

A New Cloud Detection Method Supported by GlobeLand30 Data Set

Lin Sun, Xueying Zhou[✉], Jing Wei[✉], Quan Wang[✉], Xinyan Liu, Meiyuan Shu, Tingting Chen, Yulei Chi, and Wenhua Zhang

Abstract—In terms of traditional threshold methods, uniform thresholds are used for cloud detection based on remote sensing images; however, due to complex surface structures and cloud conditions, such an approach is typically difficult to effectively implement for high-precision cloud detection. To solve this problem, a new cloud detection algorithm is proposed based on global land cover data. Specifically, a high spatial-resolution at 30-m Global Land Cover Data set with global coverage was employed as background data for image inversions, which further supported cloud detection in remote sensing images. Notably, threshold settings can be varied for different land cover types. Such an algorithm can effectively improve the accuracy of cloud pixel identification for thin and broken clouds, even over bright areas. Moreover, Landsat 5 data are used to perform cloud detection experiments based on this algorithm. The thresholds are considering land cover variations. The thresholds of land cover types spatiotemporally vary, such as vegetation, differed by latitude and over time. In addition, six common land cover types are selected for cloud detection experiments. Then, validations analyses are conducted through visual interpretation and the results indicated that the algorithm is capable of achieving a high cloud detection accuracy. Specifically, the overall RMSE of cloud cover is 4.44%, and the accuracies of cloud and clear-sky pixel identifications are 86.5% and 98.7%, respectively.

Index Terms—Cloud detection, GlobeLand30, land cover product, Landsat, threshold method.

I. INTRODUCTION

AS AN important climate change factor, clouds alter the energy transfer process of solar radiation and affect the radiation budget between the surface of Earth and the atmosphere. Therefore, clouds play a critical role in the radiation

Manuscript received October 9, 2017; revised April 23, 2018; accepted July 17, 2018. Date of publication September 19, 2018; date of current version October 15, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 41771408, by the Natural Science Foundation of Shandong Province under Grant ZR201702210379, and by the Graduate Innovation Fund of Shandong University of Science and Technology under Grant SDKDYC170209. (*Corresponding author: Xueying Zhou*.)

L. Sun, Q. Wang, X. Liu, M. Shu, T. Chen, and Y. Chi are with the Geomatics College, Shandong University of Science and Technology, Qingdao 266510, China (e-mail: sunlin6@126.com; wangquan_rs@hotmail.com; mmxinyan@163.com; 2448858578@qq.com; ctt_rs@163.com; cylilly@sina.com).

X. Zhou is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China (e-mail: zhouxueying666@hotmail.com).

J. Wei is with the College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China (e-mail: weijing_rs@163.com).

W. Zhang is with the College of Geography and Tourism, Qufu Normal University, Rizhao 273165, China (e-mail: zhangwenhua0426@163.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2018.2861755

energy balance [1], [2]. Although satellite remote sensing technology is currently a crucial means of Earth monitoring, clouds can blur remote sensing images and even block ground information, which leads to incorrect information expression and the accurate inversion of atmospheric and surface parameters [3]–[6]. Therefore, cloud detection is essential for remote sensing image processing and is significant for the improvement of both the quality and utilization ratio of remote sensing data.

Among the numerous cloud detection methods, the threshold method, statistical methods (Markov random fields, support vector machine, etc.) [7]–[11], and artificial neural network methods [12]–[14] are the major detection method categories. In detail, the threshold method has been extensively applied to the preprocessing of various remote sensing images due to its simple algorithm, user-friendly control, and high accuracy; in most cases, this method is used for vocational operations to produce cloud products. Saunders and Hunt used single channel thresholds for cloud detection without any auxiliary data at the beginning, which was referred to as a “gross cloud check” [15]. Multichannel thresholds were also applied for early cloud detections [16]–[18]. Rossow and Rossow [19], [20] developed the International Satellite Cloud Climatology Project (ISCCP) algorithm in 1983. It is assumed that radiation only originates from two atmospheric conditions: clear sky and clouds. The main thoughts regarding this algorithm are that clear-sky images are less variable than those with cloud cover and that clear skies are also darker and hotter in the visible and infrared bands. Moreover, when the radiation value is significantly greater than the variations in a clear-sky pixel, the pixel is identified as a cloud pixel. Developed by Saunders and Kriebel [21], the advanced very high resolution radiometer (AVHRR) processing scheme over cloud land and ocean algorithm used five tests formed by bands one to five of the AVHRR to perform cloud detection, and when all test objects are clear-sky pixels, the corresponding pixel is determined to be a clear-sky pixel; otherwise, the pixel is cloud pixel. Presently, the moderate resolution imaging spectroradiometer (MODIS) cloud mask is one of the commonly used cloud models. The relevant algorithm primarily utilizes multiple bands, and the mutual operations of the bands are used to conduct a series of threshold tests, which are divided into five groups. The test results of the diverse confidence values are combined in a particular way to form the final cloud mask product [22], [23]. Vittorio and Emery [24] adopted AVHRR data to automatically generate a threshold algorithm known as dynamic threshold cloud-masking (DTCM) and these data include

the reflectance at the central wavelength is $0.68 \mu\text{m}$, brightness temperature difference between the third and fourth bands (central wavelengths are 3.83 and $10.50 \mu\text{m}$, respectively), and brightness temperatures of the fourth and fifth bands (central wavelength is $12.20 \mu\text{m}$). Moreover, DTCM can be used for cloud detection over land. To improve the MODIS cloud mask algorithm for polar nights, Liu *et al.* [25] introduced a new testing method based on the physical foundation of testing and radiation transfer simulations. Zhuge *et al.* proposed a visible cloud index based on the reflectance of three bands at central wavelengths of 0.46 , 0.640 , and $0.86 \mu\text{m}$ for the advanced Himawari imager [26]. With improvements in remote sensing satellite sensors and cloud detection technology, cloud detection in higher spatial resolution images was developed. The Fmask [27] algorithm utilizes the physical characteristics of clouds and a series of threshold tests to extract potential clouds and further isolate clouds, cloud shadows, and snow. As a cloud detection algorithm, the universal dynamic threshold cloud detection algorithm (UDTCDA), which was proposed by Sun *et al.* [28] supported by the land surface reflectance database, uses the 6S model to simulate atmospheric impacts and acquire the associated dynamic thresholds. Li *et al.* [29] proposed an automatic multifeature combined method for cloud and cloud shadow detection in GaoFen-1 wide field of view imagery using spectral features, guided filtering, and geometric features to produce the cloud and cloud shadow masks.

However, although the traditional threshold methods of cloud detection are based on simple principles and can be easily implemented, some problems remain with these methods. For example, the threshold method is utilized to identify cloud pixels according to spectral characteristic differences between surface features and clouds. However, surface conditions are complicated, and spectral characteristic differences exist for different types of surface features and clouds or between dramatically different bands. For example, the reflectance of thin clouds above water bodies can be much lower than that of bare land. Therefore, an identical threshold for the same band is inapplicable to an entire remote sensing image with complex surface features. Although computer-assisted unsupervised classification can be performed with the DTCM algorithm, the corresponding classification accuracy is limited due to the complexity and the relevant classification procedure can be very sophisticated if a high classification accuracy is obtained. In recent years, cloud detection studies on Landsat series satellite images have been conducted continuously using the Fmask cloud detection algorithm [26], [27], UDTCD [28], and multitemporal cloud detection [30]. Although high levels of accuracy have been obtained by these algorithms, the associated process can be very complicated. For example, dynamic thresholds such as UDTCD are used to eliminate errors incurred by the Earth surface complexity and gain a high accuracy, and a minimum value composite, clipping, mosaic should be performed, which is targeted at MODIS surface reflectance products, to establish a surface reflectance database for different bands. Considering a heavy preparation workload in early stages, the database is applicable to long-term cloud detection batch processing and product manufacturing. Nevertheless, for processing a few or

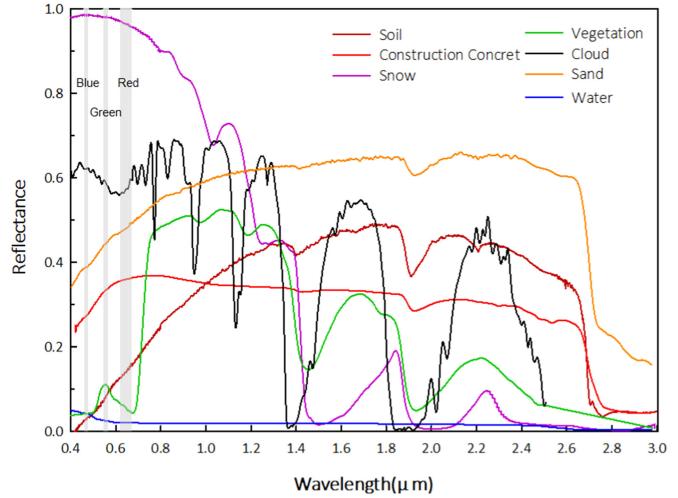


Fig. 1. Spectral curves of typical surface features.

TABLE I
MAJOR PARAMETERS OF THE LANDSAT-5 TM SENSOR

Number	Band	Wavelength range (μm)	Central wavelength (μm)	Spatial resolution (m)
1	Blue	0.45-0.52	0.50	30
2	Green	0.52-0.60	0.59	30
3	Red	0.63-0.69	0.68	30
4	Near Infrared	0.76-0.90	0.80	30
5	SWIR1	1.55-1.75	1.71	30
6	Thermal Infrared	10.40-12.50	11.49	120
7	SWIR2	2.08-2.35	2.20	30

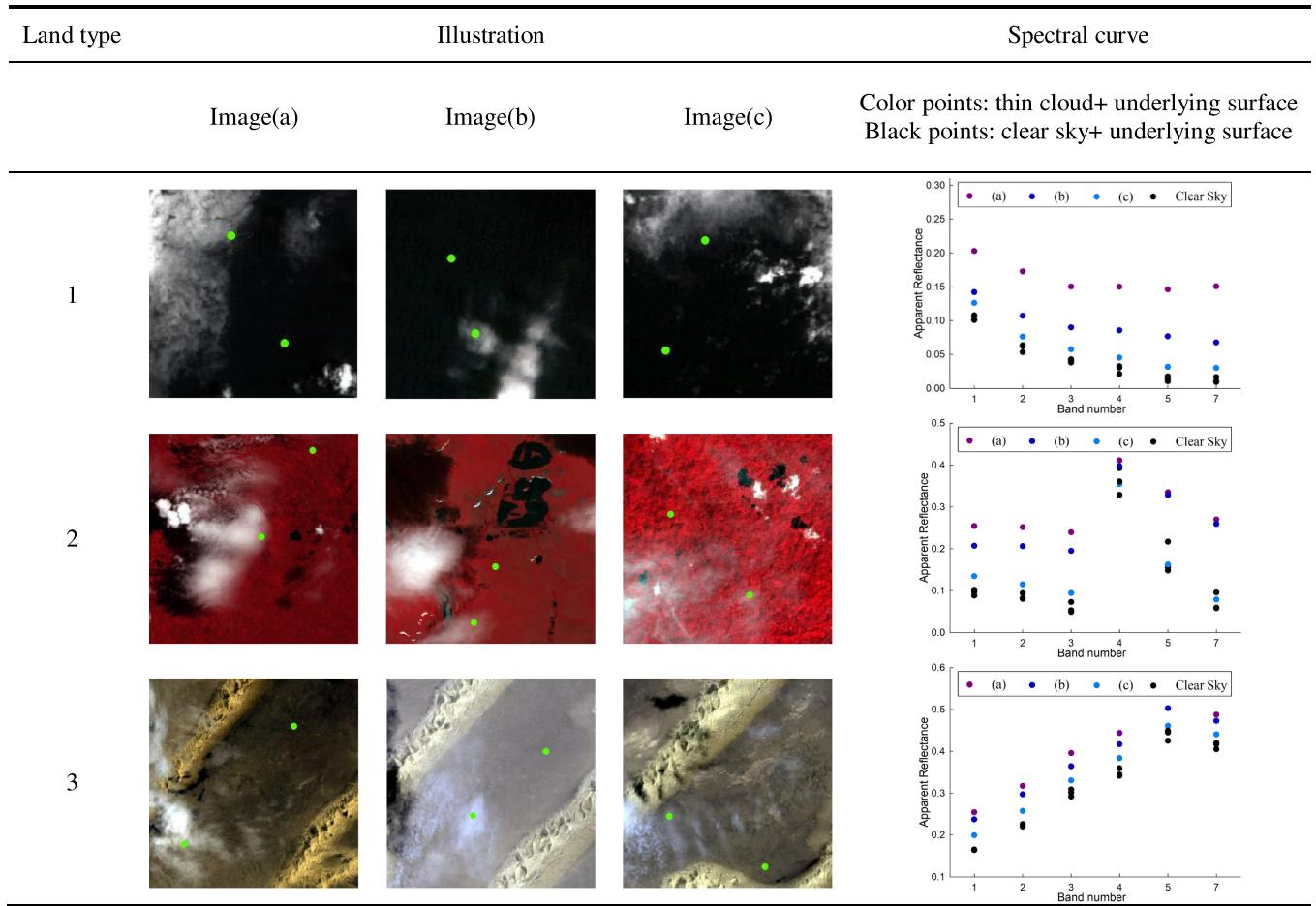
dinary remote sensing images, cloud detection only serves as a preprocessing step. Thus, such an algorithm is relatively sophisticated.

