Remote Sensing of Environment 217 (2018) 144–157

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A novel co-training approach for urban land cover mapping with unclear Landsat time series imagery

Ting Hu[a], Xin Huang[a,b,*], Jiayi Li[a], Lefei Zhang[c]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, PR China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China
[c] Computer School, Wuhan University, Wuhan 430072, PR China

ARTICLE INFO

Keywords:
Land-cover mapping
Unclear observations
Landsat
Co-training
Collaborative representation classifier
Matrix completion

ABSTRACT

Landsat time-series (LTS) imagery shows potential for dynamic mapping in urban areas. However, unclear observations (clouds, cloud shadows, snow/ice, and SLC-off data) in the dataset inevitably restrict the efficacy of many of the state-of-the-art classifiers. In this work, we present a novel co-training classification approach consisting of two steps to cope with the unclear observations. Firstly, we develop a method called the MCCR classifier that deals with the unclear observations in an error-recoverable way. The clear observations are utilized to recover the unclear ones in the training samples by the use of the matrix completion (MC) algorithm, and the collaborative representation (CR) classifier is exploited to handle unclear observations in the unlabeled data. Secondly, considering that the random forest (RF) classifier is able to cope with contaminated data in an error-tolerant way, a co-training approach (CotrRM) based on the RF and MCCR classifiers is also proposed to further improve the classification efficacy. The CotrRM method is executed by iteratively constructing semi-labeled training sets based on the crisp and soft predictions of the two individual classifiers on the unlabeled data. To validate the effectiveness of the proposed MCCR classifier and CotrRM method, LTS imagery of the city of Wuhan (a metropolitan city of China) from four years (11 images from 2000, 16 images from 2005, 13 images from 2010, and 15 images from 2015) was adopted. The experiments showed that the MCCR classifier performs as well as the RF classifier for the mapping of urban land cover with contaminated LTS imagery. Moreover, the proposed CotrRM method has the ability to further improve the classification performance. The proposed approach can not only work effectively in the classification, but can also recover the unclear observations in the LTS imagery, courtesy of the MC algorithm. The overall accuracies of the land-cover changes between each two adjacent periods are all over 85%. Given the effectiveness and flexibility, the proposed method could also be applied in other unclear data classification.

1. Introduction

Urban areas host more than half of the worlds' population and play a central role in efforts to mitigate and adapt to the effects of climate and other ecosystem changes (United Nations, 2014). As human-dominated habitats, urban areas have developed at an unprecedented rate in recent decades, especially in developing countries such as China (Long et al., 2009). Urban land-cover maps are one of the most fundamental datasets used in many scientific fields, e.g., urban heat island effects (Estoque and Murayama, 2017), air pollution (Lin et al., 2015), urban ecosystem service (Haase et al., 2014) and local climate zone (Middel et al., 2014).

Remotely sensed imagery of various spatial resolutions has been widely used to produce land-cover maps. High resolution (HR) imagery has the ability to provide fine spatial detail for urban mapping (Huang et al., 2017). However, the sparse coverage, limited access, and absence of historical data impede the use of such data (Huang et al., 2014b; Zhu, 2017). Land-cover maps generated from coarse spatial resolution images, such as Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR), have been reported to show a limited mapping accuracy (Fritz et al., 2010), especially in urban areas with complex distributions of many combinations of materials (Chen et al., 2016). Considering the issues of spatial detail, data availability, and areal coverage, medium resolution imagery such as Landsat is more appropriate at present for urban land-cover mapping and change detection. Furthermore, the Landsat satellite archive has a long and continuous record stretching over 40 years and

* Corresponding author at: School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China.
E-mail addresses: huang_whu@163.com, xhuang@whu.edu.cn (X. Huang).

https://doi.org/10.1016/j.rse.2018.08.017
Received 19 November 2017; Received in revised form 26 June 2018; Accepted 14 August 2018

has been open access since 2008, which has resulted in it being widely used in the monitoring of land-surface dynamics (Gómez et al., 2016). There have been numerous applications that have adopted Landsat imagery as the main data source, e.g., forest disturbance and recovery surveillance (Senf et al., 2015; Grogan et al., 2015; Matthew et al., 2016; White et al., 2017), agricultural expansion and intensification (Kontgis et al., 2015; Qin et al., 2015), and impervious surface cover characterization (Lu et al., 2011; Zhang and Weng, 2016).

Single-date images only reflect static land surfaces, but the multi-temporal signature extension provides dynamic observations of land cover. Moreover, it has been verified that using multi-temporal images as input can help to improve the classification accuracy (Bhandari et al., 2012), especially for the land-cover types that have similar spectral characteristics in a single-date image, such as cropland and forest (Schneider, 2012). Landsat can visit the same location every 16 days. However, due to the existence of clouds as well as their shadows, the frequency of clear observations for a specific location is generally much < 16 days. What is worse, the scan-line corrector (SLC) of Landsat 7 malfunctioned in May 2003, which causes wedge-like data gaps in the ETM+ scenes (SLC-off data). These unclear observations (i.e., clouds, cloud shadows, snow/ice, SLC-off data) inevitably influence the availability of multi-temporal images and their mapping accuracy. In order to address this problem, Grinand et al. (2013) used images with the lowest cloud cover in the study period to estimate every-five-year deforestation. The time interval of mapping results obtained by the use of such an approach is restricted to the frequency of the clear satellite observations. Zhu and Woodcock (2014) fit a time-series model by using all the clear observations for each pixel. But this algorithm requires sufficient clear observations to initialize the model. Beckschäfer (2017) produced annual best-available-pixel (BAP) composites, e.g., the least cloudy pixels, from various acquisition dates to cope with the data quality problem, but the experimental results showed that careful selection of images was mandatory in BAP compositing. The previous studies using time-series Landsat imagery are usually based on clear observations, but in Schneider (2012), where dense time series Landsat data, regardless of data quality, were stacked for urban land-cover classification. It was demonstrated that this approach could achieve a better accuracy than simply discarding the unclear datasets. In their experiments, the random forest (RF) classifier was also proved to outperform the maximum likelihood classifier and support vector machine classifier when handling contaminated Landsat data. Some other studies (Breiman, 2001; Zhu et al., 2016b) have also shown that the RF classifier can work effectively, even when the dataset contains some noise. The RF classifier is an ensemble learning method that combines K binary CART trees (Classification And Regression Trees). Each tree is created by selecting a random subset of the features or predictive variables at each node with replacement and the trees grow without pruning (Rodriguez-Galiano et al., 2012; Pelletier et al., 2016). The random and bootstrapped manner enables RF to be a robust and error-tolerant classifier when unclear observations are included (Belgiu and Drăguţ, 2016). However, this method neglects the degree of the reliability between the clear and unclear observations, and may suffer from the negative effect of the unclear ones (Rodriguez-Galiano et al., 2012).

Therefore, it is worth investigating new methods that can focus more on the utilization of both the clear and unclear observations. The improvement of the usability of unclear observations is proved to be conducive to the classification procedure (Schneider, 2012). Training samples from the same class are correlated in both the spectral and temporal dimensions. Therefore, the matrix, constituted by stacking the feature vectors of each training sample from the same class, should be approximately low rank (Chen and Yang, 2014; Cabral et al., 2015). The clear observations of training samples can then be utilized to recover the unclear ones by the matrix completion (MC) algorithm (Candès and Recht, 2009; Cai et al., 2010). To avoid the negative impact of unclear observations in the unlabeled data, the collaborative representation (CR) classifier (Zhang et al., 2012) is an appropriate method to combine with the MC algorithm since its classification hyperplane would not be seriously affected even when only partial observations are available, inferring the CR classifier has the potential to flexibly maintain the discriminative ability when discarding the unclear observations (Li et al., 2014a; Waqas et al., 2013). Therefore, we propose the MCCR classifier that exploits the merits of both MC and CR to cope with poor-quality LTS data classification.

In the MCCR classifier, the clear observations are utilized to recover the unclear ones by the MC algorithm and to then construct the CR classifier. It should be noted that the error-recoverable manner of the MCCR classifier is quite different from the error-tolerant way of the RF classifier. Co-training is a semi-supervised learning paradigm where two basic classifiers are iteratively retrained with the additional semi-labeled samples based on the predictions of either classifier of unlabeled samples (Blum and Mitchell, 1998; Xu et al., 2012). The efficacy of co-training classification method has also been validated in remote sensing imagery, including multispectral image, hyperspectral image, and high spatial resolution image (Persello and Bruzzone, 2014; Zhang et al., 2014). The predictions of RF and MCCR in the unlabeled samples could be complementary to each other, making it possible to exploit the co-training paradigm to provide a better decision than each separate classifier (Zhang and Zhou, 2011; Zhu et al., 2016a). Therefore, we further embed the two classifiers into the co-training paradigm and develop a novel approach (CotrRM), which has the potential to enhance the classification. To the best of our knowledge, this is the first time that an error-recoverable approach with a co-training scheme has been explored to handle contaminated data classification of LTS imagery.

Specifically, in this study, we attempt to investigate the following two research questions:

1) Can the MCCR classifier effectively deal with contaminated data in LTS imagery classification?
2) Is it possible to fuse the MCCR and RF classifiers in order to further raise the performance of their individual use?

2. Study site and datasets

In our study, the city of Wuhan was chosen as the study site (Fig. 1) as it is a typical city that has experienced high-speed development and urbanization in recent decades. Wuhan is the capital of Hubei province, and had a population of over 10 million in 2012 (Wuhan Municipal Statistics Bureau, 2013). The mean annual temperature ranges from 15.8 °C to 17.5 °C (Han et al., 2009), with annual average rainfall of 1050 mm to 2000 mm (Wang et al., 2015). Wuhan lies in the middle reaches of the Yangtze River, and its unique locational characteristics have made it one of the biggest metropolises in central China.

The city of Wuhan is almost entirely covered by the Landsat scene of WRS-2 Path 123 and Row 39, except for about 5% of the study area in the north of Huangpi and Xinzhou districts, as shown in Fig. 1. Since our objective was to investigate a classification method, we selected only this scene to represent the study area. All the available Level 1 Terrain (Corrected) (L1T) Landsat 5,7, and 8 surface reflectance (SR) products acquired in 2000, 2005, 2010, and 2015 with < 60% unclear observations (i.e., pixels with clouds, cloud shadows, snow/ice, and SLC-off data) were downloaded from the U.S. Geological Survey (USGS). Fig. 2 demonstrates the date distribution of the selected images for each year. The Fmask algorithm with its default setting was applied to each image to detect clouds, cloud shadows, and snow/ice. The average producer's accuracy of Fmask in cloud detection is reported to be 92.1% and user's accuracy is as high as 89.4% (Zhu and Woodcock, 2012). This Fmask layer and SR product can provide us whether the observation is clear or not.

