RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2025, **15**, 50045

Enhancing the robustness of a near-infrared (NIR) model for determining the blending proportion of cut tobacco by accounting for variations in moisture content

Jianhong Lai,[ab] Dongfu Xie,[a] Yufeng Luo,[a] Jinbang Wang,[*c] Qi Deng[*d] and Zongwen Yu[a]

Received 16th December 2024
Accepted 31st October 2025

DOI: 10.1039/d4ra08806a

rsc.li/rsc-advances

The blending proportion of cut tobacco significantly affects the intrinsic quality of the cigarette product. Variability in the moisture content of cut tobacco markedly influences its near-infrared (NIR) spectral signature and the accuracy of blending proportion prediction models. To address this critical challenge, spectral data were systematically collected from tobacco samples at various moisture levels. Four partial least squares regression (PLSR)-based correction methods were implemented to mitigate the moisture effect: global correction, orthogonal signal correction (OSC), generalized least squares weighting (GLSW), and dynamic orthogonal projection (DOP). The results indicate that moisture content strongly influences the diffuse transmission spectrum of cut tobacco. A model calibrated at a fixed 12.15% moisture content achieves satisfactory prediction accuracy under that specific condition but exhibits substantial errors when applied to samples with different moisture levels. This underscores the necessity of correcting for moisture effects to establish a more robust and generalizable blending proportion prediction model. Among the correction methods, DOP yielded the most promising performance, enhancing the coefficient of determination for prediction (R_p) from 0.39 to 0.90 and decreasing the root mean square error of prediction (RMSEP) from 5.50% to 2.22% compared to the uncorrected model. These findings have significant practical implications for advancing the application of blending proportion prediction in the tobacco industry.

1. Introduction

The blending proportion of cut tobacco is a critical process in tobacco manufacturing, significantly influencing the intrinsic quality and sensory characteristics of the final product.[1,2] Near-infrared spectroscopy (NIR) offers a fast, simple, and non-destructive method for inspecting the blending proportion of cut tobacco in the tobacco industry.[3] While previous studies have demonstrated the feasibility of NIR for this application, their practical utility in dynamic production environments remains limited. For instance, Hu[4] employed PLSR with NIR to estimate cut tobacco content in cigarette blends, but required a large calibration set to ensure model stability and accuracy. Similarly, Liu[5] utilized linear non-negative regression coefficient regression combined with NIR to determine the cut tobacco blending proportion, but the model was validated only under offline conditions for a single brand of tobacco. These limitations highlight the challenges of applying NIR models directly in real-world production scenarios. A major challenge lies in the high susceptibility of cut tobacco moisture content to environmental humidity, which can significantly alter the absorption bands and intensities associated with water's O–H groups.[6] When models developed under specific moisture conditions are applied to samples with different moisture levels, prediction accuracy often deteriorates.[7] Although temperature-compensation strategies have been explored for models of fruit soluble solids,[8] analogous work for tobacco moisture correction is scarce.

To address these challenges, this study focuses on investigating the effects of moisture content and developing robust correction methods for NIR models used in detecting internal quality indicators of cut tobacco. The ultimate goal is to establish models that maintain high accuracy, stability, and adaptability across varying moisture conditions. The main moisture content correction methods include global correction modeling,[9] exclusion of moisture-sensitive spectral bands,[10] spectral correction techniques,[11] formula correction methods[12] and multi-step modeling incorporating moisture content

[a]East China Jiaotong University, Nanchang, 330013, China
[b]China Tobacco Jiangxi Industrial Co., Ltd, Nanchang, 330006, China
[c]Zhengzhou Tobacco Research Institute of China National Tobacco Corporation, Zhengzhou, 450001, China. E-mail: wangjinbangok@126.com
[d]Hefei Design Institute of China National Tobacco Corporation, Hefei, 230041, China. E-mail: dq3493@sohu.com

information.[13] Global moisture correction modeling improves model robustness by incorporating moisture variations into the model. However, the accuracy of the model depends heavily on the size and representativeness of the calibration set, and the model's predictive power may be limited if the range of moisture variation is not adequately covered.[14] The method of excluding moisture-sensitive spectral bands uses algorithms such as simulated annealing,[15] successive projections algorithm[10] and genetic algorithm[16] to remove bands sensitive to moisture. While this helps reduce moisture interference, it may also remove important chemical information, thereby reducing model precision. Formula-based correction methods adjust for the effects of moisture through empirical formulas or by introducing intermediate variables related to moisture. However, indirect modeling using these variables can introduce secondary errors. Over-reliance on these variables can lead to cumulative errors and affect the model's final accuracy.[12] Multi-step modeling that incorporates moisture content as a key variable in the modeling process progressively corrects for moisture effects to improve model prediction accuracy. However, the use of intermediate variables in such models may introduce additional errors, especially when moisture is used as an intermediate variable, potentially leading to error propagation and inaccurate predictions.[13] In contrast, spectral correction techniques, such as OSC,[17] DOP,[18] and GLSW,[8] offer a more direct approach by transforming moisture-affected spectra to standard moisture conditions, effectively reducing the impact of moisture on model accuracy.

