chemosensors

MDPI

Article

Quantitative Analysis of Four Active Components in Wine-Processed *Scutellaria baicalensis* by Data Fusion Strategy Coupled with Multivariate Statistical Analysis

Shengsheng Wang [1], Xulong Peng [1], Tian Yao [1], Qingzhao Wan [1], Tulin Lu [2], Jing Zhu [1] and Yaqi Wang [1,*]

[1] State Key Laboratory for the Modernization of Classical and Famous Prescriptions of Chinese Medicine, Jiangxi University of Chinese Medicine, Nanchang 330004, China
[2] School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China
* Correspondence: wangyaqi_3@163.com

Abstract

Wine-processed *Scutellaria baicalensis* is a processed Chinese herbal product obtained by wine-processing of raw *S. baicalensis*. Flavonoids are its core pharmacodynamic constituents with heat-clearing, anti-inflammatory, hepatoprotective and other bioactivities. High-performance liquid chromatography (HPLC) remains the reference for their quantification but is slow and solvent-intensive. This study proposes a chromatography-free alternative that fuses near-infrared (NIR) and ultraviolet (UV) spectra through low- and mid-level data fusion, coupled with interval partial least squares (iPLS) variable screening, to quantify the four flavonoids simultaneously across five processing degrees (raw plus four stir-frying durations, $n = 90$). HPLC values served as reference and the set was split by the Kennard–Stone algorithm. Against single-spectroscopy models, data fusion improved prediction for all four analytes: low-level fusion benefited the aglycones baicalein and wogonin (e.g., wogonin: $Rc^2 = 0.973$, $Rp^2 = 0.979$), whereas mid-level fusion raised the glycosides baicalin and wogonoside to satisfactory accuracy. Across all analytes, calibration $R^2 > 0.93$, prediction $R^2 > 0.90$, and RPD > 2.5. The method supports rapid, non-destructive evaluation of wine-processed *S. baicalensis* and is extensible to other processed herbal medicines.

Keywords: wine-processed *Scutellaria baicalensis*; near-infrared spectroscopy; processing degree; spectral analysis; quantitative model; data fusion

1. Introduction

Scutellaria baicalensis Georgi is the dried root of the Lamiaceae plant, which mainly contains flavonoid bioactive constituents including baicalin, wogonoside, and baicalein and wogonin [1]. Wine processing is its classical processing technique. Moistening with rice wine followed by thermal stir-frying can moderate the cold medicinal property of raw *S. baicalensis* and enhance its therapeutic efficacy [2,3]. Nevertheless, improper temperature control during processing may trigger degradation or structural transformation of active ingredients, further affecting the stability and therapeutic performance of clinical medication. HPLC serves as the mainstream quality-control approach and benchmark technique for separation and quantification of multiple constituents in TCM. It enables accurate multi-component determination relying on chromatographic separation. However, this technique suffers from large consumption of organic solvents, labor-intensive sample pretreatment and long assay duration for single sample. It is only suitable for offline batch testing and cannot satisfy the practical requirements for on-line dynamic quality control.

Check for updates

Received: 10 September 2026
Revised: 14 October 2026
Accepted: 14 October 2026
Published: 17 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Near-infrared (NIR) and ultraviolet (UV) spectroscopy feature fast measurement, non-destructive detection, and simple operation. Molecular characteristic information can be acquired without chromatographic separation, and they have been widely applied in multi-component quantification and processing degree evaluation of traditional Chinese medicines. NIR spectroscopy mainly reflects vibrational signals of hydrogen-containing groups such as C-H and O-H, and sensitively captures overall spectral variations of chemical constituents during processing. UV spectroscopy is highly sensitive to conjugated structures of flavonoids and collects characteristic absorption signals originating from electronic transitions. Although the information obtained from the two spectroscopic techniques is inherently complementary, most existing studies merely establish quantitative models targeting a single flavonoid, leading to limited analytical dimensions. In addition, single spectroscopy exhibits insufficient discrimination capacity against complex herbal matrices, which makes it difficult to fully reveal the intrinsic correlation between wine-processing duration and the overall content variation of the four flavonoids. Meanwhile, most relevant investigations focus on qualitative classification models for sample authenticity identification or processing grade discrimination, while analytical strategies for simultaneous accurate quantification of multiple active components remain scarce. In summary, for quality control of processed herbal medicines, there is still a lack of practical alternative methods for rapid multi-component quantification, which are available for online real-time processing monitoring and high-throughput rapid detection to compensate for the drawbacks of conventional liquid chromatography.

In this work, we use HPLC baseline-separated contents as a reference and construct a quantitative model that fuses NIR and UV spectra via low- and mid-level data fusion with interval partial least squares (iPLS) screening of informative variables [4]. Redundant spectral regions and matrix background are removed so that the complementary information of the two spectra is exploited for four structurally similar flavonoids simultaneously, without chromatography. We compare the predictive performance of single-spectrum, low-level and mid-level fusion models. The resulting method quantifies baicalin, wogonoside, baicalein and wogonin in wine-processed *S. baicalensis* across processing degrees (Figure 1), providing a basis for objective, near-real-time evaluation of wine-processed herbal medicines.

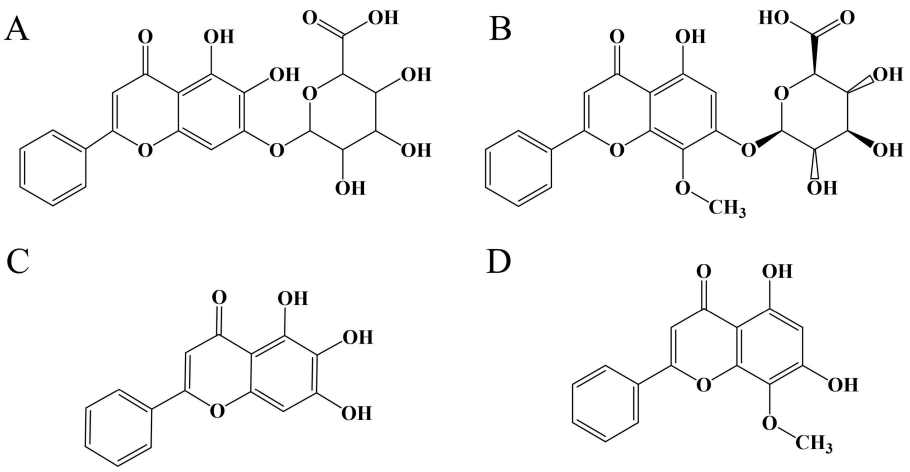


Figure 1. Structural formulas of the main chemical constituents in *S. baicalensis* ((**A**) baicalin; (**B**) wogonoside; (**C**) baicalein; (**D**) wogonin).

2. Material and Methods

2.1. Chemicals and Materials

S. baicalensis decoction pieces of different batches were purchased from multiple manufacturers to ensure sample representativeness; the raw medicinal materials originated from Liaoning, Hebei, and Inner Mongolia. Complete source information for all 90 samples, including batch data, geographical origin and processing groups, is provided in Table S1 of the Supplementary Materials. Reference standards for baicalin (Batch No. RFS-H01602105028, purity $\geq$ 98%), wogonoside (Batch No. RFS-H01911809030, purity $\geq$ 98%), baicalein (Batch No. RFS-H01811804026, purity $\geq$ 98%), and wogonin (Batch No. RFS-H02911809013, purity $\geq$ 98%) were purchased from Chengdu Rifensteidan Biotechnology Co., Ltd. (Chengdu, China). Chromatographic-grade methanol was purchased from TEDIA Company (Fairfield, OH, USA). Phosphoric acid (analytical grade) was purchased from Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China).

2.2. Instrumentation and Experimental Conditions

The instruments employed in this study were as follows: LYNX 4000 floor-standing refrigerated high-speed centrifuge (Thermo Fisher Scientific, Waltham, MA, USA); 1260 Infinity II high-performance liquid chromatography system (Agilent Technologies, Santa Clara, CA, USA); 103B high-speed herbal grinder (Yongli Pharmaceutical Machinery Co., Ltd., Taizhou, China); BT25S analytical balance (readout: 0.01 mg, Sartorius, Göttingen, Germany); Fourier-transform near-infrared (FT-NIR) spectrometer (Thermo Fisher Scientific, Waltham, MA, USA); KQ-500E ultrasonic cleaner (Kunshan Ultrasonic Instruments Co., Ltd., Kunshan, China); and UV-2550 ultraviolet–visible spectrophotometer (Shimadzu, Kyoto, Japan). MATLAB R2023a (MathWorks, Natick, MA, USA) running on a laptop equipped with an 11th Gen Intel Core i7-11800H processor, 16 GB RAM, and NVIDIA GeForce RTX 3050 Laptop GPU was used for data processing and chemometric modeling. All chemometric calculations were performed using MATLAB R2023a scripts.

Chromatographic analysis was carried out using the Agilent 1260 Infinity II HPLC system consisting of a quaternary solvent pump, autosampler, and variable wavelength detector (VWD). Separation was achieved on a Kinetex C18 column (250 mm $\times$ 4.6 mm i.d., 5 μm). The detection wavelength was set at 277 nm, and the column temperature was maintained at 30 °C; the injection volume was 10 μL. The mobile phase comprised methanol (A) and 0.2% aqueous phosphoric acid (B) with the following gradient elution program: 0–10 min, 30–40% A; 10–20 min, 40–50% A; 20–30 min, 50% A; 30–40 min, 50–60% A; 40–50 min, 60–70% A; 50–60 min, 70–30% A. The flow rate was 0.8 mL/min.

NIR spectra of each sample were acquired using a Fourier-transform near-infrared spectrometer equipped with an InGaAs detector. Spectra were recorded over the wavenumber range of 10,000–4000 cm^{-1} via integrating sphere diffuse reflectance mode at a resolution of 8 cm^{-1}. To improve the signal-to-noise ratio and measurement repeatability, each spectrum was accumulated over 64 scans [5]. The average spectrum from three replicate acquisitions was adopted for subsequent analysis.

UV spectra were recorded on a UV-2550 spectrophotometer (Shimadzu, Japan) within the wavelength range of 200–800 nm, with a sampling interval of 1 nm and a slit width of 2.0 nm. A purified water blank was scanned to mitigate baseline drift and spectral noise.

2.3. Optimized Final Method

Raw *S. baicalensis* decoction pieces were commercially available standard prepared decoction pieces that complied with the quality specifications stipulated in the Chinese Pharmacopoeia (2025 edition) [1]. No additional processing was required. The raw materials were pulverized and passed through a No. 4 sieve prior to use. Wine-processed

S. baicalensis decoction pieces with different processing degrees were prepared according to the wine processing protocol established in our previous study [2]. In brief, 10 kg of yellow rice wine was thoroughly blended with 100 kg of raw *S. baicalensis* decoction pieces. After moistening for 15 min, the mixtures were stir-fried at 120 °C, and samples were collected at 3, 6, 9, and 12 min during the frying process to obtain four different processing degrees. After cooling, each processed sample was pulverized, passed through a No. 4 sieve (65 mesh), and stored in dry, airtight containers for subsequent analysis.

