Research on Hyperspectral-Based Feature Set Construction and Machine Learning Inversion for Mixed Salts Characteristics in Murals

Wenxuan Lin [1,2], Shuqiang Lyu [1,2], Bo Ning [3], Lihong Li [3], Miaole Hou [4,*]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, No.15 Yongyuan Road, Daxing District, Beijing, China -2108570424046@stu.bucea.edu.cn, lvshuqiang@bucea.edu.cn

[2] Beijing Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring, No.15 Yongyuan Road, Daxing District, Beijing, China

[3] Yungang Research Institute, Datong 037007, China-13233616482@163.com, ningbo007@163.com

[4] Chang'an University, No.126 Yanta Road, Yanta District, Xi'an 710064, Shaanxi, China -houmiaole@bucea.edu.cn

Keywords: Hyperspectral, Mural, Mixed Salts, Feature Set Construction, Gaussian Process Regression.

Abstract:

To enable non-destructive quantitative identification of mixed salts in mural plaster layers, hyperspectral data were collected from Na_2SO_4-$CaCl_2$ mixed-salt samples. Based on these data, a method integrating spectral preprocessing, feature-set construction, and machine-learning inversion was proposed. First, the original spectra were preprocessed using Savitzky-Golay smoothing and multiplicative scatter correction. A 0.6-order fractional-order derivative (FOD) was then introduced to enhance subtle salt-related spectral features. Subsequently, 30 single-band features were selected using a two-step strategy involving competitive adaptive reweighted sampling for preliminary screening and variable importance in projection for secondary screening. On this basis, dual-band and tri-band spectral indices were further constructed, and a combined-band feature set was formed by integrating the three feature sets. Gaussian process regression (GPR) was used to compare the inversion performance of different feature-input strategies for Na_2SO_4 and $CaCl_2$ contents. The results showed that the 0.6-order FOD achieved a favorable balance between feature enhancement and noise suppression. Among the evaluated feature-input strategies, the combined-band model showed the best predictive performance for both Na_2SO_4 and $CaCl_2$. These results indicate that integrating complementary information from feature sets with different dimensions can improve the stability and accuracy of mixed-salt inversion, providing a useful reference for the hyperspectral non-destructive quantification of mixed salts in murals.

1. Introduction

Salt damage is a common issue in cultural heritage conservation, which is particularly prominent in the conservation of murals. The crystallization and migration of salts pose a serious threat to the structural integrity of murals. The synergistic crystallization of various soluble salts often leads to salt damage on the surface of murals, thereby compromising their artistic value and affecting the preservation of historical information (Li et al., 2023). Therefore, accurately identifying and quantifying the salt content in murals is crucial for early damage diagnosis and the implementation of conservation interventions.

Hyperspectral technology offers advantages such as the integration of spectral data and imagery, as well as non-contact detection, enabling the non-destructive quantitative estimation of salt content in murals (Bhargava et al., 2024). In recent years, fractional-order derivatives (FOD) have been widely used for the inversion of plant physiological parameters and the quantitative estimation of soil properties, as they can enhance subtle spectral features in hyperspectral data, improve model sensitivity, and effectively suppress noise. Previous studies have shown that under moist conditions, visible and near-infrared (VIS-NIR) spectra combined with FOD and one-dimensional convolutional neural network (1D-CNN) can be used for the simultaneous estimation of various soil properties (Liu et al., 2023). Furthermore, FOD spectra generally outperform original spectra and integer-order derivative spectra, effectively enhancing target spectral information in water-saturated environments. When using unmanned aerial vehicle (UAV) hyperspectral monitoring to assess secondary salinization in arid oasis cotton fields, FOD can highlight subtle spectral differences against a noisy background, with different orders revealing distinct hidden information. Among these, the 1.9th order is most conducive to salinization feature extraction and model accuracy improvement (Wang et al., 2024).

Although FOD and multidimensional feature set construction have been applied in other fields, traditional methods for salt content inversion in wall paintings primarily rely on a single salt content parameter or simple spectral indices, which lack a systematic evaluation that combines hyperspectral preprocessing with feature set construction. Previous studies using simulated samples of Na_2SO_4-$CaCl_2$ mixed salts from the surfaces of temple murals have found that combining a natural logarithmic transformation with competitive adaptive reweighted sampling (CARS) feature band screening and stepwise regression modelling can improve the accuracy of hyperspectral inversion of mixed salt content (Li et al., 2025). Another study compared methods for salt content inversion in simulated murals. The results indicated that first-order differential preprocessing combined with sample partitioning based on joint x-y distances (KSPXY) subsampling and a partial least squares regression (PLSR) model yields superior predictive performance, suggesting that systematic feature processing and model evaluation can improve the stability and accuracy of salt content inversion in murals (Dong et al., 2025).

Therefore, FOD was introduced in this study to enhance spectral features associated with salt content in mural samples containing mixed Na_2SO_4 and $CaCl_2$ salts. Three feature sets were then constructed, namely a single-band feature set, a dual-band spectral-index set, and a tri-band spectral-index set. In addition, these three feature sets were further integrated to form a combined-band feature set. By comparing the inversion performance of different feature-input strategies using Gaussian process regression (GPR), this study aimed to provide a methodological

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-695-2026

reference for the non-destructive quantification of mixed salts in murals.

2. Materials and methods

2.1 Sample Preparation and Data Collection

To investigate the spectral response characteristics of mural plaster layers under mixed-salt exposure, mural plaster samples containing Na_2SO_4 and $CaCl_2$ were prepared in the laboratory. Each sample consisted of two layers: a coarse plaster layer and a fine plaster layer. The base layer was the coarse plaster, prepared from a mixture of soil, sand, and grass knots. The surface layer was the fine plaster, composed of soil, sand, cotton, and hemp. All raw materials were desalinated, air-dried, and sieved prior to use (Lu et al., 2021).