Based on this analysis, the study aims to achieve the following objectives: (1) to quantitatively elucidate the interference mechanism of moisture content on the NIR spectra of tobacco; (2) to construct and evaluate new models using updated methods such as Global correction, OSC, DOP, and GLSW; (3) to establish a robust and moisture-adaptive NIR model for accurately determining the blending proportion of cut tobacco.

2. Materials and methods

2.1 Materials

In this study, a specific cigarette brand produced in Jiangxi Province was selected as the research subject. To simulate the industrial blending process, a predetermined quantity of cut tobacco and tobacco stem was collected, homogenized, and then stored separately in sealed bags before being sent to the laboratory for further analysis. Considering that the typical proportion of tobacco stem blended in cigarette production is generally below 20%, the experiment established a 25-level blending matrix with tobacco stem proportions ranging from 1% to 25% by mass. For each blending level, 160 samples were prepared and subsequently divided into three portions: Portion A (20 samples), Portion B (120 samples), and Portion C (20 samples). Each sample weighed 10 grams, forming a comprehensive baseline blending matrix. To account for the moisture variation of cut tobacco in actual production, where the moisture content typically fluctuates around 12%, occasional extreme cases may see moisture content levels as low as 5% or as high as 18%. Samples from portions A, B, and C underwent three different moisture treatments: natural air-drying for a predetermined period, no treatment, and controlled moisture adjustment by spraying an appropriate amount of water. This was followed by a 24 hours equilibration in sealed bags at a constant temperature and humidity to ensure uniform moisture distribution. The actual moisture contents of the prepared samples, measured using the oven-drying method, were standardized to 5.08%, 12.15%, and 18.09%, respectively. Following preparation, all samples were placed in sealed bags to prevent any further moisture exchange.

For model development and validation, the following sample sets were designated: calibration set: 100 samples selected from the 12.15% moisture Portion B. Prediction set: 10 samples at each of the three moisture levels, equally selected from the corresponding moisture-treated portions. Update set: another 10 samples at each of the three moisture levels, equally selected from the corresponding moisture-treated portions, intended for model updating. All sample preparation and subsequent experiments were conducted in a controlled environment of constant temperature and humidity to mitigate the impact of tobacco's hygroscopic nature on the measurement results.

2.2 NIR spectral acquisition

NIR testing was conducted at the Jiangxi Province Agricultural Product Optoelectronic Testing Technology and Equipment Engineering Laboratory, using a Bruker Fourier Transform Infrared Spectrometer (TENSOR 37) from Germany. Measurements were taken in standard transmission mode, covering a spectral range of 10 000 to 4000 cm^{-1} with a resolution of 4 cm^{-1}. Each sample was scanned for 16 seconds. Prior to spectral acquisition, the spectrometer underwent a 30 minutes preheating phase. The equilibrated samples were then placed into sample cups, compacted, and rapidly analyzed to collect spectra. The rotational feature of the sample cups was utilized to minimize the impact of sample inhomogeneity by collecting three spectra from different regions of the same sample. To prevent interference from external environmental factors, the bottom of the sample cup was firmly pressed against the spectrometer's measurement head. An average of the three spectra was calculated to obtain the final averaged spectrum.

2.3 Data processing and analysis

The spectral data from tobacco samples were imported into MATLAB and Unscrambler software to establish PLSR models for predicting the blending proportion of cut tobacco. The performance of these models was evaluated using several statistical indicators, including the coefficient of determination for cross-validation and prediction (R^2 and R_p), the root mean square error for cross-validation and prediction (RMSECV and RMSEP). Generally, a higher R_p, a lower RMSEP indicate superior predictive accuracy.

To mitigate the adverse effects of moisture content on the predictive accuracy of the model, this study employed several advanced spectral correction methods in conjunction with the PLSR algorithm.[19] These methods included Global correction, OSC, DOP, and GLSW (The framework of predicting the cut tobacco blending proportion in Fig. 1). (1) Global correction: this approach constructs a single, comprehensive calibration model by incorporating a representative subset of samples from all moisture content levels. The resulting model inherently accounts for variations caused by moisture and other external factors, allowing it to be directly applied to predict samples with different moisture contents.[20] (2) OSC: the OSC algorithm aims to standardize spectral data acquired under varying conditions by eliminating spectral variations that are not correlated with the target analyte concentration. It projects the spectral data into a subspace where the variations due to moisture content and other interferences are orthogonal to the predictive information, thereby enhancing the model's robustness and specificity.[21] (3) DOP: unlike methods that require a set of standard reference samples, DOP enables model transfer without them, requiring only partial reference measurements under new conditions. It projects the spectral data into a low-dimensional subspace, effectively separating the variations caused by instrument differences and moisture content effects from the predictive information.[22] This approach, derived from External Parameter Orthogonalization, allows a recalibrated model to be directly applied to new samples without the need for a complete re-calibration.[23] (4) GLSW: the GLSW algorithm constructs a weighting matrix by analyzing differences in the spectral data (X matrix) corresponding to similar concentration values (y matrix). This matrix is then utilized to weight the spectral data, effectively eliminating systematic variations in the X matrix caused by external variables such as moisture content, thereby enhancing the model's predictive accuracy.[24]