For quantitative analysis, a mixed standard solution was prepared by accurately weighing appropriate amounts of baicalin, wogonoside, baicalein, and wogonin reference standards, which were then dissolved in methanol. The final concentrations of the four analytes were 48.1, 29.8, 100.8, and 11.1 μg/mL, respectively.

HPLC test solution was prepared using a modified procedure based on the literature method [3]. Briefly, 1 g of wine-processed *S. baicalensis* powder was accurately weighed and transferred into a conical flask. The sample was soaked in 25 mL of 70% ethanol for 15 min and ultrasonically extracted for 45 min. The extract was vacuum-filtered with a circulating water vacuum pump. The filtrate was diluted to 50 mL with 70% ethanol. A 2 mL aliquot of the above filtrate was accurately pipetted and diluted with methanol to a final volume of 10 mL. The diluted solution was filtered through a 0.22 μm polytetrafluoroethylene syringe filter, and a 10 μL aliquot of the filtrate was injected into the HPLC system for analysis.

For UV spectral measurement, 0.1 g of the prepared sample powder was accurately weighed and soaked in 50 mL of purified water for 15 min, followed by ultrasonic extraction for 30 min. The extract was centrifuged at 7000× *g* for 10 min. Afterwards, 1 mL of the supernatant was transferred into a 50 mL volumetric flask and diluted to the mark with purified water for subsequent UV analysis [6].

NIR spectra were acquired using an Antaris™ II Fourier-transform NIR spectrometer (Thermo Fisher Scientific, USA), equipped with an integrating sphere accessory and Result software (version 9.8.208). Approximately 4 g of wine-processed *S. baicalensis* powder (passed through a No. 4 sieve) was accurately weighed and placed in a dedicated sample cup, then gently pressed to ensure uniform thickness and a flat surface. Using the instrument's built-in background as the reference, spectra were collected in diffuse reflectance mode at ambient temperature of (25 ± 2) °C and relative humidity of (45 ± 2)%.

2.4. Spectral Data Pretreatment and Variable Screening

To develop a stable and reliable partial least squares regression (PLS) quantitative model for the simultaneous determination of multiple constituents in wine-processed *S. baicalensis*, the Kennard–Stone (KS) algorithm partitioned all 90 samples into calibration and independent validation sets [7]. Each independent partitioning maintained a 70:30 split ratio: 63 samples (70%) for the calibration set and 27 samples (30%) for the validation set. This strategy guaranteed comparable sample numbers for each wine-processing time within each individual partition. The detailed sample numbers of the four flavonoids at different wine-processing times under LLDF and MLDF strategies are summarized in Table S2. The core purpose of the KS algorithm is to select uniformly distributed and representative sample subsets from the whole dataset. The selected subsets cover not only the central region of the sample space but also include extreme boundary samples, so that the distribution profiles of the calibration set and independent validation set are consistent with those of the overall dataset. Samples in the validation set were not involved in any parameter tuning and were only used to evaluate the prediction accuracy of the final model. After partitioning by the KS algorithm, the calibration set and independent validation set exhibited similar distributions in spectral features and analyte contents. Both

the calibration and validation sets contained samples collected from Liaoning, Hebei, and Inner Mongolia.

The original spectra are often influenced by interference sources such as light scattering, random noise, background, and baseline drift. Therefore, various spectral preprocessing methods were utilized to reduce interference and improve the prediction accuracy of the model, including Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), Savitzky–Golay (SG) smoothing (7-point moving window, third-order polynomial), first derivative (1D), second derivative (2D), 1st derivative combined with SG smoothing (1st + SG), and second derivative combined with SG smoothing (2nd + SG) [8].

Full-spectrum data matrices usually contain hundreds or thousands of variables, including much ineffective or redundant information. The irrelevant variables not only increase model complexity but also may introduce interference, leading to prediction errors [6,9]. Furthermore, it is a widespread phenomenon that variables collected from modern spectroscopic instruments exhibit severe multicollinearity, such as those from NIR and UV spectroscopy, which also significantly impair the prediction performance of models. Therefore, in this study, iPLS was used to screen key variables, aiming to establish an accurate and reliable calibration model.

IPLS, based on PLS regression, divides the full spectrum into multiple sub-intervals of equal width and sequentially establishes local PLS models on each sub-interval [10]. The core advantage of this method is to intuitively compare the performance of the models built in different spectral intervals. Each sub-interval model is compared with the full-spectrum model according to the root mean square error (RMSE) obtained from the cross-validation set, and then the spectral regions containing the most informative features are selected.

In this work, standard non-sliding iPLS without moving step size was adopted. The preprocessed full spectrum was equally partitioned into four groups of non-overlapping equal-width sub-intervals (10, 20, 30, and 40 sub-intervals). The actual sub-interval width was automatically determined by the total spectral variables and the preset division number. The optimal division scheme and corresponding feature sub-intervals were determined by minimizing the root Mean Square Error of Cross-Validation (RMSECV) obtained from leave-one-out cross-validation. Only spectral variables falling within the selected optimal sub-intervals were retained as feature variables for subsequent modeling.

2.5. Construction, Optimization and Evaluation of Quantitative Models

To evaluate the independent predictive performance of different spectroscopic techniques, quantitative calibration models were established based on two individual single-spectrum datasets (NIR and UV spectra) to investigate the predictive capacity and inherent model characteristics of each spectroscopic technique when used alone. NIR spectroscopy records abundant overtone and combination absorptions originating from hydrogen-containing groups (e.g., O-H, C-H, N-H), which renders it suitable for rapid, non-destructive analysis of complex matrices. UV spectroscopy provides electronic transition information of constituents possessing conjugated structures or chromophores, and is widely adopted for the quantification of specific analytes. Comparative analysis of critical indicators including root mean square error of prediction (RMSEP) and R^2 enabled intuitive evaluation of the respective applicability and limitations of NIR and UV spectroscopy, laying a foundation for subsequent multi-spectral fusion modeling [11].

On this basis, multi-source data fusion strategies were adopted in this study. Data fusion strategies integrate information from multiple spectroscopic sources into a single model. Based on the level of data processing, data fusion can be categorized into LLDF and MLDF [9]. LLDF directly concatenates multiple original data sources into a new data matrix and then performs a unified preprocessing procedure on the matrix. MLDF, also

called feature-level fusion, extracts feature variables separately from each independent data source to compress data dimensionality and eliminate redundant information. Then, the feature vectors from each data source are integrated to form a new data matrix to explore synergistic interactions between data of different origins and provide fused features for subsequent modeling. This study selected the LLDF and MLDF strategies for analysis. All variables were uniformly scaled to the 0–1 interval using min-max normalization before fusion to ensure that data from different sources have consistent dimensions and are comparable (Figure 2).

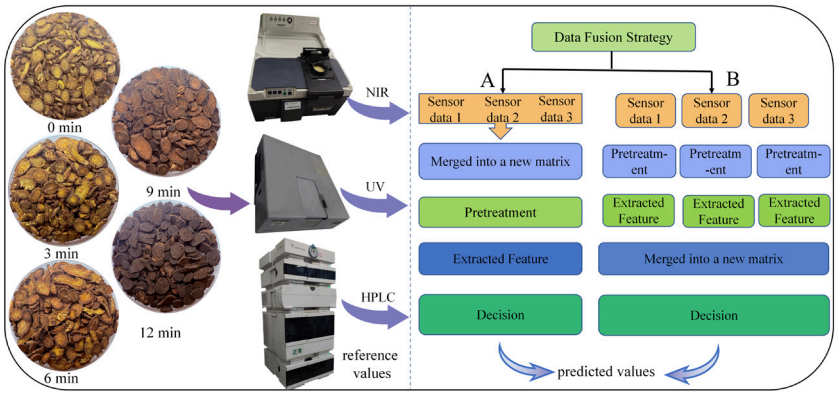


Figure 2. Schematic diagram of data fusion strategies for quality evaluation of wine-processed *S. baicalensis* ((**A**) low-level data fusion; (**B**) mid-level data fusion).

Subsequently, the optimal number of latent variables (LVs) was selected by LOOCV to optimize each model, minimizing RMSECV within a search range of 1–20; this criterion was applied uniformly to all PLS models. Model performance was comprehensively evaluated using multiple statistical metrics, including root mean square error of calibration (RMSEC), RMSECV, RMSEP, R_c^2 and R_p^2. To further examine the reliability and stability of the model on the independent validation set, additional metrics such as bias and the RPD were introduced for validation [12]. All the aforementioned data analyses were conducted in the MATLAB (R2023a) environment using self-developed programs.

3. Results and Discussion

3.1. HPLC Quantitative Analysis

An HPLC separation method was established to achieve favorable separation of four flavonoid bioactive constituents, namely baicalin, wogonoside, baicalein and wogonin, in wine-processed *S. baicalensis* [13]. Simultaneous multi-component quantitative detection was performed on raw *S. baicalensis* samples and wine-processed *S. baicalensis* samples prepared from commercial raw materials. Under the optimized chromatographic conditions, baseline separation of each component was achieved without obvious overlap or tailing of adjacent chromatographic peaks. The favorable separation performance guaranteed accurate peak area integration and yielded reliable content data, which served as reference values for subsequent spectroscopic modeling. All four flavonoids could be detected in different batches of raw *S. baicalensis* and prepared wine-processed samples, and the constituent contents showed regular variations with processing duration. The contents of baicalin and wogonoside gradually decreased with prolonged processing time. High temperature induced the hydrolysis of flavonoid glycosides, breaking glycosidic bonds to generate free aglycones (baicalein and wogonin), thereby increasing the contents of the two aglycones as processing time extended. The contents of the four flavonoids ranged from 0.05% to 16%, providing a sufficiently wide concentration gradient and rich sample variation. The broad content distribution fully covered the variation characteristics of raw

S. baicalensis and wine-processed *S. baicalensis* under different processing degrees, which effectively avoided overfitting caused by narrow concentration ranges. Therefore, the obtained dataset was reliable and representative, and could serve as high-quality reference values for the establishment and optimization of subsequent NIR and UV spectral fusion quantitative models.

3.2. NIR and UV Spectral Characteristics

The raw NIR spectra of the 90 wine-processed *S. baicalensis* samples are displayed in Figure 3A. NIR spectroscopy inherently suffers from broad bands and severe peak overlap. Moreover, the absorption signals at 8310, 6850, 5180 and 4750 cm^{-1} in raw spectra are easily masked by strong absorption from matrix components including moisture, starch, and cellulose, so functional group assignment cannot be directly performed on raw spectra.

