

# A New Hyperspectral Redundant Band Detection Method Based on Local Hurst Exponent

Jianhui Li

Foshan Polytechnic

Qiaozhi Li

Southern Medical University

Fang Wang (✉ [popwang619@163.com](mailto:popwang619@163.com))

Hunan Agricultural University <https://orcid.org/0000-0002-4756-719X>

Fan Liu

Hunan Agricultural University

---

## Research Article

**Keywords:** Local Hurst exponent, redundancy detection, multifractal feature, hyperspectral

**Posted Date:** May 4th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-442818/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# A New Hyperspectral Redundant Band Detection Method Based on Local Hurst Exponent

Jianhui Li · Qiaozhi Li · Fang Wang ·  
Fan Liu

Received: date / Accepted: date

**Abstract** Hyperspectrum reflectance is a curve in a certain wavelength range. Its complex dynamic structure reflects rich information of the object at variable bands. However, the potential redundancy will seriously affect accurate extraction of spectral features, therefore, information redundancy detection is a critical pretreatment for spectral analysis. In this paper, by using the local detrended fluctuation analysis, we propose a new method to detect the redundant bands. The method focuses on the spectral auto-correlation represented by local Hurst exponent in moving windows. Thus, the redundant band can be determined by the comparison of auto-correlation between two adjacent windows. To test our method, using the fractal feature of the removing redundant bands as augment, rapeseed oleic acid's prediction model is constructed based on random decision forest method. As comparison, the same feature of the original spectrum is also employed as augment for the model. The result shows that the feature of removing the redundant bands will bring better model performance than the feature of original spectrum does.

**Keywords** Local Hurst exponent · redundancy detection · multifractal feature · hyperspectral

---

J. H. Li  
School of Electronic Information, Foshan Polytechnic, Foshan. 528000, China

Q. Z. Li  
Department of Pathology, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China; Department of Pathology, The Fifth Peoples Hospital of Foshan Nahai District, Foshan 528231

\*Corresponding author: F. Wang  
School of Information and Intelligence, Hunan Agricultural University, Changsha, 410128, China  
E-mail: popwang619@163.com

F. Liu  
School of Agronomy, Hunan Agricultural University, Changsha, 410128, China

## 1 Introduction

Nowadays, the concept of smart agriculture has swept the world. It is an important content of the smart agriculture to use intelligent algorithm and advanced equipment to diagnose crops non-destructive. As a momentous remote sensing technology, the spectral analysis is often used for detecting the physical structure and chemical composition of substances. With the progress of information technology and precision optical technology, the spectral resolution has been improved [1]. The optical remote sensing technology has experienced the initial panchromatic, color photography and multispectral ( $\lambda/10$  order of magnitude). In the 1980s, a new photoelectric detection technology, hyperspectral ( $\lambda/100$ ) remote sensing, has brought about a revolution in the field of remote sensing [1,2]. Compared with traditional multispectral remote sensing, it can not only present the shape and spatial information of matter at different wavelengths in multiple spectral channels, but also access the spectral information of its components [2]. Due to the powerful data acquisition and analysis capabilities, the hyperspectral technique is naturally applied in intelligent agriculture [3–9]. Nevertheless, since that the hyperspectral remote sensing has the characteristics of multi band, narrow band width and large amount of data, the analysis of dynamic structure of hyperspectral signal become a critical premise to promote its effective utilization [10–12]. One of the key pretreatment task is to detect information redundancy hidden in hyperspectral, which seriously affects the extraction of real information of spectrum [13]. A lot of voluntary works have focused on this subject [13–16]. Of which, Liu et al. [13] investigated the redundancy hidden in the fluorescence spectrum through principal component regression analysis and moving windows selection in chemometrics; Liu et al. [14] proposed a weighted maximum relevance minimum redundancy waveband selection algorithm to calculate the mutual information between wavebands and target classes and wavebands, which was used for classification of soybean hyperspectral imaging datasets.

The hyperspectral reflectance is generally a continuous sequence arranged by wavelength, which is generated by the narrow band spectral information of each pixel. The reflectance at different band expresses different band structure of the object. Therefore, this technology can reveal more potential information of the object than the early near-earth remote sensing technology. However, a real problem is that the acquisition of crop hyperspectrum is easily affected by the collection mode as well as environment. The external noise leads to the information redundancy of adjacent bands more seriously. Therefore, fast and accurate redundancy detection has become particularly important, which is helpful for efficient modeling of crop diagnosis. Exactly as the particularity of crop hyperspectral acquisition, the redundancy and noise affect the dynamic structure of spectral reflectance. Specifically, the long-term correlation of reflectivity maybe changed in a certain wavelength range. Naturally, an idea of detecting the redundancy is emerged by using the auto-correlation of the local bands range. And this can be realized by a reliable method, detrended fluctuation analysis (DFA) [17], a widely used approach for estimation the Hurst

exponent that can well depict the long-term correlation of sequences. Because of the ability to deal with nonstationary measure, the method of DFA together with its extension have widely been applied into various fields [18–24]. As a microcosmic generalization of the DFA method, local DFA [6, 25–27] will provide the long-term correlated measurement for each considering windows, which is exactly what we need, i.e. to explore the correlation at each band.

In this work, we attempt to propose a method to detect rapeseeds hyperspectral redundancy by the local DFA. In addition, researches show that the crop hyperspectrum expresses multifractal nature due to its self-similarity and singularity [4–6, 28–30]. Accordingly, to validate our method, we apply the multifractal DFA (MF-DFA) [18] to extract the fractal features after removing redundant band, which are used as argument to construct an oleic acid content prediction model. As a comparison, the prediction model is also considered by the same fractal feature of original spectrum involving the redundancy.

The rest of this paper is organized as follows. In Sec.2, we first briefly review the Local MF-DFA, and then present a detection method of statistic significance of auto-correlation. By using this knowledge, we further propose a new method of spectral redundant band detection. At the end of this section, we give an account of experiment material. In Sec.3, we apply the proposed method to determine the redundant bands for the spectra of rapeseed. Furthermore, for the original spectrum and that of removing redundancy, we compare the performance of the oleic acid regression model by using the spectral fractal feature. A brief summary is provided in Sec.4.

## 2 Method and materials

### 2.1 Local multifractal detrended fluctuation analysis

The well-known MF-DFA method is a powerful tool for dealing with the non-stationary series. It is described as follows. For a time series  $\{x_t\}$ ,  $t = 1, 2, \dots, N$ , we split its profile  $X_t = \sum_{i=1}^t (x_i - \langle x \rangle)$  into  $N_s = [N/s]$  nonoverlapping segments with equal length  $s$ , denoted as  $X_{j,k}$ ,  $k = 1, 2, \dots, s$ . The same procedure is repeated starting from the opposite end to avoid disregarding a short part of the series in the end and thus  $2N_s$  segments are obtained altogether. In the  $j^{th}$  segment, we have  $X_{j,k} = X_{(j-1)s+k}$  for  $j = 1, 2, \dots, N_n$  and  $X_{j,k} = X_{N-(j-N_s)s+k}$  for  $j = N_s + 1, N_s + 2, \dots, 2N_s$ , where  $k = 1, 2, \dots, s$ . In each segment, the local linear (or other) trend can be fitted as  $\widehat{X}_{j,k}$  (in our work, we use 1 order polynomial to fit the trend). Fluctuation function  $f(s, j)$  is then defined for each segment as

$$f(s, j) = \frac{1}{s} \sum_{k=1}^s (X_{j,k} - \widehat{X}_{j,k})^2. \quad (1)$$

Average the fluctuation  $f(s, j)$  over all segments to obtain

$$\begin{aligned}
F(q, s) &= \left[ \frac{1}{2N_s} \sum_{j=1}^{2N_s} f(s, j)^{q/2} \right]^{1/q} \quad \text{for } q \neq 0; \\
F(q, s) &= \exp \left[ \frac{1}{4N_s} \sum_{j=1}^{2N_s} \ln f(s, j) \right] \quad \text{for } q = 0,
\end{aligned} \tag{2}$$

which is the so-called  $q$ -order variance function. It expresses the weighted fluctuation of every segments with the moment  $q$ . In generally, the medium and large fluctuation intervals are dominant when  $q \geq 2$ , and smaller fluctuation intervals dominated in the case of  $q < 2$ . If there is scale behavior existed in  $F(q, s)$ , the scaling exponent can be determined by the power-law between the  $F_q(s)$  and  $s$ ,

$$F_q(s) = [F(q, s)]^{1/q} \sim s^{h(q)} \tag{3}$$

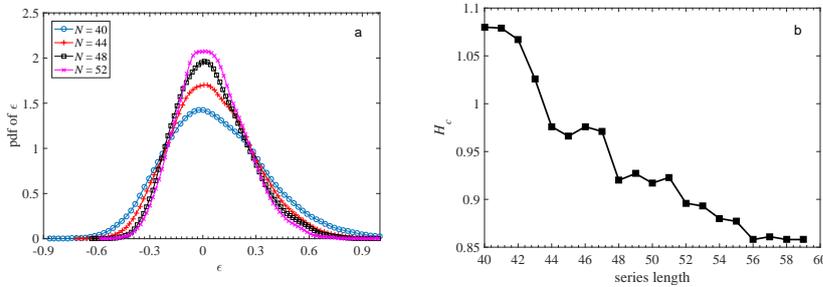
The dependence of the scaling exponent  $h(q)$  on  $q$  reflects the multifractal nature, conversely, the series is monofractal if  $h(q)$  is independent on  $q$ . When  $q = 2$ , the method of MF-DFA is degenerated to typical DFA[17]. And the  $h(2)$  is a good estimator of Hurst exponent  $H$ , i.e.  $H = h(2)$  if  $h(2) < 1$ , and  $H = h(2) - 1$  for  $h(2) > 1$  [19]. Generally, the Hurst exponent is a measurement, used to characterize the degree of auto-correlation.  $H > 0.5$  describes the persistent auto-correlation of the series  $x_t$  and  $H < 0.5$  implies it's anti-persistent. Especially, the value  $H = 0.5$  denotes the absence of auto-correlations, e.g. the  $H$  of white Gaussian noise (WGN) is 0.5.