3. Results and discussion

3.1 Tobacco spectra at different moisture contents

To investigate the effect of moisture on the spectra, we collected tobacco samples with varying moisture levels and analyzed their NIR spectra. Fig. 2 illustrates the raw averaged spectra (A) and the first derivative transformed averaged spectra (B) for various moisture contents. In panel A, the most pronounced features are the combination absorption band of water's O–H group at 1448 nm and its overtone band at 1942 nm. Moreover, combination bands linked to the C–H group, appearing at 2222–2381 nm and their overtones at 1667–1786 nm, are also distinctly visible across different components.[7,9] The application of the first derivative transformation, as shown in panel B, effectively corrects baseline drift. However, this technique does not eliminate spectral deviations caused by variations in moisture content. The most significant differences in the sample spectra are mainly seen as variations in the absorption intensity at the 1942 nm band. Although the spectral shape and trend of the same tobacco largely remain consistent across different moisture contents, the spectral intensity varies considerably. This suggests that the NIR spectra of the collected tobacco samples contain information not only related to their intrinsic components but also to their moisture content, resulting in corresponding spectral variations when the moisture content changes.

3.2 Effect of moisture content on the prediction model

A PLSR calibration model was developed using samples with a moisture content of 12.15%. The K–S algorithm classified the 2500 experimental spectra into calibration and validation sets, comprising 1750 and 750 spectra, respectively. This algorithm was selected because it maximizes the Euclidean distances among multivariate spectra, resulting in a calibration set that spans the moisture and blending proportion space more evenly than random or duplex selection methods, thereby enhancing the model's extrapolation reliability. To prevent overfitting or underfitting, the optimal number of latent variables (LVs) was determined by optimizing the model within the range of 1 to 20. Specifically, the RMSECV was evaluated across this range of 1 to 20 latent variables, and ultimately, the value that minimized RMSECV without causing overfitting was selected. As shown in Fig. 3(A), the model developed at 12.15% moisture content exhibited strong predictive performance, with a high coefficient of determination (R^2) of 0.94 for the calibration set and a R_p of 0.91. The RMSEC and RMSEP were both 2.21%. The results

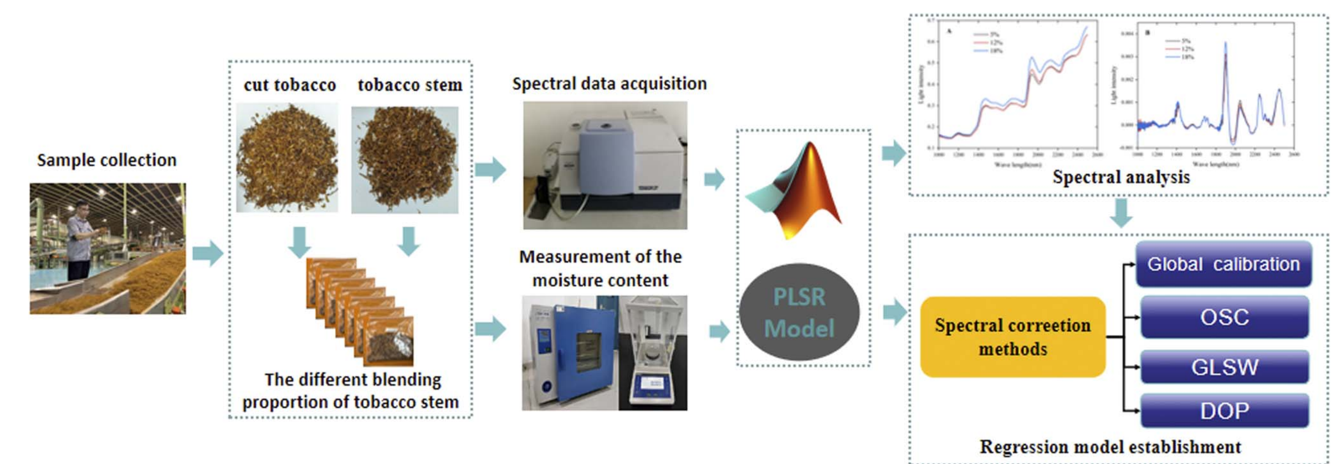

Fig. 1 The framework of predicting the cut tobacco blending proportion.