In this study, MSC combined with second derivative transformation was adopted for spectral preprocessing. MSC eliminates multiplicative baseline shift induced by light scattering of powder particles, while the second derivative resolves overlapping broad absorption bands and amplifies weak vibrational signals of low-content flavonoids. After MSC (Figure 3B), multiplicative baseline drift from powder scattering is completely removed, and spectral curves of all samples converge at 8000 cm^{-1}, retaining only spectral differences derived from flavonoid components. Single second derivative preprocessing (Figure 3C) can split overlapping absorption peaks yet fails to eliminate residual scattering noise, resulting in obvious spiky artifacts at both spectral ends and low signal contrast [14]. Figure 3D shows NIR spectra preprocessed by MSC coupled with second derivative transformation; broad overlapping background absorption originating from moisture, starch and cellulose in the powder matrix is significantly suppressed, and weak functional group vibrational signals of four flavonoids (baicalin, wogonoside, baicalein and wogonin) are effectively amplified and well-resolved [15].

The oscillatory peaks near 8310 cm^{-1} are assigned to the second overtone stretching vibration of methyl C-H in the benzene-ring methoxy group (-OCH_3) of wogonoside and wogonin, serving as characteristic responses of methoxy-substituted flavonoids; baicalin and baicalein without methoxy groups exhibit no obvious characteristic oscillations in this region. The absorption signals at 6850 cm^{-1} arise from the coupling of the first overtone of methoxy methyl C-H and phenolic hydroxyl O-H, where only wogonoside and wogonin display sharp, high-intensity characteristic oscillations, enabling discrimination of monophenolic hydroxyl methoxy-substituted flavonoids. The characteristic peak cluster at 5180 cm^{-1} originates from the combination vibration of stretching and bending of multiple hydroxyl O-H on the glucuronic acid ring of baicalin and wogonoside, representing exclusive signals of flavonoid glycosides; baicalein and wogonin without sugar rings show no prominent oscillations here, which can distinguish flavonoid glycosides from their aglycones. The oscillatory absorption at 4750 cm^{-1} corresponds to the combination vibration of ortho-diphenolic Ar-O-H in baicalin and baicalein. Benefiting from more phenolic hydroxyl groups, these two compounds present remarkably stronger signal intensity than wogonoside and wogonin bearing only monophenolic hydroxyl groups. Minor oscillations at 4320 cm^{-1} are attributed to the stretching-bending combination vibration of CH/CH_2 C-H on the glycosidic sugar ring, which are faint characteristic signals unique to baicalin and wogonoside and assist in identifying flavonoid glycosides. Continuous terminal oscillations at 4020 cm^{-1} are shared by all four flavonoids, originating from the second combination vibration of conjugated C=O carbonyl and aromatic C-H on the flavone skeleton, and can characterize the overall flavonoid framework signals of samples [16–18].

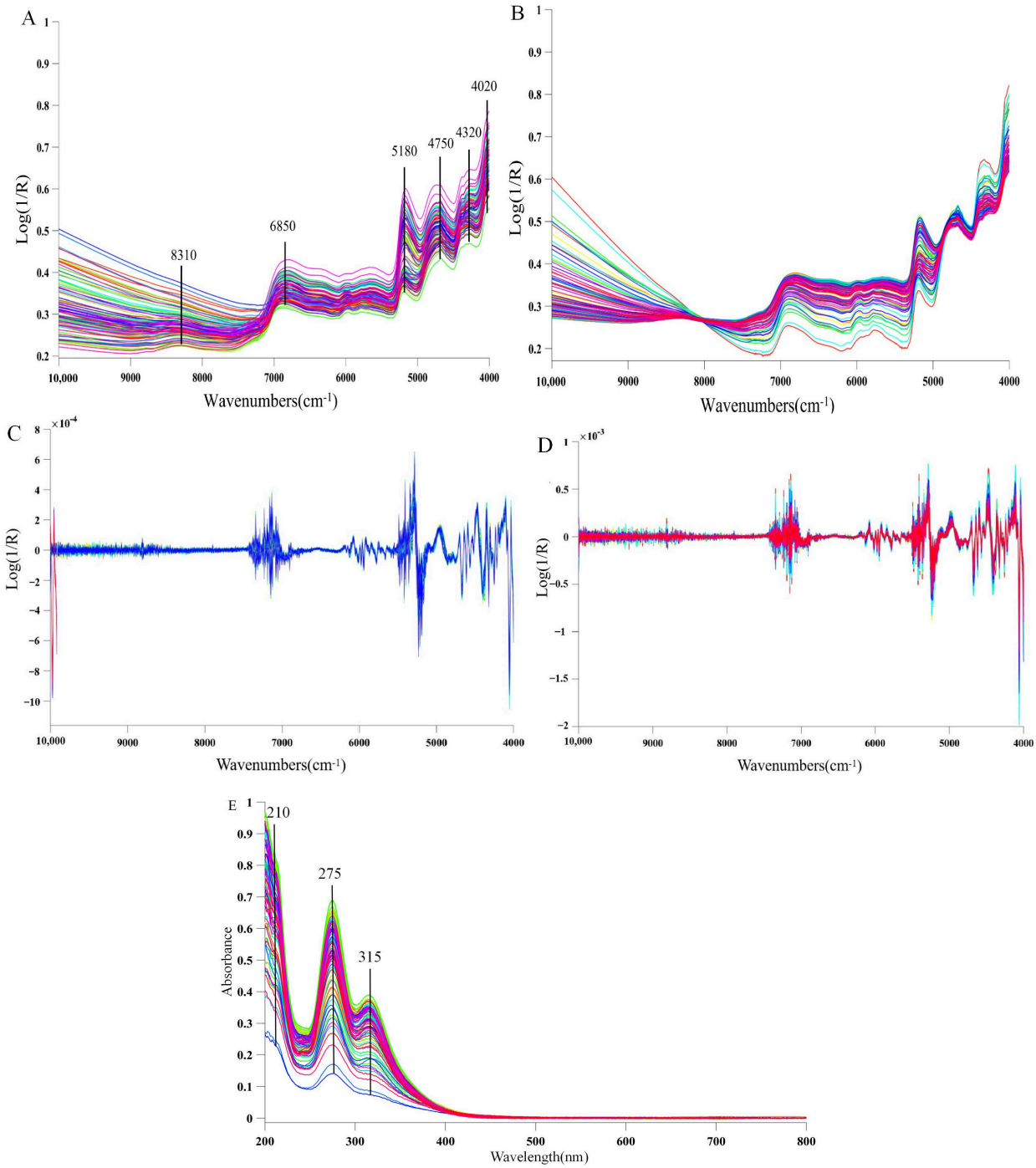


Figure 3. Full NIR and UV spectra of Wine-processed *S. baicalensis.* (**A**) Raw Log(1/R) NIR spectra; (**B**) greatly mitigated, corrected NIR spectra; (**C**) second derivative spectra of raw spectra; (**D**) second derivative spectra combined with MSC; (**E**) UV absorption spectra ranging from 200 to 800 nm.

After combined MSC and second derivative preprocessing, regular spectral variations consistent with molecular structural differences among the four active ingredients can be observed in the above wavebands, which greatly mitigated spectral interference from the sample matrix and provided reliable spectral support for quantitative analysis of multiple flavonoid components in wine-processed *S. baicalensis*.

The original UV spectra of the 90 wine-processed *S. baicalensis* samples are displayed in Figure 3E. Two major characteristic absorption bands can be observed, namely Band I (300–400 nm) and Band II (240–280 nm). The variations in peak position and profile of the four components depend on substituent types and the existence of glycosidic bonds.

Baicalin, which contains o-diphenolic hydroxyl groups and glycosidic bonds, has characteristic absorption peaks at approximately 315–320 nm (Band I) and 275–280 nm (Band II) [19]. As a flavonoid glycoside with monophenolic hydroxyl substitution, wogonoside shows absorption peaks at about 310–315 nm (Band I) and 270–275 nm (Band II) [20].

Baicalein is the aglycone of baicalin. It possesses o-diphenolic hydroxyl groups without glycosidic bonds, forming a more complete conjugated system and resulting in red shift of absorption peaks. Its Band I is located at 325–330 nm and Band II at approximately 280–285 nm. Wogonin, the aglycone of wogonoside, has monophenolic hydroxyl groups and no glycosidic bonds, with absorption peaks at roughly 320–325 nm (Band I) and 275–280 nm (Band II) [21].

Finally, 1557 variables from NIR spectra and 601 variables from UV spectra were collected.

3.3. Data Preprocessing

Different spectral pretreatment methods exerted remarkably distinct influences on the performance of PLS quantitative models for NIR and UV spectra. For the NIR spectral system, most pretreated models exhibited lower RMSEC and RMSECV, accompanied by an overall increase in prediction set R_p^2 and a decrease in RMSEP. This indicates that physical interferences such as scattering and baseline drift are critical factors restricting the generalization ability of NIR models, while spectral pretreatment can effectively eliminate irrelevant noise and enhance effective characteristic information.

Among all pretreatment schemes, SNV delivered prominent performance: for the NIR spectra of baicalin, wogonoside, and baicalein, SNV treatment yielded the lowest RMSECV and relatively low RMSEC, achieving optimal comprehensive performance in model fitting and cross-validation. Taking baicalin as an example, the RMSECV after SNV pretreatment was 1.013, lower than that of raw spectra without pretreatment and other pretreatment methods. For the NIR spectrum of wogonin, SG smoothing combined with 2D achieved the best results. This scheme required only 4 LVs with lower model complexity, and possessed the lowest RMSECV and the smallest RMSEP among all pretreatment groups. Its comprehensive prediction and generalization ability outperformed single SNV treatment and the combination of SNV with 2D.

For the UV spectral systems of the four target components, SG smoothing exhibited stable and reliable performance. SG pretreatment slightly decreased RMSECV and RMSEP for the UV spectra of baicalin, wogonoside, baicalein, and wogonin, making it the preferred pretreatment method for UV spectra in this system. MSC is an algorithm designed to correct scattering effects of solid particles. However, no obvious particle scattering interference exists in UV detection systems; forced application of MSC and its derivative-combined treatments distorted spectral baselines and amplified errors, resulting in sharp abnormal increases in RMSEP on multiple occasions. For instance, the RMSEP of the baicalin UV group with MSC combined with 1D pretreatment reached as high as 5.404, which severely degraded model prediction stability, proving MSC unsuitable for UV spectral correction in this study (Table S3, Supplementary Material).

The number of LVs reflects model complexity. None of the optimized models adopted excessively high LVs, avoiding overfitting while ensuring fitting accuracy. Parameters were screened based on leave-one-out cross-validation RMSECV, which effectively balanced model fitting ability and generalization performance. By comprehensively evaluating calibration and cross-validation indicators, with the prediction set used only for final validation, the optimal pretreatment methods for NIR and UV spectra of four flavonoid components in wine-processed *S. baicalensis* were determined, laying a solid foundation for subsequent NIR-UV data fusion modeling (Table 1).

Table 1. Optimal pretreatment methods for NIR and UV spectroscopic quantification.

