), Root Mean Standard Error of Prediction (RMSEP), Ratio of Performance Deviation (RPD), Range Error Ratio (RER) e R^2 são parâmetros utilizados para avaliar os modelos de calibração do NIR, e ainda RMSEP/RMSECV é capaz de indicar a robustez do modelo, sendo a razão mais próximo de 1 mais robusto será o modelo.

3. RESULTADOS E DISCUSSÃO

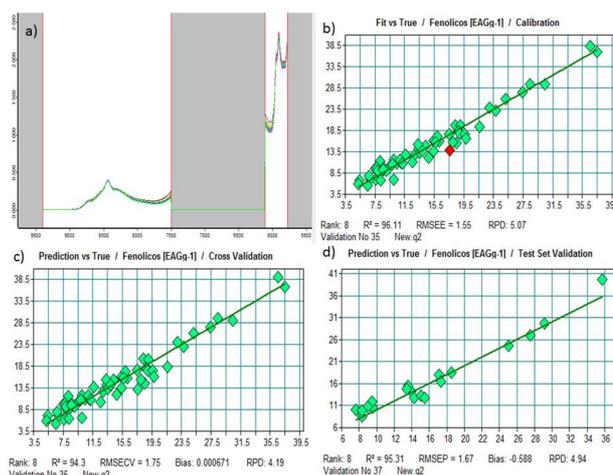


Figura 1 – Modelo 1: com pré-tratamento Constant Offset Elimination, a) região espectral utilizada $9400 - 7498,4 \text{ cm}^{-1}$ e $6102,1 - 5774,3 \text{ cm}^{-1}$, b) modelo de calibração, c) curva de validação cruzada, d) curva de validação externa

No modelo 1 (Figura 1) os dados foram tratados através do pré-processamento Constant Offset Elimination com 8 VL em região espectral específica de $9400 - 7498,4 \text{ cm}^{-1}$ e $6102,1 - 5774,3 \text{ cm}^{-1}$ (Figura 1a). Neste modelo 5 amostras foram identificadas



como outliers (ou seja, anômalas ao modelo), porém apenas 4 delas foram excluídas da curva de calibração e uma amostra (ponto em vermelho) realçada na figura 1b não foi retirada da curva pois a permanência dela na curva não resultou em melhoria na resposta do modelo, desta forma, para evitar um superajuste da curva a amostra foi mantida. As curvas de validação cruzada interna e externa são indicadas na figura 1c e 1d respectivamente, e não apresentaram outliers.

Para a construção do modelo 2 (Figura 2) os dados contaram com pré-processamento MSC com 6 VL em região espectral específica (Figura 2a). O conjunto utilizado para a curva de calibração é demonstrado na figura 2 b, a análise de variância indicou 3 amostras como anômalas que foram retiradas da curva desenvolvida, possibilitando a obtenção de um modelo com melhores respostas. As curvas de validação cruzada interna e validação externa são apresentadas na figura 2c e 2d respectivamente, essas curvas não apresentaram amostras anômalas.

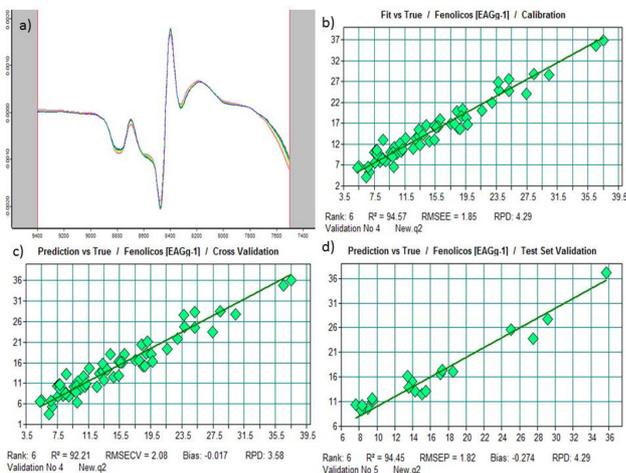


Figura 2 – Modelo 2: com pré-tratamento 1 D (SG) + MSC, a) região espectral utilizada 9400 - 7498,4 cm⁻¹, b) modelo de calibração, c) curva de validação cruzada, d) curva de validação externa

O modelo 3 (Figura 3) utilizou a região espectral de 9400 – 7498,4 cm⁻¹ (Figura 3a) e os dados contaram com o pré-processamento SNV com 9 VL. Para a construção da curva de calibração foram retiradas 3 amostras outliers, resultando em uma resposta mais confiável do modelo (Figura 3b). Para construção da curva de validação externa um outlier foi indicado

(ponto vermelho, Figura 3d), porém este ponto não foi retirado para que se evitasse um superajuste dos dados e consequente estimação errônea do modelo.

