

PREDIÇÃO DA PRODUÇÃO DE ETANOL EM MILHO VIA FUSÃO DE DADOS ESPECTRAIS NIR E FÍSICO-QUÍMICOS

Maria Lúcia Ferreira Simeone^{1*}; Maria Cristina Dias Paes²; Paulo Evaristo de Oliveira Guimarães³; Cristiane de Carvalho Guimarães⁴; Lauro José Moreira Guimarães⁵; Roberto dos Santos Trindade⁶.

¹Embrapa Milho e Sorgo, Sete Lagoas – MG, <http://lattes.cnpq.br/3365739098963519>

²Embrapa Milho e Sorgo, Sete Lagoas – MG, <http://lattes.cnpq.br/9287991917094339>

³Embrapa Milho e Sorgo, Sete Lagoas – MG, <http://lattes.cnpq.br/5658791387526684>

⁴Embrapa Milho e Sorgo, Sete Lagoas – MG, <http://lattes.cnpq.br/9862614215080679>

⁵Embrapa Milho e Sorgo, Sete Lagoas – MG, <http://lattes.cnpq.br/3900005624888128>

⁶Embrapa Milho e Sorgo, Sete Lagoas – MG, <http://lattes.cnpq.br/6034333526540637>

PALAVRAS-CHAVE: NIR. Quimiometria. Bioenergia.

ÁREA TEMÁTICA: Tecnologia e produção

DOI: 10.47094/IIICOBRAP.2026/RE/9

INTRODUÇÃO

A crescente demanda por energia renovável impulsiona a busca por maior eficiência na produção de biocombustíveis a partir do milho. A conversão do grão em etanol depende do teor de amido e de sua interação com a matriz proteica, a qual afeta a sacarificação. No entanto, os métodos tradicionais de fermentação para avaliar esse rendimento são demorados, caros e destrutivos. Essa limitação dificulta a análise de grandes volumes de amostras em programas de melhoramento genético vegetal. Como solução rápida e sustentável, a espectroscopia no infravermelho próximo (NIR) vem ganhando grande destaque. A técnica NIR destaca-se por ser não destrutiva, econômica e capaz de monitorar a qualidade da biomassa com agilidade. Ela substitui, com vantagens, os processos analíticos convencionais de pré-tratamento e de fermentação simultânea. Dessa forma, o NIR consolida-se como uma alternativa viável para prever o potencial de bioetanol em larga escala.

A integração de dados espectrais e químicos, por meio da fusão de dados, representa um salto significativo para o setor. Ao combinar a riqueza de informações da espectroscopia com a precisão dos atributos físico-químicos, é possível maximizar o conteúdo informativo e mitigar ruídos espectrais e colinearidades redundantes.

OBJETIVO

Este estudo visa explorar essa sinergia por meio de algoritmos avançados de *machine learning*, como *Partial Least Squares* (PLS), *Support Vector Machine* (SVM) e *Random Forest* (RF), para quantificar rapidamente o rendimento de etanol no milho, superando as limitações dos métodos analíticos convencionais e fornecendo uma estrutura robusta para a fenotipagem moderna.

METODOLOGIA

Foram utilizados grãos de 36 híbridos de milho provenientes do programa de melhoramento da Embrapa. Os ensaios de campo foram conduzidos em duas repetições, durante a safra 2023–2024, em duas localidades distintas: Sete Lagoas – MG e Sinop – MT. O delineamento experimental resultou em 144 amostras. As amostras foram moídas em moinho de grãos com granulometria de 0,5 mm. Foram analisados os seguintes parâmetros: umidade: determinada gravimetricamente a 105 °C por 6 horas; amido total: utilizado o kit Megazyme (AOAC 996.11); amido extraível: determinado pelo método de Eckhoff et al., 1996; peso hectolitro (PH): conforme ISO 7971-3, padronizado para 13% de umidade; produção de etanol: o método de fermentação seguiu os protocolos de Paes et al., 2023. A análise do teor de etanol foi realizada por HPLC (Waters Alliance 2695) com coluna Aminex HPX-87H e detector de índice de refração (RID), utilizando ácido sulfúrico a 5 mmol·L⁻¹ como fase móvel. Os dados espectrais foram coletados em um espectrômetro Büchi NIRFlex N-500 em modo de refletância difusa, cobrindo a faixa de 10.000 a 4.000 cm⁻¹. O conjunto de dados foi dividido em treinamento (70%) e validação externa (30%). O pré-processamento incluiu a segunda derivada de Savitzky-Golay (SG) para redução de ruído. Na estratégia de fusão de dados de baixo nível (*low-level data fusion*), os sinais espectrais NIR e os constituintes físico-químicos foram concatenados, formando uma matriz de dados independente. Os modelos foram construídos utilizando, primeiramente, os dados espectrais NIR e os algoritmos PLS, SVM e RF. Em seguida, utilizou-se a estratégia de fusão de dados, com base no melhor resultado obtido na etapa anterior, a partir dos dados NIR. Por fim, a capacidade preditiva do modelo foi avaliada calculando-se o coeficiente de determinação (R^2) e o Erro Quadrático Médio de Validação Cruzada (RMSECV) e o Erro Quadrático Médio de Predição (RMSEP), utilizando o conjunto de validação externa independente, bem como os coeficientes de correlação entre os valores de referência e os previstos (R^2_c , R^2_{cv} e R^2_p). Valores mais baixos de RMSEP indicam maior capacidade preditiva dos modelos. Além disso, o Desvio Residual de Predição (RPD) e o Desempenho em Relação ao Intervalo Interquartil (RPIQ) foram utilizados para avaliar a consistência e a confiabilidade dos diferentes modelos preditivos.

