ESTIMATION OF THE SOLUBLE SALT CONCENTRATION IN MURALS BASED ON SPECTRAL TRANSFORMATION AND FEATURE EXTRACTION MODELLING**

Z. Q. Guo, S. Q. Lyu, M. L. Hou*

School of Geomatics and Urban Spatial Informatics,
Beijing University of Civil Engineering and Architecture, Beijing, China;
Beijing Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring,
Beijing, China; e-mail: 2108570021067@stu.bucea.edu.cn

Affected by temperature and humidity in the environment, salt crystals expand and accumulate on the surface of murals, causing the pigment layer to peel off, which damages the mural. A method for the rapid and nondestructive detection of salt content in murals using hyperspectral techniques is proposed. The spectral data from mural samples were collected with a spectroradiometer and preprocessed by removing breakpoints with Savitzky–Golay smoothing. The raw spectra were subjected to continuum removal, logarithm of the reciprocal (LR) processing, multiple scattering correction, and standard normal variate transformation combined with first-order differentiation (FD) and second-order differentiation processing to obtain 15 transformed spectra of different forms. The spectra of samples at different concentration levels were classified, and the characteristic wavelengths were extracted using sample set partitioning based on the joint X–Y distance and successive projection algorithm. The pearson correlation coefficient and variable importance in the projection were used for comparison. Salt concentration estimation models were developed using partial least squares regression (PLSR), support vector regression (SVR), and a random forest (RF) model. The slopes of fit were calculated and compared. The results showed that the reflectance spectra decreased and then increased with increasing salt concentration. The accuracy of RF and SVR was better than that of PLSR, and the R_c^2, RMSEc, and RPDc values of the RF-LR-FD model were 0.9703, 0.0466, and 16.8350, respectively. Spectral analysis combined with machine learning models has potential for the nondestructive detection of salt in murals.

Keywords: *mural disruption, salt content, spectral transformation function, machine learning.*

МЕТОД НЕРАЗРУШАЮЩЕГО ОПРЕДЕЛЕНИЯ СОДЕРЖАНИЯ СОЛЕЙ НА ФРЕСКАХ С ИСПОЛЬЗОВАНИЕМ РЕГИСТРАЦИИ ГИПЕРСПЕКТРАЛЬНОГО ИЗОБРАЖЕНИЯ И МНОГОПАРАМЕТРИЧЕСКИХ МЕТОДОВ ЕГО ОБРАБОТКИ

Z. Q. Guo, S. Q. Lyu, M. L. Hou*

УДК 535.317.1

Школа геоматики и городской пространственной информатики Пекинского университета строительства и архитектуры, Пекин, Китай;
Главная лаборатория тонкой реконструкции и мониторинга состояния архитектурного наследия, Пекин, Китай; e-mail: 2108570021067@stu.bucea.edu.cn

(Поступила 31 июля 2022)

Предложен метод быстрого и неразрушающего определения содержания солей в росписях с использованием гиперспектральных изображений. Спектральные данные образцов настенной росписи собраны с помощью спектрорадиометра и предварительно обработаны путем удаления точек излома с помощью сглаживания Савицкого–Голея. Необработанные спектры подвергались удалению континуума, логарифмированию обратной (LR) обработки, коррекции множественного рассеяния

** Full text is published in JAS V. 90, No. 5 (http://springer.com/journal/10812) and in electronic version of ZhPS V. 90, No. 5 (http://www.elibrary.ru/title_about.asp?id=7318; sales@elibrary.ru).

и преобразованию стандартной нормальной переменной в сочетании с дифференцированием первого и второго порядка для получения 15 преобразованных спектров различных форм. Спектры образцов при различных уровнях концентрации классифицированы, характеристические длины волн извлечены с использованием разделения набора образцов на основе совместного расстояния X–Y и алгоритма последовательного проецирования. Для сравнения использованы коэффициент корреляции Пирсона и переменная в проекции. Модели оценки концентрации соли разработаны с использованием частичной регрессии наименьших квадратов (PLSR), регрессии опорных векторов (SVR) и модели случайного леса (RF). Наклоны подгонки рассчитаны и сопоставлены. Показано, что интенсивность спектров отражения уменьшалась, а затем увеличивалась с ростом концентрации соли. Точность RF и SVR лучше, чем PLSR, для модели RF-LR-FD R_C^2 = 0.9703, RMSE = 0.0466 и RPD = 16.8350.

Ключевые слова: *разрушение фрески, содержание солей, функция спектрального преобразования, машинное обучение.*

Introduction. Ancient murals, which represent a balance between research and artistic value, are important cultural and religious treasures from different historical periods [1]; however, they are affected by water and salt transport in the environment, which is not conducive to their survival. Soluble Na_2SO_4 in murals changes with temperature and humidity in the environment, repeatedly dissolving and swelling, which destroys the structural layer of the murals; additionally, salt crystals accumulate on the pigment surface, causing irreversible damage such as peeling [2, 3]. Existing salt tests are performed by field sampling followed by drenching analysis using ion chromatography analysers [4]. However, the costs of equipment and personnel are high, and the sampling and analysis processes generate contamination as well as damage to murals.