During sample preparation, the coarse plaster materials were first thoroughly mixed with deionized water and then packed into an iron mold with a diameter of 9 cm and a height of 1.8 cm to form the base layer. Predetermined amounts of Na_2SO_4 and $CaCl_2$ were then dissolved in 20 mL of deionized water, thoroughly mixed with the treated fine plaster materials, and packed onto the top of the base layer to form the surface layer. Using a total sample mass of 100 g per sample, 18 sets of mural plaster samples with different Na_2SO_4-$CaCl_2$ addition levels were prepared. The detailed salt-addition scheme is listed in Table 1. The preparation process of the mural plaster samples is illustrated in Figure 1.

Na_2SO_4 added (g)	$CaCl_2$ added (g)	Na_2SO_4 added (g)	$CaCl_2$ added (g)
0	0	0.45	0.25
0.05	0.02	0.5	0.28
0.1	0.04	0.55	0.3
0.15	0.09	0.6	0.32
0.2	0.12	0.65	0.35
0.25	0.15	0.65	0.45
0.3	0.18	0.7	0.38
0.35	0.2	0.75	0.27
0.4	0.22	0.8	0.16

Table 1. Na_2SO_4-$CaCl_2$ addition scheme for the mixed-salt mural plaster samples.

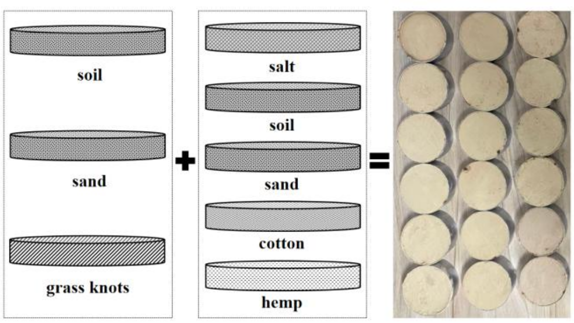


Figure 1. Preparation process of the mural plaster samples.

Hyperspectral data were collected using the FieldSpec 4 field spectrometer manufactured by Analytical Spectral Devices (ASD). Four measurement points were selected for each sample, with four repeated measurements taken at each point. The final average spectrum was used as the representative spectrum for that sample for subsequent analysis and modelling. The main technical specifications of the instrument are listed in Table 2.

Parameter	Specification
Instrument model	ASD FieldSpec 4 HI-Res
Spectral range	350-2500nm
Spectral resolution	3nm(350-1000nm); 8nm(1001-2500nm)
Sampling interval	1.4nm(350–1000nm); 1.1nm(1001–2500nm)
Number of channels	2151

Table 2. Technical specifications of the ASD FieldSpec 4 Hi-Res spectrometer used for reflectance measurements.

2.2 Data Processing Methods

To quantitatively estimate the Na_2SO_4 and $CaCl_2$ contents in mixed-salt mural plaster samples, the data-processing workflow in this study comprised four steps. First, the original spectra were preprocessed using Savitzky-Golay (SG) smoothing, multiplicative scatter correction (MSC), and fractional-order derivative (FOD) transformation. Second, CARS and variable importance in projection (VIP) were used to screen informative single-band features. Third, dual-band and tri-band spectral indices were constructed and further screened using VIP and Pearson correlation coefficient (PCC), respectively. Finally, the single-band feature set, dual-band spectral-index set, tri-band spectral-index set, and their combined feature set were input into the GPR model for training and testing to compare their predictive performance.The overall experimental workflow is presented in Figure 2.

2.2.1 Spectral Preprocessing and Enhancement: Original spectra are susceptible to instrument noise, sample surface conditions, and measurement environment effects. To reduce these interferences, SG smoothing was first applied to suppress high-frequency noise, followed by MSC to correct baseline drift and offset caused by particle-size variation and surface scattering. Subsequently, FOD was introduced to enhance subtle spectral features. For discrete spectral data, FOD was calculated as follows:

$$\frac{d^{\alpha}f(\lambda)}{d\lambda^{\alpha}}=f(\lambda)+(-\alpha)f(\lambda-1)+\frac{(-\alpha)(-\alpha+1)}{2}+\ldots+\frac{\Gamma(-\alpha+1)}{n!(-\alpha+1)}f(\lambda-n) \quad (1)$$

where α is the order, Γ is the gamma function, λ is the wavelength, $f(\lambda)$ is the original spectral function, and n is the number of discrete summands.

2.2.2 Single-Band Feature Selection: To reduce the redundancy of hyperspectral data and identify bands sensitive to mixed-salt variation, a stepwise single-band feature-selection strategy was adopted. First, CARS was used to preliminarily screen the bands associated with Na_2SO_4 and $CaCl_2$ separately from the enhanced spectra. The retained bands for the two salts were then merged by union, and VIP was subsequently applied for secondary screening. After removing variables with low contribution and high collinearity, the final single-band feature set was obtained for subsequent modelling and spectral-index construction.

2.2.3 Construction and Selection of Dual-Band Spectral Indices: A single spectral band is often insufficient to fully characterize the complex spectral information associated with mixed salts. Therefore, dual-band spectral indices were constructed based on the selected single-band features. Difference index (DI), ratio index (RI), and normalized

