Optical remote sensing cloud detection based on random forest only using the visible light and near-infrared image bands

Xudong Yaoab, Qing Guoa, An Liab and Lu Shiab

* Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; bSchool of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT
Cloud detection is of great significance for optical remote sensing images. Most existing cloud detection approaches often rely on many thresholds among multiple bands with a wide spectrum range, and normally just can be applied to specific satellite data. Cloud detection is difficult in some sensor data with limited number of available spectral bands. To tackle this challenge, we propose a cloud detection method based on random forest (RFCD) only with the most common RGB and NIR bands. The RFCD normalizes different sensor data by calculating the top-of-atmosphere (TOA) reflectance and combines stable spectral and texture features extracted from the most common bands, which improves accuracy, reduces reliance on multi-band information, and improves the extendibility of the method. Moreover, the RFCD effectively avoids the complicated threshold setting and reduces subjective factors. Few parameters and strong generalization ability of RF further improve the extendibility of RFCD and provide the application possibility in a variety of data. Experimental results show the total validity obtained by 107 Landsat-8 images is 93.46%. The accuracy of the RFCD is higher than the Function of mask method and similar to the SegNet deep learning method, while the RFCD need fewer training samples and hyperparameters which makes it easier to be used. The extended experimental results show RFCD also gets good detection results in Sentinel-2 and GaoFen-1 images. The new RFCD is an accurate cloud detection method with strong extendibility and stability, also avoiding the complicated threshold setting.

Introduction

According to the data from the International Satellite Cloud Climatology Project (ISCCP), the annual average cloud coverage accounts for more than 60% of the world’s total area (Zhang et al., 2004). In the imaging process of optical remote sensing images, the cloud hinders the transmission of signals between the sensor and the ground, resulting in the change or loss of the spectral information. Such cloud-covered images have less available information, which will cause a lot of inconvenience to the analysis and application of images (Tan et al., 2016). In addition, with the increase of remote sensing data, it has become a very urgent task to quickly eliminate useless information and retain available information. Therefore, it is necessary to study the cloud detection technology of remote sensing images. The detected cloud cover can help eliminate useless images, reduce the pressure of processing system, and it can also provide users with references when selecting and downloading data. At the same time, the accurate detection of cloud regions is conducive to subsequent image processing, and provides a basis for further image analyses and applications.

Many scholars have conducted long-term and in-depth research on cloud detection, and have obtained abundant results. The cloud detection has also experienced from the threshold methods based on physical characteristics to the methods based on texture features, and now the methods based on machine learning. The physical threshold methods determine the most suitable thresholds for separating the cloud and the ground based on the differences in the visible light band reflectivity, the infrared band brightness temperature and other normalized indexes. For Irish (2000); (2006)) designs a series of spectral thresholds by analysing the imaging characteristics of clouds in different bands, and proposed an automatic cloud cover assessment (ACCA) algorithm for Landsat TM/ETM+ images. Zhu and Woodcock (2012); (2015)) proposes the the Function of mask (Fmask) method. This algorithm builds a probability model by combining the information of multiple bands to calculate the probability that each pixel belongs to the cloud and determines the potential cloud area. Zhong et al. (2017) proposes an object-based cloud and cloud shadow detecting algorithm which needs to combine the complex corresponding relationship between the cloud and the cloud shadow. Zhai et al. (2018)
proposes a cloud detection algorithm using spectral indices based on the physical reflectivity characteristics for multi/hyperspectral optical remote sensing imagery. The analysis performed in these existing methods shows that the threshold-based methods have a wide range of applications. However, the setting of the thresholds requires much human involvement and is highly subjective. The thresholds will vary with the seasons and geographic location of remote sensing images. Therefore, it needs a lot of experiments to determine the thresholds that are suitable for most images. In addition, due to the differences in the band range and spectral response function of different remote sensing sensors, most physical threshold methods can only detect certain specific products and cannot be extended to other sensors.

With the continuous improvement of the spatial resolution in remote sensing images, the texture information contained in the images is more abundant. Therefore, some methods based on texture features and spatial characteristics are also applied to the cloud detection (Azimi-Sadjadi et al., 1996; Dong et al., 2018; Shan et al., 2009; Zhang & Xiao, 2014). At present, the main texture extraction methods in cloud detection include the fractal dimension, the gray-level co-occurrence matrix and the multiple bilateral filtering, etc. Shan et al. (2009) uses the fractal dimension and the angular second moment of the gray-level co-occurrence matrix to implement cloud detection based on a tree-like discriminant structure. Dong et al. (2018) first uses the linear iterative clustering to obtain the initial cloud detection results, then the results are refined according to the second-order matrix and texture average of the object. The introduction of texture features improves the accuracy and extendibility of the algorithm, but still requires complex threshold determination.

In recent years, some researches obtain more accurate cloud detection results by introducing deep learning (Zhang et al., 2001; Jang et al., 2006; Chai et al., 2019; Z.W. Li et al., 2019; Chen, Tang, et al., 2020) which includes the radial basis function network, the multi-scale convolutional feature fusion (MSCFF), the SegNet, the 3D convolutional neural network and so on. For example, Chai et al. (2019) propose an adaptation of SegNet to detect clouds in Landsat imagery. This method effectively improves the accuracy of detection. In addition, some deep learning methods designed for high-resolution optical remote sensing images usually only rely on the RGB and NIR bands. For example, Chen, Tang, et al. (2020) propose an end-to-end multi-scale 3D cirrocumulus detection method for high-resolution images which makes full use of the spatial and spectral information under the condition of limited spectral range. Segal-Rozenhaimer et al. (2020) present a convolutional neural network algorithm and perform cloud detection for World-View-2 and Sentinel-2 data based on the RGB and NIR bands. However, these methods based on deep learning generally have a high computational complexity and are time-consuming (Zhai et al., 2018). At the same time, these methods need to make abstract decisions at multiple levels, which leads to very complex relationships between the dependent and independent variables. That makes the model a “black box” with insufficient interpretability. Moreover, the deep learning methods require a large amount of training data and sometimes the cost of acquiring training data is very high. For some new sensor data, it is difficult to obtain a large amount of high-quality label data for training in a short time. Finally, they are highly dependent on the computer configuration, which makes difficult to realize a wide range of applications (Wei et al., 2020).