Target Component	Spectra	Optimal Pretreatment Method	LVs	Calibration		Validation		Cross-Validation
				Rc^2	*RMSEC*	Rp^2	*RMSEP*	*RMSECV*
Baicalin	NIR	SNV	6	0.980	0.834	0.927	1.205	1.013
	UV	SG	6	0.983	0.755	0.961	0.851	1.160
Wogonoside	NIR	SNV	6	0.985	0.124	0.932	0.221	0.180
	UV	SG	6	0.981	0.137	0.967	0.145	0.254
Baicalein	NIR	SNV	6	0.958	0.351	0.897	0.474	0.484
	UV	SG	5	0.948	0.387	0.975	0.279	0.451
Wogonin	NIR	SG + 2D	4	0.975	0.124	0.929	0.162	0.158
	UV	SG	6	0.969	0.132	0.965	0.133	0.177

3.4. Feature Extraction

To reduce computational complexity and minimize interference from redundant spectral information, iPLS was applied to perform variable selection and feature extraction on the NIR and UV datasets. The modeling performance of each candidate sub-interval was evaluated and compared using RMSECV [22].

For baicalin, the optimal number of sub-intervals for the NIR spectra was 30. The NIR variables selected by iPLS were mainly distributed near 6100–5800, 5200–5000, and 4800–4600 cm^{-1}, corresponding to C-H stretching vibration, O-H stretching vibration, and the combination vibration of coupled C-O stretching and C-O-H bending, respectively. The sub-interval with the minimum RMSECV was selected for subsequent MLDF modeling. For the UV spectra, the variables selected by iPLS were more uniformly distributed, which may be related to the broad absorption band of baicalin in the UV region. When the number of sub-intervals was 10, the RMSECV reached its minimum within the 261–320 nm range.

For wogonoside, the selected variables were similar to those of baicalin. When the NIR spectrum was divided into 10 segments, the minimum RMSECV obtained from the optimal sub-interval was identical to that obtained with 40 segments. The selected sub-intervals were all located near the characteristic regions of wogonoside in the NIR spectrum. After considering other factors, 40 was chosen as the optimal number of sub-intervals. The optimal UV sub-interval for wogonoside was the same as that for baicalin.

For baicalein, the characteristic vibrational regions in the NIR spectrum were located at 6100–5800, 5200–5000, and 4600–4400 cm^{-1}, corresponding to aromatic C-H stretching, phenolic O-H stretching, and aromatic C-H bending vibrations, respectively. The optimal NIR sub-interval selected was also near its characteristic regions, with an optimal number of 30 sub-intervals. The optimal number of sub-intervals for the UV spectrum was 10.

For wogonin, the optimal number of NIR sub-intervals was 30, and the minimum RMSECV was achieved in the 6005–6201 cm^{-1} band. The optimal number of UV sub-intervals near its characteristic regions was 20.

For MLDF, iPLS variable selection was performed independently on the NIR and UV spectra. After screening the optimal feature sub-intervals for each spectrum, the variables within the optimal intervals were retained and concatenated. The original total number of fusion variables was 2158, which was compressed to 82–112 after independent iPLS screening of the NIR and UV spectral data.

3.5. Data Fusion Strategies

For LLDF, two raw spectral datasets were pretreated individually and concatenated to form a novel data matrix [23]. The results demonstrated that the baicalin model obtained the minimal RMSECV in the absence of pretreatment, while SNV pretreatment provided the lowest RMSECV for wogonoside, baicalein and wogonin (Table S4 Supplementary Materials).

Compared to models using only a single NIR or UV spectrum, the PLS models established by means of data fusion strategies demonstrated better predictive performance [24]. For baicalin, LLDF improved predictive accuracy and significantly reduced prediction errors relative to the single NIR spectrum model. In contrast, LLDF showed no significant improvement in either predictive accuracy or error reduction for UV spectrum. The results observed for wogonoside are similar to baicalin, which may be attributed to the strong absorption range of UV spectroscopy. For baicalein, LLDF represented lower RMSEP compared with single spectrum models with the value of 0.249. For wogonin, LLDF with a higher R_p^2 of 0.975 and a lower RMSEP of 0.096, showed marked improvement compared to the single NIR model and the single UV model. For MLDF, iPLS variable selection was performed independently on the NIR and UV spectra. The retained variables screened from the optimal NIR-iPLS and UV-iPLS sub-intervals were concatenated to construct the MLDF matrix [25]. Compared with LLDF, which directly uses all 2158 full-spectrum variables, MLDF effectively compresses the number of variables to 82–112, substantially reducing data redundancy and model complexity. Comparison of the modeling results for the four flavonoid components reveals a clear component-dependent characteristic between MLDF and LLDF. For baicalin, the MLDF model yielded an R_c^2 of 0.979, an R_p^2 of 0.953, and an RMSEP of 0.845.

Compared with the unpretreated LLDF model (R_c^2 = 0.977, R_p^2 = 0.904, RMSEP = 1.087), the calibration ability was slightly improved, and both prediction accuracy and stability were significantly enhanced, indicating that feature selection effectively improved the generalization ability of the fused model. For wogonoside, the MLDF model (R_p^2 = 0.950, RMSEP = 0.182) performed slightly better overall than the untreated LLDF model (R_p^2 = 0.937, RMSEP = 0.170); both fusion models exhibited stable performance, with MLDF further simplifying the variable dimension while maintaining a low prediction error. For baicalein, the advantage of MLDF was most pronounced, with Rp^2 as high as 0.970, RMSEP as low as 0.214, and a RPD of 4.041, clearly outperforming the LLDF model (R_p^2 = 0.965, RMSEP = 0.249, RPD = 3.563). This demonstrates that iPLS feature selection can effectively eliminate redundant interference, maximize the complementary information from NIR and UV spectra, and significantly improve the quantitative accuracy and robustness of the model. For wogonin, the MLDF model showed better calibration performance (R_c^2 = 0.981, higher than 0.973 of the LLDF model), but its predictive performance (R_p^2 = 0.955, RMSEP = 0.094) was slightly inferior to that of the optimal SNV-pretreated LLDF model (R_p^2 = 0.979, RMSEP = 0.084). This indicates that, for this specific component, direct full-spectrum fusion may retain more subtle effective information, whereas iPLS variable selection may result in a slight loss of such information during compression (Table 2).

Table 2. Model performance for NIR and UV spectra based on iPLS algorithm.

Chemical Composition	Spectra	Methods	Pretreatment	Variables	LVs	R_c^2	R_p^2	RMSEC	RMSEP	RMSECV	Bias	RPD
Baicalin	NIR	Full-PLS	None	1557	6	0.977	0.886	0.888	1.483	1.014	0.0229	2.190
			SNV	1557	6	0.980	0.928	0.834	1.205	1.013	0.0394	2.695
		iPLS	SNV	52	6	0.968	0.938	1.050	1.108	1.111	−0.1850	2.933
	UV	Full-PLS	None	601	5	0.982	0.948	0.788	0.964	1.159	0.0312	3.065
			SG	601	6	0.983	0.961	0.755	0.851	1.160	0.1057	3.472
		iPLS	SG	60	5	0.977	0.964	0.903	0.821	1.150	−0.0418	3.582
	LLDF	Full-PLS	None	2158	6	0.977	0.904	0.876	1.087	1.090	0.0622	2.35
	MLDF		None	112	4	0.979	0.953	0.896	0.845	0.986	0.2000	3.005
Wogonoside	NIR	Full-PLS	None	1557	6	0.978	0.884	0.151	0.280	0.167	0.0172	2.165
			SG + 2D	1557	4	0.976	0.938	0.159	0.232	0.201	0.0100	2.610
		iPLS	SG + 2D	39	4	0.989	0.940	0.106	0.203	0.133	−0.00358	2.980
	UV	Full-PLS	None	601	6	0.983	0.966	0.132	0.146	0.255	0.0189	3.669
			SG	601	6	0.982	0.967	0.137	0.145	0.254	0.0200	3.700
		iPLS	SG	60	2	0.971	0.969	0.175	0.138	0.216	−0.0158	3.952
	LLDF	Full-PLS	None	2158	6	0.976	0.937	0.156	0.170	0.204	0.0282	2.751
			1D	2158	4	0.976	0.944	0.154	0.163	0.235	0.0292	2.872
	MLDF		None	99	6	0.980	0.950	0.144	0.182	0.181	−0.0414	2.827
Baicalein	NIR	Full-PLS	None	1557	5	0.926	0.847	0.461	0.578	0.531	−0.0809	1.853
			SNV	1557	6	0.958	0.897	0.351	0.474	0.484	−0.0334	2.257
		iPLS	SNV	52	6	0.941	0.914	0.414	0.444	0.418	−0.0384	2.410
	UV	Full-PLS	None	601	5	0.949	0.973	0.381	0.283	0.453	−0.0095	3.320
			MSC	601	3	0.882	0.979	0.571	0.306	0.665	−0.0761	3.071
		iPLS	MSC	60	4	0.932	0.963	0.441	0.590	0.591	0.2889	1.590
	LLDF	Full-PLS	None	2158	6	0.939	0.965	0.406	0.249	0.528	−0.0510	3.563
			SNV	2158	5	0.925	0.964	0.448	0.250	0.515	−0.0487	3.549
	MLDF		None	112	6	0.957	0.970	0.344	0.214	0.496	−0.0193	4.041
Wogonin	NIR	Full-PLS	None	1557	5	0.965	0.898	0.147	0.193	0.162	−0.0230	2.259
			SG + 2D	1557	4	0.975	0.929	0.124	0.162	0.158	−0.0135	2.683
		iPLS	SG + 2D	52	4	0.991	0.947	0.074	0.137	0.102	0.0009	3.163
	UV	Full-PLS	None	601	5	0.968	0.964	0.134	0.144	0.180	−0.0040	2.999
			SNV	601	3	0.941	0.979	0.182	0.170	0.226	−0.0120	2.538
		iPLS	SNV	30	3	0.938	0.965	0.187	0.150	0.218	−0.0051	2.871
	LLDF	Full-PLS	None	2158	6	0.973	0.975	0.125	0.096	0.178	−0.0263	3.714
			SNV	2158	6	0.973	0.979	0.124	0.084	0.165	−0.0232	4.267
	MLDF		None	82	5	0.981	0.955	0.105	0.094	0.132	0.0025	3.328

3.6. Analysis of Predictive Performance of the Models

Figures 4 and 5 display the scatter plots of measured versus predicted contents for four target flavonoids in wine-processed *S. baicalensis* based on LLDF and MLDF strategies, respectively. The diagonal line in each subplot represents the ideal fitting condition where the predicted value is equal to the measured value; samples distributed closer to the diagonal and higher determination coefficient (R^2, closer to 1) indicate superior quantitative performance of the model.