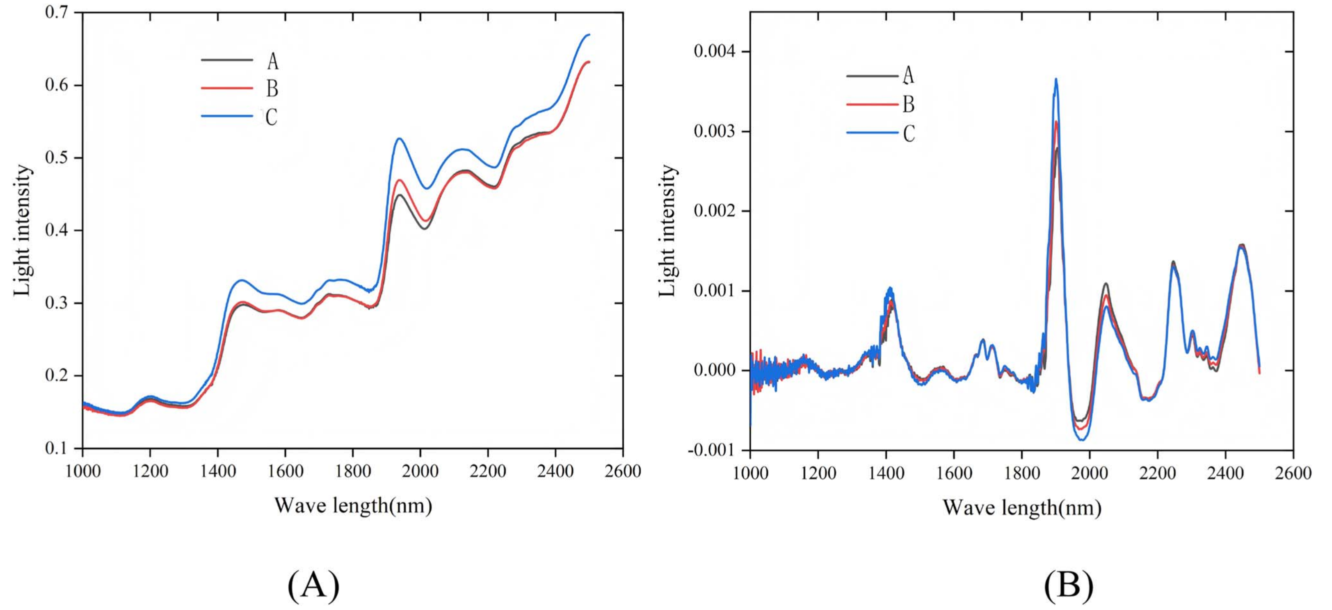

Fig. 2 (A) Averaged absorption spectra for the three moisture levels. (B) 1D spectra for the three moisture levels. In the A, B, and C represent tobacco moisture levels of 5.08%, 12.15%, and 18.09%.

indicate that the model established under a constant moisture condition of 12.15% can achieve accurate and reliable predictions. However, when applied to mixed moisture content samples (Fig. 3(B)), the model's performance significantly declines. The R_p for the validation samples' blending ratio prediction was 0.39, and the RMSEP was 5.50%. Suggesting that the model's prediction error at 12.15% increased and the accuracy of the validation samples decreased. This is primarily due to the significant impact of moisture content on the NIR spectra of the samples, which results in a decline in the predictive performance of the single moisture content model at 12.15% for mixed moisture content samples.

To further understand the spectral variations in blend samples with different moisture contents, principal component space plots were generated. These plots depict the blend samples from both the 12.15% moisture modeling set and prediction set (Fig. 3(C)), as well as the blend samples from the modeling set and a new prediction set after the addition of three moisture gradients (Fig. 3(D)). Fig. 3(C) demonstrates that the modeling set sufficiently covers the prediction set, indicating strong predictive performance. Conversely, Fig. 3(D) illustrates that the samples within the modeling set do not adequately encompass the compositions of the new prediction set after introducing varying moisture contents, leading to diminished model performance. This finding suggests that while the predictive model is effective for blend samples under consistent moisture conditions, its performance declines for samples with differing moisture levels. Thus, it is necessary to explore relevant spectral correction methods to mitigate the impact of moisture on the predictive model for cut tobacco blend prediction.

3.3 Model correction

3.3.1 Global correction. The global correction method utilizes sample spectra with varying moisture content to establish a quantitative analysis model. Samples from the update set were used as standard samples and were incorporated in different proportions into either the original calibration set or the calibrated spectra. The final number of samples introduced into the modeling set was determined by monitoring changes in the RMSEP value. As shown in Fig. 4(A) and (B), the inclusion of different quantities of update set samples generally led to enhancements in prediction outcomes. When all update set samples were included, a favorable prediction result was achieved (R_p = 0.90, RMSEP = 2.18%), due to the improved coverage of the prediction set by the new modeling set. Interestingly, incorporating just 20% of the update samples yielded an optimal prediction (R_p = 0.89, RMSEP = 2.31%), with only minor improvements observed in R_p and RMSEP upon adding more samples. This outcome is consistent with Acharya's findings.[21] Although this method is straightforward and somewhat alleviates the impact of moisture on model predictions, its accuracy significantly relies on the quantity and representativeness of the calibration set data. The proportion of new samples added remains uncertain, and a considerable amount of redundant information persists.[9] Therefore, further research is required to explore alternative methods that can effectively eliminate the influence of moisture on the prediction model.