Similar to Ying et al. (2017), in this study, we selected time-series Landsat images within a specific given year to produce an annual land-cover map, which can characterize the temporal characteristics for the

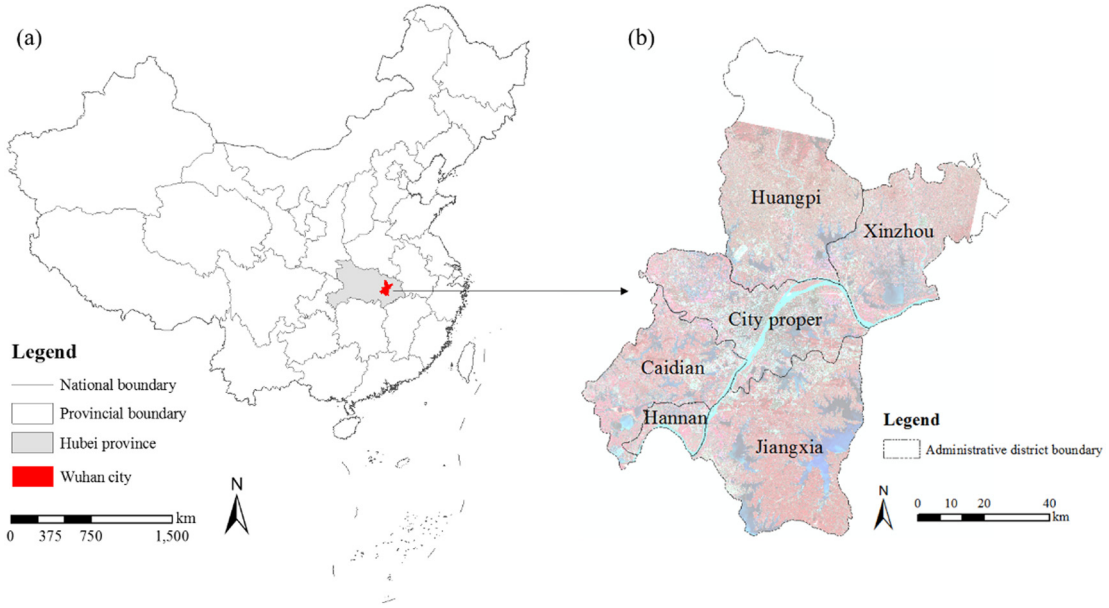


Fig. 1. Study area map. (a) The geographic location of the city of Wuhan. (b) The background image is a false-color composite (R: near-infrared, G: red, B: green) Landsat 8 image acquired on November 26, 2015. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

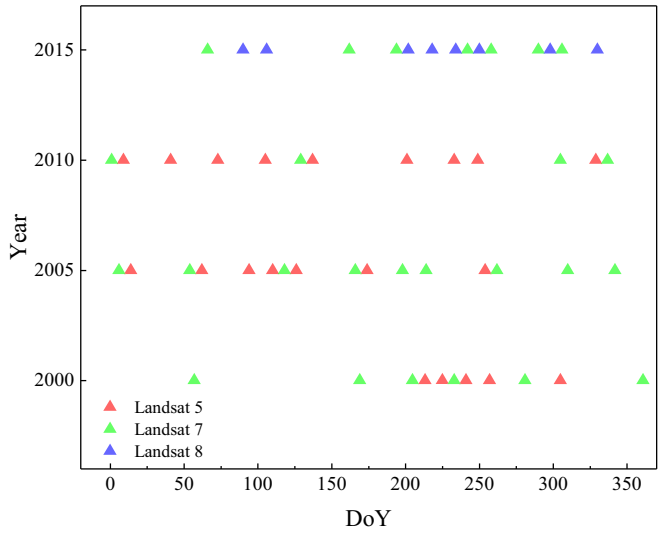


Fig. 2. The selected images of each year. DoY: day of year.

different land-cover types.

3. Methodology

In this section, we first describe how the MC algorithm and CR classifier are effectively combined to develop the MCCR classifier, which can classify the contaminated observations of LTS imagery. The co-training approach on the basis of RF and MCCR classifiers (CotrRM) is then explained in Section 3.2. Finally, the accuracy assessment is detailed in Section 3.3.

3.1. The proposed MCCR classifier

A number of the existing classifiers may fail to effectively process contaminated LTS imagery, e.g., imagery subject to clouds, cloud shadows, snow/ice, and SLC-off data. There are two key issues to be resolved: 1) unclear observations in the training data impede the successful construction of a classifier; and 2) unclear observations in each unlabeled sample have a negative impact on its accurate identification. Training data from the same category normally show similar signatures, which enables the MC algorithm to recover the unclear ones in the training data. With respect to the unclear observations in the unlabeled data, the CR classifier is employed since its discriminative ability has the potential to be guaranteed, even with incomplete observations (Li et al., 2014a; Waqas et al., 2013). Fig. 3 shows the workflow of the MCCR classifier for LTS imagery. Specific descriptions are given in the following subsections.

3.1.1. The recovery of training samples

If we suppose that K LTS images are selected within a specific year, then the spectral bands of all the images are stacked together to constitute the input feature vectors for classification. The input feature vector of a given sample $\boldsymbol{x}$ can be depicted as shown in Fig. 4, where B denotes the dimension of the feature vector.

There are M distinct classes defined in the classification scheme and N training samples, of which $N_i (i = 1, .., M)$ training samples belong to the ith class. As shown in Fig. 3(a), the matrix $\boldsymbol{X}_i^o \in \mathbb{R}^{B \times N_i}$ is composed of training samples from the ith class. Each element in the matrix can be identified as clear or not (displayed in white in Fig. 3(a)) by the Fmask product. The matrix $\boldsymbol{X}_i^o$ actually lies on a low-dimensional space (Tao et al., 2013), for the following reasons. Firstly, features composed by multi-temporal stacking of the spectral bands are correlated with each other, to some extent (Li et al., 2014b; Li et al., 2015; Schneider, 2012). Secondly, samples from the same class normally show a degree of similarity, which is useful to distinguish them from other categories (Huang et al., 2014a; Madonsela et al., 2017). Based on the low-rank constraint, it is possible to recover the contaminated observations through the clear subsets, and the problem can then be modeled as:

$$\begin{aligned} &\min\ rank(\boldsymbol{X}_i) \\ &\text{s. t. } (\boldsymbol{X}_i)_{pq} = (\boldsymbol{X}_i^o)_{pq},\ (p,q) \in \Omega \end{aligned} \tag{1}$$

where $\boldsymbol{\Omega}$ denotes the set of clear elements and $\boldsymbol{X}_i$ is the recovered matrix. In this research, fixed-point continuation with approximate (FPCA) singular value decomposition (SVD) (Ma et al., 2011) is adopted to solve the MC problem due to its efficiency. The training data from each class can be processed with the MC algorithm separately, and in this

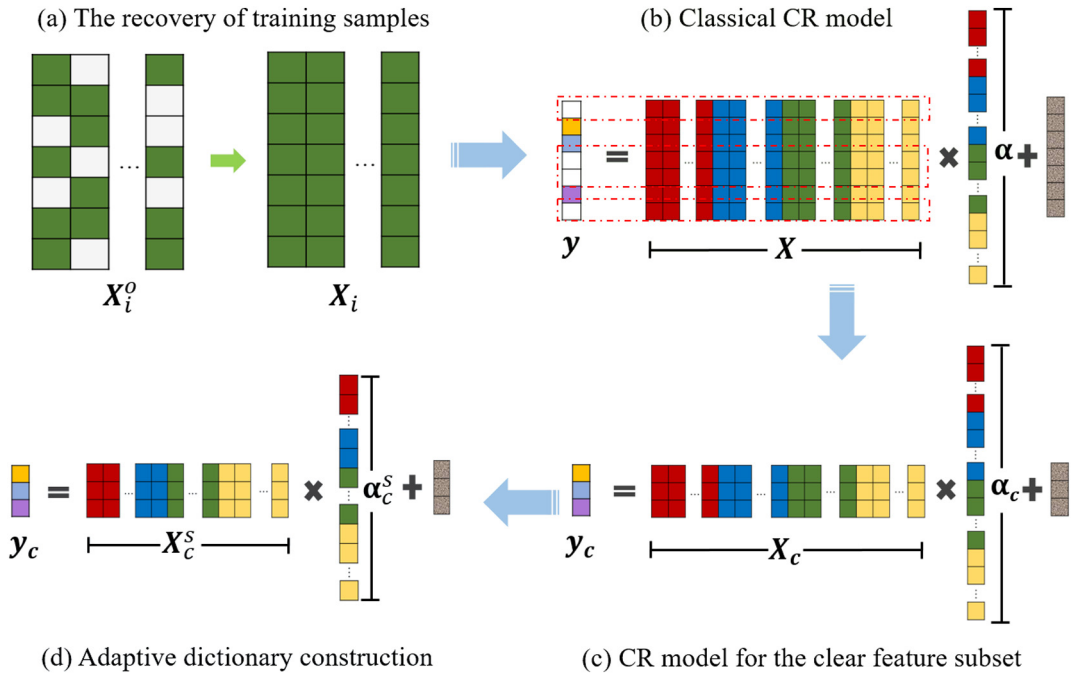


Fig. 3. Workflow showing the MCCR classifier for LTS imagery. In step (a), unclear elements (shown in white) of the sub-dictionary $\boldsymbol{X}_i$ composed of training samples from the ith ($i = 1,..,M$) class are recovered by the MC algorithm. From steps (b) to (c), the unclear sub-features $\boldsymbol{y}$ (shown in white) are abandoned, as are the corresponding feature subsets in the dictionary $\boldsymbol{X}$, represented by the red dashed lines. From steps (c) to (d), the training samples which are very different from the unlabeled sample are removed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

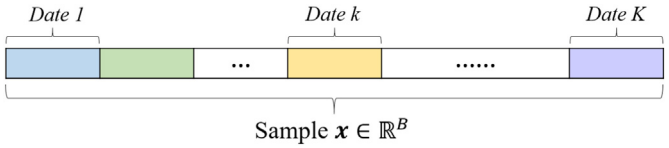


Fig. 4. Depiction of the input feature vector. The feature of each sample is formed by multi-temporal stacking of the spectral bands.

way, the contaminated observations in all the training samples are recovered.

3.1.2. The CR based classifier for processing unclear observations in the unlabeled samples

If all the K LTS images are free of contamination, an unlabeled sample $\boldsymbol{y} \in \mathbb{R}^B$ with class i can be linearly represented as follows in the classical CR model (Zhang et al., 2012), displayed in Fig. 3(b):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \boldsymbol{X}_1\boldsymbol{\alpha}_1 + ...+\boldsymbol{X}_i\boldsymbol{\alpha}_i + ...+\boldsymbol{X}_M\boldsymbol{\alpha}_M + \boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{X} \in \mathbb{R}^{B\times N}$ represents the dictionary that is constituted by all the training samples; $\boldsymbol{X}_i \in \mathbb{R}^{B\times N_i}$ denotes the ith sub-dictionary, which is formed by training samples from the ith($i = 1,..,M$) class; and the vector $\boldsymbol{\alpha} \in \mathbb{R}^N$ is called the CR coefficient.

The Fmask product can provide the location of the unclear feature subset in $\boldsymbol{y}$, which is shown as the elements highlighted by the red dashed lines in Fig. 3(b). With the unclear observations abandoned, the CR model can be modified as follows (also shown in Fig. 3(c)):

$$\boldsymbol{y}_c = \boldsymbol{X}_c\boldsymbol{\alpha}_c + \boldsymbol{\varepsilon}_c \tag{3}$$

where $\boldsymbol{y}_c \in \mathbb{R}^{B_c}$ is the clear feature subset that is composed of all the clear observations and $\boldsymbol{X}_c \in \mathbb{R}^{B_c}\times N$ denotes the flatter dictionary that is extracted by the corresponding feature subset of each training sample. From the derivation process, it can be observed that the CR model still holds for the subset of unlabeled samples, and the samples can be represented by the coefficient $\boldsymbol{\alpha}_c \in \mathbb{R}^N$, given the modified dictionary.