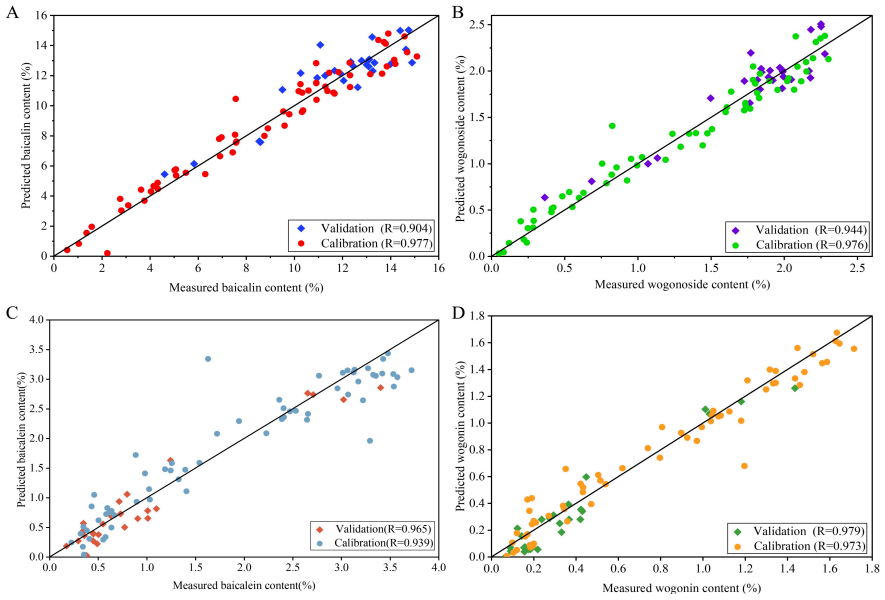


Figure 4. Comparison between predicted and measured values of the four target components in wine-processed *S. baicalensis* using LLDF ((**A**) baicalin; (**B**) wogonoside; (**C**) baicalein; (**D**) wogonin).

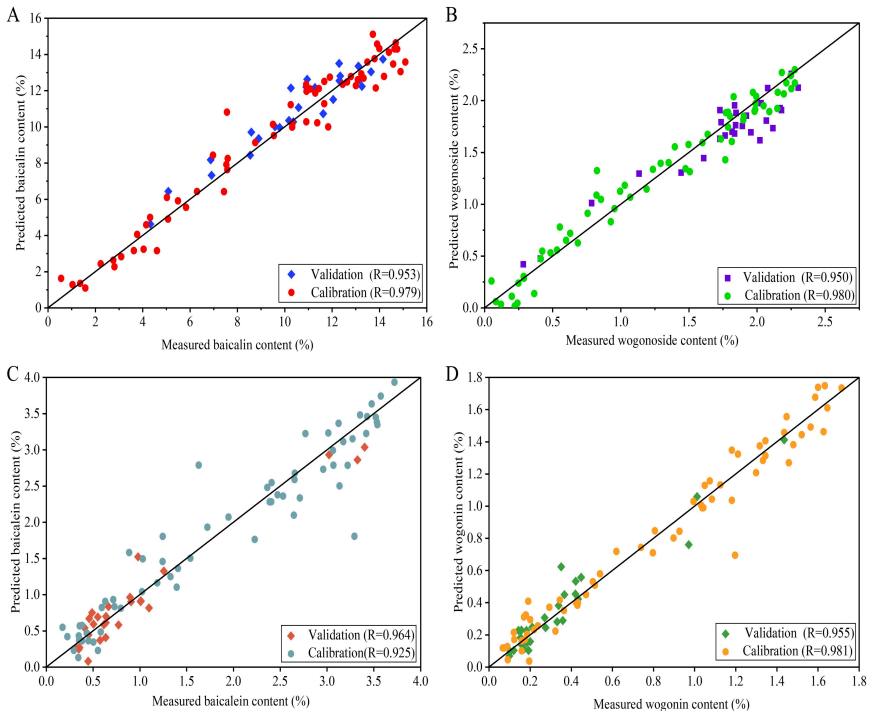


Figure 5. Comparison between predicted and measured values of the four target components in wine-processed *S. baicalensis* using MLDF ((**A**) baicalin; (**B**) wogonoside; (**C**) baicalein; (**D**) wogonin).

For baicalin, the LLDF model yielded R_c^2 = 0.977 for the calibration set and R_p^2 = 0.904 for the validation set. After mid-level feature fusion, the R_c^2 slightly increased to 0.979, and the R_p^2 was improved from 0.904 to 0.953, accompanied by less scattered data points and greatly improved external sample prediction stability. Wogonoside exhibited comparable modeling performance under both fusion schemes, with a constant R_c^2 of 0.980 and R_p^2 of 0.950, enabling accurate quantification with either strategy. For baicalein, the LLDF model achieved R_c^2 = 0.939 and R_p^2 = 0.965; the MLDF model slightly reduced the calibration-set R_c^2 to 0.925 while maintaining a high R_p^2 of 0.964, revealing strong generalization capacity. Wogonin showed the most prominent performance improvement: the LLDF model presented R_c^2 = 0.973 and R_p^2 = 0.979, and the MLDF model raised the R_c^2 to 0.981 with an R_p^2 of 0.955, showing tighter fitting for calibration samples.

To evaluate model residuals and detect potential outliers, residual analysis and Hotelling's T^2-Q diagnostics were performed. Figures 6 and 7 show the residual plots (measured value versus residual) for the four flavonoids in the LLDF-PLS and MLDF-PLS models, respectively. The residuals were randomly distributed around the zero-residual line without obvious systematic trends, and nearly all samples fell within the ±3 SD limits, demonstrating that no obvious biased outliers existed in the calibration set. Figures 8 and 9 present the Hotelling's T^2-Q residual plots for the LLDF-PLS and MLDF-PLS models, respectively. Only a small number of samples slightly exceeded the 95% confidence limit of Hotelling's T^2. For these samples exceeding the 95% confidence limit, a careful examination was conducted by combining their HPLC reference values, original spectral profiles, and sample preparation records. No measurement anomalies were found; these samples were genuine extreme samples under deep wine-processing conditions rather than erroneous outliers. This phenomenon can be attributed to the inherent chemical differences among wine-processed *S. baicalensis* samples with different processing degrees (e.g., 9 min and 12 min). Since these samples did not introduce systematic bias into the model, they were retained in the calibration set to ensure the model covers the full range of processing degrees.

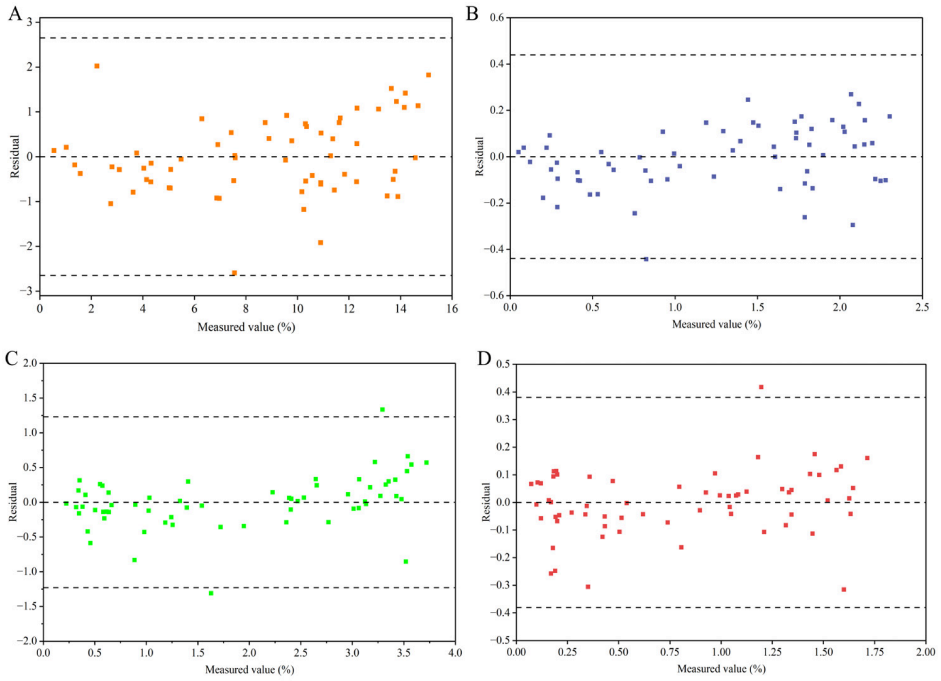


Figure 6. Residual plots (measured value versus residual) for the calibration set of the LLDF-PLS models. The dashed lines represent zero residual and the ±3 SD limits ((**A**) baicalin; (**B**) wogonoside; (**C**) baicalein; (**D**) wogonin).

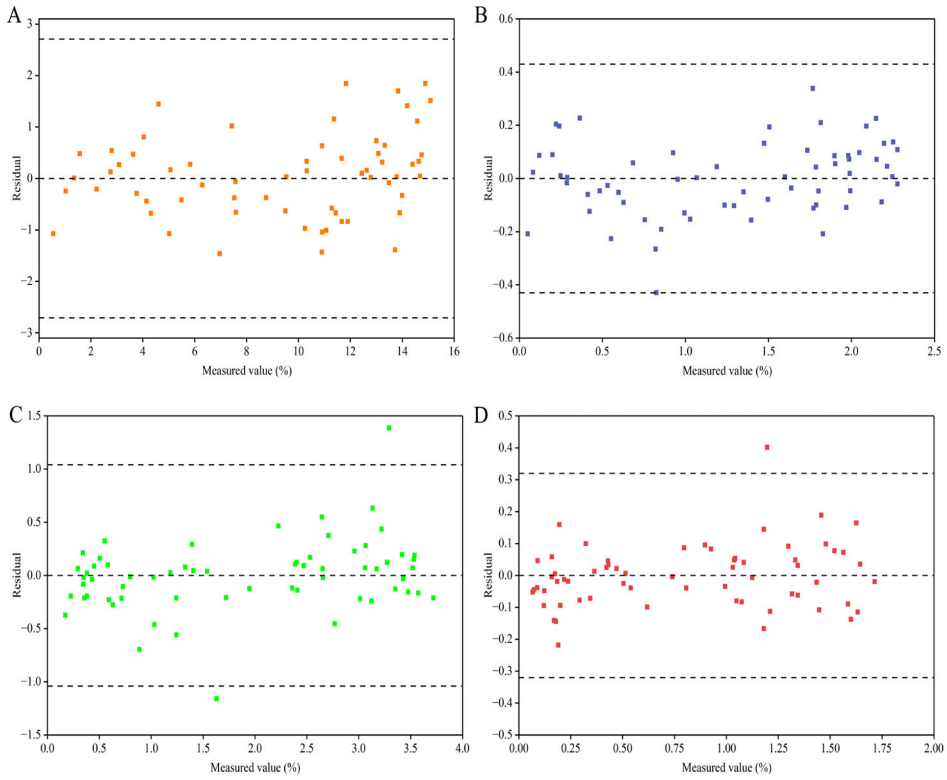


Figure 7. Residual plots (measured value versus residual) for the calibration set of the MLDF-PLS models. The dashed lines represent zero residual and the ±3 SD limits ((**A**) baicalin; (**B**) wogonoside; (**C**) baicalein; (**D**) wogonin).