3.3.2 Generalized least squares weighting. The GLSW algorithm involves the adjustment of the weighting parameter α, which determines the extent of the weighting effect. The parameter typically ranges from 0.0001 to 1, with a smaller α value corresponding to a stronger filtering effect. The optimal parameter combination was determined to be $\alpha = 1$ and LV = 5.

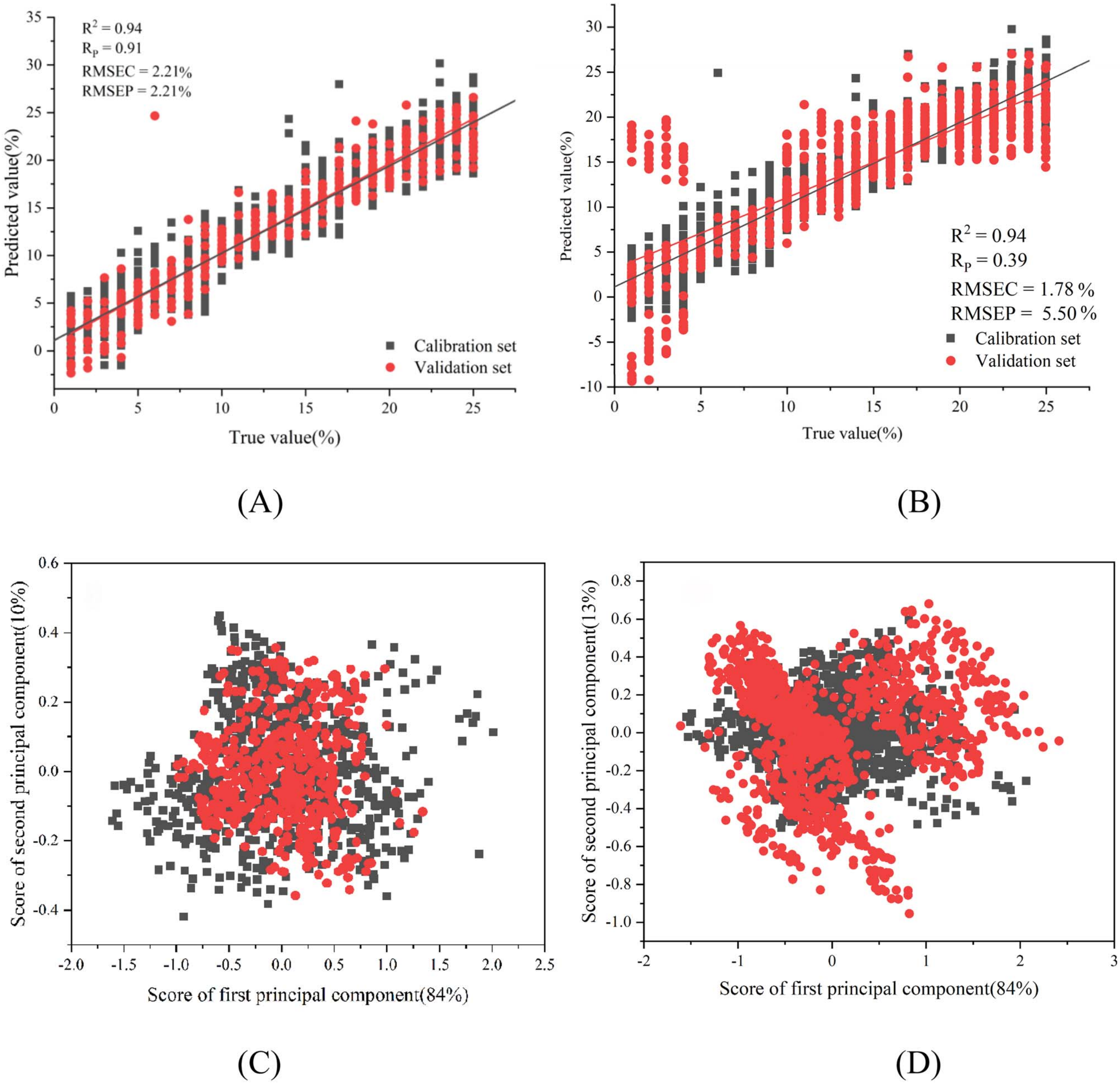

Fig. 3 (A) Scatter plot of single moisture content model prediction results at 12.15%. (B) Scatter plot of blending proportion predictions for mixed moisture content samples from a single moisture content model at 12.15%. (C) The score plot of the principal component analysis (PCA) for absorption (12.15% moisture modeling and prediction sets). (D) The score plot of the principal component analysis (PCA) for absorption (modeling set with added moisture gradients and a new prediction set). Note: In C and D, black denotes the modeling set, while red indicates the prediction set.