As reported in Zhang et al. (2012), training samples from different categories present both similarities and differences. The training samples that are very different from $\boldsymbol{y}$ can naturally result in the decline of the representation accuracy of $\boldsymbol{y}$. Based on this idea, it has been suggested that adaptive dictionary selection could be utilized to further represent the unlabeled samples and facilitate their identification (Li et al., 2014c). The L training samples that are most spectrally similar to all the training samples are picked to construct the adaptive dictionary for the specific unlabeled samples. This process is demonstrated in Fig. 3(d), where a narrower dictionary is exhibited than that shown in Fig. 3(c). We let $\boldsymbol{x}_{cj}(j = 1,...,N)$ denote the jth row in the dictionary $\boldsymbol{X}_c$, and the spectral similarity between $\boldsymbol{y}_c$ and each $\boldsymbol{x}_{cj}$ is computed. The largest L values are chosen, and then the associated training samples are selected accordingly to construct the adaptive dictionary, which is denoted as $\boldsymbol{X}_c^s \in \mathbb{R}^{B_c}\times L$. Therefore, the final CR-based model (Fig. 3(d)) can be summarized as:

$$\boldsymbol{y}_c = \boldsymbol{X}_c^s\boldsymbol{\alpha}_c^s + \boldsymbol{\varepsilon}_c \tag{4}$$

The representation coefficient can be solved by the regularized least squares (RLS) method:

$$\boldsymbol{\alpha}_c^s = \boldsymbol{Q}\boldsymbol{y}_c, \text{ with } \boldsymbol{Q} = ((\boldsymbol{X}_c^s)^T(\boldsymbol{X}_c^s) + \lambda\boldsymbol{I})^{-1}(\boldsymbol{X}_c^s)^T \tag{5}$$

where parameter λ is a trade-off between the data fidelity term and the coefficient prior, and $\boldsymbol{I}$ is an identity matrix. For a given sample $\boldsymbol{y}$, the output identity can be denoted as Zhang et al. (2012):

$$lab(\boldsymbol{y}) = \arg\min_{i=1,...,M} \|\boldsymbol{y}_c - (\boldsymbol{X}_c^s)_i(\boldsymbol{\alpha}_c^s)_i\|_2/\|(\boldsymbol{\alpha}_c^s)_i\|_2 \tag{6}$$

where $(\boldsymbol{\alpha}_c^s)_i$ is the representation vector associated with class i, and $(\boldsymbol{X}_c^s)_i$ is the sub-dictionary of class i. The CR-based model, as well as the CR-based classifier, to adapt to the clear feature subset in the unlabeled sample has been completed.

In summary, the training samples are first recovered class by class to obtain a fully reliable dictionary using the MC algorithm (Section 3.1.1). The CR model is then established according to the clear feature subset of each unlabeled sample, and the corresponding label can be identified by Eqs. (5) and (6) (Section 3.1.2). In the CR model, each pixel is collaboratively represented by training samples from all the

classes. Training samples from other classes can highlight the discriminative ability of the current class when maintaining the linear presentation process, as shown in Eq. (2). They are called as collaborative assistants, which enables the CR classifier competitive among the state-of-the-art classifiers. It is this mechanism that guarantees the potential discriminative ability of the CR classifier, even with incomplete features (Li et al., 2014a; Zhang et al., 2012).

3.2. The proposed CotrRM method

The RF classifier copes with the data quality problem in an error-tolerant way. However, the MCCR classifier treats the problem in an error-recoverable manner, where the unclear observations are recovered by the MC algorithm, and then only the clear ones are utilized to construct the CR classifier. The two learners provide different approaches to the predictions of unlabeled data (Tan et al., 2014). Inspired by the co-training paradigm, the newly labeled samples can be iteratively selected to retrain each classifier, and a strong classifier is then built on the fusion of the two retrained basic classifiers (Zhang et al., 2011). Therefore, we propose a co-training approach (CotrRM) based on the fusion of RF and MCCR classifiers.

3.2.1. The proposed CotrRM framework

The framework of the CotrRM method is displayed in Fig. 5. Both the RF and MCCR classifiers can predict a label for a given sample and provide the posterior probability for each class. The class certainty for the assigned label is computed as Huang et al. (2014b):

$$cer(\boldsymbol{y}) = prob_{max}(\boldsymbol{y}) - prob_{sub-max}(\boldsymbol{y}) \tag{7}$$

where $prob_{max}(\boldsymbol{y})$ and $prob_{sub-max}(\boldsymbol{y})$ are associated with the maximum and the second largest posterior probability values.

Both the crisp (class label) and soft (class certainty) outputs of the unlabeled data for each classifier are utilized to select the appropriate samples to enhance the classifiers. A higher class certainty generally means that the assigned label is more to be correct, but also shows the sample has limited benefit to further strengthen the classifier. Besides, the consistency of an unlabeled sample in the RF and MCCR classifiers indicates that the predicted label is probably reliable. It is reported that the reliability and variety of training samples have significant influence on the classification performance (Gray and Song, 2013), therefore, an *accurate-diverse* trade-off (ADT) selection (further details are given in Section 3.2.2) based on the predictions of the unlabeled samples is designed to generate a semi-labeled sample set with less label noise and more feature diversity (Zhang and Zhou, 2011). Each classifier is retrained iteratively with the selected semi-labeled samples until a predefined maximum iteration number *maxIter* is reached. Please kindly note that the semi-labeled samples selected for the MCCR classifier should be recovered by the MC algorithm to ensure the performance of MCCR classifier.

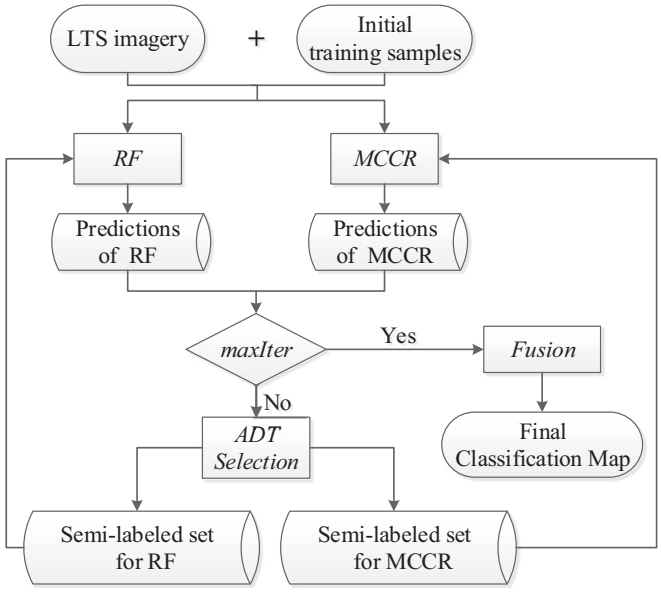


Fig. 5. Flowchart of the proposed CotrRM method. ADT selection: *accurate-diverse* trade-off selection.

Please note that the two base classifiers with their respective semi-labeled training sets may still have inconsistent predictions (i.e., conflicts) in unlabeled samples after the last iteration. In such situation, the label is determined by fusion of the two individual predictions. The posterior probabilities of the RF and MCCR classifiers are summed up and the final label is assigned to the class with the maximum posterior probability. This aggregation rule is adopted since it does not involve learning and counter-acts over-fitting of an individual classifier (Ruta and Gabrys, 2000).

3.2.2. Construction of semi-labeled sample sets

As mentioned before, the key to the success of the CotrRM method is the selection of samples with sufficient reliability and variety, which can enhance the current training set. A sample that is qualified to help improve the classification should satisfy the following two requirements (Zhu et al., 2016a): 1) *Accurate*: a mislabeled sample would harm the classification process, therefore the accuracy of the selected samples should be guaranteed; and 2) *Diverse*: the feature diversity of selected samples should be enlarged to improve the recognition ability of each classifier. Inspired by these considerations, we propose the ADT selection strategy based on the degree of class certainties in the RF and MCCR classifiers.

Firstly, the requirement for the sample to be accurate is considered. A sample $\boldsymbol{y}$ is more likely to be correctly labeled when the following constraint is met:

$$lab(\boldsymbol{y})_{\mathrm{RF}} = lab(\boldsymbol{y})_{\mathrm{MCCR}} \& cer(\boldsymbol{y})_{\mathrm{RF}} > T_1 \& cer(\boldsymbol{y})_{\mathrm{MCCR}} > T_1 \tag{8}$$

where $lab(\boldsymbol{y})_{\mathrm{RF}}$ and $lab(\boldsymbol{y})_{\mathrm{MCCR}}$ stand for the predicted labels of the RF and MCCR classifiers, respectively; $cer(\boldsymbol{y})_{\mathrm{RF}}$ and $cer(\boldsymbol{y})_{\mathrm{MCCR}}$ denote the class certainty of the predicted labels of the RF and MCCR classifiers, respectively; and T_1 is the class certainty threshold. For example, Samples A and E in Fig. 6 have different predictions, therefore they are eliminated. The class certainties of Samples D and G in the two base classifiers are lower than the predefined threshold, and hence they are also deleted. Unlabeled samples that meet the requirements are added to the candidate training set, which is denoted as $\mathscr{C}$ (e.g., Samples B, C, F, and H). The samples in $\mathscr{C}$ are identified with a reliable label, and unclear observations within them can be recovered with the MC algorithm (Section 3.1.1).

Subsequently, the diversity criterion of the samples is taken into account. For a given sample, lower class certainty normally means that this sample has potential benefit to strengthen the classifier. If this sample is assigned to a correct label, the incorporation of this sample can enlarge the intra-class diversity of training set, and hence promote the classification performance (Tuia et al., 2011; Gray and Song, 2013). Thus, samples with a lower class certainty need to be further selected from $\mathscr{C}$ with the following steps. Firstly, we define groups of spatially connected pixels in $\mathscr{C}$ from the same class as a segment. One sample with the minimum class certainty is picked out for each classifier within each segment, constituting a trimmed sample set $\mathscr{C}_s$. Next, if the class certainty of a sample for RF is lower than that for MCCR, then it is added to RF, and vice versa. With all samples in $\mathscr{C}_s$ processed, two different sample sets for the RF and MCCR classifiers are respectively generated, i.e., $\mathscr{C}_s\_RF$ and $\mathscr{C}_s\_MCCR$. For example, in Fig. 6, according to the class certainties, Samples C and F are added to the RF classifier, and Samples B and H are added to the MCCR classifier.