To acquire the Hurst exponent for each point of the studied series, we use a window with size  $wt$  to slide across the series, namely, there are  $wt$  points in each window, say, in the first window, the point number starts from 1 ends to  $wt$ , the second window contains the point number from 2 to  $wt + 1$ , the  $i^{th}$  window has the point number from  $i$  to  $i + wt - 1$ , and so on. Then we utilize above DFA method in each window and obtain a series of Hurst exponent. This is so-called local Hurst exponent, denoted as  $\{LH_1, LH_2, \dots\}$ . In practice, we denote the auto-correlation of the  $i^{th}$  point by  $LH_i$ , actually, the  $LH_i$  describe the correlation of the sub-series starts from the  $i^{th}$  point.

## 2.2 Statistic significance of auto-correlation

As mention above, the Hurst exponent of uncorrelated series is expected as 0.5 theoretically, which presumably can be calculated only for an infinitely long time series. However, for finite time series without autocorrelation, its Hurst exponent will not be 0.5 due to the size limitation. Therefore, we must get a threshold, which indicates significant auto-correlation for a real-world series. Drawing on the experience of cross-correlated significance [32], a statistic significance test is proposed for  $H$ . It allows us to determine a critical value (threshold), denoted as  $H_c$ , so that the area greater than the  $H_c$  indicates statistically significant auto-correlations.

For this purpose, we consider the null hypothesis: the time series are *i.i.d.* uncorrelated variable and the range of  $H$  for the given length series can be obtained under the assumption that the time series without correlation. To this end, 10,000 *i.i.d.* Gaussian noise with zero mean and unit variance [32], which is uncorrelated series and its Hurst exponent is expected as 0.5, are taken for our consideration. The error between the real  $H$  and 0.5 (denoted as  $\epsilon = H - 0.5$ ) is expected as 0. Hence, let the integral of probability distribution function (pdf) of  $\epsilon$  between  $-\epsilon_c$  and  $\epsilon_c$  be equal to 0.95, namely, the critical value of  $H$  (denoted as  $H_c$ ) can be determined by  $0.5 + \epsilon_c$ . We present pdf of  $\epsilon$  for four groups of WGNs with given very short length in Fig. 1a; and the  $H_c$  of series length 40 to 59 with step size 1 is shown in Fig.1b. As shown in Fig. 1a, the symmetrical pdf of  $\epsilon$  converges to normal distribution with zero mean according to the central limit theorem. The critical value decreases as  $N$  increases, which can be also clearly shown in Fig. 1b. By using this critical value, one can determine whether the auto-correlation presents in each length series is significant or not. In general, a series is auto-correlated when its  $H$  is greater than  $H_c$ .



**Fig. 1** Statistical significance test of auto-correlation for series with given lengths. (a) is the pdf of critical value  $H_c$  for the statistical test with 10,000 *i.i.d.* Gaussian series; (b) is the  $H_c$  with the increasing series length.

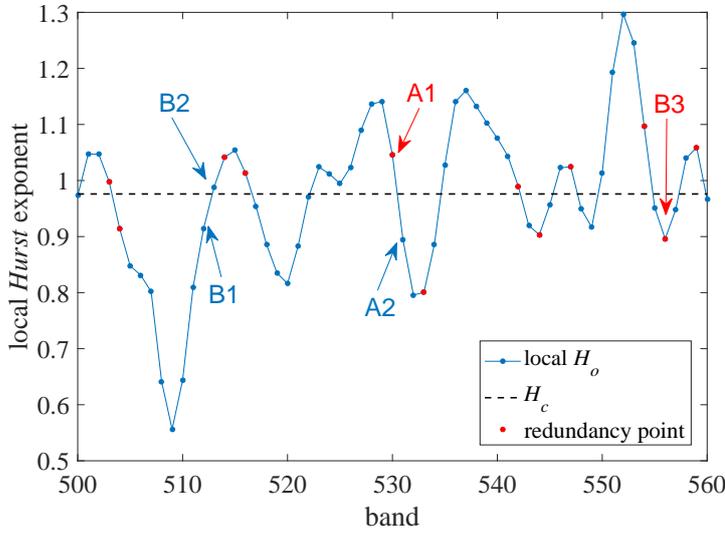
### 2.3 Working principle of redundant point detection

As mention in Introduction, correlations existed between bands lead redundancy hidden in a spectrum, which may interfere with the extraction of useful spectral features. Here, we develop a new method to detect redundancy point of spectrum from the perspective of detecting auto-correlation. To do so, we detect the redundant points by the relationship between the local Hurst exponent and  $H_c$  of two adjacent points. Firstly, use DFA method and a sliding window with size  $wt$  presented Subsect. 2.1 to calculate the  $LH_i$  for each window, thus, the  $LH_i$  characterizes the auto-correlation for the  $i^{th}$  window. Then, utilize Monte-Carlo simulation to obtain the critical value of the Hurst exponent  $H_c$  with the same length as the window. By comparing the  $LH_i$

and  $H_c$ , we can assess the auto-correlation for each window. Next, by comparing the correlation between two adjacent windows, the potential redundancy points can be detected as follows:

(i) if  $LH_i > H_c$  and  $LH_{i+1} < H_c$ , then the spectral reflectivity from band  $i$  to  $i + wt - 1$  is correlated and that from band  $i + 1$  to  $i + wt$  is uncorrelated, which suggests that the reflectivity of the  $i^{th}$  band can be represented by the bands from  $(i+1)^{th}$  to  $(i+wt-1)^{th}$ , so that means the  $i^{th}$  band is a redundant band.

(ii) On the other hand, if  $LH_i < H_c$  and  $LH_{i+1} > H_c$ , then the spectral reflectivity from band  $i$  to  $i + wt - 1$  is uncorrelated and that from band  $i + 1$  to  $i + wt$  is correlated, which suggests that the reflectivity of the  $(i + wt)^{th}$  band can be represented by band from  $i + 1$  to  $i + wt - 1$ , so that means the  $(i + wt)^{th}$  band is a redundant band.



**Fig. 2** Working principle of redundant point detection. The chosen sliding windows size is  $wt = 44$ , the corresponding critical value for the 95% confidence level of no auto-correlation is 0.976 (the black dotted line).

To show the method more intuitive, take a real spectrum as an example, we show the our detection principal in Fig. 2. The sliding window size is set  $wt = 44$ . For the adjacent points A1(the  $530^{th}$  band) and A2 (the  $531^{st}$  band), because  $LH_{A1} > H_c$  and  $LH_{A2} < H_c$ , then in the bands  $\{530, 531, \dots, 573\}$ , the spectrum is correlated, while the spectrum in bands  $\{531, 532, \dots, 574\}$  is uncorrelated. Hence, the point A1 is regarded as a redundancy point according the principal (i). For the adjacent points B1 (the  $512^{nd}$  band) and B2 (the  $513^{rd}$  band), since  $LH_{B1} < H_c$  and  $LH_{B2} > H_c$ , then the spectrum in bands  $\{512, 513, \dots, 555\}$  is uncorrelated, while the spectrum in bands

$\{513, 514, \dots, 556\}$  is correlated, and thus the last point of the bands  $\{513, 514, \dots, 556\}$ , say, the B3 (the 556<sup>th</sup> band) is a redundant band in the light of principal (ii). In this way, the potential redundant bands can be determined. Hereafter, we will remove the redundant bands and construct new spectral series for extraction of features, and further model for the object we interest in.