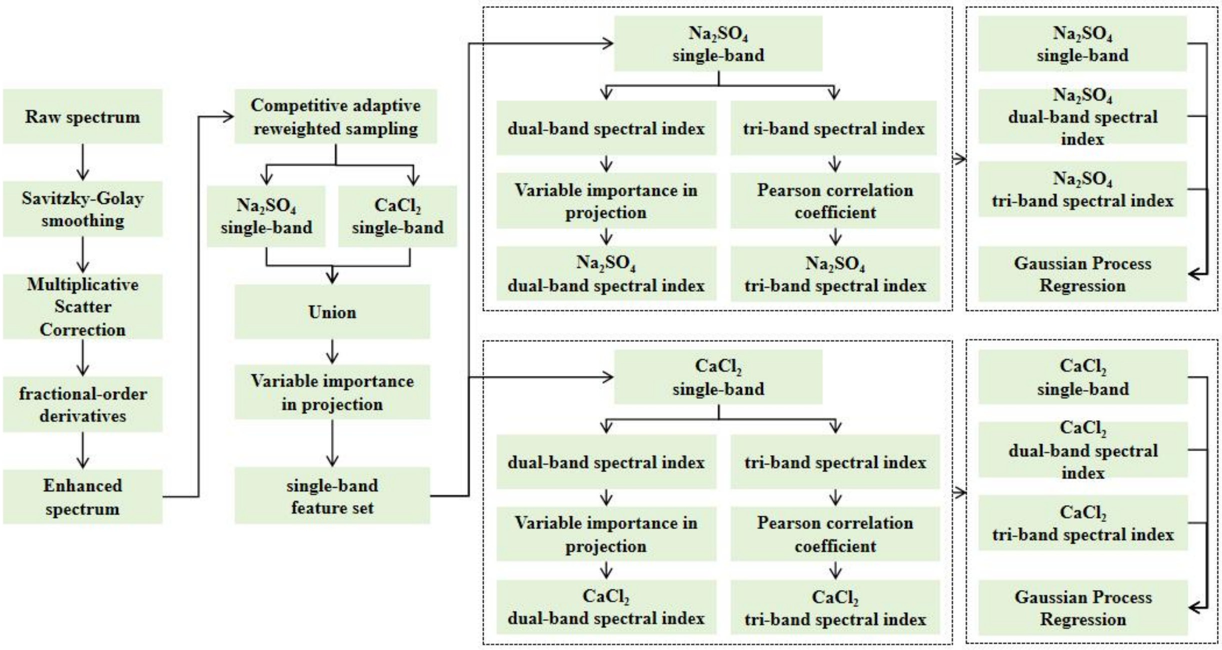


Figure 2. Experimental workflow.

difference index (NDI) were used as dual-band spectral indices, and VIP was further employed to identify the combinations with the highest explanatory power for the target salt contents.

$$DI(R_{\lambda1},\ R_{\lambda2})=R_{\lambda1}-R_{\lambda2} \tag{2}$$

$$NDI(R_{\lambda1},\ R_{\lambda2})=\frac{R_{\lambda1}-R_{\lambda2}}{R_{\lambda1}+R_{\lambda2}} \tag{3}$$

$$RI(R_{\lambda1},\ R_{\lambda2})=\frac{R_{\lambda1}}{R_{\lambda2}} \tag{4}$$

Among these, $R_{\lambda1}$ and $R_{\lambda2}$ represent reflectance at any two bands between 350 and 2500 nm, and $R_{\lambda1} \neq R_{\lambda2}$.

2.2.4 Construction and Selection of Tri-Band Spectral Indices: To further capture more complex inter-band relationships, tri-band spectral indices were constructed based on the selected single-band features. Seven tri-band formulations (TBI_1-TBI_7) were used, and PCC was applied to identify tri-band combinations that were strongly correlated with the target salt contents.

$$TBI_1(R_{\lambda1},\ R_{\lambda2},\ R_{\lambda3})=\frac{R_{\lambda1}}{R_{\lambda2}\times R_{\lambda3}} \tag{5}$$

$$TBI_2(R_{\lambda1},\ R_{\lambda2},\ R_{\lambda3})=\frac{R_{\lambda1}}{R_{\lambda2}+R_{\lambda3}} \tag{6}$$

$$TBI_3(R_{\lambda1},\ R_{\lambda2},\ R_{\lambda3})=\frac{R_{\lambda1}-R_{\lambda2}}{R_{\lambda1}+R_{\lambda3}} \tag{7}$$

$$TBI_4(R_{\lambda1},\ R_{\lambda2},\ R_{\lambda3})=\frac{R_{\lambda1}-R_{\lambda2}}{R_{\lambda1}-R_{\lambda3}} \tag{8}$$

$$TBI_5(R_{\lambda1},\ R_{\lambda2},\ R_{\lambda3})=\frac{R_{\lambda1}+R_{\lambda2}}{R_{\lambda3}} \tag{9}$$

$$TBI_6(R_{\lambda1},\ R_{\lambda2},\ R_{\lambda3})=\frac{R_{\lambda1}-R_{\lambda2}}{[(R_{\lambda1}-R_{\lambda2})-(R_{\lambda2}-R_{\lambda3})]} \tag{10}$$

$$TBI_7(R_{\lambda1},\ R_{\lambda2},\ R_{\lambda3})=(R_{\lambda1}-R_{\lambda2})-(R_{\lambda2}-R_{\lambda3}) \tag{11}$$

Among these, $R_{\lambda1}$, $R_{\lambda2}$, and $R_{\lambda3}$ represent reflectance at any three bands within the 350 to 2500 nm range, and $R_{\lambda1} \neq R_{\lambda2} \neq R_{\lambda3}$.

2.2.5 Multidimensional Feature Fusion and GPR modelling: To compare the effects of different feature-input strategies on mixed-salt inversion, GPR models were established for Na_2SO_4 and $CaCl_2$ using the single-band feature set, dual-band spectral-index set, tri-band spectral-index set, and their combined feature set as input variables, respectively (Zhang et al., 2025). The samples were divided into training and testing sets at a ratio of 8:2, with the training set used for model fitting and the testing set for independent validation. Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and residual predictive deviation (RPD) to analyze the effects of different feature-input strategies on model fitting and predictive performance.

3. Results

3.1 Spectral Preprocessing and Enhancement of Mixed Salts

As shown in Figure 3(a), the original hyperspectral curves exhibit a similar undulating trend overall. As shown in Figure 3(b), after SG smoothing, local high-frequency noise is effectively filtered out without compromising the key absorption features of the spectrum. As shown in Figure 3(c), after MSC processing, the spectral curve exhibits a more concentrated overall distribution, indicating that scattering effects caused by differences in sample particle size and surface roughness have been partially corrected, and the spectral baseline has become more stable.

