Multiscale Superpixelwise Prophet Model for Noise-Robust Feature Extraction in Hyperspectral Images

Ping Ma, Jinchang Ren, *Senior Member*, *IEEE*, Genyun Sun, *Member*, *IEEE*, Huimin Zhao,
Xiuping Jia, *Fellow*, *IEEE*, Yijun Yan, and Jaime Zabalza

Abstract—**Despite of various approaches proposed to smooth the hyperspectral images (HSIs) before feature extraction, the efficacy is still affected by the noise, even using the corrected dataset with the water absorption bands and noise bands discarded. In this study, a novel spectral-spatial feature mining framework, Multiscale Superpixelwise Prophet Model (MSPM), is proposed for noise-robust feature extraction and effective classification of the HSI. The prophet model is highly noise-robust for deeply digging into the complex structured features thus enlarging interclass diversity and improving intraclass similarity. First, a pseudo color image is generated from the first three principal components of an HSI, from which the superpixelwise segmentation is produced to group pixels into regions with adaptively determined sizes and shapes. A multiscale prophet model is then utilized to extract the multiscale informative trend components from the average spectrum of each superpixel. Taking the multiscale trend signal as the input feature, the HSI data are classified superpixelwisely, which is further refined by majority voting based decision fusion. Comprehensive experiments on three publicly available datasets have fully validated the efficacy and robustness of our MSPM model when benchmarked with eleven state-of-the-art algorithms, including six spectral-spatial methods and five deep learning ones. Besides, MSPM also shows superiority under limited training samples, due to the combined strategies of superpixelwise fusion and multiscale fusion. Our model has provided a useful solution for noise-robust feature extraction as it achieves superior HSI classification even from the uncorrected dataset without prefiltering the water absorption and noisy bands.**

Index Terms—**Hyperspectral image (HSI); Multiscale Prophet model; Spectral-spatial feature mining; Superpixel segmentation.**

I. INTRODUCTION

NOWADADYS, with the enhanced sensor and communication techniques, increasing spectral ranges and bands can be obtained in HSIs [1, 2]. The rich spectral information in an HSI allows for accurate discrimination of materials with subtle spectral differences than panchromatic and multi-spectral images [3]. As a result, the HSI has been widely used in a wide range of applications, such as salient object detection [4], peated barley malt analysis [5], beef quality evaluation [6], corneal epithelium injuries classification [7] and fingerprint

anthocyanins extraction [8] et al.

Often, pixel-level mapping is widely used for classification of the HSI [9], where a certain number of pixels in each class should be labelled in prior for training a supervised classifier [10], such as the popularly used support vector machine (SVM) [11]. Due to the limitations of the sensors, environmental factors, and even atmospheric effects, there inevitably exist bands with a low signal-to-noise ratio (SNR) [12]. Even after removal of these water absorption bands and noise bands, the resting bands still contain severe noise [13]. This would unavoidably degrade the subsequent feature extraction and classification, especially under limited training samples and severely unbalanced distributions in different classes.

To tackle these challenges, the feature mining of an HSI has gained increasing attention recently in both the spectral and spatial domains [1]. Typical techniques for spectral feature mining include the principal component analysis (PCA) [14], non-negative matrix factorization (NMF) [15], empirical mode decomposition (EMD) [16] and the singular spectrum analysis (SSA) [13], which have shown good performance in preserving the dominant spectral information whilst removing the noise. However, most of these techniques only utilize the spectral information, which fail to deal with the objects with low inter-class heterogeneity and low intra-class homogeneity [17, 18].

There are also many spatial feature mining methods, such as the Gabor filter [19], morphological profile [20], sparse representation [21], 2D-EMD [16] and 2D-SSA [3]. In addition, a supervised spectral-spatial classification approach was proposed in [10], which features a spectral data fidelity term derived from the sparse multinomial logistic regression (SMLR) and a spatially adaptive Markov Random Field (MRF) prior model constructed via spatially adaptive total variation regularization (denoted as SMLR-SpATV). This method could spatially smooth the images and gain robust results of classification even with very limited training samples.

Besides, the deep learning methods, which can extract latent structure features in spatial and spectral domains, have recently drawn increasing attention in classifying HSIs. Typical methods include the diverse region based convolutional neural network (DR-CNN) [22], the spectral-spatial cascaded

P. Ma is with the Department of Electronic and Electrical Engineering, University of Strathclyde, G1 1XW, Glasgow, U.K. and the National Subsea Centre, Robert Gordon University, AB21 0BH, Aberdeen, U.K.
J. Ren (e-mail: jinchang.ren@ieee.org) and Y. Yan are with the National Subsea Centre, Robert Gordon University, AB21 0BH, Aberdeen, U.K.

G. Sun is with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China, and also with the Laboratory for Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China.
H. Zhao is with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China. (e-mail: zhaohuimin@gpnu.edu.cn)
X. Jia is with the School of Engineering and Information Technology, University of New South Wales at Canberra, Canberra, ACT 2600, Australia.
J. Zabalza is with the Department of Electronic and Electrical Engineering, University of Strathclyde, G1 1XW, Glasgow, U.K.

recurrent neural network (SSCasRNN) [23], the Gaussian-Bernoulli restricted Boltzmann machine (GBRBM) in parallel [24], the fully convolutional segmentation network (FCSN) [2] and the segmented autoencoders using only the spectral features [25]. However, their results depend heavily on the network structure and the training strategy, where the results can be quite poor when there are insufficient training samples [26].

In most of the aforementioned approaches, spatial-spectral classification produces much improved results, where often the spatial region adopted has a fixed size and shape. This actually differs from real cases where the regions of interest are usually irregularly shaped and inconsistently sized [20]. Therefore, the spatial regions used for feature mining should be adaptive to the specific structures of the image. To this end, the *superpixel* has been employed to adaptively extract the spatial features for the object based HSI classification [9, 20, 26]. The superpixels in an image are local homogeneous regions with different sizes and shapes, where smooth areas tend to form larger superpixels while superpixels from heterogeneous areas tend to be much smaller. In fact, superpixel is found effective to represent the spatial-contextual information and improve the classification [12]. In [9], multiscale superpixel (MSP) segmentation and subspace-based support vector machine (SVM), SVMsub, are combined for spectral-spatial classification of HSIs. By using the simple linear iterative clustering (SLIC) for extracting superpixels, multiscale spatial features are combined and fused for improved classification of HSIs. In [26], the superpixel was combined with the conventional SSA and 2-D SSA for the HSI classification, where the local spatial features are adaptively extracted based on the size of the extracted superpixels.

However, most of these spectral/spatial feature mining techniques fail to deal with the noisy bands, as they are often tested on the corrected datasets with the water absorption bands and noisy bands removed. In practical applications, this will cause information loss and also extra burden to filter these noise affected bands. To tackle the noisy bands rather than simply removing them, a few advanced denoising approaches have been developed in recent years, such as low-rank representation (LRR) and superpixel segmentation based denoising (denoted as SS-LRR) [12]. In [27], the low-rank matrix approximation and an iterative regularization framework is proposed to denoise the low SNR bands while preserving high-SNR bands. In [28], the subspace LRR was used to decompose an HSI into two lower-rank sub-matrices, using superpixel based segmentation to exploit low rank in local spatial regions. However, these approaches pay more attention on the denoising rather than extraction of the discriminative features for data classification in an HSI. In [29], we proposed a multiscale 2D-SSA with principal component analysis (2D-MSSP). By fusing the spatial features of 2D-SSA in different scales, it shows promising performance in noise-robust feature extraction even on the uncorrected HSI datasets.

As a recently developed forecasting tool by Facebook, the Prophet model can deeply analyze data by decomposing it into different components whilst well preserving the trend component and the most informative part of signal [30]. The derived trend component is jointly decided by the information of former data in the series, making it robust to any missing or highly noisy data. The HSI also characterizes a sequential data though in the spectral domain [28], which contains highly nonlinear scattering noise and even missing data, thus it motivated us to apply the Prophet model in analyzing the HSI.

In this paper, a novel multiscale superpixelwise Prophet model (MSPM) is proposed for noise-robust feature extraction under small training size. With the exceptional capability of the Prophet model in noise-robust trend extraction and data prediction for enhancing features with improved interclass diversity and intraclass similarity, we aim to realize improved classification even on the uncorrected HSIs. The major contributions of this paper can be summarized as follows:

(1) The proposed MSPM model can simultaneously denoising the data whilst extracting the noise-robust features of an HSI, even from the uncorrected dataset without removing or denoising the water absorption bands and severely noisy bands;

(2) It is the first time the Prophet model is applied in HSIs, where the proposed MSPM can adaptively extract the trend components in different scales based on the characteristics of the input data. The fusion of multiscale trend components can extract effective features under unknown noise levels, which is beneficial to the HSI classification;

(3) With both multiscale superpixelwise spatial-spectral feature-level fusion and decision-level fusion, the proposed framework is found to be effective for the HSI classification even under limited training samples, which has outperformed a number of state-of-the-art approaches including conventional spectral-spatial classifiers and several deep learning models as evaluated on three publicly available datasets.

The remainder of this paper is organized as follows. Section II introduces the principles of the Prophet model. Section III presents in detail the proposed MSPM framework. The design of experiments including the testing datasets and parameter settings are discussed in Section IV. Section V presents the experimental results and analysis, followed by some concluding remarks drawn in Section VI.