Normally, a large number of samples can meet the ADT selection. To further reduce the number of semi-labeled samples and select the

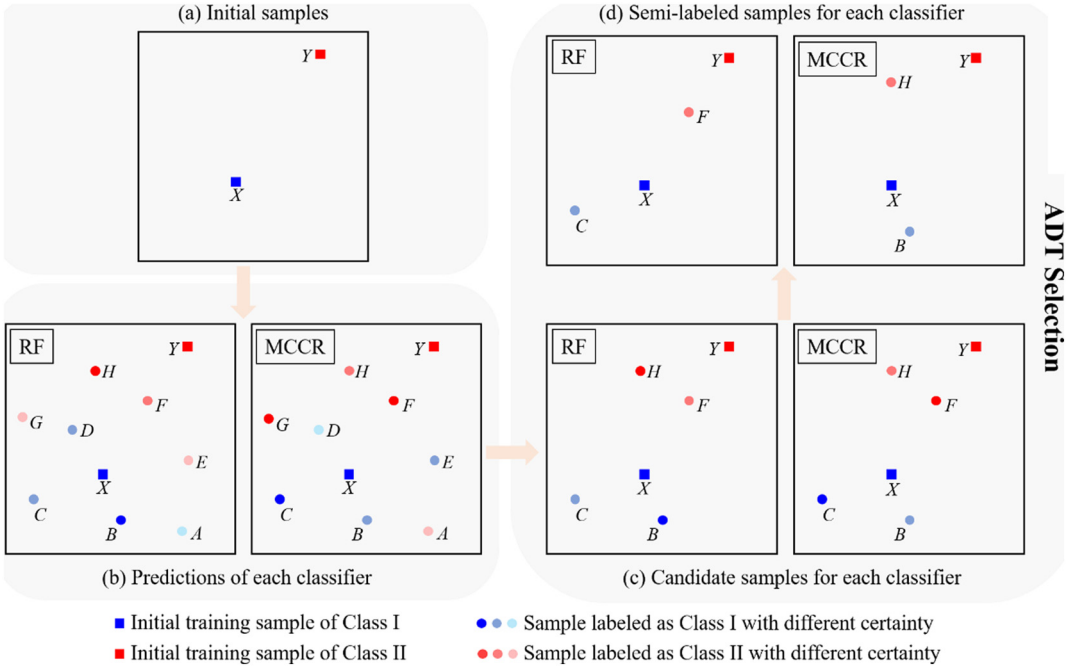


Fig. 6. The selection procedure of semi-labeled samples. (a) Initial sample for each class (in squares), e.g., *X* for Class I and *Y* for Class II; (b): predictions of unlabeled samples (i.e., samples A to H) derived by the RF and MCCR classifiers, respectively, where the shade of the color represents the classification uncertainty (deeper color signifies larger certainty); (c): the candidate samples for RF and MCCR classifiers, respectively. These samples that satisfy the *accurate* criterion of ADT selection are retained; (d): the selected semi-labeled samples for the RF (e.g., Sample *C* and *F*) and MCCR classifiers (e.g., Sample *B* and *H*), respectively. They are assigned as the *diverse* criterion of ADT selection. The ADT semi-supervised selection consists of steps (c) and (d).

representative ones, center patterns for each class are obtained by the k-means method (Lance and Williams, 1967). The number of centers is fixed to be proportional to that of the current training samples with the ratio of T_2 ($0 < T_2 < 1$) (Bruzzone et al., 2006). In this way, the semi-labeled sample sets for the RF and MCCR classifiers, i.e., $\mathscr{S}\_RF$ and $\mathscr{S}\_MCCR$, are generated in each iteration. After the iteration process, the semi-labeled set together with the initial training set are treated as the individual input of each classifier.

In summary, we first propose a method called the MCCR classifier, where the MC algorithm is effectively combined with the CR classifier to cope with the unclear observations in LTS imagery. In addition, a co-training classification approach, the CotrRM method, is also proposed to take further advantage of the RF and MCCR classifiers by iteratively absorb the reliable and diverse samples from the predictions on unlabeled data.

3.3. Accuracy assessment

The reference land cover samples for each year were independently collected with the following steps. First, the study site, Wuhan city, was divided into 15 columns and 20 rows resulting in 300 cells (Yuan et al., 2005), and from each cell, 15 polygons with 5×5 pixels were randomly chosen to represent the spatial variation of land cover. Then a minimum distance of 30 pixels (900 m) between polygons is used as a constraint condition, to suppress the spatial autocorrelation (Müller et al., 2015). The selected polygons were visually interpreted using Landsat data and high resolution (HR) imagery. As a reference to the manual interpretation, in addition to the images from Google™, we have sufficient archive of HR data in Wuhan since 2000, including aerial images, QuickBird, GeoEye-1, Worldview-2, and Ziyuan-3. Please note that only the polygons that can be labeled confidently are collected.

A random stratification procedure was then applied to the reference data to produce 10 disjoint datasets (in polygons) for training (50%) and test (50%) for each year. The training samples were then formed by randomly selecting one pixel from each training polygon, and similarly, the test samples were collected by choosing one random pixel from each test polygon. In order to reduce the possible bias induced by random sampling, 10 unique training and test sets were generated for each year. The proposed CotrRM method was then carried out and assessed independently with the 10 training and test sets for each year, and the classification map of Wuhan city for each year was generated by majority voting of the 10 results in the corresponding year.

4. Results

In this section, the performances of the RF and MCCR classifiers and the CotrRM method for land-cover mapping with LTS imagery are evaluated. A commonly used classification scheme comprising five land-use/land-cover types was adopted to implement the classification, namely, built-up land, cropland, forest, water, and barren land (Zhou et al., 2014).

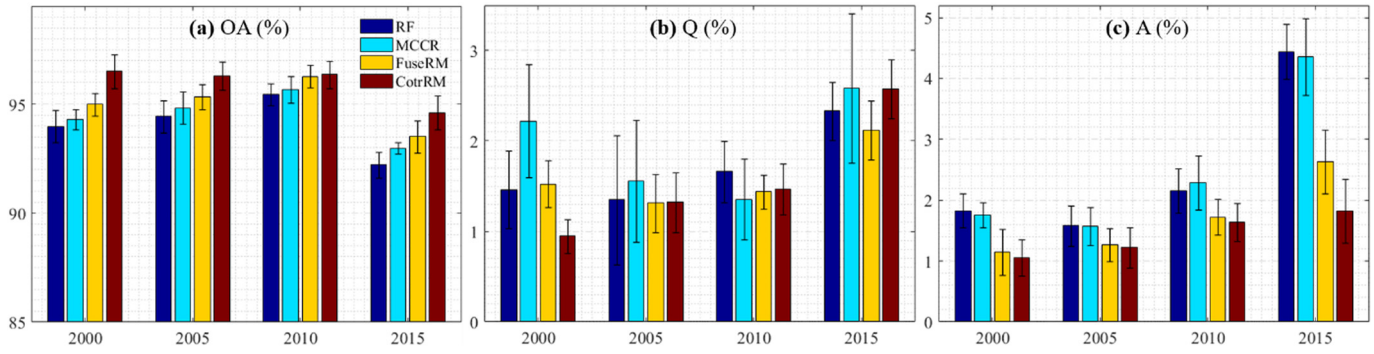
4.1. Experimental setup

Table 1 lists the numbers of training and test samples (in pixels) collected from the reference data for each year.

In addition to the three classifiers (RF, MCCR, and CotrRM), another method, FuseRM, based on the fusion of RF and MCCR, is carried out

Table 1
Numbers of training and test pixels for each year.

Category	2000		2005		2010		2015	
	Training	Test	Training	Test	Training	Test	Training	Test
Built-up	98	97	102	102	111	111	136	135
Cropland	141	141	142	141	141	140	160	160
Forest	91	91	97	97	100	99	101	100
Water	137	137	139	139	145	145	187	186
Barren	38	37	34	34	48	47	39	39



(d)

Indicator	Year	RF vs. MCCR	RF vs. FuseRM	MCCR vs. FuseRM	RF vs. CotrRM	MCCR vs. CotrRM	FuseRM vs. CotrRM
OA	2000	—	*	*	*	*	*
	2005	—	*	*	*	*	*
	2010	—	*	*	*	*	—
	2015	*	*	—	*	*	*
Q	2000	*	—	*	*	*	*
	2005	*	—	*	—	*	—
	2010	*	*	—	*	—	—
	2015	*	*	*	*	—	*
A	2000	—	*	*	*	*	—
	2005	—	*	*	*	*	—
	2010	*	*	*	*	*	—
	2015	—	*	*	*	*	*

—: non-significant; *: significant.

Fig. 7. The classification results of the different classifiers.

for the purpose of comparison. Specifically, the posterior probabilities of RF and MCCR classifiers are summed up for each class, and the final label is then assigned to the class with the maximum posterior probability (Ruta and Gabrys, 2000). It should be noted that the difference between FuseRM and CotrRM method is the co-training process with the injection of semi-labeled samples. Therefore, the impact of semi-labeled samples on the classification performance can be assessed by the comparison between these two methods.

The tree number of the RF classifier was fixed at 500 (Breiman, 2001), parameter λ in Eq. (5) of the MCCR classifier was fixed to $1e-2$ (Zhang et al., 2012), and the size of the adaptive dictionary was set to 50% (Li et al., 2014c) of the original dictionary size in all the experiments. In the CotrRM experiment, the three parameters, i.e., the maximum iteration number *maxIter*, the class certainty threshold T_1, and the clustering threshold T_2, were respectively set to 4, 0.1, and 0.1. The sensitivity of the classifiers to these parameters is discussed in Section 5.2.

4.2. Experimental results

Fig. 7 displays the classification results of the RF, MCCR, FuseRM, and CotrRM classifiers for the 10 experiments for each year with overall accuracy (OA), quantity disagreement (Q), allocation disagreement (A) (Pontius Jr and Millones, 2011), and the significant result of Student's *t*-test (Box, 1987). A lower value indicates a better result when using the indicators of Q and A. In terms of OA (Fig. 7(a)), MCCR achieves slightly higher average accuracy than RF in the four years, and the FuseRM method shows superiority over the two base classifiers. The OAs of CotrRM are better than those of FuseRM in all the experiments, indicating that the iteratively selected semi-labeled samples are conductive to the classification process. With regard to the Q and A scores (Fig. 7(b)–(c)), the comparison results between MCCR and RF differ in the experiments of the four years. The Q values of MCCR are worse than those of RF in the 2000, 2005, and 2015 experiment, but MCCR outperforms RF in 2010. The A values of MCCR are similar to those of RF in all the years. Both Q and A values of FuseRM are better than those of two base classifiers in most of the experiments. The Q value of CotrRM in 2000 is better than that of FuseRM, but opposite result can be found in 2015. The Q values of CotrRM and FuseRM are similar in 2005 and 2010. In terms of the A values, the performance of the proposed CotrRM method is better than that of FuseRM in all the four years.

To further evaluate the MCCR and CotrRM classifiers, the significance of the difference between every two classifiers are also listed in Fig. 7(d), with the *p*-value set to 0.05. With 95% confidence intervals, we can draw the following observations: 1) In terms of OA, the proposed MCCR can achieve a comparable accuracy to the RF in the 2000, 2005, and 2010 experiments, and outperform RF in 2015. The proposed CotrRM method significantly outperforms the two base classifiers in all the experiments. 2) With respect to the indicator of Q, the proposed CotrRM achieves significantly better results than RF in the 2000, 2010, and 2015 experiments, and a similar result to RF in 2005. 3) The A values obtained by CotrRM are significantly better than those of the two base classifiers.

According to the above results, it can be stated that the MCCR classifier achieves a similar or even better result than the RF classifier in LTS imagery classification. In addition, it is interesting to see that their co-training classifier (i.e., CotrRM) can further improve the classification performance compared with their individual usage. To summarize, the proposed CotrRM method has the ability to outperform the RF classifier.

5. Discussions

5.1. Performance of the matrix completion algorithm

As mentioned in Section 3.1.1, unclear observations in all the training samples should be recovered by the MC algorithm beforehand. Therefore, the performance of the MCCR classifier is affected by the quality of the recovered observations in the training samples. We first assess the quality of MC algorithm through both simulated and real data experiments in Section 5.1.1, and then the benefit of recovering unclear observations to the classification is examined in Section 5.1.2.

Table 2
Results of the simulated testing.

Class	Built-up	Cropland	Forest	Water	Barren
n	378	619	1184	1106	329
cp (%)	36.41 ± 8.09	41.24 ± 11.36	37.06 ± 12.28	34.18 ± 16.22	45.87 ± 12.31
RecErr (%)	6.47 ± 6.19	11.75 ± 8.04	9.71 ± 8.46	8.68 ± 9.83	10.18 ± 6.76

n: the number of clear samples per category.
cp: contamination percentage.
In the expression 'a ± b', 'a' means the average value and 'b' denotes the standard deviation.