Compared to traditional methods, hyperspectral remote sensing, which can be used to predict the salt content in murals, provides a low-cost, flexible, and fast detection method without reagents [5] and is highly efficient and nondestructive. Spectral analysis techniques have proven to be effective tools that are suitable for salt detection in vegetation and soil. García-Martín et al. [6] applied standard normal variate (SNV) transformation, multiplicative scattering correction (MSC), and derivative transformation to the reflectance spectra of Jatropha roots to differentiate the rootstocks based on high or low Cu concentration. Amer et al. [7] applied the normalized pigment chlorophyll ratio index and pigment-specific simple ratio to monitor trace metal salt concentrations in different wheat growth states and confirmed the feasibility of using hyperspectral indices to monitor the strong correlation between salt concentrations and vegetation spectra. Mashimbye et al. [8] established an empirical relationship between soil conductivity and spectral characteristics in South Africa using a single band, normalized difference salinity index, partial least squares, and bagging partial least squares methods. Tan et al. [9] performed random forest (RF) and support vector machine (SVM) modelling prediction after the augmented transformation of farmland spectra from Xuzhou, Jiangsu, China, and the soluble salt concentrations of Zn, Cr, As, and Pb were predicted.

However, salt on a mural differs greatly from salt in natural soil, the mural production process is complex, the mural surface is smoother and the spectral mixing between the salt in mural pigment layers is different from that between the various constituents in the ground and the pigment layers. Therefore, it is necessary to explore the relationship between the reflectance spectra and the salt content in murals. The purpose of this study is to investigate the feasibility of using hyperspectral techniques to predict salt concentrations in mural samples. The specific objectives are to explore the variation relationship between reflectance and the salt content, to analyse the effect of different spectral forms on the salt prediction accuracy after combined transformation, and to evaluate the performance of three regressors, partial least squares regression (PLSR), support vector regression (SVR) and random forest (RF) model, for mural salt concentration prediction after feature band extraction.

Materials and methods. Dunhuang sand was obtained from Dunhuang, Shanxi, China, mixed well with loess, clarified board soil, wheat straw, and hemp rope, moistened with a small amount of water, and then sealed with cling film for 24 h. The mud layer of the mural consisted of 60 moulds filled with mud. Each mould had a diameter of 9 cm and a height of 1.8 cm, and sulfuric acid paper was placed underneath the mud layer. To restore the saline environment of the real mural, anhydrous Na_2SO_4 was weighed at a salt-to-soil ratio of 0–1% and a grade difference of 0.05%. Twenty concentration levels were set, and three samples were prepared at each concentration level, corresponding to 60 mural samples. The salts were completely dissolved in ionized water, mixed well with the sample mud layer, and dried naturally under typical inside conditions during a period when the moisture and temperature of the samples were monitored using a soil

detector MDN-6813. After complete drying, salt-containing mural samples were obtained, and the concentration of Na_2SO_4 was the only variable.

The mural spectral reflectance was determined using an ASD-FieldSpec4HI-RES spectroradiometer (Analytical Spectroscopy Equipment, USA). The performance parameters of the instrument are shown in Table 1. Standard reflectance calibration was performed after instrument warm-up. The sample was padded with black flannel, and the contact probe and its light source were placed vertically upside down on the ventral surface of the sample. The probe was rotated 90° parallel to the surface of the sample, and measurements were repeatedly collected in 4 directions, with the average value used as the measured spectrum. The whole process was performed in a dark room.

TABLE 1. Performance Parameters of the ASD-FieldSpec4 Spectroradiometer

Spectral range, nm	350–2500	Band	2151
Sampling interval, nm	1.4 @ 350–1000 nm; 2 @ 1001–2500 nm	size, cm	12.7×35.6×29.2
Spectral resolution, nm	350–1000 nm @ 3 nm; 1001–2500 nm @ 8 nm	weight, kg	5.44

ASD-FieldSpec4HI-RES measures the radiance of ground objects, which is generally recorded as a digital number (DN). The DN value is influenced by the surface conditions of the real object and the accuracy of the sensor, such as the light environment at the collection site, the radiation resolution of the sensor, and the emissivity of the object. To attenuate the effect of drift error of the sensor response system with the variation in the incidence angle, the acquired DN values of the murals were converted to sample reflectance, and standard reflectance correction was performed every 10 min based on darkroom measurements. The corresponding formula [10] is as follows:

$$R_{\mathrm{M}} = R_{\mathrm{CP}} \frac{\mathrm{DN_M}(\lambda)}{\mathrm{DN_{CP}}(\lambda)}, \tag{1}$$

where R_{M} and R_{CP} are the mural and the calibration panel reflectance, respectively, the latter being 99%, and $\mathrm{DN_M}$ and $\mathrm{DN_{CP}}$ are the DN values of the mural and calibration panel, respectively. The obtained reflectance is optimized by breakpoint elimination and spectral averaging in the subsidiary software. Savitzky–Golay filtering is used to remove the high-frequency burr noise from each reflection spectrum. The filter window is set to 21, and a second-order polynomial is used for spectral smoothing while retaining the linear trend and salt signal characteristics. To improve the spectral signal-to-noise ratio and highlight the salt-sensitive spectral properties, four enhancement transformations of the reflectance spectra are performed – namely, continuum removal (CR), the logarithm of the reciprocal (LR) transformation, MSC, and SNV transformation – and combined with two differential forms – namely, a first-order derivative (FD) and a second-order derivative (SD) – to obtain 15 spectral forms with different feature enhancement strategies. The advantage of the combined transformation strategy is that the effects arising from rough surface undulations, light range scattering, and baseline drift of the sample can be removed, the local spectral response differences can be enhanced, and the specific signals of the component spectra can be amplified. By comparing the modelling accuracy for the 15 spectral forms, the most suitable spectral form for mural salt detection is identified.