The prediction results are presented in Fig. 4(C)(D), where the R_p increased to 0.40, and the RMSEP decreased to 5.4%. This finding is inconsistent with the study by Liu *et al.*[25] suggesting that the advantages observed in handling moisture variations in granular soils do not directly translate to the case of tobacco.

3.3.3 Dynamic orthogonal projection. The objective of the DOP algorithm is to derive a spectral matrix that is independent of moisture content using the difference matrix (D). The parameter optimization process is analogous to that of the GLSW algorithm. As depicted in Fig. 4(E) and (F), the optimal combination of parameters was a g value of 20 and an LV value of 5. The prediction outcomes are illustrated in Fig. 4(F), where the R_p improved to 0.90, and the RMSEP was reduced to 2.22%.

3.3.4 Orthogonal signal correction. The OSC algorithm is used to eliminate spectral variations in the spectral matrix that are uncorrelated with the target analyte and the blending proportion, thereby removing irrelevant information that could affect the prediction. The outcomes of predicting the blending proportion of validation set samples using the OSC model are depicted in Fig. 4(G). The application of the OSC algorithm

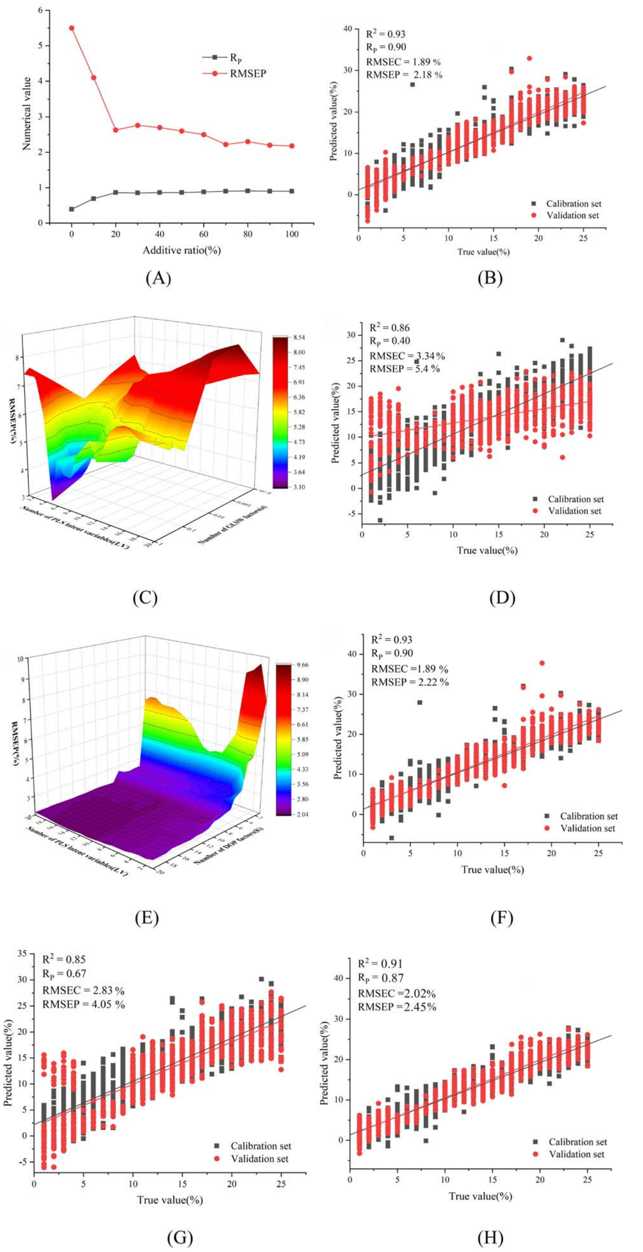

Fig. 4 (A) The relationship between RMSEP and RP values in global correction and the proportion of new samples added to the modeling set. (B) Scatter plot of global correction prediction results. (C) Prediction results of the GLSW correction model, with variation in RMSEP based on the D1 difference matrix. (D) Scatter plot of GLSW correction model prediction results. (E) Prediction results of the DOP correction model, with variation in RMSEP based on the D1 difference matrix. (F) Scatter plot of DOP correction model prediction results. (G) Scatter plot of predicted results from the OSC correction model. (H) External validation scatter plot of blending proportion predictions for cut tobacco.

significantly enhanced the model's predictive accuracy, increasing the R_p to 0.67, decreasing the RMSEP to 4.05% for the validation set samples. It is clear that the OSC model exhibits enhanced performance compared to both the original spectral model and the global correction model.

3.4 External validation

To address concerns regarding external validation, we conducted an additional experiment using a new batch of samples collected three months later from the same production line but from different tobacco lots. These samples were prepared following the same protocol and measured on a different Bruker TENSOR 37 unit located at a different laboratory. A total of 90 independent samples. The previously established DOP-corrected PLSR model was directly applied to predict the blending proportions of cut tobacco for these new samples, without any re-calibration. As depicted in Fig. 4(H),The external validation yielded: $R_p = 0.87$, RMSEP = 2.45%. These results are consistent with the internal test set performance ($R_p = 0.90$, RMSEP = 2.22%), demonstrating that the DOP-corrected model generalizes well across different batches, time points, and instruments. Furthermore, we performed a *t*-test between predicted and reference values, and no significant bias was found ($p > 0.05$), confirming the accuracy and robustness of the model under external conditions. Future research could investigate the use of larger validation sets to further substantiate the model's robustness across a wider range of conditions.