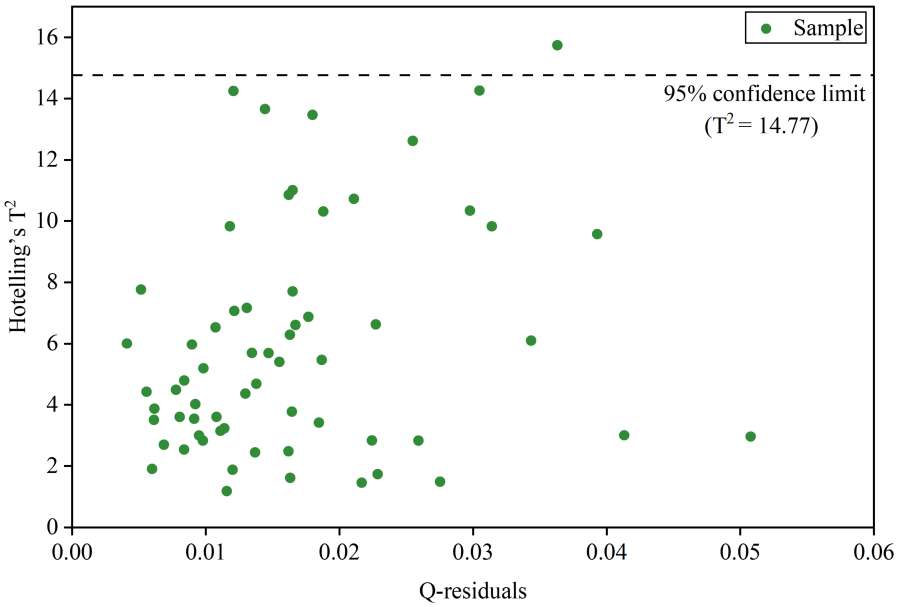


Figure 8. Hotelling's T^2 vs. Q-residuals plot for the calibration set (n = 63) of the LLDF-PLS model. The horizontal dashed line indicates the 95% confidence limit (T^2 = 14.77).

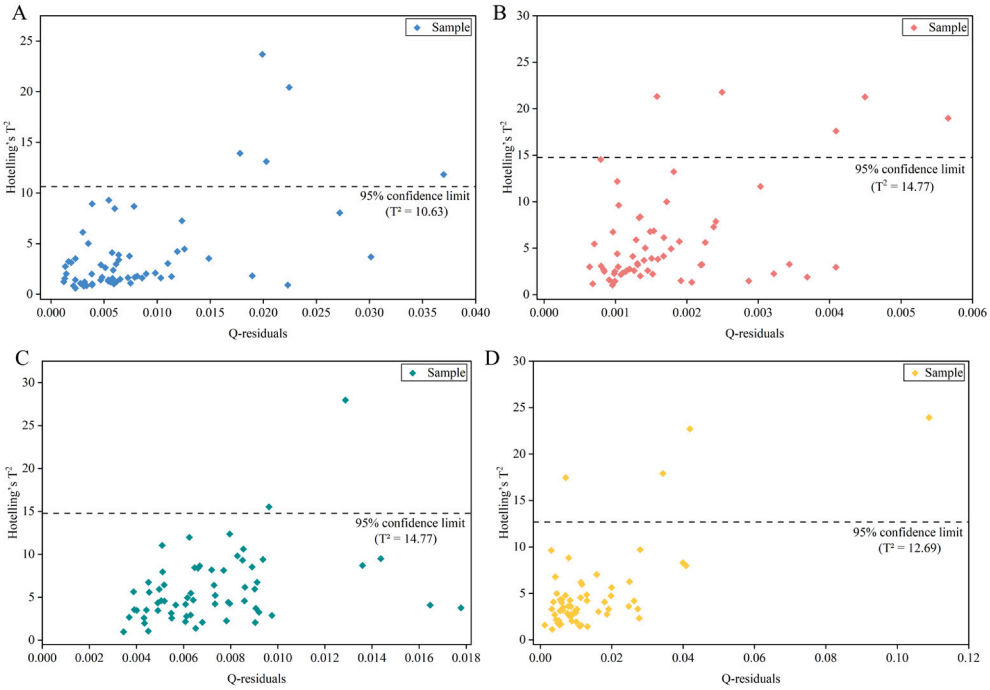


Figure 9. Hotelling's T^2 versus Q-residuals plots for the calibration set of the MLDF-PLS models. The dashed lines represent the 95% confidence limit of Hotelling's T^2 ((**A**) baicalin; (**B**) wogonoside; (**C**) baicalein; (**D**) wogonin).

To further evaluate the complexity and generalization risk of all optimized PLS models, the RMSECV/RMSEC ratio was calculated and summarized in Table 3. Most of the models established in this work yielded an RMSECV/RMSEC ratio below 1.3, indicating reasonable model complexity and favorable generalization performance. Nevertheless, several optimized models showed ratios exceeding 1.3, suggesting potential insufficient generalization ability, and this phenomenon was not limited to the original full-spectrum NIR model. The RMSECV and optimal LVs under four modelling strategies (Single-NIR, Single-UV, LLDF and MLDF) for baicalin (A), wogonoside (B), baicalein (C) and wogonin (D) are compared in Figure 10. For baicalin (A), the MLDF fusion model achieved the lowest RMSECV with the smallest number of LVs, demonstrating that mid-level data fusion simplified the model while improving prediction accuracy. For wogonoside (B), all four models maintained stable optimal LVs, and the MLDF model still presented the minimal cross-validation error. In the baicalein model (C), the Single-UV model obtained the lowest RMSECV together with fewer LVs, while the LLDF model required more latent variables and produced higher prediction error. For wogonin (D), the UV and LLDF models needed more LVs, and the MLDF model reduced both RMSECV and the number of latent variables simultaneously. Overall, the data fusion strategies generally tended to reduce prediction error and often required fewer latent variables compared with single-spectrum models. A large discrepancy between RMSECV and RMSEC is commonly attributed to multiple influencing factors. Inappropriate variable selection may retain redundant noisy spectral variables, raising model complexity and triggering overfitting. Sample partitioning also affects the gap between RMSECV and RMSEC, and inconsistent concentration distribution across calibration and cross-validation subsets can elevate cross-validation error. The presence of extreme samples may further enlarge this discrepancy; such extreme points are not necessarily measurement artifacts but authentic samples spanning the full experimental range. Careful selection of spectral variables and variable combinations provides an effective way to balance model complexity and generalization ability. Retaining too many irrelevant or noisy variables will increase model complexity and weaken its performance on unseen samples, while excessively aggressive

variable elimination may discard chemically meaningful spectral signals and result in underfitting. The iPLS algorithm adopted in this work screens informative spectral intervals and constructs rational variable combinations, which helps retain characteristic signals related to target analytes while removing redundant noise, so that the established model can capture the core spectral features without overfitting the calibration data. Although these models display a certain overfitting tendency, they still maintain acceptable predictive reliability. In contrast, most glycoside and aglycone models built by LLDF and MLDF fusion strategies achieved low RMSECV/RMSEC ratios, demonstrating that appropriate spectral pretreatment and rational iPLS variable selection effectively reduced redundant spectral information and greatly alleviated the risk of overfitting.

Table 3. Comparison of RMSEC, RMSECV, and their ratio for quantitative PLS models of four target components constructed based on optimal single spectra and data fusion strategies.

Chemical Composition	Spectra	RMSEC	RMSECV	RMSECV/RMSEC
Baicalin	NIR	0.834	1.013	1.215
	UV	0.755	1.160	1.536
	LLDF	0.876	1.090	1.244
	MLDF	0.896	0.986	1.100
Wogonoside	NIR	0.124	0.180	1.145
	UV	0.137	0.254	1.854
	LLDF	0.156	0.204	1.308
	MLDF	0.144	0.181	1.257
Baicalein	NIR	0.351	0.484	1.379
	UV	0.387	0.451	1.165
	LLDF	0.406	0.528	1.300
	MLDF	0.344	0.496	1.442
Wogonin	NIR	0.124	0.158	1.274
	UV	0.132	0.177	1.341
	LLDF	0.125	0.178	1.424
	MLDF	0.105	0.132	1.257

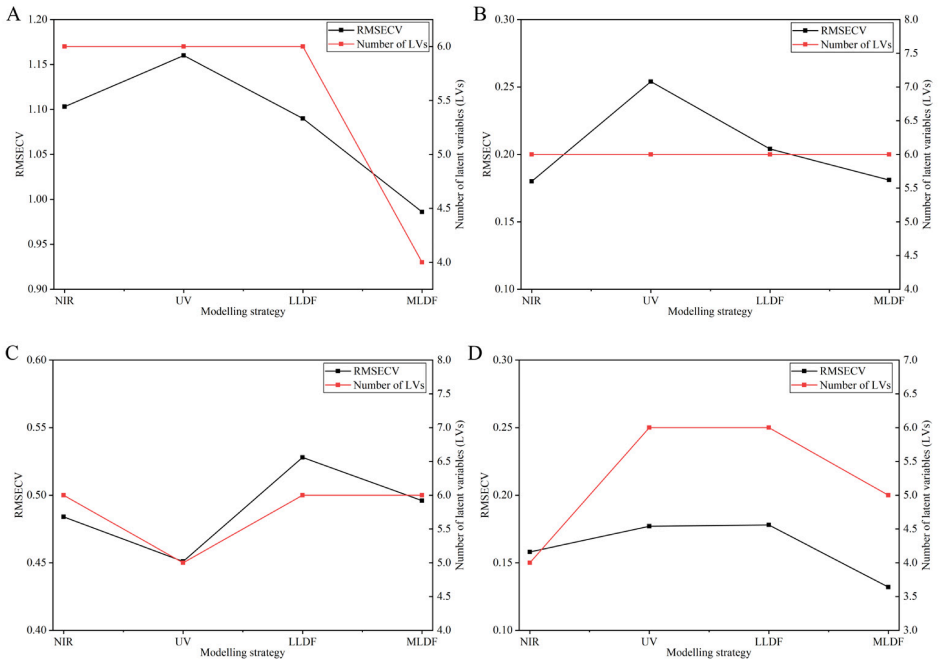


Figure 10. Comparison of RMSECV and the optimal number of LVs under four modelling strategies (Single-NIR, Single-UV, low-level data fusion, and mid-level data fusion) ((**A**) baicalin; (**B**) wogonoside; (**C**) baicalein; (**D**) wogonin).

In summary, both LLDF and MLDF models achieved reliable quantification of the four flavonoids, with all calibration and validation determination coefficients exceeding 0.90. Different from LLDF, which directly concatenates full-spectrum variables, MLDF removes redundant spectral noise by extracting informative feature variables via the iPLS algorithm.