Sample-set partitioning based on the joint *X*–*Y* distance (SPXY) was proposed by Galvão et al. [11] as an approach for spectral data segmentation. Wang et al. [12] implemented SPXY for the segmentation of watermelon NIR spectral data and validated the accuracy of the model for predicting the soluble solids content of watermelon. SPXY is used to calculate the spacing between the dependent variable, different salt concentrations in samples (*Y*), and the independent variable, a 350–2500 nm range spectral matrix for samples (*X*). It is ensured that the spectra and salt content have the same weight in sample segmentation to improve the confidence of sample division. After three consecutive days of spectral measurements, 280 spectral curves from 46 samples were obtained with SPXY as the calibration set, and 128 spectral curves from 28 samples were used as the validation set.

Hyperspectral data contain a large amount of background noise, which reduces the processing efficiency and prediction accuracy; additionally, the salt signal is removed from redundant spectral information by band screening. The successive projection algorithm (SPA) is a space vector covariance minimization and forward variable selection method that was proposed by Mario Cesar Ugulino Araujo et al. [13], whereby they extracted bands with the smallest root mean square error by calculating the magnitude of the projection

vector between band pairs. Liang et al. [14] used the SPA method to predict the level of mycotoxin deoxynivalenol (DON) in FHB-infected wheat grains, and the accuracy of the model reached 97%. The Pearson correlation coefficient (PCC) and variable importance in the projection (VIP) are two typical spectral band selection metrics for comparison, which are used to verify the practicability of SPA in the extraction of mural salt spectral bands.

Based on the feature bands extracted under different spectral forms, a classical linear regression (PLSR) model was combined with two nonlinear machine learning models (SVR and RF) and used to evaluate the salt levels in murals. PLSR, which represents the linear relationship between multiple sets of independent and dependent variables projected into a new space [15], combines the advantages of multiple linear regression and principal component regression. Li et al. [16] demonstrated the applicability of PLSR in a statistical analysis of soil AS. SVR is a structural risk minimization-based machine learning model that uses a cross-validation grid search to determine the penalty factor c and obtain a hyperplane with minimal expected risk based on multiple kernel functions for high-dimensional, nonlinear problems [17, 18]. The kernel function selected for this study was the radial basis function (RBF) with a penalty filter $c = 1$. The RF method is a modelling approach consisting of multiple decision trees for classification by voting through bootstrap resampling, where each randomly selected subsample and its corresponding subattributes can be used to generate a regression tree, thus forming a forest, with results obtained by voting [19]. High accuracy in salt concentration monitoring can be achieved using hyperspectral techniques [20]. In this paper, two important parameters were set before using the RF algorithm. One is the number of trees in the forest, which was set to 500. The other is the number of root nodes, which was adjusted based on 10 replicate experiments, was used to select the parameters when the calibration set R^2 was optimal. The values ranged between 3 and 12 for 15 data sets.

Three goodness-of-fit statistics, the coefficient of determination (R^2), root mean square error (RMSE), and relative percent deviation (RPD) were used to evaluate the calibration and validation effects of the model in terms of stability, precision and confidence. The deviation of the slope of fit (SLOPE) from the 1:1 line between the true and predicted concentrations under different conditions was calculated to assist in judging the estimation effect. The larger the R^2 and RPD are, the smaller the RMSE, and the closer to 1 the SLOPE is, the higher the model quality, and vice versa.

Results and discussion. The sample spectra of murals at seven salt concentration levels were selected from the total data, and the sample reflectance curves are shown in Fig. 1. For different salt concentrations, the peak and valley positions and curve profile trends changed minimally, and three asymmetric absorption valleys were observed at approximately 1420, 1940, and 2210 nm. The only significant change associated with the concentration was the magnitude of reflectance. With increasing salt concentration, reflectance first decreased and then increased sharply. Na^+ has a strong ability to absorb water, forming compounds that enhance the absorption of electromagnetic waves, and in high-concentration salt-water transport, the salt crystals gradually become saturated and precipitate, increasing the reflectance [21]. The reflectance is low in the visible-near infrared range at a salt concentration of 0.3% and increases sharply at concentrations higher than 0.6%, with a large difference in reflectance of 0.75 and 0.6% between the two salt concentration samples.

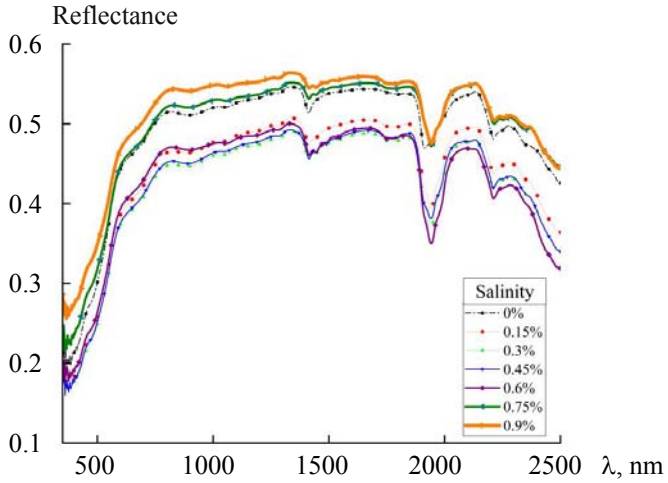

Fig. 1. Spectral reflection curves of mural samples at different salt concentrations.