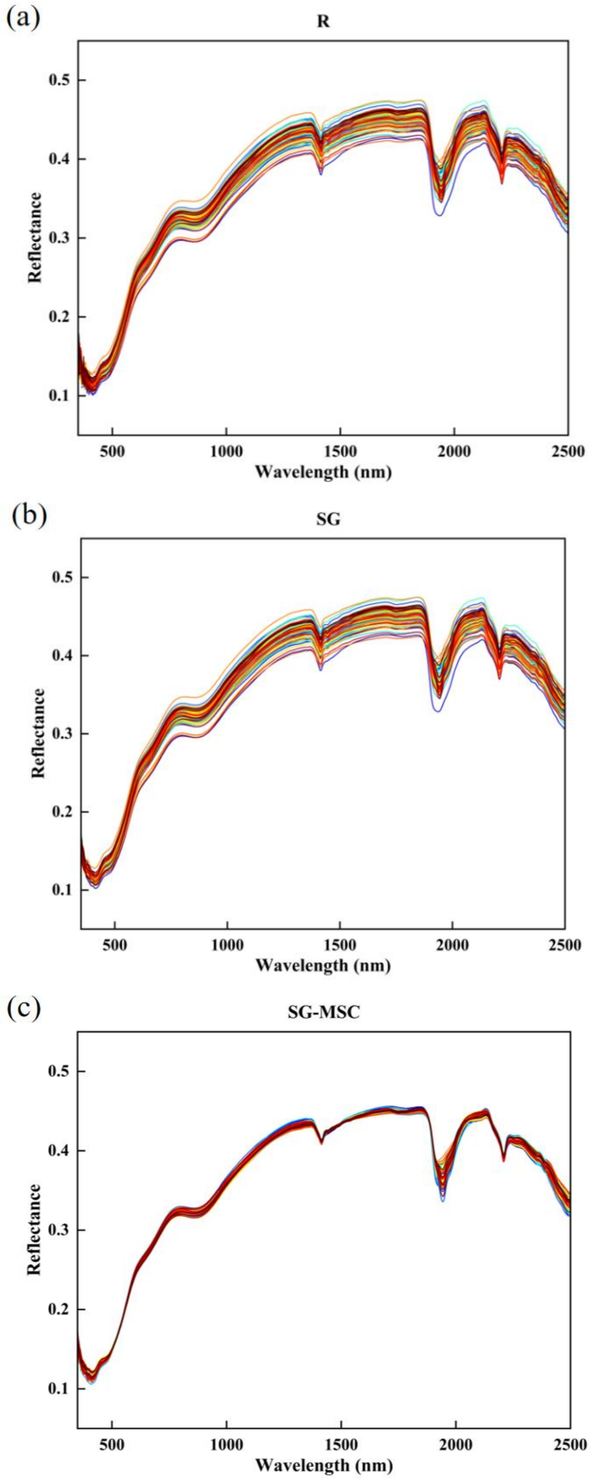


Figure 3. (a) Original spectral curve; (b) Curve after SG smoothing; (c) Curve after SG-MSC processing

To further enhance the subtle spectral features related to mixed salts, fractional-order derivative (FOD) processing was applied to the SG-MSC-preprocessed spectra over the order range from 0.1 to 2.0 at increments of 0.1. After comprehensively comparing the spectral enhancement effects and subsequent modelling performance of different fractional orders, the 0.6-order FOD showed the best overall performance in feature enhancement and salt-content inversion. Considering the page limitation of the main text, only the spectra processed with the 0.6-order FOD are presented in Fig. 4.

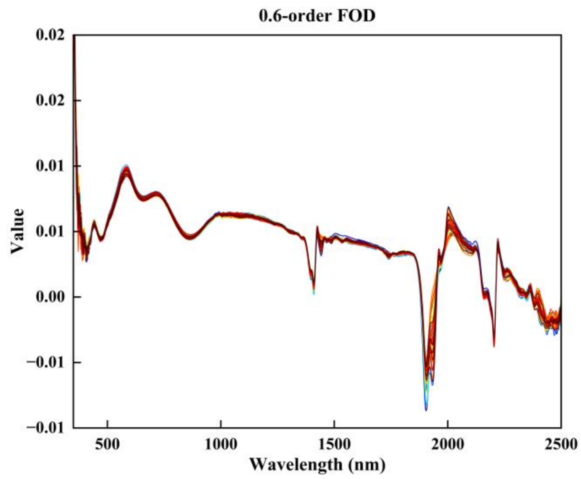
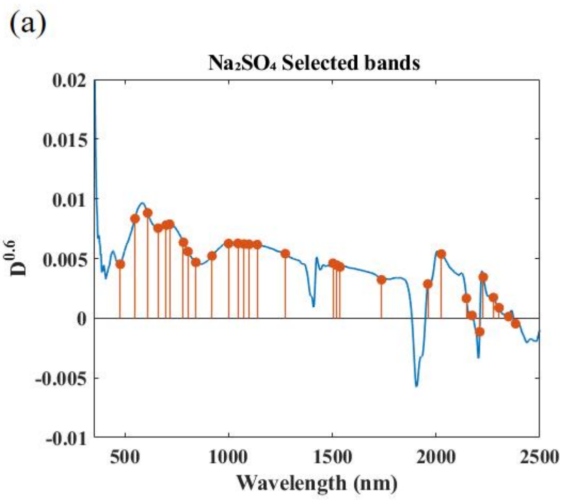


Figure 4. Spectra processed with the 0.6-order fractional-order derivative.

3.2 Single-Band Feature Selection Using CARS and VIP

To improve the inversion accuracy of the machine-learning models, a stepwise single-band feature-selection strategy was adopted. Specifically, CARS was first used for preliminary screening of bands associated with Na_2SO_4 and $CaCl_2$. The retained bands were then merged by union, and VIP was subsequently applied for secondary screening.