## 2.4 Experiment materials

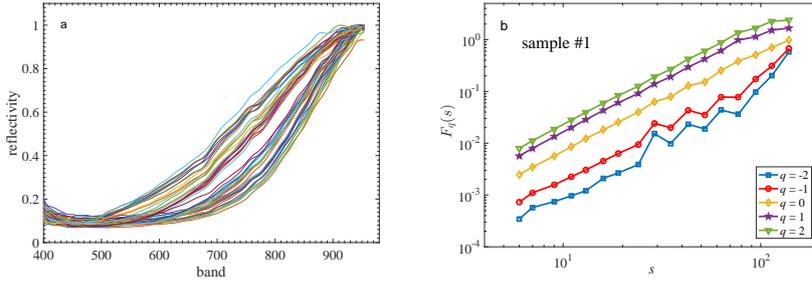
We select an ordinary hybrid (*Brassica napus L.*) rapeseed (cultivars name: Xiangyou 708 and Xiangyou 710) for our study, which was planted on October, 2018 and reaped on April, 2019, in paddy field in Yunyuan experiment basement ( $28^{\circ}10'E, 113^{\circ}4'N$ ) of Hunan Agricultural University. Five plants with uniform growth were randomly selected from each field, and five pods were taken from each plant to collect seeds. The two kinds of seeds are put in two petri dishes and collected spectrum reflectivity by SOC710 portable hyperspectral imager (band range 375 – 1041nm, resolution 4.6875nm), produced by the American Surface Optics Corporation. More specifically, using the SOC710 portable hyperspectral imager, we randomly select an area of 10 pixels  $\times$  10 pixels each time as the region of interest (ROI). 48 non-overlapping ROIs are selected. In each ROI, five points are selected randomly to be collected spectral reflectivity and averaged as a sample, labelled as 1 – 48. On the other hand, in each ROI, we grind all the seeds and use the Agilent 7890 ICP-MS produced by American Agilent Corporation to measure the rapeseeds fatty acid components. As the most important component of fatty acid, the oleic acid (%) is selected for our studied. In this way, we obtain 48 spectra samples and corresponding 48 oleic acid values.

## 3 Results and discussion

### 3.1 Detection of the spectral redundant bands

The bands range of the collected spectrum is from 375nm to 1041nm. To extract the accurate spectral features, we cut off the bands with serious noises at both ends and obtain the spectrum from bands 400 to 954, the normalized spectrum is shown in Fig. 3(a), and the multifractal nature of sample #1 is shown in Fig. 3(b).

Following, we utilize the proposed redundancy detection method to uncover the potential redundant bands for the 48 spectra. Some window sizes are used to our consideration. We show the local Hurst exponent of the reflectivity with window size  $wt = 44$  in Fig. 4. The corresponding critical value for the 95% confidence level under the assumption of no auto-correlation is 0.976, also shown in the upper panels of Fig. 4 (the black dotted line). According to the principle of redundancy detection presented in Subsect.2.3, There are 34 and 43 redundant bands (the red dots in Fig. 4) in the sample #1 and sample #25,



**Fig. 3** The normalized spectrum of rapeseed (a) and its multifractal nature (b).

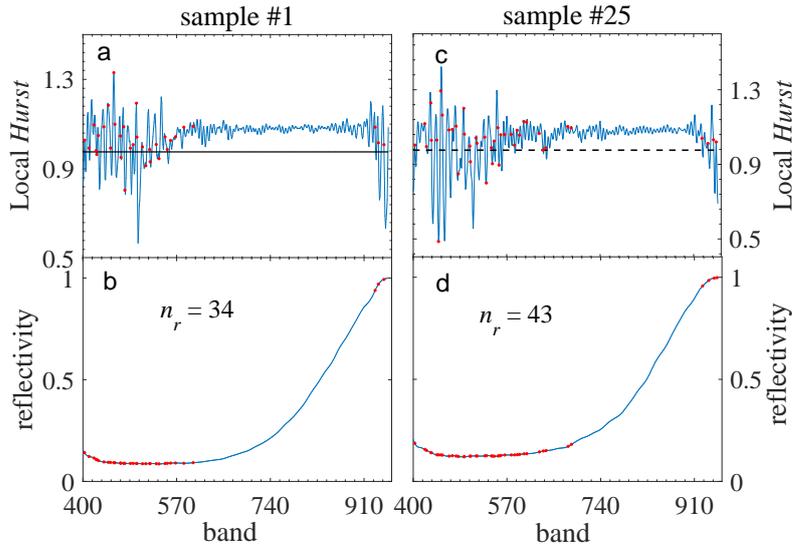
**Table 1** Statistics of the number of redundant points with the given window sizes, the number in parentheses is the corresponding sample number

	$wt = 40$	$wt = 42$	$wt = 44$	$wt = 46$	$wt = 48$	$wt = 50$
maximum	140(#23)	52(#25)	49(#5)	45(#3)	40(#27)	37(#27)
minimum	100(#36)	26(#16)	7(#23)	8(#22)	7(#23)	10(#40, #47)
mean	123.49	39.51	27.24	24.45	22.00	20.02
std.	8.34	6.31	8.74	8.56	6.76	6.94

respectively. We also show the numbers of redundant bands of every sample in Fig. 5 when window sizes are  $wt = 40$  and  $44$ . The original spectrum and the reconstructed spectrum with removing redundant bands of the samples #1 and #25 are shown in bottom panels of Fig. 5. To verify that removing redundancy does not reduce much useful information, the ratio of information entropy ( $R_{E_{tro}}$ ) between the spectrum of original and after removing redundancy is employed here, as shown in upper panels of Fig. 5. Seen from the ‘-’ line, the  $R_{E_{tro}}$  is closed to 1 for every sample, which demonstrates that the information content of the removing redundant spectrum is almost same as that of the original spectrum. For the more window sizes selection, we list the statistics of the number of redundant bands in Table 1. One notes that the number of redundant points decreases with the increasing window size and there is some difference of number of redundant bands among different samples. It implies that the different choice of window size may impact the feature extraction.

### 3.2 Regression model for oleic acid

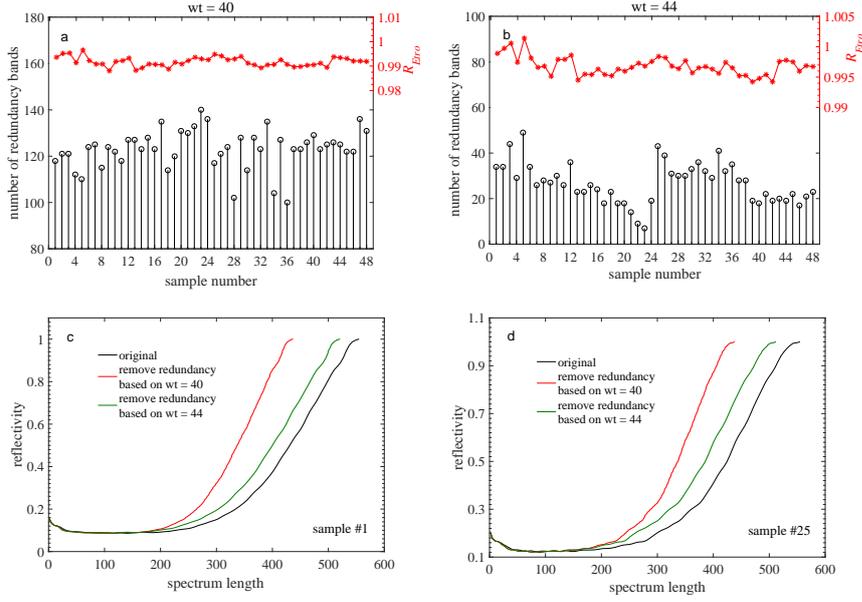
As the main product of rapeseed, rapeseed oil is becoming more and more necessary in people’s life, and its quality is directly related to people’s healthy. The composition of fatty acids is one of the most important indicators for the quality of rapeseed oil. The main composition of fatty acids is oleic acid, linoleic acid, linolenic acid and erucic acid, of which, the oleic acid is of most significance. Since that the oleic acid is conducive to human digestion, and it helps to reduce human cholesterol as well as the risk of thrombosis, its content



**Fig. 4** Local Hurst exponent of the spectrum reflectivity of sample #1 and sample #25. The black dotted line gives the critical values ( $H_c = 0.976$ ) for the 95% confidence level under the assumption of no auto-correlation, which is obtained from 10 000 *i.i.d.* Gaussian white noise with length  $N = 44$ . Red dots are the redundant points detected by local Hurst exponent corresponding band. There are 34 and 43 redundant bands in #1 and sample #25, respectively.

is regarded as critical indicator to monitor the rapeseeds quality. Therefore, research on modelling for oleic acid will be helpful to the quality analysis of rapeseed. In this subsection, we establish the oleic acid inversion model by using the spectral feature of removing redundancy. To this end, we use a compound feature (denoted as  $h_r$ ) composed of five generalized Hurst exponents  $\{h(-2), h(-1), h(0), h(1), h(2)\}$  as argument, which are calculated for the spectrum without the redundancy. As comparison, the five generalized Hurst exponents of original spectral are also computed and combined as a feature (denoted as  $h_o$ ). Since the simple linear as well as high-order polynomial models cannot exactly portray the relationship between the spectrum information and rapeseeds biochemical index [6], here, an intelligent method, namely, random decision forest (RDF) is employed [33] to do this job. In addition, three indicators, namely root-mean-square error ( $Rmse$ ) [6], correlation coefficient ( $R$ ) [6], and relative error ( $Re$ ) [6], are employed to evaluate the models, defined as in Eqs.(4)-(6), respectively.