Random forest (RF) is a machine learning algorithm which has developed rapidly in recent years. Belgiu and Dragut (2016) compare RF with other machine learning algorithms. The comparison results show that RF can get better classification results, especially when dealing with high-dimensional or small sample data. Moreover, RF is faster than the support vector machine (SVM) or other integrated learning algorithms such as the adaptive boosting (AdaBoost). In addition, RF consists of decision trees and can provide the importance of different features. RF feature importance can be used as a global interpretability measure (Hall et al., 2017) and the interpretability makes the results relatively controllable. Many scholars have applied RF to the cloud detection. Wei et al. (2020) use the top-of-atmosphere reflectance and brightness temperature to realize cloud detection of Landsat-8 (L8) data based on RF. Chen, Liu, et al. (2020) realize cloud detection of L8 and Sentinel-2 data based on RF without thermal bands. In addition, according to the large number theory, increasing the number of weak learners in RF can reduce the generalization error. Fu et al. (2019) train the RF model relying on source data at one moment, and then the trained model is applied to images of other times. This experiment proves that RF has certain extendibility in different images.

Aiming at these problems in the above analysis and combining multiple advantages of the RF, we use the most common RGB and NIR bands to extract a variety of spectral and texture features which are relatively stable in different optical remote sensing images. And a cloud detection method based on RF (RFCD) is proposed based on these features. The RFCD can avoid the complicated threshold determination process in the traditional threshold method, and improves the detection accuracy. Secondly, the RFCD uses the most common bands which reduces the reliance on multi-band information and lays the foundation for the expansion of different sensor data. The RFCD converts digital number (DN) value to top-of-
atmosphere (TOA) reflectance value, which initially realizes the normalization among different sensors and further lays the foundation for extendibility. Thirdly, these extracted features which are stable in different sensor data ensure the extendibility while improving accuracy. Finally, the random forest has a few parameters, relatively simple training samples, strong interpretability, and generalization capabilities, which provides a certain degree of usability.

Section 2 introduces the cloud detection method based on RF. In section 3, we first evaluate the validity and the quantity performance of the RFCD in L8 data. Then, we apply the RFCD to Sentinel-2 and GaoFen-1 (GF-1) images to verify the extendibility and stability. In section 4, we give a summary and conclusion of this study.

Random forest cloud detection

RF is a multi-classifier model based on classification and regression tree (CART; Breiman, 2001a). It generates a strong learner through combining a large number of weak learners, which has higher accuracy than a single classifier and can effectively prevent overfitting. At present, RF has been widely used in many fields due to its advantages such as small number of training samples, strong generalization ability, less manual intervention, fast calculation speed and high classification accuracy (Breiman, 2001b).

In this paper, we build a classification model based on RF, and combine the extracted multiple classification features to perform cloud detection in optical remote sensing images. It is mainly divided into two phases: training and testing. In the training phase, we first perform radiation correction to obtain the reflectivity at the top of the atmosphere from the RGB and NIR bands. Then, we select and extract a series of spectral and texture features based on reflectivity. At the same time, we select representative cloud and non-cloud pixels from the training images to form a training sample set. Finally, we make them as input of the RF to train and get the trained RF model. In the testing phase, we first input the images to be detected into the RF model and obtain the initial cloud detection results. Then we use guided filtering to refine the results to get more accurate cloud distribution maps and calculate the cloud covers. The flow chart is shown in Figure 1.

Radiometric calibration

Optical remote sensing images normally display the DN not the reflectivity. In order to eliminate the error generated by the sensor and determine the radiation at the entrance pupil of the sensor, it is necessary to perform the radiation calibration to convert the DN into the reflectivity at the top of the atmosphere. This process provides the basis for normalization between various parts of data in one image or between different images. The calibration formula is as follows:

\[ L_{\lambda} = (\text{Gain})(\text{DN}) + (\text{Bias}) \]  

\[ \rho_{\text{TOA}} = \frac{\pi L_{\lambda} d^{2}}{\text{(ESNU)} \cos \theta} \]

where \( L_{\lambda} \) denotes the radiance of the sensor; Gain denotes the calibration slope; Bias denotes the calibration intercept; \( \rho_{\text{TOA}} \) denotes the reflectivity at the top of the atmosphere; \( d \) denotes the solar-earth astronomical unit distance; ESNU denotes the solar average spectral radiation at the top of the atmosphere; \( \theta \) is the sun zenith angle. The parameters used in the above calibration process are included in the image header file.

Feature selection and extraction

Features are the basis of classification and directly affect the classification accuracy. In view of the limitations of a single feature in distinguishing the cloud from the ground surface, the combination of multiple features can effectively improve the accuracy of cloud detection. Based on the difference in reflectivity and texture characteristics between the cloud and the ground, we extract a series of new classification features from common four bands of the RGB and NIR. Multiple classification features increase the dimensionality of the data and improve the accuracy of classification to a certain extent.

Band feature

The band feature is the most intuitive feature in remote sensing images, and the images can be directly interpreted and analysed based on this feature. In the process of the remote sensing image interpretation, it is generally believed that each type of ground surface has its corresponding spectral characteristic curve in each band. Therefore, the reflectivity in each band can be used as the main basis for judgment. Due to the unique reflectivity of clouds, they often show high brightness and continuous coverage in optical remote sensing images. The reflected information of the RGB and NIR bands can be used as feature to distinguish clouds from most ground surface types.