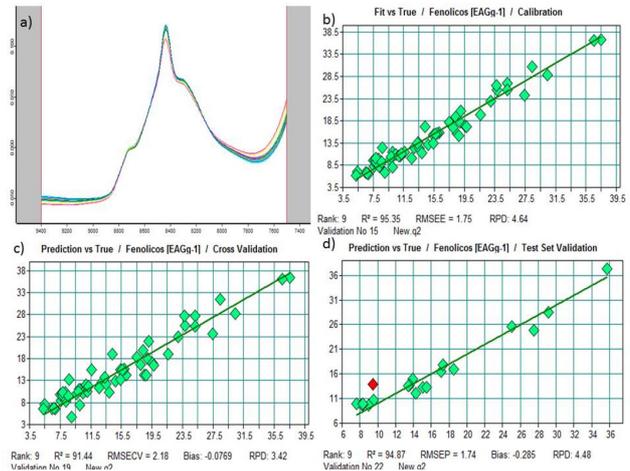


Figura 3 – Modelo 3: com pré-tratamento SNV, a) região espectral utilizada 9400 - 7498,4 cm⁻¹, b) modelo de calibração, c) curva de validação cruzada, d) curva de validação externa

A tabela 1 apresenta os parâmetros de validação dos modelos desenvolvidos.

Tabela 1 – Comparação de diferentes modelos de calibração com diferentes métodos de pré-tratamento espectral

Md	Pré-tratamento	Calibração			Validação Cruzada			Validação Externa		
		R ²	RP	D	R ²	RP	D	R ²	RP	D
1	Constant offset elimination	0,96	5,07	8	0,94	4,19	18,41	0,95	4,94	19,30
2	1D (SG) + MSC	0,95	4,29	6	0,92	3,58	15,49	0,94	4,29	17,70
3	SNV	0,95	4,64	9	0,91	3,42	14,78	0,95	4,48	18,52

Md.: modelo.

Todos os modelos gerados (Tabela 1) apresentaram valores de R² para calibração acima de 0,95 e para validação acima de 0,91, sendo que o maior valor de R² foi obtido no modelo 1, para estes mesmo modelo o valor de RPD também foi o mais elevado, próximo a 5. Segundo Dal Zotto et al. (2008) modelos que apresentam R² > 0,90 são considerados excelentes para quantificação e, valores de RPD acima de 2 indicam boas equações de calibração.

Os valores de erro para calibração e validação (RMSEC, RMSECV e RMSEP) foram 1,44; 1,75 e 1,67 respectivamente para o modelo 1, sendo este os menores valores do conjunto indicando curvas mais capazes de prever o modelo. Já para o modelo 2 e 3 os valores de erro para calibração e validação



(RMSEC, RMSECV e RMSEP) foram: 2,21; 2,08; 1,82 e 2,34; 2,18; 1,74 respectivamente.

Segundo Páscoa et al. (2013) valores acima de 10 para RER são indicativos de que o modelo apresenta uma boa estimativa. Para as curvas desenvolvidas todos os valores estão acima do indicado, destacando-se os valores de 18,41 e 19,30, para validação cruzada e externa do modelo 1 respectivamente, o que indica uma melhor resposta.

A razão RMSEP/RMSECV para o modelo 1 apresentou valor de 0,95; para o modelo 2 de 0,87 e para o modelo 3 de 0,79; quanto mais próximo a 1 for esta razão mais robusto será o modelo (PÁSCOA et al., 2013), desta forma os resultados indicam que o modelo 1, além de apresentar boa estimativa apresentou-se também como o mais robusto.

REFERÊNCIAS

DAL ZOTTO, Riccardo; DE MARCHI, Massimo; CECCHINATO, A; PENASA, Mauro; CASSANDRO, Martino; CARNIER, Paolo; GALLO, Luigi; BITTANTE, Giovanni. Reproducibility and repeatability of measures of milk coagulation properties and predictive ability of mid-infrared reflectance spectroscopy. **Journal of dairy science**, v. 91, n. 10, p. 4103–12, 2008.

LI, Aifeng; XUAN, Hongzhan; SUN, Ailing; LIU, Renmin; CUI, Jichun. Preparative separation of polyphenols from water-soluble fraction of Chinese propolis using macroporous absorptive resin coupled with preparative high performance liquid chromatography. **Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences**, v. 1012-1013, p. 42–49, 2016.

LI, Wenlong; QU, Haibin. Rapid quantification of phenolic acids in Radix Salvia Miltorrhiza extract solutions by FT-NIR spectroscopy in transreflective mode. **Journal of Pharmaceutical and Biomedical Analysis**, v. 52, n. 4, p. 425–431, 2010.

MORELLATO, Saulo Almdeida. **Modelos de regressão PLS com erros heteroscedásticos**. 2010. 49 f. Dissertação (Mestrado em Estatística) - UFSCar, São Carlos, 2010.