$$Rmse = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}, \quad (4)$$



**Fig. 5** Result of redundant band detection of 48 spectra samples. (a) and (b) are the number of redundant bands tested by sliding windows with size 40 and 44, respectively, the relative information entropy between the original spectrum and removing redundant spectrum are also shown in them. (c) and (d) are spectrum length of original spectrum and removing redundant spectrum of sample #1 and sample #25, respectively.

$$R = \frac{\sum_{i=1}^n (Y_i - \bar{Y}_i)(\hat{Y}_i - \tilde{Y}_i)}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \sqrt{\sum_{i=1}^n (\hat{Y}_i - \tilde{Y}_i)^2}}, \quad (5)$$

$$Re = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right| \times 100\%. \quad (6)$$

where  $Y_i$  and  $\hat{Y}_i$  denote the  $i^{th}$  observed value and predicted value, respectively. Summarize our model as follows:

- First step: Use a window with size  $wt$  (we try different  $wt$ ) to slide through the bands of every spectrum sample. In each window, the local Hurst exponent  $LH$  is calculated by using MF-DFA. On the other hand, 10000 *i.i.d.* white noise series with the same length ( $N = wt$ ) is used to generate the critical value  $H_c$  for 95% confidence level of no correlation.
- Second step: According the principle of redundancy detection, the redundant bands of all sample spectra are found. For the removing redundant spectrum, MF-DFA is used again to calculated the five  $h(q)$  ( $q = -2, -1, 0, 1, 2$ ). The five  $h(q)$  are combined as the feature  $h_r$ .
- Third step: Regarding the  $h_r$  as independent variable and oleic acid contends as dependent variable (denoted as  $Y$ ), RDF model is constructed to predict the oleic acid contends (denoted as  $\hat{Y}$ ).

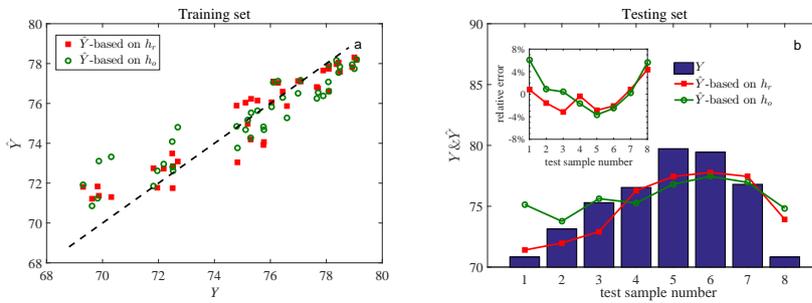
**Table 2** Statistics of the number of redundant points with the given window sizes, the number in parentheses is sample number

	Training set			Testing set		
	<i>Rmse</i>	<i>R</i>	<i>Re</i> (%)	<i>Rmse</i>	<i>R</i>	<i>Re</i> (%)
$h_r$ -based	1.0095	0.9498**	1.11	1.9929	0.7465*	2.22
$h_o$ -based	1.2304	0.9340**	1.30	2.2826	0.6358	2.48

\*\* denotes 0.01 significant level, \* denotes 0.05 significant level.

• Fourth step: Model test. Three indicators of *Rmse*, *R*, and *Re* are used to evaluate the model performance. Changing the window size as well as number of training samples and repeat the above steps.

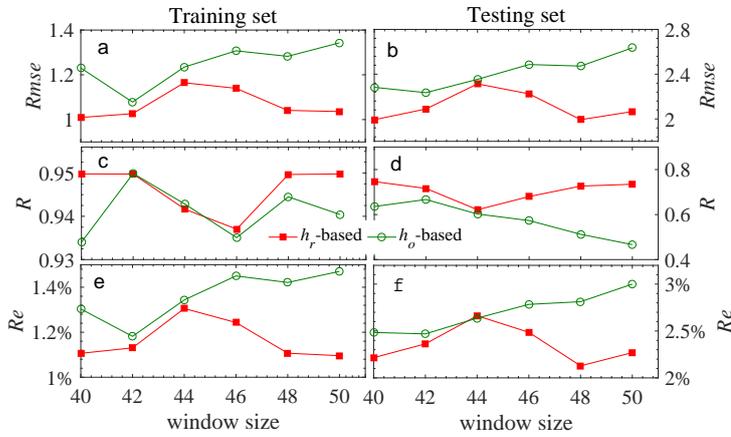
In the following model, 40 samples are chosen as training set randomly and 8 remaining samples constitute the testing set. The process is repeated 100 times to eliminate the impact of randomness. Fig. 6 shows the observed and predicted oleic acid values for training set and testing set. Table 2 lists the three evaluation indicators of the RDF regression models. Fig. 6 clearly shows that the predicted values are very close to the observed values. By contrast, the results based on  $h_r$  is better than that based on  $h_o$ , which is also confirmed in Table 2, the three indicators acquired from  $h_r$  are superior to those from the  $h_o$ , especially for the testing set. We note that the *R* of the testing set based on  $h_o$  is 0.6358, which is of no significance under the 95%, while the *R* = 0.7465 of  $h_r$ -based is significant. The nice results are attributable to the removing redundant bands, which makes the extracted spectral information more accurate and effective.



**Fig. 6** Comparison between observed and predicted oleic acid values based on fractal features ( $h_o$  and  $h_r$ ) calculated from the spectral reflectivity of original and removing redundancy. The redundancy points are detected in the sliding window with size 40. (a) and (b) are for training set and testing set, respectively. In (a), the dotted line means ideal state of  $\hat{Y} = Y$ . In (b), the inner panel is relative errors between the observed and predicted values for eight samples in test set.

### 3.3 Model test

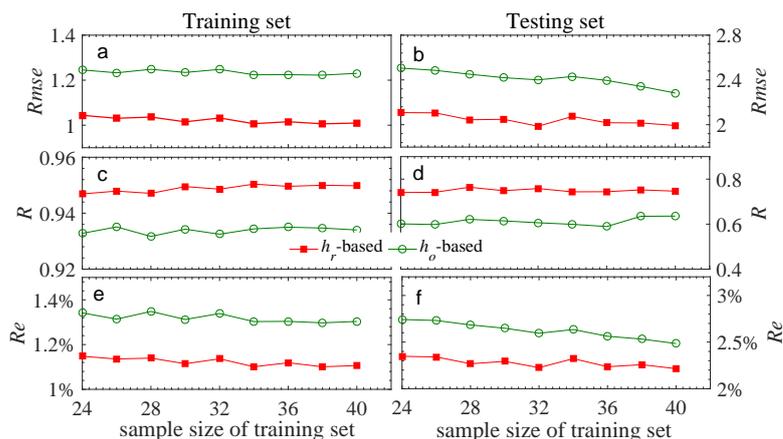
In this subsection, we test our model by changing the windows size and number of training samples. On the one hand, we still set the number of training sample is 40 and change the  $wt$  from 40 to 50 with step size 2. Conducting the model from the first step to the fourth step, we report the result in Fig. 7. The lower  $Rmse$  and  $Re$  together with the higher  $R$  illustrate that the  $h_r$ -based result is superior to that of  $h_o$ -based. The results were slightly affected by the different window sizes. On the other hand, we fix the window size  $wt = 40$  and vary the number of training sample from 24 to 40 with step size 2. The model result is shown in Fig. 8. The  $h_r$ -based results can be observed are significantly better than the  $h_o$ -based results regardless impact of training set size. In fact, the average improvement rates of  $Rmse$ ,  $R$ , and  $Re$  over the considering training set size are 22.58%, 14.84% and 13.13% for the testing set, respectively.



**Fig. 7** Three evaluating indicators obtained by fractal features of original spectrum and removing redundant spectrum with given window sizes. Left panel is for training set and right panel is for testing set

## 4 Conclusions

In the modeling with spectral feature, the potential redundancy will interfere with the extraction of accurate information, hence, redundant band detection is a critical job before feature extraction. In this work, we develop a new method of redundant band detection. By calculating the local Hurst exponent, which describes the auto-correlation, we can examine the relationship of auto-correlation between two adjacent bands of the spectrum, and thus determine the redundant bands. Applying this method to the rapeseed spectral, different degrees of redundancy can be checked for considered samples. In order to test



**Fig. 8** Three evaluating indicators obtained by fractal features of original spectrum and removing redundant spectrum with different sizes of training set. The redundancy points are tested in the sliding window with size 40. Left panel is for training set and right panel is for testing set.

the proposed method, we calculate the fractal features of the original spectrum and the removing redundant bands spectrum, respectively, and construct prediction models for oleic acid by random decision forest method. It shows that the feature based on the removing redundant bands spectrum will bring better model performance.