According to the principle of RMSE minimization, the SPA algorithm extracts different numbers of characteristic bands from 15 spectral forms, and the results are shown in Table 2. The extracted bands are concentrated in the following ranges 350–420, 1880–1950, 1100–1180, and 2080–2150 nm and the corresponding positions can be found at the peaks and valleys in the full spectrum. LR is relatively sensitive to the SPA, and a total of 32 bands are proposed, accounting for 29.4% of all results. The process of characteristic band extraction based on SPA is shown. Figure 2 shows that when the number of SPA-based variables is 12, the RMSE is only 0.0841, the minimum value. Figure 3 shows the labelled positions of the 12 bands extracted based on SPA-LR-FD in the spectrum.

TABLE 2. SPA Feature Wavelength Selection Based on Different Spectral Transformation Strategies

Spectral transformation form	Differential orders	Wavelength range, nm
R	–	1988, 1800, 2085, 1907, 1945
	FD	1915, 1943, 372, 1885
	SD	1918, 1794, 1115, 1667, 402, 2103, 633, 1256, 2451
CR	–	1444, 1907, 403, 1944, 391, 376
	FD	1937, 1914, 1884
	SD	1896, 1730, 351, 359, 372, 350, 376, 354, 368, 1933
LR	–	1738, 1093, 2086, 391, 587, 376, 361
	FD	1422, 2067, 1928, 1830, 402, 1679, 399, 1031, 2456, 1256, 1659, 2282
	SD	1919, 1794, 1115, 2268, 1102, 884, 922, 830, 1641, 1343, 1016, 423, 2156
MSC	–	418, 1946, 1907, 2100, 1100, 1987, 858
	FD	795, 1708, 579, 1681, 384, 2419, 372, 410
	SD	1355, 727, 788, 2152, 660, 922, 2328, 649, 1115
SNV	–	1737, 391, 1397, 2090, 382, 1943
	FD	2109, 1362, 1181, 1314, 1661, 2477
	SD	1794, 369, 1013, 489

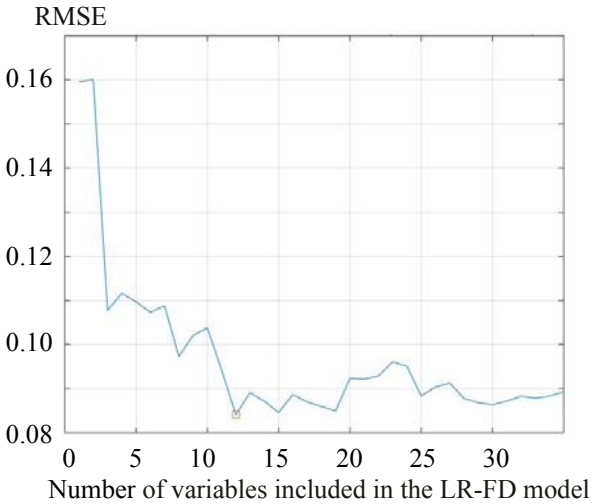

Fig. 2. Relationship between the number of SPA selection variables and the RMSE of the LR-FD model.

The feature bands extracted by SPA were used as independent variables to establish a mapping relationship with the salt concentration. A total of 45 prediction models were developed based on 15 spectral forms using three regressors: PLSR, SVR, and RF. The models with the best accuracy were selected to compare the effects of different spectral transformation strategies on the models and find the most suitable spectral forms

for mural salt detection. The accuracy of PLSR is shown in Table 3. R_C^2 values of 0.9086 and 0.9028 were achieved using the LR-FD and MSC-FD models, respectively. These were the only two groups in the PLSR model that achieved an R_C^2 that exceeded 0.9. For the validation set of MSC-FD, values of 0.9298, 0.0778, and 7.1255 were obtained for R_V^2, RMSE$_V$, and RPD$_V$, respectively, and are the best in each case. This approach was considered the most applicable for PLSR prediction. The SVR accuracy results in Table 4 suggest that the indices after FD or SD treatment were improved compared with those for PLSR. In particular, an R_C^2 of 0.9354 and an RMSE$_C$ of only 0.0802 were achieved using SVR-LR-FD. However, the improvement of SVR in terms of the accuracy of the validation set was not obvious, and there were three sets of data that could be applicable. The accuracy of the applied RF model in Table 5 was higher than that of PLSR and SVR, with an R_C^2 value that exceeded 0.9 for 12 of the 15 sets of spectra. RF-R-SD compensated for the lower R_V^2 of the other regressors and achieved a double-digit RPD$_V$ of 11.6932. RF is more reliable in mural salt estimation, with the calibration set of RF-LR-FD leading to three indicators with the best results among the 45 datasets.