RESULTADOS E DISCUSSÃO

Os constituintes e as propriedades físico-químicas apresentaram variabilidade suficiente para a calibração de modelos de alta precisão (Tabela 1).

Tabela 1: Estatística descritiva dos constituintes do milho e rendimento de etanol

Constituintes	Média	Mínimo	Máximo	DP
Umidade (%)	10,84	9,96	12,33	0,45
Proteína (%)	9,13	7,29	10,86	0,63
Fibra Bruta (%)	1,80	0,89	2,23	0,25
Extrato Etéreo (%)	4,00	3,35	5,06	0,44
Amido (%)	73,19	61,42	89,18	5,85
Amido Extraível (%)	56,26	51,52	60,73	1,90
Peso hectolítro (kg. hL ⁻¹)	77,05	68,20	82,80	2,45
Rendimento de Etanol (L.t ⁻¹)	430,65	374,67	458,27	13,65

DP: Desvio Padrão.

A Tabela 2 apresenta as estatísticas de desempenho para a comparação entre todos os modelos, utilizando os algoritmos PLS, SVR e RF, bem como a fusão dos dados de RF.

Tabela 2: Parâmetros estatísticos para a comparação entre os modelos PLS, SVR, RF com dados de espectroscopia NIR e RF (*low data-fusion* com dados NIR e físico-químicos).

Algoritmo	R ² _c	RMSEC	R ² _{cv}	RMSECV	R ² _p	RMSEP	RPD	RPIQ
PLS	0,96	2,35	0,81	5,05	0,65	6,48	1,7	2,8
SVR	0,99	0,68	0,61	8,14	0,71	5,34	1,9	2,9
RF	0,96	2,80	0,62	7,42	0,81	4,27	3,2	4,4
RF <i>low data fusion</i> *	0,94	0,50	0,65	1,20	0,65	1,19	3,0	3,4

Partial Least Squares (PLS) com 5 variáveis latentes; SVR: kernel RBF, $c = 0,0006662$, utilizando dados NIR pré-processados pela segunda derivada de Savitzky-Golay (SG); Random Forest (RF), com 500 árvores. Erro quadrático médio de calibração (RMSEC); Erro quadrático médio de validação cruzada (RMSECV); Erro quadrático médio de predição (RMSEP), Desvio Residual de Predição (RPD); Desempenho em Relação ao Intervalo Interquartil (RPIQ).

* Pré-processamento Standard Normal Variate (SNV).

A análise comparativa dos resultados mostra que o modelo RF (fusão de dados) apresentou R² mais elevado (0,94) e RMSEP mais baixo (1,19) do que os modelos PLS, SVR e RF. O valor de RPD foi de 3 e o de RPIQ foi de 3,4, conforme recomendado por Williams e Norris (2001), em que valores superiores a 3 e a 2 para RPD e RPIQ, respectivamente, indicam a robustez do modelo e sua capacidade preditiva quantitativa.

CONSIDERAÇÕES FINAIS

A fusão de dados espectrais NIR e físico-químicos, por meio do algoritmo RF, permitiu prever com alta precisão o rendimento de etanol no milho. Essa abordagem superou os modelos que consideravam apenas os dados espectrais no NIR, reduzindo os erros de predição em cerca de 70%.

Com excelente desempenho preditivo (RMSEP de 1,19, RPD de 3 e RPIQ de 3,4), o método substitui de forma ágil os demorados ensaios de fermentação. Dessa forma, a metodologia consolida-se como a escolha ideal para orientar decisões estratégicas de alta precisão no melhoramento genético.

REFERÊNCIAS

- ECKHOFF, S. R.; SINGH, S.; ZEHR, B. E. *et al.* A 100-g laboratory corn wet-milling procedure. *Cereal Chemistry*, Saint Paul, v. 73, n. 1, p. 54–57, 1996.
- WILLIAMS, P.; NORRIS, K. *Near-infrared technology: in the agricultural and food industries*. 2nd ed. St. Paul: American Association of Cereal Chemists (AACC), 2001.
- CARDOSO, W. S.; PAES, M. C. D.; DIAS, S. R.; CAMPOREZI, J. D.; MODOLO, M. U.; PINHEIRO, F. de A. *Metodologia científica: determinação do rendimento de produção de etanol de grãos de milho por via seca em escala laboratorial*. Sete Lagoas: Embrapa Milho e Sorgo. 2020, 17 p.

AGRADECIMENTOS

Os autores agradecem à Embrapa pelo financiamento do projeto SEG 20.23.01.011.00.07.012.