Hence, a threshold-based cloud detection algorithm, which is supported by the 30-m global land cover type product, is presented in this study to improve the universality of fixed threshold values for a complex surface of Earth, as well as more effectively utilize relatively accurate surface classification and realize convenient and rapid cloud detection. This algorithm is called land-cover-based cloud detection (LCCD). Moreover, accuracy validation was performed for the cloud detection results of diverse underlying surfaces.

II. PRINCIPLES

Cloud identification is conducted according to spectral differences between the clouds and typical surface features or brightness temperature differences [31]–[33]. Fig. 1 shows the spectral curves of various, typical surface features. Among the features, the spectral characteristics of surface features are obtained from the United States Geological Survey (USGS) spectral library [34], and the cloud features are acquired from a cloud pixel library established with airborne visible infrared imaging spectrometer hyperspectral data. In addition, the main parameters of Landsat 5 are listed in Table I. In the visible near-infrared

TABLE II
APPARENT REFLECTANCE OF THIN CLOUDS ABOVE TYPICAL SURFACE FEATURES



1: water, 2: vegetation, 3: bareland

band, except for snow, the reflectance difference between typical surface features and clouds is very high. As a result, the threshold method can be adopted to achieve a favorable detection effect for thick clouds, especially those over a dark surface. However, thin cloud detection is of the same importance, because thin clouds combine with the underlying surface to form mixed pixels.

Different typical Earth surfaces are selected in this study to obtain the statistics of the apparent reflectance of thin clouds. The corresponding statistical results are given in Table II, in which the reflectance of thin clouds above water bodies, vegetation, and bare land are presented. Notably, the thin clouds in the selected images (a), (b), and (c) tend to become increasingly thinner. Dependent on three categories of surface feature spectral curves, the underlying surface has a dramatic influence on the reflectance of thin clouds. The apparent reflectance of thin clouds is typically consistent with the spectral characteristics of the underlying surface. For an identical surface feature, the thinner the cloud is, the closer the apparent reflectance will be to the spectral surface feature. In the figure, the apparent reflectance of thin clouds above water bodies ranges approximately between 0 and 0.2 in six bands. By contrast, in terms of bare land, the surface

feature reflectance is higher than 0.2. In this case, if clouds over bare land must be detected, the corresponding threshold should be set to 0.2 and above; otherwise, misjudgment may occur for bright surfaces, and thin clouds with an apparent reflectance less than the threshold above water bodies (or even vegetation) could result in a false negative. Likewise, the visible band is optimal for the detection of cloud pixels above vegetation. The red band (Band 3) can be used as an example. Since the reflectance of bare land itself is close to 0.3, which is higher than the apparent reflectance of thin clouds above vegetation (<0.25), preferable cloud detection results cannot be obtained for both in the red band. In addition, cloud pixel identification above water bodies achieves a good effect in the near-infrared (Band 4) and short-wave infrared (Bands 5 and 7) bands. However, vegetation and bare land also have high reflectance values in Band 4 and Band 5. Consequently, the corresponding threshold should be relatively lower to ideally detect cloud pixels above water bodies, which may result in bare land being misjudged as clouds. Such a phenomenon signifies that the traditional threshold method cannot comprehensively consider the complex spectral characteristics of the Earth's surface. Additionally, land cover type data must be used to define a reasonable threshold range for cloud detection.

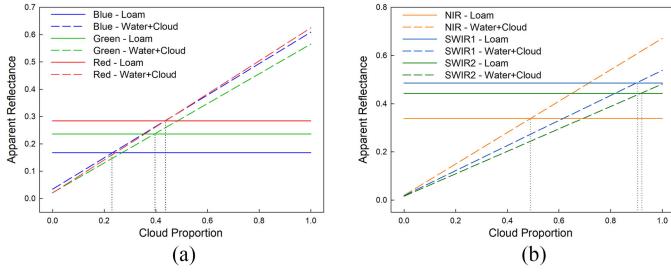


Fig. 2. Apparent reflectance variation simulation of thin clouds over water bodies and bareland.

Taking water bodies and bare land as examples, the mixed pixel decomposition principle [35] (refraction of thin clouds is equivalent to the proportion occupied by pixels) is utilized to perform simulated calculations for the apparent reflectance of thin clouds above water bodies. According to the spectral curves, we select the reflectance values of clouds, water, and reddish brown, sandy loam for the simulation. The corresponding determination and calculation results are given in Fig. 2, in which (a) and (b) represent the outcomes of the visible and infrared bands. In the figure, the solid line refers to the apparent reflectance of bare land at a certain band, and the apparent reflectance variation tendencies of mixed pixels, which correspond to the water bodies along with the increase in cloud cover, are denoted by dashed lines. When the proportion occupied by the clouds above water bodies in the pixels is 0, this is the apparent reflectance of the water bodies. As this proportion rises, the apparent reflectance tends to change linearly until it reaches 1 (pure cloud pixel), and thus, the apparent reflectance of the cloud appears. In Fig. 2(a), when the proportion taken by the cloud is 0.23 (Blue), 0.40 (Green) or 0.44 (Red), the reflectance of mixed pixels formed by water bodies and clouds reaches a value identical to that of bare land. In Fig. 2(b), in the case where the proportion of clouds is 0.49 (NIR), 0.90 (SWIR1) or 0.92 (SWIR2), the reflectance of mixed pixels is also the same as that of bare land. Such simulation indicates that the apparent reflectance of bare land can be higher than that of the mixed pixels combined with thin clouds and water bodies under certain circumstances, in which case, thin clouds over water bodies still cannot be identified if the cloud detection threshold of bare land is adopted.

In summary, the LCCD algorithm uses the global land cover data set as a basis to support the underlying surface types of remote sensing images and to improve both the classification efficiency and accuracy. In addition, a rational threshold is also given according to the classification.

III. 30-M GLOBAL LAND COVER DATA SET

The 30-m global land cover data set (GlobeLand30) is a global land cover product generated from 30-m multispectral images, including those obtained from Landsat TM5, ETM+ multispectral images, and multispectral images from the Chinese Environmental Disaster Alleviation Satellite (HJ-1). This data set is also used in combination with large quantities of auxiliary and reference data (e.g., global ecological and ge-

TABLE III
BRIEF INTRODUCTION TO THE LAND TYPES OF GLOBELAND30

Type	ID	Definition
Cultivated land	10	Land for crop cultivation
Forest	20	Land with more than 30% of arbor cover and tree canopy cover
Grassland	30	Land with more than 10% natural herbosa cover
Shrubland	40	Land with more than 10% shrub cover
Wetland	50	Areas that are inundated or saturated by surface or groundwater at a frequency and duration sufficient
Water bodies	60	Areas covered by liquid water within the scope of land surface
Tundra	70	Land covered by lichen, moss, long hardy herbs and shrub vegetation under frigid climate
Artificial Surfaces	80	Land surface built by artificial construction activities
Bareland	90	Natural land cover with less than 10% of vegetation cover
Permanent snow and ice	100	Land covered by permanent snow, ice and ice sheet
Ocean	255	A body of saline water that composes much of a planet's hydrosphere

ographical division data, global basic geographic information data and global DEM data, etc.). Data are selected for this product based on being cloudless or partly cloudy. Images acquired during the vegetation growing season are preferentially chosen as the classified images. Concerning most of the images, the acquisition time is within the year 2010 ± 1 , and regarding areas where it is difficult to collect images, the time condition of image acquisition is properly less restrictive. GlobeLand30 [36] uses the WGS-84 coordinate system, the UTM projection, and 6 degree zoning. It contains ten surface features, which include cultivated land, forest, grassland, shrubland, wetland, water bodies, tundra, artificial surfaces, bare land, and permanent snow and ice. The assignment and color for each type are shown in Table III.

The GlobeLand30 product was derived using a hierarchical extraction method. Each land cover type was classified individually, with mask constraints for other land cover types. The workflow is to extract only one land cover type at a time, and then, the class is masked after extraction. The classification is conducted for the next land cover type and then masked until all classes are derived. According to a third-party accuracy assessment, the classification accuracy of GlobeLand30 was evaluated using a spatial data two-level sampling scheme. The first level is frame sampling, which uses map tiles as the sampling unit. The second level is feature sampling, which uses a selected pixel in the tile as the sampling unit. In total, 9 types and over 150 000 test samples were evaluated in terms of accuracy. The overall accuracy of GlobalLand 30–2010 can reach 83.51% [36]. The relevant Kappa indicator is 0.78. The accuracies of diverse types are shown in Fig. 4. Thus, this land cover type classification product achieves high accuracy, which provides an ac-

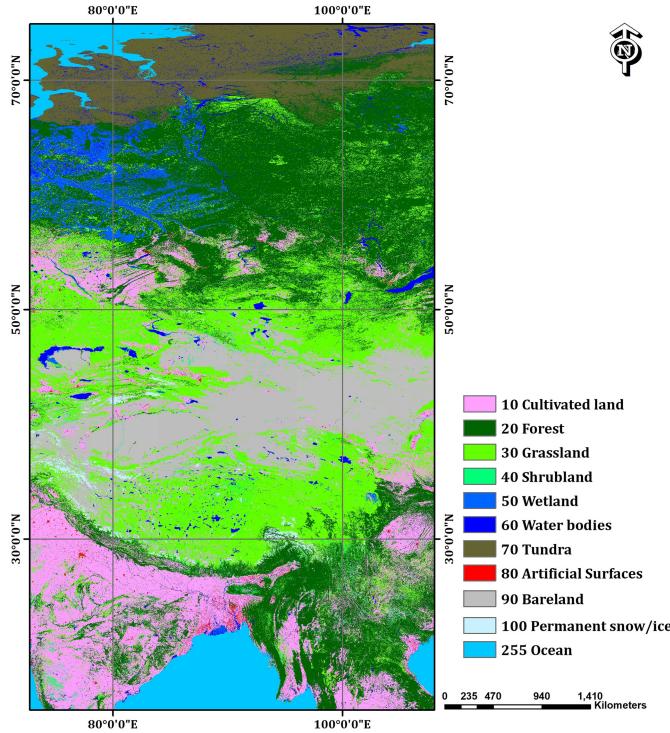


Fig. 3. Introduction of GlobeLand30.

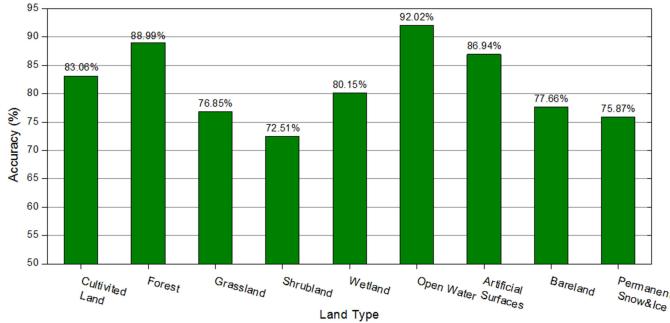


Fig. 4. Accuracies of GlobeLand30 for all land cover types.

curacy guarantee for its application in cloud detection in this study.

IV. METHODOLOGY

Because the Landsat-5 TM data are consistent with the years that correspond to the GlobeLand30 land cover type database temporal distribution, the algorithm was applied to the TM data. Landsat-5 is a satellite in the sun-synchronous orbit, which was launched by NASA on March 1, 1984. The satellite's orbital altitude and revisiting period are 705 km and 16 days, respectively. In addition, the satellite's TM sensor has seven bands, which can cover visible and thermal infrared. Moreover, the LCCD algorithm utilizes 1–7 bands of this sensor.