5.1.1. The recovery quality of unclear observations

A simulated experiment was designed by selecting the samples whose observations were all clear in the LTS images of 2015. The number of pixels per category is displayed in Table 2. In order to generate a synthetic dataset that was close to the distribution of clouds, cloud shadows, ice/snow, and SLC-off data, two guidelines were considered for simulating the contamination process: 1) the unclear observations usually appear in a spatial cluster; and 2) the locations of unclear observations in different images are independent of each other. The contamination percentage of each sample was randomly assigned, but limited to between 0 and 80%. The simulation process was repeated 10 times. The recovery errors of the MC algorithm are summarized in Table 2. The recovery error is computed as:

$$\mathrm{RecErr} = \frac{\sum_{p=1}^{num}\sum_{q=1}^{B}\left|(x_{pq} - \overline{x}_{pq})/\overline{x}_{pq}\right|}{num \times B} \times 100\% \tag{9}$$

where *num* is the number of pixels in the given category, B denotes the number of feature dimensions, x_{pq} represents the recovered element in the qth dimension of the pth pixel, and $\overline{x}_{pq}$ is the corresponding real value of the element. A smaller value of RecErr indicates a better recovery result. For all categories, the average values of RecErr are about 10%, which shows that the result of the simulation testing is satisfactory and the effectiveness of the MC algorithm can be confirmed.

The recovery quality of the MC algorithm was also checked by visual inspection in real data testing. Considering the spectral signature of built-up class is relatively more stable than those of other classes, we randomly selected a sample from the built-up class to assess the recovery quality. The original and recovered feature curves, representing the temporal stack of the spectral bands of this sample, are compared in Fig. 8(a), where the blue dashed line denotes the original feature curve and the red line is the recovered one. The overlapping parts in the two curves correspond to the clear observations, and the differences indicate that unclear observations are recovered. It can be noted that the recovered curve looks more regular than the original one.

Additionally, three examples with unclear observations are displayed. The first one (Fig. 8(b)) was induced by SLC-off data on the date of 2015306. Fig. 8(c) displays the recovered image, and Fig. 8(d) is the image acquired on an adjacent date (2015290) as a reference. It can be seen that the recovered observation is very similar to the reference. The second example is contaminated by clouds as well as their shadows (Fig. 8(e)). As in the first example, the recovered image in Fig. 8(f) is close to the reference in Fig. 8(g). When using the observations from neighboring dates as benchmarks, the recovery errors of the two examples are 4.87% and 13.03%, respectively, showing that the recovery performance is reliable. In the last example, a sample whose land cover changed over a year is demonstrated (Fig. 8(h)). Referring to Google™ imagery, we found that this change was due to flooding caused by heavy rain (Fig. 8(j)). In this case, the recovery quality is not as satisfactory as the two previous ones (Fig. 8(i)), since the low-rank condition cannot be fully met.

The results of both the simulated and real data experiments verify the recovery quality of the MC algorithm. In the CotrRM process, unclear observations in the semi-labeled samples can be recovered once they are labeled, and the unselected samples are labeled by the aggregation of the class probabilities obtained by the RF and MCCR classifiers. With the labels of all the samples assigned, all the LTS imagery can be recovered by utilizing the MC algorithm.

5.1.2. The benefit of recovering unclear observations

In this section, additional experiments were performed to assess the benefit of MC algorithm to the classification, and the results are presented in Fig. 9, where for instance the combination 'Clear + CR' indicates the CR classifier is used to deal with only the clear observations, i.e., simply discarding the unclear observations, and 'All + CR' means all observations, including the unclear ones, are used as the input to the CR classifier. It can be observed that 'Clear + CR' outperforms 'All + CR' because unclear observations can impede the collaborative linear representation of the CR classifier. However, when the unclear observations are recovered by the MC algorithm, the collaborative linear representation of CR (i.e., MCCR) is capable of substantially improving the classification accuracy. In summary, the comparison between MCCR and 'Clear + CR' explicitly indicates that the inclusion of unclear observations is conducive to raising the mapping accuracy. In addition, 'All + RF' achieves better OA than 'Clear + RF', inferring that when the RF classifier is used, there is something in unclear data that can positively add to the classifier's ability.

The stacked multi-temporal spectral bands can correlate with each other for each class, which enables MC to effectively recover the unclear observations. The experimental results of MC also validated the improvement of image quality in LTS data. With more useful observations added to the classifier, the model performance can be improved.

5.2. Parameter analysis

5.2.1. Parameters of ADT selection

This section aims at assessing the influence of the two important parameters related to construction of the semi-labeled samples of ADT selection in each iteration. The class certainty threshold T_1 affects the reliability of the candidate training samples, and the clustering threshold T_2 controls the representativeness of the final semi-labeled samples. Fig. 10 shows how the OA values of the three classifiers vary as a function of T_1 with different T_2, based on LTS imagery of 2015. With a fixed T_2 value, nearly all the OA values of the three classifiers show a tendency to decline when T_1 increases from 0.1. When T_1 is relatively large, the *accuracy* criterion is sufficiently strict to guarantee the reliability of selected samples, however, the class certainties would be too high to capture samples with more diversity, leading to the reduced accuracy (Gray and Song, 2013). Among these experiments, the highest accuracy is achieved when T_1 is in the range of [0, 0.2], regardless of the classifier or the value of T_2. In addition, it can be seen that when T_1 is fixed, the variation range of Kappa with different T_2 values is relatively small, inferring that the influence of T_1 on the classification performance is generally greater than that of T_2. The range of [0.1, 0.3] is recommended for the selection of T_2.

As noted in Section 3.2.1, with a relatively low T_1, the diversity of the training samples has the potential to be enhanced. However, on the other hand, it is possible to introduce some mislabeled samples at the

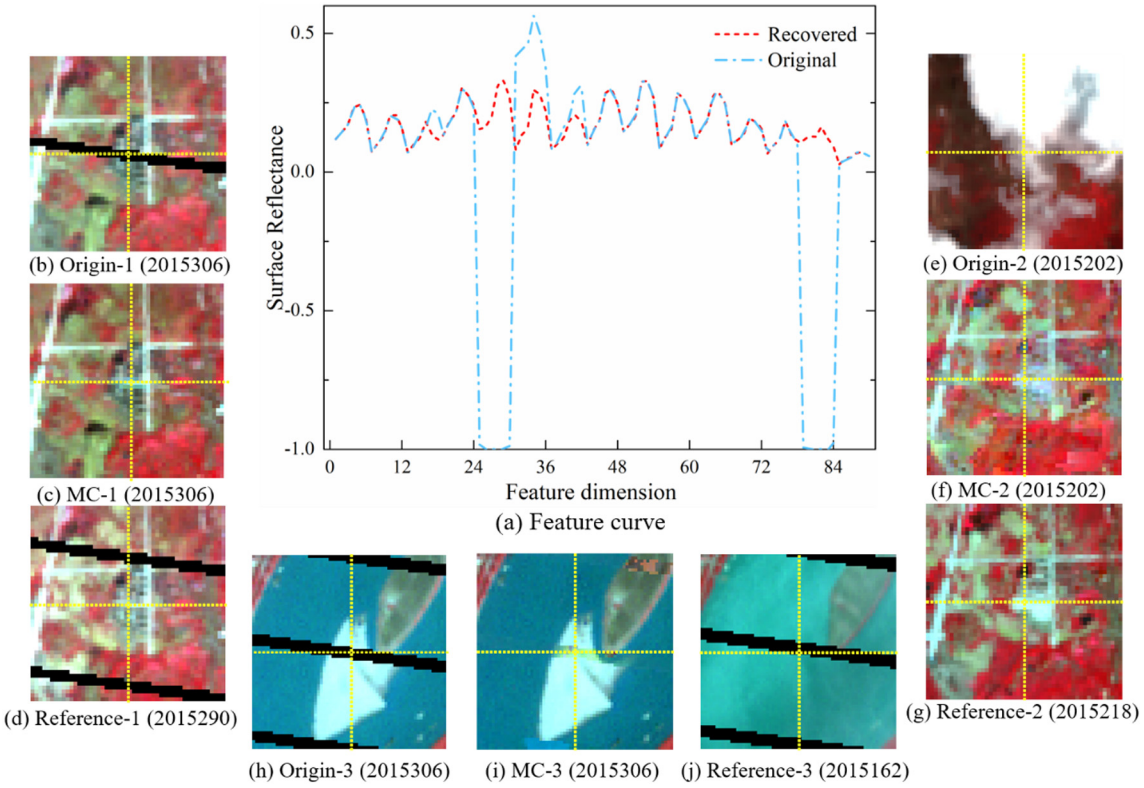


Fig. 8. The recovery results of the real data experiment: (a) The recovered feature curve compared to the original one for the selected sample. The images in (b)–(j) are displayed as near-infrared, red-visible, and green-visible false-color composite images, respectively. The intersections of the yellow dashed line represent the contaminated samples. The three examples are shown in (b)–(j), and the acquisition date of each image is listed in brackets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

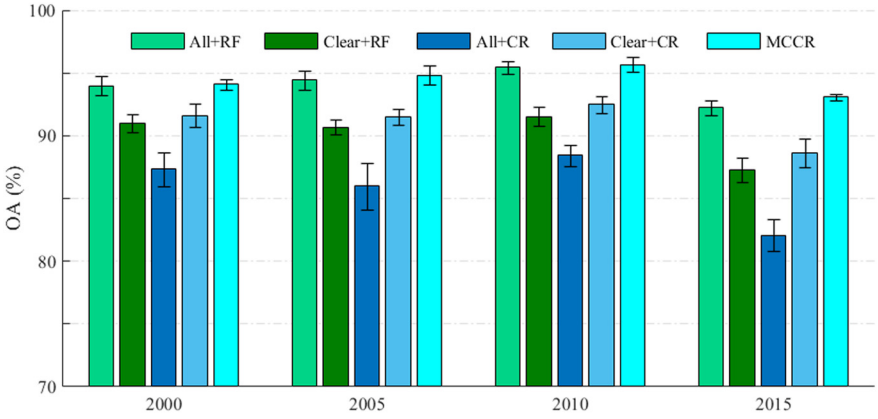


Fig. 9. The overall accuracy of different combinations of features and classifiers for each year.

same time. Therefore, we set up an experiment to investigate the possible reasons why a high accuracy can be obtained when T_1 is relatively low (e.g., $T_1 = 0.1$). The sets of $\mathscr{C}_s$_*RF* and $\mathscr{C}_s$_*MCCR* were made up of samples that satisfied the ADT selection. To measure the accuracy of the selected samples, we checked the intersection set of $\mathscr{C}_s$_*RF* and the test sample set, and then examined the intersection set of $\mathscr{C}_s$_*MCCR* and the test sample set. The accuracy in the two intersection sets was about 97.9% and 97.1%, respectively. That is, mislabeled samples do exist after ADT selection. Next, to assess the effect of k-means method in selecting the representative samples, the final semi-labeled samples (i.e., $\mathscr{S}$_*RF* and $\mathscr{S}$_*MCCR*, see Section 3.2.2) were constructed by randomly selecting samples with the same size of center patterns (generated from the k-means method) for each class. The random selection process was repeated 100 times. The classification results after the first iteration of CotrRM process for 2015 are shown in Table 3, where the column 'Initial' represents the classification result obtained by the initial training samples, and the columns 'Max' and 'Min' represent the best and worst classification results among the 100 random selections, respectively.