II. Principles of the Prophet Model

A. Concept and Algorithm of the Prophet Model

As a recently developed forecasting tool by Facebook for analyzing time-series data of business [30], the Prophet model blends the advantages of judgmental and statistical modelling to convert the forecasting problem as a curve-fitting task. As a nonlinear and decomposable model, the Prophet model can deeply analyze data by decomposing a given time-series $y(t)$ into different components, in which the trend component represents the most informative part of signal.

$$y(t) = g(t) + p(t) + h(t) + \varepsilon_t \qquad (1)$$

where $g(t)$ is the trend component modelling the non-periodic changes; $p(t)$ is the periodic seasonality component i.e., daily, weekly or annual seasonality; $h(t)$ denotes the holiday effects which usually appear on irregular schedules over one or more days; ε_t is the error term assumed in the normal distribution.

The trend component represents the most informative part of signal, which is widely defined by a logistic function [30].

$$g(t) = C / \left(1 + e^{-r(t-m)} \right) \qquad (2)$$

where C is the carrying capacity determining the maximum value of $g(t)$, r denotes the growth rate and m is an offset parameter. In $g(t)$, the growth rate r is defined to characterize the changes in trend component. More specifically, there are S

changepoints are set over the whole time t. At each changepoint $s(s = 1, ..., S)$, the growth rate r is set to change for better curve-fitting. The rate change is defined as $\boldsymbol{\delta} \in \mathbb{R}^S$, where δ_s denotes the rate change at the s-th changepoint. Therefore, the growth rate r_s at any changepoint is calculated by adding the rate adjustment to the former growth rate.

$$r_s = r_{s-1} + \delta_s \tag{3}$$

where the base rate r_1 is determined by the Stan's L-BFGS [32] and the rate adjust vector $\boldsymbol{\delta}$ is represented by a sparse prior via a Laplace distribution: $\delta_q \sim \text{laplace}(\tau)$. The more explanations are given in Section IV.B. Based on the growth rate, the parameter m is correspondingly updated for connecting the trend at different changepoints. The adjustment on m at the s-th changepoint is defined as $\emptyset_s$. Thus, the offset value m at s-th changepoint is given by:

$$m_s = m_{s-1} + \emptyset_s \tag{4}$$

where the base offset m_1 is also defined using the Stan's L-BFGS. The $\emptyset_s$ is calculated as:

$$\emptyset_s = \left(s - m_1 - \sum_{q<s} \emptyset_q\right)(1 - r_{s-1}/r_s) \tag{5}$$

where $\emptyset_1$ is set to 0.

Business time series usually possess periodic phenomena due to the associated repeated seasonal behaviors [31]. To achieve this, the period function of t is employed as the Seasonality in the model. It employs a standard Fourier series to flexibly approximate the seasonal effects.

$$p(t) = \left[\cos\left(\frac{2\pi(1)t}{M}\right), ..., \sin\left(\frac{2\pi(N)t}{M}\right)\right]\boldsymbol{\beta} \tag{6}$$

where M represents the expected regular period of data (e.g., $M = 7$ for weekly series when using days as the time scale). The parameter vector $\boldsymbol{\beta}$ is set to satisfy $\boldsymbol{\beta} \sim \text{Normal}(0, \sigma^2)$. The factor N affects the fitting performance on the seasonality, where a large N results in more decomposed terms in $p(t)$. It allows seasonal patterns change more quickly with an increased risk of overfitting [30, 31].

In the forecasting of business time series, the holiday's effects are incorporated in a straightforward way based on the prior knowledge of the events as follows:

$$h(t) = Z(t)\boldsymbol{\kappa} \tag{7}$$

where $Z(t) = [1(t \in Day_1), ..., 1(t \in Day_L)]$ is an indicator signifying if a time t is in the period of holiday events. The factor Day is the predefined holiday days. The parameter $\boldsymbol{\kappa}$ denotes the magnitude of the change caused by the corresponding holiday events in the forecast, which is defined as $\boldsymbol{\kappa} \sim \text{Normal}(0, v^2)$. Further details about the Prophet Model can be found in [30, 32].

B. Enhancing Features of HSI with the Prophet Model

From Eqs. (3-5), it is clear that the derived trend component $g(t)$ is decided not only by the input data at time t but also by all information of the former data jointly in the time-series. Therefore, when the input data at time t is corrupted by noise, the obtained $g(t)$ can avoid the negative effect of noisy data and still derive the correct information. As such, the Prophet model has great potential to analyze the complex uncorrected HSI with noisy and water absorption bands. By taking the spectral data in hundreds of bands as a time-series, we gain insights into an HSI, including decomposing the spectral profiles into different components to derive the informative parts i.e., the trend, and mitigate the noise in an HSI.

For effective feature extraction, it is necessary to deal with the subtle interclass difference and large intraclass variations of an HSI [33]. In order to evaluate the feature characterization ability of Prophet on HSIs, the interclass and intraclass variations are calculated and assessed in this section. Here, some classic feature extraction approaches, such as 1D-SSA [13] and Mean filtering are compared with Prophet model. The raw spectral profiles are taken as the baseline for comparison, too. Besides, the Prophet model with and without seasonality settings are also evaluated to test the effects of seasonality components in feature extraction of an HSI. For 1D-SSA, the window size is set to 5×5. Only the first component is used for eigenvalue grouping according to the recommended configurations [3, 13]. The size for the mean filtering is set to 5×5. In the Prophet Model, the values of parameters τ, σ and N are set to 25, 30 and 3 empirically in this test.

For evaluating how the Prophet model may affect the interclass diversity and intraclass variance, the Kullback-Leibler divergence [34] is employed, which is based on the information entropy theory to quantify the dissimilarity (or difference, distance and discrimination) between different data. Here, the Salinas dataset [35] is taken as an example. For each comparing approach, ten samples are randomly selected from each class of the 16 labelled classes. The Kullback-Leibler divergences among the samples from every two classes, each with all other classes on average, and within each class are calculated. The results of interclass and intraclass Kullback-Leibler divergences are shown in Fig.1.
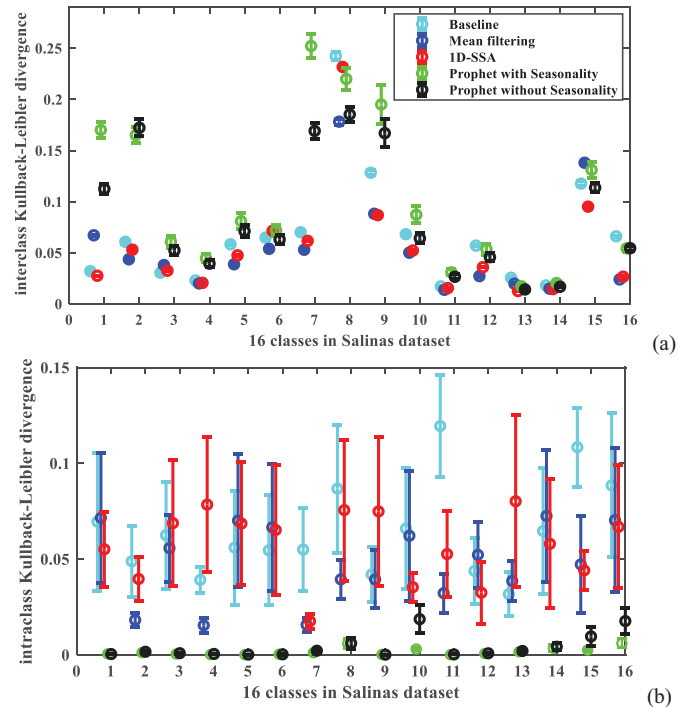

Fig.1. The Kullback-Leibler divergence of classes in Salinas dataset: (a) interclass Kullback-Leibler divergence, (b) intraclass Kullback-Leibler divergence.

As seen in Fig. 1, the Prophet model achieves the highest interclass Kullback-Leibler divergence and lowest intraclass Kullback-Leibler divergence in most classes. That is, the

Prophet model can effectively improve the disparity between classes and similarity within each class, which would help enhance further data classification [36]. This validates the potential of Prophet in an HSI feature extraction. The raw spectral profiles (Baseline) contain much noisy information. As shown in Fig.1, compared with Mean filtering and 1D-SSA, the Baseline exhibits higher interclass and intraclass Kullback-Leibler divergence. As for Mean filtering, it performs better in classes #1, #3, and #15 on interclass disparity and classes #2-#4, #7-#9, and #11 on intraclass similarity. The 1D-SSA shows higher interclass Kullback-Leibler divergence on the class #6 and better intraclass Kullback-Leibler divergence on the classes #1, #10, #12, and #14-#16. The main reason for the inferior performances of 1D-SSA and Mean filtering is that they filter out the noisy contents only using the local information within each sliding window size. On the contrary, the Prophet model can suppress the noisy and reconstruct data based on the contextual spectral information in former bands. In addition, from Fig.1, we can find that the Prophet model without Seasonality shows better results, i.e., higher interclass Kullback-Leibler divergence and lower intraclass Kullback-Leibler divergence in most cases. This may be due to that it is difficult to use the periodic Seasonality settings to characterize the complicated non-periodic changes in HSI spectra. The results from the Indian Pines and PaviaU datasets are given in Figs. S1 and S2 of the Supplementary material, which are consistent with the findings here. Thus, in this paper, we use the Prophet model without Seasonality for the further feature extraction and data classification.