In this study, surface feature types except areas covered by snow/ice are primarily selected to perform cloud detection based on the land cover type classification product. Six surface features, which include wetland, water bodies, tundra, artificial surface, bare land, and ocean, have reflectance values that slightly

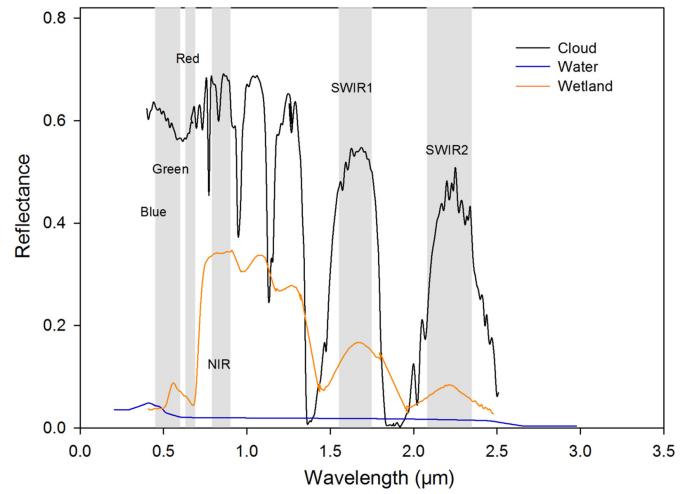


Fig. 5. Spectral reflectance of wetland and water bodies.

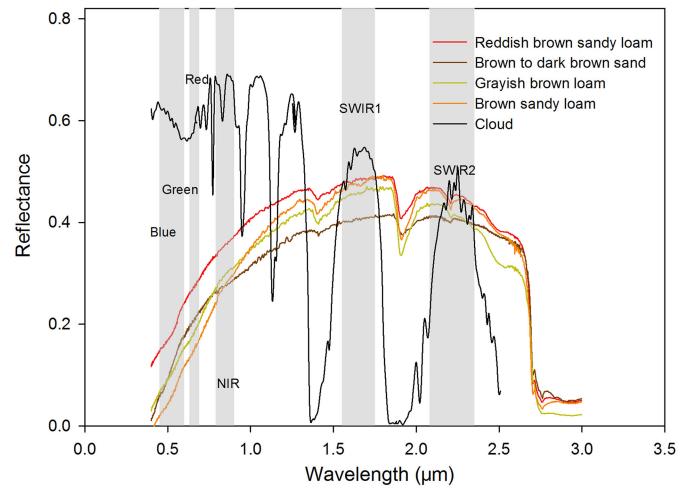


Fig. 6. Spectral reflectance of typical soil.

change as the season varies, and thus, these surface features are deemed constant attribute surfaces. Therefore, cloud detection can be conducted for these features using a fixed threshold value. However, the reflectance values of cultivated land, forest, grassland, and shrubland are dramatically different from each other at different latitudes, in different seasons and for different vegetation types. As Earth's attribute surfaces change, spatial-temporal changes should be comprehensively considered to determine the cloud detection threshold. In this study, the USGS spectral library is employed to statistically analyze the spectral characteristics of the typical surface features of diverse land cover types. Additionally, a reasonable cloud detection threshold is obtained, which is based on the global land cover data.

A. Constant Attribute Surfaces

In Figs. 4–6, typical spectral curves of water bodies, wetland, soil, and urban material are presented. Considering that the overall reflectance of water bodies is relatively low (see Fig. 5) and the reflectance values of the blue and green bands are slightly

high and gradually decreases until close to 0 in the infrared band, the blue, green, red, and near-infrared bands are employed to identify cloud pixels over the ocean. In addition, due to the influence of phytoplankton and sediments in inland rivers and lakes, their reflectance can rise somewhat in the near-infrared band. Considering this, only the blue, green, and red bands are utilized to perform cloud identification for pixels over water bodies. For example, if the reflectance values of underlying surfaces of ocean or water bodies satisfy discriminant (1) or (2), respectively, the corresponding pixel is a cloud pixel; otherwise, the pixel is a clear-sky pixel.

$$\text{Band 1} > 0.13 \cup \text{Band 2} > 0.15 \cup \text{Band 3} > 0.1 \quad (1)$$

$$\text{Band 1} > 0.12 \cup \text{Band 2} > 0.12 \cup \text{Band 3}$$

$$> 0.1 \cap \text{Band 4} > 0.1. \quad (2)$$

Wetlands dominated by aquatic organisms have a moist soil layer, rich vegetation species, and stable ecological systems. The material underlying of a wetland is moist soil or shallow water, the reflectance is very low, and the influence of seasonal variations on the reflectance change in mixed pixels is insignificant. Therefore, the impact of mixed pixels has been disregarded in this study. When the underlying surface is wetland, the bands with a corresponding apparent reflectance that satisfies the following condition are identified as clouds

$$\text{Band 1} > 0.13 \cup \text{Band 2} > 0.15 \cup \text{Band 3} > 0.13. \quad (3)$$

For bare land, brown-red, sandy loam soil, brown to dark brown sandy soil, grayish brown soil, and brown sandy loam soil are selected as the typical bare land types in this study. As revealed by the reflectance spectral curves in Fig. 6, the reflectance of different types of soil increases in the visible band. Between 1.0 and 2.7 μm is a high-reflection region with a slight reflectance difference from clouds. Hence, the blue, green, red, and near-infrared bands are adopted to identify whether pixels over bare land can be classified as cloud pixels. Additionally, the reflectance in the bare land brightness temperature band is quite different from that of clouds, although they both have high reflectance values in the visible band. In detail, the cloud brightness temperature is low, whereas bare land has a high brightness temperature. For this reason, pixels meeting discriminant (4) are identified as cloud pixels. Otherwise, the pixels are clear-sky pixels

$$\begin{aligned} & (\text{Band 1} > 0.15 \cup \text{Band 2} > 0.15 \cup \text{Band 3} \\ & > 0.2) \cap \text{BT}_6 < 298. \end{aligned} \quad (4)$$

In Fig. 7, the spectral curves of typical artificial materials in urban areas are presented. There is a large difference in the reflectance values of different artificial materials; however, they have a consistent rule in general. That is, the spectral difference between artificial materials and clouds is the largest in the visible range. Therefore, the visible band is selected in this study to conduct cloud identification for pixels over artificial surface. The following discriminant has been established according to the low brightness temperature characteristic of clouds, and

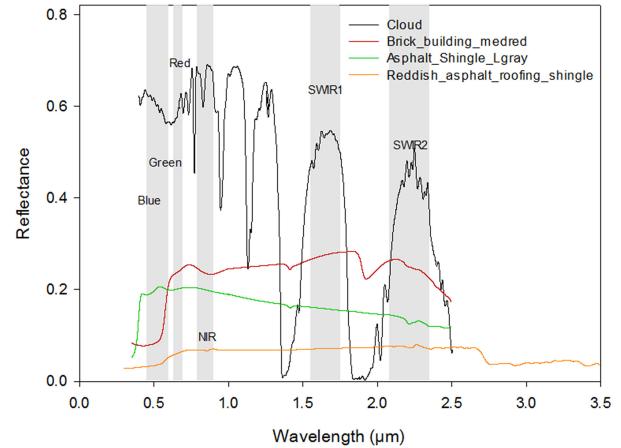


Fig. 7. Spectral reflectance of artificial surface.

those up to conditions of this discriminant can be identified as cloud pixels

$$\begin{aligned} & (\text{Band 1} > 0.2 \cup \text{Band 12} > 0.25 \cup \text{Band 3} > 0.3) \\ & \cap \text{BT}_6 < 296. \end{aligned} \quad (5)$$

B. Changing Attribute Surfaces

For cultivated land, forest, grassland, and shrubland, different seasons and geographical locations may lead to different vegetation growth states and such differences further induce changes in their reflectance values. For example, boreal forest is commonly predominated by coniferous forest, whereas temperate forests are predominated by deciduous forest and broad-leaved forest evergreen during all four seasons, which can be seen in the tropics in most cases. During summer, vegetation has plentiful leaves, which is a vegetation characteristic; however, during winter, most leaf falling phenomena are considered to be a bare land feature. Therefore, when the threshold method is used to identify cloud pixels, the effects of spatiotemporal variations and mixed pixels on reflectance should be considered. In this study, the mixed pixel effects of vegetations across diverse seasons and latitudes can be used to obtain the corresponding cloud detection thresholds. Furthermore, latitudes were divided into the tropical zone (0° – 23.5°), temperate zone (23.5° – 66.5°), and frigid zone (66.5° – 90°).

1) *Cultivated Land:* Actual cultivated land situations are complex. As different crops are planted on different plots, the plant growth cycle fails to exhibit an obvious wholeness law such as forest or grassland. Therefore, regions with low reflectance (vegetation cultivated regions) in cultivated land are first removed from the visible band (blue, green and red bands) to extract clouds and bare land with reflectance thresholds of 0.2, 0.25, and 0.2. Subsequently, a low cloud brightness temperature in the infrared band is employed as a criterion to distinguish bare land from cloud pixels. Brightness temperature thresholds in different temperature zones and seasons are shown in Table IV

$$\begin{aligned} & \text{Band 1} > 0.2 \cup \text{Band 2} > 0.25 \cup \text{Band 3} > 0.2 \cap \text{BT}_6 \\ & < \text{TBT}_6. \end{aligned} \quad (6)$$

TABLE IV
TM6 BRIGHTNESS TEMPERATURE THRESHOLD TBT_6 (K) IN THE CULTIVATED LAND AREA

	Tropic	Temperate zone	Frigid zone
Spring	290	285	280
Summer	298	298	285
Autumn	290	285	280
Winter	285	275	275

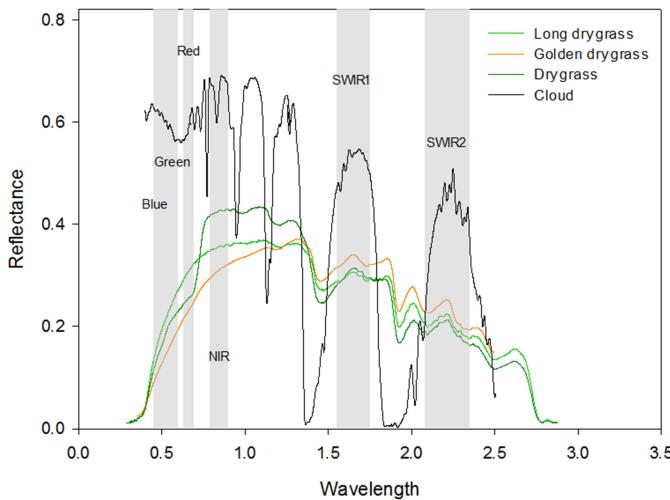


Fig. 8. Spectral curves of various grasslands.

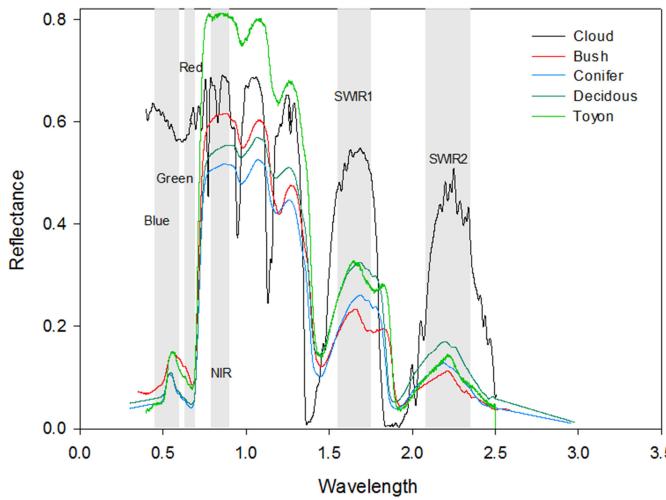


Fig. 9. Spectral reflectance of various tree and vegetation types.

Equation (6) is a threshold discriminant for cloud detection over cultivated land, where BT_6 refers to the brightness temperature of the thermal infrared band (Band 6) and TBT_6 is the brightness temperature threshold. The pixels that satisfy such a condition are identified as cloud pixels.

2) *Forest, Grassland, and Shrubland*: Unlike cultivated land, forest, grassland, and shrubland growth rates are uniform within a relative range. In Figs. 8 and 9, the reflectance feature curves of grasslands and trees in diverse types and clouds are given. Dependent on these figures, the reflectance values of

vegetation types in the visible band are significantly different from those of clouds, despite only slightly different reflectance curves. As a result, the blue, green, and red bands are selected to identify clouds over grassland and forest areas. For cloud detection over shrubland areas, blue, green, and shortwave infrared bands (Band 7) are adopted.