Color feature

In the RGB color space, there is a big difference between thick clouds and ground surfaces. But thin clouds are usually darker and contain part of the ground surface information, so the detection results are often unsatisfactory (Y. X. Xu et al., 2006). In the IHS color space, both thick and thin clouds can show significant characteristics. The clouds are slightly
white in the RGB color space, and the cloud reflectivity of each band is relatively large and similar. After the RGB color space is converted to the IHS color space, the pixel values of clouds in the intensity channel (I) will be significantly larger than the ground surface and the pixel values in the saturation channel (S) will be significantly smaller than the ground surface. The calculation process of intensity and saturation is as follows:

\[
I = \frac{R + G + B}{3} \tag{3}
\]

\[
S = 1 - \frac{3 \min(R, G, B)}{R + G + B} \tag{4}
\]

where \(R, G, B\) represent the reflectivity of the three visible light bands; \(I\) and \(S\) represent the value of the intensity channel and the saturation channel respectively.

In addition, in order to enhance the contrast between clouds and ground objects, we construct a basal figure as a new feature (Kai. et al., 2016), and the calculation process is as follows:

\[
J' = \frac{I' + \tau}{S' + \tau} \tag{5}
\]

where \(I'\) denotes the normalized intensity channel; \(S'\) denotes the normalized saturation channel; \(\tau\) is the buffer coefficient, which usually takes a value between 0 and 1 to prevent the value from being too large. This article sets it to 0.05. \(J'\) is stretched into the corresponding gray scale to get the basal figure.

**Dark channel feature**

He et al. (2010) proposed the dark channel theory. Through the observation and analysis of a large number of prospect images, it is found that there would always be some pixels in the image that contain a very low value in the three color channel components of RGB. Research has shown that after the remote sensing images are processed by the dark channel, the pixel values of some areas with a single color drop a lot, and even tends to 0, while the pixel values of clouds can still be maintained at a high level (Zhu, 2018). The dark channel feature extraction process can be expressed by the following formula.

![Flow chart of cloud detection](image-url)
\[ B_{\text{dark}} = \min_{x \in \{R,G,B\}} (\text{Band})_x \]  

(6)

where \((\text{Band})_x\) represents the reflectivity of the image in the RGB bands; \(B_{\text{dark}}\) represents the value of the dark channel.

**Whit ness index feature**

The cloud has a relatively flat reflectivity in the RGB and NIR bands and usually displays as white. Zhu and Woodcock (2012) proposed the whitness index and applied it to the cloud detection of Landsat images, which effectively eliminated a large number of non-cloud pixels. The definition of the whitness is shown in the following formula.

\[
    m = \frac{B_1 + B_2 + B_3}{3} 
\]  

(7)

\[
    W = \sum_{i=1}^{3} \frac{|B_i - m|}{m} 
\]  

(8)

where \(B_1, B_2, B_3\) are the reflectivity of the three RGB bands, respectively, \(m\) denotes the average reflectivity of the three bands. \(W\) is the whitness index. The whitness index value of the cloud pixel is generally small, while the one of ground is relatively large since the color of the ground is rich and the ground reflectivity of each band varies greatly.

**Cloud index feature**

Zhai et al. (2018) proposed a cloud index (CI) based on the physical reflectivity of cloud to distinguish between clouds and ground surface. CI is expressed as follows.

\[
    \text{CI} = \frac{3B_{\text{NIR}}}{B_R + B_G + B_B} 
\]  

(9)

where \(B_R, B_G, B_B\) and \(B_{\text{NIR}}\) denote the reflectivity of the RGB and NIR bands, respectively.

CI is used to measure the similarity of the reflectivity between the RGB and NIR bands. In most cases, the reflectivity of clouds is relatively similar in the RGB and NIR bands, so the CI value of the cloud is usually around 1. The CI value of areas with high vegetation coverage is usually higher than 1, and areas with low vegetation coverage, such as bare land, have a lower CI value than 1. We consider the absolute value of the difference between the CI value and 1 as a feature. The smaller the difference value, the greater the probability that the pixel belongs to the cloud. Therefore, in the feature map, the cloud has a lower pixel value than the ground object.

**Gabor transform feature**

Gabor transform is a local Fourier transform (Lee, 1996), which divides the signal into multiple intervals, and performs Fourier transform separately in each interval to obtain the local characteristics of the signal. The 2D Gabor filter is similar to the visual stimulus response of human visual cells. It is sensitive to edges and insensitive to light changes. It can well extract the local spatial and frequency domain information of the target, and is very suitable for the texture analysis. Clouds in optical remote sensing images usually have simple textures. We can use the Gabor filtering to extract texture information to exclude some areas with high brightness but complex textures.

Gabor transform has good direction and scale characteristics. However, the cloud does not have obvious directivity. The features extracted by using different filters with different scales and directions are mutually redundant. Therefore, through the comparative analysis between features, we only select the feature with a large separation to reduce the feature redundancy. We select six scales with values of 7, 9, 11, 13, 15, 17 and four angles with values of 0°, 45°, 90°, 135° to extract texture features of the intensity map and finally obtain 24 texture feature maps. By comparing the differences between the clouds and the ground surface in the feature maps, the resulting image with a scale of 7 and an angle of 45° is finally selected as the classification feature.

In order to observe the effect of different features more intuitively, an example of the above classification features for cloud detection is shown in Figure 2.

### 2.3 Training sample set construction

Abundant training samples can improve the classification performance of RF, so the establishment of the training sample set is an important part of this research. In this study, we choose the “L8 Biome” cloud validation masks for cloud detection experiment and verification (U.S. Geological Survey, 2016). The “L8 Biome” dataset has a total of 96 Level-1 T scenes (images size 7000 × 7000) and the corresponding manually generated cloud masks, which are evenly distributed around the world. Moreover, it contains a variety of seasons and land surface types, including barren, forest, grass, shrubland, snow, water, and wetland. L8 Biome is a relatively comprehensive and recognized dataset, and many researches are based on it (Skakun et al., 2019; Joshi et al., 2019; H et al., 2019; K Xu et al., 2019; Chai et al., 2019; Ma et al., 2020; Chen, Liu et al. 2020; Wei et al., 2020). Our experiments are also done based on L8 Biome.