III. THE PROPOSED APPROACH

The flowchart of the proposed MSPM approach is depicted in Fig. 2, which has three main stages, i.e., multiscale superpixelwise spatial constrained smoothing, multiscale Prophet model based feature extraction, and superpixelwise HSI classification and fusion, as detailed below.

A. Superpixelwise Spatial Segmentation and Smoothing

In general, adjacent pixels in an HSI often share the similar spectral profiles because of the same surface material. Through superpixelwise segmentation, these pixels can be grouped to one region (superpixel) whilst the whole image is segmented into homogeneous regions with adaptively determined sizes and shapes. To ease the computational cost, the PCA is applied onto the original HSI dataset to extract the first three principal components to form a pseudocolour image. The SLIC [37] method is then selected and applied to this pseudocolour image for superpixelwise segmentation of the HSI, as the SLIC can produce accurate boundary adherence yet with a low computational complexity [9].

Specifically, the SLIC starts with k initial cluster centres distributed on an evenly spaced grid, where k is the desired number of superpixels. It clusters on a square region sized of $2R \times 2R$ around the centres, where $R = \sqrt{n/k}$ denotes the segmentation scale, and n is the total number of pixels of the HSI. For measuring the homogeneity of a superpixel, a distance between two pixels $Dist(i,j)$ is defined, which includes the spectral distance to measure the spectra similarity and the

spatial distance to estimate the regularity, proximity and compactness of these two pixels, respectively:

$$Dist(i,j) = \sqrt{\sum_{d=1}^{D}(x_{i,d} - x_{j,d})^2 + G/R^2 ((p_i - p_j)^2 + (q_i - q_j)^2)} \quad (8)$$

where the spatial coordinates of the two pixels i and j are (p_i, q_i) and (p_j, q_j); their intensity values in the d-th band are denoted as $x_{i,d}$ and $x_{j,d}$, where $1 \leq d \leq D$, and D is the total number of bands; G is a geometric factor to balance the relative importance between the spectral and spatial features.
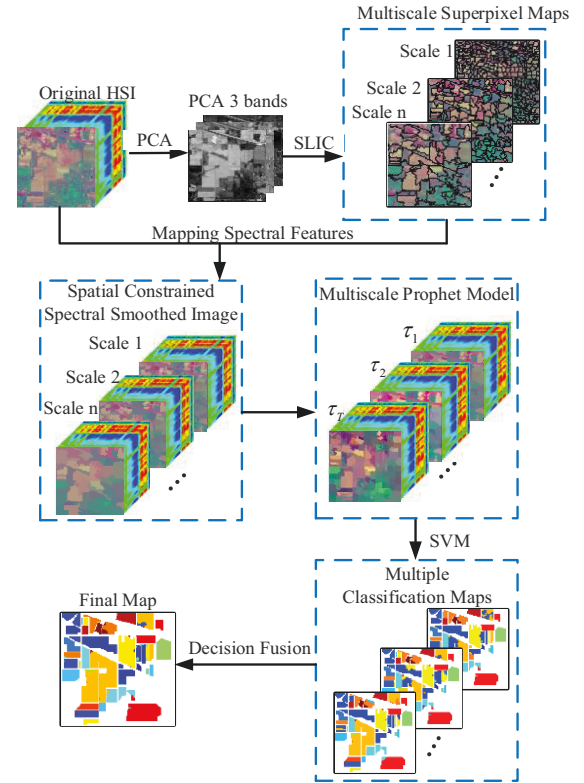

Fig.2. The flowchart of the proposed MSPM method.

For each superpixel in the generated segmentation results of the superpixel maps, the location indices of all its pixels will be recorded and mapped to all the spectral bands, resulting in non-overlapping 3-D superpixels [20]. In this way, each superpixel contains a group of adjacent pixels in D dimensions. As shown in Fig.2, superpixelwise spatially constrained image smoothing is then employed to remove the noise. Specifically, for each spectral band, the mean value of all pixels within the superpixel is taken as the value for that superpixel. In other words, superpixelwise mean spectral vectors are calculated to replace all pixels within each superpixel. In this paper, rather than setting a fixed segmentation scale for each dataset, we adopt a multiscale superpixel strategy, in which a set of segmentation scales are applied to different datasets when generating multiple superpixel maps for improved feature extraction.

The superpixelwise image smoothing can remove the noise within each superpixel to some extent. Besides, the next Prophet operation will only need to be applied to the mean spectral vector of each superpixel, which can significantly improve the efficiency of the following-on analysis.

B. Superpixelwise Multiscale Prophet Model

The Prophet model without the Seasonality was used for the further image process. Specifically, for the k-th superpixel in one superpixel map, let $SP_{k,d}$ denote its superpixelwise mean spectra at the d-th band gained in Section III.A, and the spectra data from all bands would form a vector $\boldsymbol{SP}_k$. The trend component in $SP_{k,d}^g$ can be derived via Eq. (3) as:

$$SP_{k,d}^g = C_k / \left(1 + e^{-r_d(d-m_d)}\right) \qquad (9)$$

where C_k is the carrying capacity of the k-th superpixel, which is determined based on the input spectra, i.e., the maximum value among the spectra of all bands, $C_k = \max\limits_{d \in \{1,...,D\}} SP_{k,d}$. For better capturing the statistical spectral-spatial structures in an HSI throughout the whole bands, we define each band as a changepoint. This can help to extract the local shifts of the superpixelwise profile among consecutive bands. Based on the Eqs. (3)-(5), the r_d and m_d in an HSI data can be calculated.

During this process, as explained in Eq. (3), the factor τ significantly affects the value of r in each pair of consecutive bands, which further determines the magnitude of output trend components. That is, a large τ leads to the derived trend components overfitting while a small τ causes trend components under-fitting with the input superpixelwise spectral profile. In other words, the parameter τ determines the informativeness and noise level of the input data. Generally, the noise level varies throughout the hyperspectral bands, which is also correlated to sensor characteristics, image contents and weather conditions whilst the image was acquired [38].

Taking one pixel from the Indian Pines dataset [3] as an example, Fig. 3 (a) shows the extracted trend signals under different values of τ, indicating the effect of various noise levels on the derived features. The trend function with a larger τ generates more similar estimates to the original data, although it may risk of potential overfitting. In essence, for an HSI data with a low-SNR, smaller τ values can remove most of the noisy content. However, if τ is too small, important characteristics of the spectral profile may be lost. Thus, applying a small τ to a dataset with a high-SNR can lead to the loss of discriminating information in the samples. Besides, it also shows that the Prophet model can derive the informative trend from this HSI data, which is robust to the noisy content. This will be further validated in Section V.

In addition, the effectiveness of superpixelwise Prophet model on the water absorption band is also analyzed as shown Figs. 3(b-d). Although water absorption bands are severely affected by noise, they still contain some useful information, as shown below. We take the band image with wavelength of 2499nm as an example, which is severely affected by water absorption. The reconstructed image from Fig. 3 (b), using our proposed superpixelwise Prophet model with a superpixel scale of 100 and τ of 20 is shown in Fig. 3 (c). Although Fig. 3 (b) seems to contain no useful information, the denoised version from our approach shown in Fig. 3 (c) has dramatically improved the usefulness of the features. In other words, thanks to the strong prediction capability of the Prophet model, the severely noise-affected water-absorption band now becomes reusable, in particular the well preserved and enhanced local spatial structures as compared with the GT in Fig.3(d). Overall, it has demonstrated the great value and noise-robustness of the

proposed model to extract useful information from the water-absorption bands and further improve the classification of HSI.

For the noisy HSI data, in real cases the signal and noise ratio varies differently in each band [38]. Therefore, it is a challenging or even impossible task to estimate an optimal scale of τ for the trend components for all the bands. To tackle this challenge, a multiscale approach is adopted in our MSPM framework, where a set of τ values are used, $\tau = \{\tau_1, \tau_2, ..., \tau_T\}$, representing different scales, where T is the number of scales. For various values of τ, in total T trend components can be extracted from the smoothed mean spectral vector of each superpixel. These are then taken as multiscale features of the HSI for data classification in the next stage, where the Prophet R package is used for implementation in our experiments [39].
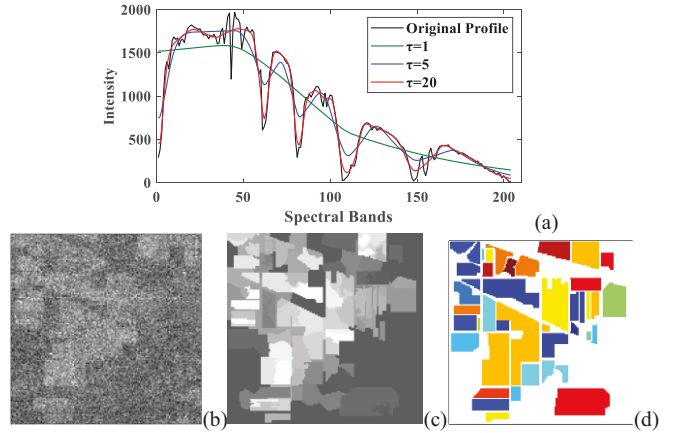

Fig. 3. The application of the proposed approach on spectral and spatial domain. (a) The original and derived (by the Prophet model) profiles with different τ values for a pixel selected from the Indian Pines scene, (b) one water absorption band at 2499nm, (c) reconstructed image from (b), and the Ground truth (d).