It is worthy to note that most existing methods of spectral redundant band detection are based on the model results, say, the determination of redundant bands depends on the modeling objects. However, our method is designed based on the auto-correlation. It is the essential characteristics of spectrum itself, which is independent on modeling objects. Therefore, our method is suitable for spectra of different materials theoretically.

**Acknowledgements** This work was partially supported by National Natural Science Foundation of China (no. 61973111), High-level talent research project of Foshan Polytechnic (no. KY2020G01), Natural Science Foundation of China of Hunan Province (CN) (Grant no. 2020JJ4377)

**Conflict of Interest:** The authors declare that they have no conflict of interests.

## References

1. Thenkabail, P. S. and Lyon, J. G. Hyperspectral remote sensing of vegetation. CRC press, 2016.
2. Tong, Q. X., Zhang, B. and Zhang, L. F.: Current progress of hyperspectral remote sensing in China. *Journal of Remote Sensing*, 20(5): 689-707 (2016).
3. Zhang, L. F., Sun, X. J., Wu, T. X., et al.: An analysis of shadow effects on spectral vegetation indexes using a ground-based imaging spectrometer. *IEEE Geoscience and Remote Sensing Letters*, 12(11): 2188-2192 (2015)

4. Wang, X. Q., Wang, F., Liao, G. P., et al.: Multifractal analysis of rape spectrum and modeling of chlorophyll diagnosis, *Spectrosc. Spectr. Anal.* 36(11), 36573663 (2016) (in Chinese with English abstract).
5. Jiang, S., Wang, F., Shen, L. M., et al.: Extracting sensitive spectrum bands of rapeseed using multiscale multifractal detrended fluctuation analysis, *J. Appl. Phys.* 121, 104702 (2017)
6. Jiang, S., Wang, F., Shen, L. M., et al.: Local detrended fluctuation analysis for spectral red-edge parameters extraction, *Nonlinear Dyn.* 93, 995-1008 (2018)
7. Lin, F., Guo, S., Tan, C., et al.: Identification of Rice Sheath Blight through Spectral Responses Using Hyperspectral Images. *Sensors*, 20(21): 6243 (2020)
8. Jin, H., Ma, Y., Li, L., et al.: Rapid and non-destructive determination of oil content of peanut (*arachis hypogaea* L.) using hyperspectral imaging analysis. *Food Analytical Methods*, 9(7), 2060-2067 (2016)
9. Li, L.T., Ren, T., Ma, Y., et al.: Evaluating chlorophyll density in winter oilseed rape (*Brassica napus* L.) using canopy hyperspectral red-edge parameters. *Comput. Electron. Agric.* 126, 21-31 (2016)
10. Wang, N., Du, B., Zhang, L. P., et al.: An abundance characteristic-based independent component analysis for hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing*, 53(1): 416-428 (2015).
11. Li, N., Zhao, H., Huang, P., et al.: A novel logistic multiclass supervised classification model based on multi-fractal spectrum parameters for hyperspectral data. *Int. J. Comput. Math.* 92(4), 836849 (2015)
12. Combrexelle, S., Wendt, H., Tourneret, J.Y., et al.: Hyperspectral image analysis using multifractal attributes. In: *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, (JAP), 18 (2015)
13. Liu, J., Shang, L.P., Qu, W.W., et al.: The Preliminary Research on Information Redundance in Fluorescence Spectral Data Process. *Spectroscopy & Spectral Analysis*, 30(10): 2685 (2010) (in Chinese with English abstract).
14. Liu, Y., Li, M., Wang, S., et al.: Hyperspectral waveband selection algorithm based on weighted maximum relevance minimum redundancy and its stability analysis. *Measurement Science and Technology*, 31(8): 085501 (2020)
15. Zhang, W., Li, X., Zhao, L.: Discovering the Representative Subset with Low Redundancy for Hyperspectral Feature Selection. *Remote Sensing*, 11(11): 1341 (2019).
16. Zhang, M., Gong, M. and Chan, Y.: Hyperspectral band selection based on multi-objective optimization with high information and low redundancy. *Applied Soft Computing*, 70: 604-621 (2018).
17. Peng, C.K., Buldyrev, S.V., Havlin, S., et al.: Mosaic organization of DNA sequences. *Phys. Rev. E* 49, 1685-1689 (1994)
18. Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, et al.: Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A* 316(1/2/3/4), 87-114 (2002)
19. Movahed, M.S., Jafari, G.R., Ghasemi, F., et al., Multifractal detrended fluctuation analysis of sunspot time series, *Journal of Statistical Mechanics: Theory and Experiment* 2, 115 (2006)
20. Wang, F., Liao, G.P., Li, J.H., et al.: Multifractal detrended fluctuation analysis for clustering structures of electricity price periods. *Phys. A* 392(22), 5723-5734 (2013)
21. Oświęcimka P., Kwapień J., Drożdż S.: Wavelet versus detrended fluctuation analysis of multifractal structures. *Phys. Rev. E* 74(1), 016103 (2006)
22. Gieraltowski, J., Zebrowski, J.J., Baranowski, R.: Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia. *Phys. Rev. E* 85, 021915 (2012)
23. Wang, J., Shang, P.J., Cui, X.R.: Multiscale multifractal analysis of traffic signals to uncover richer structures. *Phys. Rev. E* 89, 032916 (2014)
24. Wang, F., Fan, Q.J., Stanley, H.E.: Multiscale multifractal detrended fluctuation analysis of two-dimensional surfaces. *Phys. Rev. E* 93, 042213 (2016)
25. Xia, J., Shang, P., Wang, J.: Estimation of local scale exponents for heartbeat time series based on DFA. *Nonlinear Dyn.* 74(4), 11831190 (2013)
26. Ihlen, E.A.F., Vereijken, B.: Detection of co-regulation of local structure and magnitude of stride time variability using a new local detrended fluctuation analysis. *Gait Posture* 39(1), 466-471 (2014)

27. Wang, F., Li, Z.S., Li, J.W.: Local multifractal detrended fluctuation analysis for non-stationary images texture segmentation. *Appl. Surf. Sci.* 233, 116-25 (2014)
28. Wang, T., Zhou, C., Gu, Y., et al.: Hyperspectral Estimation of Cadmium Content in Tumorous Stem Mustard Based on the Wavelet-Fractal Analysis. *Scientia Agricultura Sinica*, 51(1): 71-81 (2018)
29. Chavez, P., Yarlequ, C., Loayza, H., et al.: Detection of bacterial wilt infection caused by *Ralstonia solanacearum* in potato (*Solanum tuberosum* L.) through multifractal analysis applied to remotely sensed data. *Precision agriculture*, 13(2): 236-255 (2012)
30. Qi, J. B., Fan, F. L. and Guo, Z. X.: Monitoring Tobacco Canopy Growth Status Based on Hyperspectral Fractal Analysis. *Journal of South China normal University (Natural Science Edition)*, 48(1): 94-100 (2016)
31. Podobnik, B., Grosse, I., Horvatic, D., et al.: Quantifying cross-correlations using local and global detrending approaches, *Eur. Phys. J. B* 71, 243250 (2009).
32. Podobnik, B., Jiang, Z.Q., Zhou, W.X.et al.: Statistical tests for power-law cross-correlated processes, *Phys. Rev. E* 84, 066118 (2011)
33. Ho, T.K.: Random decision forest. In: *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, Canada, pp. 278282, 1418 August 1995

# Figures

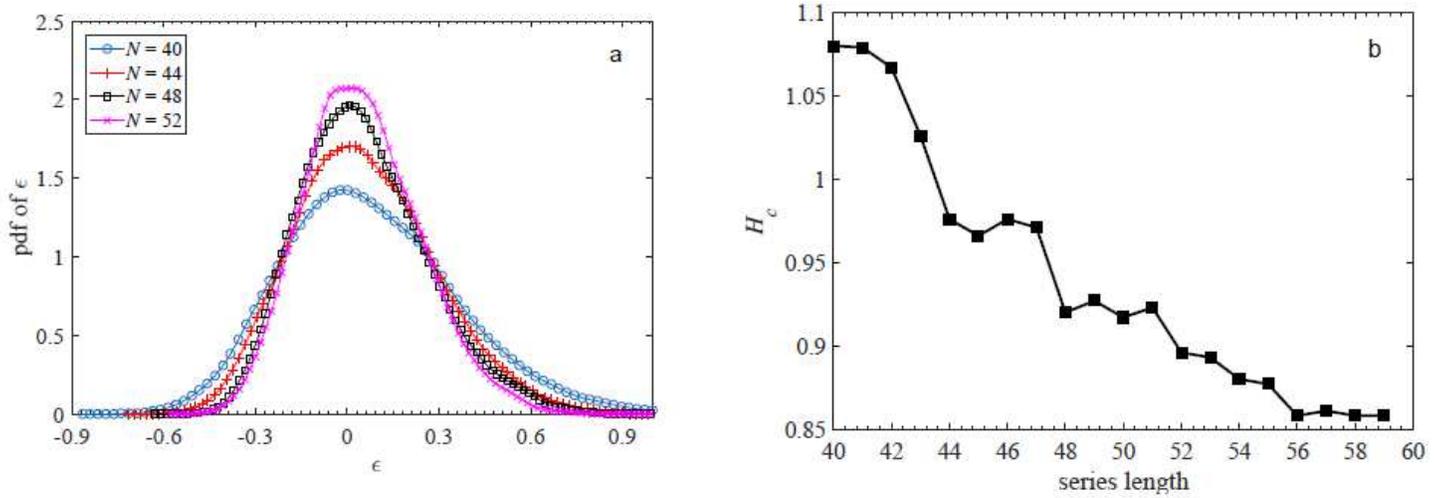


Figure 1

Statistical significance test of auto-correlation for series with given lengths. (a) is the pdf of critical value  $H_c$  for the statistical test with 10,000 i.i.d. Gaussian series; (b) is the  $H_c$  with the increasing series length.