4. Discussion

Glycosidic components in medicinal herbs are easily hydrolyzed during wine-processing (Figure 11). Glycosidic bond cleavage transforms glycosides into aglycones, reducing glycoside levels and increasing aglycone contents, which is the primary mechanism for chemical changes in processed *S. baicalensis* [26]. HPLC analysis revealed that wine-processing time markedly affects its chemical composition. No significant ingredient variation was observed within 0–3 min. After 6 min of processing, glycoside contents decreased rapidly while aglycones increased obviously. At 12 min, aglycones reached the highest level, and glycosides decreased to the lowest content. The pharmacological value of this conversion depends on the clinical application of wine-processed *S. baicalensis*. Excessive transformation is unfavorable when retaining glycoside-dominated effects, such as heat-clearing and dampness-drying activities. In contrast, glycoside-to-aglycone conversion benefits the antibacterial, antioxidant, and anti-inflammatory capacities and improves the dissolution and absorption of active components. Based on the dynamic changes of baicalin, wogonoside, baicalein, and wogonin, 9 min was determined as the optimal processing duration. At this condition, baicalein and wogonin increased significantly to 26.06 mg/g and 10.30 mg/g, whereas baicalin and wogonoside were well retained at 77.20 mg/g and 10.10 mg/g. Although 12 min processing yielded the maximum aglycone content, it caused excessive degradation of glycosidic components. Therefore, 9 min is the preferred processing time. These quantitative HPLC data support further model establishment.

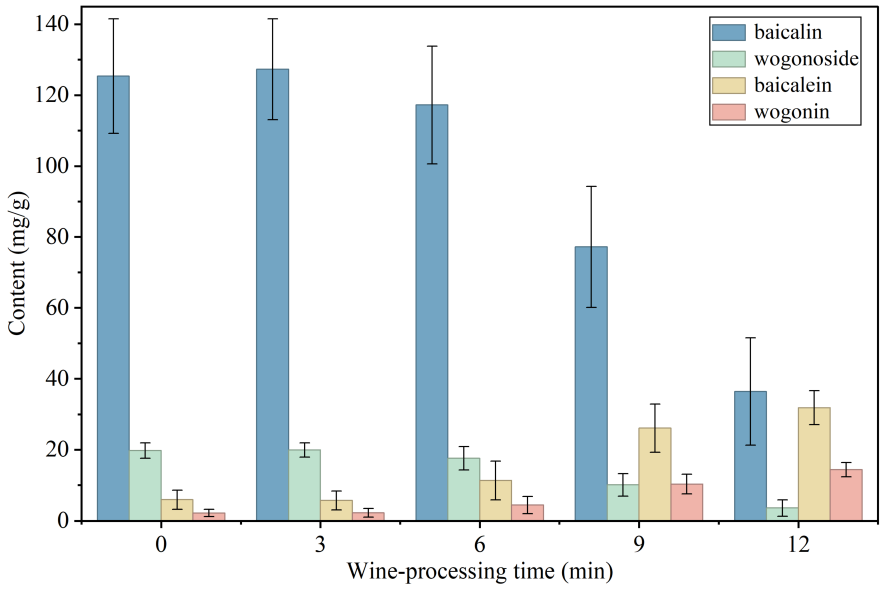


Figure 11. Changes in the contents of baicalin, wogonoside, baicalein and wogonin in raw *S. baicalensis* and wine-processed samples treated with different wine-processing durations. Data are presented as mean ± SD (n = 18).

In this study, spectral prediction models were established based on accurate HPLC quantitative data. The iPLS algorithm was adopted for characteristic variable selection and dimensionality reduction to improve model accuracy. The results indicated that reliable single NIR and UV quantitative models could be achieved after spectral pretreatment and iPLS

optimization [27]. Optimized single UV models presented excellent stability and prediction performance. In contrast, raw full-spectrum NIR models exhibited serious overfitting. For the baicalin NIR model, the high R_c^2 (0.977) but low R_p^2 (0.886) and discrepant RMSEC (0.888) and RMSEP (1.483) indicated favorable calibration but poor generalization. Similarly, the untreated wogonoside NIR model also showed obvious overfitting ($R_c^2 - R_p^2 = 0.094$). Appropriate pretreatment methods, including SNV, SG smoothing, and second derivative transformation combined with iPLS selection, effectively eliminated spectral noise, reduced irrelevant variables, and significantly alleviated overfitting, thereby improving the accuracy and stability of single-spectrum models. Figure 12 quantitatively compares the modeling performance of single-spectrum and data fusion strategies for four target flavonoids. Simple full-spectrum LLDF failed to exert complementary spectral advantages owing to the superposition of redundant and overlapping noise. For glycosides (baicalin and wogonoside), LLDF provided satisfactory calibration but inferior prediction compared with optimized single-UV models. The baicalin LLDF model (R_c^2 = 0.977, R_p^2 = 0.904, RPD = 2.35) and wogonoside LLDF model (R_c^2 = 0.976, R_p^2 = 0.944, RPD = 2.751) showed considerably lower predictive capacity than their corresponding single-UV models. In contrast, LLDF greatly improved the prediction of aglycones. The baicalein LLDF model (R_c^2 = 0.925, R_p^2 = 0.964) yielded higher R_p^2 than the single NIR model and comparable accuracy to the single-UV model. The wogonin LLDF model (R_c^2 = 0.973, R_p^2 = 0.979) even outperformed both single-spectrum models. Compared with LLDF, MLDF generally enhanced the prediction of glycosides, with favorable fitting and prediction results for baicalin (R_c^2 = 0.979, R_p^2 = 0.953) and wogonoside (R_c^2 = 0.980, R_p^2 = 0.950). For aglycones, MLDF exhibited inconsistent improvement: it further promoted baicalein prediction (R_c^2 = 0.957, R_p^2 = 0.970) but slightly reduced the performance for wogonin.

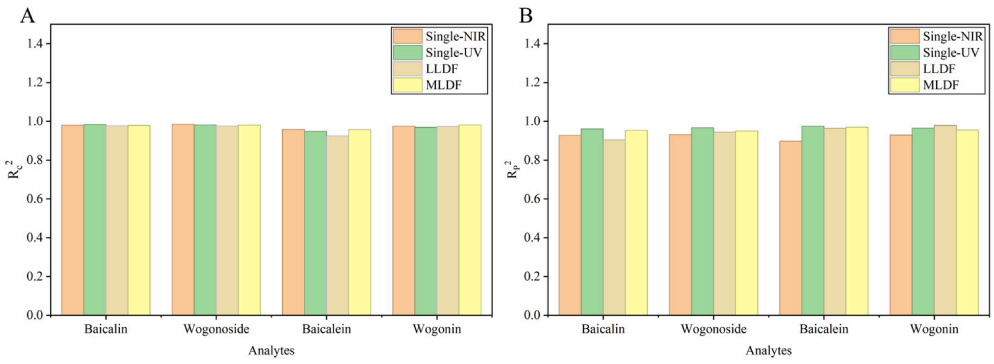


Figure 12. Comparison of determination coefficients R_c^2 and R_p^2 of Single-NIR, Single-UV, LLDF and MLDF models for four analytes, namely baicalin, wogonoside, baicalein and wogonin. (**A**) R_c^2 of calibration set; (**B**) R_p^2 of prediction set.

In summary, direct fusion of raw full-spectrum data without spectral pretreatment and characteristic variable selection introduces a large amount of invalid and overlapping redundant information, which fails to effectively resolve characteristic component signals in complex matrices. This approach cannot improve model prediction performance; on the contrary, it significantly reduces the generalization ability and quantitative accuracy of the model. This study takes full advantage of both chromatographic separation and rapid spectroscopic detection. Using highly accurate HPLC chromatographic separation results as reference true values, virtual resolution in the spectral domain is realized via spectral pretreatment and iPLS characteristic variable selection. The MLDF strategy is adopted to fully exploit the information complementarity between NIR and UV spectra, substantially improving the prediction accuracy for simultaneous multi-component determination as well as model robustness. The complementary combination of precise chromatographic

separation and rapid spectroscopic characterization effectively compensates for the drawbacks of single-spectrum techniques, such as signal overlap and limited resolution capacity, and overcomes the disadvantages of conventional chromatographic methods including time-consuming measurement and cumbersome operation. This work demonstrates a feasible integration strategy of multi-source spectral data-fusion and chromatographic quantification techniques for the simultaneous analysis of multiple active constituents in wine-processed *S. baicalensis* with different processing degrees. It provides a novel and reliable method for quality evaluation of processed herbal medicines toward the integration of precise separation-based quantification and rapid non-destructive detection.

Despite the satisfactory quantitative performance of the spectral fusion models established in this study, several practical limitations remain. Although experimental samples were obtained from multiple manufacturers, the sample set still needs to be expanded to incorporate samples from authentic production areas, general production areas, different manufacturers and multiple batches for external validation, so as to verify the generalizability of the models. With advances in modern manufacturing technology, automatic stir-frying equipment for traditional Chinese medicine has been widely adopted. Samples prepared by automatic stir-frying machines and manual processing may require separate model validation, which poses a major challenge to the universality of the predictive models. All spectral measurements of wine-processed *S. baicalensis* were performed under stable laboratory conditions (temperature: 25 ± 2 °C; relative humidity: 45 ± 2%). In practical applications, fluctuations in temperature and humidity can alter the moisture status of powder samples and introduce spectral interference, reducing model robustness under real-world conditions. Furthermore, the current models were trained and validated solely on laboratory-grade spectrometers. Hardware discrepancies such as differences in light sources and detectors hinder model transfer to industrial on-site environments, representing an area requiring long-term research efforts. Meanwhile, spectral repeatability is susceptible to preprocessing factors including sample particle size, sample filling thickness and sample storage duration. Although rigorous and unified sample preparation protocols were adopted in this study, human operational variations in practical manipulation can still cause spectral drift. In addition, the sample size was limited for the deeply processed groups at 9 min and 12 min (these groups have the highest aglycone contents and correspond to critical technological evaluation indicators). After uniform KS partitioning, only 1–2 samples of these extreme processing groups were assigned to the validation set, which may restrict thorough verification of model generalizability for deep-processing samples. This uneven sample distribution originates from the inherent characteristics of the original sample population rather than inappropriate partitioning. Future research will focus on expanding the sample size of deeply processed specimens, conducting multi-origin external validation, optimizing environmental interference correction, developing efficient model transfer approaches, and establishing standardized sample preparation workflows to promote the practical application of multi-spectral quantitative techniques in the online quality monitoring of traditional Chinese medicine processing.