C. Superpixelwise HSI Classification and Decision Fusion

At a given scale, the trend components extracted from the Prophet model can be taken as the feature for the corresponding superpixel to produce a superpixelwise classification map. For T scales, we will have T classification maps, denoted as $\boldsymbol{M} = \{M_1, M_2, ..., M_T\}$. The final classification map M_f is determined by majority voting, i.e., the label that has the maximum number of appearances within $\boldsymbol{M}$:

$$M_f = \arg \max_{l=1,...,L} \sum_{t=1}^{T} F(M_t, l) \qquad (10)$$

where l represents the possible class label in HSI, and F is an indicator function given by $F(i,j) = 1 \; if \; i = j$, or $F(i,j) = 0$.

IV. EXPERIMENTAL SETTINGS AND DATASETS DESCRIPTION

For performance assessment, three publicly available HSI datasets are used in our paper. The descriptions of the datasets and the experimental settings as well as some ablation studies are presented in detail as follows.

A. Description of the Datasets

The first dataset is the Indian Pines, which was collected using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor [35] over the Indian Pines study site in Northwest Indiana, USA. This dataset contains 145×145 pixels

with a low spatial resolution of 20 m/pixel and 220 spectral bands in the wavelength range of $0.4 - 2.5\mu m$ after removing 4 invalid bands. In addition, twenty water absorption bands (104-108, 150-163 and 220) were often discarded [40] before data classification, leaving 200 spectral reflectance bands for analysis and testing. There are 16 manually defined land-cover classes, which are mostly different kinds of crops.

The second dataset is Salinas, which was also acquired via AVIRIS [35] in the Salinas Valley in California, USA. It has 512×217 pixels and 224 spectral bands with a spatial resolution of 3.7m. Similar to the first dataset, twenty water absorption bands are usually discarded. In addition, the corresponding ground-truth map also has 16 classes.

The third is the Pavia University dataset, namely PaviaU, which was acquired by the Reflective Optics System Imaging Spectrometer sensor over Pavia, Northern Italy [41]. It contains 114 spectral bands in a spectral range of $0.43 - 0.86\mu m$ and 610×340 pixels with a spatial resolution of 1.3m. Similarly, for avoiding the effect of water absorption, the available number of bands was reduced from 114 to 103. The ground truth has nine classes of land covers. Further detail including the number of samples in each class within these datasets can be found in [20].

B. Experimental Settings

For performance evaluation, two groups of experiments are designed, using respectively the uncorrected HSI dataset (without removing any bands), and the corrected dataset (with the noisy and water absorption bands removed). This is to validate the robustness of our MSPM as most of the existing approaches only work well on the corrected HSI datasets.

TABLE I
LIST OF PARAMETER SETTINGS FOR OUR PROPOSED MSPM FRAMEWORK AND OTHER BENCHMARKING APPROACHES

Method	Settings
SVM-spe	N/A
2D-EMD	Stop threshold: 0.2; IMF Grouping (IMFG): 1-3rd for Indian Pines, 1-4th for Salinas and PaviaU
SMLR-SpATV	Smoothness parameter: 2
MSP-SVMsub	superpixel number: {5, 10, 15, 25, 50, 75, 100}
SS-LRR	superpixel number: Indian Pines/Salinas (40), PaviaU (190)
FS^{2}LRL	subspace dimension: 10; superpixel number: 100; sparse parameter: 0.13; fidelity parameter: 0.040
2D-MSSP	Window size: $\{5 \times 5, 10 \times 10, 20 \times 20, 40 \times 40, 60 \times 60\}$ NPCs: 40
Proposed MSPM	τ: {5, 10, 20, 25, 30} superpixel scale: {50, 100, 200, 400, 600}

First, comparison analysis was conducted and benchmarked with six state-of-the-art spectral-spatial HSI classification methods, using both the corrected and uncorrected HSI images. These include two signal decomposition-based methods, 2D-EMD [16] and 2D-MSSP [29], the multiscale superpixel-based method, MSP-SVMsub [9], an MRF weighted spatial-spectral classifier, SMLR-SpATV [10] and two superpixelwise low rank representation based denoising techniques, SS-LRR [12] and fast superpixel based subspace low rank learning method (FS^{2}LRL) [28], which exhibited superiority in directly classifying the uncorrected HSI data. Our MSPM is also compared with some deep learning approaches. Besides, the approach to apply the SVM classifier on the raw spectral profiles of HSI for classification, denoted as SVM-spe, is taken

as the baseline. The ablation experiments for comparing the effectiveness of different components of the proposed approach are given in the Table S1 of Supplementary.

The parameter settings of relevant methods are summarized in Table I. For our MSPM, the superpixel segmentation scales were set to $\{50, 100, 200, 400, 600\}$ for all HSI datasets according to the parameter analysis in Subsection C. As for the scale τ of the Prophet model, it varies within the range of $\{1, 5, 10, 15, 20, 25, 30, 35\}$ in different permutations and combinations. For balancing the accuracy and efficiency, the optimal set of $\{5, 10, 20, 25, 30\}$ was employed based on the parameter analysis in Subsection C. The base rate r_0 and the base offset value m_0 are set by using Stan's L-BFGS in the Prophet R package by determining a maximum a *posteriori* estimate [32]. For SVM, the LIBSVM library was utilized for implementing multiclass classification [42, 43]. According to [16, 44], the Gaussian radial basis function (RBF) was employed, where the kernel factor and penalty parameter were optimally determined via a grid search to 0.125 and 1024, respectively. To ensure a fair comparison, we keep the SVM parameters consistent for all experiments. The other parameters for the benchmarking approaches are set according to their recommended default values [13] as shown in Table I.

To reduce the random discrepancies and avoid systematic errors, all experiments were independently run ten times. In each run, the training and testing sets were randomly divided with no overlapping in between. To fully evaluate the performance of these methods, experiments were also conducted with different numbers of training samples. In particular, we mainly concern the problem of a small training set. Hence, the number of selected samples m set to 5, 10 and 20 for each land cover class in the training set via stratified sampling. If the total number of samples within the land cover class is less than 30, such as grass/pasture-mowed and oats in Indian Pines, 50% of samples in each class will be used for training [45]. In addition, the overall accuracy (OA), Kappa coefficients (κ) and average accuracy (AA) are used for quantitative evaluation of the classification results.

All involved approaches were implemented using the Matlab 2018a platform on a computer with an Intel (R) Core (TM) i7-8700 CPU (3.20 GHz) and 16.0 GB of memory. The Prophet model was implemented on the R 3.5.3.

C. Parameter Analysis

1) Segmentation scale

The scale of superpixel segmentation is a key parameter that affects the generated superpixels and the classification results [17]. Here, the optimal scale values for the three datasets were tested, which were set to $\{10, 50, 100, 200, 400, 600, 800\}$. Note that the numbers of training and test samples were set the same as aforementioned in Subsection IV(B).

Fig. 4 illustrates the mean OA over ten independent runs of the proposed MSPM using different segmentation scales on the three corrected datasets, under various numbers of training samples per class. In general, more training samples will lead to better classification results in terms of higher OA values. For a given training size, the segmentation scale in MSPM determines the homogeneity within a local region and thus affects the classification accuracy. An appropriate scale value can generate more accurate object boundaries. The landcovers

in three datasets have different characteristics and complexity. As seen in Fig.4, the best segmentation scales for Indian Pines, Salinas and PaviaU are found to be 100, 600 and 200, respectively, as they can help to generate the highest OA for a given number of training samples. Specifically, the proper range of segmentation scales for Indian Pines, Salinas and PaviaU are supposed to be 50~400, 50~800 and 50~600, respectively. Therefore, in the multiscale superpixel segmentation, the set $\{50, 100, 200, 400, 600\}$ is adopted for three datasets in this paper in the remaining experiments.
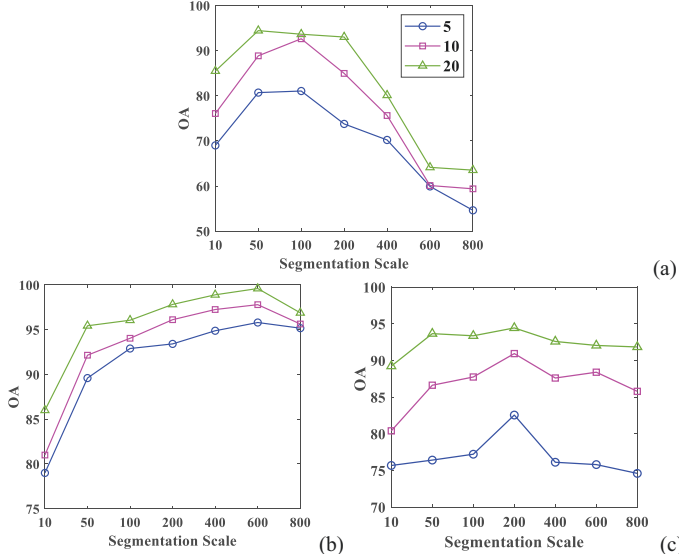

Fig. 4. The OA of MSPM with different superpixel segmentation scales for three datasets of Indian Pines (a), Salinas (b) and PaviaU (c).