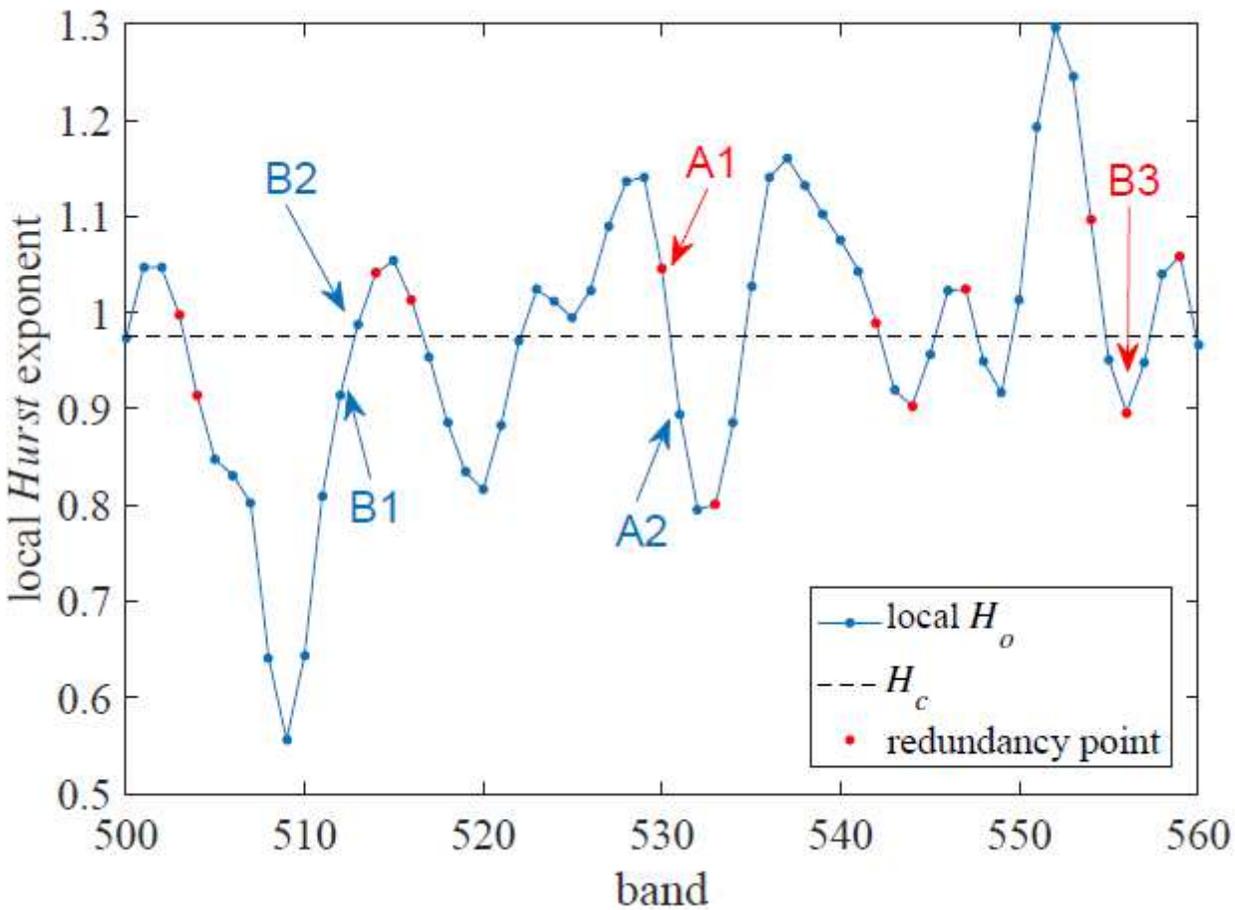


Figure 2

Working principle of redundant point detection. The chosen sliding windows size is  $w_t = 44$ , the corresponding critical value for the 95% confidence level of no auto-correlation is 0.976 (the black dotted line).

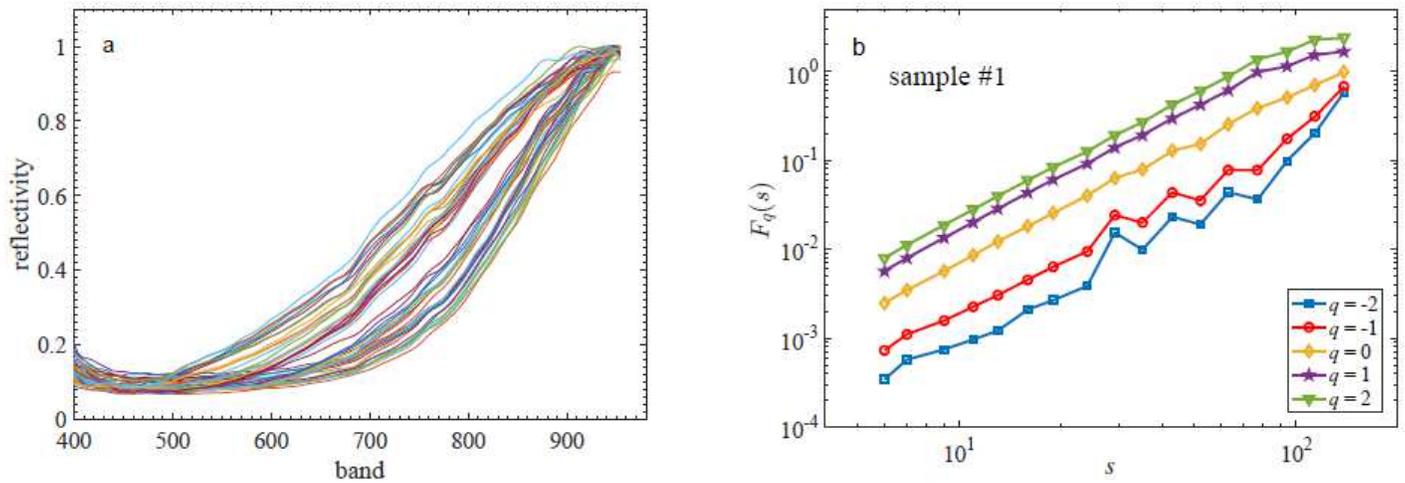
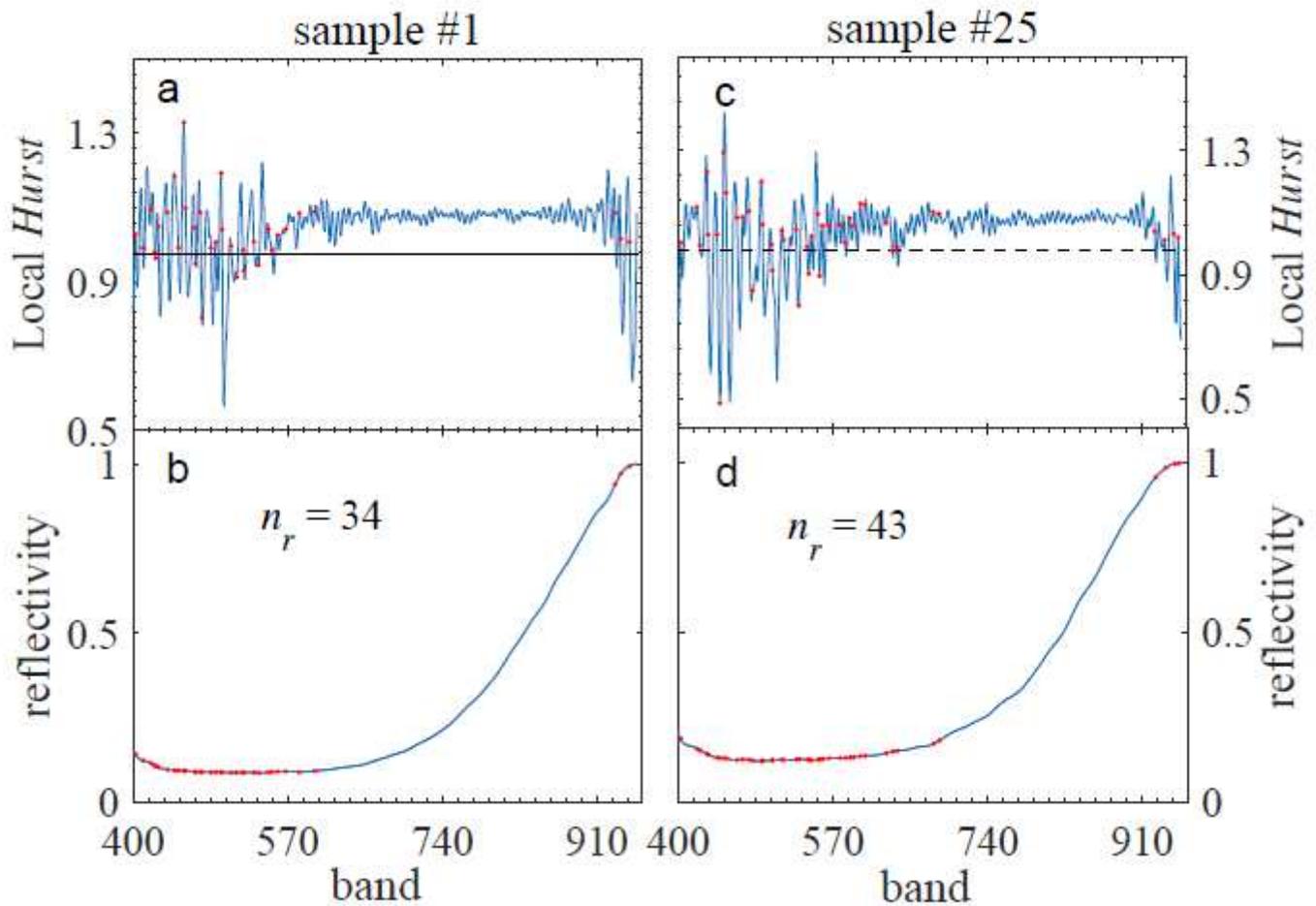