3.5 Model comparison

The prediction results for the single moisture content model, as well as the Global, OSC, GLSW, and DOP correction models, are summarized in Table 1. These results indicate that all moisture content correction methods enhance the model's predictive capability compared to the uncorrected model, with the corrected models exhibiting reduced prediction errors and improved adaptability to varying moisture conditions. Among the correction approaches, the DOP method achieved the highest performance, yielding a validation set R_p of 0.90 and an RMSEP of 2.22%. This underscores the superior moisture content correction ability of the DOP method relative to the Global, OSC, and GLSW methods.

The key advantage of the DOP method lies in its ability to eliminate non-target variations in spectral modeling caused by physical factors, chemical properties, and environmental interferences. Once applied, this method requires only a small set of reference measurements under new physical, chemical, or environmental conditions to recalibrate the model, without the need for extensive additional standard measurements. This flexibility makes the DOP method particularly valuable in practical applications, especially when dealing with variable factors such as moisture, temperature, external lighting conditions, or changes in sample state. The key advantage of the DOP method lies in its ability to eliminate non-target variations in spectral modeling caused by physical factors, chemical properties, and environmental interferences. Once applied, this method requires only a small set of reference measurements under new physical, chemical, or environmental conditions to recalibrate the model, without the need for extensive additional standard measurements. This flexibility makes the DOP method particularly valuable in practical applications, especially when dealing with variable factors such as moisture, temperature, external lighting conditions, or changes in sample

Table 1 Prediction results of the PLSR model for cut tobacco blending proportion using different moisture content correction methods

		Calibration		Prediction	
Method	Parameters	R^2	RMSECV (%)	R_p	RMSEP (%)
Single moisture content	LVs = 9	0.94	1.78	0.39	5.50
Global	LVs = 8	0.93	1.89	0.90	2.18
OSC	LVs = 6	0.85	2.83	0.67	4.05
GLSW	LVs = 5, $\alpha = 1$	0.86	3.34	0.40	5.40
DOP	LVs = 5, $K = 20$	0.93	1.89	0.90	2.22

state. Therefore, future studies could explore the application of the DOP method to recalibrate new datasets for other critical tobacco quality indicators under these varying conditions, thereby further enhancing the robustness and applicability of the models. The validation scope can also be extended to multiple brands and different processing environments to further evaluate the universality of the calibration method.

4. Conclusions

In this study, we established a blending proportion prediction model for cut tobacco and analyzed the impact of moisture content on both the tobacco spectrum and the model's accuracy. The results revealed that moisture content significantly influences tobacco spectra, leading to decreased model prediction accuracy and limited model applicability across varying moisture conditions. To enhance the robustness of the PLSR model against variations in sample moisture content, we implemented moisture content correction methods, including DOP, GLSW, OSC, and global correction. Among these, the DOP method yielded the best outcomes, with the model achieving an RMSEP of 2.22% and an R_p of 0.90. The effectiveness of the DOP method in correcting for moisture content is evident. These findings have practical implications for the application of NIR in predicting tobacco quality in industrial settings.

Author contributions

Jianhong Lai conducted all the experiments, analyzed the data, and prepared the graphs. Jinbang Wang and Zongwen Yu performed the chemometric analysis. Dongfu Xie developed the idea and research methodology. Yufeng Luo provided logistical support, and Qi Deng provided guidance and assistance for the revision of the paper. All authors have read and approved the manuscript for publication.

Conflicts of interest

The authors declare that they have no conflicts of interest. The authors are solely responsible for the content and the writing of this manuscript.

Data availability

All of the data generated in this study is presented in this manuscript.

Acknowledgements

We thank China Tobacco Corporation for financial support through the project "Research on key core technology requirements and foresight in the tobacco industry" (No. 11020210248).