2) The parameter τ

To test the optimal sets of parameter τ on three datasets, the achieved class-based accuracy and OA of the proposed method with varying τ within $\{1, 5, 10, 15, 20, 25, 30, 35\}$ are analysed in this section. The results of class-based accuracy are given in in Fig. S3 of the Supplementary. The number of training samples are set to 5 for each land cover for all tests.
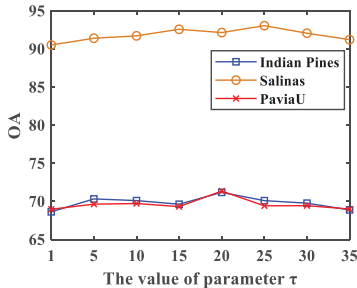

Fig. 5. The effect of the parameter τ of MSPM to the OA.

From Fig. 5, it is clear that the value of parameter τ in the range [5,30] enables the MSPM to achieve better classification results in terms of higher OA values. As for the class-based classification performance shown in Fig. S3, on the Indian Pines dataset, the value 5, 10, 20 and 25 for parameter τ leads to the best accuracy in the three classes (the classes #3, #5 and #16), two classes (the classes #6 and #15), two classes (the classes #10 and #14) and two classes (the class #2 and #11), respectively. On the corrected Salinas dataset, apart from the classes #1, #9 and #12, where the 100% accuracy of MSPM are produced on all cases, the best results are generated when the parameter τ equal to 20 and 25, in which MSPM achieve the highest accuracy in three classes, respectively. The proposed MSPM also realize better performance when parameter τ are set to 5, 15 and 30. The best value of parameter τ for PaviaU are found to be 1 and 5, as they can help to generate the highest accuracy on the classes #1 and #4, and #3 and #5, respectively. The value 10, 20, 25, 30 and 35 for parameter τ also lead to the best classification performance on the classes #9, #2, #6, #8 and #7, respectively. To sum up, considering the OA and class-based performance of MSPM on three datasets, we employ the set of $\{5, 10, 20, 25, 30\}$ for the parameter τ in this paper.

V. EXPERIMENTAL RESULTS

In this section, qualitative and quantitative results from the three HSI datasets are presented for performance assessment of the proposed MSPM approach. Relevant results from uncorrected and corrected HSI datasets as well as the efficacy of different strategies in MSPM and further comparison with several deep learning approaches are detailed as follows.

A. Results from the Indian Pines Dataset

Indian Pines dataset is heavily corrupted by noise and water absorption [28]. Here, corrected and uncorrected Indian Pines datasets, where noisy and water absorption bands are removed or kept, respectively, are used for quantitative and qualitative performance evaluation as detailed below.

1) Quantitative evaluation

For quantitative evaluation, the overall accuracy (OA) and the Kappa coefficient (κ) are utilized. The proposed MSPM approach has been compared with 2D-EMD, 2D-SSA, SMLR-SpATV, MSP-SVMsub, SS-LRR, FS²LRL, SVM-spe, and SP-SVM (only use superpixelwise smoothing yet without Prophet model). Relevant results are reported in Table II. Note that the best results in each row are highlighted in bold for comparison.

First of all, all the results in Table II are positively correlated with the increasing number of training samples. With the introduced spatial features, 2D-EMD, 2D-MSSP and SMLR-SpATV have all yielded better classification results than SVM-spe, in which no spatial features are used. With superpixel

TABLE II
THE OA (%) AND κ (IN PARENTHESES) OF DIFFERENT METHODS UNDER VARIOUS NUMBERS OF TRAINING SAMPLES
FOR THE INDIAN PINES DATASET WITH OR WITHOUT REMOVING THE NOISY AND WATER ABSORPTION BANDS

Samples	Dataset	SVM-spe	2D-EMD	SMLR-SpATV	MSP-SVMsub	SS-LRR	FS²LRL	2D-MSSP	MSPM
5	Uncorrected	41.77 (0.35)	54.16 (0.49)	54.69 (0.50)	62.39 (0.58)	63.51 (0.59)	55.98 (0.51)	82.89 (0.81)	**87.92 (0.86)**
	Corrected	46.85 (0.41)	57.71 (0.53)	56.08 (0.52)	68.09 (0.68)	60.95 (0.56)	56.13 (0.51)	83.27 (0.82)	**87.91 (0.86)**
10	Uncorrected	49.80 (0.44)	67.35 (0.63)	73.63 (0.70)	82.01 (0.80)	75.63 (0.73)	67.30 (0.63)	90.15 (0.89)	**93.54 (0.93)**
	Corrected	54.91 (0.50)	71.88 (0.68)	73.95 (0.71)	83.57 (0.81)	74.26 (0.71)	69.60 (0.66)	90.13 (0.89)	**92.03 (0.91)**
20	Uncorrected	57.22 (0.52)	79.77 (0.77)	83.12 (0.81)	88.99 (0.88)	85.11 (0.83)	77.64 (0.75)	94.92 (0.94)	**95.74 (0.95)**
	Corrected	63.44 (0.59)	84.16 (0.83)	83.12 (0.81)	90.12 (0.89)	83.64 (0.81)	78.95 (0.76)	94.31 (0.94)	**95.18 (0.95)**

based smoothing, more effective spatial features can be extracted thus the further improved results from MSP-SVMsub. This superior performance is mainly due to the utilization of adaptive local structures in superpixelwsie segmentation and smoothing, which has enhanced the consistency of the spectral features better than using a fix-sized window. Besides, SS-LRR and FS²LRL also improve the classification accuracy due to the effect of their denoising operations. Overall, the proposed MSPM remains the best and constantly produces the highest accuracy in terms of OA and κ. It is worth noting that the 22D-MSSP method also works effectively for the HSI feature extraction, which shows competitive results to MSPM on both corrected and uncorrected Indian Pines dataset, and ranks second among all. Compared with 2D-MSSP, thanks to the introduced Prophet model, our MSPM has further improved the spectral feature extraction. The MSPM achieves better performance than 2D-MSSP especially when the training size is small. In the case of 5 training samples, the classification accuracy of MSPM is about 5% higher than that of 2D-MSSP. This has validated the efficacy of our proposed MSPM approach in addressing the Hughes phenomenon under very limited training samples.

When comparing the performance of each method on corrected and uncorrected Indian Pines dataset, we can find that in most cases, SVM-spe, 2D-EMD, SMLR-SpATV, MSP-SVMsub and FS²LRL achieve better results on corrected dataset. The 2D-MSSP and advanced noise reduction method, SS-LRR, can effectively remove the effect of noise while preserving the spectral features. As a result, they achieve higher accuracy on the uncorrected HSI in most cases. For our proposed MSPM approach, in almost all cases the results on the uncorrected dataset have outperformed those from the corrected one. Thanks to our multiscale superpixelwise smoothing and multiscale Prophet modelling, the proposed MSPM model can successfully suppress the noise and produce satisfactory results, even without removing the low-SNR and water absorption bands. This has provided additional benefits for efficient and effective interpretation of HSI in an automatic way, as no extra work is needed to filter out those unwanted bands.

2) Visual and class-based performance assessment

Although the superior performance of our MSPM has been validated in quantitative assessment, here visual and class-based performance evaluation is used to show how exactly the classification errors has been reduced from the uncorrected dataset. Three methods in Table II with promising performance, i.e., SMLR-SpATV, MSP-SVMsub and 2D-MSSP, are selected for further comparison. For visual assessment, the classification maps from our approach and the three others on both the uncorrected and corrected dataset under 20 training samples per class are compared in Fig. S4 of the Supplementary, where black and magenta circles are used to mark respectively the

correct and incorrect classification. In Table III, class-by-class classification results are also compared in terms of the OA, average accuracy (AA) and Kappa coefficient.

First of all, as shown in Table III, MSPM has achieved the best classification accuracy in 13 and 11 out of 16 classes on the uncorrected and corrected datasets, respectively, especially in the classes 3, 10 and 11 (*Corn-min till*, *Soybeans-no till* and *Soybeans-min till*). Overall, MSPM ranks first in terms of OA, AA and Kappa coefficient, and shows better accuracy on the uncorrected dataset. This has demonstrated again the efficacy of MSPM on classification of the Indian Pines dataset with limited and unbalanced samples. Note that the recently proposed 2D-MSSP also shows high efficiency in noise-robust feature extraction with promising class-based performance.