We select one-third of the images from the data set for the establishment of the training sample set, and there are four training images for each land surface type. Except the snow biome, it is currently difficult to identify a large range of weakly textured snow from the RGB and NIR bands. But in other biomes, there are still some images that contain some snow for training and testing. Considering the diversity of
cloud samples, the training images also contain abundant clouds with different shapes and densities. Then, we select a large number of representative pixels from the training images to form the training sample set. When selecting samples that can typically represent different types of clouds and ground surface from L8 images, in order to achieve the optimal classification effect, the number of samples for each type should be allocated in proportion to the area. Therefore, the type that occupies a large area should correspond to more samples than that of the small one. In the end, our training set includes 300,000 cloud pixels and 300,000 non-cloud pixels.

**Random Forest model construction**

The RF consists of a large number of decision trees whose construction process is as follows. It first performs bootstrap sampling on the input samples to obtain k sample subsets. The size of each subset is about two-thirds of the original training set. Then the RF constructs a decision tree corresponding to the subset one by one by randomly selecting features. Each decision tree selects the optimal attribute to split based on the minimum Gini coefficient criterion, and prunes the classification tree to reduce the amount of calculation. The obtained k decision trees constitute a RF. The final detection result is determined by voting on the output results of all decision trees. In the above process, the random selection of sample subsets and the minimum Gini coefficient principle used in the establishment of decision trees can effectively reduce the correlation between decision trees, enhance the prediction ability of a single decision tree, improve the generalization ability and the classification accuracy of the entire RF. The cloud detection process of remote sensing image based on random forest is shown in Figure 3. First, we extract the spectrum and texture features from the original image to be detected. Then, the decision trees select the corresponding

![Figure 2. The classification features for cloud detection. (a) The reflectivity in RGB bands; (b) The reflectivity in NIR band; (c) l; (d) S; (e) The basal figure; (f) The dark channel; (g) The whitness index; (h) The cloud index; (i) Gabor transform feature.](image-url)
feature subset for detection. Finally, k results are obtained, and the final cloud detection result is determined by voting on the results of all decision trees.

In the bootstrap sampling process of the sample data, about 30% of the sample is not selected. This part of the sample is usually called out of bag (OOB) data. We use the OOB data corresponding to each decision tree as the test data to calculate the OOB error of the tree. The average of all OOB errors can be used as the generalization error of the RF to evaluate the performance. Therefore, the RF does not need to use an independent test set to obtain an unbiased estimate of the error. On the contrary, it can establish an unbiased estimate of the error during the generation process, that is, calculate the OOB error during the construction of the RF. Different from the cross-validation required by other machine learning methods, this validation method has high operating efficiency and takes up less resources, especially, the results are similar to the cross-validation results (Fang et al., 2011). This paper also uses the OOB error as the evaluation index in subsequent experiments. The smaller the OOB error, the better the performance of the algorithm.

In the RF construction process, only two parameters need to be defined: the number of decision trees (Ntree) and the maximum feature number of decision trees (Mtry). Research shows (2017) that with the increase of Ntree, the classification performance and the generalization ability of RF gradually improve, but the computational complexity is also increased. The larger the Mtry, the more information a single decision tree has and the easier to overfit. As the Mtry value decreases, the prediction accuracy of the decision trees will decrease. In remote sensing applications, the most common sets are Ntree to 500 and Mtry to the square root of the number of input features (Belgiu and Dragut, 2016). In order to find the optimal Ntree and Mtry values, we use the OOB error of the RF as the evaluation standard, and analysis the influence of the two parameters.

The relationship between Ntree and the OOB error of the RF model is shown in Figure 4 where the value range of Ntree is set to 1~500. It can be seen that when the Ntree is between 100 and 500, the OOB error value basically tends to be stable, and there is no obvious change. This phenomenon shows that the performance of RF at this time tends to be stable.

In the same way, we fix the number of decision trees to 100, and adjust the size of Mtry. As shown in Figure 5, we get the relationship between the OOB error and the Mtry. When the value of Mtry is 6, the OOB error is the smallest. Therefore, considering the two aspects, the optimal parameters of the RF model are finally determined as Ntree = 100 and Mtry = 6.

Another advantage of RF is that it can evaluate the importance and contribution of input features, so that the model has a certain degree of interpretability (Calle & Urrea, 2011). Figure 6 shows the importance score of each feature. The higher the importance scores, the greater the influence and contribution of the feature is to the classification result.

The results show that discrete spectral channels in the visible light bands are important, especially the blue band. The dark channel and the basal figure are also important. This is mainly because the dark channel combines the information of the visible light bands. The basal figure combines the intensity and saturation information, and the contrast between the cloud and the non-cloud pixels in the basal figure is great. The Gabor transform feature that contains the texture information also has a high importance score, which shows that the texture feature plays an important role in the classification process. The importance scores of other features are not high, which may be due

Figure 3. Flowchart of the RF classification.
to a certain degree of the information redundancy between different features. But they still contribute to the classification.

**Refined processing**

In remote sensing images, clouds usually appear as continuous coverage, rather than the isolated cloud pixels. However, due to the influence of factors such as the illumination, the detection results obtained by the RF will have a number of holes in the cloud area, or some misjudgment points in the non-cloud area. In order to effectively remove the isolated noise points, further improve the accuracy of the cloud detection algorithm, and ensure the integrity of the cloud area, we use the guided filtering to refine the detection results.