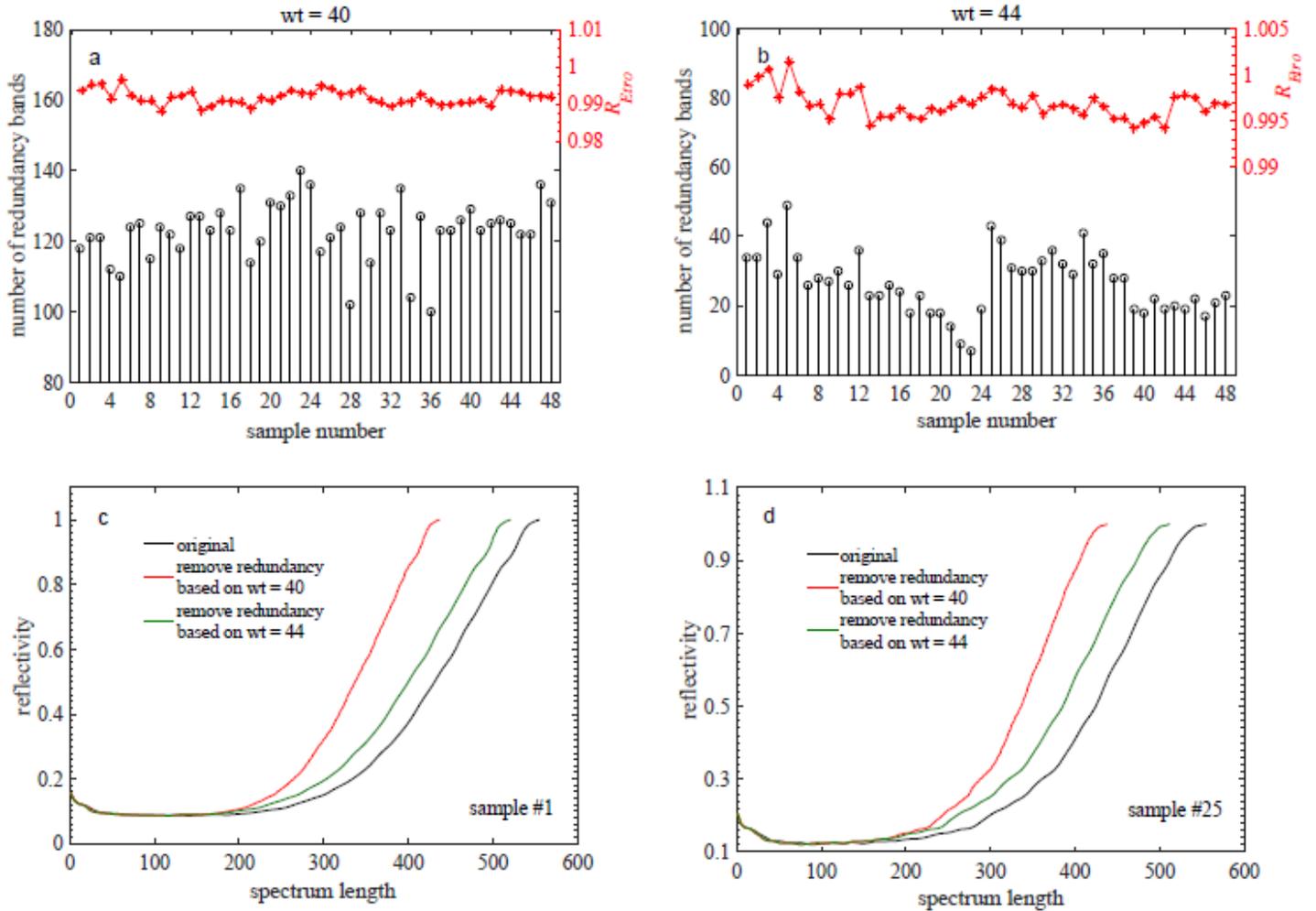
Figure 3

The normalized spectrum of rapeseed (a) and its multifractal nature (b).



**Figure 4**

Local Hurst exponent of the spectrum reflectivity of sample #1 and sample #25. The black dotted line gives the critical values ( $H_c = 0.976$ ) for the 95% confidence level under the assumption of no auto-correlation, which is obtained from 10 000 i.i.d. Gaussian white noise with length  $N = 44$ . Red dots are the redundant points detected by local Hurst exponent corresponding band. There are 34 and 43 redundant bands in #1 and sample #25, respectively.



**Figure 5**

Result of redundant band detection of 48 spectra samples. (a) and (b) are the number of redundant bands tested by sliding windows with size 40 and 44, respectively, the relative information entropy between the original spectrum and removing redundant spectrum are also shown in them. (c) and (d) are spectrum length of original spectrum and removing redundant spectrum of sample #1 and sample #25, respectively.

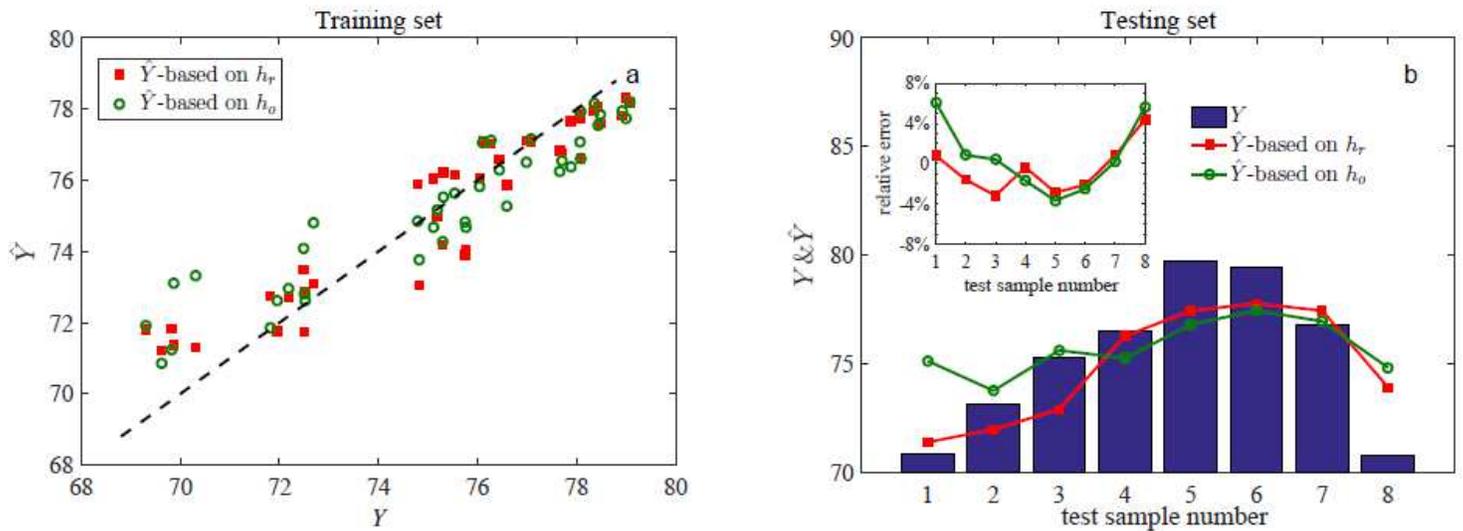


Figure 6

Comparison between observed and predicted oleic acid values based on fractal features ( $h_o$  and  $h_r$ ) calculated from the spectral reflectivity of original and removing redundancy. The redundancy points are detected in the sliding window with size 40. (a) and (b) are for training set and testing set, respectively. In (a), the dotted line means ideal state of  $\hat{Y} = Y$ . In (b), the inner panel is relative errors between the observed and predicted values for eight samples in test set.