TABLE III
CLASSIFICATION ACCURACY (%) OF ALL INVOLVED METHODS FOR THE INDIAN PINES DATA WITH 20 TRAINING SAMPLES PER CLASS

Class	Uncorrected				Corrected			
	SMLR-SpATV	MSP-SVMsub	2D-MSSP	MSPM	SMLR-SpATV	MSP-SVMsub	2D-MSSP	MSPM
1	99.28	**100.00**	100.00	100.00	100.00	100.00	100.00	100.00
2	73.80	85.58	93.62	**94.29**	71.05	88.99	**97.66**	96.28
3	77.25	88.27	96.75	**99.88**	76.65	90.49	94.69	**99.75**
4	99.80	**100.00**	100.00	100.00	98.06	**100.00**	100.00	100.00
5	88.02	94.38	97.69	**99.81**	89.37	84.45	**96.69**	93.54
6	98.61	98.87	**99.87**	98.79	98.94	97.89	98.10	**99.44**
7	99.76	100.00	100.00	100.00	100.00	100.00	100.00	100.00
8	99.95	99.13	100.00	100.00	99.91	100.00	100.00	100.00
9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
10	79.80	86.66	89.89	**99.47**	82.02	90.02	94.75	**99.38**
11	71.73	80.29	90.34	**90.83**	71.14	80.99	85.13	**88.14**
12	86.41	90.58	**92.95**	81.87	89.32	91.27	**91.36**	83.49
13	99.95	96.22	**100.00**	100.00	100.00	98.92	100.00	100.00
14	93.66	93.09	99.52	**100.00**	94.48	96.39	100.00	100.00
15	94.10	98.63	**100.00**	99.73	95.55	97.81	100.00	99.73
16	99.86	**100.00**	100.00	100.00	100.00	100.00	100.00	99.04
OA	83.12	88.99	94.92	**95.74**	83.12	90.12	94.31	**95.18**
κ	0.81	0.88	0.94	**0.95**	0.81	0.89	0.94	**0.95**
AA	91.37	94.48	97.54	**97.79**	91.66	94.83	97.40	**97.42**

B. Results from the Salinas Dataset

Similarly, the Salinas dataset with or without noisy and water absorption bands, denoted as uncorrected Salinas and corrected Salinas, respectively, are employed for performance evaluation. Relevant results under different numbers of training samples are reported in Table IV for quantitative comparison.

1) Quantitative evaluation

In Table IV, the similar factual conclusions can be made as from the Indian Pines dataset. The difference here is that the addition of various spatial features contributes less than that on the Indian Pines dataset. The possible reason is the high intra-class homogeneity caused by low-level noise and geometrically structured simple agricultural objects in the image. Overall, the proposed MSPM remains the best and constantly produces the

TABLE IV
THE OA (%) AND κ (IN PARENTHESES) OF DIFFERENT METHODS UNDER VARIOUS NUMBERS OF TRAINING SAMPLES FOR THE SALINAS DATASET WITH OR WITHOUT REMOVING THE NOISY AND WATER ABSORPTION BANDS

Samples	Dataset	SVM-spe	2D-EMD	SMLR-SpATV	MSP-SVMsub	SS-LRR	FS²LRL	2D-MSSP	MSPM
5	Uncorrected	79.79 (0.78)	80.22 (0.78)	86.54 (0.85)	91.73 (0.91)	84.12 (0.82)	85.64 (0.84)	95.81 (0.95)	**96.14 (0.96)**
	Corrected	80.99 (0.79)	80.70 (0.79)	86.42 (0.85)	89.18 (0.88)	82.46 (0.81)	83.85 (0.82)	96.33 (0.96)	**96.54 (0.96)**
10	Uncorrected	83.53 (0.82)	83.76 (0.82)	90.77 (0.90)	95.70 (0.95)	87.33 (0.86)	89.35 (0.88)	97.06 (0.96)	**97.64 (0.97)**
	Corrected	83.62 (0.82)	84.05 (0.82)	90.77 (0.90)	93.30 (0.93)	87.08 (0.86)	88.93 (0.88)	96.83 (0.97)	**97.55 (0.97)**
20	Uncorrected	86.27 (0.85)	86.03 (0.85)	92.89 (0.92)	97.81 (0.98)	92.23 (0.91)	93.51 (0.93)	99.00 (0.99)	**99.16 (0.99)**
	Corrected	86.38 (0.85)	86.33 (0.85)	93.12 (0.92)	97.49 (0.97)	90.71 (0.90)	93.02 (0.92)	98.65 (0.99)	**98.67 (0.99)**

highest OA and κ except one occasion on the corrected dataset with 5 training samples per class. In most cases, MSPM has better results on the uncorrected dataset than the corrected one. These results have clearly shown the efficacy of our MSPM for HSI classification, which is again mainly due to combining multiscale superpixelwise segmentation and smoothing along with the multiscale Prophet model based noise-robust feature mining. In the classification of Salinas dataset, both the 2D-MSSP and MSPM achieve enough accuracy when the training size is 20. The 2D-MSSP shows competitive results. When comparing the OA and κ of other methods on the corrected and uncorrected Salinas datasets, in all cases, SVM-spe and 2D-EMD generate better results on the corrected Salinas than the uncorrected one. As for the SMLR-SpATV and 2D-MSSP, they have inconsistent results though achieve better results in more cases from the uncorrected dataset. The two denoising methods, SS-LRR and FS²LRL, and MSP-SVMsub produce better classification on uncorrected Salinas in almost all cases.

2) Visual and class-based performance assessment

In addition, results of class-by-class classification from SMLR-SpATV, MSP-SVMsub, 2D-MSSP and MSPM with 20 training samples per class are compared in Table V, with the classification maps given in Fig. S5 of the Supplementary for visual comparison. In Table V, the proposed MSPM produces the best results in 14 and 13 out of 16 land cover classes on the uncorrected and corrected dataset, respectively. The 2D-MSSP ranks the second, which achieves the best results in 10 out of 16 land cover classes on both uncorrected and corrected datasets. Apart from SMLR-SpATV, all the methods in this subsection ha ve achieved better OA on the uncorrected Salinas dataset. This has validated the efficacy of the useful information in water vapor absorption bands.

TABLE V
CLASSIFICATION ACCURACY (%) OF ALL INVOLVED METHODS FOR THE SALINAS DATA WITH 20 TRAINING SAMPLES PER CLASS

Class	Uncorrected				Corrected			
	SMLR-SpATV	MSP-SVMsub	2D-MSSP	MSPM	SMLR-SpATV	MSP-SVMsub	2D-MSSP	MSPM
1	99.91	99.80	**100.00**	**100.00**	99.94	**100.00**	**100.00**	**100.00**
2	**100.00**	97.30	**100.00**	**100.00**	99.98	99.38	**100.00**	**100.00**
3	95.43	100.00	**100.00**	**100.00**	94.28	99.90	**100.00**	**100.00**
4	99.10	97.31	99.85	**99.93**	98.56	97.02	**99.93**	**99.93**
5	98.38	96.31	**99.56**	98.95	98.59	98.87	97.36	**98.95**
6	100.00	98.45	99.79	**100.00**	100.00	98.98	99.92	**100.00**
7	99.94	98.96	99.64	**99.92**	99.94	99.38	99.89	**99.92**
8	80.18	96.20	95.78	**96.96**	81.22	**95.63**	94.74	94.63
9	99.97	**100.00**	**100.00**	**100.00**	**100.00**	98.84	**100.00**	**100.00**
10	95.80	96.13	**99.82**	**99.82**	93.93	95.33	**99.94**	99.88
11	99.88	99.62	**100.00**	**100.00**	99.87	99.90	**100.00**	**100.00**
12	99.97	96.54	99.74	**100.00**	99.97	99.21	**100.00**	**100.00**
13	99.23	98.88	**100.00**	**100.00**	99.69	99.78	99.90	**100.00**
14	99.88	96.29	**100.00**	**100.00**	99.84	98.76	99.91	**100.00**
15	82.23	98.05	**99.72**	99.03	83.46	94.34	**99.29**	98.93
16	99.29	98.43	**100.00**	**100.00**	99.16	98.10	**100.00**	**100.00**
OA	92.89	97.81	99.00	**99.16**	93.12	97.49	98.65	**98.67**
κ	0.92	0.98	0.99	**0.99**	0.92	0.97	0.99	**0.99**
AA	96.83	98.02	99.62	**99.66**	96.78	98.34	99.43	**99.51**

C. Results from the PaviaU Dataset

In this subsection, the performance of our proposed MSPM is further evaluated using the PaviaU dataset. Here the performances of all comparing approaches on corrected PaviaU

dataset are analyzed. As the uncorrected PaviaU dataset with water absorption bands is unavailable, a simulated noisy dataset is produced with results given in Table S2 of Supplementary.

1) Quantitative evaluation

In Table VI, our proposed MSPM is compared with SVM-spe, 2D-EMD, 2D-MSSP, SMLR-SpATV, MSP-SVMsub, SS-LRR and FS²LRL, where class-based classification results are also shown. Similarly, the best quantitative results are labelled in bold for clarity. As seen in Table VI, the proposed MSPM has produced the best results in terms of the highest OA, κ and AA, which is followed by the 2D-MSSP. The proposed method MSPM and 2D-MSSP are significantly better than other comparing methods. For the class-based performance, MSPM produces the highest accuracy in 6 out of 9 classes and realizes correct classification of all testing data in class #6 and #7.