The guided filter is a kind of adaptive weight filter. In the image processing process, by constructing the local linear relationship between the guided image and the output image, the boundary of the guided image can be maintained in the output image, so that noise points can be effectively removed (He et al., 2013). The guided filtering can be expressed as:

\[ q_i = \sum_j W_{ij}(I) S_j \]  \hspace{1cm} (10)

where \( I \) is the guided image, \( S \) is the input image, \( q \) is the output image, \( i \) and \( j \) are the pixel indexes, and \( W_{ij} \) is the filter kernel function, which is defined as:

\[ W_{ij}(I) = \frac{1}{|\omega|^2} \sum_{(i,j) \in \omega} \left( 1 + \frac{(I_i - \mu_{\omega})(I_j - \mu_{\omega})}{\sigma_{\omega}^2 + \epsilon} \right) \]  \hspace{1cm} (11)
where $\omega_k$ is the k-th kernel function window, $|\omega|$ is the number of pixels in the window, $\mu_j$ and $\sigma_j^2$ are the mean and variance of the guided filtering, respectively, and $\epsilon$ is a regularization parameter usually with a value between 0 and 1 to prevent the denominator from being too small.

In order to maintain the boundary information in the output image, we use the dark channel as the guided image, and the result of the RF as the input image, because the dark channel obtained by selecting the minimum value among the RGB channels contains the abundant boundary information of the cloud and the ground. Finally, the output image is calculated through the guided filtering (Zhang & Xiao, 2014). In the implementation of filtering, the window radius is set to 30 and $\epsilon$ is set to 0.09.

The output image after the guide filter is not a binary image, so a threshold needs to be set to obtain the binary cloud detection result. Here, the threshold value is set to 80.

The guided filtering effectively removes most of the misjudgment areas, and adds some semi-transparent cloud pixels at the edge of the cloud, which improves the detection accuracy to a certain extent. In addition, although there are ground pixels in some small cloud gaps, they usually do not have a complete scene. These gaps can also be included in the cloud range through the guided filtering.

**Results and analyses**

This section mainly evaluates the cloud detection algorithm from three aspects of the effectiveness, the accuracy and the extendibility. The test images used in the experiment are independent of the training samples.

**Validity evaluation**

We randomly selected 107 L8 satellite images at different times and locations to verify the performance of our RFCD. The imaging time of the images covers four seasons of the year, and the imaging area covers a variety of ground surface types, such as barren, water, wetland, forest, grass, snow, etc. In this paper, the cloud cover value marked in the L8 data is used as the reference, and the cloud cover value obtained from our experimental result is compared with the reference value. We reference the standard of effectiveness evaluation in Landsat7 cloud detection (Irish, 2000; Y Li et al., 2001) and use the difference between our cloud cover value and the reference value to measure the effect of detection. If the difference value is less than 10%, the evaluation is considered accurate. If the difference value is less than 20%, the evaluation is considered valid. If the difference value is greater than 20%, it is considered as invalid. In Table 1, the cloud cover difference is divided into five grades: <5%, 5%-10%, 10%-15%, 15%-20%, >20%, and we count the number of images in these five grades, respectively.

From Table 1, for the test images, the calculated total validity is 93.46% which meets the requirement in the real remote sensing application. The analysis of the detection effect shows that the proposed RFCD method is suitable for most types of ground surface. It can detect most of thick clouds, thin clouds, scattered clouds, and can also effectively eliminate some confusing water or some snow with complex textures. However, when there are bright features with weak texture on the ground, such as the fresh snow, the bright buildings, etc., the detection results of this method may be biased. The main reason is that...
a series of the spectral and texture features extracted from the RGB and NIR bands cannot clearly distinguish clouds from such type of surface.

**Quantitative evaluation**

The remote sensing image contains a large area, and the overall cloud cover value cannot reflect the detection effect in detail. Therefore, we further evaluate the performance of the RFCD pixel by pixel. We use the precision ratio (PR), the recall ratio (RR), the accuracy ratio (AR) and the Kappa coefficient as measurement indicators to quantitatively evaluate the cloud detection accuracy on different types of ground surface. The calculation formulas are as follows.

\[
AR = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
PR = \frac{TP}{TP + FP}
\]

\[
RR = \frac{TP}{TP + FN}
\]

\[
P_e = \frac{(TP + FN)(TP + FP) + (TN + FN)(TN + FP)}{N^2}
\]

\[
Kappa = \frac{AR - P_e}{1 - P_e}
\]

where \(TP\) (true positive) and \(TN\) (true negative) denote the total number of cloud pixels and non-cloud pixels correctly predicted, respectively, \(FP\) (false positive) and \(FN\) (false negative) denote the total number of pixels with an incorrect outcome from the cloud and non-cloud recognition, respectively. \(N\) denotes the total number of pixels. \(P_e\) is the intermediate variable involved in calculating the \(Kappa\).

Seven "L8 Biome" examples of barren, forest, grass, shrubland, urban, water and wetlands are presented in Figures 7–13 for visual demonstration, respectively, which include the original images, the results of Fmask, the results of the proposed RFCD and the manually generated cloud masks from "L8 Biome" dataset. The cloud cover assessment system of L8 product uses the Fmask algorithm (Zhu et al., 2015; Zhu & Woodcock, 2012) to identify cloud confidence for representation in the quality assessment (QA) band as bit-mapped values. We use the “Landsat Quality Assessment Tools” provided by USGS to obtain the result of Fmask generated by pixels with high cloud confidence. The manually generated cloud masks from the "L8 Biome" dataset are used as the ground truth to verify these two methods. In order to better compare the cloud detection results of different methods with visual interpretations, the zoomed-in images taken from the full-scene images are also displayed. On the whole, the spatial distribution of our RFCD results is closer to the manually generated cloud masks than that of Fmask method. Moreover, in some case, the Fmask misjudges some ground objects as clouds.