It can be seen that the accuracy rankings of the four strategies (i.e., 'Min', 'Initial', 'K-means', 'Max') in the three methods are the same. The difference is about 2% between the maximum and minimum OA in the 100 experiments. We visually interpreted the samples that led to the minimum Kappa and found mislabeled samples in both semi-labeled sets, i.e., $\mathscr{S}$_*RF* and $\mathscr{S}$_*MCCR*. In other words, the injection of mislabeled samples decreases the accuracy. With this k-means strategy, the OA values increase for all three classifiers when compared with 'Initial', showing that the k-means clustering method can effectively eliminate

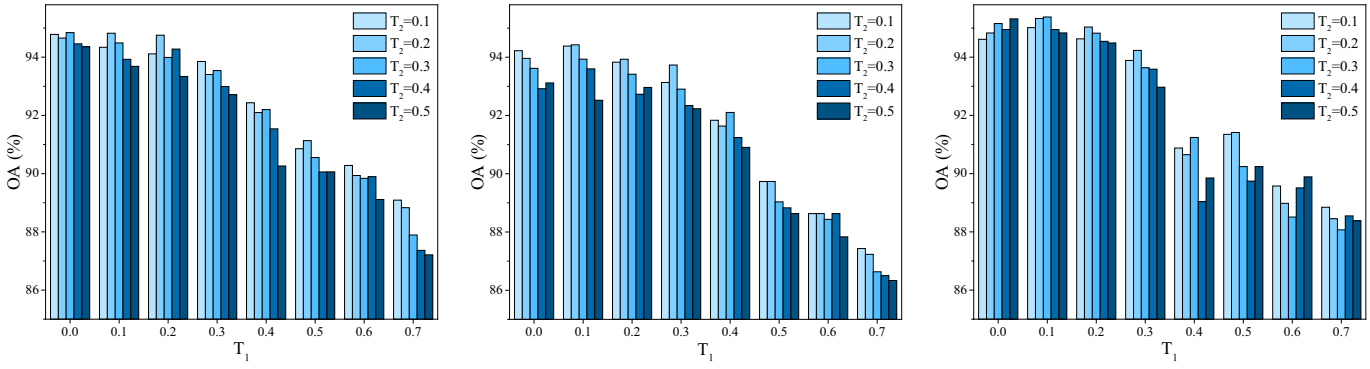


Fig. 10. Sensitivity analysis for the three classifiers of parameters T_1 and T_2 in 2015.

Table 3
OA values with different strategies to select semi-labeled samples in 2015.

(%)	*K*-means	Initial	Min	Max	ave ± std
RF	92.81	92.28	91.64	93.15	92.73 ± 0.39
MCCR	93.16	92.65	91.72	93.49	92.81 ± 0.66
CotrRM	94.04	93.33	92.49	94.42	93.92 ± 0.48

the negative effect of mislabeled samples and maintain the sample diversity at the same time. Therefore, a relatively low threshold in class certainty is preferred in the proposed approach.

5.2.2. The maximum iteration number

The experimental results show that the proposed CotrRM method further improves the performance compared to the two individual classifiers. To further investigate the performance of the CotrRM method, the performance during the iteration for the year of 2015 is presented in Fig. 11, including the number of training samples, diversity of training samples, the average certainty of the classification map, and the overall accuracy.

Fig. 11(a) shows that the number of training samples for each class increases linearly during the iteration. The diversity of the current training samples in each iteration for MCCR classifier (Fig. 11(b)) is calculated for each class by considering the standard deviation of the sample set, and the diversity of the training samples for each class gradually rises along with the iteration. The average certainties of the classification map are increased for both RF and MCCR classifier (Fig. 11(c)), and the OAs of RF and MCCR also present a gradually rising trend (Fig. 11(d)). In summary, with reliable and diverse training samples added, the OAs of the base classifiers, i.e., RF and MCCR, increase iteratively, and the CotrRM method also presents a rising trend during the iteration. It can be observed that the OA values of CotrRM remain relatively stable after 4 iterations, and, therefore, the maximum iteration number was fixed to 4 in all the experiments of this research, considering both performance and efficiency.

5.3. Comparisons

5.3.1. Comparison between MCCR and RF classifiers

The experimental results verify that both the MCCR and RF classifiers can effectively deal with contaminated data in LTS imagery. In order to comprehensively investigate their performances, in this subsection, the performances of the MCCR and RF classifiers are compared with different contamination degrees. The contamination degree of each sample was determined by the frequency of occurrence of unclear observations in all the observations, which is defined as the contamination percentage. Hence, we calculated the OA values of the RF and MCCR classifiers in identifying the test samples with different contamination percentages for each year. Fig. 12 displays the OA differences between the MCCR and RF classifiers, where a positive value denotes that the MCCR classifier performs better than the RF classifier, and a negative value indicates that RF outperforms MCCR. Please kindly note that when there is no test sample whose contamination percentage is in a given range, for instance, Year 2000 with a percentage of 50%, the corresponding bar is missing in this figure.

In general, more positive values can be observed from the figure, which shows the superiority of the proposed MCCR classifier over the RF classifier when processing LTS data with different contamination percentages. In the case of low contamination rates (e.g., < 30%), MCCR steadily outperforms RF, but when the data contamination becomes more severe (e.g., above 30%), the results of MCCR show more fluctuations. This is because, with a low contamination rate, only a few unclear observations are discarded, and hence the CR classifier can be brought into full play. Instead, when the data are heavily contaminated, the clear observations for restoring the unclear ones are insufficient, which may lead to uncertainty or instability for the construction of the CR classifier.

5.3.2. Comparison with existing methods

The “best-available-pixel” (BAP) approach is considered in this research for a comparison. BAP aims to select the “best” observation for each pixel among several candidate images, and enables the generation of a composite image free of clouds and shadows (Hermosilla et al., 2015). The pixel scoring rule for selecting the “best” observations can be referred to White et al. (2014). With the composite image, the RF, CR, and SVM (support sector machines) classifiers are adopted to conduct the classification. In addition, SVM was also selected as a comparison with the proposed CotrRM method. Fig. 13 displays the classification results of different methods, in terms of OA.

When all observations (including the unclear ones) are utilized as the input feature, both the proposed MCCR and CotrRM method perform significantly better than SVM. It can also be observed that the OA values of ‘BAP + RF’, ‘BAP + CR’ and ‘BAP + SVM’ are lower than those of MCCR and CotrRM, indicating that the recovery of unclear observations is beneficial to LTS imagery classification.

5.4. Land-cover change analysis

The classification results obtained by the proposed CotrRM method are displayed in Fig. 13. In order to further verify the efficacy of CotrRM, the multi-temporal change detection accuracy between neighboring years was also investigated. We randomly generated 60 samples for the change class and 140 samples for the non-change class based on the classification maps of two neighboring years (Olofsson et al., 2014). These reference samples were carefully labeled by

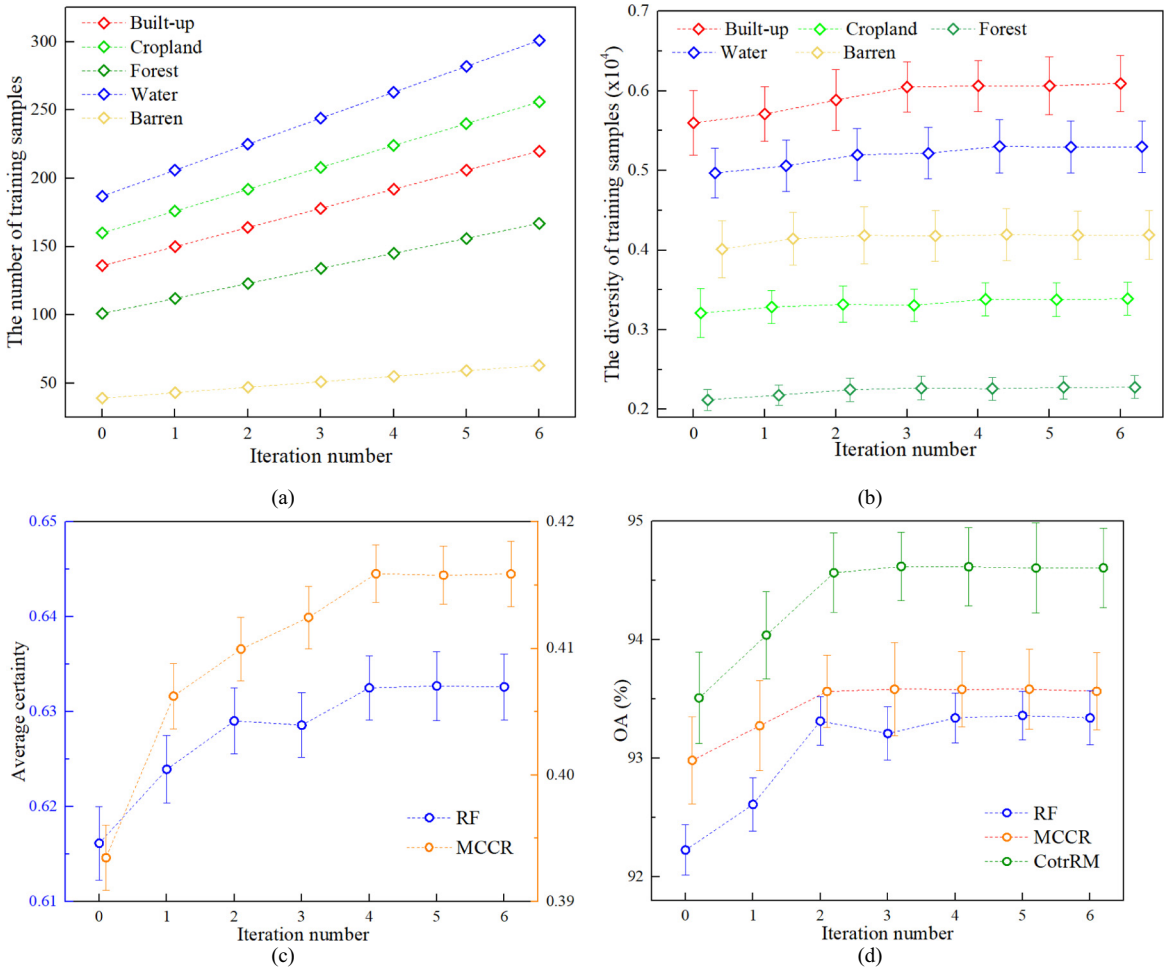


Fig. 11. The iteration process in the CotrRM method for 2015.

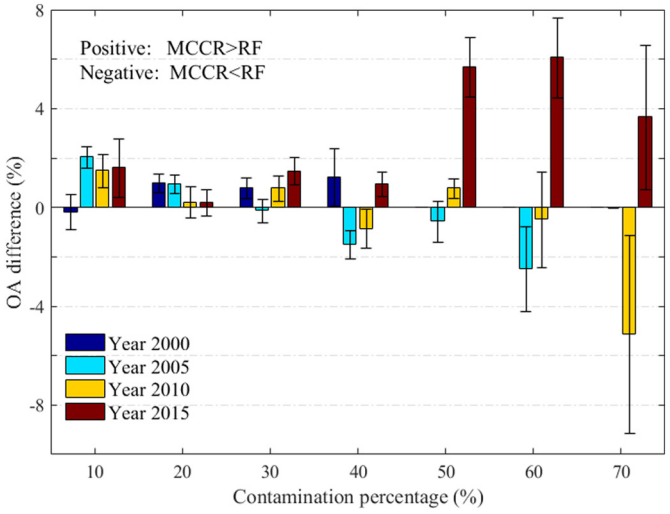


Fig. 12. The OA difference between the MCCR and RF classifiers when identifying samples with different contamination percentages.

referring to the high-resolution images, and the accuracy assessment is shown in Table 4. The results show that reliable accuracies for the change detection are achieved, with > 85% for the overall accuracies (OAs) and over 0.70 for the Kappa in the three periods.