TABLE VI
CLASSIFICATION ACCURACY (%) OF ALL INVOLVED METHODS FOR THE CORRECTED PAVIAU DATA WITH 20 TRAINING SAMPLES PER CLASS

Class	SVM-spe	2D-EMD	SMLR-SpATV	MSP-SVMsub	SS-LRR	FS2 LRL	2D-MSSP	MSPM
1	67.87	69.08	83.62	92.80	73.19	77.20	**93.11**	90.38
2	76.87	77.78	86.11	89.22	76.13	85.59	99.12	**99.31**
3	74.54	72.24	87.50	92.74	79.54	77.18	98.42	**99.04**
4	91.37	90.98	92.45	90.97	90.10	90.59	83.91	**92.95**
5	99.09	99.08	**100.00**	99.70	99.23	99.28	96.82	99.78
6	70.60	75.83	86.58	97.34	77.73	89.44	99.88	**100.00**
7	89.42	90.70	99.11	96.41	93.06	92.16	**100.00**	**100.00**
8	75.93	77.69	77.93	95.88	81.66	88.02	94.89	**99.02**
9	**99.91**	99.85	60.70	91.15	99.90	99.86	87.03	79.84
OA	77.15	78.41	85.88	92.19	79.26	85.83	96.49	**97.12**
κ	0.71	0.72	81.71	0.90	0.74	0.82	0.95	**0.96**
AA	82.84	83.69	86.00	94.02	73.20	77.20	94.80	**95.59**

Table VII shows the classification results from the corrected PaviaU dataset using different numbers of training samples for all the involved algorithms. As seen, MSPM still achieves the highest OA and κ in all the experiments. With a lower number of spectral bands than the other two datasets, the Hughes effect in PaviaU dataset is not apparent [9]. As a result, spectral-spatial classification strategies have shown less contributions than those on the other two datasets. Nevertheless, the superpixel based approaches, MSP-SVMsub still exhibit superior performance due to the preserved local homogeneous regions especially when the number of training samples is above 5. Besides, the proposed MSPM and 2D-MSSP show significantly better results than other methods, which validates their effectiveness in feature extraction. In general, the proposed MSPM has again produced effective and reliable results on the PaviaU dataset.

TABLE VII
THE OA (%) AND κ (IN PARENTHESES) OF DIFFERENT METHODS ON THE CORRECTED PAVIAU DATASET UNDER VARIOUS TRAINING SIZES

No. Samples	5	10	20
SVM-spe	63.79 (0.55)	71.42 (0.64)	77.15 (0.71)
2D-EMD	63.76 (0.55)	71.41 (0.64)	78.41 (0.72)
SMLR-SpATV	68.02 (0.60)	78.03 (0.72)	85.88 (0.82)
MSP-SVMsub	66.26 (0.59)	82.65 (0.78)	92.19 (0.90)
SS-LRR	63.21 (0.55)	73.62 (0.67)	79.26 (0.74)
FS2LRL	65.86 (0.58)	78.55 (0.73)	85.83 (0.82)
2D-MSSP	87.66 (0.84)	90.66 (0.88)	96.49 (0.95)
MSPM	**89.43 (0.86)**	**95.74 (0.94)**	**97.12 (0.96)**

2) *Visual comparison*

Visual comparison of the classification maps from our MSPM and other comparing methods on the corrected PaviaU dataset are shown in Fig. S6, where black and magenta circles denote correct and incorrect classified pixels to show the best results that have been achieved by the proposed MSPM method.

D. *Running time Comparison*

The running time of different methods are compared in this subsection. As seen in Table VIII, the methods 2D-EMD, 2D-MSSP, MSP-SVMsub, and MSPM have longer running time than SMLR-SpATV, SS-LRR and FS^{2}LRL. Therein 2D-MSSP, MSP-SVMsub, and MSPM all use the multiscale strategy, inevitably leading to a high computational complexity. However, these three methods have produced the best results on all the datasets. The proposed MSPM unfortunately requires the longest processing time, followed by 2D-EMD without using the multiscale strategy hence quite poor results. In this paper, the improvement of the classification accuracy under small number of training samples is first emphasized. The fast Prophet model will be explored in the future study.

TABLE VIII
THE RUNNING TIME (S) OF DIFFERENT METHODS UNDER 20 TRAINING SAMPLES PER CLASS ON THREE HSI UNCORRECTED DATASETS

Dataset	2D-EMD	SMLR-SpATV	MSP-SVMsub	SS-LRR	FS^{2}LRL	2D-MSSP	MSPM
Indian Pines	843	105	374	43	116	476	862
Salinas	3948	530	1190	237	600	3704	6554
Pavia	3209	409	876	304	519	3964	3575

E. *Comparing with other Methods include Deep Learning*

We further compare our MSPM with five state-of-the-art approaches including four deep learning approaches, i.e., GLapLRR [46], SSCasRNN [23], GBRBM in parallel [24], FSCN [2], and DR-CNN [22] on the corrected Indian Pines, Salinas, and PaviaU datasets, and the results are given in Table IX. Note for the three datasets, we only use 20 samples per class for training, in comparison to 10% training samples used in GLapLRR, SSCasRNN, and GBRBM in parallel, which equals to on average 65, 338 and 475 samples per class.

TABLE IX
COMPARING MSPM WITH FOUR DEEP LEANING APPROACHES USING OA (%) ON THREE CORRECTED DATASETS WITH THE NUMBER OF TRAINING SAMPLES

Corrected Datasets	Training samples	Glap LRR	SSCas-RNN	GB-RBM	FSCN	DR-CNN	MSPM
Indian Pines	65/64/50/20	**96.42**	91.79	-	90.45	94.94	95.18
Salinas	338/-/50/20	-	-	95.94	-	95.54	**98.67**
PaviaU	475/238/50/20	86.39	90.30	96.22	98.19	**98.67**	97.12

The FSCN adopts 64 training samples per class on the corrected Indian Pines dataset, and 238 samples per class on the corrected PaviaU dataset. For DR-CNN, 50 training samples per class are used in each dataset. Obviously, in most cases MSPM has produced quite comparable or even slightly better results than those methods, including four deep learning models. In summary, this has validated again the superiority of our proposed MSPM approach in noise-robust spectral-spatial feature extraction and data classification of the HSI.

VI. Conclusion

In this paper, we have proposed Multiscale Superpixelwise Prophet model (MSPM), a novel spectral-spatial feature mining framework for noise-robust feature extraction and effective data classification of the HSI. First, we demonstrate that the Prophet model is able to enhance the HSI features in terms of reduced intraclass variance and enlarged interclass diversity. Second, superpixelwise image segmentation has found particularly useful for grouping local spectrally similar pixels and reducing the high intra-class heterogeneity and inter-class homogeneity of different land cover classes in the HSI. Third, our MSPM model has successfully exploited spectral data at different noise levels. The Prophet model has also contributed noticeably to the classification, especially to the uncorrected datasets. The superpixelwise segmentation and the Prophet model can supplement to each other in enhancing the features in the spatial and spectral domains, respectively. As a result, the joint spectral-spatial features can more effectively characterize both the corrected and uncorrected HSI datasets, especially with very limited training samples. Our MSPM model has significantly outperformed a few state-of-the-art approaches, including several deep learning models along with much more training samples. The improved classification results from the uncorrected datasets have enabled potentially a new and fully automatic roadmap for interpreting the HSI where conventional wisdom of pre-filtering of unwanted bands can be skipped.

Future work will focus on the analysis of the non-periodic seasonality effects in the spectral profile of the HSI. Besides, fast and adaptive Prophet model will be investigated, where an adaptive scale will be exploited to replace the multiscale strategy for improved computational efficiency as well as new applications in missing data restoration and HSI reconstruction. Since the proposed method performs well with limited training sample size, the unsupervised Prophet model for the HSI data classification will be explored in the future work.

References

[1] L. Zhang, Q. Zhang, B. Du, *et al.*, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.,* vol. 48, no. 1, pp. 16-28, Jan. 2016.

[2] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Trans. Image Process.,* vol. 30, pp. 2810-2825, Feb. 2021.

[3] J. Zabalza, J. Ren, J. Zheng, *et al.*, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 8, pp. 4418-4433, Aug. 2015.

[4] J. Liang, J. Zhou, L. Tong, *et al.*, "Material based salient object detection from hyperspectral images," *Pattern. Recogn.,* vol. 76, pp. 476-490, Apr. 2018.

[5] Y. Yan, J. Ren, J. Tschannerl, *et al.*, "Nondestructive phenolic compounds measurement and origin discrimination of peated barley malt using near-infrared hyperspectral imagery and machine learning," *IEEE Trans. Instrum. Meas.,* vol. 70, pp. 1-15, May 2021.

[6] T. Qiao, J. Ren, J. Zabalza, *et al.*, "Singular spectrum analysis for hyperspectral imaging based beef eating quality evaluation: a new pre-processing method," *Comput. Electron. Agric.,* vol. 115, pp. 21-25, Jul. 2015.

[7] M. N. Siti, J. Ren, M. Stephen, *et al.*, "Hyperspectral Image Enhancement and Mixture Deep-Learning Classification of Corneal Epithelium Injuries," *Sensors,* vol. 17, no. 11, pp. 2644, Nov. 2017.

[8] M. D. Santamaria, J. Fernandez-Novales, A. M. Fernandes, *et al.*, "Use of visible and short-wave near-infrared hyperspectral imaging to

fingerprint anthocyanins in intact grape berries," *J. Agric. Food. Chem.,* vol. 64, no. 40, pp. 7658-7666, Sep. 2016.

[9] H. Yu, L. Gao, W. Liao, *et al.*, "Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.,* vol. 14, no. 11, pp. 2142-2146, Nov. 2017.

[10] L. Sun, Z. Wu, J. Liu, *et al.*, "Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 3, pp. 1490-1503, Mar. 2015.