In Figure 7, the Fmask misjudges non-cloud pixels at the cloud boundary as cloud pixels. The RFCD effectively avoids this situation and achieves a more accurate result. Figure 8 shows that in the case of high vegetation coverage, both methods can achieve good results. It can be seen from Figure 9 that the RFCD is better than the Fmask for detecting some low-brightness transparent thin clouds. But there are still some thin clouds that have not been detected. For the red arrow in Figure 10 pointing to the piece of snow, neither method correctly detects. Both methods do not correctly distinguish between cloud and snow. However, the cloud and snow separation has always been a difficult point in the cloud detection, especially in the RGB and NIR bands. In Figure 11, the result of the RFCD is significantly better than the Fmask. For the coastline area pointed by the red arrow in Figure 12, the Fmask misjudges it, but the proposed RFCD algorithm effectively avoids this situation. In Figure 13, the Fmask misjudges some non-cloud pixels as cloud pixels, and the detection result of the RFCD is relatively better.

Quantitative evaluations of the cloud detection in “L8 Biome” dataset are presented in Table 2. For the proposed RFCD algorithm, the average PR, RR, AR, and Kappa coefficient of cloud detection are 95.04%, 90.26%, 96.22%, and 0.8962, respectively, which are all at a high level. The evaluation indexes of Fmask are 85.64%, 86.58%, 91.23% and 0.7823, respectively. No matter which index is compared, our RFCD is better than the Fmask algorithm. But there are still some deficiencies. For example, in the detection results of the grass scene, the RR of the RFCD is 73.15% which is low but still higher than that of the Fmask. This is mainly because some low-brightness light-transmitting thin clouds are omitted. The RR of the RFCD in the water scene is 79.30%, which is also low. The main reason is that the transparent thin cloud above the water contains part of the water information, resulting in omissions.

**Comparison with Deep Learning Method**

We perform another comparison with the SegNet deep learning method (Chai et al., 2019) using the same validation sources of “L8 Biome” reference masks. Note that the training data of the two methods are not totally the same (the RFCD uses pixels as training data, while the SegNet uses the entire images), which could make the accuracy comparison not particularly fair.
The size of the input image in the SegNet deep learning method is 512 × 512. For the sake of comparison, we crop the results of the RFCD to the corresponding size. Figure 14 shows the comparison between the RFCD results and the SegNet results. Similar to Section 3.2, we use the manually generated cloud masks from the “L8 Biome” dataset as the verification data. Table 3 provides the quantitative comparison of those results. The accuracy of the SegNet is slightly higher than that of our RFCD. The main reason is that the edges in the deep learning results are closer to the ground truth. The RFCD is a pixel-level detection, which requires guided filter to remove isolated noise points. This post-processing causes part of the cloud edge information to be lost in the detection result, and causes cloud edges to be relatively smoother.
Overall, although our RFCD is slightly lower in accuracy than the SegNet, it can still accurately detect most cloud areas. Compared with our RFCD, the SegNet requires a large number of training samples which contain about 2400 images with the size of $512 \times 512$. While the training set of the RFCD only contains about 600,000 pixels which are easier to obtain. In addition, the SegNet needs a long training time and complex parameters to iteratively optimize the deep learning model. But the training of the RFCD only includes the process of building decision trees which needs less time. Finally, the SegNet requires the support of high-performance GPU. However, the RFCD has a small amount of calculation, the training and testing process can also be completed on the ordinary CPU.
Extended experiment

The method in this paper only uses the RGB and NIR bands to realize the cloud detection, and has little dependence on the spectral information. Moreover, RF has strong generalization ability. So, compared with general methods, our RFCD has better extensibility. In order to verify the extendibility, we use the RFCD to perform cloud detection in the Sentinel-2 images and the GF-1 images. Due to the lack of true cloud distribution maps, we only analysis and evaluate the results from the intuitive visual effect which is also important.

Sentinel-2 cloud detection

The Sentinel-2 multispectral imager (MSI) has 13 channels, of which the RGB and NIR bands are bands 2/3/4/8, and the spatial resolution is 10 m. The
coverage of the Sentinel-2 image is too large, so we intercept the sub-image with a size of 2000 × 2000 for cloud detection.

Figure 15 illustrates four typical examples of the standard-false-color composite images and the RFCD results for Sentinel-2 imagery on different surface types. From Figure 15, it can be seen that the RFCD can detect clouds above the ground with different vegetation coverage. Moreover, in Figure 15(d) and Figure 15(f), the thin clouds that are easily confused with the ground surface are also correctly detected. Figure 15(h) shows that the RFCD can also achieve good detection results above water. Even some low-brightness thin clouds on the water surface are correctly detected.

**GF-1 cloud detection**

GF-1 has two high-resolution PMS cameras and four medium-resolution WFV cameras. In this article, we choose the PMS images with a spatial resolution of 8 m for experiments.

Figure 16 shows the standard-false-color composite images of GF-1 and the corresponding cloud detection results. From Figure 16, it can be seen that the overall visual effect of the cloud detection results obtained by the RFCD are good. Figure 16(b) shows that the RFCD can accurately detect thick clouds when the vegetation coverage is high. Figure 16(d) shows that the RFCD can distinguish between clouds and bright barrens. In Figure 16(f) with relative low vegetation coverage and some thin clouds, the RFCD can also achieve good detection results. But in the left part of the image, some of the bright buildings are misjudged as clouds. In Figure 16(h), some snow is still misjudged as clouds. In addition, the overall performance of the RFCD in GF-1 data is slightly worse than Sentinel-2. Maybe this is mainly because the reflectivity at the top of the atmosphere retrieved from Sentinel-2 is closer to that form L8.