In terms of forest, toyon, deciduous, and coniferous trees are the major vegetation types in tropical, temperate, and frigid zones, respectively. The spectral curves of the above three typical vegetation types are used as references to set the forest cloud detection thresholds. Due to the latitudinal influence of tropical and frigid zones, the temperatures in those locations are relatively stable year round, and the four seasons in two of the zones are not distinctive. In contrast, tropical forests are hot and humid throughout the year and predominated by evergreen forests with wide and flourishing leaves. On this basis, the proportions of pixels occupied by bare land and vegetation are assumed to be 0 and 1, respectively. In the frigid zone where the temperature is perennially low, coniferous forests with evergreens and needle-like leaves are distributed during all four seasons. Hence, the percentage occupied by bare land and vegetation are supposed to be 0.2 and 0.8, respectively. In the temperate zone with four distinctive seasons, most areas are deciduous forests. In this study, different bare land and vegetation mixing ratios are used according to various degrees of leaves falling in forests during spring (March–May), summer (June–August), autumn (September–November), and winter (December–February) to acquire cloud detection thresholds in all four seasons.

Grassland was not classified because it is less susceptible to altitude. The grassland reflectance thresholds in the blue, green, and red bands are set to 0.2, 0.23, and 0.3, respectively, in this study. For temperate zone grassland, different thresholds are given for different grass-bare land proportions according to spring, summer, autumn, and winter. Due to climatic factors, grassland decay in autumn and winter is neglected for the tropical zone, and the ratio between grass and bare land is set to 1. Vegetation is sparse in the frigid zone where the temperature is low year round, but the temperature during the summer in this zone is similar to that during spring in the temperate zone. As a result, the threshold established for summer in the frigid zone should be consistent with that of spring in the temperate zone. In the remaining cold periods, the temperature conformed with the winter thresholds of the temperate zone.

There is a wide variety of shrubs, leaf colors, flower colors, and growth cycles, which indicates poor uniformity. Furthermore, the red reflectance of some shrub varieties is higher than that of other species, due to the influence of leaves or flowers, and fruits. Therefore, cloud detection in this study is conducted over shrubland based on the blue, green, and shortwave infrared (Band 7) bands, which have reflectance thresholds of 0.16, 0.18, and 0.25, respectively. Additionally, the shortwave infrared band has a relative reflectance difference from that of clouds. In the temperate zone, different thresholds were determined according to the four distinctive seasons of spring, summer, autumn, and winter. In the tropical zone where shrubs are thick and evergreen, threshold settings should be defined similarly to those of summer in the temperate zone. Compared with shrubs in warmer

TABLE V
THRESHOLDS FOR FOREST, GRASSLAND, AND SHRUBLAND AREAS

	Forest	Soil	Tropic	Frigid zone	Temperate zone			
					Spring	Summer	Autumn	Winter
N _s /N _f			0.0/1.0	0.2/0.8	0.6/0.4	0.0/1.0	0.4/0.6	0.9/0.1
T _{Band1}	0.12	0.18	0.15	0.132	0.144	0.120	0.156	0.174
T _{Band2}	0.18	0.20	0.20	0.184	0.188	0.180	0.192	0.198
T _{Band3}	0.13	0.25	0.18	0.154	0.178	0.130	0.202	0.238
Grass	Soil	Tropic		Frigid zone	Temperate zone			
				Summer	Other season	Spring	Summer	Autumn
N _s /N _g			0.0/1.0	0.6/0.4	0.9/0.1	0.6/0.4	0.0/1.0	0.4/0.6
T _{Band1}	0.20	0.18	0.20	0.192	0.182	0.192	0.200	0.188
T _{Band2}	0.23	0.20	0.23	0.218	0.203	0.218	0.230	0.212
T _{Band3}	0.30	0.25	0.0	0.280	0.255	0.280	0.30	0.270
Shrub	Soil	Tropic		Frigid zone	Temperate zone			
					Spring	Summer	Autumn	Winter
N _s /N _b			0.0/1.0	0.4/0.6	0.6/0.4	0.0/1.0	0.4/0.6	0.9/0.1
T _{Band1}	0.16	0.18	0.162	0.168	0.168	0.162	0.172	0.176
T _{Band2}	0.18	0.20	0.182	0.188	0.188	0.182	0.192	0.196
T _{Band7}	0.25	0.40	0.265	0.310	0.310	0.265	0.340	0.370

regions, shrubs in the frigid zone are more cold-resistant. Moreover, because the temperature variation range is small in the frigid zone, the growth conditions of shrubs are stable. As a result, the spring threshold in the temperate zone is selected as the cloud detection threshold for shrublands in the frigid zone.

In Table V, the forest column refers to reflectance values based on spectral characteristic curves of major vegetation, the grass column gives the reflectance values of grass, the bush column is those of typical shrubs, and the soil column lists the reflectance values of bare land. In addition, T_{Band1} , T_{Band2} , T_{Band3} , and T_{Band7} are the reflectance thresholds of the blue, green, red, and the near-infrared, which are worked out after mixed pixel calculations. Also, N_s , N_f , N_g , and N_b are the proportions of bare land, forest cover, grass, and shrub pixels. The discriminants of forest and grassland cloud pixels are expressed in (7). In addition, the discriminant for shrubland cloud pixels is presented in (8)

$$\text{Band 1} > T_{Band 1} \cup \text{Band 2} > T_{Band 2} \cup \text{Band 3} > T_{Band 3} \quad (7)$$

$$\text{Band 1} > T_{Band 1} \cup \text{Band 2} > T_{Band 2} \cup \text{Band 7} > T_{Band 7} \quad (8)$$

C. Snow/Ice Surfaces

In this study, cloud detection identification for permanent snow and ice and tundra regions is not considered. Regarding areas covered by snow year round, such as the summit of a mountain, the snow coverage degrees during diverse periods can be considerably different due to the influences of the air temperature, sun exposure, etc. However, GlobeLand30, which only reflects the coverage in a single state, fails to support data detection based on multitemporal data sets. In most cases of tundra affected by snow and ice coverage, it is difficult to perform accurate cloud detection, because various, complex landforms at the same latitude have dramatically different reflectance values. Moreover, the practical significance of cloud detection in large areas of snow and ice at high altitudes and in polar regions, etc. are also limited. Hence, no cloud detection identification method is suggested in this study for areas of permanent snow and ice cover.

For temporary winter snow/ice coverage in nonpermanent snow and ice areas of a remote sensing image, the NDSI [37], [38] is utilized in this study to distinguish these areas from clouds and then to remove them. The clouds to be detected in the image may be ice clouds [39] with high NDSI values, and thus, the threshold of NDSI is set at 0.7 to eliminate the effects of ice clouds. In the case that the NDSI value is greater than 0.7, snow/ice can be identified.

D. Supplementary Test

1) *Fragmented Pixel Removal*: Due to the complex reflectance distribution of the ground surface and edge deviations caused by the land type database, some small misidentifications are generated that result in inaccurate cloud detection. Therefore, after the above cloud detection steps are completed, fragmented pixel removal is performed. The method traverses all pixels identified as clouds. When the number of cloud pixels in the eight neighborhoods of one cloud pixel is less than or equal to 2, the pixel is determined to be a fragmented pixel and is removed.

2) *Correction of Artificial Surface*: Bright urban areas and clouds are difficult to separate in terms of reflectance due to their similar spectral properties (high reflectance). In addition, the spectra of high-reflectance artificial surfaces vary for different materials. Therefore, it is easy to identify the artificial surface as a cloud pixel by extracting cloud pixels only based on the reflectance information. Notably, temperature information is a significant parameter for cloud detection. Although the cloud brightness temperatures all show a high level of brightness in the images, they will be much lower than that of an artificial surface. However, there are also large uncertainties in the brightness temperature. For example, in the middle or high latitudes, the surface itself will have lower brightness temperatures (especially in winter), and it is difficult to meet global conditions with a fixed brightness temperature threshold. We assume that the artificial surface brightness temperature values in a certain image range are relatively uniform. Based on the detection results above an artificial surface with a reasonably broad thresholds, the procedure uses a dynamic brightness temperature threshold to remove urban pixels that have been mistakenly recognized as clouds from previous results.

In this correction, the previously completed cloud detection results are used as input data to exclude the urban pixels that have been misidentified as clouds in the detected results over an artificial surface. This test is only conducted when the ratio of the number of “clear artificial surface” (CAS) pixels to that of the total effective observation pixels was greater than 0.1%, which ensures that there are enough clear-sky pixels to calculate the statistics. For the completely clear-sky pixels, the pixels that have been identified as clouds are removed from all urban pixels. In addition, pixels with reflectance values of less than 0.1 in the red and near-infrared bands are also removed to remove the effects of vegetation and cloud shadows, respectively, from the artificial surface, which will result statistically in a lower brightness temperature.

The brightness temperature of the CAS pixels is summed from the minimum to maximum values at intervals of 0.1 K, counting the number of pixels in each interval and calculating the frequency. The formula to exclude misidentified cloud pixels using the brightness temperature is as follows:

$$T_{\text{cor_arti}} = \sum_{i=1}^{\left\lceil \frac{T_{\max} - T_{\min}}{0.1} \right\rceil} \left(\frac{N_i}{N_{\text{total}}} \right) \times (T_{\min} + 0.1 \times i) \quad (9)$$

where $T_{\text{cor_arti}}$ is the dynamic brightness temperature correction threshold, T_{\max} and T_{\min} are the maximum and minimum brightness temperatures in the CAS pixels, respectively; N_{total} is the total number of pixels, and N_i is the number of pixels in the i th interval.

V. CLOUD DETECTION RESULTS AND ACCURACY EVALUATION

To maintain temporal consistency between the detected data and the GlobeLand30 product, the experimental data acquired for this study are from 2009 to 2011. For details, please refer to Table VI. For the Landsat TM data, the DN value of the image should be converted to the apparent radiance with definite physical meanings to correspondingly figure out the apparent reflectance or brightness temperature (10–12). Subsequently, the image to be detected could be further processed. The relevant conversion formulas are as follows [40]:

$$L_{\lambda} = \text{Gain} \cdot Q_{\text{cal}} + \text{Bias} \quad (10)$$

$$\rho = \frac{\pi \cdot L_{\lambda} \cdot D^2}{\text{ESUN}_{\lambda} \cos \theta_S} \quad (11)$$

$$T = \frac{K_2}{\ln \left(\frac{K_1}{L_{\lambda}} + 1 \right)} \quad (12)$$

where Gain refers to the gain value, Bias is the bias value, Q_{cal} is the DN value, L_{λ} is the apparent radiance (unit: $\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \mu\text{m}^{-1}$), ρ is the apparent reflectance of the top of the atmosphere, D is the Earth–Sun distance, ESUN_{λ} is the mean solar exoatmospheric irradiance in band λ , θ_S is the solar zenith angle, T is the brightness temperature (units: K), and K_1 and K_2 are calibration constants. Regarding Landsat 5, the values of K_1 and K_2 are $607.76 \text{ W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \mu\text{m}^{-1}$ and 1260.56 K , respectively.

In this study, Landsat TM remote sensing images from 2009 and 2012 are selected. During preprocessing and cloud detection for these images, two scene images are chosen from the detection results, which is according to diverse land cover types, to perform visual interpretations for general and local results and false color images (RGB, Bands: 432). This approach yields an intuitive and clear contrasting effect. Then, 60 sample areas are selected from all the images listed in Table VI to map cloud areas based on the false color images. Subsequently, quantitative validation is conducted for the LCCD results. Ultimately, the LCCD results, Fmask products, and false color images are visually compared at a local scale, and the associated uncertainties are analyzed.

TABLE VI
DATA SOURCE INFORMATION

ID	Land Type	Date	Corner coordinate(UL) Latitude
1		2010/07/19	39.8
2		2010/09/13	49.9
3	Cultivated land	2010/07/11	48.4
4		2010/04/13	37.0
5		2010/02/23	15.4
6		2011/03/06	14.0
7		2011/01/26	-0.50
8	Forest	2010/02/03	28.4
9		2010/07/20	69.3
10		2011/08/29	3.8
11		2011/02/02	14.0
12		2010/05/15	48.4
13	Grassland	2010/10/23	47.0
14		2010/07/05	47.0
15		2011/07/12	69.3
16		2010/02/11	32.7
17	Shrubland	2010/04/23	-32.2
18		2010/08/22	58.3
19		2010/05/04	-7.7
20		2011/01/02	-4.9
21	Wetland	2010/08/13	52.7
22		2011/08/05	-26.5
23		2011/03/06	12.5
24		2009/08/24	48.4
25	Water bodies	2010/06/07	54.1
26		2010/07/31	55.5
27		2010/06/18	44.4
28		2010/07/10	42.7
29		2010/09/02	42.7
30	Artificial Surfaces	2010/06/06	56.9
31		2009/11/13	49.8
32		2011/01/31	-22.2
33		2010/02/06	-22.2
34		2010/06/04	42.7
35	Bareland	2010/08/20	41.3
36		2010/08/05	28.4
37		2011/05/07	18.3
38		2011/04/10	18.3
39		2011/01/18	45.6
40	Ocean	2009/12/05	29.8
41		2009/12/08	18.3
42		2010/11/15	28.4
43		2010/02/02	16.8
44		2010/08/04	42.7
45	Complex surface	2010/06/02	58.3
46		2010/09/20	48.4
47		2010/10/30	28.4

A. Cloud Detection Results

1) *Constant Attribute Surfaces*: The cloud detection results over constant attribute surfaces are presented in Fig. 10. On the left in each figure is a false color image, and the LCCD results are shown on the right. To clearly visualize the effects of the LCCD algorithm, images of two scenes are selected for each type of underlying surface type to comparatively analyze the general results and local details between cloud detection and the false color images, respectively. Fig. 10(a), (b), (c), and (d) represents the underlying surfaces of water bodies, wetland, bare land, and artificial surfaces, respectively.