5. Conclusions

This study investigated the variation of flavonoid constituents during wine-processing of *S. baicalensis* of different batches from three geographical origins. With the extension of processing time, the contents of baicalin and wogonoside gradually decreased, while the contents of baicalein and wogonin increased. Considering both glycoside retention and aglycone formation, 9 min was finally determined as the optimal stir-frying duration for wine-processed *S. baicalensis*. On the basis of clarifying the variation patterns of flavonoid constituents in *S. baicalensis* during wine-processing, NIR-UV spectral quantitative models

were established using HPLC-measured results as the reference. Combined with spectral pretreatment and the iPLS variable-screening algorithm, low-level and mid-level NIR-UV data-fusion quantitative models were separately constructed. Appropriate spectral pretreatment and feature-variable selection were able to effectively reduce spectral redundancy and alleviate model overfitting. LLDF is more suitable for the prediction of aglycone components, whereas MLDF improves the quantification performance of flavonoid glycosides. The proposed spectral-fusion strategy enables rapid and non-destructive multi-component detection, providing an alternative analytical option besides HPLC for the quality evaluation of wine-processed *S. baicalensis*. Nevertheless, practical application of this technique is restricted by limited sample coverage, environmental interferences and instrumental differences. Further research should expand the sample size, carry out external validation, correct environmental disturbances and improve model-transfer techniques to promote the application of this technique in online quality monitoring of herbal-medicine processing.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/chemosensors14100235/s1, Table S1: Sample source information of wine-processed *S. baicalensis*; Table S2: Number of calibration and validation samples at different wine-processing times for four flavonoid components under LLDF and MLDF strategies; Table S3: Effects of different spectral pretreatments on the performance of PLS model for NIR and UV spectra; Table S4: Effects of different spectral pretreatments on the performance of PLS model for LLDF.

Author Contributions: S.W.: Writing—original draft, Investigation, Formal analysis, Data curation, Writing—review and editing. X.P.: Methodology. T.Y.: Data curation. Q.W.: Experimental Methods. T.L.: Conceptualization, Funding acquisition. J.Z.: Project administration. Y.W.: Conceptualization, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number 2023YFC3504204. The APC was funded by the National Key Research and Development Program of China, grant number 2023YFC3504204.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available from the authors upon reasonable request, and the open-source code is available at https://ucphchemometrics.com/186-2/algorithms/ (accessed on 18 April 2026).

Acknowledgments: This work was supported by the National Key Research and Development Program of China (No. 2023YFC3504204).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China*; China Medical Science Press: Beijing, China, 2025.
2. Xiong, Y.; Wang, Y.Q.; Cao, L.J.; Yang, Y.Z.; Hu, L.Q.; Yang, M. Quality Evaluation of *Scutellariae radix* and Wine-Processed *Scutellariae radix* from Different Habitats. *J. Chin. Med. Mater.* **2018**, *41*, 2127–2133. [CrossRef]
3. Wang, L.; Min, W.J.; Bian, S.J.; Liu, Y.X.; Liu, L.; Qi, B. Effects of Different Types of Yellow Rice Wine on Quality of Wine-Processed *Scutellariae radix* Based on Fingerprint and Multi-Component Determination. *Chin. Tradit. Herb. Drugs* **2025**, *56*, 79–88. [CrossRef]
4. Cao, L.; Zhao, Y.W.; Chen, X.L.; Zhang, H.R.; Zhao, N.; Ding, Y.Q.; Duan, S.Y.; Qian, M.Y.; Ren, G.Q.; Wang, Z.Z.; et al. Integrated Quality Evaluation of Traditional Chinese Medicine Based on Chemical Composition and Biological Effects Using High-Level Data Fusion of Near-Infrared and Ultraviolet Spectroscopy Combined with Machine Learning. *Measurement* **2025**, *258*, 118972. [CrossRef]
5. Yao, W.Q.; Ye, J.H.; Sun, X.N.; Chang, P.J.; Guo, J.L.; Liu, X.H.; Zhong, G.H. Determination of Fatty Acids in Peanuts by NIR Combined with CARS-PLS. *Infrared Phys. Technol.* **2025**, *153*, 106316. [CrossRef]
6. Qian, Z.M.; He, Z.B.; Lian, D.Y.; Huang, Q.; Tan, G.Y.; Xie, J.Y. A Low-cost and Rapid Method for Determination of Five Flavonoids in Herbal Medicines with Rutin by HPLC-UV at Equal Absorption Wavelength. *Curr. Anal. Chem.* **2025**, *21*, 124–133. [CrossRef]

7. Yang, Y.L.; Zhang, Y.; Alimujiang, S.; Guan, Z.Q.; Zhang, Y.Z.; Li, W.L. Portable NIR Spectroscopy for On-site Origin Discrimination and Alkaloid Quantification in *Stephania tetrandra* Roots. *Anal. Methods* **2025**, *17*, 7251–7265. [CrossRef]
8. Wang, Y.Q.; Yang, Y.Z.; Sun, H.J.; Dai, J.P.; Zhao, M.X.; Teng, C.Z.; Ke, Z.H.; Yang, M.; Zhong, L.Y.; Zhu, W.F. Application of a Data Fusion Strategy Combined with Multivariate Statistical Analysis for Quantification of Puerarin in *Radix puerariae*. *Vib. Spectrosc.* **2020**, *108*, 103057. [CrossRef]
9. Liu, Y.; Zhang, L.W.; Zhang, X.Z.; Bian, X.H.; Tian, W. Modern Spectroscopic Techniques Combined with Chemometrics for Process Quality Control of Traditional Chinese Medicine: A Review. *Microchem. J.* **2025**, *213*, 113605. [CrossRef]
10. Liu, X.S.; Zhang, S.Y.; Ni, H.F.; Xiao, W.; Wang, J.; Li, Y.R.; Wu, Y.J. Near Infrared System Coupled Chemometric Algorithms for the Variable Selection and Prediction of Baicalin in Three Different Processes. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *218*, 33–39. [CrossRef]
11. Cheng, P.H.; Jiang, X.G.; Liu, Y.D. FT-NIR Spectroscopy Combined with PLS-DA-SVM for Adulteration Detection in Xinyang Maojian Tea. *J. Food Compos. Anal.* **2025**, *148*, 108667. [CrossRef]
12. Chen, H.Z.; Liu, X.K.; Chen, A.; Cai, K.; Lin, B. Parametric-scaling Optimization of Pretreatment Methods for the Determination of Trace/Quasi-trace Elements Based on Near Infrared Spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *229*, 117959. [CrossRef]
13. Hu, L.Q.; Xiong, Y.; Zou, Z.R.; Yang, Y.Z.; He, J.W.; Zhong, L.Y.; Wang, Y.Q.; Yang, M. Identifying the Chemical Markers in Raw and Wine-processed Scutellaria baicalensis by Ultra-performance Liquid Chromatography/quadrupole Time-of-flight Mass Spectrometry Coupled with Multiple Statistical Strategies. *Biomed. Chromatogr.* **2020**, *34*, e4849. [CrossRef]
14. Wei, Q.Z.; Yu, M.H.; An, M.Y.; Li, R. A Study of Flavonoids in Perilla Leaves from Different Areas Based on LC-MS and UV-Vis-NIR Spectroscopy with Multivariate Algorithms. *Microchem. J.* **2026**, *224*, 117470. [CrossRef]
15. Jakubíková, M.; Sádecká, J.; Kleinová, A. On the Use of the Fluorescence, Ultraviolet-Visible and Near Infrared Spectroscopy with Chemometrics for the Discrimination Between Plum Brandies of Different Varietal Origins. *Food Chem.* **2018**, *239*, 889–897. [CrossRef]
16. Zhang, L.; Li, Z.; Bian, X. Rapid and Simultaneous Quantification of Four Components in Radix Scutellariae by Near-Infrared Spectroscopy Combined with Chemometrics. *Microchem. J.* **2025**, *219*, 115836. [CrossRef]
17. Peng, G.; Xue, Q.; Liu, H.; Chen, D.; Liu, Q.; Zhang, S.; Lu, T.; Wu, Y.; Zhang, X.; Yin, F. Construction of a multidimensional quality evaluation system for *Scutellariae radix*: Integrated research on chemical components, biological potency, near-infrared spectroscopy technology, and intelligent algorithms. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2026**, *355*, 128642. [CrossRef]
18. Qiu, T.; Yang, Y.; Sun, H.; Hu, T.; Wang, X.; Wang, Y.; Wu, Z.; Zhong, L.; Zhu, W.; Yang, M. Rapid discrimination and quantification of kudzu root with its adulterant part using FT-NIR and a machine learning algorithm. *Vib. Spectrosc.* **2021**, *116*, 103289. [CrossRef]
19. Zhou, Y.; Deng, P.; Chen, W. Facilely Promoting the Concentration of Baicalin in Polylactic Acid Fiber for UV Shielding and Antibacterial Functions: A Customized and Sustainable Approach. *Materials* **2024**, *17*, 3734. [CrossRef]
20. Taniguchi, M.; LaRocca, C.A.; Bernat, J.D.; Lindsey, J.S. Digital Database of Absorption Spectra of Diverse Flavonoids Enables Structural Comparisons and Quantitative Evaluations. *J. Nat. Prod.* **2023**, *86*, 1087–1119. [CrossRef]
21. Muradov, R.; Fayzullaev, N.; Ergashev, I.; Norkulov, U. The Study of the Phenolic Substances of the Aerial Part of the Plant "Kokamaron" (*Scutellaria leptosiphon* Juz.). In *E3S Web of Conferences*; EDP Sciences: Les Ulis, France, 2023; Volume 431, p. 01061. [CrossRef]
22. Wang, T.; Zheng, Y.; Xu, L.L.; Yun, Y.H. Comprehensive Comparison on Different Wavelength Selection Methods Using Several Near-infrared Spectral Datasets with Different Dimensionalities. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2025**, *331*, 125767. [CrossRef]
23. Zhang, S.J.; Zhang, S.; Gong, X.C.; Qu, H.B. Real-Time Monitoring of the Column Chromatography Process of *Ginkgo biloba* Using Near-Infrared and Raman Spectroscopy Combined with Spectral Fusion Strategy. *Process Biochem.* **2024**, *145*, 50–62. [CrossRef]
24. Feng, D.; Long, R.L.; Li, J.M.; Song, X.M.; Sun, J. Online Monitoring for Extraction of Tibetan Medicine *Meconopsis quintuplinervia* Regel. Based on Near Infrared Spectroscopy Coupled with Chemometrics. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2024**, *321*, 124695. [CrossRef]
25. Beier, K.; Dutschmann, T.M.; Beuerle, T.; Lubienski, M.; Baumann, K. Classification of Horsetails Using Predictive Modelling on NIR Spectra. *J. Chemom.* **2025**, *39*, e3634. [CrossRef]
26. Yang, B.; Zhang, Z.; Song, J.; Qi, T.; Zeng, J.; Feng, L.; Jia, X. Interpreting the Efficacy Enhancement Mechanism of Chinese Medicine Processing from a Biopharmaceutic Perspective. *Chin. Med.* **2024**, *19*, 14. [CrossRef]
27. Yang, J.; Liu, Y.; Li, J.Y.; Chen, P.; Xu, Y.P.; Liu, D.; Chu, X.L. Research Progress and Prospects of Modern Spectral Fusion Analysis Technology. *Prog. Chem.* **2024**, *36*, 1874–1892.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.