MORGANO, Marcelo Antonio. **Aplicação de métodos quimiométricos em análise de alimentos**. 2005. 109f. Tese (Doutorado em Química) – Unicamp, Campinas, 2005.

4. CONCLUSÕES

Através dos parâmetros de predição, este trabalho evidenciou que a utilização de pré-tratamentos espectrais distintos, seleção do número de variáveis latentes e utilização de região espectral específica na construção dos modelos, influenciam na obtenção de um modelo mais confiável para quantificação de compostos fenólicos em extrato etanólico de própolis. Neste estudo o pré-tratamento que possibilitou a construção de um modelo mais robusto e com uma capacidade preditiva maior que os demais foi o Constant Offset Elimination com 8 variáveis latentes (VL). Desta forma o modelo 1 tem capacidade de melhor prever o teor de compostos fenólicos totais em EEP que os demais.

AGRADECIMENTOS

Os autores gostariam de agradecer ao CNPq e à Central de Análises / UTFPR – Câmpus Pato Branco.

PÁSCOA, Ricardo; MAGALHÃES, Luís; LOPES, João. FT-NIR spectroscopy as a tool for valorization of spent coffee grounds: Application to assessment of antioxidant properties. **Food Research International**, v. 51, n. 2, p. 579–586, 2013.

SINGLETON, Vernon; ROSSI Joseph Jr. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. **American Journal of Enology and Viticulture**, v. 16, p. 144–58, 1965.

USMAN, Umar Zayyanu; BAKAR, Aainul Bahiyah Abu; MOHAMED, Mahaneem. Phytochemical screening and comparison of antioxidant activity of water and ethanol extract propolis from Malaysia. **International Journal of Pharmacy and Pharmaceutical Sciences**, v. 8, n. 5, p. 413–415, 2016.

VIEGAS, Thayná Rua; MATA, Ana; DUARTE, Marcia Maria Lima; LIMA, Kassio. Determination of quality attributes in wax jambu fruit using NIRS and PLS. **Food Chemistry**, v. 190, p. 1–4, 2016.

WAGH, Vijay. Propolis: A wonder bees product and its pharmacological potentials. **Advances in Pharmacological Sciences**, v. 2013, 2013.

YE, Mengqi; GAO, Zhenpeng; LI, Zhao; YUAN, Yahong; YUE, Tianli. Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy. **Food Chemistry**, v. 190, p. 701–708, 2016.



Influence of pre-processing and choice of variable latent in construction models PLS for measurement of total phenolic in propolis

Matheus Augusto Calegari⁵ Anaclara Prasniewski⁶ Cleidiane da Silva⁷
Tatiane Luiza Cadorin Oldoni⁸

17 maio 2017

Abstract – Propolis is a resinous substance produced by bees and its main function is to defend the hive because the bees use to seal small cracks, protecting it from micro-organisms and the weather, mummify invading and disinfecting bodies. Moreover, propolis has various pharmacological properties acting to promote human health as antioxidant capacity which acts by inhibiting the oxidation of macromolecules. The main compounds that are related to antioxidant are phenolic compounds, and are therefore commonly determined in propolis extracts, as a way of measuring the quality of the material. The aim of this study was to develop multivariate calibration models, based on analysis using near infrared (NIR) combined with chemometric tool for PLS, for determination of total phenolics, checking which preprocessing showed better response to build the calibration model. Among the pre-processing analyzed Constant Offset Elimination showed the greatest predictive ability and robustness to the PLS model for quantification of phenolic compounds.

Keywords: Antioxidants. *Apis mellifera*. Multivariate Calibration. Constant Offset Elimination. Chemometrics.

Correspondência:

Mauris Malesuada Malesuada Elit

Rua José Cardoso Bonifácio nº 77, Bairro: Cristo Rei, Pato Branco, Paraná, Brasil.

Recebido: 25/10/2016

Aprovado: 17/05/2017

Como citar: Calegari, Matheus Augusto, et al. Influência do pré-processamento e escolha das variáveis latentes na construção de modelos PLS para quantificação de fenólicos totais em própolis. **Syn. scy. UTFPR**, Pato Branco, v. 12, n. 1, p. 14–18. 2017. ISSN 2316-4689 (Eletrônico). Disponível em: <<https://periodicos.utfpr.edu.br/synscy>>. Acesso em: DD mmm. AAAA.

DOI: "registro apenas quando a revista for depositada no portal do PERI"



Direito autoral: Este artigo está licenciado sob os termos da Licença **Creative Commons** Atribuição 4.0 Internacional.

⁵ matheus_augustocalegari@hotmail.com, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.

⁶ anaclaraprasniewski@hotmail.com, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.

⁷ cleidi_dsv@hotmail.com, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.

⁸ tatiianeoldoni@utfpr.edu.br, Universidade Tecnológica Federal do Paraná-UTFPR, Câmpus Pato Branco, Pato Branco, Paraná, Brasil.