References

1 R. L. Li, Y. Y. Liu, *et al.*, Study on rapid determination of tobacco blending uniformity by Near Infrared Spectroscopy, *Food Mach.*, 2019, **35**(5), 83–87.
2 K. Zhao, S. Chen, *et al.*, Determination of blending uniformity of cut tobacco based on near-infrared spectroscopy, *Food Mach.*, 2020, **36**(11), 183–188.
3 Y. Qin, H. Gong, *et al.*, NIR models for predicting total sugar in tobacco for samples with different physical states, *Infrared Phys. Technol.*, 2016, **77**, 239–243.
4 L. Z. Hu, *et al.*, Estimation of cut stem and reconstituted tobacco content in different cigarette blends by uniform design-PLS-NIR, *Acta Tab. Sin.*, 2010, **16**(2), 26–30.
5 J. Liu, *et al.*, Detection of mixing proportion of stem by near infrared spectroscopy combined with linear non-negative regression coefficient regression method, *Food Mach.*, 2021, **37**(11), 188–192.
6 P. Büning, *et al.*, Analysis of water in food by near infrared spectroscopy, *Food Chem.*, 2003, **82**(1), 107–115.
7 J. Luypaert, D. L. Massart, Y. Vander Heyden, *et al.*, Near-infrared spectroscopy applications in pharmaceutical analysis, *Talanta*, 2007, **72**(3), 865–883.
8 X. Jiang, J. Yao, M. Zhu, B. Li, Y. Liu, *et al.*, Achieving robustness to temperature change of a NIR model for apple soluble solids content, *Food Qual. Saf.*, 2023, **7**, 1–7.
9 M. Clavaud, Y. Roggo, K. Dégardin, P. Y. Sacré, P. Hubert and E. Ziemons, Global regression model for moisture content determination using near-infrared spectroscopy, *Eur. J. Pharm. Biopharm.*, 2017, **119**, 343–352.
10 X. Peng, T. Shi, Y. Chen, A. Song and W. Gao, Estimating Soil Organic Carbon Using VIS/NIR Spectroscopy with SVMR and SPA Methods, *Remote Sens*, 2014, **6**, 2699–2717.

11 F. D. Barboza and R. J. Poppi, Determination of alcohol content in beverages using short-wave near-infrared spectroscopy and temperature correction by transfer calibration procedures, *Anal. Bioanal. Chem.*, 2003, **377**(4), 695–701.
12 X. Sun, D. Deng, J. Liu, *et al.*, Model development and update of portable NIRS instrument for assessment of internal quality attributes of two navel orange varieties, *Front. Nutr.*, 2022, **9**, 976178.
13 J. M. Roger, F. Chauchard and V. Bellon-Maurel, EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits, *Chemom. Intell. Lab. Syst.*, 2003, **66**(2), 191–204.
14 N. Y. Ji, M. Li, W. B. Lü, R. Liu, Y. Y. Zhang and D. H. Han, Study on NIR model transfer between similar kinds of fruits based on slope/bias algorithm, *Guang pu*, 2017, **37**(1), 227–231.
15 Y. F. Bao, X. F. Wang, G. L. Liu, *et al.*, NIR detection of alcohol content based on GA-PLS, *Appl. Mech. Mater.*, 2012, **128**, 200–204.
16 E. Bouveresse and D. L. Massart, Standardisation of near-infrared spectrometric instruments: a review, *Vib. Spectrosc.*, 1996, **11**(1), 3–15.
17 Q. Fu, H. Suo, X. He, *et al.*, Transfer calibration for alcohol determination using temperature-induced shortwave near infrared spectra, *Spectrosc. Spectral Anal.*, 2012, **32**(8), 2080–2084.
18 L. Pan, H. Li, Z. Hu, M. Zhang and J. Zhao, Standard-free sample model transfer of visible/near-infrared spectral model of apple ripeness under seasonal variation, *J. Food Compos. Anal.*, 2024, **128**, 106028.
19 G. Mateos-Aparicio, Partial least squares (PLS) methods: Origins, evolution, and application to social sciences, *Commun. Stat. - Theory Methods*, 2011, **40**(13), 2305–2317.
20 C. P. S. Kuda-Malwathumullage and G. W. Small, Determination of temperatures of polyamide 66 directly from near-infrared spectra, *J. Appl. Polym. Sci.*, 2014, **131**(13), 378–387.
21 U. K. Acharya, K. B. Walsh, P. P. Subedi, *et al.*, Robustness of partial least-squares models to change in sample temperature: I. A comparison of methods for sucrose in aqueous solution, *J. Near Infrared Spectrosc.*, 2014, **22**(4), 279–286.
22 F. Z. Bai, W. Zheng, H. Luo, Y. H. Li, X. Yang and B. H. Zhang, Model robustness in estimation of blueberry SSC using NIRS, *Comput. Electron. Agric.*, 2022, **198**, 107073.
23 J. A. Diaz-Olivares, S. Grauwels, X. Fu, *et al.*, Temperature correction of near-infrared spectra of raw milk, *Chemom. Intell. Lab. Syst.*, 2024, **255**, 105251.
24 K. Haroon, A. Arafeh, T. Rodgers, F. Mendoza, M. Baker and P. Martin, Evaluation of temperature compensation methods for a near-infrared calibration to predict the viscosity of micellar liquids, *J. Chemom.*, 2020, **34**(11), e3301.
25 Y. Liu, Q. Jiang, T. Shi, T. Fei, J. Wang, G. Liu and Y. Chen, Prediction of total nitrogen in cropland soil at different levels of soil moisture with Vis/NIR spectroscopy, *Acta Agric. Scand. - B Soil Plant Sci.*, 2014, **64**, 267–281.