3.2.1 Preliminary Screening of Single-Band Features Using CARS: In this study, CARS was first used to identify the single-band features of Na_2SO_4 and $CaCl_2$ separately. As shown in Figure 5(a), the characteristic bands of Na_2SO_4 are primarily distributed in the visible–near-infrared region (500–900 nm) and near the absorption valleys at 1900 nm and 2200 nm. These regions are associated with the characteristic absorption of hydrated salts in the near-infrared spectrum. As shown in Figure 5(b), the characteristic bands of $CaCl_2$ are primarily concentrated around 950–1350 nm and exhibit a pronounced response near water absorption bands such as 1400 nm. This may be related to the strong hygroscopicity of $CaCl_2$ and its hydration behavior.



https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-695-2026 | © Author(s) 2026. CC BY 4.0 License.

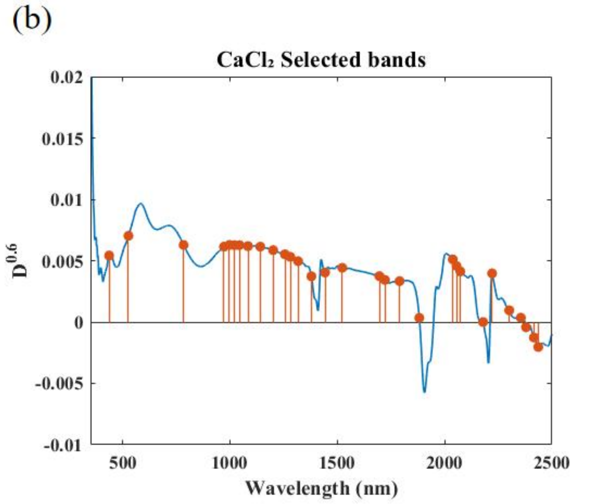


Figure 5. (a) Single band selected by Na_2SO_4; (b) Single band selected by $CaCl_2$

These results indicate that the sensitive spectral regions of the two salts were not completely consistent in the mixed system. Therefore, both sets of sensitive bands should be considered in subsequent mixed-salt feature construction.

3.2.2 Union of Salt-Specific Single-Band Features and Secondary Screening by VIP: To account for the possible co-crystallization and interaction between Na_2SO_4 and $CaCl_2$ in the mural plaster system, the sensitive bands identified for the two salts were merged by union after duplicate bands were removed. The merged band set was then subjected to secondary screening using VIP. The results showed that the top 30 bands effectively captured the key spectral information of the mixed-salt system. Therefore, these bands were retained as the final single-band feature set for subsequent modelling and spectral-index construction.

3.3 Construction and Selection of Dual-Band and Tri-Band Spectral Indices

To further amplify the spectral-response differences between Na_2SO_4 and $CaCl_2$, dual-band and tri-band spectral indices were constructed by mathematically combining the selected single-band features.

3.3.1 Construction and Selection of Dual-Band Spectral Indices: For dual-band feature construction, difference index (DI), ratio index (RI), and normalized difference index (NDI) were generated by combining the selected single-band features. VIP-based secondary screening showed that, for Na_2SO_4, combinations involving bands near 2179 nm and 2349 nm had relatively high explanatory power, whereas for $CaCl_2$, combinations involving bands near 2179 nm and 1882 nm performed well among the dual-band candidates. Based on the screening results, the top 10 dual-band indices were retained to form the dual-band feature set.

These results indicate that dual-band indices introduce relative information between bands beyond that provided by single-band features. This can reduce the influence of local noise or background fluctuations to some extent and provide richer feature representations for subsequent modelling.

3.3.2 Construction and Selection of Tri-Band Spectral Indices: Compared to dual-band indices, tri-band spectral indices incorporate an additional band into the calculation, allowing them to capture more complex inter-band relationships to a certain extent. As shown in Figure 6(a) and (b), darker colors indicate higher correlation. For Na_2SO_4, the screened combinations such as TBI_2, TBI_3, and TBI_7 exhibited high correlation with the target salt content. For $CaCl_2$, the constructed combinations such as TBI_2 and TBI_3 similarly displayed a large area of high correlation. Based on the PCC results, the tri-band combinations with the highest correlations were retained as the tri-band feature set for subsequent modelling.

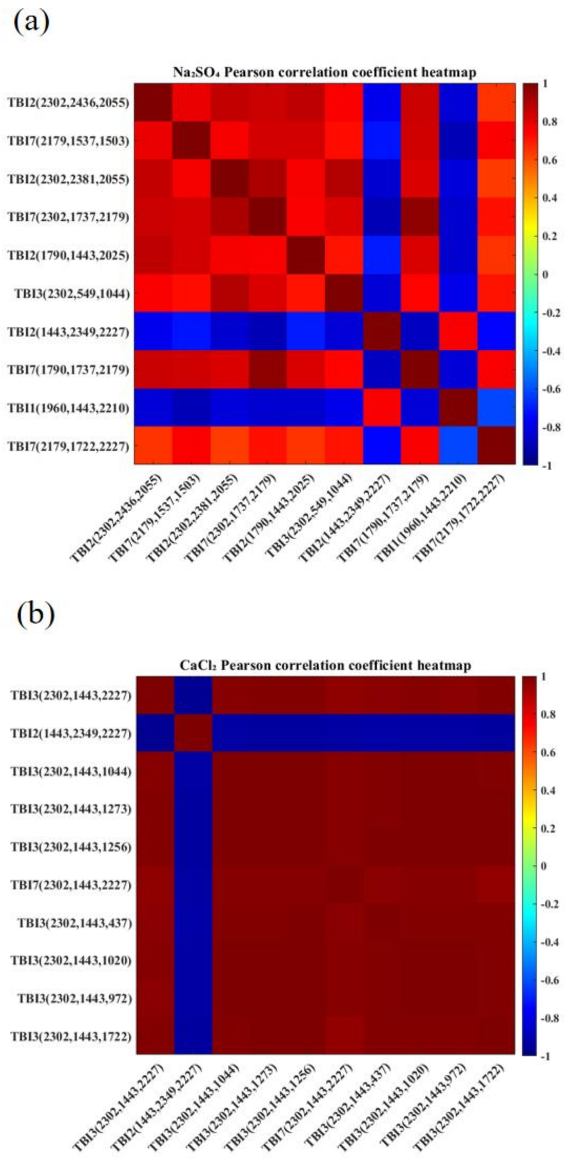


Figure 6. Pearson correlation coefficient heatmap: (a) tri-band spectral indices of Na_2SO_4, (b) tri-band spectral indices of $CaCl_2$