[11] L. Zhang, Q. Zhang, L. Zhang, *et al.*, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern. Recogn.,* vol. 48, no. 10, pp. 3102-3112, Oct. 2015.

[12] F. Fan, Y. Ma, C. Li, *et al.*, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.,* vol. 397, pp. 48-68, Aug. 2017.

[13] J. Zabalza, J. Ren, Z. Wang, *et al.*, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geosci. Remote Sens. Lett.,* vol. 11, no. 11, pp. 1886-1890, Nov. 2014.

[14] J. Xia, J. Chanussot, P. Du, *et al.*, "(Semi-) Supervised probabilistic principal component analysis for hyperspectral remote sensing image classification," *IEEE J-STARS,* vol. 7, no. 6, pp. 2224-2236, Jun. 2014.

[15] N. Kumar, P. Uppala, K. Duddu, *et al.*, "Hyperspectral tissue image segmentation using semi-supervised NMF and hierarchical clustering," *IEEE Trans. Med. Imaging,* vol. 38, no. 5, pp. 1304-1313, May 2019.

[16] B. Demir, and S. Erturk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 48, no. 11, pp. 4071-4084, Nov. 2010.

[17] S. Jia, X. Deng, J. Zhu, *et al.*, "Collaborative representation-based multiscale superpixel fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 57, no. 10, pp. 7770-7784, Oct. 2019.

[18] W. Zhao, and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.,* vol. 54, no. 8, pp. 4544-4554, Aug. 2016.

[19] S. Jia, L. Shen, J. Zhu, *et al.*, "A 3-D Gabor phase-based coding and matching framework for hyperspectral imagery classification," *IEEE Trans. Cybern.,* vol. 48, no. 4, pp. 1176-1188, Apr. 2018.

[20] L. Fang, S. Li, W. Duan, *et al.*, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 12, pp. 6663-6674, Dec. 2015.

[21] Y. Yuan, J. Lin, and W. Qi, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.,* vol. 46, no. 12, pp. 2966-2977, Dec. 2016.

[22] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.,* vol. 27, no. 6, pp. 2623-2634, Jun. 2018.

[23] R. Hang, Q. Liu, D. Hong, *et al.*, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 57, no. 8, pp. 5384-5394, Aug. 2019.

[24] K. Tan, F. Wu, Q. Du, *et al.*, "A parallel Gaussian–Bernoulli restricted boltzmann machine for mining area classification with hyperspectral imagery," *IEEE J-STARS,* vol. 12, no. 2, pp. 627-636, Feb. 2019.

[25] J. Zabalza, J. Ren, J. Zheng, *et al.*, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing,* vol. 185, pp. 1-10, Apr. 2016.

[26] G. Sun, H. Fu, J. Ren, *et al.*, "Spassa: Superpixelwise adaptive ssa for unsupervised spatial-spectral feature extraction in hyperspectral image," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6158-6169, Jul. 2022.

[27] W. He, H. Zhang, L. Zhang, *et al.*, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J-STARS,* vol. 8, no. 6, pp. 3050-3061, Jun. 2015.

[28] L. Sun, B. Jeon, B. N. Soomro, *et al.*, "Fast superpixel based subspace low rank learning method for hyperspectral denoising," *IEEE Access,* vol. 6, pp. 12031-12043, Feb. 2018.

[29] P. Ma, J. Ren, H. Zhao, *et al.*, "Multiscale 2D Singular Spectrum Analysis and Principal Component Analysis for Spatial-Spectral Noise-Robust Feature Extraction and Classification of Hyperspectral Images," *IEEE J-STARS,* vol. 14, pp. 1233-1245, Nov. 2020.

[30] S. J. Taylor, and B. Letham, "Forecasting at scale," *The American Statistician,* vol. 72, no. 1, pp. 37-45, Apr. 2018.

[31] N. Zhao, L. Ying, J. K. Vanos, *et al.*, "Day-of-week and seasonal patterns of PM 2.5 concentrations over the United States: Time-series analyses using the Prophet procedure," *Atmos. Environ.,* vol. 192, pp. 116-127, Nov. 2018.

[32] A. Ibrahim, R. Kashef, and L. Corrigan, "Predicting market movement direction for bitcoin: A comparison of time series modeling methods," *Comput. Electr. Eng.,* vol. 89, pp. 106905, Jan. 2021.

[33] B. Yang, H. Li, and Z. Guo, "Learning a deep similarity network for hyperspectral image classification," *IEEE J-STARS,* vol. 14, pp. 1482-1496, Nov. 2020.

[34] J. M. Joyce, "Kullback-Leibler divergence," *International Encyclopedia of Statistical Science*, M. Lovric, ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 720-722.

[35] R. O. Green, M. L. Eastwood, C. M. Sarture, *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.,* vol. 65, no. 3, pp. 227-248, Sep. 1998.

[36] K. K. Huang, C. X. Ren, H. Liu, *et al.*, "Hyperspectral image classification via discriminative convolutional neural network with an improved triplet loss," *Pattern. Recogn.,* vol. 112, no. 2, pp. 107744, Apr. 2021.

[37] R. Achanta, A. Shaji, K. Smith, *et al.*, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE. Trans. Pattern Anal. Mach. Intell.,* vol. 34, no. 11, pp. 2274-2282, Nov. 2012.

[38] Y. Zhao, and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 1, pp. 296-308, Jan. 2015.

[39] S. Taylor, and B. Letham, "prophet: Automatic forecasting procedure," *R package version 0.2,* vol. 1, 2017.

[40] B. Du, and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 11, pp. 6844-6857, Nov. 2014.

[41] S. Holzwarth, A. Muller, M. Habermeyer, R. Richter, A. Hausold, S. Thiemann, and P. Strobl, "HySens-DAIS 7915/ROSIS imaging spectrometers at DLR." in *Proc. 3rd EARSeL,* Herrsching, Germany, 2003, pp. 225-235.

[42] J. Wang, G. Yao, and G. Yu, "Semi-supervised classification by discriminative regularization," *Appl. Soft Comput.,* vol. 58, pp. 245-255, Sep. 2017.

[43] X. J. Shen, L. Mu, Z. Li, *et al.*, "Large-scale support vector machine classification with redundant data reduction," *Neurocomputing,* vol. 172, pp. 189-197, Jan. 2016.

[44] M. Pal, and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.,* vol. 48, no. 5, pp. 2297-2307, May 2010.

[45] H. Xu, H. Zhang, W. He, *et al.*, "Superpixel-based spatial-spectral dimension reduction for hyperspectral imagery classification," *Neurocomputing,* vol. 360, pp. 138-150, Sep. 2019.

[46] Z. Feng, S. Yang, M. Wang, *et al.*, "Learning dual geometric low-rank structure for semisupervised hyperspectral image classification," *IEEE Trans. Cybern.,* vol. 51, no. 1, pp. 346-358, Jan. 2021.

Ping Ma received the Ph.D. degree at the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK, in 2022. She is currently a Research Fellow with National Subsea Centre, Robert Gordon University, Aberdeen, U.K. Her research interests include multimodal remote sensing, hyperspectral imaging, machine learning.

Jinchang Ren (Senior Member, IEEE) received the Ph.D. degree in electronic imaging from the University of Bradford, Bradford, U.K., in 2009.
He is currently a Professor of Computing Science with Robert Gordon University, Aberdeen, U.K. His research interests include hyperspectral imaging, image

processing, computer vision, big data analytics, and machine learning.

Genyun Sun (Member, IEEE) received the B.Sc. degree from Wuhan University, Wuhan, China, in 2003, and the Ph.D. degree from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2008. He is currently a Professor with the China University of Petroleum (East China), Qingdao, China. He is also with the Laboratory for Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao. His research interests cover remote sensing image processing, including hyperspectral and high-resolution remote sensing and intelligent optimization.

Huimin Zhao was born in Shaanxi, China, in 1966. He received the B.Sc. and M.Sc. degrees in signal processing from NWPU, Xi'an, China, in 1992 and 1997, respectively, and the Ph.D. degree in electrical engineering from Sun Yat-sen University, Guangzhou, China, in 2001. He is currently a Professor and Dean with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China. His research interests include image/video and information security technologies, and applications.

Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from The University of New South Wales at Canberra, Canberra, ACT, Australia, in 1996. Since 1988, she has been with the School of Engineering and Information Technology, University of New South Wales at Canberra, where she is currently an Associate Professor. Her research interests include remote sensing, hyperspectral image (HSI) processing and spatial data analysis.

Yijun Yan received the Ph.D. degrees in electrical and electronic engineering from the University of Strathclyde, Glasgow, U.K., in 2018. He is currently a Research Fellow with National Subsea Centre, Robert Gordon University, Aberdeen, U.K. His research interests include computer vision, hyperspectral imagery, pattern recognition, and machine learning.

Jaime Zabalza (Member, IEEE) received the M.Eng. degree in industrial engineering from Universitat Jaume I (UJI), Castellón, Spain, and the M.Sc. and Ph.D. degrees in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K. He is currently a Chancellor's Fellow (Lecturer) with the Advanced Nuclear Research Centre (ANRC), Department of Electronic and Electrical Engineering, University of Strathclyde. His research interests include hyperspectral data analysis and signal and image processing in a wide range of applications.