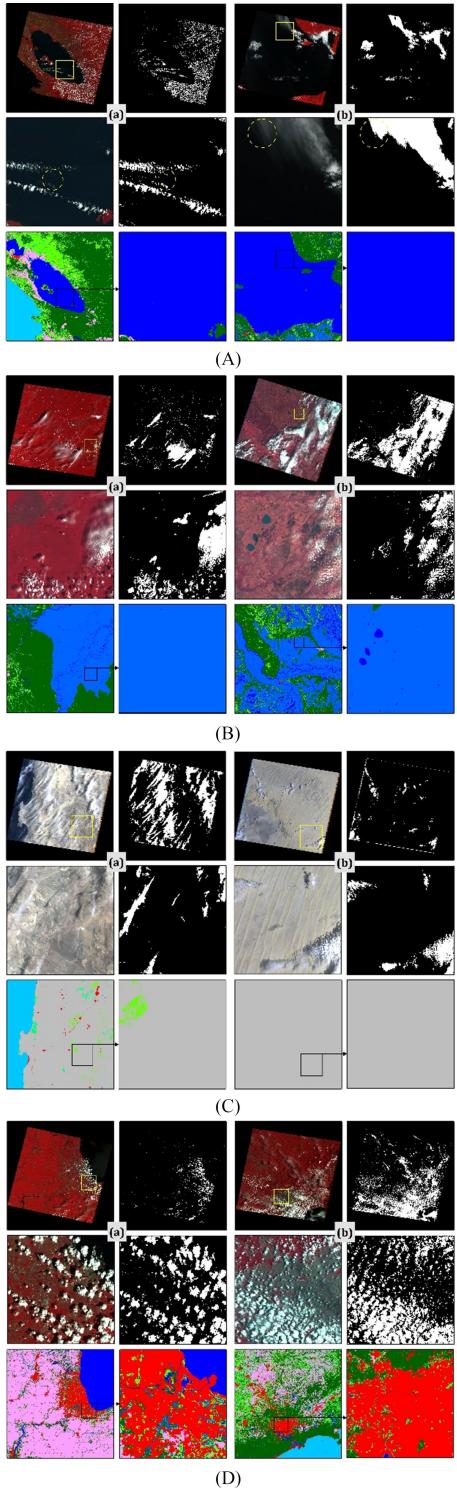


Fig. 10. Cloud detection results over constant attribute surfaces underlying surface type. (A) Water bodies. (B) Wetland. (C) Bareland. (D) Artificial surface.

Both water bodies and wetlands are underlying low reflectance surfaces. Specifically, water bodies with low overall reflectance values show dark colors in the image; furthermore, almost no attribute variation occurs in water bodies, and thus, the reflectance curves are relatively stable and suffer little seasonal or latitudinal influence, except during the freezing period.

Wetland areas are covered by aquatic plants in most cases, and the ground in their growing regions is usually wet and soft or may even be water bodies. In addition, wetland ecological systems are very stable and lack dramatic internal change. Hence, the long-term overall wetland reflectance is relatively low. According to Fig. 10(a) and (b), favorable detection effects are achieved for thick and thin clouds over water bodies and wetland. The corresponding detection results are also consistent with the tendencies revealed by false color images. For ordinary cloud detection threshold methods, thresholds for clouds over dark surfaces may be subjected to the impacts of other land cover types in images, which may result in an excessively high value and cause the omission of pixels composed of thin clouds and water bodies from mixed pixel identification. However, the LCCD algorithm can be adopted to establish a proper cloud detection threshold for dark surfaces, which is in line with the classification of underlying surfaces, to avoid the occurrence of the phenomena described above. The local results given in Fig. 10(a) and (b) indicate that good LCCD-based detection effects can be acquired for thick, thin, and broken clouds, which as shown by the dotted line circle in these results.

Fig. 10(c) and (d) presents the cloud detection results over bare land and artificial surfaces, both of which have relatively high and stable reflectance values. Bare land and artificial surfaces are characterized by stability, as their reflectance values remain within an extremely small range over time and latitude changes. Similar to dark areas, the threshold set for bright surfaces through traditional cloud detection threshold methods may also be inappropriate. For example, although a synthesized threshold can be used to achieve a good effect in dark areas by considering various surface reflectance features, the threshold may be slightly lower in bright areas. As a result, there could be some false detection because of the inappropriate threshold. However, the LCCD algorithm can set particular thresholds for bright areas of bare land and artificial surfaces to reduce the misidentification rate, provided that the LCCD algorithm is supported by known land cover types. In these figures, the detection results obtained using the LCCD algorithm remained consistent overall with the cloud distribution, as shown on the false color image, and favorable detection effects are also achieved according to the corresponding thresholds of thick clouds over bare land. The local detection results also show that thin cloud identification over bare land is also very accurate because the LCCD algorithm has been employed. As indicated by the dotted line circle, thin clouds and even edges of clouds in this area can be clearly detected. Moreover, for clouds and their edges can also be accurately identified over artificial surfaces. Nevertheless, excessively bright areas of artificial surfaces may occasionally be misidentified as clouds.

2) Changing Attribute Surfaces: Changing attribute surfaces are shown in Fig. 11 and Fig. 11 (a), (b), (c) and (d) are the underlying surfaces of cultivated land, forest, grassland, and shrubland. These underlying types are classified in the vegetation category. These surfaces change with seasons and climate and exhibit different growing statuses, which further affects the reflectance characteristics of surface features. Because the traditional threshold methods of cloud detection seldom consider

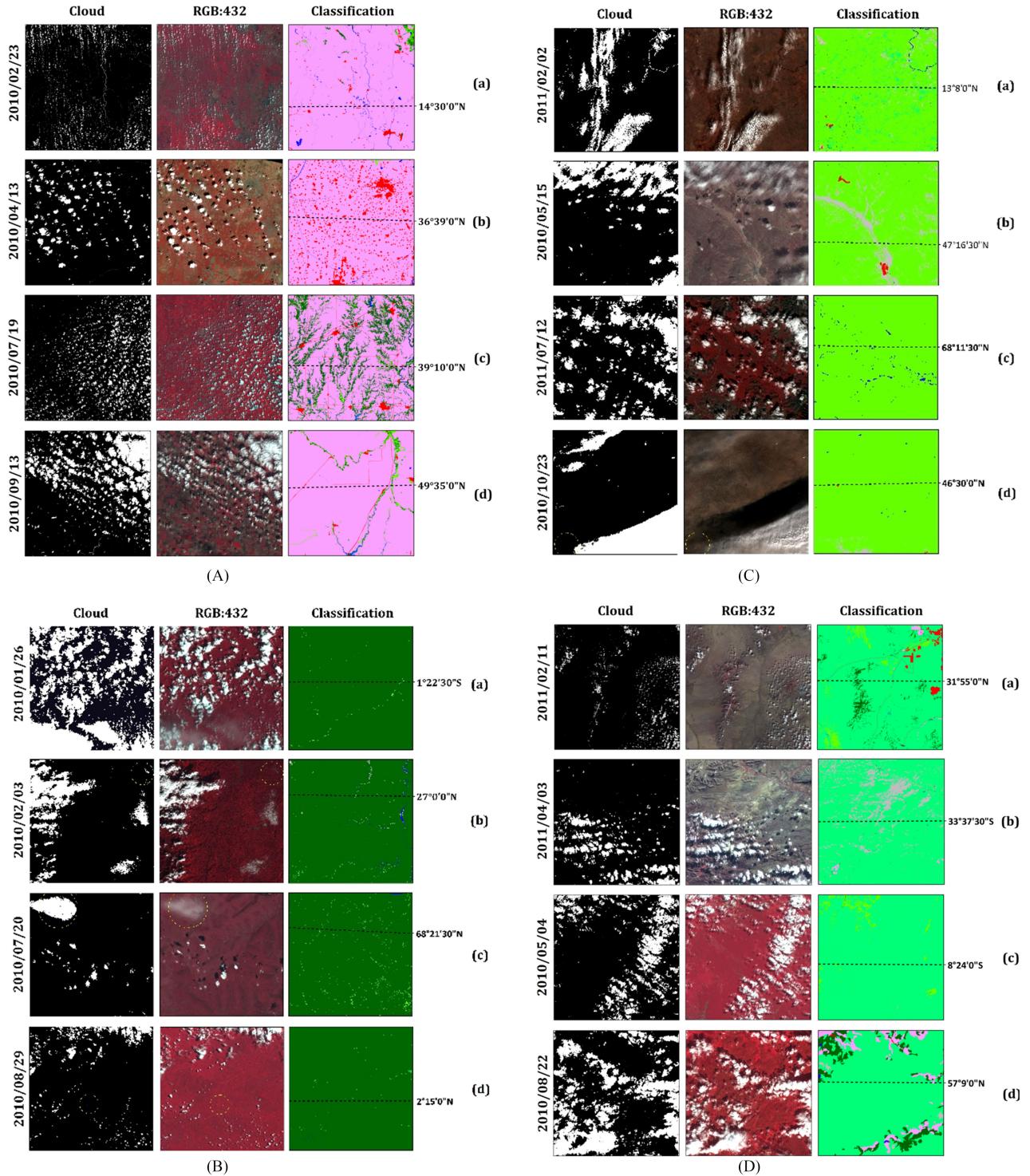


Fig. 11. Cloud detection results over changing attribute surfaces underlying surface type. (A) Cultivated land. (B) Forest. (C) Grassland. (D) Shrubland.

the different reflectance characteristics of vegetation during various growing periods, four categories of vegetation surfaces are subdivided in this study according to the season and latitude to obtain the corresponding cloud detection thresholds. Therefore, the application of cloud detection thresholds is improved accordingly.

In Fig. 11(a), (b), (c), and (d), four scene images with distinctive latitudes and seasonal features, which are convenient for verifying the threshold applicability for different seasons or

latitudes in detection areas, were selected for cloud detection. The land cover type of cultivated land is shown in Fig. 11(a). Due to little cultivated land surfaces in the frigid zone, this type is mostly distributed in temperature and tropical zones. For this reason, cultivated land is not utilized as a typical area to be demonstrated. Furthermore, Fig. 11 also indicates that preferable detection outcomes can be obtained for thick, thin, and broken clouds over cultivated land in a variety of seasons or latitudes if the LCCD algorithm is adopted. Even for areas where

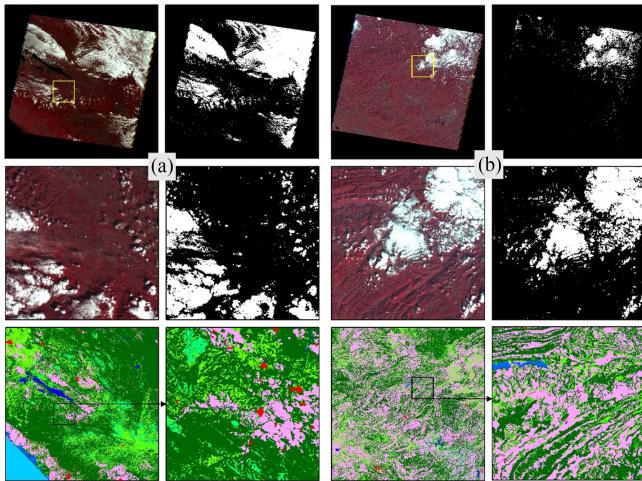


Fig. 12. Cloud detection results over complex surfaces.

city and cultivated land overlap each other, such as in Fig. 11(b) and (c), the obtained detection results are still complete and accurate. In Fig. 11(b), the cloud detection results over forest areas are also very precise. For example, the LCCD algorithm is used to identify thin sheet clouds presented in Fig. 11(a), (b), and (c) as well as thin, sporadic, and broken clouds in Fig. 11(d), which is within the range encircled by the yellow dotted line. The sparsity of grassland and shrubs is closely correlated with seasonal variations. In leaf withering or decaying periods, the characteristics of bare land become much more obvious than those of forest. As shown by the false color in the figure, the more distinctive the red band is, the more vigorously the vegetation will grow. In Fig. 11(c) and (d) where vigorous vegetation growth and sparsity periods are included, good detection results, which are consistent with cloud distribution presented by the false color image, can be acquired using the LCCD algorithm in different seasons and at different latitudes. Moreover, for thick, thin, and broken clouds, the detection capacity is even stronger, and the number of corresponding inaccurate and missing extractions is also very low. However, there is still a small portion of thin cloud edges that have not been extracted (shown by the yellow dotted line) in Fig. 11(d) and (c).