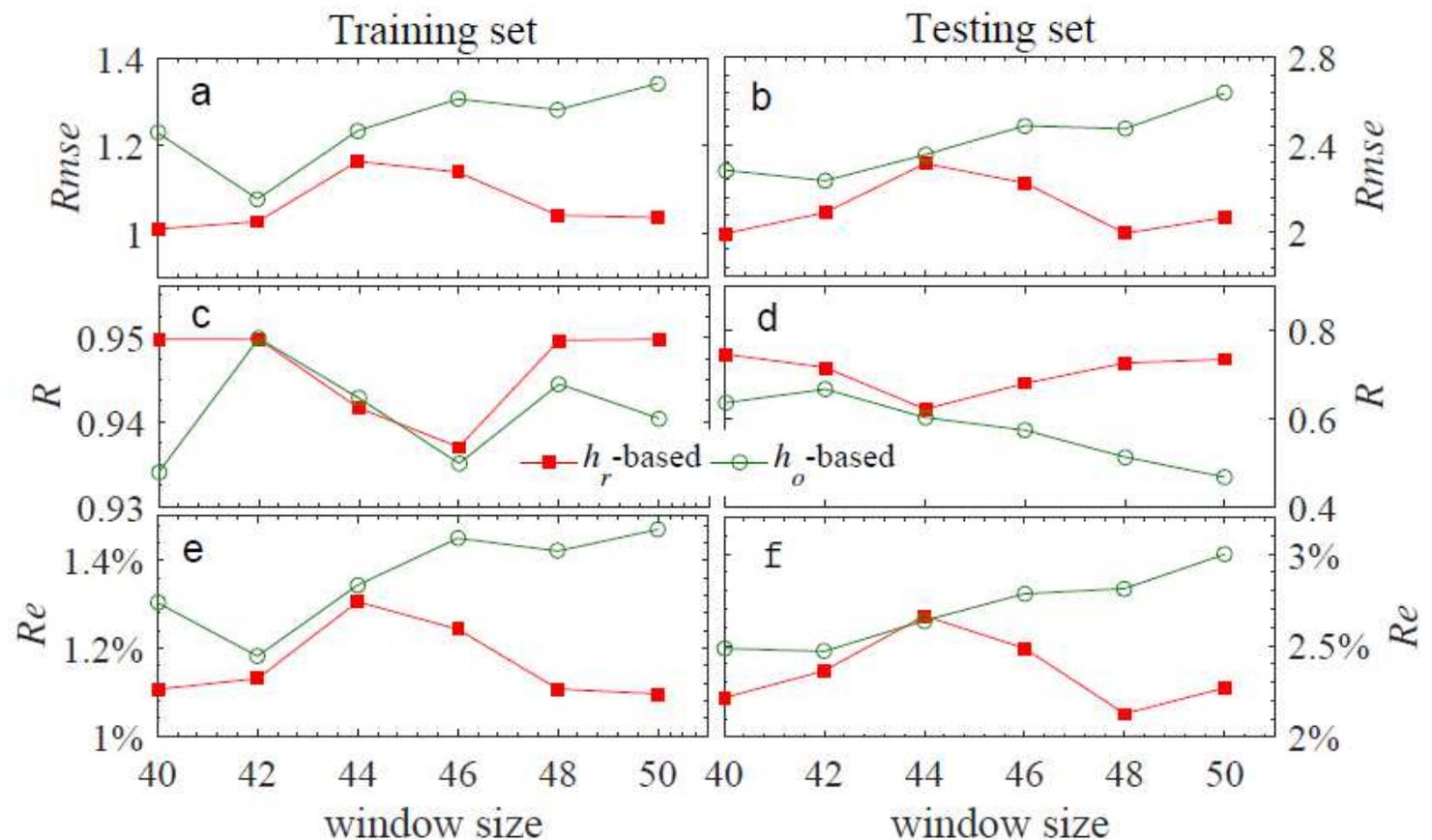


Figure 7

Three evaluating indicators obtained by fractal features of original spectrum and removing redundant spectrum with given window sizes. Left panel is for training set and right panel is for testing set

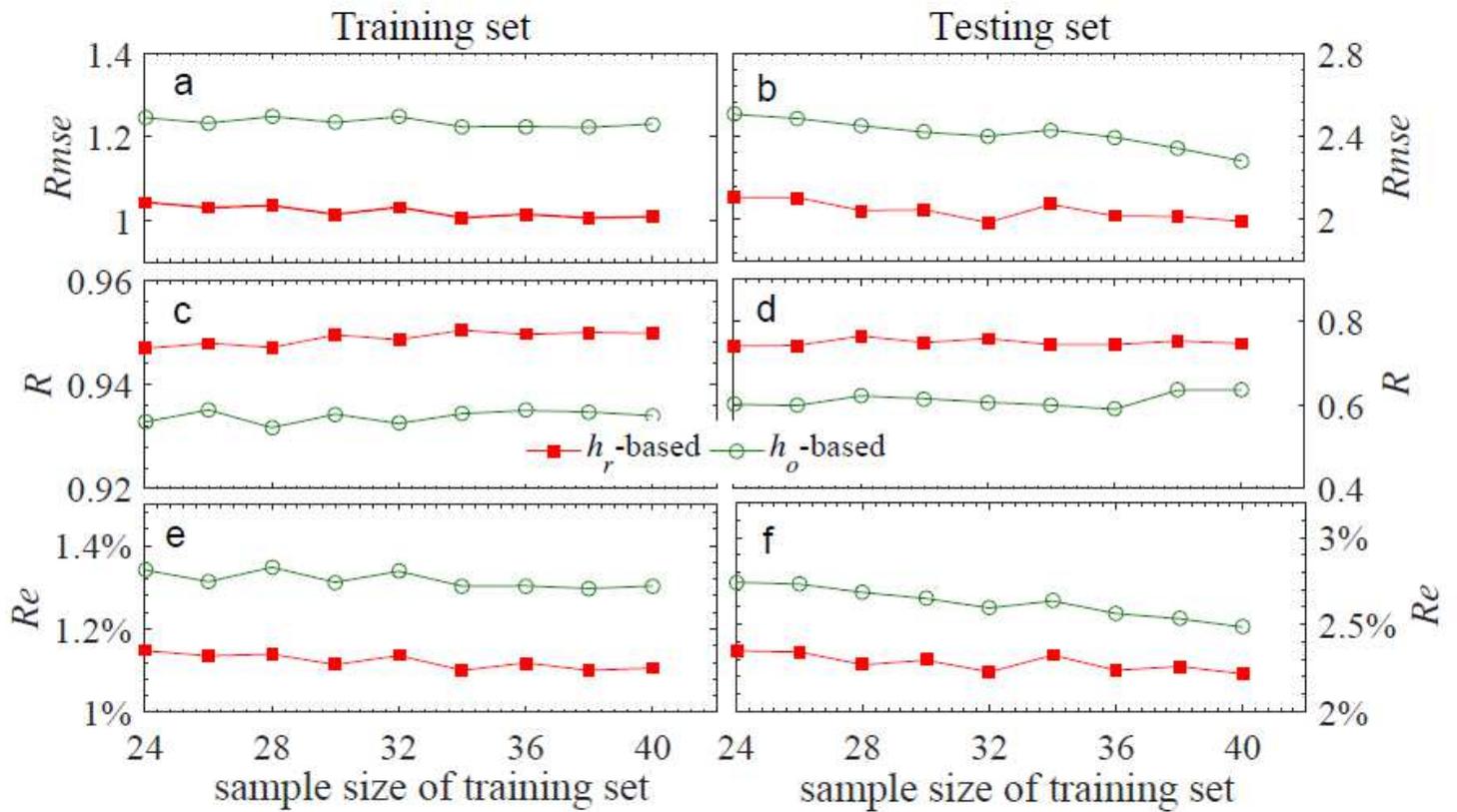


Figure 8

Three evaluating indicators obtained by fractal features of original spectrum and removing redundant spectrum with different sizes of training set. The redundancy points are tested in the sliding window with size 40. Left panel is for training set and right panel is for testing set.