**Summary and Conclusions**

There are many cloud detection methods for optical remote sensing images. However, traditional methods usually rely on a large amount of band information and complex threshold determination, which makes these methods difficult to apply to different sensor data. In this paper, we propose a cloud detection method based on the RF for optical remote sensing images which is named RFCD. The RFCD only uses the TOA reflectance of the most common RGB and NIR bands to reduce the dependence on band information and to lay the foundation for the extendibility of this method. The generalization of RF also improves the extendibility to a certain extent. The introduction of the machine learning avoids the difficulty of manually determining thresholds by a large number of statistics in the traditional threshold method.

The effectiveness and quantitative evaluations of a variety of surface types and cloud types show that the proposed RFCD has a good detection effect in most
cases. Especially when identifying broken clouds and thin clouds with low brightness, the detection effect has been greatly improved. It can also distinguish brighter surface more accurately such as coastlines. In the validity evaluation results, the total validity of the RFCD in 107 L8 images is 93.46% which satisfies the requirement in the application. The quantitative evaluation results show that the RFCD can obtain a high degree of detection accuracy. Compared with the Fmask, the RFCD can detect clouds more accurately with the PR of 95.04%, the RR of 90.26%, the AR of 96.22% and the Kappa coefficient of 0.8942. Compared with the deep learning method SegNet, the accuracy of the RFCD is comparable to the SegNet. However, the proposed RFCD requires fewer samples, simpler parameters, less training time and it is less dependent on computing device performance and has certain interpretability. Finally, the extended experiments in the Sentinel-2 and the GF-1 images verify that the RFCD has strong extendibility in different satellite images.

Although the proposed RFCD can achieve good results in the cloud detection, there are still some deficiencies to be further explored. When the local surface is covered with a large area of snow with weak texture information, it is difficult for the RFCD to accurately distinguish the clouds. This is mainly
because the current features extracted from the RGB and NIR bands cannot accurately distinguish clouds from such objects. In the next step, we will improve the detection accuracy by extracting and selecting more suitable features. In addition, we will use more satellite images for cloud detection to further verify the extendibility of this method.

Data availability statement

The "L8 Biome" data that support the findings of this study are openly available in Landsat 8 Cloud Cover Assessment Validation Data at http://doi.org/10.5066/F7251GDH, reference number 19.

The other data used to support the findings of this study are available from the corresponding author upon request.
Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 61771470; in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA19010401 and XDA19060103; in part by the Key Research Program of Frontier Sciences, Chinese Academy of Sciences under Grant QYZDY-SSW DQC026.<#AWARD-ID;>

References

Azimisadjadi, M. R., Shaikh, M. A., Tian, Bin, E., K. C. Reinken, D. “Neural network-based cloud detection classification using textural and spectral features.” 1996. IGARSS’96. 1996 International Geoscience and Remote Sensing Symposium. Vol. 2 1105–1107. IEEE.  
Belgui, M., & Dragut, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114(4), 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Breiman, L. (2001a). Using iterated bagging to debias regressions. Machine Learning, 45(3), 261–277. https://doi.org/10.1023/A:1017934522171

Breiman, L. (2001b). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:101093404324

Calle, M. L., & Ureá, V. (2011). Letter to the editor: Stability of random forest importance measures. Briefings in Bioinformatics, 12(1), 86–89. https://doi.org/10.1093/bib/bbr011

Chai, D., Newsam, S., K. Z. H., Qiu, Y., & Huang, J. (2019). Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. Remote Sensing of Environment, 225, 307–316. https://doi.org/10.1016/j.rse.2019.03.007

Chen, X., Liu, L., Gao, Y., Zhang, X., & Xie, S. (2020). A Novel Classification Extension-Based Cloud Detection Method for Medium-Resolution Optical Images. Remote Sensing, 12(15), 2365. https://doi.org/10.3390/rs12152365

Chen, Y., Tang, L., Kan, Z., Latif, A., Yang, X., Bilal, M., & Li, Q. (2020). Cloud and cloud shadow detection based on multiscale 3D-CNN for high resolution multispectral imagery. IEEE Access, 8, 16505–16516. https://doi.org/10.1109/ACCESS.2020.2967590

Dong, Zhipeng, Wang, Mi, Li, Deren, Wang, Yanli, Zhang, Zhiqi, et al. (2018). A Cloud Detection Method for High Resolution Remote Sensing Imagery Based on the Spectrum and Texture of Objects. Acta Geodaetica et Cartographica Sinica, 47(7), 996–1006. https://doi.org/10.101197/j.AGCs.2018.20170690

Fang, K. N., Wu, J. B., Zhu, J. P., & Xie, B. C. (2011). A Review of Technologies on Random Forests. Statistics & Information Forum, 26(3), 32–38. https://doi.org/10.3969/j.issn.1017-3116.2013.03.006

Fu, H., Shen, Y., Liu, J., He, G., Chen, J., Liu, P., Qian, J., & Li, J. (2019). Cloud detection for FY meteorology satellite based on ensemble thresholds and random forests approach. Remote Sensing, 11(1), 44. https://doi.org/10.3390/rs11010044

H. J. J., H. J. R., Inceoglu, F., & Tofftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep learning. Remote Sensing of Environment, 229, 247–259. https://doi.org/10.1016/j.rse.2019.03.039

Hall, P., Gill, N., Kurka, M., Phan, W. “Machine learning interpretability with h2o driverless ai” 2017. H2O.ai, Inc.

He, K., Sun, J., & Tang, X. (2010). Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(12), 2341–2353. https://doi.org/10.1109/TPAMI.2010.168

He, K., Sun, J., & Tang, X. (2013). Guided image filtering. IEEE Transactions on Pattern Analysis & Machine Intelligence, 35(6), 1397–1409. https://doi.org/10.1109/TPAMI.2012.213

Irish, R. R. “Landsat 7 automatic cloud cover assessment.” 2000. Proceedings of SPIE-The International Society for Optical Engineering. Orlando, FL, United States 4049 (: AeroSense 2000):348–355. https://doi.org/10.1117/12.410358.