The LCCD is an algorithm based on land cover type support, and ideal cloud detection results have been achieved when targeting a single land cover type. However, global land cover types are diverse. In addition to large areas of identical surfaces, multiple land cover types can be adjacent, intersect, and inlaid, which are commonly seen. Complex surfaces may complicate the spectral characteristics of surfaces in such areas. Thus, the land cover boundary problem is unavoidable. However, on the one hand, thresholds are set slightly higher instead of selecting the spectral curve reflectance values as absolute thresholds, which leaves room for uncertainties (such as surface changes, atmospheric effects, and so on). On the other hand, the reflectance fluctuation caused by a boundary change is far less than the cloud influence. To verify the applicability of such an algorithm to clouds over complex surfaces, the images in Fig. 12 are selected for validation analysis. The results revealed by these images show that the LCCD algorithm generates a high

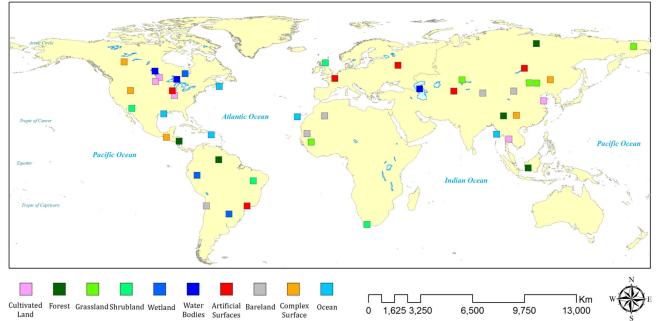


Fig. 13. Diagram of the distribution of quantitative validation data.

level of accuracy for cloud identification over complex underlying surfaces regardless of the general or local cloud detection results, and a favorable fitting degree exists among the diverse land cover types. Such a phenomenon indicates high integrality. Without a doubt, the LCCD algorithm is preferably applicable to all types of underlying surfaces and exhibits a satisfactory cloud detection capability.

B. Accuracy Evaluation

Images of 47 scenes are evenly selected around the world to verify the accuracy of the LCCD algorithm in this paper and six sample areas with 500×500 pixels are randomly chosen from images corresponding to all types to perform visual interpretations for tasks conducted in the cloud region, and then the areas judged to be clouds are mapped out. Data source information related to the images adopted has been given in Table VI, and their distributions are shown in Fig. 13. Because North America and Asia have not only large areas but also diverse land cover types, more images have been selected from these areas. Images selected for each type should have a uniform distribution to the greatest extent possible. For land cover types with vegetation, latitudinal variations were considered in selecting the relevant images. Then, the results of visual interpretation are regarded as truth values to compare with cloud detection results obtained through the LCCD algorithm.

In this study, six indexes, including the cloud proportion (CP), correct rate of cloud (CRC), correct rate of clear-sky (CRS), total correct rate (TCR), error rate (ER) and missing rate (MR) are used to evaluate the accuracy of the LCCD algorithm. The corresponding computational formulas are as follows:

$$CP = \frac{TNC}{NT} \quad (13)$$

$$CRC = \frac{NC}{NC_V} \quad (14)$$

$$CRS = \frac{NS}{NS_V} \quad (15)$$

$$TCR = \frac{NC + NS}{NC_V + NS_V} \quad (16)$$

$$ER = \frac{NE}{NS_V} \quad (17)$$

$$MC = \frac{NM}{NC_V} \quad (18)$$

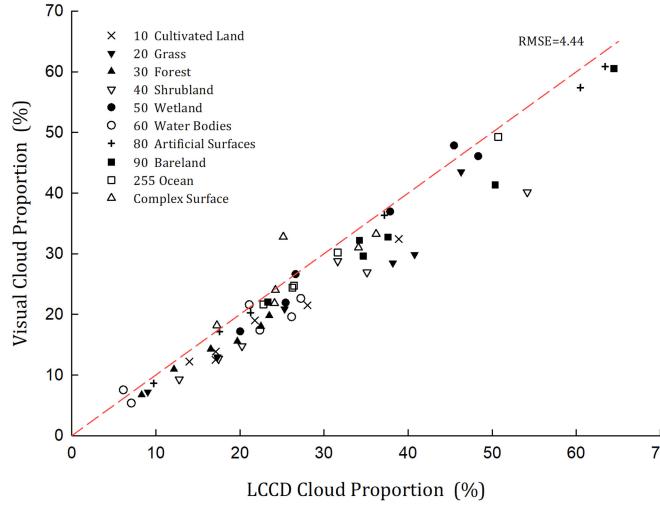


Fig. 14. Statistical regression diagram of the CP.

TABLE VII
RMSE OF THE CP OVER DIVERSE LAND COVER TYPES

Land Type ID	10	20	30	40	50	60	80	90	255	Complex Surfaces
RMSE (%)	5.05	7.27	3.45	8.21	2.54	4.35	1.97	5.52	1.75	4.07

where TNC refers to the number of cloud pixels; NT is the total number of pixels; and NC and NS are the numbers of cloud and clear-sky pixels, respectively, that are clouds in both the LCCD algorithm and visual interpretation (true value) results. NC_V and NS_V represent the numbers of cloud pixels and clear-sky pixels in the visual interpretation results, respectively. Furthermore, NE and NM are the numbers of cloud pixels with false or missing identification information among the results obtained by the LCCD algorithm.

In this paper, a regression analysis is conducted for CP, which is shown the visual interpretation and the LCCD cloud results in the sample area, to determine the RMSE of the LCCD cloud results. Subsequently, the CRC, CRS, ER, and MR indexes are used to perform a detailed statistical analysis and precision evaluations of cloud and clear-sky pixels according to the cloud detection results. Ultimately, the TCR values are calculated for all sample points, i.e., the TCR values of accurate cloud and clear-sky pixels are identified based on the LCCD algorithm.

1) *Validation of the CP Index:* A regression analysis of the CP statistical results from the LCCD and visual interpretation is given in Fig. 14, and Table VII reveals the RMSE calculation results of the CP over various surfaces. According to Fig. 14, when each surface type is compared with the true values, most of the CPs obtained by the LCCD are underestimated. Although the corresponding underestimation deviation is small, it has an overall tendency to be consistent with the visual interpretation results and is close to the reference line. The overall RMSE is 4.44%, which indicates a high correlation and a relatively high accuracy. As shown in Fig. 14 and Table VII, the RMSE values of shrubland, grassland, and bare land have substantially deviated from the reference line and are significantly underestimated.

Specifically, the RMSE values are 8.21%, 7.27%, and 5.52%, respectively. Clearly, the deviation exhibited by shrubland is the largest, and cultivated land, with an RMSE value of 5.05%, ranks behind bare land. Additionally, the land cover type with the minimum error for the CP index is ocean (RMSE of 1.75%).

These results indicate that the CP index error for shrubland is at the maximum level, which reflects a significant underestimation. The reason for this underestimation is that a great difference lies in the spectral information and growth cycles of shrubs in diverse varieties, and thus, it is difficult to cover all situations with a uniform threshold criterion. Concerning the land cover type of obviously changing shrubland, the LCCD algorithm generates the maximum error of CP. Due to its low reflectance, the spectral information of the ocean is tremendously distinguishable from that of clouds. Therefore, the statistical value of CP is nearest the true value for oceans.

2) *Validation of Evaluation Indexes:* In this section, the CRC, CRS, ER, and MR indexes are selected for more specific accuracy evaluations. Fig. 15 presents the accuracies of the indexes in six sample areas with various land cover types and the accuracy distribution ranges for cloud detection. Different land cover types can be observed in this figure. The sums of both CRC and MR and CRS and ER are equal to 1.

The accuracy distributions of the CRC and MR indexes reveal that the correct rate of detection for various cloud categories ranges from 0.7 to 1.0, which represents a rather high cloud identification accuracy. Simultaneously, the corresponding MR range is between 0 and 0.3. Among them, the overall accuracy of cloud pixel identification for oceans (255) is high which also exhibits the minimum MR. Moreover, the accuracy distributions of the six sample areas in the image are highly coincident with each other. Overall, the cloud identification accuracy for wetlands is high, and the relevant MR is low. Conversely, the overall cloud pixel identification accuracy is low and concentrated for shrubland areas, which also exhibits a high MR. In addition, the CRC and MR results conform to the verification trends of CP.

Based on the accuracy distributions of CRS and ER, the CRS pixel detection is high and lies between 0.9 and 1. Likewise, the MR of clouds is also low (0–0.1). Among the land cover types, forest is proven to have the highest correct rate of all clear-sky pixels, and the accuracies of the six samples are also very close to each other. In addition, both grassland and ocean give high and correct clear-sky pixel identification accuracies and low MRs. Moreover, these accuracies are similar to that for forests. This illustrates that the above land cover types are highly stable under the circumstances of high correct rates for clear-sky pixels. By contrast, the clear-sky correct rate of wetland is relatively low and the accuracies for all samples substantially differ and vary in an unstable manner.

By investigating the six sample areas for all land cover types, the overall accuracy of all indexes is obtained for each surface (see Fig. 16). Notably, CRC pixel identification reaches 70% and higher. Specifically, the land cover type with the highest cloud pixel identification rate and lowest MR is ocean (93.96%), which is successively followed by artificial surfaces (92.13%) and wetland (91.81%). Shrubland exhibits the min-

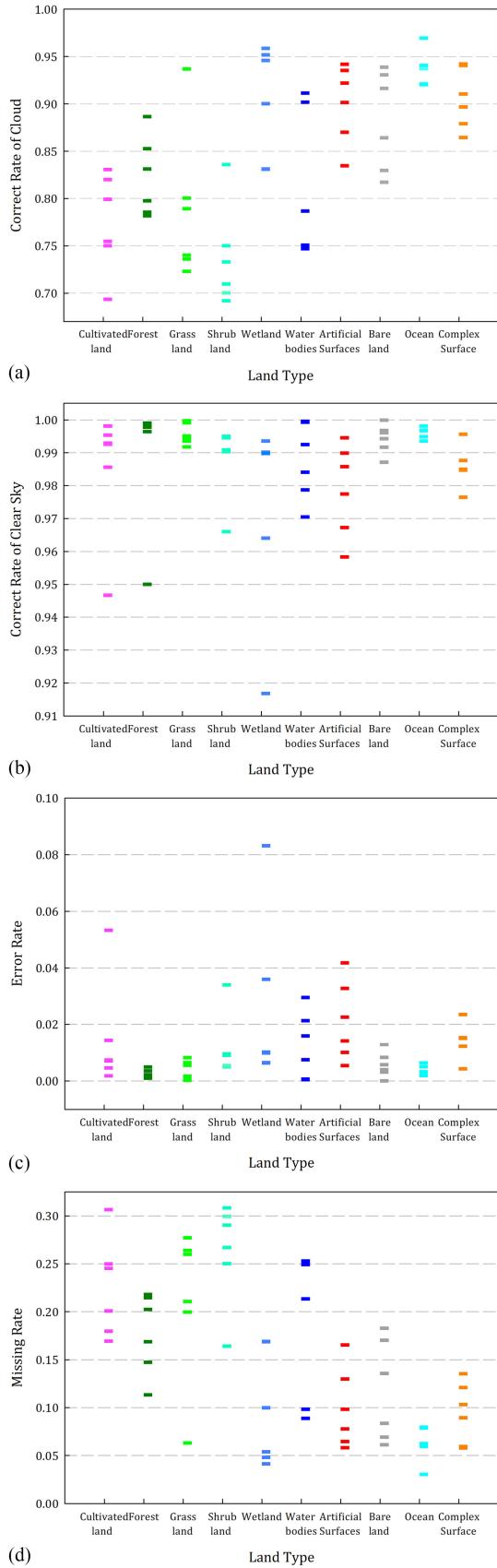


Fig. 15. Statistical diagram of all sample evaluation indexes.