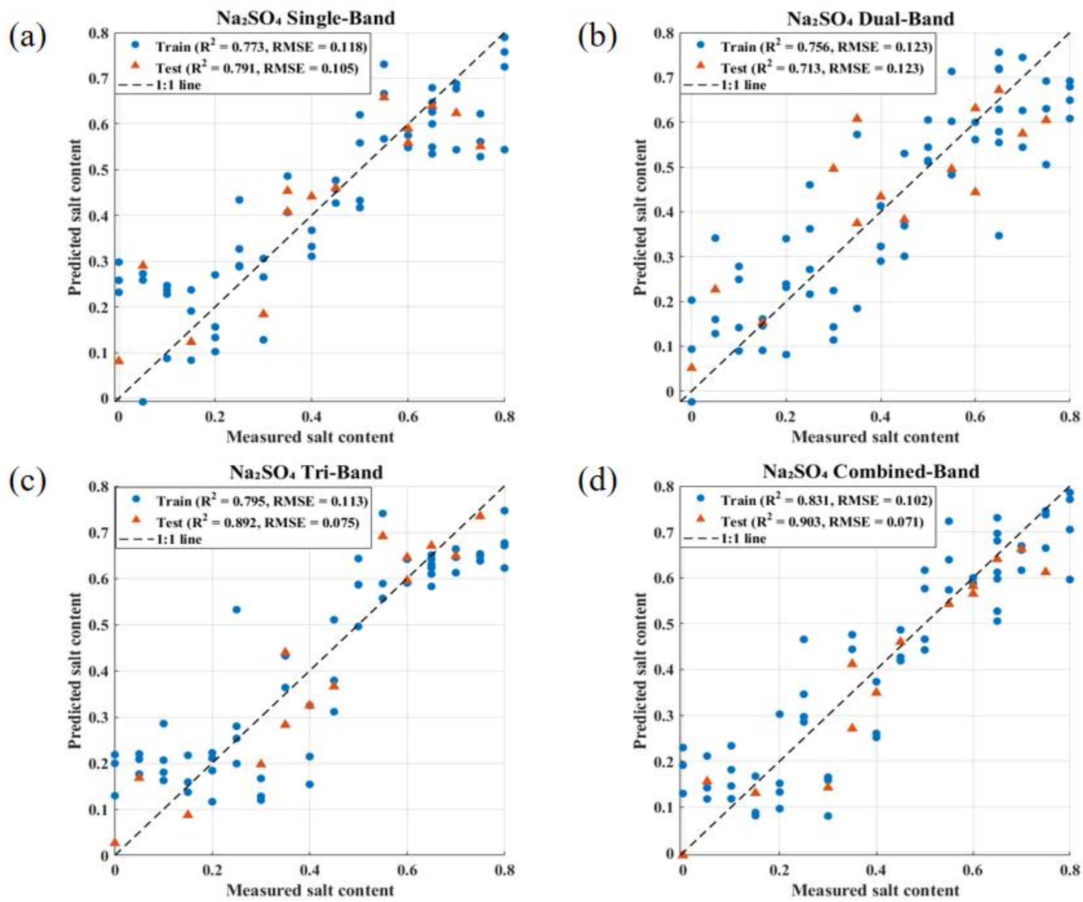
These results indicate that by incorporating additional bands into the calculation beyond single-band features, the tri-band index can integrate information from different bands to some extent, thereby enhancing the correlation between the features and the target salt content. However, this enhanced correlation is primarily reflected at the feature representation level. Whether it can be further translated into higher model prediction accuracy requires evaluation in conjunction with subsequent modelling results.

3.4 Comparison of the Effects of Different Feature Sets on GPR Inversion Accuracy

To systematically evaluate the effects of different feature-input strategies on inversion performance, the single-band feature set, dual-band spectral-index set, tri-band spectral-index set, and combined-band feature set were input into the GPR model for training and testing.The scatter plots and detailed accuracy metrics of the GPR models are shown in Figure 7 and Table 3, respectively.

3.4.1 Comparison of the Accuracy of Na_2SO_4 Inversion Models: For Na_2SO_4 prediction, the combined-band GPR model achieved the best overall performance, with RMSE = 0.0713, R^2 = 0.9029, and RPD = 3.3301. The tri-band model ranked second (RMSE = 0.0764, R^2 = 0.8916, RPD = 3.1515), followed by the single-band model (RMSE = 0.1045, R^2 = 0.7915, RPD = 2.2726), whereas the dual-band model showed the lowest performance (RMSE = 0.1226, R^2 = 0.7130, RPD = 1.9372). These results indicate that, for Na_2SO_4 inversion, integrating feature sets from different dimensions improved predictive performance more effectively than using any individual feature set alone.