It can be observed in Fig. 14 that built-up areas experienced successive development and expansion in Wuhan. The proportion of pixels changed to built-up from other classes was further analyzed (Table 5). During the research period, the main source of rapid urban sprawl was the loss of cropland (486.3 ± 87.0 km^2 with a 95% confidence interval). The conversion from water bodies to built-up land was also significant (163.1 ± 21.8 km^2 with a 95% confidence interval). All these land-cover changes have resulted in deterioration of the ecosystem (Zhou et al., 2014). Some representative land-cover changes are shown in Fig. 15. Sites 1 and 2 are two examples of land-cover change from 2000 to 2005, when cropland and water were converted to built-up area. The changes in Sites 3 and 4 happened between 2005 and 2010, when cropland experienced conversion to barren land. Site 5 shows the transition from cropland to built-up area and barren land from 2010 to 2015, and Site 6 indicates that water surface was converted to built-up area and barren land in this period.

6. Conclusions

Landsat time-series (LTS) imagery can provide dynamic information for accurate urban land-cover mapping. However, the observations are unavoidably contaminated due to the existence of clouds, cloud shadows, snow/ice, and SLC-off data. In this study, the MCCR classifier has been proposed to take advantage of the matrix completion (MC) algorithm and collaborative representation (CR) classifier, in order to handle the classification of unclear data in LTS imagery. The MCCR classifier deals with the unclear observations in an error-recoverable way, rather than the error-tolerant manner of the random forest (RF) classifier. RF treats the unclear observations as well as the clear ones, ignoring the degrees of reliability between them. However, the MCCR

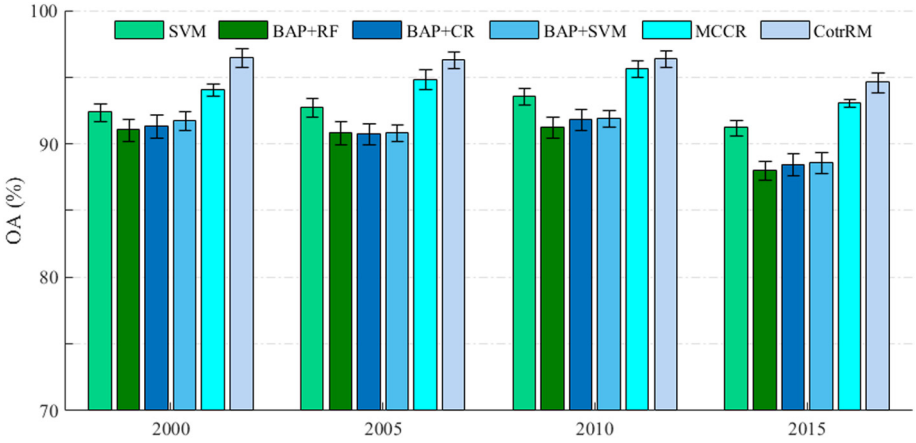


Fig. 13. Comparison between the proposed method and other ones. The 'BAP + RF', 'BAP + CR' and 'BAP + SVM' denote the BAP image is classified by RF, CR and SVM classifier, respectively.

Table 4
Accuracy assessment for the land-cover change transitions from 2000 to 2015.

	Ground reference											
	From 2000 to 2005				From 2005 to 2010				From 2010 to 2015			
	C	NC	Total	UA(%)	C	NC	Total	UA(%)	C	NC	Total	UA(%)
C	51	9	60	85.0	48	12	60	80.0	45	15	60	75.0
NC	15	125	140	89.3	10	130	140	92.9	8	132	140	94.3
Total	66	134	200		58	142	200		53	147	200	
PA(%)	77.3	93.3			82.8	91.5			84.9	89.8		
OA	88.0				89.0				88.5			
Kappa	0.72				0.74				0.72			

C: change; NC: non-change.

classifier focuses more on the utilization of clear observations to eliminate the adverse impact of unclear ones. The contaminated data in the training samples are recovered by the MC algorithm, and with respect to the unclear observations in the test samples, only the clear observations are utilized to construct the CR classifier, without loss of discriminative ability. Furthermore, inspired by the co-training paradigm, disagreements between the RF and MCCR classifiers are adopted as the source of mutual learning, and hence a novel co-training approach (CotrRM) based on semi-labeled sample selection is proposed to improve the classification performance. To the best of the authors' knowledge, it is an innovative attempt to investigate an error-recoverable method with a co-training scheme to deal with unclear LTS imagery classification.

The proposed methodology was tested with LTS imagery from the city of Wuhan, using annual time-series imagery from 2000 to 2015 for

Table 5
Statistics of the built-up area expansion.

	From 2000 to 2005	From 2005 to 2010	From 2010 to 2015
Cropland	53.05% (102.5 ± 9.5 km^2)	48.61% (181.4 ± 22.1 km^2)	38.80% (202.4 ± 22.0 km^2)
Forest	5.30% (10.2 ± 1.3 km^2)	5.80% (21.6 ± 1.9 km^2)	2.72% (14.2 ± 1.4 km^2)
Water	7.13% (13.8 ± 1.54 km^2)	14.58% (54.4 ± 3.4 km^2)	18.18% (94.9 ± 6.8 km^2)
Barren	34.52% (66.7 ± 14.7 km^2)	31.01% (115.7 ± 18.0 km^2)	40.30% (210.3 ± 62.7 km^2)

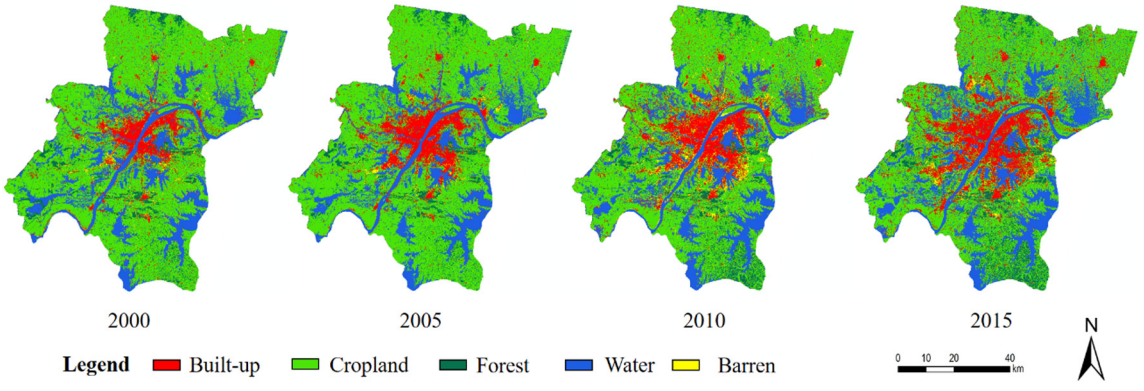


Fig. 14. The classification results for the city of Wuhan for each year.

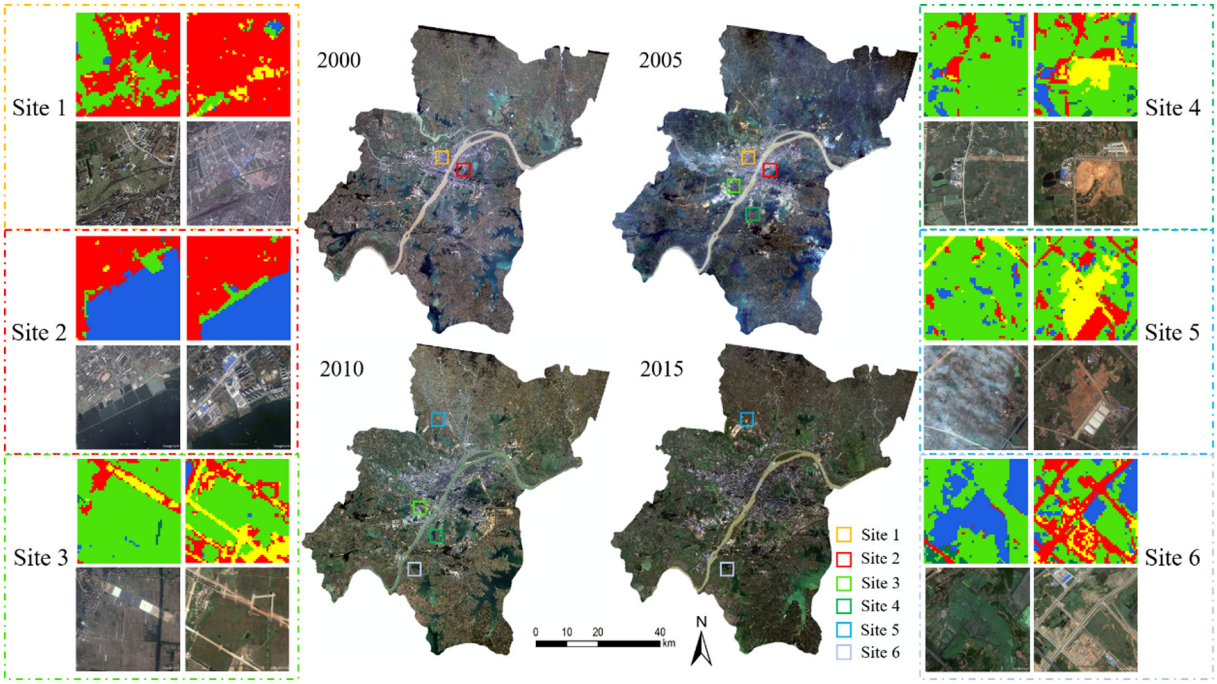


Fig. 15. Representative land-cover changes of the city of Wuhan. The images for each year are all displayed in true-color composite (R: red; G: green; B: blue), and the acquisition dates are 2000305, 2005254, 2010316, and 2015330, respectively. For each site, a 2 × 2 grid is displayed, where the top left image is the classification result before the change, the bottom left one is the Google™ image of the same year, the top right is the classification result after the change, and the bottom right is the Google™ image of the same year. The legend of the classification results is shown in Fig. 14. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

land-cover mapping. The results of the experiments confirmed that the MCCR classifier has the ability to work as effectively as the RF classifier in the presence of data noise, and the proposed CotrRM method can further enhance the classification due to the iterative learning from unlabeled data. Moreover, with all the samples labeled, the unclear observations in the whole image can be recovered with the MC algorithm, which is beneficial to the LTS data processing and other applications.

It should be noted that the MCCR classifier could be applied to other remote sensing data with contaminated observations. Furthermore, base classifiers of the co-training paradigm are not limited to the RF and MCCR, and other classifiers that provide crisp and soft predictions can also be adopted to strengthen the time-series urban land-cover mapping.

Acknowledgements

The research was supported by the National Natural Science Foundation of China under Grants 41522110 and 41771360, the Hubei Provincial Natural Science Foundation of China under Grant 2017CFA029.