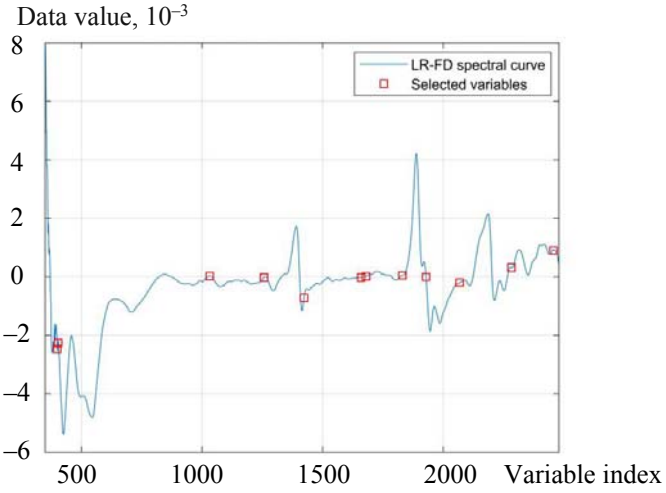

Fig. 3. SPA characteristic band selection results obtained with the LR-FD model.

TABLE 3. PLSR Model and Accuracy Parameters

Spectral transformation form	Differential orders	Calibration set			Validation set		
		R_C^2	RMSE$_C$	RPD$_C$	R_V^2	RMSE$_V$	RPD$_V$
R	–	0.8176	0.1266	2.7406	0.8245	0.1231	2.8497
	FD	0.8413	0.1182	3.1512	0.8901	0.0974	4.5487
	SD	0.8619	0.1101	3.6210	0.8920	0.0965	4.6305
CR	–	0.7969	0.1335	2.4615	0.8241	0.1232	2.8418
	FD	0.8246	0.1241	2.8504	0.8728	0.1048	3.9308
	SD	0.8781	0.1035	4.1002	0.8122	0.1273	2.6630
LR	–	0.7865	0.1369	2.3419	0.8552	0.1118	3.4528
	FD	0.9086	0.0896	5.4687	0.8938	0.0957	4.7087
	SD	0.7888	0.1362	2.3669	0.8758	0.1035	4.0256
MSC	–	0.8838	0.1010	4.3024	0.9047	0.0907	5.2474
	FD	0.9028	0.0924	5.1417	0.9298	0.0778	7.1255
	SD	0.7829	0.1380	2.3036	0.8833	0.1004	4.2836
SNV	–	0.8700	0.1068	3.8456	0.9167	0.0848	6.0032
	FD	0.8579	0.1117	3.5178	0.9003	0.0928	5.0152
	SD	0.2201	0.2616	0.6411	0.4264	0.2225	0.8717

TABLE 4. SVR Model and Accuracy Parameters

Spectral transformation form	Differential orders	Calibration set			Validation set		
		R_C^2	$RMSE_C$	RPD_C	R_V^2	$RMSE_V$	RPD_V
R	–	0.5358	0.2145	1.0771	–1.7260	0.2424	0.1834
	FD	0.8377	0.1197	3.0805	0.8117	0.1178	2.6553
	SD	0.8964	0.0962	4.8280	0.8981	0.0860	4.9068
CR	–	0.7694	0.1449	2.1685	0.7081	0.1261	1.7129
	FD	0.8520	0.1162	3.3784	0.8433	0.0996	3.1908
	SD	0.8940	0.0987	4.7153	0.7290	0.1342	1.8450
LR	–	0.6002	0.1899	1.2506	–0.5220	0.1908	0.3285
	FD	0.9354	0.0802	7.7375	0.8817	0.0881	4.2265
	SD	0.8938	0.1012	4.7092	0.8117	0.1069	2.6553
MSC	–	0.8246	0.1337	2.8501	0.6974	0.1170	1.6523
	FD	0.9156	0.0899	5.9230	0.9084	0.0800	5.4585
	SD	0.8466	0.1233	3.2584	0.8769	0.0941	4.0617
SNV	–	0.8748	0.1099	3.9935	0.8251	0.0968	2.8588
	FD	0.8848	0.1045	4.3395	0.8718	0.0892	3.9002
	SD	0.2926	0.2502	0.7068	–2.6380	0.2688	0.1374

TABLE 5. RF Model and Accuracy Parameters

Spectral transformation form	Differential orders	Calibration set			Validation set		
		R_C^2	$RMSE_C$	RPD_C	R_V^2	$RMSE_V$	RPD_V
R	–	0.8413	0.0943	3.1506	0.7993	0.0953	2.4907
	FD	0.9542	0.0590	10.9266	0.9374	0.0625	7.9911
	SD	0.9594	0.0503	12.3122	0.9572	0.0567	11.6932
CR	–	0.9456	0.0638	9.1844	0.9239	0.0677	6.5720
	FD	0.9617	0.0551	13.0378	0.9519	0.0566	10.3864
	SD	0.9683	0.0492	15.7629	0.9447	0.0569	9.0481
LR	–	0.8321	0.0971	2.9783	0.8825	0.0788	4.2557
	FD	0.9703	0.0466	16.8350	0.9451	0.0576	9.1075
	SD	0.9270	0.0699	6.8521	0.9101	0.0701	5.5599
MSC	–	0.9193	0.0702	6.1920	0.9004	0.0717	5.0191
	FD	0.9354	0.0646	7.7399	0.9166	0.0669	5.9945
	SD	0.9120	0.0714	5.6792	0.8700	0.0779	3.8473
SNV	–	0.9189	0.0724	6.1667	0.9245	0.0640	6.6181
	FD	0.9366	0.0650	7.8827	0.9355	0.0613	7.7471
	SD	0.5197	0.1323	1.0410	0.5752	0.1213	1.1769