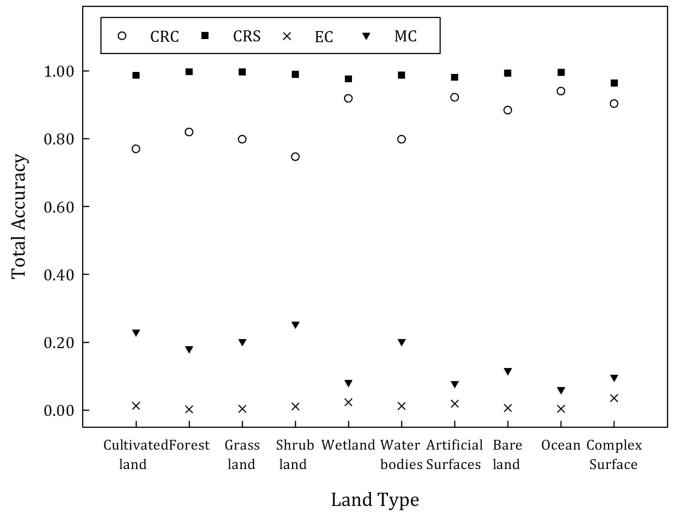


Fig. 16. Statistical diagram of the overall evaluation indexes for all land cover types.

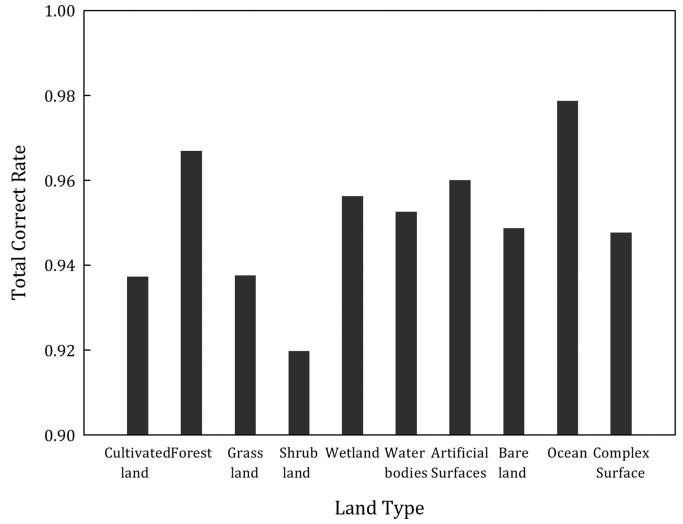


Fig. 17. Statistical TCR results for all land cover types.

imum cloud pixel identification rate (74.64%). All clear-sky pixels have high correct rates, which are greater than 90%, indicating that this algorithm can be adopted to very accurately identify clear-sky and generate stable identification effects.

Statistics are also obtained for TCR based on the LCCD, and the relevant statistical results are given in Fig. 17. Combining the cloud and clear-sky pixel identification details, the TCR is obtained and shown in Fig. 17 of the statistical results. Based on this figure, the TCR of the LCCD ranges from 90% to 100%. As differences in growth cycles and vegetation colors are incurred by different varieties and scarcities, the identification of shrubland is substantially uncertain, which further leads to the minimum accuracy of 91.98%. As ocean and forest are dark surfaces and the cloud spectral information over these areas can be extremely distinctive, a high associated level of accuracy is

TABLE VIII
TOTAL ACCURACY EVALUATION FOR THE LCCD ALGORITHM

Evaluation Index	CRC	CRS	TCR	ER	MR
Result	0.8650	0.9870	0.9508	0.0130	0.1350

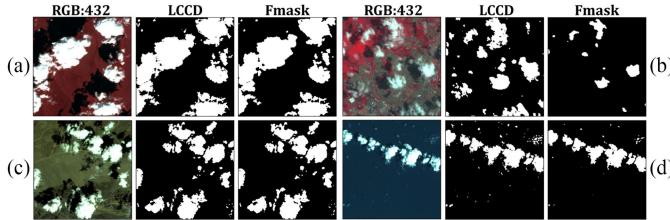


Fig. 18. Comparison of the LCCD and Fmask results for thick clouds.

achieved. The two land cover types with the highest accuracy are ocean and forest, and their TCRs are 97.88% and 96.69%, respectively.

The cloud detection results for all land cover types are summarized to perform a statistical analysis of all pixels. Then, the overall accuracy of the LCCD algorithm is determined for further calculations and evaluations. Table VIII presents the total accuracy of the five evaluation indexes (CRC, CRS, TCR, ER, and MR). These accuracies indicate that the CRS identification is high, and the TCR for cloud identification can reach 86.50%.

C. Comparison With the Fmask Cloud Results

Fmask is an algorithm commonly used for the generation of Landsat cloud and cloud shadow products at present. Fmask cloud detection first uses a series of spectral thresholds to identify obvious clouds and obtain the potential cloud pixel layer. Subsequently, probability statistics are utilized to perform the cloud probability parameter computation for the remaining clear-sky pixels. In combination with the potential cloud pixel layer acquired during the first step, the potential cloud layers can be obtained. In this section, the cloud detection results from the LCCD algorithm are compared with those obtained using the Fmask algorithm.

A comparison of the cloud detection results based on the LCCD and Fmask for both thick and thin clouds is performed. The results of the Fmask are dilation processed, and those of the LCCD are not. Therefore, no dilation results for the Fmask are selected to compare with those of the LCCD algorithm. Moreover, the quantitative validation data are obtained by visual interpretation.

Good results are obtained for thick clouds, as shown in Fig. 18, since thick clouds and the surface exhibit a large difference in reflectance values. According to the statistics, the TR and SR accuracies are both higher than 0.9, whereas most of the accuracies are higher than 0.95. Additionally, CR displays satisfactory results, as illustrated in Fig. 19.

Compared with thick clouds, which are easy to identify because of their high reflectance values, thin clouds have always been difficult to identify in cloud detection due to their much lower reflectance values (simultaneously includes the surface

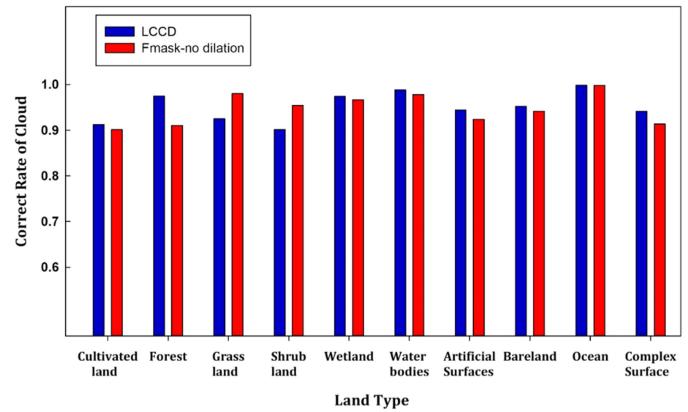


Fig. 19. Statistical comparison of the LCCD and Fmask for thick clouds.

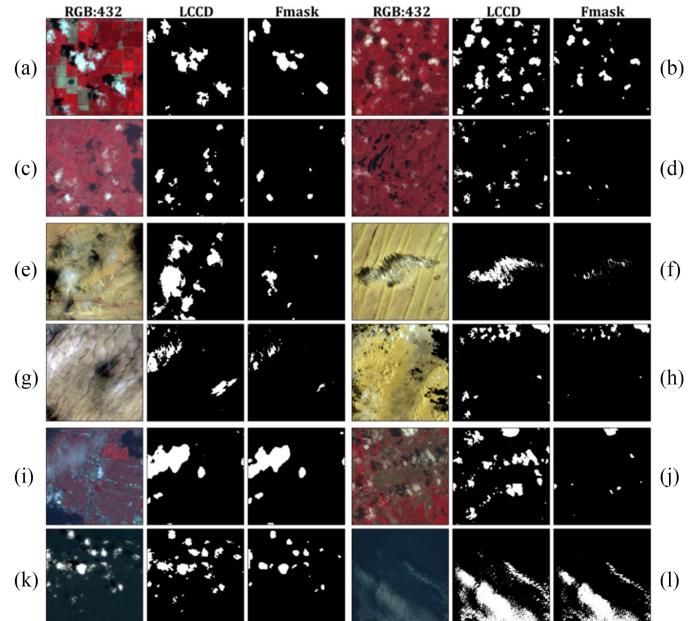


Fig. 20. Comparison of the LCCD and Fmask results for thin clouds.

information) than thick clouds and have less influence on reflectance increasing than thick clouds, making extraction difficult. For thin clouds, the LCCD cloud results are more accurate than those of the Fmask, which can detect more thin clouds. Fig. 20 shows the thin clouds over different land types, including vegetated, bare soil, artificial surface, and water areas.

As the figure illustrates, the LCCD algorithm yields better results in thin clouds areas. The Fmask algorithm performs a series of spectrum-based tests (including single-band and multi-band combinations). These tests are usually based on the difference between clouds and general surface features. However, thin clouds over different objects have different sensitivities to these tests, which leads to inaccuracies for thin clouds. For example, the brightness temperature in the basic test and the HOT test may lead to false negatives. Subsequently, the Fmask performs brightness temperature based tests based on statistical conditions. Although the brightness temperature is an important parameter for cloud detection, when using a single-band bright

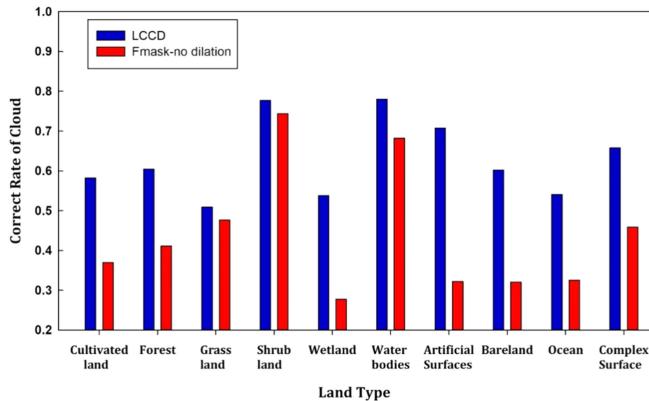


Fig. 21. Statistic comparison of the LCCD and Fmask for thin clouds.

temperature, there are still some uncertain factors linked to different land types. The Fmask algorithm calculates the brightness temperature and performs a series of operations to obtain the threshold according to the image, and thus, there will be some errors. Then, the brightness temperature tests will affect the cloud detection results.

The LCCD algorithm distinguishes the judgment conditions according to different surface conditions and reduces the influence from the consideration of various surface complexities; therefore, thin clouds can be better identified. In addition, in the LCCD algorithm, the brightness temperature is only used in areas where there is a bright surface or is counted for one kind land type, and the brightness temperature is used as an auxiliary parameter. Thus, there are fewer errors related to the brightness temperature.

In Fig. 21, quantitative validation of the thin cloud detection results is conducted. Only thin clouds are selected in the interpretation for validation. Both algorithms have high accuracies in SR, which is higher than 0.95 for each land type. However, Fig. 21(a) shows that the LCCD algorithm yields better results than the Fmask for CR, although neither of them has a high CR due to the objective difficulty of thin cloud detection. The better LCCD cloud results mainly due to the establishment of appropriate and specific thresholds for each type of land surface. Therefore, the LCCD algorithm performs generally better for thin clouds.

D. Uncertainty Analysis

The LCCD algorithm is a cloud detection algorithm proposed based on 30-m land cover data. Although the algorithm is supported by prior land cover data and yields satisfactory detection effects, the detection results still occasionally deviate from the true values. The uncertainties associated with the LCCD algorithm can be described as follows.

- 1) Formed by thin clouds and shadows generated by nearby clouds, the mixed pixels have lower reflectance values, which does not match the characteristics of clouds. In this case, some cloud pixels may be occasionally missed by the LCCD algorithm.
- 2) Despite considering the effects of seasonal and latitudinal variations in vegetation on thresholds in this study

and despite the fact that ideal cloud detection results can be acquired in most cases, some deviation may exist in the thresholds in a few areas where vegetation conditions are special, e.g., shrubs of diverse species with various characteristics and growth cycles.

- 3) Excessively bright areas of small rivers that are dry or contain high concentrations of suspended solids can be falsely extracted as clouds because the surface properties are not consistent with those in the land type database.

VI. CONCLUSION

A simple cloud detection algorithm (LCCD) supported by GlobeLand30 land cover type data is proposed in this study. Based on the surface classification products, this algorithm classifies surfaces into constant attribute surfaces and changing attribute surfaces. With respect to the former, a surface feature is assigned a fixed threshold, which is dependent on spectral characteristics in the spectral library. For the vegetation type, variable thresholds are established according to different latitudes and seasons based on the mixed pixel principle.