3.4.2 Comparison of the Accuracy of $CaCl_2$ Inversion Models: For $CaCl_2$ prediction, the combined-band GPR model also showed the best predictive performance, with RMSE = 0.04593, R^2 = 0.8359, and RPD = 2.5619. The single-band model ranked second (RMSE = 0.0514, R^2 = 0.7948, RPD = 2.2907), followed by the tri-band model (RMSE = 0.0536, R^2 = 0.7765, RPD = 2.1952), whereas the dual-band model again showed the lowest performance (RMSE = 0.0607, R^2 = 0.7135, RPD = 1.9386). These results suggest that, for $CaCl_2$ inversion, integrating features from different dimensions provided more effective predictive information than any individual feature set.



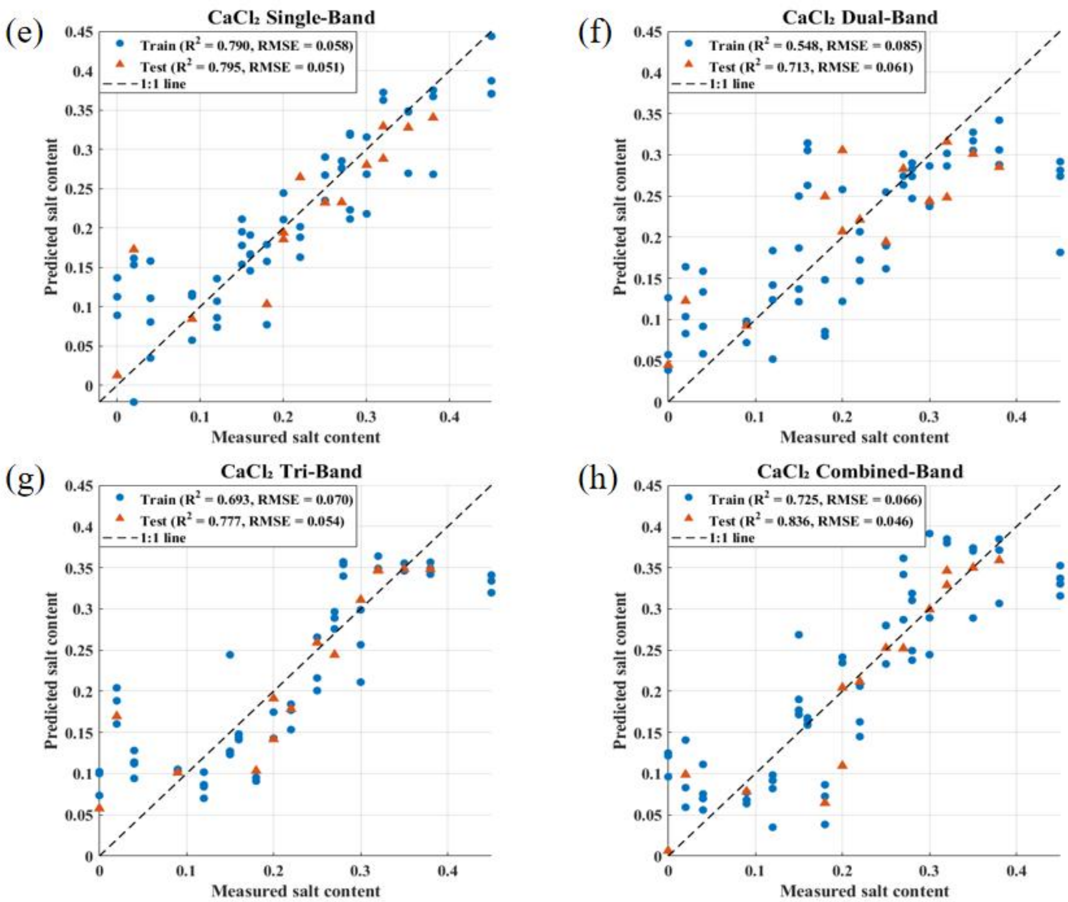


Figure 7. Scatter plot of the predictive power of feature sets from different dimensions in the GPR model

Salt	Feature Set	RMSE	R^2	RPD
Na_2SO_4	single-band	0.1045	0.7915	2.2726
	dual-band	0.1226	0.7130	1.9372
	tri-band	0.0764	0.8916	3.1515
	combined-band	0.0713	0.9029	3.3301
$CaCl_2$	single-band	0.0514	0.7948	2.2907
	dual-band	0.0607	0.7135	1.9386
	tri-band	0.0536	0.7765	2.1952
	combined-band	0.04593	0.8359	2.5619

Table 3. Performance of GPR models for Na_2SO_4 and $CaCl_2$ based on different feature sets.

Taken together, the results for Na_2SO_4 and $CaCl_2$ indicate that the combined-band feature set achieved the best overall predictive performance in the GPR model. This suggests that integrating complementary feature information from different dimensions can improve mixed-salt inversion more effectively than using single-band, dual-band, or tri-band features alone.

4. Discussion

4.1 The Effect of Fractional Order Differentiation on the Spectral Characteristics of Mixed Salts

The results indicate that 0.6-order fractional differentiation performs well in extracting spectral features of mixed salt in murals. Compared with integer-order differentiation, although first- or second-order differentiation can mitigate the effects of baseline drift, it is also more likely to amplify high-frequency noise, thereby affecting the identification of weak salt signals. In contrast, the 0.6-order fractional derivative preserves the overall morphology of the original spectrum while effectively suppressing slowly varying background information and enhancing subtle differences in local spectral bands, thereby achieving a suitable balance between feature enhancement and noise control.

The 30 characteristic bands identified through CARS and VIP screening are primarily concentrated around 1400 nm, 1900 nm, and 2200 nm. The absorption features near 1400 nm and 1900 nm are primarily associated with O-H vibrational modes, reflecting changes in adsorbed water, crystalline water, or structural water within the samples. The spectral response near 2200 nm, in addition to being related to hydroxyl-related absorption, may also be influenced by the mineral composition of the substrate and its chemical bond vibrations. Due to differences in the crystallization, hygroscopicity, and hydration behavior of Na_2SO_4 and $CaCl_2$ in the mural materials, these differences may manifest as moisture-related absorption features in the near-infrared region. After fractional differential processing, these subtle changes were enhanced, thereby providing a clearer spectral basis for subsequent mixed salt feature extraction and modelling.