Irish, R. R., Barker, J. L., Goward, S. N., & Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. Photogrammetric Engineering & Remote Sensing, 72(10), 1179–1188. https://doi.org/10.14358/PERS.72.10.1179

Jang, J., Viau, A. A., Ancilll, F., & Bartholomé, E. (2006). Neural network application for cloud detection in SPOT VEGETATION images. International Journal of Remote Sensing, 27(4), 719–736. https://doi.org/10.1080/01431160500106892

Joshi, P. P., Wynne, R. H., & Thomas, V. A. (2019). Thomas V A. Cloud detection algorithm using SVM with SWIR2 and tasseled cap applied to Landsat 8. International Journal of Applied Earth Observation and Geoinformation, 82, 101898. https://doi.org/10.1016/j.jag.2019.101898

Lee, Tae Sing, & . (1996). Image representation using 2D gabor wavelets. IEEE Transactions on Pattern Analysis & Machine Intelligence 18 10 959–971 . https://doi.org/10.1109/16.541406

Li, Y., Ye, X., & Feng, Z. (2001). Landsat7 automatic cloud cover assessment. Remote Sensing Information, 1, 39–41.

Li, Z. W., Shen, H., Cheng, Q., Liu, Y., You, S., & He, Z. (2019). Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. ISPRS J. Photogramm. Remote Sens, 150, 197–212. https://doi.org/10.1016/j.isprsjprs.2019.02.017

Ma, N., Sun, L., Wang, Q., Yu, Z., & Liu, S. (2020). Improved cloud detection for Landsat 8 images using a combined neural network model. Remote Sensing Letters, 11(3), 274–282. https://doi.org/10.1080/2150704X.2019.1708503

Segal-Roehaimer, M., Li, A., Das, K., & Chirayath, V. (2020). Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (CNN). Remote Sensing of Environment, 237, 111446. https://doi.org/10.1016/j.rse.2019.111446

Shan, N., Zheng, T., & Wang, Z. (2009). High-speed and high-accuracy algorithm for cloud detection and its application. Journal of Remote Sensing, 13(6), 1138–1146. https://doi.org/10.3321/j.issn.1007-4619.2009.06.012

Skakun, S., F. V., E., C. R. J., Justice, C. O., & Masek, J. G. (2019). Validation of the LaSRC cloud detection algorithm for Landsat 8 images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7), 2439–2446. https://doi.org/10.1109/JSTARS.2019.2894553
Tan, Kai, Zhang, Yongjun, Tong, Xin, Kang, Yifei. (2016). Automatic Cloud Detection for Chinese High Resolution Remote Sensing Satellite Imagery. Acta Geodaetica et Cartographica Sinica, 45(5), 581. https://doi.org/10.11947/j.AGCS.2016.20150500

U.S. Geological Survey. (2016). L8 Biome Cloud Validation Masks. U.S. Geological Survey. Data Release. https://doi.org/10.5066/F7251GDH.

Wei, J., Huang, W., Li, Z., Sun, L., Zhu, X., Yuan, Q., Liu, L., & Cribb, M. (2020). Cloud detection for Landsat imagery by combining the random forest and superpixels extracted via energy-driven sampling segmentation approaches. Remote Sensing of Environment, 248, 112005. https://doi.org/10.1016/j.rse.2020.112005

Xu, K., Guan, K., Peng, J., Luo, Y., Wang, S. “DeepMask: An algorithm for cloud and cloud shadow detection in optical satellite remote sensing images using deep residual network 1–17.” 2019. arXiv preprint arXiv:1911.03607.

Xu, Y. X., Zhu, Z. H., & Liu, Y. Y. (2006). Developing Technique for the Detection and Removal of Cloud and Haze in Satellite Images. Remote Sensing for Land & Resources 3 23–28 . 3. https://doi.org/10.3969/j.issn.1001-070X.2006.03.006

Zhang, Q., & Xiao, C. (2014). Cloud detection of RGB color aerial photographs by progressive refinement scheme. IEEE Transactions on Geoscience and Remote Sensing, 52 (11), 7264–7275. https://doi.org/10.1109/TGRS.2014.2310240

Zhang, W. D., He, M. X., & Mak, M. W. “Cloud detection using probabilistic neural networks.” 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217) Sydney, NSW, Australia. IEEE, 5: 2373–2375. https://doi.org/10.1109/IGARSS.2001.978006

Zhang, Y., Rossow, W. B., Lacis, A. A. et al. (2004). Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. Journal of Geophysical Research: Atmospheres, 109(D19 1–27). https://doi.org/10.1029/2003JD004457

Zhong, R., Chen, W., Wu, S., Hu, L., Luo, X., & Liu, Q. (2017). A cloud detection method based on relationship between objects of cloud and cloud-shadow for Chinese moderate to high resolution satellite imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(11), 4898–4908. https://doi.org/10.1109/JSTARS.2017.2734912

Zhou, T. N., Ming, D. P., & Zhao, R. (2017). Land cover classification based on algorithm of parameter optimization random forests. Science of Surveying and Mapping, 42 02 , 92–98. https://doi.org/10.16251/j.cnki.1009-2307.2017.02.017.

Zhu, L. “A Cloud Detection Algorithm Based on the Dark Channel and Image Semantics.” 2018. IEEE International Conference on Information and Automation (ICIA) (Wuyishan, China: IEEE) (pp. 1350–1354). https://doi.org/10.1109/ICInfA.2018.8812350.

Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4-7, 8, and sentinel 2 images. Remote Sensing of Environment, 159, 269–277 . https://doi.org/10.1016/j.rse.2014.12.014

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in landsat imagery. Remote Sensing of Environment, 118, 83–94 . https://doi.org/10.1016/j.rse.2011.10.028