By comparing cloud pixels with false color images and conducting visual interpretations of Fmask cloud products, these pixels are shown to be accurately extracted by the LCCD algorithm, which suggests that the cloud detection results are satisfactory. As demonstrated via quantitative validation, the LCCD algorithm reaches a high level of accuracy overall. In detail, while the overall RMSE of the CP is 4.44%, the overall accuracies of CRC, CRS, TCR, ER, and MR are 0.8650, 0.9870, 0.9508, 0.0130, and 0.1350, respectively. Among all the land cover types, the highest accuracies are linked to forest and ocean areas. By contrast, shrubland exhibits the lowest accuracy. Moreover, the LCCD algorithm is preferably applicable to complex surfaces to obtain ideal cloud detection results. By comparison with the Fmask results, it is obvious that the LCCD algorithm has advantages in thin cloud detection. In general, the LCCD algorithm can obtain accurate detection results, especially for thin clouds.

In addition to the clear principle and ease of implementation of the proposed method, basing the algorithm on previously obtained land cover data provides a guarantee for reasonable threshold accuracy in cloud detection. This approach can be applied for various domains with the GlobeLand30 product and offers a new method of cloud detection.

ACKNOWLEDGMENT

The Landsat data were made available from the Earth Resources Observation and Science Center (<http://glovis.usgs.gov/>). The authors also thank Z. Zhu and C. E. Woodcock for providing the codes for the Fmask algorithm.

REFERENCES

- [1] Harshvardhan, D. A. Randall, and T. G. Corsetti, "Earth radiation budget and cloudiness simulations with a general circulation model," *J. Atmos. Sci.*, vol. 46, no. 13, pp. 1922–1942, Jul. 1988.
- [2] A. Kazantzidis, K. Eleftheratos, and C. S. Zerefos, "Effects of cirrus cloudiness on solar irradiance in four spectral bands," *Atmos. Res.*, vol. 102, no. 4, pp. 452–459, Dec. 2011.

- [3] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *J. Atmos. Oceanic Technol.*, vol. 28, no. 10, pp. 1286–1296, May 2011.
- [4] A. Kazantzidis, P. Tzoumanikas, A. F. Bais, S. Fotopoulos, and G. Economou, "Cloud detection and classification with the use of whole-sky ground-based images," *Atmos. Res.*, vol. 113, pp. 80–88, Sep. 2012.
- [5] K. S. Carslaw, R. G. Harrison, and J. Kirkby, "Cosmic rays, clouds, and climate," *Science*, vol. 94, nos. 1/2, pp. 1732–1737, 2000.
- [6] J. Yang, W. Lu, Y. Ma, and W. Yao, "An automated cirrus cloud detection method for a ground-based cloud image," *J. Atmos. Oceanic Technol.*, vol. 29, no. 4, pp. 527–537, Feb. 2012.
- [7] C. Papin, P. Bouthemy, and G. Rochard, "Unsupervised segmentation of low clouds from infrared METEOSAT images based on a contextual spatio-temporal labeling approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 1, pp. 104–114, Jan. 2002.
- [8] G. Vivone, P. Addesso, R. Conte, M. Longo, and R. Restaino, "A class of cloud detection algorithms based on a MAP-MRF approach in space and time," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5100–5115, Aug. 2014.
- [9] P. Addesso, R. Conte, M. Longo, R. Restaino, and G. Vivone, "MAP-MRF cloud detection based on PHD filtering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 919–929, Jun. 2012.
- [10] L. Xu, A. Wong, and D. A. Clausi, "A novel bayesian spatial-temporal random field model applied to cloud detection from remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4913–4924, Sep. 2017.
- [11] L. Gomez-Chova, G. Camps-Valls, J. Munoz-Mari, and J. Calpe, "Semi-supervised cloud screening with Laplacian SVM," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 1521–1524.
- [12] M. R. Azimi-Sadjadi, M. A. Shaikh, B. Tian, K. E. Eis, and D. Reinke, "Neural network-based cloud detection/classification using textural and spectral features," in *Proc. Geosci. Remote Sens. Symp.*, 1996, vol. 2, pp. 1105–1107.
- [13] O. Slawinski, J. G. Kowalski, and P. C. Cornillon, "A neural network approach to cloud detection in AVHRR images," in *Proc. IJCNN-91-Seattle Int. Joint Conf. Neural Netw.*, 1991, vol. 1, pp. 283–288.
- [14] J. Jang, A. A. Viau, F. Ancil, and E. Bartholomé, "Neural network application for cloud detection in SPOT VEGETATION images," *Int. J. Remote Sens.*, vol. 27, no. 4, pp. 719–736, Feb. 2006.
- [15] R. W. Saunders and G. E. Hunt, "METEOSAT observations of diurnal variation of radiation budget parameters," *Nature*, vol. 283, no. 5748, pp. 645–647, Feb. 1980.
- [16] P. Minnis and E. F. Harrison, "Diurnal variability of regional cloud and clear-sky radiative parameters derived from GOES data. Part I: Analysis method," *J. Climate Appl. Meteor.*, vol. 23, no. 7, pp. 993–1011, Jul. 1984.
- [17] P. Minnis and E. F. Harrison, "Diurnal variability of regional cloud and clear-sky radiative parameters derived from GOES data. Part II: November 1978 cloud distributions," *J. Climate Appl. Meteor.*, vol. 23, no. 7, pp. 1012–1031, Jul. 1984.
- [18] P. Minnis and E. F. Harrison, "Diurnal variability of regional cloud and clear-sky radiative parameters derived from GOES data. Part III: November 1978 radiative parameters," *J. Appl. Meteorol.*, vol. 23, pp. 1032–1051, Jul. 1984.
- [19] R. A. Schiffer and W. B. Rossow, "The international satellite cloud climatology project (ISCCP): The first project of the world climate research programme," *Bull. Amer. Meteor. Soc.*, vol. 64, no. 7, pp. 779–784, Jul. 1983.
- [20] W. B. Rossow *et al.*, "ISCCP cloud algorithm intercomparison," *J. Climate Appl. Meteor.*, vol. 24, no. 9, pp. 877–903, Sep. 1985.
- [21] R. W. Saunders and K. T. Kriebel, "An improved method for detecting clear sky and cloudy radiances from AVHRR data," *Int. J. Remote Sens.*, vol. 9, no. 1, pp. 123–150, Jan. 1988.
- [22] S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, "Discriminating clear sky from clouds with MODIS," *J. Geophysical Res. Atmos.*, vol. 103, no. D24, pp. 32141–32157, 1988.
- [23] S. Ackerman *et al.*, *Discriminating Clear-Sky From Cloud With MODIS. Algorithm Theoretical Basis Document (MOD35)*, vol. 103. Madison, WI, USA: Univ. Wisconsin-Madison, Jan. 2006.
- [24] A. V. D. Vittorio and W. J. Emery, "An automated, dynamic threshold cloud-masking algorithm for daytime AVHRR images over land," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1682–1694, Aug. 2002.
- [25] Y. Liu, J. R. Key, R. A. Frey, S. A. Ackerman, and W. P. Menzel, "Nighttime polar cloud detection with MODIS," *Remote Sens. Environ.*, vol. 92, no. 2, pp. 181–194, Aug. 2004.
- [26] X. Y. Zhuge, X. Zou, and Y. Wang, "A fast cloud detection algorithm applicable to monitoring and nowcasting of daytime cloud systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6111–6119, Nov. 2017.
- [27] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, no. 6, pp. 83–94, 2012.
- [28] L. Sun *et al.*, "A universal dynamic threshold cloud detection algorithm (UDTCDA) supported by a prior surface reflectance database," *J. Geophysical Res. Atmospheres*, vol. 121, no. 12, pp. 7172–7196, 2016.
- [29] Z. Li, H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sens. Environ.*, vol. 191, pp. 342–358, 2017.
- [30] O. Hagolle, M. Huc, D. Villa Pascual, and G. Dedieu, "A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images," *Remote Sens. Environ.*, vol. 114, pp. 1747–1755, Aug. 2010.
- [31] J. K. Roskovensky and K. N. Liou, "Detection of thin cirrus from 1.38 μm /0.65 μm reflectance ratio combined with 8.6–11 μm brightness temperature difference," *Geophysical Res. Lett.*, vol. 30, no. 19, 2003.
- [32] K. D. Hutchison and J. M. Jackson, "Cloud detection over desert regions using the 412 nanometer MODIS channel," *Geophysical Res. Lett.*, vol. 30, no. 23, 2003.
- [33] L. Sun *et al.*, "A cloud detection algorithm-generating method for remote sensing data at visible to short-wave infrared wavelengths," *ISPRS J. Photogrammetry Remote Sens.*, vol. 124, pp. 70–88, Feb. 2017.
- [34] R. N. Clark *et al.*, "USGS digital spectral library splib06a," *Digital Data Series*, vol. 231, Jan. 2007.
- [35] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [36] J. Chen, A. Liao, J. Chen, S. Peng, L. Chen, and H. Zhang, "30-meter global land cover data product- Globe Land30," *Geomatics World*, 2017.
- [37] D. K. Hall, G. A. Riggs, and V. V. Salomonson, "Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data," *Remote Sens. Environ.*, vol. 54, no. 2, pp. 127–140, Nov. 1995.
- [38] V. V. Salomonson and I. Appel, "Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1747–1756, Jul. 2006.
- [39] H. Choi and R. Bindschadler, "Cloud detection in Landsat imagery of ice sheets using shadow matching technique and automatic normalized difference snow index threshold value decision," *Remote Sens. Environ.*, vol. 91, no. 2, pp. 237–242, May 2004.
- [40] G. Chander and B. Markham, "Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2674–2677, Nov. 2003.



Lin Sun received the B.S. degree in geography from Qufu Normal University, Qufu, China, in 1999, the M.S. degree in meteorology from Nanjing University of Information Science and Technology, Nanjing, China, in 2003, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing, China, in 2006.

Since 2006, he has been with the College of Geomatics, Shandong University of Science and Technology, Qingdao, China, where he is currently a Professor. His research interests include cloud detection and aerosol optical depth retrieval from satellite data.



Yueying Zhou was born in Taian, Shandong, China, in 1992. She received the B.S. and M.S. degrees from Shandong University of Science and Technology, Qingdao, China. She is currently working toward the Ph.D. degree in photogrammetry and remote sensing with Wuhan University, Wuhan, China.

Her research interests include thermal infrared remote sensing, cloud and cloud shadow detection, and the quantitative inversion of airborne hyperspectral data.



Jing Wei was born in Zibo, Shandong, China, in 1991. He received the B.E. degree in 2014 and M.E. degree in 2017 from Shandong University of Science and Technology, Qingdao, China. He is currently working toward the Ph.D. degree in global environmental change with Beijing Normal University, Beijing, China.

He served as a Research Assistant with the Chinese University of Hong Kong and Tsinghua University. He focuses on aerosol and cloud remote sensing, as well as aerosol-cloud-radiation interactions.



Tingting Chen received the B.S. degree in geographic information system from the Shandong Jianzhu University, Jinan, China, in 2016. Since 2016, she has been working toward the M.S. degree with the College of Geomatics, Shandong University of Science and Technology, Qingdao, China.

Her research interests include cloud detection and soil moisture retrieval using SAR.



Quan Wang received the B.Eng. degree in remote sensing science and technology from Shandong University of Science and Technology, Qingdao, China, in 2016.

He is a Postgraduate majoring in quantitative remote sensing with the Geomatics College, Shandong University of Science and Technology, Qingdao, China. His research interests include remote sensing image processing and cloud and cloud shadow detection in remote sensing images.



Yulei Chi received the B.S. degree from Shandong University of Science and Technology, Qingdao, China, in 2013. Since 2017, she has been working toward the M.S. degree with Shandong University of Science and Technology, Qingdao, China.

Her research interests include cloud detection and atmospheric gas research.



Xinyan Liu received the B.S. degree in civil engineering from Qindao College, Qingdao University of Technology, Qingdao, China, in 2014. Since 2016, she has been working toward the M.S. degree with the Shandong University of Science and Technology, Qingdao, China.

Her research interests include cloud and cloud shadow detection, aerosol retrieval and dust detection based on remote sensing data.



Meiyang Shu received the B.S. degree in surveying and mapping sciences from Zhengzhou Technology and Business University, Zhengzhou, China, in 2016. Since 2016, she has been working toward the M.S. degree with the Institute of Surveying Science and Engineering, Shandong University of Science and Technology, Qingdao, China.

Her research interests include crop lodging and the quantitative inversion of physical parameters based on remote sensing data.

Wenhua Zhang received the B.S. degree in the science of geography from Qufu Normal University, Rizhao, China, in 2016. Since 2016, she has been working toward the M.S. degree in physical geography from Qufu Normal University, Rizhao, China.

Her research interests include atmospheric remote sensing, meteorology, and climate and land administration.