4.2 The Impact of Spectral Characteristics with Different Feature-Input Strategies on Inversion Results

Spectral features with different dimensional structures played different roles in the GPR models. Single-band features are simple in form and rely on reflectance information from individual wavelengths. Dual-band indices introduce relative relationships between two bands and can reduce the influence of local fluctuations to some extent. Tri-band indices further incorporate a third band, allowing more complex inter-band relationships to be represented. On this basis, the combined-band feature set integrates complementary information from single-band, dual-band, and tri-band features, thereby providing a richer representation of mixed-salt spectral characteristics.

The modelling results showed that the combined-band feature set achieved the best predictive performance for both Na_2SO_4 and $CaCl_2$. For Na_2SO_4, the combined-band model outperformed the single-band, dual-band, and tri-band models in terms of RMSE, R^2, and RPD. For $CaCl_2$, the combined-band model likewise showed the highest predictive accuracy, whereas the dual-band model remained the least effective. These results indicate that, under the present experimental conditions, integrating feature information from different dimensions is more effective than using any individual feature set alone. Therefore, for mixed-salt inversion, the key is not merely to increase feature dimensionality, but to combine complementary feature information in a way that improves predictive stability and accuracy.

5. Conclusions

This study developed a non-destructive quantitative inversion framework for Na_2SO_4-$CaCl_2$ mixed salts in mural plaster based on hyperspectral preprocessing, stepwise feature selection, spectral-index construction, feature-set integration, and Gaussian process regression. The main conclusions are as follows:

(1) Based on SG-MSC preprocessing, 0.6-order fractional differentiation is effective in enhancing the spectral features of mixed salts. This method mitigates the effects of baseline drift while effectively preserving the original spectral information and enhancing the characteristic responses near 1400 nm, 1900 nm, and 2200 nm.

(2) A stepwise single-band feature-selection strategy was adopted, in which CARS was first used for preliminary screening of Na_2SO_4- and $CaCl_2$-related bands, the retained bands were then merged by union, and VIP was subsequently applied for secondary screening. Based on this procedure, 30 sensitive bands were identified for mixed-salt inversion, reducing data redundancy and providing effective spectral information for subsequent modelling.

(3) Different feature-input strategies played different roles in mixed-salt inversion. Although single-band, dual-band, and tri-band feature sets each captured useful spectral information, the combined-band feature set consistently showed better predictive performance than any individual feature set. This indicates that integrating complementary information from different feature dimensions is more effective than relying on a single feature type alone.

(4) Among the GPR models developed in this study, the combined-band feature set achieved the best predictive performance for both Na_2SO_4 and $CaCl_2$. For Na_2SO_4, the RMSE, R^2, and RPD values were 0.0713, 0.9029, and 3.3301, respectively. For $CaCl_2$, the corresponding values were 0.04593, 0.8359, and 2.5619. These results demonstrate that combining feature sets from different dimensions can improve the stability and accuracy of mixed-salt inversion.

In summary, the proposed framework provides a feasible approach for the non-destructive quantification of mixed salts in mural plaster and highlights the importance of integrating complementary spectral features for improving inversion performance.

Acknowledgements

This work was supported by the Open Project of the Chinese Academy of Cultural Heritage, "Salt Spectral Response Mechanism and Quantitative Retrieval Method of Typical Temple Murals - Taking Cheng'en Temple in Beijing and Yiju Temple in Shanxi Province as Examples". It was also supported by the National Natural Science Foundation of China (No.42171356 and No.42171444).

References

Bhargava, A., Sachdeva, A., Sharma, K., Alsharif, M.H., Uthansakul, P., Uthansakul, M., 2024: Hyperspectral imaging and its applications: A review. Heliyon, 10, e33208. doi.org/10.1016/j.heliyon.2024.e33208.

Dong, Q., Lyu, S., Hou, M., Jin, Y., Li, X., 2025: An inversion approach for salt content in simulated murals based on spectral enhancement and sample partitioning. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-M-9–2025, 415–422. doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-415-2025.

Li, F., Liu, H., Yang, H., Ayyamperumal, R., Liu, Y., 2023: Influence of water vapor sorption of NaCl-Na_2SO_4-containing mural earthen plaster and preliminary deterioration analysis. Total Environ. Res. Themes, 6, 100040. doi.org/10.1016/j.totert.2023.100040.

Li, S., Lyu, S., Hou, M., Lin, W., Meng, F., 2025: Hyperspectral prediction model for mixed salts on the surfaces of temple murals in northern China based on simulated samples. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-M-9–2025, 835–842. doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-835-2025.

Liu, Y., Lu, Y., Chen, D., Zheng, W., Ma, Y., Pan, X., 2023: Simultaneous estimation of multiple soil properties under moist conditions using fractional-order derivative of vis-NIR spectra and deep learning. Geoderma, 438, 116653. doi.org/10.1016/j.geoderma.2023.116653.

Lu, Q., Si, W., Wei, L., Li, Z., Xia, Z., Ye, S., Xia, Y., 2021: Retrieval of water quality from UAV-borne hyperspectral imagery: A comparative study of machine learning algorithms. Remote Sens., 13, 3928. doi.org/10.3390/rs13193928.

Wang, Z., Ding, J., Tan, J., Liu, J., Zhang, T., Cai, W., Meng, S., 2024: UAV hyperspectral analysis of secondary salinization in arid oasis cotton fields: Effects of FOD feature selection and SOA-RF. Front. Plant Sci., 15, 1358965. doi.org/10.3389/fpls.2024.1358965.

Zhang, X., Sun, X., Lin, Z., 2025: Improving soil moisture prediction using Gaussian process regression. Smart Agric. Technol., 11, 100905. doi.org/10.1016/j.atech.2025.100905.