The predicted concentration values from the calibration and validation sets obtained for different estimation models were compared with the true concentration values; then, scatter plots were constructed, and the slopes of the fitted lines were calculated. The spectra generated using the LR-FD strategy display high applicability among those obtained with the three estimation methods, and Fig. 4 shows the effects of the three regressors. The SLOPE variable for the SVR regressor is close to the true value of 1 but biased in the 0.4–0.7% concentration prediction. PLSR yields the most pronounced dispersion trend for the scattered points across the full concentration range. Although the SLOPE of the two fits of the RF regressor is slightly lower than that of SVR, SVR is more reliable in terms of the measured and predicted scatter distributions. The four metrics R_C^2, $RMSE_C$, RPD_C, and $SLOPE_C$ are shown to comprehensively assess the model quality. The following conclusions can be drawn from the plots of the four indicators in Fig. 5. The spectra generated using the SNV-SD strategy are unsuitable for all three methods of salt estimation and have low sensitivity to salt transformation. The difference between FD and SD combined with spectral transformation is not obvious, and the high-order differential transformation has little effect on accuracy improvement. LR-FD is a reliable form of spectral transformation for mural salt estimation and is applicable to all three regressors. The SPA feature-based RF-LR-FD model with waveband extraction provides stability, accuracy, and credibility.

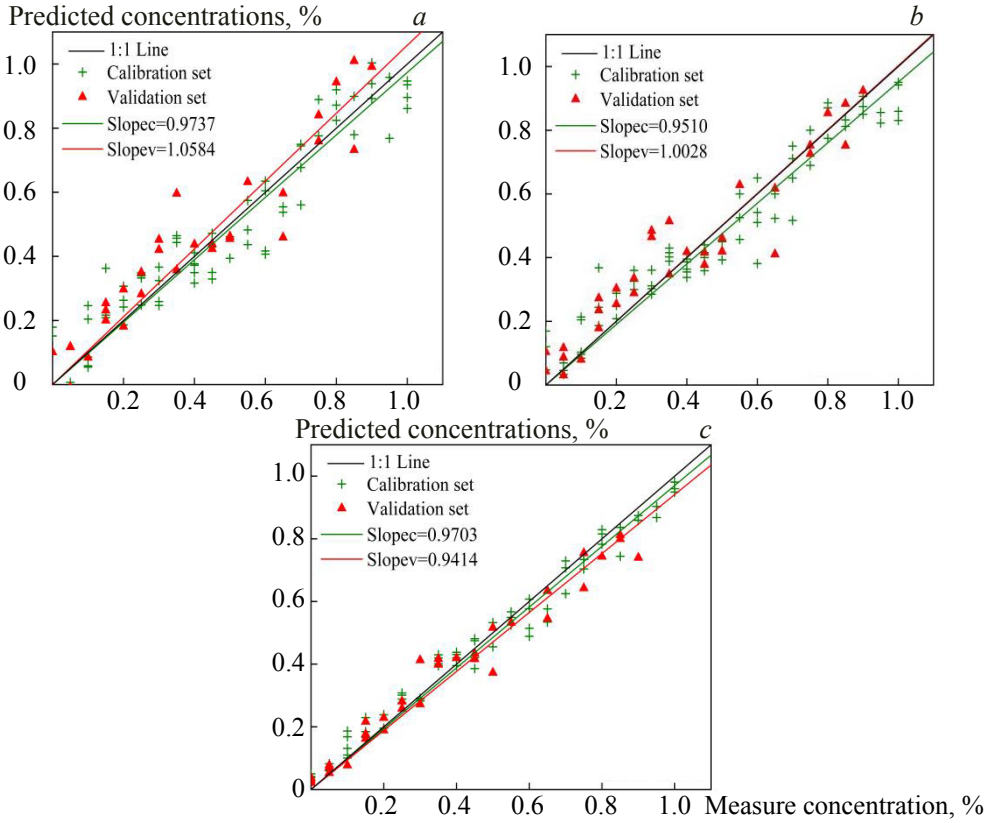

Fig. 4. Scatter plots of measured and predicted salt concentrations using (a) PLSR, (b) SVR, and (c) a RF model for spectra considering the LR-FD form.