References

Beckschäfer, P., 2017. Obtaining rubber plantation age information from very dense Landsat TM & ETM + time series data and pixel-based image compositing. Remote Sens. Environ. 196, 89–100.
Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.
Bhandari, S., Phinn, S., Gill, T., 2012. Preparing Landsat Image Time Series (LITS) for monitoring changes in vegetation phenology in Queensland, Australia. Remote Sens. 4, 1856.
Blum, A., Mitchell, T., 1998. Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, pp. 92–100 (USA).
Box, J.F., 1987. Guinness, Gosset, Fisher, and small samples. Stat. Sci. 2, 45–52.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Bruzzone, L., Chi, M., Marconcini, M., 2006. A novel transductive SVM for semisupervised classification of remote-sensing images. IEEE Trans. Geosci. Remote Sens. 44, 3363–3373.
Cabral, R., Torre, F.D.L., Costeira, J.P., Bernardino, A., 2015. Matrix completion for weakly-supervised multi-label image classification. IEEE Trans. Pattern Anal. Mach. Intell. 37, 121–135.
Cai, J.-F., Candès, E.J., Shen, Z., 2010. A singular value thresholding algorithm for matrix completion. SIAM J. Optim. 20, 1956–1982.
Candès, E.J., Recht, B., 2009. Exact matrix completion via convex optimization. Found. Comput. Math. 9, 717.
Chen, J., Yang, J., 2014. Robust subspace segmentation via low-rank representation. IEEE Trans. Cybern. 44, 1432–1445.
Chen, Y., Song, X., Wang, S., Huang, J., Mansaray, L.R., 2016. Impacts of spatial heterogeneity on crop area mapping in Canada using MODIS data. ISPRS J. Photogramm. Remote Sens. 119, 451–461.
Estoque, R.C., Murayama, Y., 2017. Monitoring surface urban heat island formation in a tropical mountain city using Landsat data (1987–2015). ISPRS J. Photogramm. Remote Sens. 133, 18–29.
Fritz, S., See, L., Rembold, F., 2010. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. Int. J. Remote Sens. 31, 2237–2256.
Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogramm. Remote Sens. 116, 55–72.
Gray, J., Song, C., 2013. Consistent classification of image time series with automatic adaptive signature generalization. Remote Sens. Environ. 134, 333–341.
Grinand, C., Rakotomalala, F., Gond, V., Vaudry, R., Bernoux, M., Vieilledent, G., 2013. Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier. Remote Sens. Environ. 139, 68–80.
Grogan, K., Pflugmacher, D., Hostert, P., Kennedy, R., Fensholt, R., 2015. Cross-border forest disturbance and the role of natural rubber in mainland Southeast Asia using annual Landsat time series. Remote Sens. Environ. 169, 438–453.
Haase, D., Larondelle, N., Andersson, E., Artmann, M., Borgström, S., Breuste, J., Gomez-Baggethun, E., Gren, Å., Hamstead, Z., Hansen, R., Kabisch, N., Kremer, P., Langemeyer, J., Rall, E.L., McPhearson, T., Pauleit, S., Qureshi, S., Schwarz, N., Voigt, A., Wurster, D., Elmqvist, T., 2014. A quantitative review of urban ecosystem service assessments: concepts, models, and implementation. Ambio 43, 413–433.
Han, Y., Li, S., Zheng, Y., 2009. Predictors of nutritional status among community-dwelling older adults in Wuhan, China. Public Health Nutr. 12, 1189–1196.
Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. Remote Sens. Environ. 158, 220–234.
Huang, X., Lu, Q., Zhang, L., 2014a. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. ISPRS J. Photogramm. Remote Sens. 90, 36–48.
Huang, X., Lu, Q., Zhang, L., Plaza, A., 2014b. New postprocessing methods for remote

sensing image classification: a systematic study. IEEE Trans. Geosci. Remote Sens. 52, 7140–7159.
Huang, X., Wen, D., Li, J., Qin, R., 2017. Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery. Remote Sens. Environ. 196, 56–75.
Kontgis, C., Schneider, A., Ozdogan, M., 2015. Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data. Remote Sens. Environ. 169, 255–269.
Lance, G.N., Williams, W.T., 1967. A general theory of classificatory sorting strategies: II. Clustering systems. Comput. J. 10, 271–277.
Li, J., Zhang, H., Zhang, L., 2014a. Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification. ISPRS J. Photogramm. Remote Sens. 94, 25–36.
Li, J., Zhang, H., Zhang, L., Huang, X., Zhang, L., 2014b. Joint collaborative representation with multitask learning for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 52, 5923–5936.
Li, J., Zhang, H., Huang, Y., Zhang, L., 2014c. Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary. IEEE Trans. Geosci. Remote Sens. 52, 3707–3719.
Li, X., Shen, H., Zhang, L., Li, H., 2015. Sparse-based reconstruction of missing information in remote sensing images from spectral/temporal complementary information. ISPRS J. Photogramm. Remote Sens. 106, 1–15.
Lin, C., Li, Y., Yuan, Z., Lau, A.K.H., Li, C., Fung, J.C.H., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2.5. Remote Sens. Environ. 156, 117–128.
Long, H., Liu, Y., Wu, X., Dong, G., 2009. Spatio-temporal dynamic patterns of farmland and rural settlements in Su–Xi–Chang region: implications for building a new countryside in coastal China. Land Use Policy 26, 322–333.
Lu, D., Li, G., Moran, E., Batistella, M., Freitas, C.C., 2011. Mapping impervious surfaces with the integrated use of Landsat Thematic Mapper and radar data: a case study in an urban–rural landscape in the Brazilian Amazon. ISPRS J. Photogramm. Remote Sens. 66, 798–808.
Ma, S., Goldfarb, D., Chen, L., 2011. Fixed point and Bregman iterative methods for matrix rank minimization. Math. Program. 128, 321–353.
Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O., 2017. Remote sensing of species diversity using Landsat 8 spectral variables. ISPRS J. Photogramm. Remote Sens. 133, 116–127.
Matthew, C.H., Alexander, K., Alexandra, T., Peter, V.P., Svetlana, T., Bryan, Z., Suspense, I., Belinda, M., Fred, S., Rebecca, M., 2016. Humid tropical forest disturbance alerts using Landsat data. Environ. Res. Lett. 11, 034008.
Middel, A., Häb, K., Brazel, A.J., Martin, C.A., Guhathakurta, S., 2014. Impact of urban form and design on mid-afternoon microclimate in Phoenix local climate zones. Landsc. Urban Plan. 122, 16–28.
Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A.J., Hostert, P., 2015. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. Remote Sens. Environ. 156, 490–499.
United Nations, 2014. World Urbanization Prospects: The 2014 Revision, Highlights. Department of Economic and Social Affairs. Population Division, United Nations.
Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57.
Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. Remote Sens. Environ. 187, 156–168.
Persello, C., Bruzzone, L., 2014. Active and semisupervised learning for the classification of remote sensing images. IEEE Trans. Geosci. Remote Sens. 52, 6937–6956.
Pontius Jr., R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int. J. Remote Sens. 32, 4407–4429.
Qin, Y., Xiao, X., Dong, J., Zhou, Y., Zhu, Z., Zhang, G., Du, G., Jin, C., Kou, W., Wang, J., Li, X., 2015. Mapping paddy rice planting area in cold temperate climate region through analysis of time series Landsat 8 (OLI), Landsat 7 (ETM+) and MODIS imagery. ISPRS J. Photogramm. Remote Sens. 105, 220–233.
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104.
Ruta, D., Gabrys, B., 2000. An overview of classifier fusion methods. J. Comput. Inf. Syst. 7, 1–10.
Schneider, A., 2012. Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. Remote Sens. Environ. 124, 689–704.
Senf, C., Pflugmacher, D., Wulder, M.A., Hostert, P., 2015. Characterizing spectral–temporal patterns of defoliator and bark beetle disturbances using Landsat time series. Remote Sens. Environ. 170, 166–177.
Tan, K., Li, E., Du, Q., Du, P., 2014. An efficient semi-supervised classification approach for hyperspectral imagery. ISPRS J. Photogramm. Remote Sens. 97, 36–45.
Tao, D., Jin, L., Yang, Z., Li, X., 2013. Rank preserving sparse learning for Kinect based scene classification. IEEE Trans. Cybern. 43, 1406–1417.
Tuia, D., Pasolli, E., Emery, W.J., 2011. Using active learning to adapt remote sensing image classifiers. Remote Sens. Environ. 115, 2232–2242.
Wang, L., Gong, W., Xia, X., Zhu, J., Li, J., Zhu, Z., 2015. Long-term observations of aerosol optical properties at Wuhan, an urban site in Central China. Atmos. Environ. 101, 94–102.
Waqas, J., Yi, Z., Zhang, L., 2013. Collaborative neighbor representation based classification using l2-minimization approach. Pattern Recogn. Lett. 34, 201–208.
White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A., Guindon, L., 2014. Pixel-based image compositing for large-area dense time series applications and science. Can. J. Remote. Sens. 40, 192–212.
White, J.C., Wulder, M.A., Hermosilla, T., Coops, N.C., Hobart, G.W., 2017. A nationwide annual characterization of 25 years of forest disturbance and recovery for Canada using Landsat time series. Remote Sens. Environ. 194, 303–321.
Wuhan Municipal Statistics Bureau, 2013. Wuhan Statistical Yearbook. China Statistics Press, Wuhan.
Xu, J., He, H., Man, H., 2012. DCPE co-training for classification. Neurocomputing 86, 75–85.
Ying, Q., Hansen, M.C., Potapov, P.V., Tyukavina, A., Wang, L., Stehman, S.V., Moore, R., Hancher, M., 2017. Global bare ground gain from 2000 to 2012 using Landsat imagery. Remote Sens. Environ. 194, 161–176.
Yuan, F., Sawaya, K.E., Loeffelholz, B.C., Bauer, M.E., 2005. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. Remote Sens. Environ. 98, 317–328.
Zhang, L., Weng, Q., 2016. Annual dynamics of impervious surface in the Pearl River Delta, China, from 1988 to 2013, using time series Landsat imagery. ISPRS J. Photogramm. Remote Sens. 113, 86–96.
Zhang, M.-L., Zhou, Z.-H., 2011. CoTrade: confident co-training with data editing. IEEE Trans. Syst. Man Cybern. B Cybern. 41, 1612–1626.
Zhang, L., Yang, M., Feng, X., Ma, Y., Zhang, D., 2012. Collaborative representation based classification for face recognition. In: arXiv Preprint arXiv:1204.2358.
Zhang, X., Song, Q., Liu, R., Wang, W., Jiao, L., 2014. Modified co-training with spectral and spatial views for semisupervised hyperspectral image classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 2044–2055.
Zhou, K., Liu, Y., Tan, R., Song, Y., 2014. Urban dynamics, landscape ecological security, and policy implications: a case study from the Wuhan area of central China. Cities 41, 141–153.
Zhu, Z., 2017. Change detection using Landsat time series: a review of frequencies, preprocessing, algorithms, and applications. ISPRS J. Photogramm. Remote Sens. 130, 370–384.
Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94.
Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171.
Zhu, L., Xiao, P., Feng, X., Zhang, X., Huang, Y., Li, C., 2016a. A co-training, mutual learning approach towards mapping snow cover from multi-temporal high-spatial resolution satellite imagery. ISPRS J. Photogramm. Remote Sens. 122, 179–191.
Zhu, Z., Fu, Y., Woodcock, C.E., Olofsson, P., Vogelmann, J.E., Holden, C., Wang, M., Dai, S., Yu, Y., 2016b. Including land cover change in analysis of greenness trends using all available Landsat 5, 7, and 8 images: a case study from Guangzhou, China (2000–2014). Remote Sens. Environ. 185, 243–257.