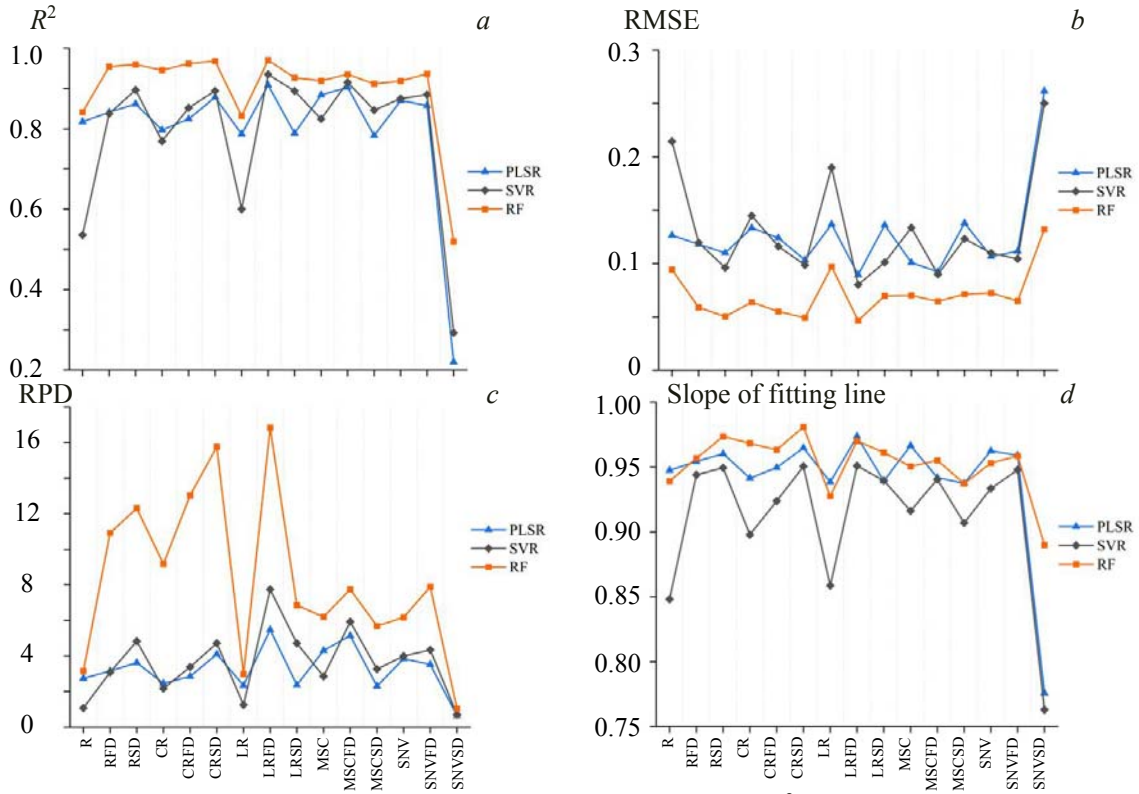

Fig. 5. Comparison of the magnitudes of the four metrics (a) R^2, (b) RMSE, (c) RPD, and (d) SLOPE for the calibration set under different spectral change strategies.

To verify the applicability of SPA, two other variable selection methods, namely, PCC and VIP, were used to extract the salt signal. PLSR is applied to the original output of spectral inversion, and the results are shown in Table 6. All band extraction methods are proven to be effective, and SPA displays the best modelling effect, with comparatively fewer bands extracted.

TABLE 6. Model Comparison Based on Different Spectral Variable Extraction Methods

Variables	Model	Calibration set				Validation set			
		R_C^2	RMSE$_C$	RPD$_C$	SLOPE$_C$	R_V^2	RMSE$_V$	RPD$_V$	SLOPE$_V$
601	Full band	0.6297	0.1803	1.6433	0.8935	0.6611	0.1710	1.7177	0.8629
12	PCC	0.6657	0.1713	1.7294	0.9039	0.6775	0.1668	1.7609	0.8706
11	SPA	0.8508	0.1144	2.5886	0.9571	0.8594	0.1102	2.6665	0.9987
40	VIP	0.6436	0.1769	1.6751	0.8976	0.5582	0.1953	1.5044	0.8206

Conclusions. The highly time-efficient and nondestructive determination of salt accumulation levels in murals is important for assessing the survival status of murals and level of peeling protection. In the spectral range of 350–2500 nm, 45 mural salt concentration prediction models were developed by using 15 enhanced transformation strategies and SPA to extract features and combine three regressors: PLSR, SVR, and RF. A correlation exists between the reflectance spectra and salt concentration, with the first decrease followed by a significant increase in mural reflectance as the salt concentration increases. Through analytical comparison, SPA displayed more advantages than PCC and VIP in the extraction of mural salt feature bands. The RF-LR-FD model yielded high R_C^2 and R_V^2 accuracies of 0.9703 and 0.9451, respectively, indicating that it is suitable for quantitative salt estimation in murals. The overall results confirm the preliminary feasibility of using hyperspectral techniques to determine the salt content of murals in a highly time-efficient and nondestructive manner. Based on this approach, visual assessments and inversions of mural salt distribution can be explored in depth in the future.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Nos. 42171356 and 42171444) and the BUCEA Post Graduate Innovation Project (PG2022111).

REFERENCES

1. J. Cao, Z. Zhang, A. Zhao, et al., *Herit Sci*., **8**, 7–20 (2020), doi: 10.1186/s40494-020-0355-x, https://heritagesciencejournal.springeropen.com/articles/10.1186/s40494-020-0355-x.
2. Z. Zheng, *J. Nat. Disasters*., **29**, 158–163 (2020), doi: 10.13577/j.jnd.2020.0616, http://zrzh.paperonce.org/oa/darticle.aspx?type=view&id=20200616.
3. C. Liu, W. Sun, J. Huang, et al., *Sci. Conserv. Archaeol.*, **32**, 70–81 (2020), http://www.wwbhkgkx.net/wwbh/article/abstract/20200310?st=search.
4. Y. Zhang, Z. Yu, L. Wang, et al., *Dunhuang Res*., **48**, 148–155 (2021), doi: 10.13584/j.cnki.issn1000-4106.2021.01.020, http://www.dhyj.net.cn/index.php?m=content&c=index&a=show&catid=17&id=963.
5. F. Wang, J. Gao, Y. Zha, *ISPRS J. Photogramm. Rem. Sens*., **136**, 73–84 (2018), doi: 10.1016/j.isprsjprs.2017.12.003, https://www.sciencedirect.com/science/article/abs/pii/S0924271617303702?via%3Dihub.
6. J. F. García-Martín, A. T. Badaró, D. F. Barbin, et al., *Processes*, **8**, 823–832 (2020), doi: 10.3390/pr8070823, https://www.mdpi.com/2227-9717/8/7/823.
7. M. Amer, A. Tyler, T. Fouda, et al., *Sci. Pap. Ser. Manag. Ec*., **17**, 47–56 (2017), https://managementjournal.usamv.ro/index.php/scientific-papers/1451-spectral-characteristics-for-estimation-heavy-metals-accumulation-in-wheat-plants-and-grain-1451#spucontentCitation5.
8. Z. E. Mashimbye, M. A. Cho, J. P. Nell, et al., *Pedosphere*, **22**, 640–649 (2012), doi: 10.1016/S1002-0160(12)60049-6, https://www.sciencedirect.com/science/article/abs/pii/S1002016012600496?via%3Dihub.
9. K. Tan, W. Ma, F. Wu, et al., *Environ. Mon. Assess*., **191**, 446–459 (2019), doi: 10.1007/s10661-019-7510-4, https://link.springer.com/article/10.1007/s10661-019-7510-4.
10. O. Mutanga, A. K. Skidmore., *Int. J. Appl. Earth Obs. Geoinf*., **5**, 87–96 (2004), doi: 10.1016/j.jag.2004.01.001, https://www.sciencedirect.com/science/article/abs/pii/S0303243404000029?via%3Dihub.

11. R. K. Galvão, M. C. Arauj, G. E. José et al., *Talanta*, **67**, 736–740 (2005), doi: 10.1016/j.talanta.2005.03.025, https://www.sciencedirect.com/science/article/abs/pii/S003991400500192X?via%3Dihub.
12. S. Wang, P. Han, G. Cui, et al., *Spectrosc. Spectr. Anal*., **39**, 738–742 (2019), doi: 10.3964/j.issn.1000-059303-0738-05, http://www.gpxygpfx.com/CN/10.3964/j.issn.1000-059303-0738-05.
13. M. C. Araújo, T. C. Saldanha, R. K. Galvão, et al., *Chem. Intell. Lab. Syst*., **57**, 65–73 (2001), doi: 10.1016/S0169-7439(01)00119-8, https://www.sciencedirect.com/science/article/abs/pii/S0169743901001198?via%3Dihub.
14. K. Liang, Q. Liu, J. Xu, et al., *J. Appl. Spectrosc*., **85**, 953–961 (2018), doi: 10.1007/s10812-018-0745-y, https://link.springer.com/article/10.1007/s10812-018-0745-y.
15. S. Wold, M. Sjostrom, L. Eriksson, *Chem. Intell. Lab. Syst*., **58**, 109–130 (2001), doi: 10.1016/S0169-7439(01)00155-1, https://www.sciencedirect.com/science/article/abs/pii/S0169743901001551.
16. Z. Li, F. Deng, J. He, et al., *Spectrosc. Spectr. Anal*., **41**, 2872–2878 (2021), doi: 10.3964/j.issn.1000-059309-2872-07, http://www.gpxygpfx.com/CN/10.3964/j.issn.1000-059309-2872-07.
17. F. Melgani, L. Bruzzone, *IEEE Trans. Geosci. Rem. Sens*., **42**, 1778–1790 (2004), doi: 10.1109/TGRS.2004.831865, https://ieeexplore.ieee.org/document/1323134.
18. T. Qiao, J. Ren, C. Craigie, et al., *J. Appl. Spectrosc*., **82**, 137–144 (2015), doi: 10.1007/s10812-015-0076-1, https://link.springer.com/article/10.1007/s10812-015-0076-1.
19. J. Ham, Y. Chen, M. Crawford, et al., *IEEE Trans. Geosci. Rem. Sens*., **43**, 492–501 (2005), doi: 10.1109/TGRS.2004.842481, https://ieeexplore.ieee.org/document/1396322.
20. W. Zhou, H. Yang, L. Xie, et al., *CATENA*, **202**, No. 12 (2021), doi: 10.1016/j.catena.2021.105222, https://www.sciencedirect.com/science/article/abs/pii/S0341816221000813?via%3Dihub.
21. W. Xu, Y. Han, Y. Qin, et al., *Spectrosc. Spectr. Anal*., **35**, 2856–2861 (2015), doi: 10.3964/j.issn.1000-059310-2856-06, http://www.gpxygpfx.com/CN/10.3964/j.issn.1000-059310-2856-06.